



UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
ESCOLA DE ENGENHARIA  
DEPARTAMENTO DE ENGENHARIA QUÍMICA  
TRABALHO DE CONCLUSÃO DE CURSO EM ENGENHARIA  
QUÍMICA



# **Detecção de rompimentos de linhas de gases usando técnicas de cartas de controle estatístico**

*Autor: Rickson Peralta Magalhães*

*Orientador: Prof. Dr. Marcelo Farenzena*

*Co-orientador: Me. Fabio Cesar Diehl*

Porto Alegre, novembro de 2021



Autor: Rickson Peralta Magalhães

## Detecção de rompimentos de linhas de gases usando técnicas de cartas de controle estatístico

*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*

Prof. Dr. Marcelo Farenzena  
Co-orientador: 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, Angelo e Clenedi, um agradecimento especial por todo o amor, incentivo e oportunidades que me foram dados por toda a minha vida. Obrigado por estarem comigo sempre que precisei e por me apoiarem nas decisões que tomo. À minha avó, Rosa, por representar toda a bondade que há nesse mundo, sendo o meu exemplo de que podemos sim esperar coisas boas das pessoas. Agradeço também a Laurene e aos nossos dois “filhos cachorros”, Ted e Joaquim. Obrigado por ter sido tão parceira e amiga nessas últimas semanas.

Agradeço a toda a Família Peralta, por todo o amor e toda a união da nossa família. Que nos reencontremos e que aproveitemos os momentos que virão. Agradeço também aos meus amigos mais queridos: Rossi, Júlia, Pedro, Ig, Dilha, Chado e Marcelo. Estou animado para poder revê-los e para botarmos as novidades em dia.

Às pessoas que a UFRGS colocou na minha vida. Ao Futsal Masculino da UFRGS, obrigado por toda a dedicação dentro das quadras, por todos os treinos, por todas as batatas em Esteio e por todas as brincadeiras. Ao pessoal da Engenharia Química (do nosso amado grupo Só Química), obrigado pela ajuda e amizade nesses 6 anos, por todos os encontros de Natal e por todos “eu não fiz ainda”. Também, ao pessoal do videogame, Malvezzi, Braga, Sabba e especialmente ao Dudu, por toda a ajuda no *Python*. E ao pessoal do CEUE, principalmente do setor Acadêmico, obrigado por todas as quintas-feiras de muita entrega e amizade.

Aos servidores, terceiros, técnicos e professores. Vocês são aqueles que movem a nossa querida universidade e deixo registrado aqui a minha gratidão por todos vocês. Não é fácil fazer o todo que vocês fazem com os recursos que são investidos em educação no nosso país. Se a UFRGS é uma das melhores universidades do país, esse mérito é de vocês.

À toda a equipe de Engenharia de Processos da Braskem, em especial à área de polipropileno, por terem me acolhido e me ensinado tanto durante os 18 meses que trabalhei lá. Alyne e Suzana, obrigado de verdade por terem me ensinado e me dado autonomia durante o meu período com vocês. Não tive outra experiência de estágio, mas duvido que eu estaria em melhores mãos.

Ao meu orientador, Marcelo, e ao meu co-orientador, Fabio, por toda a dedicação, vontade e ajuda que vocês me deram, seja em feriado ou em final de semana. Vocês foram de extrema importância para a realização desse trabalho e expressei aqui a minha gratidão aos senhores.

E, por fim, a todos os profissionais da área da saúde do nosso país. Eu não tenho a mínima ideia de como foi difícil para vocês lidar com toda essa pandemia, mas queria que vocês soubessem que vocês são o orgulho do nosso país.

## RESUMO

Vazamentos e rompimentos em tubulações nos segmentos de óleo e gás podem trazer diversos problemas graves, tanto ambientais quanto econômicos. Devido a isso, vários métodos de detecção de rompimentos nessas tubulações são propostos na literatura. Devido a grande parte desses rompimentos ocorrerem no fundo do mar, é difícil encontrar um método de detecção que tenha uma boa relação custo-benefício, pois ou o método de detecção não é satisfatório ou a sua implementação é muito cara. Dessa forma, o estudo em questão possui como objetivo a criação de um algoritmo para detectar rompimentos de fácil implementação e baixo custo. O método escolhido é um método estatístico que se baseia na Análise de Componentes Principais e na distribuição  $T^2$  de Hotelling. No estudo, foram comparadas duas plantas: uma com sintonia rápida e outra com sintonia lenta. O algoritmo proposto para a primeira planta identificou 100% dos casos de rompimentos e apresentou 24% de falsos positivos. A planta de sintonia lenta também identificou 100% dos casos de rompimento, mas apresentou um número menor de falsos positivos, apenas 12%. Os dois algoritmos foram confrontados com a Distância de Mahalanobis e com as Redes Neurais Artificiais. A Distância de Mahalanobis teve um resultado próximo ao da planta de sintonia rápida e as Redes Neurais Artificiais detectaram todos os rompimentos sem a presença de falsos positivos. Como a base de dados da literatura era muito menor que a base do estudo em questão, o método proposto se mostra tão eficiente quanto as Redes Neurais. O método proposto pode ser utilizado como monitoramento dos rompimentos nas linhas de injeção e compressão de gases desde que sejam avaliados os seus limites de confiabilidade e os riscos associados à presença de falsos positivos.

**Palavras-chave:** *Detecção de vazamentos, Tubulações de óleo e gás, Off-shore, PCA, Hotelling.*

## ABSTRACT

Leaks and ruptures in pipelines in the oil and gas segments can cause several serious problems, both environmental and economic. Because of this, several methods of leakage detection in these pipes are proposed in the literature. Because most of these disruptions occur on the seabed, it is difficult to find a good cost-effective detection method, as either the detection method is not satisfactory or its implementation is too expensive. Thus, the goal of this study is to create an algorithm to detect breaks that is easy to implement and low cost. The method chosen is a statistical method that is based on Principal Component Analysis and on Hotelling's  $T^2$  distribution. In the study, two plants were compared: one with better performance controllers and the other with worst ones. The proposed algorithm for the first plant identified 100% of the rupture cases and presented 24% of false positives. The second plant also identified 100% of the rupture cases, but had a lower number of false positives, only 12%. The two algorithms were confronted with the Mahalanobis Distance and the Artificial Neural Networks. The Mahalanobis Distance had a result close to the better controllers plant and the Artificial Neural Networks detected all the ruptures without the presence of false positives. As the literature database was much smaller than the base of this study, the proposed method proves to be as efficient as Neural Networks. The proposed method can be used to monitor disruptions in gas injection and compression lines as long as its reliability limits and the risks associated with the presence of false positives are evaluated.

**Palavras-chave:** *Leakage detection, Oil and gas pipelines, Off-shore, PCA, Hotelling.*

## LISTA DE FIGURAS

Figura 1: Descrição do método NPW. ....	3
Figura 2: Fluxograma do método NPW com sensor FBW. ....	4
Figura 3: Diagrama esquemático do sistema de detecção acústico de vazamentos. ....	5
Figura 4: Sistema BOTDA. ....	6
Figura 5: Passos para a implementação do PCA para as fases <i>off</i> e <i>on-line</i> . ....	9
Figura 6: Diagrama simplificado da planta de interesse ....	11
Figura 7: Matriz de confusão genérica. ....	16
Figura 8: Distribuição t-Student para o PCA utilizando todas as variáveis. ....	18
Figura 9: Distribuição $T^2$ para o PCA utilizando todas as variáveis. ....	18
Figura 10: Distribuição t-Student para o PCA com os maiores <i>scores</i> . ....	19
Figura 11: Distribuição $T^2$ para o PCA com os maiores <i>scores</i> . ....	19
Figura 12: Distribuição t-Student para o PCA com as variáveis de pressão. ....	20
Figura 13: Distribuição $T^2$ para o PCA com as variáveis de pressão. ....	20
Figura 14: Caso 1.3: distribuição t-Student para a planta de sintonia lenta. ....	22
Figura 15: Caso 1.3: distribuição $T^2$ para a planta de sintonia lenta. ....	22
Figura 16: Caso 1.3: distribuição t-Student para a planta de sintonia rápida. ....	22
Figura 17: Caso 1.3: distribuição $T^2$ para a planta de sintonia lenta. ....	23
Figura 18: Caso 7.1: distribuição t-Student para planta de sintonia rápida. ....	24
Figura 19: Carta de controle para a planta de sintonia lenta utilizando dados de operação normal. ....	25
Figura 20: Carta de controle para a planta de sintonia rápida utilizando dados de operação normal. ....	25
Figura 21: Carta de controle para a planta genérica utilizando dados de operação normal. ....	26
Figura 22: Matriz de confusão para a planta com sintonia rápida. ....	26
Figura 23: Matriz de confusão para a planta com sintonia lenta. ....	27
Figura 24: Matriz de confusão para a Distância de Mahalanobis: dados teste. ....	27
Figura 25: Matriz de confusão para as Redes Neurais Artificiais: dados teste. ....	28

## LISTA DE TABELAS

<b>Tabela 1:</b> Sumário dos resultados dos dois métodos para os dois casos.....	10
<b>Tabela 2:</b> Lista de variáveis utilizadas.....	12
<b>Tabela 3:</b> Comparação entre as distribuições para o caso de sintonia lenta.....	23
<b>Tabela 4:</b> Comparação entre as distribuições para o caso de sintonia rápida.....	24
<b>Tabela A.1:</b> Comparação entre as distribuições para o caso de sintonia lenta.....	33
<b>Tabela A.2:</b> Comparação entre as distribuições para o caso de sintonia rápida.....	33



## SUMÁRIO

1	Introdução	1
2	Revisão Bibliográfica	3
2.1	Métodos de detecção de rompimentos em linha	3
2.1.1	Ondas de Pressão Negativa (Negative Pressure Wave)	3
2.1.2	Medições acústicas	5
2.2	Métodos de detecção de rompimentos em sistemas <i>off-shore</i>	6
2.2.1	Sensores de fibra ótica específicos	6
2.2.2	Método Vibroacústico	7
2.2.3	Distância de Mahalanobis e Redes Neurais Artificiais	7
2.3	Análise de Componentes Principais (PCA)	8
2.3.1	Detecção de vazamentos utilizando o PCA	9
3	Metodologia	11
3.1	Descrição do Processo	11
3.2	Origem dos dados	12
3.3	Variáveis utilizadas	12
3.4	Sensibilidade do sistema	13
3.5	Método de detecção a partir do PCA e de cartas de controle estatístico	15
4	Resultados	18
4.1	Comparação entre os PCAs para a operação normal	18
4.1.1	PCA utilizando todas as variáveis	18
4.1.2	PCA utilizando os maiores scores do PCA anterior	19
4.1.3	PCA utilizando somente as variáveis de pressão	20
4.1.4	Escolha do PCA	21
4.2	Comparação das distribuições para os cenários de rompimento	21
4.3	Comparação entre os algoritmos propostos e a Distância de Mahalanobis <sup>24</sup>	24
4.3.1	Comparação entre os algoritmos propostos	25
4.3.2	Comparação com os resultados da literatura	27
5	Conclusões e Trabalhos Futuros	29
	REFERÊNCIAS	31
	APÊNDICE A	33
	APÊNDICE B	35



## 1 Introdução

O pré-sal é de fundamental importância para a economia brasileira e, em fevereiro deste ano, foi registrado o maior percentual de sua participação na produção de petróleo e gás natural (“Produção de petróleo e gás do Pré-sal representa 71,27% do total nacional em fevereiro — Português (Brasil),” [s.d.]). As rochas do pré-sal são reservatórios que se encontram abaixo de uma extensa camada de sal, que abrange o litoral do Estado do Espírito Santo até Santa Catarina, ao longo de mais de 800 km de extensão por até 200 km de largura, em lâmina d’água que varia de 1.500 m a 3.000 m e soterramento entre 3.000 e 4.000 metros (Riccomini et al., 2012).

Apesar do gás natural e do petróleo terem alto valor comercial, não são somente esses os compostos encontrados nos reservatórios do pré-sal. As suas reservas possuem grande razão gás-óleo, onde o gás em questão possui grande quantidade de CO<sub>2</sub>. Nas reservas nacionais, a quantidade de CO<sub>2</sub> varia de 8% a 12% no gás associado (Lima, 2009), entretanto, para ser comercializado no Brasil, o gás natural deve conter no máximo 3% em mol, conforme Resolução ANP nº 16 de 2008.

As tubulações são comumente utilizadas no transporte de substâncias perigosas devido ao seu baixo custo e à sua fácil instalação (Pei et al., 2018). Nas plataformas *off-shore*, o CO<sub>2</sub> sobe na tubulação juntamente com os gases de interesse, passa por processos de separação líquido-gás e gás-gás, depois retornando ao reservatório. Com a reinjeção, há a recuperação de parte da pressão perdida na depleção, mantendo a pressão do reservatório elevada, garantindo uma produtividade elevada e facilitando o escoamento do óleo (“Petrobras - Fatos e Dados - Vamos operar o quarto sistema de separação e reinjeção de gás carbônico no pré-sal,” [s.d.]).

Como dito anteriormente, há uma grande quantidade de CO<sub>2</sub> nas correntes gasosas que chegam nas plataformas *off-shore*. Por influência desta concentração elevada, do ambiente contendo também H<sub>2</sub>O e de tensões residuais no duto, houve um rompimento de um duto flexível de injeção de gás (*riser*) no seu segundo ano de uso, o que representa apenas 10% da sua vida útil. Este incidente foi tipificado como Corrosão sob Tensão por CO<sub>2</sub> (*CO<sub>2</sub> Stress Corrosion Cracking*) e foi comunicado à ANP em janeiro de 2017. (“Alerta de Segurança 001-ANP/SSM: Corrosão sob Tensão por CO<sub>2</sub> (SCC-CO<sub>2</sub>),” [s.d.]).

Nas plataformas *off-shore*, um dos maiores problemas que se tem é o risco do rompimento das tubulações de petróleo, como no caso do Golfo do México, em 2010. O caso em questão se refere à explosão de uma plataforma de petróleo na fase final de perfuração de um poço, que é um processo que requer extremo cuidado, visto que os fluidos do poço podem ser liberados violentamente. O evento deixou 22 trabalhadores feridos e 11 mortos, além de espalhar no mar uma quantidade aproximada de 5 milhões de barris de petróleo, causando incontáveis prejuízos ao ecossistema (“BP pagará 18,7 bi de dólares por vazamento no Golfo do México - ((o))eco,” [s.d.]).

