

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
CAMPUS DOIS VIZINHOS
CURSO DE ESPECIALIZAÇÃO EM CIÊNCIA DE DADOS

GABRIEL STANKEVIX SOARES

**APLICAÇÃO DE APRENDIZADO DE MÁQUINA PARA
PROTEÇÃO CATÓDICA EM GASODUTO**

TRABALHO DE CONCLUSÃO DE CURSO DE ESPECIALIZAÇÃO

DOIS VIZINHOS
2022

GABRIEL STANKEVIX SOARES

APLICAÇÃO DE APRENDIZADO DE MÁQUINA PARA PROTEÇÃO CATÓDICA EM GASODUTO

Trabalho de Conclusão de Curso de Especialização apresentado ao Curso de Especialização em Ciência de Dados da Universidade Tecnológica Federal do Paraná, como requisito para a obtenção do título de Especialista em Ciência de Dados.

Orientador: Prof. Dr. Yuri Kaszubowski Lopes

DOIS VIZINHOS
2022



4.0 Internacional

Esta licença permite remixe, adaptação e criação a partir do trabalho, mesmo para fins comerciais, desde que sejam atribuídos créditos ao(s) autor(es) e que licenciem as novas criações sob termos idênticos. Conteúdos elaborados por terceiros, citados e referenciados nesta obra não são cobertos pela licença.

GABRIEL STANKEVIX SOARES

**APLICAÇÃO DE APRENDIZADO DE MÁQUINA PARA
PROTEÇÃO CATÓDICA EM GASODUTO**

Trabalho de Conclusão de Curso de Especialização apresentado ao Curso de Especialização em Ciência de Dados da Universidade Tecnológica Federal do Paraná, como requisito para a obtenção do título de Especialista em Ciência de Dados.

Data de Aprovação: 08/janeiro/2022

Yuri Kaszubowski Lopes
Doutorado
Universidade do Estado de Santa Catarina

Rafael Alves Paes de Oliveira
Doutorado
Universidade Tecnológica Federal do Paraná - Câmpus Dois Vizinhos

Rodolfo Adamshuk Silva
Doutorado
Universidade Tecnológica Federal do Paraná - Câmpus Dois Vizinhos

DOIS VIZINHOS
2022

AGRADECIMENTOS

A todos que participaram, direta ou indiretamente do desenvolvimento deste trabalho de pesquisa, enriquecendo o meu processo de aprendizado acadêmico e profissional.

RESUMO

A proteção catódica é um método elétrico de prevenção de corrosão em estruturas metálicas. Em instalações de grande porte, como dutos enterrados, utiliza-se o sistema de proteção catódica do tipo corrente contínua para garantir esta proteção. Entretanto, existe a necessidade melhorar a eficiência operacional que proporcione a visibilidade e qualidade deste tipo de operação em tempo real para que melhore a segurança operacional, qualidade na distribuição e redução de custos. Este estudo propõe um modelo de predição para contribuir com as equipes de engenharia de automação e proteção catódica nas análises e manutenção das tubulações de gás natural.

Palavras-chave: Industria 4.0. Aprendizado de Máquina. Proteção Catódica. Gás Natural.

ABSTRACT

The Cathodic protection is an electrical method of preventing corrosion in metallic structures. In large installations, such as buried pipelines, a direct current type cathodic protection system is used to ensure this protection. However, there is a need to improve operational efficiency that provides visibility and quality of this type of operation in real time to improve operational safety, quality in distribution and cost reduction. This study proposes a prediction model to help automation and cathodic protection engineering teams in the analysis and maintenance of natural gas pipelines.

Keywords: Industry 4.0. Machine Learning. Cathodic protection. Natural gas

LISTA DE FIGURAS

Figura 1 – Rede de Proteção Catódica	11
Figura 2 – Potencial de corrosão retificadores de tensão	16
Figura 3 – Análise Autocorrelação - Termo AR	19
Figura 4 – Análise Autocorrelação - Termo MA	20

LISTA DE TABELAS

Tabela 1 – Variáveis de sensoriamento Proteção Catódica	17
Tabela 2 – Classificação potencial corrosão	17
Tabela 3 – Correlação de Pearson - Potencial	18
Tabela 4 – Auto Arimax Parâmetros	20
Tabela 5 – Resultados Finais	21

LISTA DE ABREVIATURAS E SIGLAS

Icc	Tensão
Vca	Corrente Alternada
Vcc	Corrente continua
ARIMA	AutoRegressive Integrated Moving Average
ARIMAX	AutoRegressive Integrated Moving Average Exogenous
SARIMA	Seasonal AutoRegressive Integrated Moving Average
SARIMAX	Seasonal AutoRegressive Integrated Moving Average Exogenous

