



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

Alexandre Henrique Brescovitt

**Aprendizado de máquina aplicado à caracterização do ponto de operação
empregando o campo magnético externo em máquinas síncronas**

Florianópolis

2024

Alexandre Henrique Brescovitt

**Aprendizado de máquina aplicado à caracterização do ponto de operação
empregando o campo magnético externo em máquinas síncronas**

Dissertação submetida ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Santa Catarina como requisito parcial para a obtenção do título de Mestre em Engenharia Elétrica.

Orientador: Prof. Nelson Jhoe Batistela, Dr.

Coorientador: Prof. Leandro dos Santos Coelho, Dr.

Florianópolis

2024

Brescovitt, Alexandre Henrique

Aprendizado de máquina aplicado à caracterização do ponto de operação empregando o campo magnético externo em máquinas síncronas / Alexandre Henrique Brescovitt ; orientador, Nelson Jhoe Batistela, coorientador, Leandro dos Santos Coelho, 2024.

166 p.

Dissertação (mestrado) - Universidade Federal de Santa Catarina, Centro Tecnológico, Programa de Pós-Graduação em Engenharia Elétrica, Florianópolis, 2024.

Inclui referências.

1. Engenharia Elétrica. 2. Gerador síncrono. 3. Ponto de operação de gerador síncrono. 4. Campo magnético externo. 5. Modelos de regressão linear. I. Batistela, Nelson Jhoe. II. Coelho, Leandro dos Santos. III. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Engenharia Elétrica. IV. Título.

Alexandre Henrique Brescovitt

**Aprendizado de máquina aplicado à caracterização do ponto de operação
empregando o campo magnético externo em máquinas síncronas**

O presente trabalho em nível de Mestre foi avaliado e aprovado, em 08 de julho de 2024,
pela banca examinadora composta pelos seguintes membros:

Eng. Rômulo Gonçalves Bainy, Dr.
University of Idaho

Prof. Walter Pereira Carpes Junior, Dr.
Universidade Federal de Santa Catarina

Prof. Carlos Alexandre Correa Wengerkievicz, Dr.
Instituto Federal de Santa Catarina

Certificamos que esta é a versão original e final do trabalho de conclusão que foi julgado
adequado para obtenção do título de Mestre em Engenharia Elétrica

Insira neste espaço a
assinatura digital

Coordenação do Programa de Pós-Graduação

Insira neste espaço a
assinatura digital

Prof Nelson Jhoe Batistela, Dr.
Orientador

Florianópolis, 2024

AGRADECIMENTOS

Aos meus pais, que sempre me deram o suporte necessário e estiveram torcendo pelo êxito deste trabalho.

À minha companheira Barbara, que esteve me acompanhando de perto em todos os momentos durante o desenvolvimento deste trabalho. Sua companhia foi fundamental e de grande importância

Aos meus amigos de longa data que sempre estiveram fomentando a minha melhora profissional e pessoal.

Aos meus orientadores, Prof. Jhoe e Prof. Leandro, por todo o direcionamento, apoio e boas discussões sobre o tema deste trabalho.

Aos meus colegas do LAESP, por toda a contribuição, seja de maneira direta ou indireta. Pela participação dos experimentos, discussões técnicas que renderam ótimas conclusões e conversas sobre os mais diversos temas em momentos de descontração.

O trabalho de pesquisa apresentado nesta dissertação foi motivado e parcialmente custeado no âmbito do programa de P&D ANEEL e desenvolvido no programa de P&D da Engie Brasil Energia e Itá Energética S.A. no projeto PD-00403-0048/2019, intitulado “Equipamento não Invasivo para Detecção de Falhas em Geradores Síncronos através do Campo Magnético Externo” e no projeto PD 0403-0057/2023, intitulado “Equipamento não Invasivo para Detecção de Falhas em Geradores Síncronos através do Campo Magnético Externo – Lote Pioneiro e Inserção no Mercado”, realizado em parceria com a empresa AQTech Engenharia e Instrumentação S.A. A CAPES (Código de Financiamento 001) também custeou parcialmente esta pesquisa.

RESUMO

Os geradores síncronos de rotor bobinado são parte fundamental da geração de energia elétrica. O monitoramento dessas máquinas é importante, uma vez que sua operação precisa ser confiável. Neste contexto, há pesquisas recentes na área de monitoramento de máquinas elétricas rotativas por meio do campo magnético externo com o objetivo de desenvolver métodos capazes de detectar falhas elétricas e mecânicas em estado inicial. Em processos de detecção de falhas em estágio inicial por meio do campo magnético externo, há uma lacuna em não se considerar a influência de variações no ponto de operação no comportamento do campo magnético externo de geradores síncronos conectados à rede elétrica. Investigações experimentais que consideram o ponto de operação em suas análises relatam que as alterações causadas no campo externo pelo ponto de operação precisam ser levadas em conta em processos de monitoramento. O ponto de operação provoca alterações significativas no campo magnético externo da máquina, e podem ser maiores do que aquelas provocadas por falhas incipientes. Este trabalho apresenta uma modelagem do ponto de operação de um gerador síncrono de polos salientes tendo como dado de entrada o campo magnético externo. Experimentos foram realizados em um gerador sincronizado à rede para coletar formas de onda associadas ao campo magnético externo e dados relativos ao ponto de operação, como potência ativa, potência reativa e tensão. Este conjunto de dados foi tratado, obtendo dois conjuntos de dados: um conjunto de dados independentes do tempo e um conjunto de dados dependentes do tempo. Uma parcela dos dados coletados foi analisada e utilizada como treinamento para gerar um modelo capaz de estimar o ponto de operação do gerador apenas com informações do campo magnético externo. Obteve-se como resultado um modelo de regressão que utiliza as amplitudes das harmônicas da taxa de variação do campo magnético externo para estimar o ponto de operação do gerador. Os resultados obtidos foram considerados satisfatórios apresentando desvios na ordem de 1% da potência nominal do gerador, o que indica a capacidade do método proposto.

Palavras-chave: Gerador síncrono; campo magnético externo; ponto de operação, modelos de regressão.

ABSTRACT

Wound rotor synchronous generators are a fundamental part of power generation. The monitoring of these machines is important as their operation must be reliable. In this context, the monitoring of rotating electrical machines by means of the external magnetic field has been studied recently with the aim of developing methods capable of detecting electrical and mechanical faults at an early stage. There is a gap in the early-stage fault detection methods using the external magnetic field, as they do not take into account the influence of operating point variations on the behavior of the external magnetic field of synchronous generators connected to the grid. Experimental studies that consider the operating point in their analysis report that the changes in the external field caused by the operating point need to be considered in the monitoring processes. The operating point causes significant changes in the external magnetic field of the machine, which can be greater than those caused by incipient faults. This dissertation presents a modelling of the operating point of a salient pole synchronous generator using the external magnetic field as input data. Experiments were carried out to obtain a reliable data set. This data set was processed to obtain two sets of data: a set of time-independent data and a set of time-dependent data. A portion of the collected data was analyzed and used as training to generate a model capable of estimating the generator operating point using only information from the external magnetic field. The result was a regression model that used the amplitudes of the harmonics of the rate of change of the external magnetic field to estimate the operating point of the generator. The results obtained were considered satisfactory, indicating the capability of the proposed method.

Keywords: Synchronous generator; external magnetic field, leakage magnetic field, stray magnetic field; operating point, regression models.