Devido aos prejuízos citados, é de grande interesse das empresas detectar rompimentos nas tubulações precocemente, porém este é um cenário ainda distante. Atualmente, procura-

se identificar a tubulação e o local do vazamento o mais rápido e mais preciso possível, minimizando os impactos econômicos e ambientais dos incidentes.

Em consequência da deficiência do monitoramento de rompimentos que ocorrem no fundo dos mares - que não são solucionados facilmente - este trabalho tem como objetivo propor uma metodologia de fácil implementação para detectar esses rompimentos em tubulações de transporte de gases no menor tempo possível através da Análise de Componentes Principais (PCA) e comparar o método proposto com outros métodos já utilizados para o caso de estudo: a Distância de Mahalanobis e as Redes Neurais Artificiais (Damiani, 2021).

Para atingir o objetivo proposto um algoritmo que utilize a PCA deve ser proposto, juntamente com Cartas de Controle Estatístico, para a detecção dos rompimentos das tubulações dos *risers* a partir de um banco de dados de referência já feito. Diferentes conjuntos de variáveis serão testados, e os resultados confrontados com trabalho da literatura (Damiani, 2021). O presente trabalho de conclusão é segmentado em cinco capítulos. Neste, a motivação e objetivo são apresentados; no Capítulo 2, são mostrados outros métodos de detecção de rompimentos bem como uma explicação sobre no que consiste a PCA; no capítulo 3 são mostrados a breve descrição do processo estudado e o modo como os resultados foram construídos; no capítulo 4 são mostrados os resultados dos algoritmos e a sua comparação com os resultados da literatura; e, por fim, no capítulo 5 são apresentadas a conclusão e sugestões para trabalhos futuros.

## 2 Revisão Bibliográfica

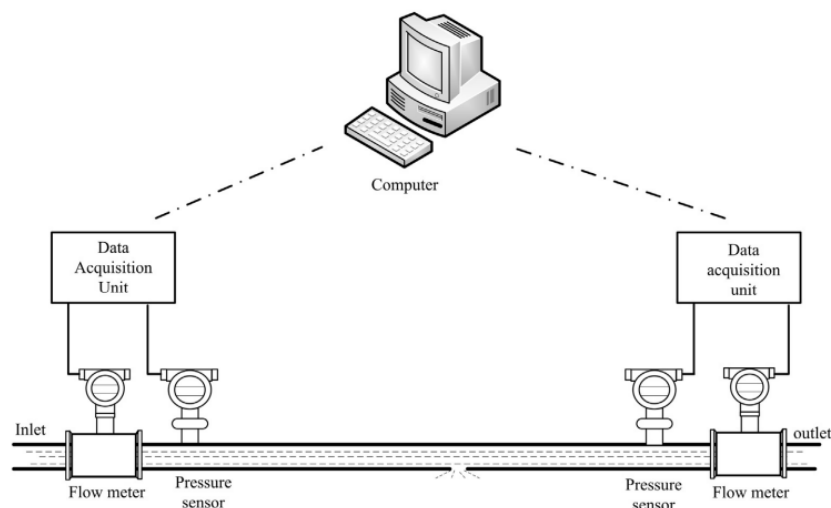
O objetivo deste capítulo é permitir a identificação das informações expostas, fundamentando a pesquisa feita, e apresentar uma relação de obras afins, na qual há a possibilidade do leitor complementar a sua leitura e se aprofundar nos assuntos em questão. Portanto, tratar-se-á aqui em expor diferentes abordagens encontradas na literatura a respeito de métodos de detecção de: rompimentos em linhas, rompimentos e vazamentos em sistemas *off-shore* e de falhas.

### 2.1 Métodos de detecção de rompimentos em linha

#### 2.1.1 Ondas de Pressão Negativa (Negative Pressure Wave)

O método de ondas de pressão negativa, do inglês, *Negative Pressure Wave (NPW)*, é um dos métodos de detecção de rompimentos em tubulações mais utilizados devido a sua simples operação e baixo custo de implementação (Ge et al., 2008). O método baseia-se em sensores de pressão a montante e a jusante do local do rompimento: quando ocorre o vazamento, ocorre uma queda de pressão no sistema e é gerada uma frente de onda que se propaga na tubulação até encontrar os sensores de pressão. A partir de filtros de correlação, essa queda de pressão é diferenciada dos sinais de pressão normal, dos ruídos e de outras manobras normais de operação, como abertura e fechamento de instrumentos. Baseado na velocidade dessa onda de pressão negativa, na diferença do tempo de chegada da onda nos sensores e na distância entre eles, o local do rompimento pode ser encontrado. A descrição do método encontra-se na Figura 1.

Figura 1: Descrição do método NPW.



Fonte: Adaptado de Sun e Chang (2014)

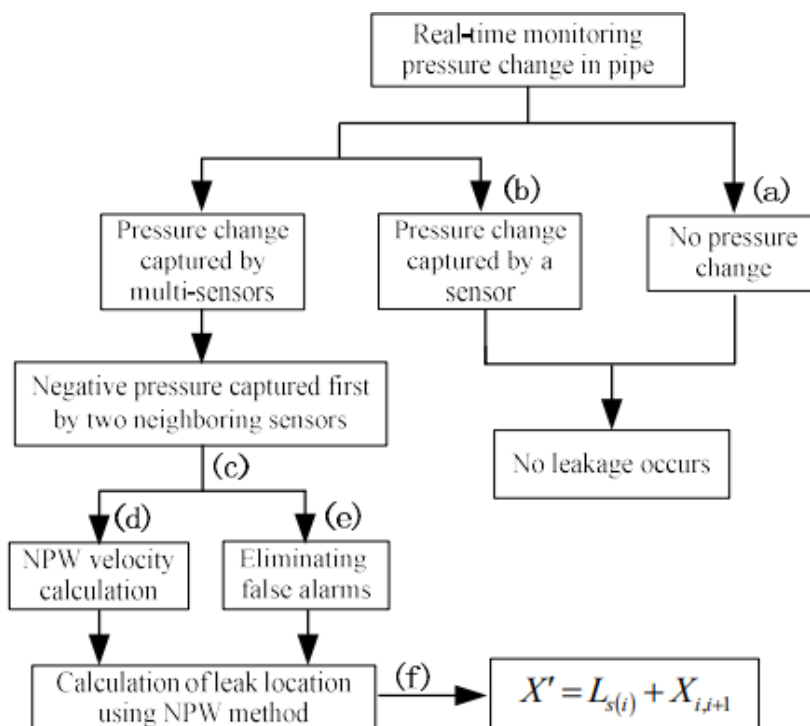
Se o método é vantajoso devido à fácil instalação em sistemas mais antigos, a sua maior desvantagem está na presença de ruídos dos sinais de pressão. Por isso, diversas correlações matemáticas foram feitas para atenuação desses ruídos. Nesse aspecto, Sun e Chang (2014) propuseram um modelo de NPW baseado na combinação dos sinais de pressão e de vazão, com uma transformada de ondeleta (*wavelet transform*) para redução dos ruídos.

Nos dois cenários de validação, um com tubulação de saída ligada a um tanque e outro envolvendo um lento vazamento na tubulação, o método utilizando a combinação dos sinais de pressão e vazão chegou a resultados melhores. Enquanto o método utilizando exclusivamente o NPW não apresentou resultados válidos do local do vazamento, o outro modelo encontrou os dois locais muito próximos aos resultados reais.

Wang et al. (2017) utilizaram um método que combina o NPW com um sensor de fibra denominado *Fiber Bragg Grating (FBG)*. Os sensores foram distribuídos uniformemente na tubulação e os sinais de saída foram conectados ao sistema de controle. O fluxograma do processo é descrito na Figura 2 e explicado a seguir.

Se o sistema de monitoramento não encontrar alguma mudança na pressão ou caso encontre apenas em um sensor, não há vazamento. Porém, se a mudança de pressão é capturada por vários sensores, a onda de pressão negativa é captada primeiro pelos sensores vizinhos e depois disso, a velocidade da onda de pressão é calculada e os falsos alarmes são eliminados, encontrando-se o local do vazamento.

Figura 2: Fluxograma do método NPW com sensor FBW.



Fonte: Adaptado de Wang et al. (2017)

Utilizando dados reais, esse método foi comparado ao NPW tradicional, obtendo resultados mais satisfatórios na localização dos vazamentos para as pressões de 0,32, 0,25 e 0,13 Mpa – erros abaixo de 1,1% para todos os cenários – mostrando que o FBG auxilia na redução dos ruídos.

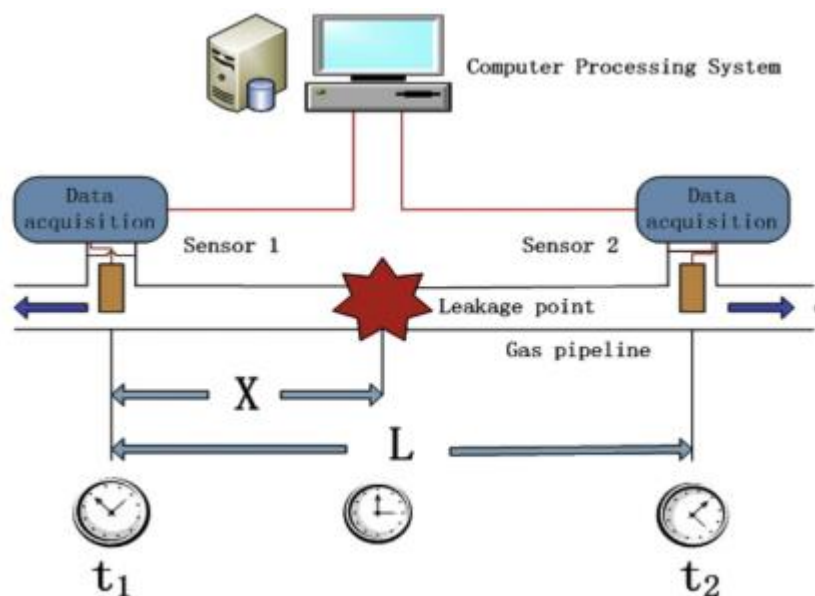
Além das combinações mencionadas, Lu et al. (2016) propuseram uma abordagem para a redução do ruído do NPW através de um método de decomposição do modo empírico melhorado (*Small Noise Reduction – Empirical Mode Decomposition, SNR-EMD*). O SNR-EMD

foi comparado a outros dois métodos de redução de ruídos, o SVD e o WPT, apresentando resultados melhores que ambos. As reduções de ruídos foram satisfatórias, haja vista que o método encontrou o vazamento situado a 2,88 km da estação A quando na realidade o vazamento estava a 3,0 km.

### 2.1.2 Medições acústicas

As medições acústicas possuem muitas vantagens quando comparadas a outros métodos de detecção e localização de vazamentos, tais como alta sensibilidade, alta precisão na localização, baixa taxa de falsos alarmes, menor tempo de testagem e alta adaptabilidade. Os métodos tradicionais de medições acústicas baseiam-se no seguinte princípio: quando um vazamento ocorre, ondas sonoras são propagadas a jusante e a montante do local, sendo capturadas pelos sensores acústicos, como mostrado na Figura 3. Como o tempo de chegada nos sensores é diferente, os vazamentos são detectados e localizados com base nessas diferenças de tempo e de velocidade entre as ondas (Liu et al., 2015).

Figura 3: Diagrama esquemático do sistema de detecção acústico de vazamentos.



Fonte: Liu et al., 2015.

Para obter informações úteis, esse método de correlação cruzada precisa que os sinais acústicos das ondas sejam filtrados já que essas ondas possuem características dispersivas, ou seja, diferentes comprimentos de onda viajam diferentes velocidades de fase. Entretanto, justamente por essa característica dispersiva, é difícil selecionar a largura de banda para um filtro (Gao et al., 2005). Kim e Lee (2009) propuseram a identificação dessas características, dispersivas para um tubo de aço através da análise de frequências de corte usando o método experimental tempo-frequência (laboratório) e o método do elemento de fronteira (BEM). Tanto os resultados de laboratório como os da simulação trouxeram valores de corte próximos aos valores teóricos para o tubo de aço, provando serem dois métodos recomendados para a definição de uma frequência de corte.

Li et al. (2018) realizaram um estudo para a detecção de vazamentos em um sistema de distribuição de água com duas tubulações de ferro unidas por uma junta de encaixe. O experimento foi montado para detectar as falhas na junta. As características acústicas dos sinais de vazamento foram investigadas para os sensores presentes nas duas tubulações. Como a taxa de amostragem dos sinais de emissão acústica são normalmente muito elevados, foi necessário realizar um método de seleção de variáveis para compressão dos dados. Após a seleção das variáveis, treinou-se uma Rede Neural Artificial (ANN). A rede neural apresentou uma acurácia de 97,2%, mostrando que combinação das variáveis de tempo e de frequência reduzem os falsos alarmes significativamente.

## 2.2 Métodos de detecção de rompimentos em sistemas *off-shore*

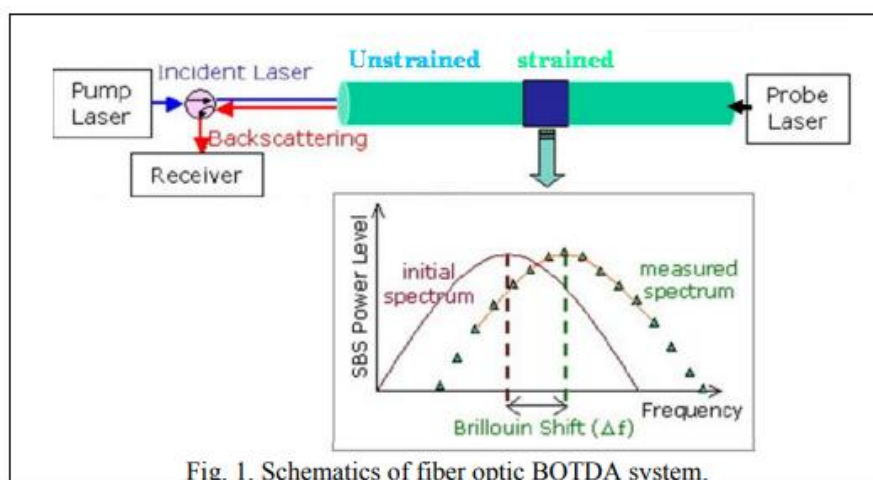
Esta seção tem por objetivo apresentar diferentes métodos de detecção de rompimentos encontrados na literatura, como a utilização de sensores de fibra ótica específicos, o método vibroacústico, a Distância de Mahalanobis e as Redes Neurais Artificiais.