SUMÁRIO

1	INTRODUÇÃO	10
1.1	Problema de Pesquisa	11
1.2	Objetivos	12
1.2.1	Objetivo Geral	12
1.2.2	Objetivos Específicos	12
1.3	Justificativa	12
1.4	Organização do Trabalho	12
2	REVISÃO DA LITERATURA	14
2.1	Trabalhos Relacionados	14
2.2	Modelos Autoregressivos	15
3	MATERIAIS E MÉTODOS	16
3.1	Materiais	16
3.2	Métodos	17
3.2.1	Análise exploratória	17
3.2.2	Construção do Modelo	18
4	RESULTADOS	21
5	CONCLUSÃO	22
	REFERÊNCIAS	23

1 Introdução

Conforme (KOCH et al., 2002), a corrosão é definida como um fenômeno que ocorre de forma natural, descrito como uma deterioração de uma substância como o metal em razão de uma reação com o meio ambiente. Por exemplo, um processo de destruição de metais por meio de uma interação química ou eletroquímica com o meio ambiente, produzindo uma deterioração do material e conseqüentemente de suas propriedades da corrosão (Wantuch; Kurgan; Gas, 2016).

Expostos ao meio ambiente, com a exceção do ouro e platina, o processo de oxidação dos metais é termodinamicamente espontâneo. Quando o ferro sofre o processo de corrosão devido ao processo eletroquímico da presença de água, oxigênio e uma solução de sal eletrolítica, é formado o óxido metálico hidratado vermelho-castanho, conhecido como ferrugem, que descama continuamente expondo novas camadas de metal vulneráveis a reação de oxigênio e água. Uma forma de controlar este processo de corrosão é mitigar pelo menos uma das fontes reagentes para reduzir a taxa de corrosão (AL, 2021).

Existem algumas técnicas para combater a corrosão e a mais simples delas é a aplicação de uma capa protetora utilizando outro metal mais difícil de oxidar. Todavia, em alguns casos, a junção de dois metais podem acelerar o processo de corrosão. Conforme explicado por (FLOWERS et al., 2015), em casos onde existam tubulações subterrâneas, comum conectá-las a metais mais ativos como zinco ou magnésio, denominados também como metal de sacrifício.

A proteção catódica é uma das técnicas mais comumente utilizadas para controlar o processo de corrosão em uma estrutura metálica, fazendo com que a mesma esteja protegida como um cátodo (Nagy; Abdallah; Abbasy, 2012). Os leitos de anodos, quando monitorados adequadamente, podem prolongar a vida útil das estruturas de armazenamento de ferro. A corrosão interna destas estruturas, neste caso em gasodutos, afetam a segurança de equipamentos em casos de furos, reduzindo a confiabilidade e qualidade na distribuição de gás no gasoduto (Zhou; Wu, 2017).

Contextualizando, o ânodo é um eletrodo para onde a eletricidade se move e o cátodo é o eletrodo por onde a eletricidade flui ou é distribuída. Geralmente o ânodo é definido como o lado positivo e o cátodo como o lado negativo (Lattarulo; Amoruso, 2014). Isso faz com o metal de sacrifício contenha a corrosão no ânodo e a tubulação sirva como cátodo com menor risco de corrosão (FLOWERS et al., 2015).

Os retificadores de proteção catódica são uma fonte de energia externa, usada em sistemas de proteção catódica de corrente impressa, para converter corrente alternada em corrente contínua. Nos sistemas ICCP, a corrente é descarregada do ânodo e na estrutura para evitar corrosão. Conforme ilustrado na Figura 1, a fonte de energia elétrica deve fluir pelo equipamento de proteção catódica para controlar o processo de corrosão conforme a definição dos anodos e catodos. Uma vez que este ecossistema está configurando, o leito de anodo pode

agir como metal de sacrifício para controlar de maneira segura o processo de corrosão e agir como uma proteção catódica.

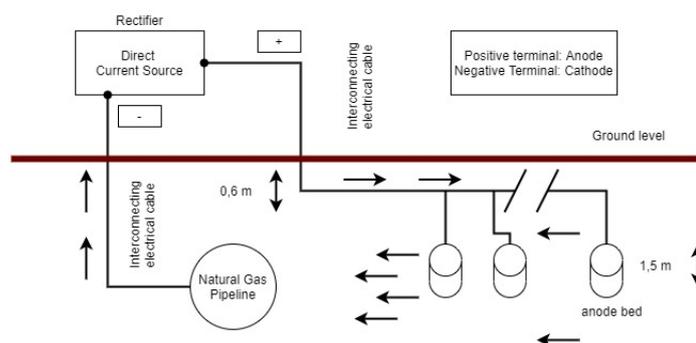


Figura 1 – Rede de Proteção Catódica

Os segmentos de serviços públicos, também conhecidos como utilities, que fornecem gás natural, água, eletricidade e telecomunicações, que foram avaliados por (KOCH et al., 2002), dependem custos direto no combate a corrosão. Avaliar um método de previsão do potencial de corrosão em gasodutos é complexo, perante a ausência de subsídios computacionais. Tem-se como decorrência tomada de ações tardias, reativas, não aprimoramento do controle por parte da equipe de engenharia e técnicos de campo. Por consequência, reduz a confiabilidade no controle operacional da corrosão nos gasodutos e potencializa os incidentes.