LISTA DE FIGURAS

Figura 1 - Fluxograma das diretrizes propostas para detecção e análise de faltas em máquinas síncronas.....	18
Figura 2 - Composição do método SSF.....	24
Figura 3 - Crescimento do interesse na pesquisa sobre monitoramento de geradores síncronos por meio do campo magnético externo.....	27
Figura 4 – Variação da densidade espectral de potência com a porcentagem de carregamento do gerador e a intensidade da falta de curto-circuito entre espiras.	28
Figura 5 - Variação da amplitude da densidade de fluxo magnético para a componente axial (a) e tangencial (b) de 50 Hz com o ponto de operação do gerador.	31
Figura 6 - Histórico da amplitude da 141ª componente harmônica da frequência mecânica do gerador durante a evolução de uma falta de vibração mecânica.....	32
Figura 7 – Arquitetura do sistema (equipamento e tratamento de dados) proposto para monitoramento do campo externo.....	33
Figura 8 - Evolução da amplitude da componente harmônica de frequência de 862,5 Hz em função das medições ao longo do tempo para o gerador de 8 polos operando a 70% de sua carga nominal, tensão nominal e 50 Hz.	35
Figura 9 - Esquemático simplificado de um sistema de alimentação do campo do gerador sem o uso de escovas (<i>brushless excited synchronous machine</i>).....	36
Figura 10 - Resultado da aplicação do método VAE para o conjunto de dados (dados reais de máquina sadia e dados sintéticos de máquina com falta).	38
Figura 11 – Ilustração da disposição dos sensores indutivos ao redor do GSPS.	41
Figura 12 -Grupo gerador especial.....	43
Figura 13 - Painel de controle do grupo gerador especial.	43
Figura 14 - Sensor indutivo utilizado para medição da taxa de variação do campo magnético externo.	44
Figura 15 - Sistema de aquisição de dados para o campo externo do GSPS.	45
Figura 16 - Fluxograma do procedimento de operação do grupo gerador....	49

Figura 17 - Representação das grandezas elétricas para a potência ativa de 7 kW com dados obtidos do sistema de monitoramento VibraOne.	50
Figura 18 – Representação gráfica dos dados de ponto de operação apresentados com dados de temperatura.	52
Figura 19 - Representação gráfica dos sinais proporcionais à derivada temporal do campo magnético externo do GPS em um trecho 80 ms.	53
Figura 20 – Exemplo de seleção dos pontos obtidos do sistema de monitoramento VibraOne.	55
Figura 21 - Desvio dos dados de potência obtidos.	56
Figura 22 - Dados de temperatura selecionados para o ponto de operação de 7 kW e 6 kvar.	57
Figura 23 - Histogramas da variação da tensão terminal e da frequência da tensão terminal.	58
Figura 24 - Formas de onda representativas da taxa de variação do campo magnético externo obtidas pela UMM.	59
Figura 25 - Pontos de operação obtidos experimentalmente.	60
Figura 26 – Variação do valor RMS da tensão induzida nos sensores de campo para diferentes durações de janela de cálculo.	63
Figura 27 - Valor RMS dos vetores de tensão para os três canais, canal 1 (CH1), canal 2 (CH2) e canal 3 (CH3).	64
Figura 28 - Espectro de frequência do vetor de tensão obtido do sensor CH1.	65
Figura 29 - Espectro de frequência do vetor de tensão obtido do sensor CH1 com filtro de corte por amplitude mínima com aplicação de filtro por média móvel.	66
Figura 30 - Aplicação do filtro de média móvel para a harmônica de 1650 Hz.	66
Figura 31 - Desvio padrão, em dBV, das amplitudes das harmônicas filtradas para CH1.	67
Figura 32 - Flutuação da frequência fundamental observada na harmônica de 1620 Hz.	68
Figura 33 - Dados gerais do GPS comparados com os dados de potência ativa e reativa.	72
Figura 34 - Atributos mais relacionadas com a potência ativa e reativa de acordo com a correlação de Pearson.	74

Figura 35 - Correlações entre atributos mais correlacionados com as potências pelo método de Pearson.	75
Figura 36 - Atributos mais relacionadas com a potência ativa e reativa de acordo com a correlação de Spearman.	77
Figura 37 - Correlações entre atributos mais correlacionados com as potências pelo método de Spearman.	78
Figura 38 - Atributos mais relacionadas com a potência ativa e reativa de acordo com a correlação de PhiK.	79
Figura 39 - Correlações entre atributos mais correlacionados com as potências pelo método de PhiK.	80
Figura 40 – Escolha dos modelos a ser utilizado.	83
Figura 41 - Visão geral do processo de criação do modelo de regressão.	85
Figura 42 – Fluxo de processamento de dados: validação cruzada <i>k-fold</i> com teste final <i>hold-out</i>	86
Figura 43 – Processo adotado para engenharia de atributos.	88
Figura 44 - Dados originais, padronizados e normalizados do canal 1 para a harmônica de 45 Hz.	90
Figura 45 - Exemplo dos três primeiros componentes principais dos dados padronizados.	91
Figura 46 – Atributos gerados com t-SNE.	92
Figura 47 - Atributos gerados com UMAP.	94
Figura 48 - Curva do R^2 médio em cada validação cruzada para os atributos ranqueados por <i>Selective</i> e correlação para a potência ativa.	97
Figura 49 - Curva do MAE médio em cada validação cruzada para os atributos ranqueados por <i>Selective</i> e correlação para a potência ativa.	98
Figura 50 - Curva do RMSE médio em cada validação cruzada para os atributos ranqueados por <i>Selective</i> e correlação para a potência ativa.	98
Figura 51 - Curva do R^2 médio em cada validação cruzada para os atributos ranqueados por <i>Selective</i> e correlação para a potência reativa.	98
Figura 52 - Curva do MAE médio em cada validação cruzada para os atributos ranqueados por <i>Selective</i> e correlação para a potência reativa.	99
Figura 53 - Curva do RMSE médio em cada validação cruzada para os atributos ranqueados por <i>Selective</i> e correlação para a potência reativa.	99

Figura 54 - Diagrama em blocos do processo de seleção do modelo, treinamento, validação e ajuste de hiperparâmetros.....	102
Figura 55 - Resultado para o modelo de regressão linear <i>baseline</i> no conjunto de teste para a potência ativa.	112
Figura 56 - Resultado para o modelo de regressão linear <i>baseline</i> no conjunto de teste para a potência reativa.	113
Figura 57 - Resultado para o modelo <i>Lasso</i> no conjunto de teste para a potência ativa.	113
Figura 58 - Resultado para o modelo <i>Lasso</i> no conjunto de teste para a potência reativa.	114
Figura 59 - Resultado para o modelo <i>Ridge</i> no conjunto de teste para a potência ativa.	114
Figura 60 - Resultado para o modelo <i>Ridge</i> no conjunto de teste para a potência reativa.	115
Figura 61 - Resultado para o modelo <i>Elastic net</i> no conjunto de teste para a potência ativa.	115
Figura 62 - Resultado para o modelo <i>Elastic net</i> no conjunto de teste para a potência reativa.	116
Figura 63 - Aplicação do modelo final escolhido para prever o ponto de operação do conjunto de teste	120
Figura 64 - Varredura dos pontos de operação para a potência ativa de 7 kW.	130
Figura 65 - Varredura dos pontos de operação para a potência ativa de 6,6 kW.	131
Figura 66 - Varredura dos pontos de operação para a potência ativa de 6,2 kW.	131
Figura 67 - Varredura dos pontos de operação para a potência ativa de 5,8 kW.	131
Figura 68 - Varredura dos pontos de operação para a potência ativa de 5,4 kW.	132
Figura 69 - Varredura dos pontos de operação para a potência ativa de 5 kW.	132
Figura 70 - Varredura dos pontos de operação para a potência ativa de 4,6 kW.	132