### 2.2.1 Sensores de fibra ótica específicos

Sensores elétricos convencionais não suportariam as condições do fundo do mar e tampouco conseguiriam colher resultados precisos referentes ao local e ao instante do rompimento frente a grande extensão das tubulações. Visando contornar esse problema, Kwon et al. (2009) realizaram um estudo utilizando o sistema de fibra ótica BOTDA (Análise Ótica de Brillouin no Domínio do Tempo).

O sistema é composto por 2 lasers (um operado continuamente e outro de pulso), pelo sensor de fibra ótica BOTDA e de um receptor para recepção do laser retroespelhado, como mostra a Figura 4. Quando a diferença de frequência entre os dois feixes de lasers é igual a frequência de Brillouin da fibra, a amplificação de dispersão de Brillouin irá ocorrer. Se alguma deformação ocorre na fibra, a frequência de Brillouin é alterada conforme a quantidade da deformação. Os autores encontraram um bom resultado utilizando este método e afirmaram que o sistema pode ser aplicado no monitoramento das condições de estruturas de longas distâncias e tamanhos.

Figura 4: Sistema BOTDA.



Fonte: Kwon et al. (2009)



### 2.2.2 Método Vibroacústico

Assim como mostrado na subseção 2.1.2., os métodos acústicos para a detecção de rompimentos também estão presentes em sistemas *off-shore*. Giunta et al. (2016) fizeram um estudo para ver como se dá a performance do sistema frente aos vazamentos de líquido simulados pela abertura das válvulas de descarga nos terminais. A pressão transiente de deslocamento associada, registrada por várias estações de monitoramento, foi comparada com o ruído do ambiente – gerado pelo equipamento de regulagem de fluxo – e com modelos teóricos.

O sistema de detecção visa explorar as ondas de pressões negativas e os princípios de análises estatísticas para monitoramento remoto das tubulações em tempo real. Na prática, uma rede de sensores de pressão e de vibração são instalados na tubulação a uma distância relativa de dezenas de quilômetros. Ondas acústicas e elásticas são produzidas pelas interferências de terceiros e por variações no fluxo (como vazamentos e regulagem de válvulas, por exemplo), se propagando ao longo da tubulação e sendo registradas nas estações de monitoramento. Esse processamento multicanal dos sinais coletados permite a detecção, a localização e a classificação do evento desencadeador. O ponto-chave do sistema de monitoramento é a transmissão em tempo real dos dados vibroacústicos sincronizados das estações remotas localizadas ao longo do conduíte para a central unidade permitindo, assim, o processamento multicanal completo com procedimentos de remoção de ruído e estimativa adaptativa dos parâmetros de propagação de pressão.

Foram feitos dois casos para a distância de detecção do vazamento teórica: um a nível do mar, com pressão externa de 1 bar, e outro a 70 m no fundo do oceano, com pressão externa de 7 bar. Para o primeiro caso foi possível detectar o vazamento a uma distância de 65 km do local da fonte, enquanto para o segundo caso foi possível detectar o vazamento a uma distância de 61,8 km, mostrando a eficácia do modelo.

### 2.2.3 Distância de Mahalanobis e Redes Neurais Artificiais

A Distância de Mahalanobis é um método estatístico que visa medir a proximidade entre duas amostras através da atribuição de um valor à diferença entre os dados de duas amostras distintas (“Reprint of: Mahalanobis, P.C. (1936) ‘On the Generalised Distance in Statistics.’” 2018). Esse método possui diversas finalidades na análise de dados multivariados, dentre elas a investigação da representatividade entre dois conjuntos de dados, a seleção de amostras de calibração a partir de um grande conjunto de dados e a detecção de amostras que se diferenciam das demais (De Maesschalck et al., 2000).

As Redes Neurais Artificiais são métodos, que podem ser usados na detecção de falhas, que utilizam a inteligência artificial e seu uso pode ser interessante para o aumento da confiabilidade de processos industriais (Yang, 2008). Podem ser descritas como um conjunto de modelos matemáticos baseados em uma estrutura biológica (as estruturas neurais dos seres vivos) que, a partir de um conjunto de dados já conhecidos, é treinado para reconhecer padrões. Seus principais elementos são divididos entre Neurônios de Entrada, Intermediários e de Saída. Esses elementos são responsáveis por receberem as informações de entrada para o treinamento, interligar os neurônios de entrada e de saída, e inferir as respostas para cada uma das classes de resposta permitidas (FAUSETT, 1994).

Damiani (2021) buscou uma comparação entre esses métodos para a detecção de rompimentos em linhas de compressão e injeção de gases utilizando como variáveis de interesse a vazão de alimentação de gás e o output do controlador dessa vazão. Esse trabalho foi a motivação para a elaboração desse estudo e os resultados encontrados pelo autor vão ser comparados com os algoritmos propostos na Seção 4.3.2.

O autor realizou, no método da Distância de Mahalanobis, o cálculo para duas distâncias: uma para condições normais de operação e outra para condição de rompimento. Para cada uma das condições, duas amostras eram requeridas: a primeira utilizando os dados de referência do processo e a segunda com dados coletados em uma janela móvel de operação. O método detecta o rompimento quando a distância da condição de rompimento é menor que a distância para a operação normal. Através desse método, o autor chegou a detectar 90% dos rompimentos e apresentou 22,2% de falsos positivos.

O modelo de Rede Neural escolhido pelo autor foi um do tipo de reconhecimento de padrões, criada a partir de um algoritmo do *software* Matlab, com 10 neurônios intermediários do tipo Sigmóide. Foram criadas 100 Redes Neurais e escolhida aquela que teve a melhor capacidade de classificar os dados corretamente. O método apresentou 100% de acurácia e 0% de falsos positivos.

### 2.3 Análise de Componentes Principais (PCA)

A Análise de Componentes Principais, criada por Karl Pearson em 1901, é um método matemático que permite a interpretação de grandes bases de dados. O método possui 4 objetivos principais (Pearson, 1901, apud Charlton et al., 2010):

1. Extrair as informações mais importantes dos dados;
2. Comprimir o tamanho do conjunto de dados pegando apenas as informações mais importantes;
3. Simplificar a descrição do conjunto de dados; e
4. Analisar a estrutura das amostras e das variáveis.

Para atingir esses objetivos, o método computa essas novas variáveis, chamadas de componentes principais, as quais são obtidas pela combinação linear das variáveis originais. O primeiro componente principal obrigatoriamente terá a maior variância do sistema, portanto, este componente irá explicar (ou extrair) a maior parte dos dados originais, o segundo maior componente terá a segunda maior variância, e assim por diante. Os valores das novas variáveis para as amostras originais são chamados de *fator scores*, e podem ser interpretados como as projeções das amostras nos componentes principais. Outra interpretação para esse método é a que ele separa o domínio das variáveis iniciais em dois: um com as principais variações do processo e outro com a variação natural (ruído) do conjunto de dados do processo (Jolliffe e Cadima, 2016).

O PCA determina um conjunto de vetores ortogonais, ordenados pela quantidade de variação contabilizada ao longo da direção de cada vetor de carregamento. Dado um conjunto de  $n$  amostras e de  $m$  variáveis de processo dispostas em uma matriz  $X$ , é necessário realizar

a normalização dos seus dados para a matriz de covariância  $S$ , como mostra a Equação 1 (Pinzón et al., 2017).

$$S = \frac{1}{n-1} X^T X \quad (1)$$

Na Equação 1,  $X$  representa a matriz com os dados originais,  $X^T$  representa a sua matriz transposta e  $n$  o número de amostras. O próximo passo é determinar os autovalores e autovetores da matriz de covariância  $S$ , como mostrado na Equação 2.

$$V^{-1} S V = D \quad (2)$$

Na Equação 2,  $V$  são os autovetores que diagonalizam a matriz de covariância  $S$  e  $D$  é matriz diagonal de  $S$ , que possui os autovalores de  $S$  na sua diagonal principal. Existem métodos de álgebra matricial para encontrar os autovalores e os autovetores desejados, mas graças as linguagens de programação, esses resultados podem ser encontrados em poucos segundos, mesmo para grandes bases de dados. O autovetor associado ao maior autovalor corresponderá ao primeiro componente principal. Então, para reduzir a nossa base de dados original, pegam-se os valores transpostos dos autovetores e os multiplicam pela matriz original.

### 2.3.1 Detecção de vazamentos utilizando o PCA

Pinzón et al. (2017) realizaram um estudo para a detecção de falhas em sistemas de transmissão de gás natural. Embora tradicionalmente a detecção de falhas nesses sistemas dependa do sistema de gerenciamento de alarme e da experiência do operador, devido ao grande tamanho do sistema, um método de detecção baseado em técnicas orientadas por dados é mais adequado. Os autores compararam dois algoritmos na realização desse trabalho: a Análise de Componentes Principais e a Análise de Componentes Principais Dinâmico (DPCA). Os dois algoritmos foram comparados na taxa de falsos positivos, falsos negativos e no tempo de detecção dos vazamentos.

Para o modelo da PCA, os autores mantiveram os passos que foram mencionados na seção 2.3, mas adicionaram um método estatístico para a detecção de falhas para qualquer processo multivariado: o método  $T^2$  de Hotelling. A correlação usada é apresentada na Equação 3.

$$T^2 = x^T P (\Sigma_a)^{-2} P^T x \quad (3)$$

Na Equação 3,  $x$  corresponde a um vetor da amostra,  $\Sigma$  é tal que  $\Sigma \Sigma^T = D$  (matriz diagonal) e  $P$  é o vetor associado à componente principal  $a$ .

A implementação do PCA como um esquema para o monitoramento de falhas compreende duas fases, a *off-line* e a *on-line*, como mostra a Figura 5.

Figura 5: Passos para a implementação do PCA para as fases *off* e *on-line*.

Offline	Online
1. Removing variables.	1. Auto scaling.
2. Auto scaling.	2. $T^2$ Statistic calculation.
3. Removing outliers.	3. Threshold comparison.
4. Covariance matrix calculation.	
5. Spectral decomposition.	
6. Order reduction.	
7. Threshold estimation ( $T_\alpha^2, Q_\alpha$ ).	

Fonte: PINZÓN, H. (2017).

Além dos ruídos aleatórios e das relações estáticas entre as variáveis de processo, muitas das flutuações nos processos químicos podem também ser explicadas pelas inerentes relações dinâmicas entre elas. Por isso, o DPCA inclui no PCA valores temporais das variáveis, como a derivada delas em relação ao tempo, por exemplo. A fim de conseguir implementar essa variação dinâmica, os dados de observação da matriz foram reestruturados para adotar o modelo de Hankel e, então, aplicando todos os passos listados pelo algoritmo da PCA.

Para essa comparação entre os modelos, foram analisados dois casos. O primeiro corresponde a uma ruptura na tubulação causada por uma máquina de escavar, causando uma repentina despressurização ocorreu devido ao grande vazamento de gás para a atmosfera. O outro cenário corresponde ao fechamento de uma válvula de emergência devido a ela ter sido atingida por um raio, deixando o atuador fora de comunicação e, portanto, o sistema não levou em consideração esse status.

A Tabela 1 mostra a comparação entre os métodos evidenciando também a taxa de falsos alarmes. Chega-se à conclusão, portanto, de que o DPCA apresenta resultados melhores que o PCA, tanto na taxa de falsos alarmes quanto na melhor detecção no tempo dos vazamentos. Entretanto, o PCA apresentou resultados bem positivos, mostrando o seu potencial para ser usada em estudos posteriores.

**Tabela 1:** Sumário dos resultados dos dois métodos para os dois casos

Case & Technique		Hotelling's $T^2$ Statistic	
		FAR (%)	Detection time (hr)
Case I	PCA	8.5	469.25
	DPCA	7.2	468.75
Case II	PCA	2.0	64.00
	DPCA	1.5	59.25

Fonte Pinzón et al. (2017).

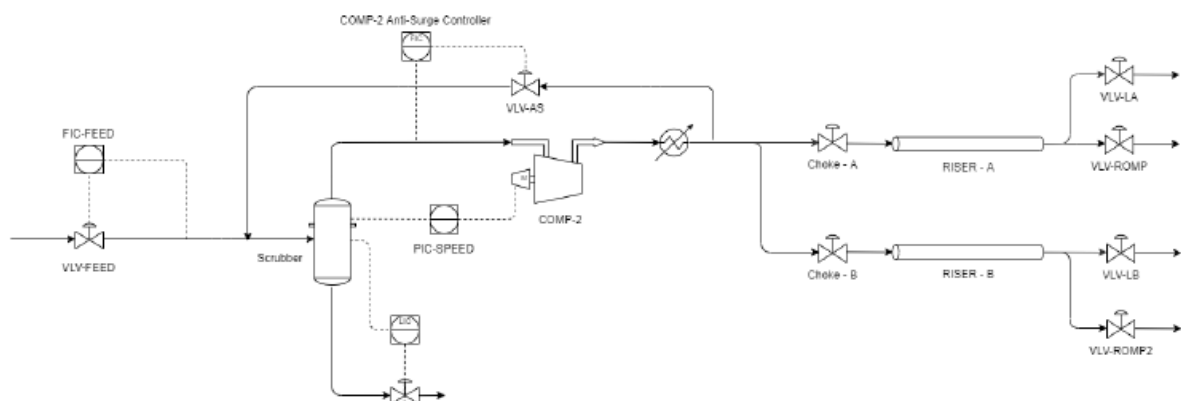
### 3 Metodologia

Neste capítulo será feita a explicação de como foram construídos os resultados que serão apresentados posteriormente, assim como uma contextualização sobre o processo. Esses pontos serão apresentados nesta ordem: descrição do processo, origem dos dados, variáveis utilizadas, sensibilidade do sistema e descrição do método.