Em vista disso, supõe-se desenvolver uma hipótese de prever o potencial de corrosão, com base nas séries temporais das medições de sensores de tensão, corrente e potencial de corrosão. Dentro deste contexto, o aprendizado de máquina aplicado a engenharia elétrica, proteção catódica para uma malha de distribuição de gás natural, é um tema moderno para indústria e literatura. Seus resultados podem antecipar eventos de não proteção, ampliar os experimentos para indústria 4.0, contribuindo para futuros projetos e pesquisas.

1.1 Problema de Pesquisa

Em projetos de indústria 4.0, existem situações onde a baixa qualidade de dados pode indicar uma ou mais limitações no desenvolvimento dos modelos, decorrentes de má experimentação, equipamento com defeito, falha de medição, erro de cálculo ou ausência de variáveis exógenas suficientes para o desenvolvimento completo do modelo (COTTIS et al., 1999). Problemas de dados são rotineiros em projetos de Big Data e fazem parte do processo de transformação digital das indústrias. Desta forma, este é um caso de melhoria contínua nos processos de captura, tratamento e transformação dos dados para viabilizar estudos avançados e geração de valor dos dados.

Alem de questões técnicas, com o avanço tecnológico em empresas, existem casos onde os profissionais vê essas mudanças como um sinal de perigo. Pois novas tecnologias interferem e mudam os hábitos de trabalho, o que é muito difícil de mudar a curto ou médio prazo.

1.2 Objetivos

1.2.1 Objetivo Geral

Este estudo objetiva propor uma confrontação entre modelos de previsão de potencial de corrosão, baseado em dados elétricos de sensoriamento de gasodutos para retificadores de tensão. Os modelos aplicados utilizam a estrutura de modelos ARIMA (HARVEY, 1990) e ARIMAX (BOX GWILYM M. JENKINS; LJUNG, 2016) construídos a partir do pacote SARIMAX do Python comparando as métricas de performance do modelo de acordo com as variáveis exógenas (variáveis externas ao modelo e que representam as entradas do modelo) disponíveis para o estudo.

1.2.2 Objetivos Específicos

O desenvolvimento do modelo de previsão para o potencial de corrosão é bastante desafiador e dividido em algumas fases exploratórias para a sua concretização do modelo. Neste estudo, inicialmente deve ser construída uma análise de tendência e distribuição dos dados de sensoriamento para contribuir no conhecimento da base e entendimento da característica de cada série temporal dos sensores.

Esta fase exploratória, compete também analisar o extrapolamento dos dados e, caso necessário, definir os melhores métodos de remoção dos dados extrapolados, combinado a testes estatísticos de correlação linear, análise de fatorial e multicolinearidade, além dos processos de normalização e agregação. Cada estado deste desenvolvimento é muito importante, pois precede a construção e confronto dos modelos ARIMA e ARIMAX.

1.3 Justificativa

As indústrias dependem de custos diretos no combate à corrosão, que muitas vezes não possuem subsídios computacionais para agilizar no recebimento de informações, decorrendo tomada de ações tardias e reativas. Desta forma, proporcionar ferramentas e recursos computacionais permitem que os usuários tomem decisões com alto grau de confiabilidade e limitem a quantidade de incidentes.

Para o segmento de utilities, como a distribuição de gás natural, a segurança é um dos pilares estratégicos das empresas. Sendo assim, o desenvolvimento e investimento em novos projetos de engenharia e computação que auxiliem as tomadas de decisões e melhorias de processo são amplamente incentivados e relevantes.

1.4 Organização do Trabalho

A proposta deste trabalho está organizada seguindo os seguintes tópicos. A Seção 1 contém a revisão de literatura e trabalhos relacionados ao tema proposto. Na Seção 2, análise

exploratória dos dados disponíveis seguindo as tratativas realizadas. Na Seção 3, contém os métodos e materiais aplicados, Seção 4, os resultados e Seção 5 as considerações finais.

2 Revisão da Literatura

Existe uma dificuldade de construir um modelo matemático universal para previsão de corrosão que possa representar todos os fatores, refletindo os resultados reais com alta precisão, devido à grande quantidade de discretização de dados nos testes de laboratório ou em ambientes reais (LIU; ZHAO; YI, 2009). Na literatura podem ser encontradas algumas abordagens baseadas em redes neurais e modelos autorregressivos, porém os projetos de automação e sensoriamento são diferentes dependendo do tipo de indústria e objetivo de projeto.

2.1 Trabalhos Relacionados

Conforme abordado por (MA et al., 2009), o desenvolvimento de modelos de previsão para corrosão de equipamentos de aço é estabelecido com relativo sucesso utilizando redes neurais, que encontram as relações complexas entre os fatores que influenciam a corrosão dos equipamentos e a probabilidade de corrosão. Todavia, deve-se considerar a limitação de amostras experimentais e dados, de cada estudo.