Figura 71 - Varredura dos pontos de operação para a potência ativa de 4,2 kW.....	133
Figura 72 - Varredura dos pontos de operação para a potência ativa de 3,8 kW.....	133
Figura 73 - Varredura dos pontos de operação para a potência ativa de 3,4 kW.....	133
Figura 74 - Varredura dos pontos de operação para a potência ativa de 3 kW.....	134
Figura 75 - Varredura dos pontos de operação para a potência ativa de 2,6 kW.....	134
Figura 76 - Varredura dos pontos de operação para a potência ativa de 2,2 kW.....	134
Figura 77 - Varredura de pontos de operação extra.....	135
Figura 78 - Varredura dos pontos de operação para a potência ativa de 0 kW. Gerador operando como compensador síncrono.....	135
Figura 79 - Importância dos top 5 atributos ranqueados por <i>Selective</i> para a potência ativa.....	141
Figura 80 - Importância dos top 15 atributos ranqueados por <i>Selective</i> para a potência ativa.....	142
Figura 81 - Importância dos top 30 atributos ranqueados por <i>Selective</i> para a potência ativa.....	142
Figura 82 - Importância dos top 60 atributos ranqueados por <i>Selective</i> para a potência ativa.....	143
Figura 83 - Importância dos top 5 atributos ranqueados por <i>Selective</i> para a potência reativa.....	144
Figura 84 - Importância dos top 15 atributos ranqueados por <i>Selective</i> para a potência reativa.....	144
Figura 85 - Importância dos top 30 atributos ranqueados por <i>Selective</i> para a potência reativa.....	144
Figura 86 - Importância dos top 60 atributos ranqueados por <i>Selective</i> para a potência reativa.....	145
Figura 87 - Importância dos top 5 atributos ranqueados por correlação para a potência ativa.....	146

Figura 88 - Importância dos top 15 atributos ranqueados por correlação para a potência ativa.	146
Figura 89 - Importância dos top 30 atributos ranqueados por correlação para a potência ativa.	146
Figura 90 - Importância dos top 60 atributos ranqueados por correlação para a potência ativa.	147
Figura 91 - Importância dos top 5 atributos ranqueados por correlação para a potência reativa.	148
Figura 92 - Importância dos top 15 atributos ranqueados por correlação para a potência reativa.	148
Figura 93 - Importância dos top 30 atributos ranqueados por correlação para a potência reativa.	148
Figura 94 - Importância dos top 60 atributos ranqueados por correlação para a potência reativa.	149
Figura 95 - Importância dos top 60 atributos ranqueados por correlação para a potência reativa com atributos removidos.	150
Figura 96 - Funções escolhidas para estudo de correlação.	151
Figura 97 - Funções escolhidas para estudo de correlação com adição de ruído.	151
Figura 98 - Coeficientes de correlação para funções sem adição de ruído.	159
Figura 99 - Coeficientes de correlação para funções com adição de ruído.	159
Figura 100 - Cálculo dos coeficientes de correlação para diferentes funções.	161

LISTA DE TABELAS

Tabela 1 – Lista de referências selecionadas da busca sistemática de literatura.....	38
Tabela 2 - Lista ordenada dos pontos de operação para o ponto de potência ativa de 7 kW.....	49
Tabela 3 - Comparação entre os coeficientes de correlação para a potência ativa.....	82
Tabela 4 - Comparação entre os coeficientes de correlação para a potência reativa.....	82
Tabela 5 - Métodos utilizados para o ranqueamento de atributos.....	95
Tabela 6 - Atributos ranqueados em ordem decrescente por <i>Selective</i> e por correlação.....	95
Tabela 7 - Métricas de performance para o modelo de regressão linear baseline no conjunto de validação para a potência ativa.....	106
Tabela 8 - Métricas de performance para o modelo de regressão linear baseline no conjunto de validação para a potência reativa.....	106
Tabela 9 - Métricas de performance para o modelo <i>Lasso</i> no conjunto de validação para a potência ativa.....	107
Tabela 10 - Métricas de performance para o modelo <i>Lasso</i> no conjunto de validação para a potência reativa.....	107
Tabela 11 - Métricas de performance para o modelo <i>Ridge</i> no conjunto de validação para a potência ativa.....	108
Tabela 12 - Métricas de performance para o modelo <i>Ridge</i> no conjunto de validação para a potência reativa.....	109
Tabela 13 - Métricas de performance para o modelo <i>Elastic net</i> no conjunto de validação para a potência ativa.....	110
Tabela 14 - Métricas de performance para o modelo <i>Elastic net</i> no conjunto de validação para a potência reativa.....	110
Tabela 15 - Métricas de performance para o modelo de regressão linear <i>baseline</i> no conjunto de teste.....	112
Tabela 16 - Métricas de performance para o modelo <i>Lasso</i> no conjunto de teste.....	113

Tabela 17 - Métricas de performance para o modelo <i>Ridge</i> no conjunto de teste.	114
Tabela 18 - Métricas de performance para o modelo <i>Elastic net</i> no conjunto de teste.	115
Tabela 19 - Análise de resultados dos modelos de regressão para a potência ativa.....	119
Tabela 20 - Análise de resultados dos modelos de regressão para a potência reativa.	119
Tabela 21 - Pontos de operação utilizados.....	136
Tabela 22 - Valores numéricos dos dados para estudo de correlação.	152
Tabela 23 - Valor médio das variáveis para cálculo de correlação de Pearson.	152
Tabela 24 - Desvios padrão das variáveis para cálculo de correlação de Pearson.....	153
Tabela 25 – Cálculo dos <i>ranks</i> das variáveis.	154
Tabela 26 – Continuação do cálculo dos <i>ranks</i> das variáveis.	155

SUMÁRIO

1	INTRODUÇÃO	16
1.1	OBJETIVOS.....	19
1.1.1	Objetivo geral	20
1.1.2	Objetivos específicos	20
1.2	ESTRUTURA DA DISSERTAÇÃO.....	21
2	BUSCA SISTEMÁTICA DE LITERATURA	22
2.1	OBJETIVO DA BUSCA DE LITERATURA.....	22
2.2	O MÉTODO SSF.....	24
2.3	SÍNTESE DOS RESULTADOS DA BUSCA SISTEMÁTICA DE LITERATURA 27	
2.4	CONSIDERAÇÕES FINAIS.....	39
3	TESTES EM GERADOR SÍNCRONO DE POLOS SALIENTES	41
3.1	OBJETIVO DOS TESTES.....	41
3.2	A BANCADA DE TESTES.....	42
3.3	PROCEDIMENTOS DE ENSAIOS ADOTADOS.....	46
3.3.1	Dados obtidos do sistema de monitoramento VibraOne	49
3.3.2	Dados obtidos do <i>datalogger</i> de temperatura	52
3.3.3	Dados obtidos da Unidade de Monitoramento Móvel (UMM)	53
3.4	TRATAMENTO DOS DADOS OBTIDOS.....	54
3.4.1	Dados dependentes do tempo	58
3.4.2	Dados independentes do tempo	59
3.5	CONSIDERAÇÕES FINAIS.....	60
4	ANÁLISE DOS DADOS DO CAMPO EXTERNO DO GERADOR SÍNCRONO DE POLOS SALIENTES	61
4.1	INFORMAÇÕES EXTRAÍDAS DO CAMPO MAGNÉTICO EXTERNO.....	61
4.1.1	Valor RMS	62
4.1.2	Atributos obtidos do espectro de frequência	64
4.1.2.1	<i>Considerações finais sobre os atributos de frequência</i>	68
4.1.3	Correlação entre potência elétrica e campo magnético externo	70
4.1.3.1	<i>Correlação de Pearson</i>	73
4.1.3.2	<i>Correlação de Spearman</i>	76
4.1.3.3	<i>Correlação PhiK</i>	78
4.1.3.4	<i>Considerações sobre a correlação dos dados</i>	80

4.2	MODELAGEM DOS DADOS DE CAMPO	83
4.2.1	Engenharia de atributos	87
4.2.1.1	<i>PCA – Principal Componente Analysis</i>	89
4.2.1.2	<i>t-SNE – t-distributed Stochastic Neighbor Embedding</i>	91
4.2.1.3	<i>UMAP – Uniform Manifold Approximation and Projection</i>	92
4.2.1.4	<i>Ranqueamento de atributos</i>	94
4.2.1.5	<i>Seleção de atributos</i>	96
4.2.2	Seleção do modelo, treinamento, validação e ajuste de hiperparâmetros	101
4.2.2.1	<i>Modelo de regressão linear Baseline</i>	103
4.2.2.2	<i>Modelo de regressão com penalização L1 (Lasso regression)</i>	106
4.2.2.3	<i>Modelo de regressão com penalização L2 (Ridge regression)</i>	108
4.2.2.4	<i>Modelo de regressão com penalização L1 e L2 (Elastic net)</i>	109
4.2.3	Análise de desempenho no conjunto de teste	111
4.3	BREVE ANÁLISE E CONCLUSÕES DOS RESULTADOS ATINGIDOS ...	116
5	CONCLUSÃO	121
	REFERÊNCIAS	125
	APÊNDICE A – DADOS OBTIDOS DO SISTEMA DE MONITORAMENTO	
	E DADOS DE TEMPERATURA	130
	APÊNDICE B – PONTOS DE OPERAÇÃO	136
	APÊNDICE C – VISUALIZAÇÃO DA IMPORTÂNCIA DE ATRIBUTOS .	141
	APÊNDICE D – ANÁLISE E APLICAÇÃO DOS CÁLCULOS DE	
	CORRELAÇÃO	151