#### 3.1 Descrição do Processo

Como foi visto anteriormente, a mistura contendo óleo, água e gás natural chega à plataforma através dos *risers*, onde é destinada para os sistemas de separação de fluidos. A seguir, os gases passam por um mais um sistema de separação - como a filtração por membrana, por exemplo - onde o gás natural é separado de outros gases, como o dióxido de carbono. A planta utilizada nesse estudo é uma planta de compressão e injeção de gases, cujo objetivo é pegar a corrente de gases rica em CO<sub>2</sub> e reinjetá-la nos próprios reservatórios dos quais a mistura é trazida pelos *risers*, reduzindo a emissão dos gases do efeito estufa e auxiliando na manutenção da pressão dos mesmos. O diagrama simplificado da planta é mostrado a seguir na Figura 6.

Figura 6: Diagrama simplificado da planta de interesse



Fonte: DAMIANI, E., 2021

A corrente com os gases a serem reinjetados chega até a unidade de compressão através da válvula de alimentação de gás. A válvula recebe sinal de um controlador de vazão que controla a vazão volumétrica dos gases, a partir de um set-point fixo.

Passando pela válvula de alimentação, os gases passam por um *scrubber*, também conhecido como *demister* ou depurador, que serve para remover gotículas ainda presentes dos processos de separação anteriores. Associados ao *scrubber*, há um controlador de nível, que permite a abertura ou o fechamento da válvula de corrente de líquidos, e um controlador de pressão, que recebe o sinal do vaso, permitindo o controle da velocidade de rotação do compressor.

A corrente de topo do depurador segue para o compressor, seguido de um trocador de calor para redução da sua temperatura. Após o trocador, parte da corrente vai para o reciclo

do compressor. Um controlador recebe sinal da linha a montante do compressor, que indica à válvula de reciclo qual deverá ser a sua abertura. Este controle de vazão opera em malha fechada, evitando a ocorrência de *surge*, um fenômeno que ocorre nos compressores centrífugos, onde a vazão de sucção não é suficiente para manter o fluido estável na carcaça do compressor, causando vibrações excessivas e podendo até causar inversão momentânea no fluxo de gás.

A outra parte da corrente de saída do compressor segue para as linhas de injeção, as quais possuem válvulas *choke*, instaladas antes dos *risers*, que permitem ao operador o controle da abertura para ajuste de vazão e a redução da pressão dos gases. Passando pelas válvulas *choke*, os gases passam pelos *risers* e chegam ao fim da tubulação, onde são reinjetados no reservatório.

### 3.2 Origem dos dados

Para este estudo, foi utilizado um banco de dados de uma simulação da planta real, realizada via Unisim. Os motivos de fazer a simulação foram a falta de um banco de dados reais para os distúrbios de processo avaliados e a falta de um banco de dados reais para os eventos de rompimento, haja vista que são bastante raros.

Os 10 primeiros minutos de cada estudo de caso, todas as variáveis mantiveram-se inalteradas a fim de garantir a condição de estado estacionário antes de serem feitas quaisquer alterações no processo. Assim, os efeitos dos desvios estudados puderam ser isolados. Após os 10 primeiros minutos, realizaram-se alterações no processo com o objetivo de coletar e analisar as variáveis selecionadas para o estudo do sistema. Os resultados das simulações foram salvos em arquivos '.csv', permitindo a manipulação dos dados, passando-os para gráficos e entendendo a dinâmica de cada sistema.

Os cenários de rompimentos foram feitos situando uma válvula no final de cada linha de injeção de gás, assumindo valores de Coeficientes de Vazão ( $C_v$ ) tão elevados – acima de  $10^5$  US GPM – que a abertura dessas válvulas representa um vazamento de grande magnitude como no caso de um rompimento completo da linha de injeção.

Para a simulação, foi considerada a operação da planta com injeção de gás somente pela linha B da Figura 6. Também foram feitas duas simulações para a fase de operação normal: uma para plantas com sintonia “rápida” e outra para plantas com sintonia “lenta”. Dizer que uma planta possui sintonia “rápida” significa que o ajuste dos controladores da planta tem maior performance, rejeitando mais rapidamente os distúrbios.

### 3.3 Variáveis utilizadas

A lista de variáveis utilizadas no estudo é apresentada na Tabela 2.

**Tabela 2:** Lista de variáveis utilizadas.

Variável	Unidade	Descrição
ENTRADA_2 - Pressure	bar	Pressão de entrada no <i>riser</i> .
S019 - Temperature	°C	Temperatura na linha da válvula <i>choke</i> .

S019 - Phase Correlation (Std. Gas Flow-Overall)	STD_m <sup>3</sup> /d	Vazão de gás na linha da válvula <i>choke</i> .
S015 <sup>1</sup> - Pressure	bar	Pressão depois da válvula <i>choke</i> localizada logo após o compressor.
S015 - Pressure	bar	Pressão antes da válvula <i>choke</i> localizada logo após o compressor.
S015 - Phase Correlation (Std. Gas Flow-Overall)	STD_m <sup>3</sup> /d	Vazão de gás na linha de saída do compressor em direção aos <i>risers</i> .
PV-2430-07 - Percentage open	%	<i>Output</i> da válvula <i>choke</i> localizada logo após o compressor.
S011 - Phase Correlation (Std. Gas Flow-Overall)	STD_m <sup>3</sup> /d	Vazão de gás na saída do <i>scrubber</i> .
VLV-104 - Percentage open	%	<i>Output</i> da válvula de reciclo do compressor.
K-100 - Duty	KJ/h	Trabalho realizado pelo compressor.
K-100 - Speed	Rpm	Velocidade de rotação do compressor.
S011 - Pressure	Bar	Pressão de gás na saída do <i>scrubber</i> .

### 3.4 Sensibilidade do sistema

Testes de sensibilidade foram feitos no sistema a fim de entender o seu comportamento diante dos distúrbios ou alterações nas variáveis manipuladas. Além disso, foram simulados rompimentos de diversas magnitudes, sendo estudados cada um dos seus efeitos. Como a metodologia de trabalho proposta é baseada no PCA, é necessário um *dataset* para o cenário normal e um para cada cenário de rompimento

Os distúrbios do sistema para a operação normal têm como objetivo afastar as variáveis utilizadas dos seus respectivos *set-points*. As simulações rodadas foram feitas com distúrbios com desvios positivos e negativos em comparação aos dados de operação normal. Esses distúrbios são importantes para conseguir identificar os motivos dos falsos positivos (dados que o sistema entende como rompimento, mas que na verdade são de operação normal), caso ocorram. A lista completa dos 32 desvios realizados – que representam os dados de operação normal - segue abaixo.

1. Variação na pressão da carga
  - 1.1 Aumentar a pressão 20 bar
  - 1.2 Aumentar a pressão 30 bar
  - 1.3 Descer a pressão -20 bar
  - 1.4 Descer a pressão -30 bar
  - 1.5 Aumentar a pressão 50 bar
  - 1.6 Descer a pressão -50 bar
2. Mudança de pressão no set point de pressão do *scrubber*
  - 2.1 Subir pressão 5 bar
  - 2.2 Subir pressão 15 bar

- 2.3 Reduzir a pressão -5 bar
- 2.4 Reduzir a pressão -15 bar
  
- 3. Mudança de pressão no set point do *anti-surge*
  - 3.1 Subir SP em 5000 m<sup>3</sup>/d
  - 3.2 Subir SP em 10000 m<sup>3</sup>/d
  - 3.3 Descer SP em -5000 m<sup>3</sup>/d
  - 3.4 Descer SP em -10000 m<sup>3</sup>/d
  
- 4. Mudança de pressão no set point do controle PIC (*header*)
  - 4.1 Descer a pressão em -10 bar
  - 4.2 Descer a pressão em -20 bar
  - 4.3 Descer a pressão em -30 bar
  - 4.4 Subir a pressão em 10 bar
  - 4.5 Subir a pressão em 50 bar
  
- 5. Variação de abertura da HCV do poço injetor
  - 5.1 Abrir +10%
  - 5.2 Abrir +20%
  - 5.3 Fechar -30%
  - 5.4 Fechar -10%
  - 5.5 Abrir +10%
  
- 6. Alinhamento do segundo poço injetor
  - 6.1 Pressão de fim da linha (reservatório) de injeção -100 bar
  - 6.2 Pressão de fim da linha (reservatório) de injeção -50 bar
  - 6.3 Pressão de fim da linha (reservatório) de injeção +50 bar
  
- 7. Variação brusca na vazão da carga
  - 7.1 Redução de vazão de 15%
  - 7.2 Aumento de vazão de 15%
  - 7.3 Redução de vazão de 25%
  - 7.4 Aumento de vazão de 25%
  - 7.5 Redução de vazão de 30%

Os cenários de rompimento foram feitos a partir de alterações na abertura das válvulas e para diferentes valores de Cv das válvulas colocadas no final das linhas de injeção, bem como casos de rompimento para diferentes vazões de alimentação de gás na planta. A lista completa dos 18 cenários de rompimento segue abaixo.

- 1. Velocidade do PIC (*header*)
  - 1.1 Kc = 3
  - 1.2 Kc = 0,75
  - 1.3 Kc = 9
  - 1.4 Malha Aberta
  
- 2. Abertura da HCV do poço injetor



- 2.1 Abertura maior:  $z = 30\%$
- 2.2 Abertura menor:  $z = 10\%$
  
- 3. Vazão total da planta
  - 3.1 Vazão menor: redução da pressão da carga de 240 bar para 225 bar
  - 3.2 Vazão maior: aumento da pressão da carga de 240 bar para 270 bar
  
- 4 Alinhamento de segundo poço
  - 4.1 Pressão de fim da linha de injeção A (reservatório) = 450 bar
  - 4.2 Pressão de fim da linha de injeção A (reservatório) = 400 bar
  
- 5 Rompimento estagiado
  - 5.1 Transição do rompimento com válvula de  $\tau = 5$  min
  - 5.2 Transição do rompimento com válvula de  $\tau = 60$  min
  
- 6. Rompimento em distâncias diferentes do flowline
  - 6.1 Distância menor: rompimento em  $-2.867$  m da ANM
  - 6.2 Distância maior: rompimento em  $+2.867$  m da ANM
  
- 7. Teor de CO<sub>2</sub> na carga
  - 7.1 Maior teor:  $y = +30\%$  CO<sub>2</sub>
  - 7.2 Menor teor:  $y = -30\%$  CO<sub>2</sub>
  
- 8. Velocidade do compressor
  - 8.1 Velocidade maior:  $K_c = 0.25$  e  $T_i = 1.00$
  - 8.2 Velocidade muito maior:  $K_c = 0.50$  e  $T_i = 0.50$

### 3.5 Método de detecção a partir do PCA e de cartas de controle estatístico

Um dos objetivos deste trabalho é a elaboração de um algoritmo capaz de detectar os rompimentos que foram simulados anteriormente. Para a construção deste algoritmo, utilizou-se o Google Colaboratory, um ambiente virtual que permite a escrita do código Python no navegador, com nenhuma configuração necessária e fácil compartilhamento.

As bases de dados consistiam em um total de 18 arquivos com cenários de rompimento (um para cada caso) e em 4 arquivos de operação normal (dois para cada tipo de planta). Com o intuito de simplificar a metodologia, o presente trabalho descreverá nesta seção como o algoritmo foi montado para somente um tipo de planta e para um cenário de rompimento, mas trará na seção de resultados o comportamento das duas plantas e de todos os cenários de rompimento.

O primeiro passo na implementação dos algoritmos foi de abrir os arquivos no Google Colaboratory para ter uma primeira visualização dos dados. Nos arquivos, cada coluna correspondia a uma das doze variáveis da seção 3.2., tendo uma coluna a mais referente ao tempo decorrido. Depois da visualização prévia, uniram-se as duas bases de dados em uma única base de dados, como mostra o código exemplo do Apêndice B, para facilitar a execução de comandos posteriores.

Como há muitos dados de diferentes escalas ( $^{\circ}\text{C}$ , bar, %, STD  $\text{m}^3/\text{d}$ ,  $\text{kJ}/\text{h}$  e rpm), foi feita a normalização dos dados para evitar a sensibilidade do PCA à diferença de escala entre as dimensões, ou seja, para que o método atribua uma mesma unidade a todas as variáveis, evitando com que variáveis tenham pesos maiores que outras somente pelo fato de suas unidades serem diferentes. A normalização foi feita utilizando o *MinMaxScaler* (Apêndice B).

O passo a seguir foi fazer o PCA para os dados já normalizados, pegando a variância que cada componente principal representa do todo. Nesse estudo, foram feitas 3 versões do PCA para posterior comparação. O primeiro PCA foi utilizando todas as colunas do banco de dados; o segundo utilizando somente as colunas com maior *score*, ou seja, faz-se o PCA com todas as colunas, analisa-se quais tem menor impacto, retiram-se essas colunas da matriz normalizada (antes do PCA) e refaz-se o PCA; e o terceiro utilizando apenas as colunas das pressões, pois julga-se serem as variáveis que mais refletem quando há algum rompimento. Em todos os 3 casos, utilizou-se somente o primeiro componente principal para a realização das cartas de controle por ser mais fácil de trabalhar com um sistema univariado do que com um sistema multivariado.

Após fazer o PCA, utilizou-se um dos arquivos referentes aos rompimentos. Então, utilizou-se a mesma normalização (com os mesmos limites de máximo e de mínimo que os dados de operação normal) para esse conjunto de dados – através do comando *transform*, mostrado no Apêndice B - a fim de deixar os dados na mesma escala. Como visto anteriormente, para se obter o primeiro componente principal do PCA, tem-se que pegar a matriz original (normalizada) e multiplicá-la pelo primeiro autovetor. Pegou-se, então, a matriz com os dados de rompimento normalizados e multiplicou-se ela pelo primeiro autovetor (para cada um dos três casos do PCA), garantindo que todos os dados estão na mesma escala.

Por último, foram comparadas duas cartas de controle estatístico: uma utilizando a distribuição t de Student e a outra, a distribuição  $T^2$  de Hotelling. Primeiramente, foram feitos os limites superior e inferior para a distribuição com os dados de operação normal e depois pegou-se estes valores para usar nos limites das distribuições com os dados de rompimentos. Assim, sempre que a distribuição normal sair dos limites, tem-se a presença de um falso positivo, e se nenhum dos dados de rompimento sair dos limites, tem-se um falso negativo.