Os modelos de redes neurais baseados em retropropagação e função de base radial podem se tornar um novo método de avaliação eficaz para a taxa de corrosão com melhores resultados de previsão e maior precisão de reconhecimento (LIU; ZHAO; YI, 2009). No entanto, esses métodos requerem conjuntos de dados consistentes (COTTIS et al., 1999). Dentro deste contexto, a baixa qualidade de dados pode indicar uma ou mais limitações no desenvolvimento dos modelos, decorrentes de experiências ruins, equipamento com defeito, falha de medição, controle ou erro de cálculo (COTTIS et al., 1999).

Além dos modelos baseados em redes neurais, existem vertentes na literatura com foco no uso de modelos autorregressivos. Conforme (CHEN; YANG; LU, 2021), a previsão e o potencial do estado de corrosão, podem ser realizados por monitoramento online aplicados em modelo de predição baseado na média móvel integrada autorregressiva (ARIMA). Porém a construção e universalização de modelos autorregressivos que consideram variáveis exógenas, na medição de potencial de corrosão em gasodutos, continua em aberto para pesquisa e desenvolvimento científico.

Nos trabalhos relacionados citados acima, o uso de outros fatores para previsão de potencial de corrosão estão relacionados à composição química e ambiental presente na medição dos equipamentos. Todavia, diferentemente da literatura, neste estudo foram utilizados fatores como componentes elétricos pertencentes ao ecossistema de proteção catódica de uma rede de gasodutos.

2.2 Modelos Autoregressivos

A origem dos dados de série temporal surgiram por meio de monitoramento de processos industriais ou rastreamento de métricas de negócios. Esta análise explica o fato de que os pontos de dados obtidos ao longo do tempo podem ter uma estrutura interna (como autocorrelação, tendência ou variação sazonal) que deve ser contabilizada. Existem alguns métodos de aprendizado de máquina utilizados para modelar problemas de séries temporais, por exemplo os modelos autoregressivos ARIMA e ARIMAX (BOX GWILYM M. JENKINS; LJUNG, 2016).

O modelo de previsão proposto neste estudo é considerado uma expansão do ARIMA, composto pelos parâmetros (p, d, q) (P, D, Q) s onde, p e P representam o componente autoregressivo e autoregressivo sazonal, respectivamente. O d e D , as diferenças não sazonais e diferenciação sazonal. Por fim, q e Q os parâmetros da média móvel e os parâmetros da média móvel sazonal, sendo s representando a duração do período sazonal.

Uma série temporal estacionária é uma série temporal cujas estatísticas não mudam com o tempo, sendo essas estatísticas normalmente a média e a variância. O resultado do teste estatístico para medir estacionariedade resulta na possibilidade de previsão dos dados da nossa série temporal. Nesta pesquisa, a série de dados era não estacionária, mas após tirar a primeira diferença, a média ficou constante (HARVEY, 1990).

O modelo ARIMAX é uma extensão da modelagem ARIMA na tentativa de prever a corrosão usando as variáveis exógenas tais como corrente elétrica, corrente alternada, tensão e dados temporais (dia, mês, ano, hora e minuto).

3 Materiais e Métodos

3.1 Materiais

Os dados deste estudo têm a característica de serem séries temporais e com a taxa de captura de um hora, sendo: corrente contínua, corrente alternada, tensão e potencial tubo solo referentes a malha de gasodutos da maior distribuidora de gás em São Paulo, Brasil. Os mesmos foram coletados no período de novembro de 2020 até janeiro de 2021.

Pelo fato da base de dados ser uma série temporal multivariada, ou seja, existe mais de uma observação por tempo de medição, busca-se, inicialmente, entender os fatores de correlação, causalidade e fatores entre as variáveis do estudo. Além disso, a avaliação da qualidade das medições e controle do sensoriamento fazem parte do processo de análise e construção da base de dados deste estudo.

Estas informações foram coletadas e transmitidas utilizando um conjunto de sensores e transmissores de uma empresa multinacional e armazenada em um sistema PIMs em tempo real. A extração dos dados pôde ser feita através de conectores em planilhas excel ou ferramentas de visualização.

Foram fornecidos os dados de sensoriamento de onze retificadores de tensão, representados na [Figura 2](#). Cada linha representa o valor de potencial de corrosão para cada retificador ao longo do tempo. Os retificadores, trabalham com níveis de corrente elétrica alternada, contínua e tensão diferentes e medidos em estagios de corrosão distintos. A origem dos dados, são de uma companhia de gás natural, foi selecionado um equipamento para o desenvolvimento e confrontamento dos modelos ARIMA e ARIMAX. O critério de seleção foi baseado na qualidade de persistência dos dados da série histórica.