1 INTRODUÇÃO

A capacidade de geração de energia elétrica brasileira reside, em sua maior parte, em usinas hidroelétricas que representam cerca de 53,2% da geração. Dos 46,8% restantes, 21,4% da geração é proveniente de usinas termelétricas, 11,5% de fontes eólicas, 11,8% de fontes solares e 1% de fontes nucleares. O restante da geração pertence ao gás natural, biomassa, derivados de petróleo e carvão [1].

As fontes hídricas e térmicas utilizam o gerador síncrono para converter a energia mecânica produzida por uma máquina primária em energia elétrica. Por este motivo, há uma necessidade de estudos contínuos sobre esse tipo de máquina visto sua importância na geração de energia elétrica no país.

O gerador síncrono é um dispositivo eletromecânico rotativo utilizado para a conversão de energia mecânica em energia elétrica. Esta máquina também pode ser operada como motor, fazendo a conversão contrária de energia: de elétrica para mecânica. Muitas vezes, sua aplicação é fundamental no contexto de geração de energia elétrica.

Ao se tratar de geradores síncronos de grande potência, como os utilizados nas usinas, uma parada não programada pode trazer como consequências a perda do fornecimento de energia elétrica para uma determinada população e a perda de faturamento juntamente com multas para a concessionária.

O gerador síncrono é projetado de maneira a proporcionar a maior confiabilidade possível durante sua vida útil em regime de operação contínua. Seus projetos devem garantir robustez e minimizar os efeitos deteriorantes de seu uso ao longo do tempo [2]. Mesmo sendo projetado para tal tarefa, a operação em regimes contínuos fadiga seus componentes internos que, por sua vez, podem resultar no surgimento de faltas. Atualmente existem empresas geradoras de energia elétrica que possuem geradores com várias décadas de idade. Ao apresentar alguma falha, esses geradores são consertados e colocados novamente em operação. Também é comum que esses geradores sejam modernizados com novas tecnologias. A estratégia de realizar reparos ao ocorrer uma falha é preferível visto o grande valor monetário agregado nestas máquinas. O tempo necessário para manufaturar um novo gerador e os esforços para o transportar e realizar a sua substituição na usina também possuem um impacto financeiro para a empresa [3].

Ao longo dos últimos anos, o Grupo de Concepção e Análise de Dispositivos Eletromagnéticos (GRUCAD) do Departamento de Engenharia Elétrica e Eletrônica

da Universidade Federal de Santa Catarina (UFSC) investiga e estuda métodos não invasivos de monitoramento de geradores síncronos utilizando o campo magnético externo produzido pela máquina durante a sua operação.

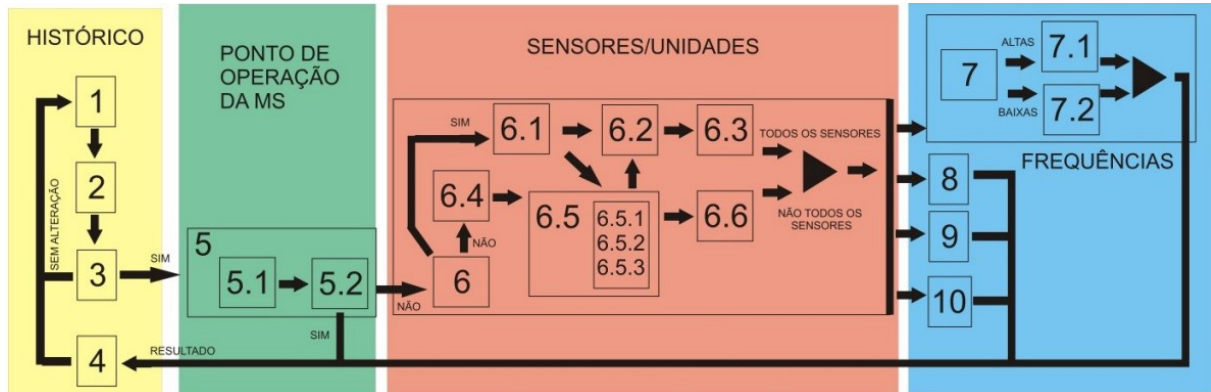
O primeiro trabalho a ser indicado com estudos do campo magnético externo é apresentado em [4]. O autor descreve o desenvolvimento de sensores magnéticos e seus circuitos de condicionamento de sinais para o monitoramento do campo externo. Como resultados, são apresentadas as detecções de alterações no campo magnético externo de motores de indução com barra rompida utilizando o sistema descrito por meio das chamadas componentes sub-harmônicas (harmônicas menores que 60 Hz). Em trabalho posterior [5], o autor estende seu trabalho para o desenvolvimento de um equipamento e uma metodologia para a detecção de faltas em geradores síncronos, também utilizando o campo magnético externo. O equipamento foi utilizado em testes de bancada e em usinas hidroelétricas. Porém, as análises são feitas para as componentes harmônicas para frequências menores que 100 Hz. Em seu trabalho, o autor indica como continuidade de pesquisa o desenvolvimento de uma ferramenta de análise de defeitos com algoritmo inteligente que leve em consideração o ponto de operação da máquina.

O trabalho apresentado em [6] também relata o desenvolvimento de um equipamento para monitoramento e aquisição de dados do campo magnético externo de geradores. A diferença reside no fato de este segundo apresentar em detalhes toda a eletrônica de condicionamento de sinais do sistema, que foi desenvolvido para ficar fixo na usina onde ocorrerá a coleta de dados durante a operação do gerador. Outro ponto a ser diferenciado é que as análises feitas com as componentes harmônicas do campo magnético externo são realizadas para frequências superiores a 100 Hz, atingindo as centenas de hertz. Em trabalho posterior [3], o autor apresenta um método de monitoramento de componentes harmônicas do campo magnético externo e as diretrizes para a detecção de faltas em geradores síncronos, conforme apresentado pela Figura 1.

Nesta figura são apresentados quatro blocos. O primeiro bloco, em amarelo, descreve como deve ser analisado o histórico das componentes harmônicas, levando em consideração as possíveis interferências eletromagnéticas nas componentes do espectro, bem como as variações em cada um dos sensores de campo magnético e seus circuitos de condicionamento. Também é apresentado como proceder com a

análise dos históricos por meio de média móvels em janelas de dias, semanas e meses.

Figura 1 - Fluxograma das diretrizes propostas para detecção e análise de faltas em máquinas síncronas.



Fonte: Adaptado de [3].

O segundo bloco, em verde, descreve como deve ser feito o acompanhamento do ponto de operação do gerador síncrono para eliminar a possibilidade de detecção de alterações no campo que sejam provenientes de variações no ponto de operação da máquina. Isto evidencia a importância de um estudo para caracterização do ponto de operação do gerador síncrono utilizando a informação do campo magnético externo.

O terceiro bloco, em vermelho, descreve como deve ser feita a análise em função da comparação das alterações detectadas nos diferentes sensores localizados ao redor da máquina. Uma vez verificado que a alteração no nível do sinal do campo externo não é relacionada com alterações no ponto de operação da máquina, deve-se analisar sua variação ao longo do tempo e comparar as alterações com os outros sensores.