Para comparação do algoritmo com os outros modelos foram comparados os resultados das suas respectivas matrizes de confusão. A matriz de confusão é uma métrica voltada para modelos de classificação e tem como objetivo de calcular a quantidade de falsos negativos e falsos positivos, fornecendo a acurácia e a sensibilidade do sistema. A Figura 7 representa um modelo genérico para uma matriz de confusão. Para o sistema do presente estudo: os verdadeiros positivos (TP) representam cenários de rompimentos nas quais o modelo acertou a detecção; os verdadeiros negativos (TN) representam os distúrbios de operação normal para os quais o modelo não detectou rompimento; os falsos positivos (FP) representam os distúrbios de operação normal para os quais o modelo detectou rompimento; e os falsos negativos (FN) representam a não detecção de rompimento quando houve um cenário de rompimento.

Figura 7: Matriz de confusão genérica.

Valor previsto \ Valor real	Positivo	Negativo
Positivo	Verdadeiro Positivo (TP)	Falso Positivo (FP)
Negativo	Falso Negativo (FN)	Verdadeiro negativo (TN)

Analisando a matriz genérica, observa-se que os acertos do modelo estão presentes na sua diagonal principal, enquanto os seus erros estão na outra diagonal. A matriz pode ser apresentada com a quantidade de TPs, TNs, FPs e FNs ou com a porcentagem de cada termo. As matrizes para cada caso apresentarão tanto a quantidade dos termos como a porcentagem deles, melhorando a visualização, a compreensão e a comparação dos resultados.

## 4 Resultados

Este capítulo apresentará os resultados da implementação dos algoritmos para detecção de rompimentos. Os gráficos para os PCAs feitos para cada uma das três combinações das variáveis utilizadas serão analisados, bem como as comparações das distribuições para cada caso de rompimento e para os dados de operação normal. Por fim, serão apresentadas matrizes de confusão comparando o melhor tipo de distribuição para cada uma das plantas, fazendo também uma comparação com os resultados se utilizando a distância de Mahalanobis e Redes Neurais Artificiais.

### 4.1 Comparação entre os PCAs para a operação normal

Como dito anteriormente, foram feitos três PCAs diferentes com o intuito de obter o menor valor possível de falsos positivos possível. A seguir, o resultado de cada será apresentado, bem como a escolha de qual o melhor PCA.

#### 4.1.1 PCA utilizando todas as variáveis

Para este caso, foram utilizadas as 12 variáveis listadas na Seção 3.3. Fazendo o PCA, para este caso chega em um primeiro componente com valor de 0,6546, o que significa que analisando somente esse componente, encontra-se 65,46% da variabilidade do sistema. As Figuras 8 e 9 mostram as cartas de controle estatístico para o componente utilizando o banco de dados de operação normal para as plantas de sintonia lenta, já que os resultados são análogos aos da planta de sintonia rápida.

Figura 8: Distribuição t-Student para o PCA utilizando todas as variáveis.

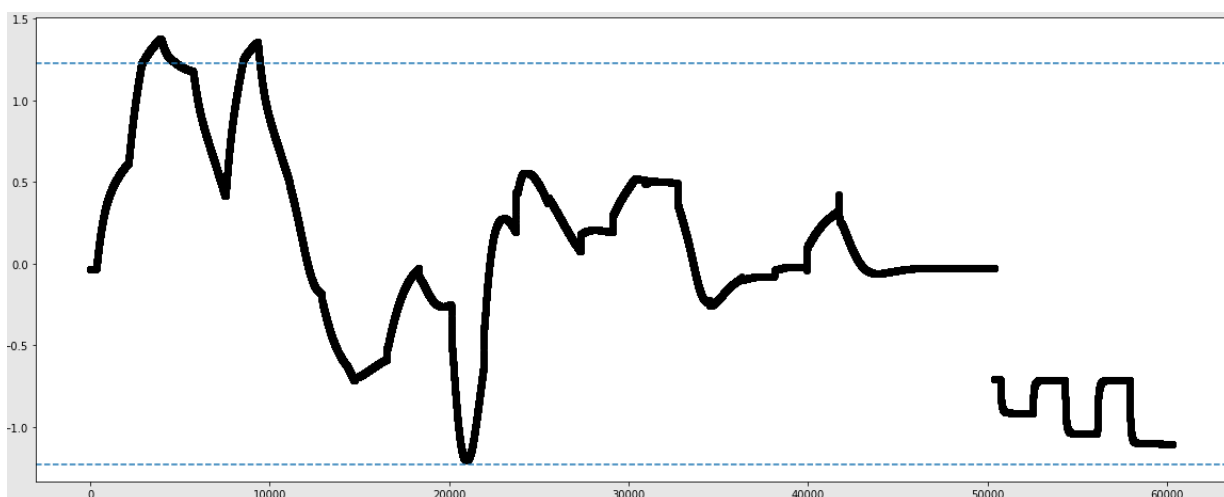
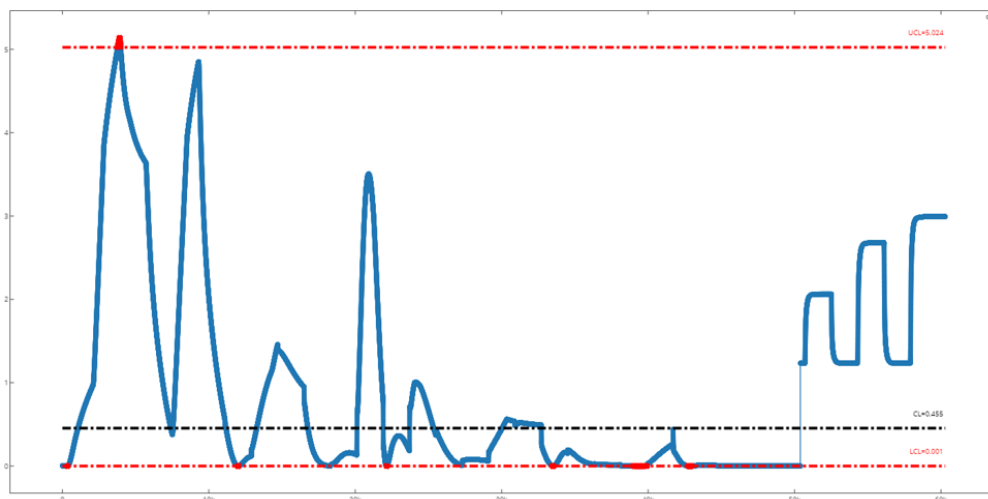


Figura 9: Distribuição  $T^2$  para o PCA utilizando todas as variáveis.



#### 4.1.2 PCA utilizando os maiores scores do PCA anterior

Neste segundo caso, foram utilizadas somente as variáveis que possuem as maiores variabilidades do sistema. Essa escolha foi possível analisando os valores das componentes do autovetor que multiplicou a matriz original dos dados normalizados para que se formassem os valores do primeiro componente principal.

Cada um dos valores das componentes do autovetor representa uma das doze variáveis utilizadas. O valor do *score* de corte foi definido em 0,2 em módulo. O passo a seguir foi retirar as colunas com pouca variabilidade antes de fazer a normalização, refazendo a mesma e o PCA. As variáveis utilizadas para o novo PCA foram: 'ENTRADA\_2 – Pressure', 'S019 - Temperature', 'S019 - Phase Correlation (Std. Gas Flow-Overall)', 'S015' - Pressure', 'S015 - Phase Correlation (Std. Gas Flow-Overall)', 'PV-2430-07 - Percentage open' e 'VLV-104 - Percentage open'. O primeiro componente desse novo PCA resultou em um valor que representa 84,45% da variabilidade dos dados e suas cartas de controle são mostradas nas Figuras 10 e 11.

Figura 10: Distribuição t-Student para o PCA com os maiores *scores*.

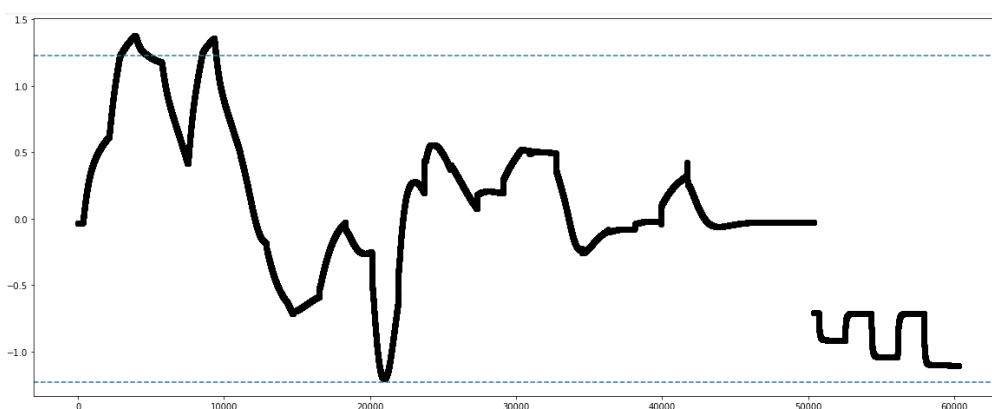
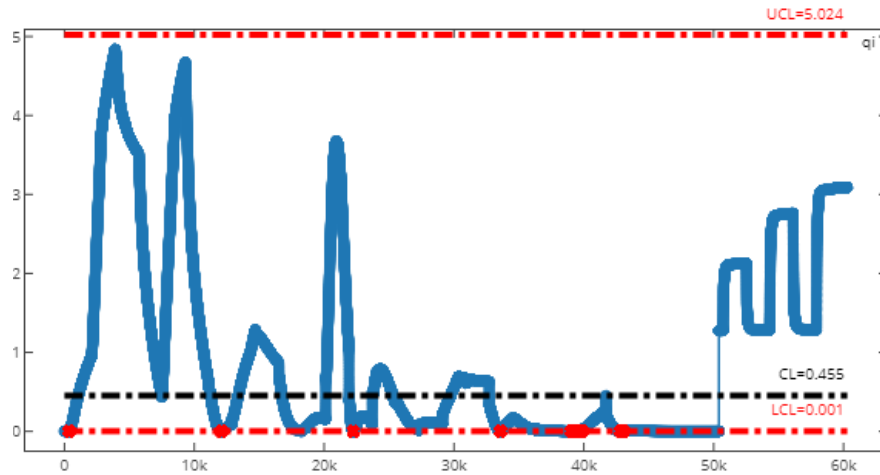


Figura 11: Distribuição  $T^2$  para o PCA com os maiores *scores*.



#### 4.1.3 PCA utilizando somente as variáveis de pressão

Para esta análise, foram considerados somente os efeitos das variáveis de pressão do sistema, pois há a hipótese de que elas seriam as variáveis mais afetadas no caso de ocorrer um rompimento na planta. O primeiro componente principal dessa análise resultou numa variabilidade de 70,32% dos dados e suas cartas de controle são mostradas nas Figuras 12 e 13.

Figura 12: Distribuição t-Student para o PCA com as variáveis de pressão.

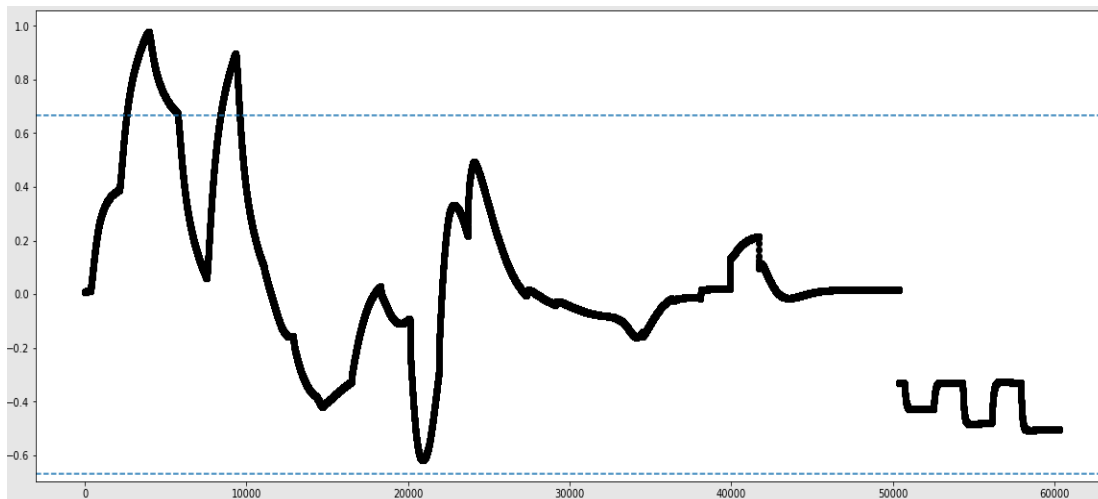
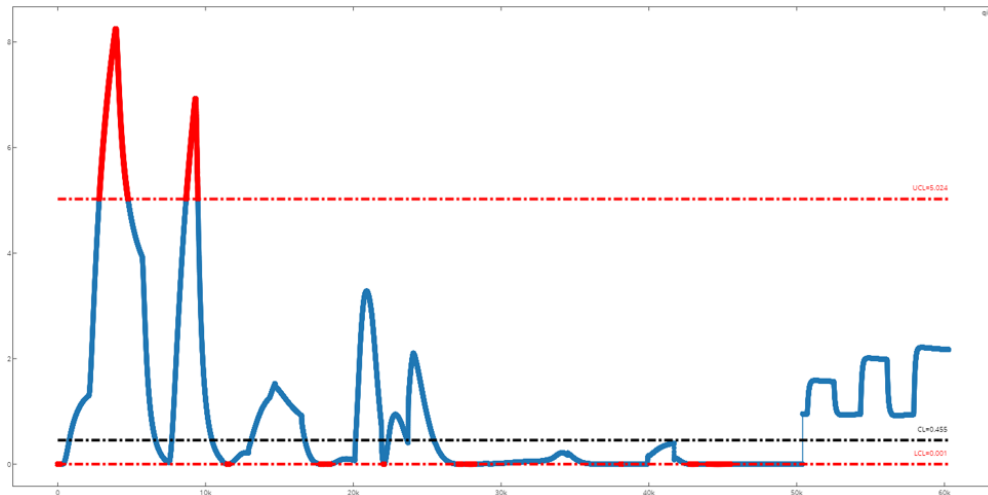


Figura 13: Distribuição  $T^2$  para o PCA com as variáveis de pressão.