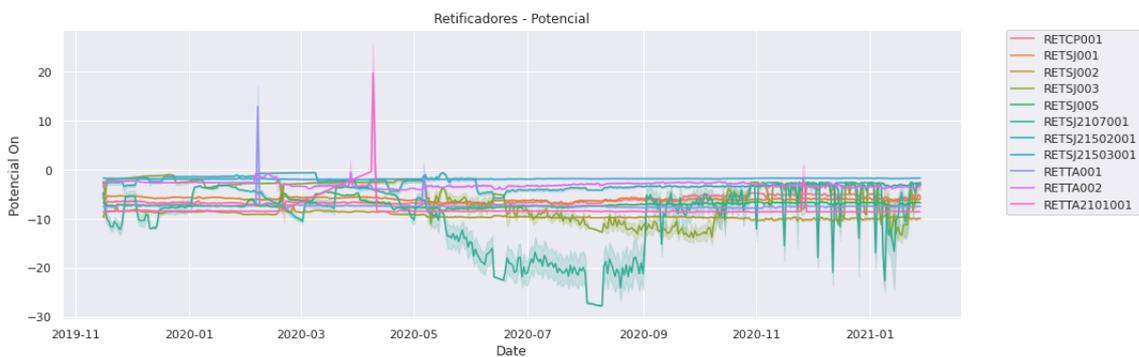


Figura 2 – Potencial de corrosão retificadores de tensão

Contextualizando, os dados destes retificadores tem origem em um projeto de indústria 4.0, onde ocorreram diversos testes de campo e ajustes na transmissão, armazenamento e transformação nos dados, que prejudicaram significativamente a qualidade das amostras.

Tabela 1 – Variáveis de sensoriamento Proteção Catódica

Variável	Descrição	Unidade
Date/Time	Data e Hora	-
Retificador	Nome do Retificador	-
Potencial On	Potencial de corrosão tubo solo	Volts
ICC	Tensão corrente elétrica	Ampere
VCA	Corrente alternada	Ampere
VCC	Corrente Continua	Volts

Portanto, o retificador selecionado apresentava informações consistentes com pouca ausência de dados em todas as suas séries temporais.

Os dados abrangem o período de novembro de 2019 até fevereiro de 2021. Todos os dados estão descritos na [Tabela 1](#), com sua respectiva unidade de medida e periodicidade, totalizando 66338 amostras. Adicionalmente, foi fornecido pela equipe de engenharia de proteção catódica os dados de classificação de potencial de corrosão tubo solo conforme procedimento da empresa [Tabela 2](#).

Para o desenvolvimento e análise de dados, foi utilizada a linguagem de programação Python bem como seus módulos pandas, numpy, scikit-learn ([PEDREGOSA et al., 2011](#)) e stats-model ([SEABOLD; PERKTOLD, 2010](#)) para construção da análise exploratória, interpolação de linear em dados faltantes, normalização dos dados e construção de modelos autoregressivos.

Dentro do processo de análise exploratória dos dados, foram selecionadas as principais variáveis correlacionadas para treinamento e teste conforme o teste de correlação de Pearson.

Tabela 2 – Classificação potencial corrosão

Classe	Tensão
Superprotegido	< -4V
Protegido	<= -0,85V
Corrosão Moderada	> -0,85V e <= 0V
Corrosão Avançada	> 0V

3.2 Métodos

3.2.1 Análise exploratória

Inicialmente, todos os dados de sensoriamento, que estavam armazenados no banco de dados temporal do sistema PIMS, foram extraídos de maneira adhoc em formato Microsoft excel para análise exploratória e desenvolvimento dos modelos de maneira local.

Devido a complexidade do estudo, quantidade e qualidade de dados, foi selecionado o retificador RETSJ21502001 para desenvolvimento final conforme análise de tendência, menores

índices de outliers e completude dos dados.

As variáveis Potencial, Icc, Vcc e Vca foram agrupadas através da agregação da média diária, transformando os dados de hora para dia, sobre o potencial de corrosão totalizando 439 linhas. Esta transformação foi necessária devido a quantidade de dados horarios faltantes ao longo de 24 horas e baixa variação ao decorrer das horas. O resultado da transformação melhorou o tempo de processamento e precisão nas predições. Além disso, foram realizados testes estatísticos sobre as variáveis identificando correlação baixa e moderada negativa do potencial de corrosão com as variáveis Year, Quarter, Month, Icc e Vcc conforme [Tabela 3](#).

Tabela 3 – Correlação de Pearson - Potencial

Variável	r
Potencial	1
Year	-0,060
Quarter	-0,365
Month	-0,336
Icc	-0,185
Vcc	-0,483
Vca	-0,068

A colinearidade é o estado em que duas variáveis são altamente correlacionadas e contêm informações semelhantes sobre a variação em um determinado conjunto de dados. A multicolinearidade, por outro lado, é mais difícil de detectar porque surge quando três ou mais variáveis, que são altamente correlacionadas, são incluídas em um modelo. A multicolinearidade pode surgir mesmo quando pares isolados de variáveis não são colineares. Ela afeta e inflaciona o desvio padrão de cada variável, aumenta o valor de p-value e consequentemente o intervalo de confiança.