Por fim, o quarto bloco, em azul, descreve como deve ser feita a análise em função das regiões de frequências que possuem alterações.

Em seu trabalho o autor indica como continuidade de pesquisa avaliar a influência do ponto de operação sobre as harmônicas do campo externo para definir se alguma alteração de amplitude de harmônicas foi causada por variação de operação ou por alguma falta.

A abordagem apresentada em [7] compara técnicas baseadas em dados para detecção de alterações em componentes harmônicas do campo magnético externo, avançando os estudos de detecção de alterações nas componentes de frequência sugerido por [3]. Dentro deste trabalho, as técnicas aplicadas para detecção de

alterações são utilizadas com dados sintéticos obtidos por [8] e também nos dados coletados de usina com o sistema desenvolvido por [6].

As pesquisas publicadas em [9] e [10] apresentam um estudo do desempenho entre sensores de campo magnético instalados dentro do gerador, chamados de bobinas sonda e que são colocados nos dentes da coroa do estator do gerador, e sensores de campo magnético externo instalados na parte externa do gerador. Como resultado, é possível concluir que os sensores de campo externo podem realizar a detecção de faltas incipientes em geradores síncronos tão bem quanto os sensores de entreferro. Em trabalho posterior [11], o autor propõe estratégias de análises no domínio do tempo e da frequência para detecção e diferenciação de faltas incipientes em geradores síncronos por meio do fluxo magnético no entreferro, com o uso de bobinas sonda em conjunto com a corrente de excitação. Como resultado da análise de diferentes indicadores propostos, mostrou-se a possibilidade de detectar, mensurar e diferenciar os defeitos em enrolamentos do estator e de excitação, bem como de excentricidade estática, independentemente do carregamento da máquina. Assim, a referência [11] traz maior robustez na comparação das alterações detectadas nos diferentes sensores instalados em diferentes pontos do gerador e avança os estudos de detecção de alterações nas informações coletadas pelos sensores de campo sugerido por [3].

Dentro dos trabalhos apresentados, existe a necessidade de avaliar e estudar o impacto do ponto de operação no campo magnético externo, conforme verificado nas sugestões de trabalhos futuros e falta de trabalhos sobre o tema até o momento. Por este motivo, o objeto de estudo deste trabalho é o gerador síncrono de polos salientes (GSPS) e sua operação, aprofundando os estudos na diretriz 5 do fluxograma proposto por [3].

1.1 OBJETIVOS

O objetivo principal deste trabalho é dar continuidade ao trabalho de [3], [5], [7], ao aplicar técnicas para estimar o ponto de operação do GSPS utilizando as componentes harmônicas do campo magnético externo geradas durante a operação da máquina. Espera-se obter como resultado desta dissertação primeiramente a viabilidade de aplicar a técnica e obter conclusões sobre as informações relevantes que o campo magnético externo de um gerador pode carregar sobre seu ponto de

operação. Também, espera-se obter um modelo para estimar o ponto de operação do gerador apenas com a informação campo magnético externo, após um treinamento prévio, visando sua aplicação em um sistema de monitoramento de condição.

1.1.1 Objetivo geral

O objetivo geral deste trabalho é continuar com os trabalhos desenvolvidos pelo GRUCAD e contribuir com o estudo e modelagem do ponto de operação do gerador síncrono de polos salientes utilizando apenas a informação contida no campo magnético externo da máquina.

1.1.2 Objetivos específicos

- a) Analisar a relação do ponto de operação com o campo magnético externo de geradores síncronos em artes correlatas a este trabalho.
- b) Avaliar a influência do ponto de operação sobre as amplitudes das componentes harmônicas do campo magnético externo.
- c) Indicar e avaliar as metodologias existentes de análise do campo magnético externo.
- d) Avaliar modelos de aprendizado de máquina para problemas de regressão.
- e) Compreender o funcionamento e operação da bancada para experimentos em geradores síncronos disponível no LABMAQ da UFSC.
- f) Realizar experimentos para coleta de dados na bancada de geradores.
- g) Documentar os procedimentos adotados durante os experimentos.
- h) Realizar o tratamento dos dados coletados durante os experimentos e construir um, ou mais, conjuntos de dados.
- i) Analisar e extrair informação dos dados coletados experimentalmente.
- j) Analisar e comparar o desempenho da(s) técnica(s) de regressão.
- k) Concluir sobre os modelos produzidos com os dados coletados experimentalmente.
- l) Propor e documentar uma metodologia de estimação de ponto de operação de geradores síncronos com a informação coletada do campo magnético externo.

1.2 ESTRUTURA DA DISSERTAÇÃO

O texto desta dissertação de mestrado é dividido em cinco partes. O capítulo 1 é esta introdução que consiste em apresentar a temática abordada durante o texto, apresentar os objetivos desta dissertação e o método que foi utilizado para conduzir a pesquisa. O capítulo 2 apresenta a busca sistemática de literatura que compreende o estudo e pesquisa de artes correlatas ao tema deste trabalho, bem como, relatar o resultado da busca. O capítulo 3 relata os experimentos realizados em uma bancada experimental para a coleta dos dados de campo externo e ponto de operação. Ainda no capítulo 3 é apresentada a metodologia implementada para o tratamento dos dados e como foram obtidos os conjuntos de dados dependentes do tempo e de dados independentes do tempo. O capítulo 4 apresenta a exploração dos conjuntos de dados obtidos no capítulo 3 com investigações sobre a relação existente entre os dados de campo e os dados de ponto de operação. O capítulo 4 também apresenta o procedimento de modelagem do ponto de operação onde os dados de entrada do modelo são provenientes do campo magnético externo. Por fim, enumerada como capítulo 5, aborda-se a conclusão do trabalho, onde são apresentadas, de maneira resumida, as conclusões parciais de cada capítulo e discutida as possibilidades de continuações e melhorias deste trabalho.

O trabalho de pesquisa apresentado nesta dissertação foi motivado e parcialmente custeado no âmbito do programa de P&D ANEEL e desenvolvido no programa de P&D da Engie Brasil Energia e Itá Energética S.A. no projeto PD-00403-0048/2019, intitulado “Equipamento não Invasivo para Detecção de Falhas em Geradores Síncronos através do Campo Magnético Externo” e no projeto PD 0403-0057/2023, intitulado “Equipamento não Invasivo para Detecção de Falhas em Geradores Síncronos através do Campo Magnético Externo – Lote Pioneiro e Inserção no Mercado”, realizado em parceria com a empresa AQTech Engenharia e Instrumentação S.A. A CAPES (Código de Financiamento 001) também custeou parcialmente esta pesquisa.

2 BUSCA SISTEMÁTICA DE LITERATURA

Para o desenvolvimento deste trabalho de pesquisa foi necessário entender o panorama atual do tema em questão. Portanto, uma busca sistemática de literatura será aqui proposta com a finalidade de averiguar os avanços recentes das técnicas de monitoramento de condição de máquina por meio do campo magnético externo. Sendo assim, esta busca foi focada em trabalhos relacionados com o monitoramento do campo magnético externo de um gerador síncrono conectado à rede elétrica com enfoque para o ponto de operação.

Vale comentar que, apesar do escopo deste trabalho ser voltado para máquinas síncronas de polos salientes, tomou-se a liberdade de pesquisar sobre monitoramento de condição por meio do campo magnético externo para as demais máquinas elétricas com o propósito de expandir o conhecimento sobre as técnicas de medição do campo externo utilizadas e as técnicas utilizadas para extrair informação sobre a operação da máquina a partir de dados coletados.

2.1 OBJETIVO DA BUSCA DE LITERATURA

Esta busca tem por objetivo mostrar o ponto de partida da investigação a ser apresentada na seção 3, trazendo respostas sobre alguns pontos importantes para este trabalho, como, por exemplo:

- a) Quais trabalhos utilizam o campo magnético externo para monitoramento de condição de geradores síncronos conectados à rede;
- b) Quais trabalhos utilizam o campo magnético externo para estimar o ponto de operação;
- c) Quais são as técnicas e os sensores utilizados para medir o campo magnético externo de geradores síncronos;
- d) Qual o tratamento, eletrônico ou digital, os dados coletados de campo magnético devem receber para se tornarem utilizáveis;
- e) Quais informações são possíveis de se obter dos dados do campo magnético externo;

Estes cinco tópicos foram utilizados como norteadores para a busca sistemática na literatura publicada.