#### 4.1.4 Escolha do PCA

Como mostrados nas seções anteriores, pela carta de controle utilizando a distribuição t-Student todos os PCAs possuem 2 falsos positivos, valores que ultrapassam os limites da carta de controle, representado pelas linhas tracejadas em azul para a distribuição t-Student e em vermelho para a  $T^2$ . Porém, na carta de controle da distribuição  $T^2$ , o primeiro PCA possui 7 falsos positivos, o segundo PCA possui 6 falsos negativos e o último PCA possui 9 falsos positivos. Por isso, resolveu-se utilizar o segundo PCA para comparar as duas plantas e as duas distribuições nos casos de rompimento.

#### 4.2 Comparação das distribuições para os cenários de rompimento

Esta seção tem por objetivo comparar as duas distribuições, para cada tipo de planta, nos cenários de rompimento, analisando qual distribuição consegue detectar os rompimentos mais rapidamente e com a menor quantidade possível de falsos positivos – valores que ultrapassam os limites das cartas, sendo considerados rompimentos, mas que na verdade não são rompimentos – e de falsos negativos – valores que não ultrapassam os limites das cartas, sendo considerados operação normal, mas que na realidade representam rompimentos – garantindo ao algoritmo uma maior confiabilidade.

Para as duas distribuições foi usado o valor do nível de significância,  $\alpha$ , igual a 0,05. O nível de significância representa a possibilidade de rejeitar a hipótese nula quando é verdadeira, ou seja, indica a probabilidade de concluir que existe uma diferença entre os resultados do estudo e a hipótese nula, quando na verdade não há diferença. Esse valor de 0,05 foi escolhido porque valores maiores aumentam o número de falsos positivos e valores menores aumentam o número de falsos negativos.

Para cada um dos 18 casos de rompimento, foram comparadas as duas distribuições para a planta de sintonia rápida e para a de sintonia lenta, resultando em 36 comparações entre as distribuições. Em todos os cenários, os rompimentos foram iniciados com  $t=300s$ , ou seja, na amostra de número 300.

Para facilitar a visualização, pegou-se um exemplo: o item 1.3 dos casos de rompimento, referentes a um Kc igual a 9 na válvula que simula o rompimento no final da linha de injeção, ou seja, um rompimento de grandes proporções. Na planta de sintonia lenta, os tempos

encontrados para o rompimento foram de 495s e de 462s para a distribuição t-Student e  $T^2$ , respectivamente, como mostram as Figuras 14 e 15. Na planta de sintonia rápida, os tempos encontrados para o rompimento foram de 484s e de 459s para a distribuição t-Student e  $T^2$ , respectivamente, como mostram as Figuras 16 e 17.

Figura 14: Caso 1.3: distribuição t-Student para a planta de sintonia lenta.

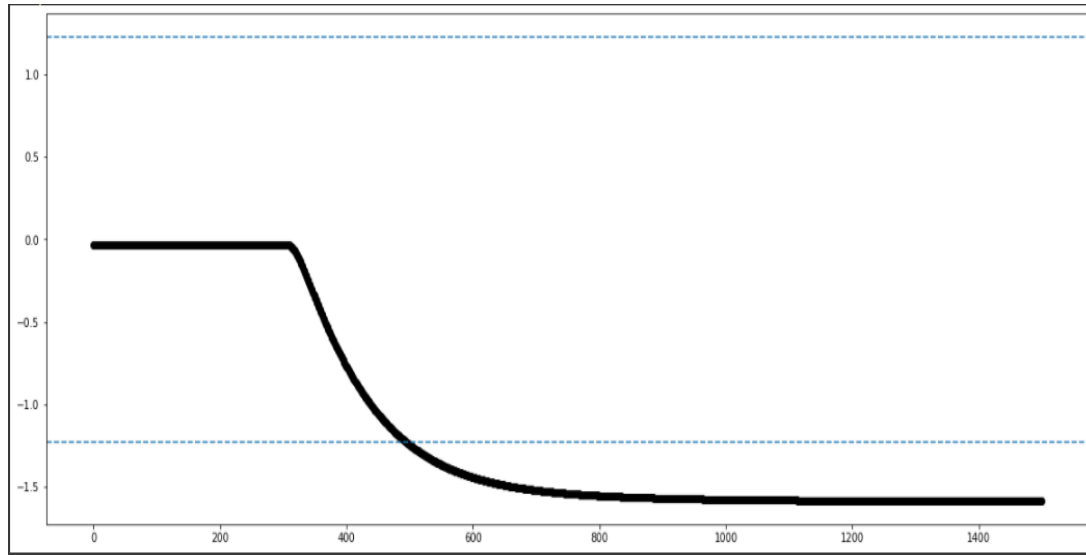


Figura 15: Caso 1.3: distribuição  $T^2$  para a planta de sintonia lenta.

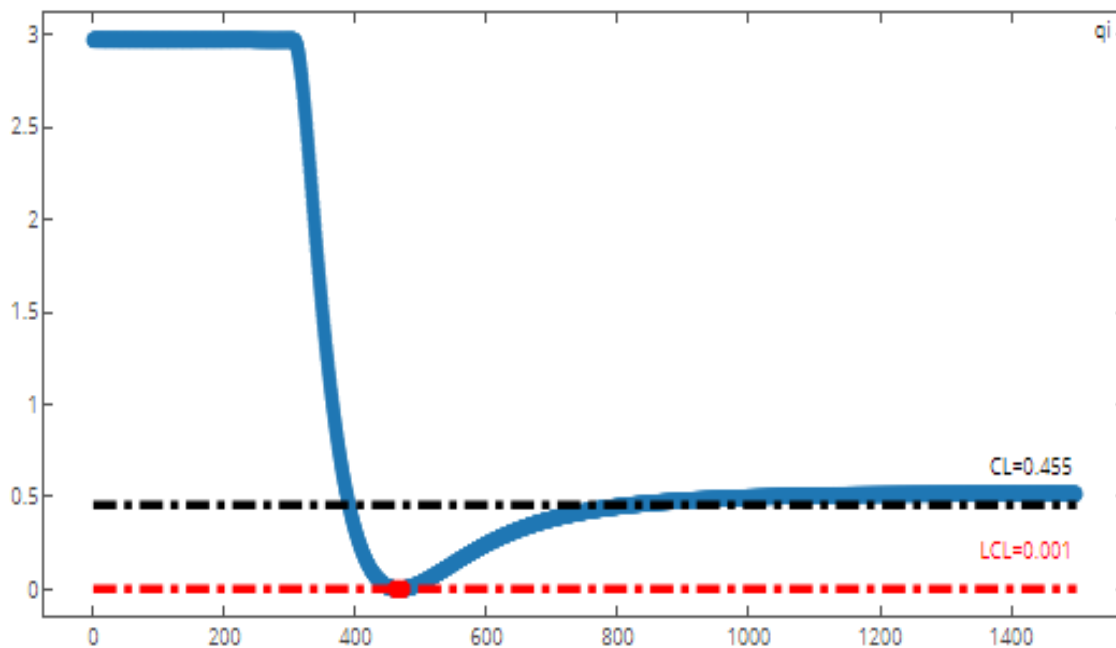


Figura 16: Caso 1.3: distribuição t-Student para a planta de sintonia rápida



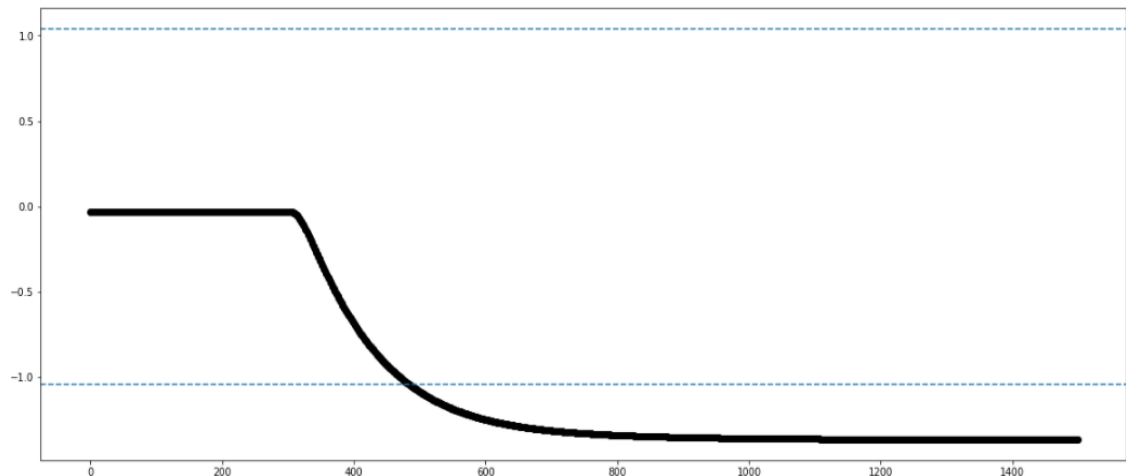
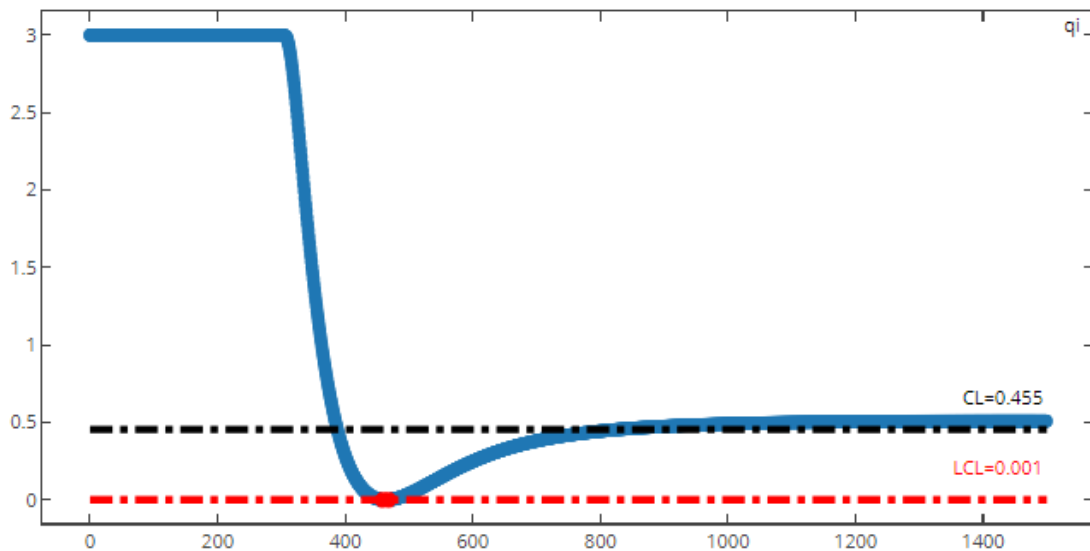


Figura 17: Caso 1.3: distribuição  $T^2$  para a planta de sintonia rápida.



A seguir, são mostradas as Tabelas 3 e 4, que representam a comparação entre as distribuições para as plantas de sintonia lenta e rápida, respectivamente. Percebe-se que, comparando as duas sintonias para uma mesma distribuição, a detecção do rompimento ocorre mais rapidamente na sintonia rápida, o que é esperado já que os controladores da planta são mais rápidos, assim como a dinâmica da planta.

**Tabela 3:** Comparação entre as distribuições para o caso de sintonia lenta.

	t-Student	Hotelling $T^2$
Média	4,5min	3,02min
Falsos Negativos	5	0

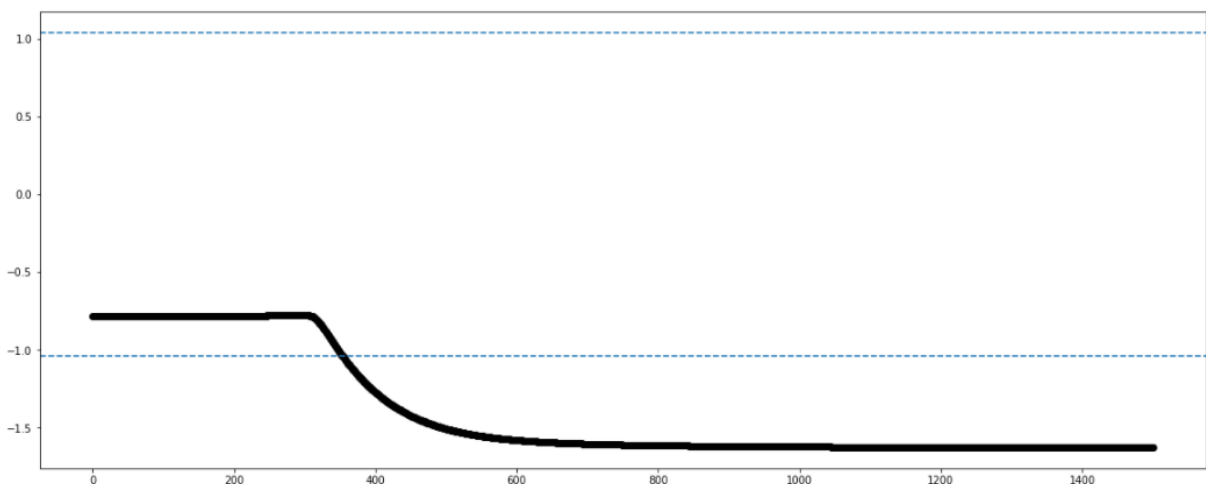
**Tabela 4:** Comparação entre as distribuições para o caso de sintonia rápida.

	t-Student	Hotelling T <sup>2</sup>
Média	4,03min	2,93min
Falsos Negativos	5	0

Comparando as duas distribuições, nota-se que em alguns casos, a distribuição t-Student apresenta falsos negativos. Isso significa que os rompimentos não foram detectados para os casos 1.4, 3.2, 5.1, 5.2 e 7.2. Na distribuição T<sup>2</sup>, nenhum falso negativo foi detectado, e, portanto, a carta de controle dessa distribuição apresentou-se como sendo a mais indicada na detecção dos rompimentos. As tabelas completas com o tempo de detecção para cada caso encontram-se no Apêndice A.

Apesar de apresentar falsos negativos, a distribuição normal apresentou alguns resultados muito favoráveis, identificando a presença de um rompimento menos de um minuto após o seu início, como no caso 7.1. (Figura 18), quando há aumento no teor de CO<sub>2</sub>. Infelizmente, não é possível ter uma distribuição para cada cenário de rompimento, pois não há como prever qual o cenário de rompimento que ocorrerá na planta.

Figura 18: Caso 7.1: distribuição t-Student para planta de sintonia rápida.