Para medir índices de multicolinearidade entre as variáveis selecionadas após o teste de correlação, aplica-se o teste Fator de inflação de variância. Definindo o valor de corte igual a 10, o teste indicou altos índices de multicolinearidade para as variáveis Quarter 23.54 e Month 22.66. Considerando que as variáveis tinham baixo nível de correlação, não se fez necessário uma análise fatorial exploratória. Desta forma foram mantidas as variáveis Icc e Vcc como exógenas ao potencial com valores de índices de 2,80.

3.2.2 Construção do Modelo

Para identificar inicialmente os primeiros valores para os termos AR, I e MA do modelo ARIMA, desenvolve-se uma análise gráfica utilizando gráfico de Autocorrelação Parcial (PACF) e o grafico de autocorrelação automática (ACF), que representam a quantidade de defasagens (lags) ideais para o termo AR e MA respectivamente, além disso, ambas análise contem dois níveis de derivação para no termo I.

Conforme ilustrado na [Figura 3](#) e [Figura 4](#), pode-se observar que o atraso 1 do PACF termo AR é bastante significativo, pois está bem acima da linha de significância. Porém o segundo atraso, não é significativo pois não cruza o limite de significância (região azul). Para o termo MA, temos um comportamento distinto onde até o termo 20 temos uma correlação moderada. Para graus de diferenciação, termo I igual a 1 ou 2, os termos AR e MA são negativos e significância moderada para baixo.

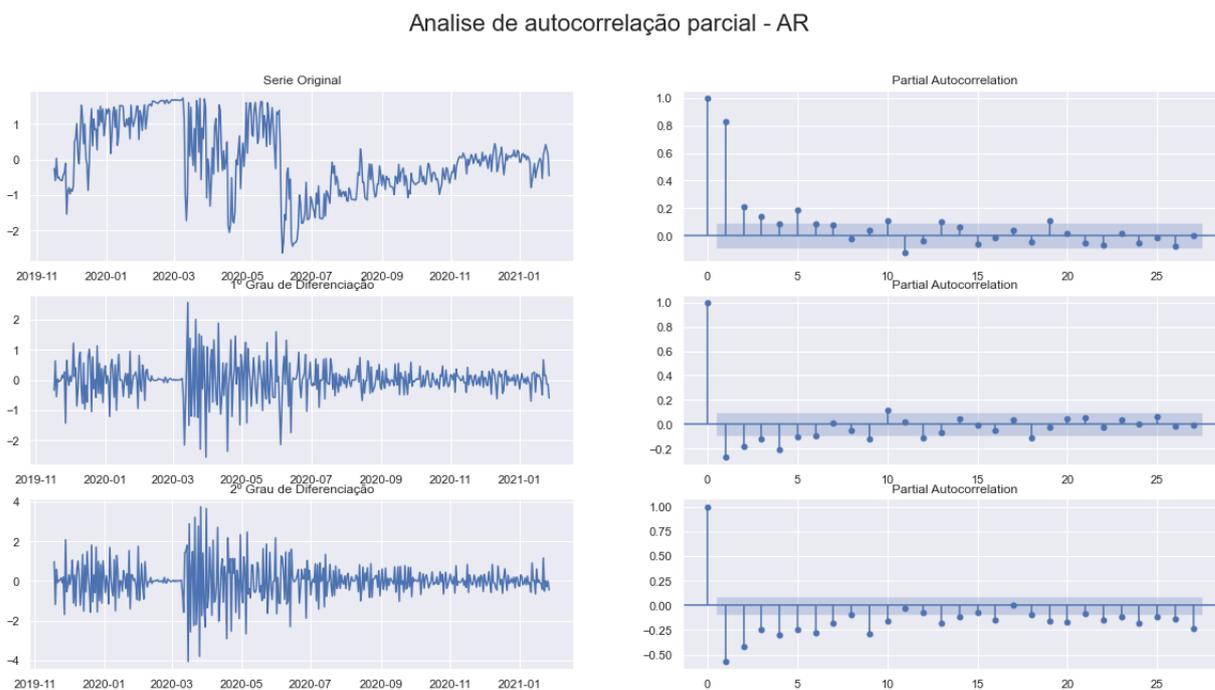


Figura 3 – Análise Autocorrelação - Termo AR

Como parte do processo de construção do modelo aplicou-se o teste estatístico de Dickey Fuller para identificar se a série era estacionária ou um caminho aleatório. O primeiro resultado com p-value de 0.15 indicava que a série temporal do potencial de corrosão não era estacionária, todavia, ao aplicar 1 grau de diferenciação sobre a série, rejeitamos a hipótese nula com p-value de 0.0000. Desta forma, o modelo deve considerar pelo menos 1 grau de diferenciação no termo I para o ARIMA e ARIMAX.

Como as variáveis exógenas e endógenas estão em escalas distintas, aplicou-se a normalização dos dados por meio do módulo do scikit-learn StandardScaler. Posteriormente a divisão de 300 linhas para treinamento e 7 para teste, baseado no horizonte de 7 dias para previsão.