A busca de literatura geralmente é um trabalho que demanda tempo. Encontrar o escopo correto da busca, tal como um conjunto correto de palavras-chave

ou autores de referência, requer algumas iterações. Por este motivo, para minimizar o tempo despendido ao realizar esta etapa da pesquisa, utilizou-se o método *systematic search flow* (SSF) proposto por [12], onde os autores descrevem uma distinção entre os tipos de revisão de literatura:

- a) Revisão narrativa, também chamada de "busca exploratória", que não tem critérios explícitos e a seleção de artigos é feita de maneira arbitrária. Esta revisão é caracterizada por grande viés do pesquisador.
- b) Revisão sistemática, que é um método de investigação científica com um processo rigoroso e explícito para identificar, selecionar e coletar os dados sobre um tema, seguido da análise dos dados e descrição das contribuições relevantes a pesquisa. Deve ser feito um planejamento inicial seguido de uma síntese dos resultados de várias investigações, e assim, compor um conjunto de estudos originais. Este tipo de revisão é caracterizado pela necessidade de dois ou mais pesquisadores, em que todos devem ler todos os artigos e realizar discussões sobre quais artigos se enquadram no grupo de trabalhos com contribuições originais ao tema da pesquisa.
- c) Revisão integrativa, que é um método cujo objetivo é reunir, explorar e resumir resultados da pesquisa sobre um tema. Este método visa contribuir com o aprofundamento do conhecimento geral do tema de maneira sistemática.

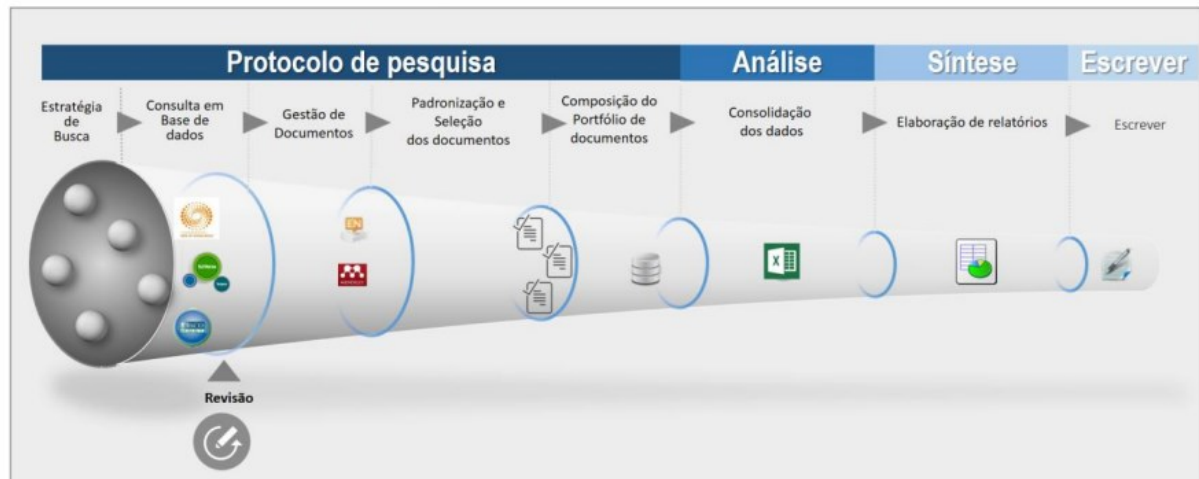
Em todos os tipos de revisão, narrativa, sistemática e integrativa, existe a necessidade de realizar uma busca. Destaca-se que a busca sistemática é um método de investigação científica caracterizado pela sistematização da busca em bases de dados a fim de organizar estudos originais em um portfólio bibliográfico. A busca sistemática é utilizada tanto em uma revisão sistemática quanto em uma revisão integrativa.

O que se pretende com esta seção é apresentar uma revisão integrativa na literatura ao apresentar, de maneira geral, o panorama do estado da arte em monitoramento de condição pelo campo magnético externo em máquinas elétricas. Dentro desta revisão será feita uma busca sistemática.

2.2 O MÉTODO SSF

A utilização do método SSF visa apontar uma maneira sistemática de realizar buscas em bases de dados e auxiliar em revisões sobre o tema de pesquisa, garantindo a repetibilidade, e diminuir a influência do viés do pesquisador no resultado. O método é composto por quatro etapas: o protocolo de pesquisa, a análise, a síntese e a escrita, conforme apresentado na Figura 2

Figura 2 - Composição do método SSF



Fonte: elaborado pelo autor.

Dentro da primeira etapa, a definição do protocolo de pesquisa, existem cinco atividades que representam a grande parte dos esforços da busca sistemática. A primeira atividade foi definir qual a estratégia de busca. A estratégia utilizada foi realizar buscas preliminares e iterativas para encontrar um conjunto de palavras (*strings*) de busca que apresentam aderência ao tema da pesquisa até que os artigos resultantes da busca se apresentem satisfatórios. Ao final da quinta atividade, espera-se obter um portfólio bibliográfico que contenha artigos relacionados com o monitoramento de máquina por campo magnético externo.

Os artigos a serem selecionados devem tangenciar, necessariamente, os três tópicos seguintes:

1. Gerador síncrono com enrolamento de campo.
2. Campo magnético externo.
3. Ponto de operação.

A segunda atividade foi definir qual a base de dados a ser utilizada. A base de dados escolhida foi a Scopus por conter a indexação de várias bases de dados diferentes (ACM Digital Library, Emerald Insight, Hindawi, IEEE Xplore, Inderscience Enterprises, ScienceDirect (Elsevier), Springer Link, Taylor & Francis Online, Wiley

Online Library, World Scientific, dentre outras) se mostrando abrangente o suficiente para uma busca sobre o tema. Pode-se sumarizar a estratégia de busca nos seguintes pontos:

- Base de dados utilizada: Scopus;
- Data da busca: abril de 2024;
- Data das publicações: de 2000 até o momento atual (abril de 2024);
- Língua: inglês;
- A busca foi limitada para artigos, revisões e artigos de conferências;
- A *string* de busca utilizada foi: ("synchronous generator") AND ("monitor*") AND (magnetic AND field AND flux AND (external OR leakage)) AND PUBYEAR > 1999 AND PUBYEAR < 2025 AND (LIMIT-TO (DOCTYPE , "ar") OR LIMIT-TO (DOCTYPE , "re") OR LIMIT-TO (DOCTYPE , "cp")) AND (LIMIT-TO (LANGUAGE , "English")) AND (LIMIT-TO (EXACTKEYWORD , "Synchronous Generators") OR LIMIT-TO (EXACTKEYWORD , "Condition Monitoring") OR LIMIT-TO (EXACTKEYWORD , "Magnetic Fields"))

Os resultados das primeiras iterações de busca utilizaram referências ao ponto de operação na *string* de busca, como por exemplo (*operating AND point*). Apesar de o ponto de operação ser fundamental para esta pesquisa, os documentos encontrados não apresentaram aderência ao tema da pesquisa, portanto, optou-se por retirar as palavras relacionadas ao ponto de operação da *string* de busca.

Em seguida, a busca foi refeita e foram adicionados filtros para palavras-chave como “*Synchronous Generators*”, “*Condition Monitoring*” e “*Magnetic Fields*”.

É importante ressaltar que as palavras-chave, assim como a *string* de busca, foram obtidas após algumas iterações, ou seja, após uma busca inicial. As palavras-chave e a *string* de busca foram sendo refinadas até se obter resultados satisfatórios e julgados aderentes ao tema da pesquisa.