### 4.3 Comparação entre os algoritmos propostos e a Distância de Mahalanobis

Nesta seção, serão comparadas as plantas de sintonias diferentes com o objetivo de identificar se os algoritmos propostos para cada uma das plantas apresentam resultados satisfatórios, isto é, com baixos intervalos entre os rompimentos e a sua detecção, e com o menor número possível de falsos negativos e de falsos positivos.

### 4.3.1 Comparação entre os algoritmos propostos

Os gráficos dos casos de rompimento para as duas plantas já foram mostrados na seção anterior, faltando apenas mostrar as cartas de controle estatístico para os dados de operação normal. A carta da planta de sintonia lenta, mostrada na Figura 19, apresenta 6 falsos positivos, ou seja, 6 dos 32 distúrbios realizados apresentaram falsa indicação. Para a planta de sintonia rápida, mostrada na Figura 20, o número de falsos positivos dobra, apresentando uma falsa indicação para 12 dos 32 distúrbios.

Figura 19: Carta de controle para a planta de sintonia lenta utilizando dados de operação normal.

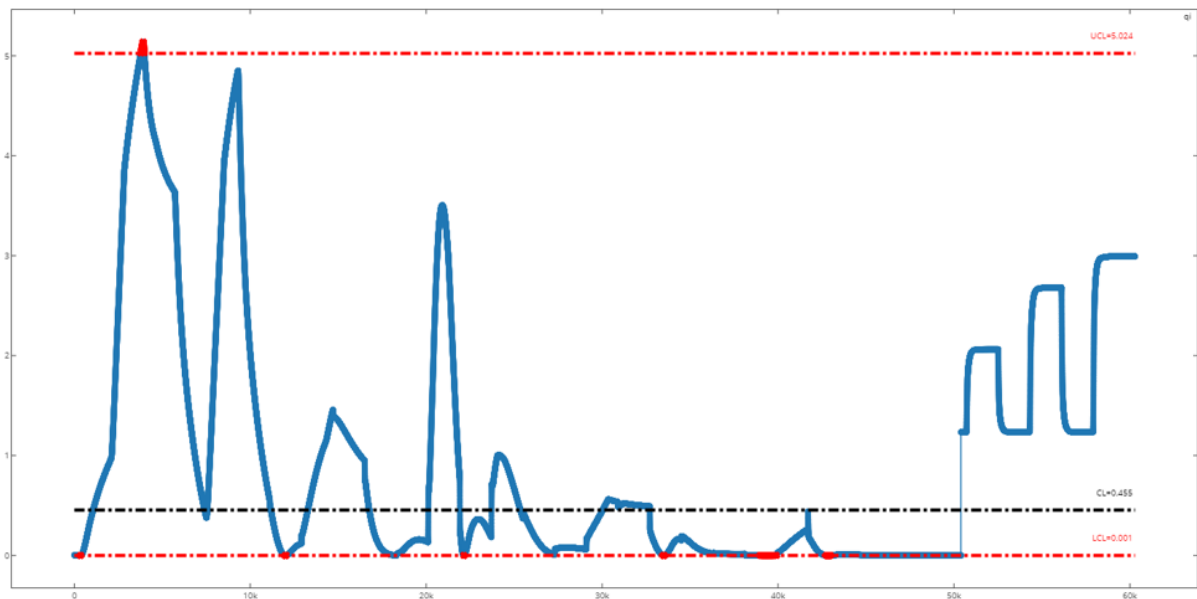
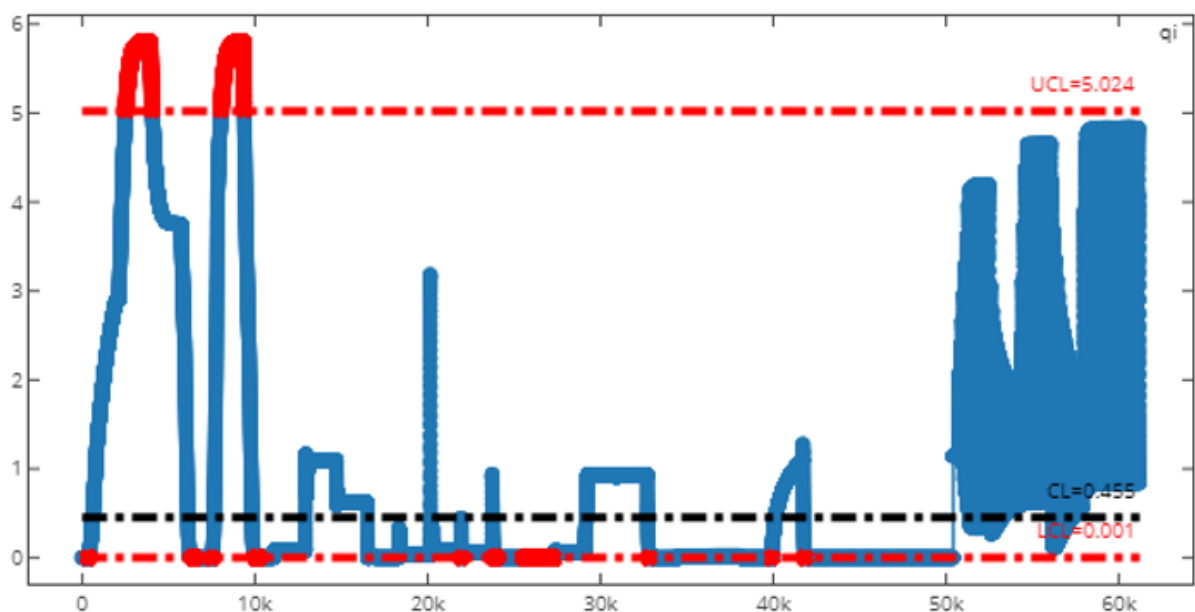
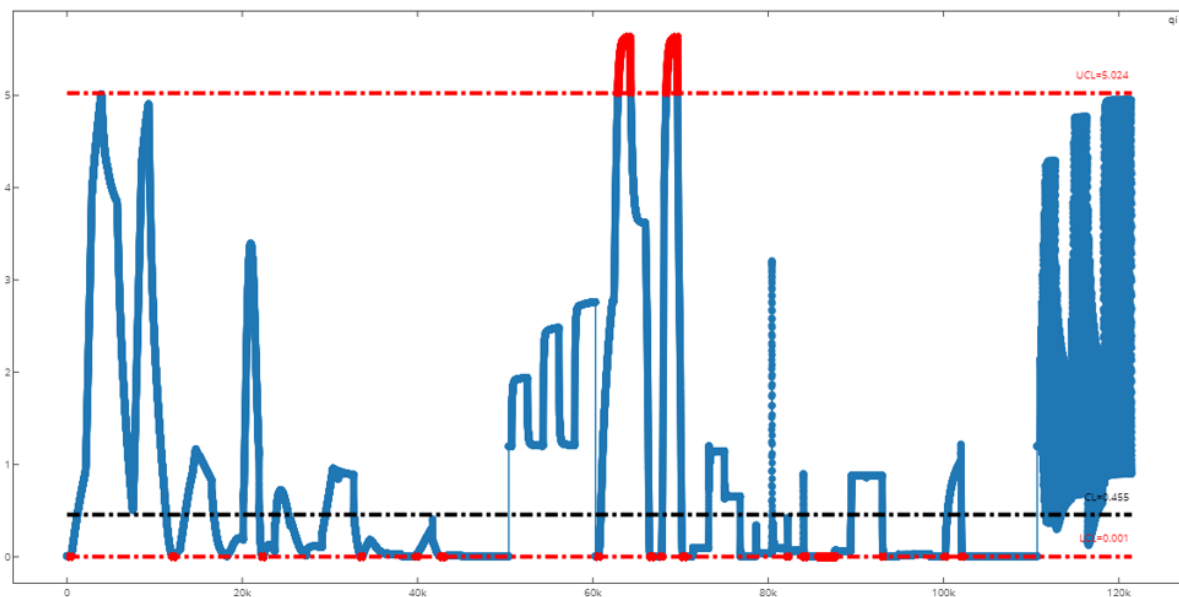


Figura 20: Carta de controle para a planta de sintonia rápida utilizando dados de operação normal.



Também foi feita uma carta de controle juntando as bases de dados de operação normal dos dois modelos de planta para um só (Figura 21), para ver se um algoritmo só consegue apresentar resultados satisfatórios para qualquer tipo de planta sem que haja uma diferenciação prévia. A carta de controle apresentou um resultado intermediário aos resultados das duas plantas, apresentando 18 falsos positivos para 64 distúrbios que estavam presentes. Como não apresentou um resultado melhor para os dados de sintonia rápida, o algoritmo juntando as duas bases de dados não se mostrou eficiente e, portanto, seus resultados não serão comparados posteriormente neste estudo.

Figura 21: Carta de controle para a planta genérica utilizando dados de operação normal.



Para melhor análise dos resultados, foram feitas 3 matrizes de confusão e além de compará-las entre si, elas foram comparadas às matrizes de confusão feita para a detecção de rompimentos utilizando a Distância de Mahalanobis e as Redes Neurais Artificiais.

Como pode-se ver pela matriz de confusão para a planta de sintonia rápida, Figura 22, o algoritmo conseguiu prever corretamente 38 dos 50 cenários estudados (18 cenários de rompimento e 20 de operação normal), o que corresponde a uma acurácia de 76%. Os casos que representam falsos positivos são os seguintes: 1.1, 1.2, 1.3, 1.4, 1.5, 3.1, 3.2, 3.3, 3.4, 4.3, 5.2 e 5.3. Os distúrbios do caso 1, onde há variação na pressão da carga, e do caso 3, onde há mudança do *set-point* do controlador *anti-surge*, mostram que há uma grande deficiência do modelo para esses casos. Em contrapartida, o modelo não apresentou falsos negativos, conseguindo prever todos os 18 casos de rompimento.

Figura 22: Matriz de confusão para a planta com sintonia rápida

		REAL	
		Operação Normal	Rompimento
PREDIÇÃO	Operação Normal	20 40%	0 0%
	Rompimento	12 24%	18 36%

O algoritmo aplicado para planta de sintonia lenta, por outro lado, conseguiu prever corretamente 44 dos 50 cenários, como pode-se ver pela sua matriz de confusão representada na Figura 23. A acurácia para essa planta é de 88% e, assim como na planta anterior, não são encontrados falsos negativos, predizendo todos os cenários de rompimento. Os falsos negativos são encontrados nos casos 1.1, 1.5, 3.1, 4.3, 5.2 e 5.3. Esse modelo apresenta os mesmos falsos negativos que o modelo anterior, mas proporcionalmente, a variação na abertura da HCV do poço injetor representa uma maior deficiência do modelo do que a mudança do *set-point* do controlador *anti-surge*.

Figura 23: Matriz de confusão para a planta com sintonia lenta.

		REAL	
		Operação Normal	Rompimento
PREDIÇÃO	Operação Normal	26 52%	0 0%
	Rompimento	6 12%	18 36%

#### 4.3.2 Comparação com os resultados da literatura

No estudo mencionado na Seção 2.2.3, foram utilizados 33 cenários de operação normal e 10 de rompimentos para a Distância de Mahalanobis e a sua matriz de confusão de teste é apresentada na Figura 24. Já para as Redes Neurais Artificiais foram utilizados 29 cenários de operação normal e 9 de rompimentos. Sua matriz de confusão de teste é mostrada na Figura 25.

Figura 24: Matriz de confusão para a Distância de Mahalanobis: dados teste.

		PREDIÇÃO	
		Operação normal	Rompimento
REAL	Operação normal	<b>3</b> 50,0%	<b>0</b> 0,0%
	Rompimento	<b>0</b> 0,0%	<b>3</b> 50,0%

Fonte: DAMIANI, E., 2021.

Figura 25: Matriz de confusão para as Redes Neurais Artificiais: dados teste.

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

Fonte: DAMIANI, E., 2021.

O algoritmo proposto para a planta de sintonia lenta apresentou o segundo melhor resultado dentre os 4 métodos analisados, com acurácia de 88%. O método das Redes Neurais apresentou o melhor resultado – acurácia de 100% - porém tanto o seu banco de dados quanto o da Distância de Mahalanobis são muito menores quando comparados ao banco de dados desse estudo. Já as acurácias para os modelos da Distância de Mahalanobis da planta de sintonia rápida ficaram bem próximas, 77,8% e 76%, respectivamente.

## 5 Conclusões e Trabalhos Futuros

Através desse trabalho, foi possível alcançar o objetivo proposto de desenvolver um algoritmo capaz de detectar os cenários de rompimento da planta utilizando técnicas de cartas de controle estatístico. O método identificou 100% dos rompimentos tanto para os dois modelos de planta, apresentando um tempo médio de detecção de 9,81min e de 9,76 min para as plantas de sintonias lenta e rápida, respectivamente. Esses valores são mais altos devido aos cenários de rompimento estagiado, cuja detecção do modelo foi muito tardia. Tirando esses dois cenários de rompimento estagiado, os tempos médios de detecção são de 3,38min e de 3,36min para as plantas de sintonias lenta e rápida, respectivamente.

Por mais que o algoritmo possua tempos baixos na detecção de rompimentos, o modelo para a planta de sintonia rápida apresentou 24% de falsos positivos, principalmente para os cenários onde há variação na pressão da carga, variação na abertura da HCV do poço injetor e mudanças no *set-point* do controlador *anti-surge*. Para a planta de sintonia lenta, os resultados foram mais satisfatórios, apresentando apenas 12% de falsos positivos, destacando-se nos cenários onde há variações na pressão da carga e variações na abertura da HCV do poço injetor.

O método para a planta de sintonia lenta apresentou um resultado melhor quando comparada à detecção através da Distância de Mahalanobis, tanto na acurácia – 88% contra 77,8% - quanto no tempo de detecção, excluindo-se os cenários de rompimento estagiado – 3,38min contra 4,30min. Já o algoritmo para a planta de sintonia rápida apresentou uma acurácia semelhante – 76% contra 77,8% - mas apresentou um tempo de detecção melhor – 3,36min contra 4,30min.

É preciso de uma análise mais detalhada para estudar a viabilidade de se implementar qualquer um dos algoritmos. Deve-se analisar a incidência dos cenários de falsos positivos durante a operação normal da planta, o que resultará em uma nova eficácia dos algoritmos. A partir dessa nova eficácia, deverá se analisar os riscos de implementação dos algoritmos e se a presença de falsos positivos não trará um prejuízo econômico à planta, haja vista que há um custo elevado para cada parada ou redução de carga.