Para construção e otimização do modelo ARIMAX, considerou-se a modelagem da versão do autoarimax baseado no SARIMAX do modulo statsmodels do python ([SEABOLD; PERKTOLD, 2010](#)), buscando dos melhores parâmetros com base nas métricas de validação do modelo BIC, AIC e RMSE. Desta forma, existe um ranking das 5 melhores combinações de parâmetros do ARIMAX ordenados pela RMSE. [Tabela 4](#) exemplifica o modelo utilizado como

Análise de autocorrelação - MA

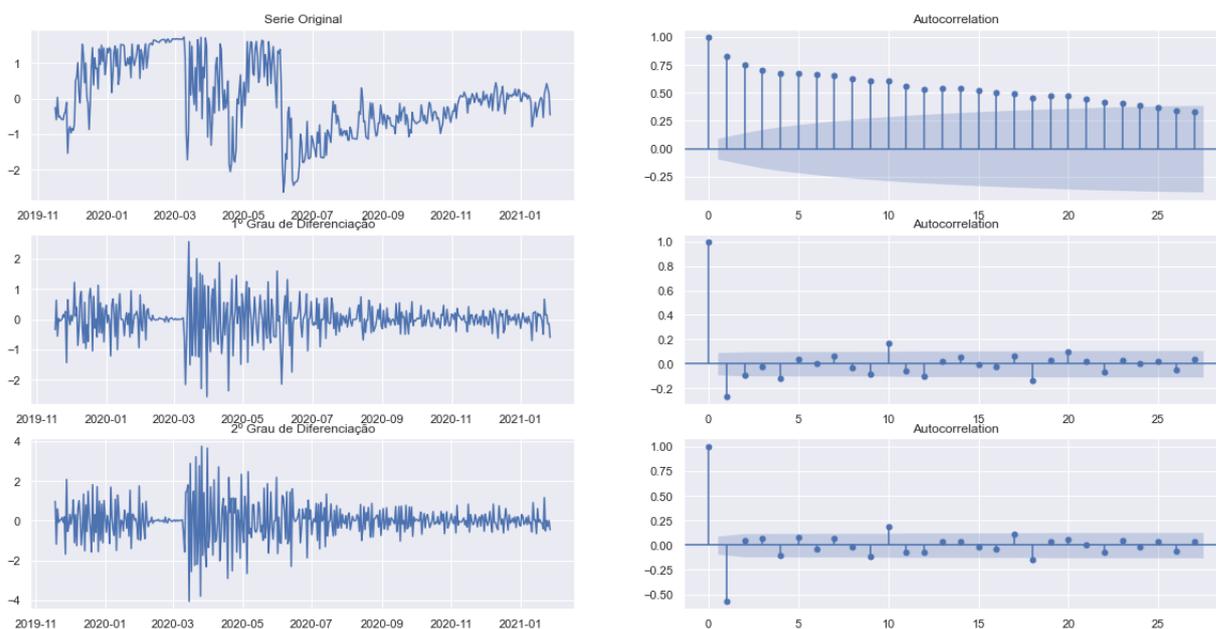


Figura 4 – Análise Autocorrelação - Termo MA

ranking de melhores termos ARIMA.

Tabela 4 – Auto Arimax Parâmetros

PDQ	BIC	AIC	RMSE
(3, 0, 7)	620,21	572,0610	0,21695
(5, 0, 7)	631,52	575,9599	0,21989
(3, 0, 6)	620,86	576,4172	0,22162
(3, 0, 8)	625,73	573,8811	0,22207
(7, 0, 4)	627,76	575,9055	0,22252

4 Resultados

Em sistemas de sensoriamento, é muito comum os dados terem ruídos. Portanto, o modelo de previsão de potencial de corrosão deve ser robusto ao ponto de evitar o sobre-ajuste durante o treinamento e teste, considerando uma base de qualidade com poucos ruídos.

Dentro dos cenários propostos, foram realizados diversos experimentos para avaliar o critério de parada ideal, selecionando amostras entre picos e vales para treinamento e teste. Desta maneira, foram desenvolvidos quatro cenários para os modelos ARIMA e ARIMAX com horizonte de previsão de 7 dias. Combinando duas variáveis exógenas com melhores níveis de correlação com o potencial de corrosão.

Para o primeiro cenário, não foram consideradas as variáveis exógenas. Aplicou-se apenas um modelo ARIMA puro com 7 atrasos com melhor resultado possível de RMSE 0,7 em fase de treinamento. Conseqüentemente, de acordo com a [Tabela 5](#) para os resultados da fase de teste, ou seja, para dados não conhecidos, uma RMSE de 0,38.

Alternativamente, foram experimentados outros modelos ARIMAX combinando variáveis exógenas como ICC ou VCC, todavia os resultados foram indiferentes ao primeiro modelo ARIMA, com pequenas variações na RMSE de treinamento e teste. Ademais, a pesquisa mostra que o modelo não foi capaz de mapear de forma assertiva momentos de picos e vales ao longo das séries com grande precisão, sempre alternando com um ponto de atraso ao começo do ponto de inflexão.