A terceira atividade foi a gestão e a organização dos documentos obtidos pela busca na base de dados. Todos os documentos foram exportados da base de dados Scopus para o *software* Zotero (organizador bibliográfico).

A quarta atividade foi a padronização da seleção dos artigos. Dentro desta atividade foi realizada a leitura dos títulos e dos resumos dos 90 documentos resultantes da busca. Destes, 16 artigos sugerem adesão ao tema da pesquisa. O

critério de escolha das publicações foi feito selecionando artigos que continham em seus títulos textos que os relacionassem com o monitoramento de condição, máquinas elétricas e campo magnético externo, seguido da leitura dos resumos (*abstract*) verificado se existe aderência ao tema da pesquisa. Foi verificado por meio dos resultados da pesquisa realizada que existe um grande esforço da comunidade científica em determinar o comportamento das grandezas de máquinas elétricas para monitorar suas condições de operação e explorar técnicas de detecção de faltas em máquinas. O monitoramento de condição utilizando vibração mecânica da máquina foi tópico recorrente durante a seleção dos trabalhos.

Apesar da *string* de busca conter as palavras "*synchronous generator*", ainda houve a ocorrência de artigos cujo foco da pesquisa fosse motores de indução e transformadores. Esses trabalhos foram considerados inaptos para leitura completa por apresentarem uma máquina de natureza diferente do foco deste trabalho. Alguns outros artigos apresentaram, além do campo magnético externo, outras grandezas avaliadas em paralelo, como, por exemplo, a vibração da estrutura da máquina, correntes e tensões elétricas.

Por fim, na quinta atividade foi realizada a composição do portfólio bibliográfico. Isto envolveu a leitura integral dos artigos, que possibilitou uma última filtragem e verificação se todos os artigos selecionados contêm exatamente os dados e informações inicialmente propostos na estratégia de busca (atividade 1).

A segunda etapa é caracterizada pela análise e consolidação dos resultados da busca. Nesta etapa é possível verificar as lacunas de conhecimento existentes e sugerir pautas para futuras pesquisas sobre o tema de interesse, bem como, levantar os dados bibliográficos encontrados em caso destes serem necessários para a pesquisa.

A terceira etapa compreende a síntese de resultados da pesquisa, cuja atividade realizada foi observar nos trabalhos encontrados as respostas para as indagações iniciais ou alguma outra questão que surgiu em decorrer do amadurecimento do tema de pesquisa ao longo da busca sistemática.

A quarta etapa é caracterizada pela redação do texto com as conclusões da busca sistemática englobando os resultados das etapas 1) protocolo de pesquisa, 2) Análise e 3) síntese.

2.3 SÍNTESE DOS RESULTADOS DA BUSCA SISTEMÁTICA DE LITERATURA

De maneira direta, o resultado da busca sistemática de literatura apresentou o crescente interesse da comunidade científica sobre o tema de monitoramento de máquinas elétricas utilizando o campo magnético externo como grandeza de monitoramento, conforme representado na Figura 3. Na figura é observado uma tendência crescente desde 2008 no número de publicações a cada ano.

Figura 3 - Crescimento do interesse na pesquisa sobre monitoramento de geradores síncronos por meio do campo magnético externo.



Fonte: elaborado pelo autor.

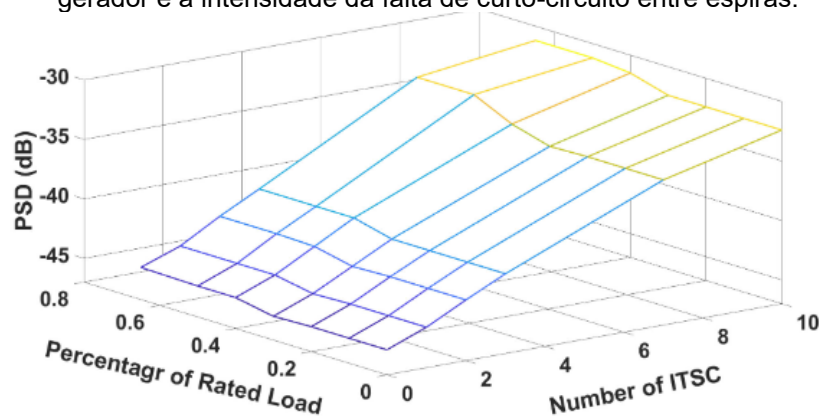
Dentro do conjunto de trabalhos encontrados na busca sistemática, apenas 16 trabalhos possuíam conteúdo relevante para esta dissertação. Dentro do contexto de monitoramento de condição de máquina síncrona, não foi encontrado nenhum trabalho dedicado exclusivamente ao estudo do campo magnético externo e como o ponto de operação influencia o seu comportamento. No entanto, ao longo dos trabalhos apresentados a seguir, são apresentadas algumas análises quanto ao campo externo e sua relação com o ponto de operação.

Nos trabalhos [13], [14] os autores apresentam uma bancada experimental contendo um gerador síncrono de 100 kVA, 14 polos, acionado por um motor de indução de 90 kW. A bancada é utilizada para realizar testes com curto-circuito entre espiras do rotor com possibilidade de reduzir aproximadamente 30% o número de espiras úteis em um polo. O autor apresenta os resultados experimentais e os compara com simulações realizadas para os mesmos casos. Uma conclusão

interessante encontrada em seu trabalho é que para a falta de curto-circuito entre espiras, a densidade espectral de potência (PSD) na banda de 50 Hz até 70 Hz apresenta leve alteração com o carregamento do gerador e uma significativa alteração com a intensidade da falta, conforme apresentado na Figura 4. Nesta figura, PSD significa densidade espectral de potência para a banda de 50 Hz até 70 Hz, “*Percentage of rated load*” significa o carregamento do gerador e “*Number of ITSC*” significa o número de espiras curto-circuitadas durante os experimentos.

Apesar de os resultados parecerem promissores, o autor não evidencia em números de potência ativa e de potência reativa o carregamento do gerador. Sabe-se apenas que foi utilizado um banco de resistores trifásicos que varia de 120Ω até $2,78 \Omega$ e um indutor trifásico fixo de 22 mH como carga para gerador. Desta maneira não se tem os resultados do gerador operando com carga capacitiva e nem com o gerador conectado à rede elétrica.

Figura 4 – Variação da densidade espectral de potência com a porcentagem de carregamento do gerador e a intensidade da falta de curto-circuito entre espiras.



Fonte: retirado de [13], [14].

No trabalho [15], os autores apresentam um algoritmo de detecção de excentricidade do eixo do gerador síncrono por meio do campo magnético externo utilizando a técnica da análise tempo-frequência (STFT – *Short Time Fourier Transform*), sem que haja a necessidade de dados coletados com a máquina operando sem defeitos. A técnica proposta foi inicialmente avaliada com dados de geradores simulados. Posteriormente, a técnica foi avaliada com dados coletados experimentalmente, tanto com um gerador de 100 kVA, quanto com geradores de usinas de 22 MVA e 42 MVA. Todos os experimentos e simulações são realizados com as máquinas operando em vazio. Os autores apresentam um refino de seu trabalho em [16], onde as características das faltas de excentricidade são exploradas

por meio de outras técnicas de processamento de sinais. O trabalho relata que utilizar a entropia de Shanon [17] combinada com a transformada *wavelet* se mostrou ser uma maneira de quantificar a excentricidade com baixa dependência do carregamento do gerador. Outro ponto a ser notado do trabalho é que os sensores indutivos utilizados são posicionados na metade do pacote estatórico do gerador com a área de seção do sensor posicionada para mensurar a componente radial (direção do centro do eixo para fora do núcleo do gerador) do campo magnético externo.

As técnicas de processamento de sinais utilizadas em [16] são descritas e analisadas em [18], sendo elas a análise de média e desvio padrão da tensão gerada nos sensores indutivos, análise no domínio da frequência por PSD (*Power Spectral Density*) e FFT (*Fast Fourier Transform*), análise de energia do sinal, análise de tempo-frequência por meio da transformada *wavelet* e análise da entropia combinada com a transformada *wavelet*.