Em termos de viabilidade de implementação, o algoritmo pode, facilmente, ser inserido no sistema de segurança da planta de compressão e injeção de gases. Há a possibilidade de deixar a carta de controle para a distribuição  $T^2$  de Hotelling em uma tela do SDCD (Sistema Digital de Controle Distribuído) com alarmes de atenção quando os dados estiverem se aproximando dos limites da carta e com alarmes de rompimento quando os dados ultrapassarem esses limites.

Monitorar cenários de rompimentos no fundo dos oceanos ainda é uma dificuldade muito grande nos sistemas *off-shore*, pela dificuldade de encontrar uma boa relação de custo-benefício: ou a tecnologia a ser empregada é muito alta ou esta não traz detecções confiáveis. Por isso, metodologias estatísticas para a identificação de padrões representam a alternativa mais viável para a resolução desse problema. Como recomendações de trabalhos futuros, destaca-se aqui a utilização de um método de cartas de controle estatístico utilizando o PCA para mais de um componente principal e de outro utilizando as Redes Neurais Artificiais,

sempre monitorando se os custos computacionais das técnicas para o monitoramento online do processo não serão muito elevados,



## REFERÊNCIAS

- Alerta de Segurança 001-ANP/SSM: Corrosão sob Tensão por CO<sub>2</sub> (SCC-CO<sub>2</sub>). [s.d.].
- BP pagará 18,7 bi de dólares por vazamento no Golfo do México - ((o))eco.**  
Disponível em: <<https://oeco.org.br/noticias/29214-bp-pagara-18-7-bi-de-dolares-por-vazamento-no-golfo-do-mexico/>>. Acesso em: 16 nov. 2021.
- CHARLTON, M.; BRUNSDON, C.; DEMŠAR, U.; et al. *Principal Components Analysis: from Global to Local*. 2010.
- DAMIANI, E. **Uso da Distância de Mahalanobis para a Detecção de Rompimentos em Linhas de Injeção de Gás**. Porto Alegre: [s.n.].
- DE MAESSCHALCK, R.; JOUAN-RIMBAUD, D.; MASSART, D. L. The Mahalanobis distance. **Chemometrics and Intelligent Laboratory Systems**, v. 50, n. 1, p. 1–18, 4 jan. 2000.
- FAUSETT, L. **FAUSETT, L. Fundamentals of Neural Networks: Architecture, Algorithms, and Applications**. 1994. [s.l: s.n.].
- GAO, Y.; BRENNAN, M. J.; JOSEPH, P. F.; et al. On the selection of acoustic/vibration sensors for leak detection in plastic water pipes. **Journal of Sound and Vibration**, v. 283, n. 3–5, p. 927–941, 20 maio 2005.
- GE, C.; WANG, G.; YE, H. Analysis of the smallest detectable leakage flow rate of negative pressure wave-based leak detection systems for liquid pipelines. **Computers and Chemical Engineering**, v. 32, n. 8, p. 1669–1680, 22 ago. 2008.
- GIUNTA, G.; MORREA, S.; GABBASSOV, R.; et al. Performance of vibroacoustic technology for pipeline leak detection. **Proceedings of the International Conference on Offshore Mechanics and Arctic Engineering - OMAE**, v. 5, 2016.
- JOLLIFE, I. T.; CADIMA, J. Principal component analysis: a review and recent developments. **Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences**, v. 374, n. 2065, 13 abr. 2016.
- KIM, M. S.; LEE, S. K. Detection of leak acoustic signal in buried gas pipe based on the time–frequency analysis. **Journal of Loss Prevention in the Process Industries**, v. 22, n. 6, p. 990–994, 1 nov. 2009.
- KWON, I.-B.; JIN, G.; SEO, D.-C.; et al. Feasibility study for monitoring of off-shore pipelines using BOTDA system. **Ocean Sensing and Monitoring**, v. 7317, p. 731700, 1 maio 2009.
- LI, S.; SONG, Y.; ZHOU, G. Leak detection of water distribution pipeline subject to failure of socket joint based on acoustic emission and pattern recognition. **undefined**, v. 115, p. 39–44, 1 fev. 2018.
- LIMA, P. C. R. O pré-sal e o aquecimento global. 2009.
- LIU, C. W.; LI, Y. X.; YAN, Y. K.; et al. A new leak location method based on leakage acoustic waves for oil and gas pipelines. **Journal of Loss Prevention in the Process Industries**, v. 35, p. 236–246, 1 maio 2015.

LU, W.; LIANG, W.; ZHANG, L.; et al. Title: A Novel Noise Reduction Method Applied in Negative Pressure Wave for Pipeline Leak Localization. 2016.

PEARSON, K. n the Criterion That a Given System of Deviations from the Probable in the Case of a Correlated System of Variables Is Such That It Can Be Reasonably Supposed to Have Arisen from Random Sampling. **Philosophical Magazine Series 5 X**, p. 559–572, 1901.

PEI, J.; WANG, G.; LUO, S.; et al. Societal risk acceptance criteria for pressure pipelines in China. **Safety Science**, v. 109, p. 20–26, 1 nov. 2018.

**Petrobras - Fatos e Dados - Vamos operar o quarto sistema de separação e reinjeção de gás carbônico no pré-sal.** Disponível em:

<<https://petrobras.com.br/fatos-e-dados/vamos-operar-o-quarto-sistema-de-separacao-e-reinjecao-de-gas-carbonico-no-pre-sal.htm>>. Acesso em: 16 nov. 2021.

PINZÓN, H.; AUDIVET, C.; ALEXANDER, J.; et al. A COMPARATIVE ASSESSMENT ON STATIC AND DYNAMIC PCA FOR FAULT DETECTION IN NATURAL GAS TRANSMISSION SYSTEMS. 2017.

**Produção de petróleo e gás do Pré-sal representa 71,27% do total nacional em fevereiro — Português (Brasil).** Disponível em: <[https://www.gov.br/anp/pt-br/canais\\_atendimento/imprensa/noticias-comunicados/producao-de-petroleo-e-gas-do-pre-sal-representa-73-do-total-nacional-em-fevereiro](https://www.gov.br/anp/pt-br/canais_atendimento/imprensa/noticias-comunicados/producao-de-petroleo-e-gas-do-pre-sal-representa-73-do-total-nacional-em-fevereiro)>. Acesso em: 13 out. 2021.

Reprint of: Mahalanobis, P.C. (1936) “On the Generalised Distance in Statistics.” **Sankhya A**, v. 80, n. S1, p. 1–7, dez. 2018.

RICCOMINI, C.; SANT’ANNA, L. G.; TASSINARI, C. C. G. Pré-sal: geologia e exploração. **Revista USP**, v. 0, n. 95, p. 33–42, 30 nov. 2012.

SUN, L.; CHANG, N. Integrated-signal-based leak location method for liquid pipelines. **Journal of Loss Prevention in the Process Industries**, v. 32, p. 311–318, 1 nov. 2014.

WANG, J.; ZHAO, L.; LIU, T.; et al. Novel Negative Pressure Wave-Based Pipeline Leak Detection System Using Fiber Bragg Grating-Based Pressure Sensors. **Journal of Lightwave Technology**, v. 35, n. 16, p. 3366–3373, 15 ago. 2017.

YANG, K. T. Artificial Neural Networks (ANNs): A new paradigm for thermal science and engineering. **Journal of Heat Transfer**, v. 130, n. 9, set. 2008.

## APÊNDICE A

**Tabela A.1:** Comparação entre as distribuições para o caso de sintonia lenta.

Sintonia Lenta		
Casos	t-Student	Hotelling T <sup>2</sup>
Caso 1.1	524s	481s
Caso 1.2	707s	540s
Caso 1.3	495s	462s
Caso 1.4	-	470s
Caso 2.1	490s	478s
Caso 2.2	1201s	522s
Caso 3.1	388s	461s
Caso 3.2	-	670s
Caso 4.1	465s	479s
Caso 4.2	452s	498s
Caso 5.1	-	1286s
Caso 5.2	-	6274s
Caso 6.1	385s	402s
Caso 6.2	943s	569s
Caso 7.1	364s	436s
Caso 7.2	-	685s
Caso 8.1	506s	471s
Caso 8.2	492s	460s

**Tabela A.2:** Comparação entre as distribuições para o caso de sintonia rápida.

Sintonia Rápida		
Casos	t-Student	Hotelling T <sup>2</sup>
Caso 1.1	512s	476s
Caso 1.2	658s	526s
Caso 1.3	484s	459s
Caso 1.4	-	455s
Caso 2.1	478s	473s
Caso 2.2	1033s	511s
Caso 3.1	388s	458s
Caso 3.2	-	635s
Caso 4.1	453s	473s
Caso 4.2	440s	488s
Caso 5.1	-	1280s
Caso 5.2	-	6250s
Caso 6.1	379s	399s

Caso 6.2	885s	561s
Caso 7.1	354s	434s
Caso 7.2	-	646s
Caso 8.1	494s	468s
Caso 8.2	482s	458s

## APÊNDICE B

Código para a Detecção de Rompimentos a partir do PCA e de Cartas de Controle Estatístico:

```
#  
  
Título: Detecção de Rompimentos a partir do PCA e de Cartas de Controle  
Estatístico  
  
Autor: Rickson Peralta Magalhães  
  
Data: 14/11/2021  
  
Código abaixo:  
  
#  
  
#importando bibliotecas necessárias  
  
import glob  
import os  
  
import pandas as pd  
import numpy as np  
from sklearn.decomposition import PCA  
import matplotlib.pyplot as plt  
  
#importando arquivos do Drive  
  
from google.colab import drive  
drive.mount('/content/drive')  
  
#Botando o diretório de interesse  
  
os.chdir("/content/drive/MyDrive/TCC/AS_Rompidos.zip (Unzipped Files)")  
  
# Atribuindo nomes aos DataFrames  
# headers pra ler direito  
  
df1 = pd.read_csv("CSP_Norm_Len_Medidas.csv",header=[8,9])  
df2 = pd.read_csv("CSP_Norm_Len_DisturbiosBruscos.csv",header=[8,9])  
  
# .head para ver as 5 primeiras linhas do df  
  
df1.head()  
df2.head()
```

```
#iloc para retirar últimas 3 colunas

df1 = df1.iloc[:, :-3]
df2 = df2.iloc[:, :-1]

#.concat para unir os 2 df's (acaba a última linha do 1° df e passa para o 2° df)

df_len = pd.concat([df1, df2], ignore_index=True)

# .isna().mean() para calcular o percentual de dados incompletos

df_len.isna().mean()

# Unindo o nome das duas primeiras linhas em uma só

df_len.columns = [" ".join(cols) for cols in df1.columns.values]

# .shape para ver o tamanho da matriz

df_len.shape

# Colocando o tempo como índice para a coluna de tempo não influenciar no cálculo do PCA

df_len.set_index("Time [seconds]", inplace=True)

# .drop para deletar as colunas especificadas

df_len = df_len.drop(['S015 - Pressure [bar]', 'S011 - Phase Correlation (Std. Gas Flow-Overall) [STD_m3/d]', 'K-100 - Duty [kJ/h]', 'K-100 - Speed [rpm]', 'S011 - Pressure [bar]'], axis=1)

# normalizando as colunas

from sklearn import preprocessing
scaler = preprocessing.MinMaxScaler()
names = df_len.columns
d = scaler.fit_transform(df_len)
scaled_df = pd.DataFrame(d, columns=names)
scaled_df.head(60290)

# Fazendo o PCA e printando a contribuição de cada componente de 1 até n, com n sendo o número de colunas da matriz de interesse

n_comps = np.arange(1, scaled_df.shape[1]+1)
log = {}
pca_log = {}
```

```
for n in n_comps:
    pca = PCA(n_components=n)
    pca.fit(scaled_df)
    pca_log[f"pca_{n}"] = pca
    log[f"n_{n}"] = pca.explained_variance_ratio_
    print(pca.explained_variance_ratio_)

Q = pca_log["pca_1"]

len_pca = Q.transform(scaled_df)

# Abrindo o dataframe de um cenário de rompimento
romp = pd.read_csv("CSP_Case_1.1.csv", header=[8,9])

romp.columns = ["".join(cols) for cols in romp.columns.values]

romp = romp.iloc[:, :-3]

# Colocando o tempo como índice
romp.set_index("Time[seconds]", inplace=True)

romp = romp.drop(['S015 - Pressure[bar]', 'S011 - Phase Correlation (Std
. Gas Flow-Overall) [STD_m3/d]', 'K-100 - Duty[kJ/h]', 'K-
100 - Speed[rpm]', 'S011 - Pressure[bar]'], axis=1)

romp_norm = scaler.transform(romp)

romp_norm = pd.DataFrame(romp_norm, columns=names)

romp_test = Q.transform(romp_norm)

# Importando as bibliotecas dos modelos estatísticos

import scipy.stats as sct
import statsmodels.api as sm

dist = sct.norm()

confidence_level = 0.95
degrees_freedom = len_pca.size - 1
sample_mean = np.mean(len_pca)
sample_standard_error = len_pca.std()

confidence_interval = sct.t.interval(confidence_level, degrees_freedom,
sample_mean, sample_standard_error)
```

```
LB, UB = confidence_interval
[LB,UB]

# Fazendo as cartas de controle t-Student para o rompimento

plt.figure(figsize=(20,8))
plt.plot(len_pca, "ko")
plt.axhline(LB, ls="--")
plt.axhline(UB, ls="--")

romp_data = pd.DataFrame(data = ydata, columns = ['principal component
1'])

# Instalando o Hotelling

!pip install hotelling

from hotelling.plots import control_chart, control_stats, univariate_co
ntrol_chart
from hotelling.stats import hotelling_t2

# Fazendo as cartas de controle T2 para o rompimento e para a operação
normal

ab = control_chart(romp_test, phase=1, alpha=0.05, x_bar=None, s=None,
legend_right=True, interactive=True, width=10, cusum=False, template='n
one', marker='o', ooc_marker='x', random_state=42, limit=70000, no_disp
lay=False)

ax = control_chart(len_pca, phase=1, alpha=0.05, x_bar=None, s=None, le
gend_right=True, interactive=True, width=10, cusum=False, template='non
e', marker='o', ooc_marker='x', random_state=42, limit=70000, no_displa
y=False)
```