Tabela 5 – Resultados Finais

Variaveis	Modelos	AIC	BIC	RMSE Treino	RMSE Teste
-	ARIMA	582,785	638,341	0,769	0,387
ICC	ARIMAX	577,668	629,521	0,764	0,369
VCC	ARIMAX	575,606	627,459	0,761	0,387
(ICC, VCC)	ARIMAX	579,723	635,279	0,760	0,360

5 CONCLUSÃO

O estudo mostra que apesar dos resultados serem medianos para RMSE, o modelo não conseguiu mapear assertivamente momentos de picos e vales ao longo da série temporal. Isso significa que as variáveis exógenas (i.e., variáveis de entrada) não conseguem responder em plenitude o comportamento da variável potencial de corrosão pois as exógenas são poucos explicativas. Além disso, o horizonte de previsão muito distante não se adequaria a este tipo de modelo autoregressivo dado o erro acumulativo.

Como próximos passos, deve-se considerar uma quantidade maior de retificadores que estão conectados na mesma malha de distribuição. Investigar a correlação entre os retificadores a nível geográfico na malha de distribuição dos gasodutos, explorar novas variáveis relacionadas à pressão, ao volume, aos dados climáticos e geográficos. Não menos importante, deve-se avaliar a possibilidade de instalação de novos sensores para medição destas novas variáveis e iniciar um processo de melhoria contínua na coleta e transmissão destes dados para o banco de dados.

Referências

- AL, B. et. Chemistry: The central science. libretexts, 2021. Disponível em: <<https://chem.libretexts.org/@go/page/21797>>. Citado na página 10.
- BOX GWILYM M. JENKINS, G. C. R. G. E. P.; LJUNG, G. M. **Time Series Analysis: Forecasting and Control, 5th Edition**. [S.l.]: John Wiley, 2016. Citado 2 vezes nas páginas 12 e 15.
- CHEN, L.; YANG, J.; LU, X. Research on time series prediction model for the trend of corrosion rate. In: **2021 IEEE Asia Conference on Information Engineering (ACIE)**. [S.l.: s.n.], 2021. p. 78–81. Citado na página 14.
- COTTIS, R. et al. Neural network methods for corrosion data reduction. **Materials & Design**, v. 20, n. 4, p. 169–178, 1999. ISSN 0261-3069. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0261306999000266>>. Citado 2 vezes nas páginas 11 e 14.
- FLOWERS, P. et al. **Chemistry**. Reading, Massachusetts: OpenStax, 2015. Disponível em: <<https://openstax.org/books/chemistry/pages/1-introduction>>. Citado na página 10.
- HARVEY, A. C. Arima models. In: _____. **Time Series and Statistics**. London: Palgrave Macmillan UK, 1990. p. 22–24. ISBN 978-1-349-20865-4. Disponível em: <https://doi.org/10.1007/978-1-349-20865-4_2>. Citado 2 vezes nas páginas 12 e 15.
- KOCH, G. et al. Historic congressional study: Corrosion costs and preventive strategies in the united states. **Mater. Perform.**, p. 1–11, 01 2002. Citado 2 vezes nas páginas 10 e 11.
- Lattarulo, F.; Amoruso, V. Fundamentals of electrical discharges. In: _____. **Filamentary Ion Flow: Theory and Experiments**. [S.l.: s.n.], 2014. p. 1–24. Citado na página 10.
- LIU, Y.; ZHAO, S.-L.; YI, C. The forecast for corrosion of reinforcing steel based on rbf neural network. In: **2009 International Conference on Wavelet Analysis and Pattern Recognition**. [S.l.: s.n.], 2009. p. 195–199. Citado na página 14.
- MA, Z. et al. A forecast model based on the bp neural network used in refinery's steel equipment's corrosion. In: **2009 4th International Conference on Computer Science Education**. [S.l.: s.n.], 2009. p. 1138–1141. Citado na página 14.
- Nagy, K. M.; Abdallah, E. N.; Abbasy, N. H. Developed software for cathodic protection of storage tanks. In: **2012 IEEE International Conference on Computer Science and Automation Engineering (CSAE)**. [S.l.: s.n.], 2012. v. 3, p. 88–93. Citado na página 10.
- PEDREGOSA, F. et al. Scikit-learn: Machine learning in Python. **Journal of Machine Learning Research**, v. 12, p. 2825–2830, 2011. Citado na página 17.
- SEABOLD, J.; PERKTOLD, J. Statsmodels: Econometric and statistical modeling with python. In: **Proceedings of the 9th Python in Science Conference**. [S.l.: s.n.], 2010. Citado 2 vezes nas páginas 17 e 19.
- Wantuch, A.; Kurgan, E.; Gas, P. Numerical analysis on cathodic protection of underground structures. In: **2016 13th Selected Issues of Electrical Engineering and Electronics (WZEE)**. [S.l.: s.n.], 2016. p. 1–4. Citado na página 10.

Zhou, Y.; Wu, Z. Risk index assessment for urban natural gas pipeline leakage based on artificial neural network. In: **2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)**. [S.l.: s.n.], 2017. p. 1261–1263. Citado na página [10](#).