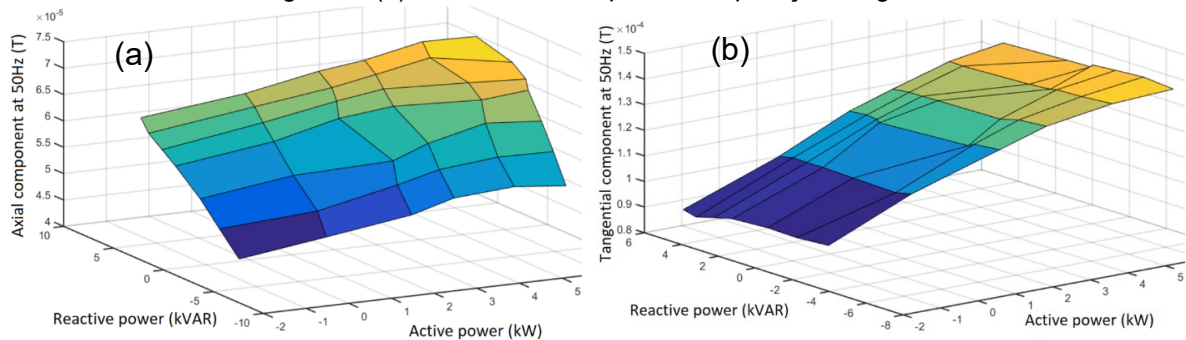
No trabalho [19], o autor apresenta análises baseadas na transformada *wavelet* para comparar a viabilidade da utilização do campo magnético externo ou do campo do entreferro para a detecção de faltas de curto-circuito entre espiras. Como resultado, o autor relata que utilizar o campo do entreferro possui dificuldades práticas na instalação do sensor de efeito hall e que a análise não apresenta grande acurácia, mesmo na presença de faltas severas. Em contrapartida, utilizar o campo magnético externo se mostra prático e apresenta bons resultados, mesmo na presença de faltas leves. Para esta análise o autor está com o gerador conectado à rede elétrica, porém não comenta sobre o carregamento da máquina.

No trabalho [20], o autor apresenta uma análise tempo-frequência para diagnosticar as faltas de curto-circuito entre espiras do rotor e excentricidade dinâmica em geradores síncronos. O autor apresenta um breve comentário sobre o impacto do ponto de operação do gerador no campo magnético externo. Nota-se que a magnitude do campo externo é ligeiramente maior quando comparado com o caso sem carga. Também é observado que o campo magnético externo com o gerador carregado tem uma mudança de fase em comparação com o caso sem carga. No entanto, o conteúdo de frequência do campo magnético externo não se altera devido às condições de carga. Em trabalho posterior [21], os autores apresentam um método para inferência de falta de excentricidade dinâmica e curto-circuito entre espiras do rotor utilizando as amplitudes das harmônicas do campo externo no domínio da frequência obtidas por FFT. O autor concentra as harmônicas utilizadas em baixas frequências (harmônicas

de até 300 Hz). Como um de seus resultados, o autor indica que quando o gerador opera conectado à rede, os espectros de frequência da tensão e da corrente terminais do gerador são sujeitos ao impacto dos harmônicos da rede. No entanto, ele afirma que as harmônicas de interesse para a detecção das faltas mencionadas, obtidas do campo externo, sofrem baixa influência da rede elétrica. Assim, o autor conclui que seu método é robusto às variações de carga na rede e ao ponto de operação do gerador.

No trabalho [22], os autores utilizam a informação contida no campo magnético externo em conjunto com a informação contida na vibração do gerador para detectar curto-circuito entre espiras do rotor. Para medir o campo magnético externo, o autor utiliza um sensor indutivo na direção tangencial no centro do pacote do gerador. Um ponto importante a ser notado é que o autor realiza suas análises com o espectro em frequência obtida por FFT para harmônicas do campo da ordem de até 1600 Hz, contrastando com trabalhos mencionados anteriormente. Em trabalhos posteriores [23], [24], os autores expandem sua análise para faltas de curto-circuito entre espiras do estator, bem como expande os sensores de campo externo para captar as componentes radial, tangencial e axial do campo. O trabalho apresenta uma análise matemática de como as componentes do campo magnético externo devem se comportar na presença de uma falta e com a máquina saída. Os autores também dedicam parte do trabalho para ressaltar a importância de se estudar o ponto de operação do gerador síncrono dentro do contexto de detecção de faltas e concluem que existem algumas harmônicas que são mais sensíveis a variações de carga, conforme apresentado na Figura 5. Nesta figura, a superfície representada na esquerda (a) é a magnitude da componente axial da densidade de fluxo magnético obtida do sensor indutivo e a superfície da direita (b) é a magnitude da componente tangencial da densidade de fluxo magnético obtida também com sensor indutivo, porém com sua área de seção apontando para outra direção.

Figura 5 - Variação da amplitude da densidade de fluxo magnético para a componente axial (a) e tangencial (b) de 50 Hz com o ponto de operação do gerador.

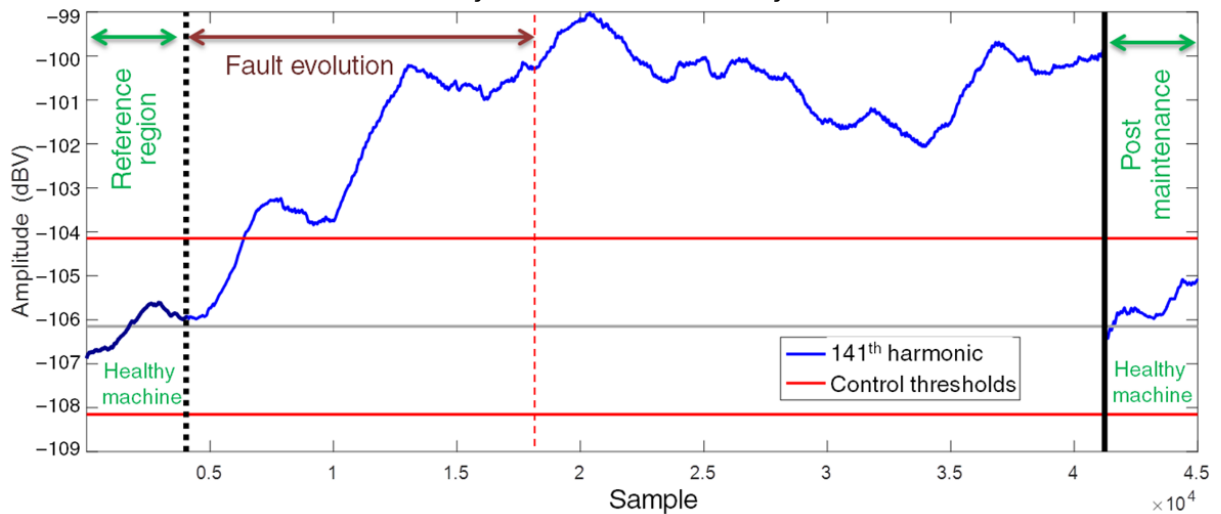


Fonte: retirado de [24].

Os autores de [23], [24], descrevem a dificuldade de identificar os pontos de funcionamento em todo o mapa de operação, em termos de potência ativa e potência reativa, do gerador síncrono de uma forma não intrusiva. De fato, a assinatura do campo magnético externa medida por um sensor varia com a potência ativa e a potência reativa, conforme observado na Figura 5. É por isso que a assinatura magnética pode traduzir uma mudança de carga ou um defeito que aparece. Assim, para separar as duas possibilidades, é necessário identificar o ponto de funcionamento da máquina. Desta maneira, o autor conclui que antes de prosseguir com as discriminações das faltas é necessário se associar cada ponto de operação com uma assinatura do campo externo da máquina sadia.

No trabalho [25], os autores propõem a aplicação de uma estratégia simples e de baixo custo computacional para o tratamento de dados oriundos do campo magnético externo para efeitos de monitoramento de condição da máquina e detecção de faltas. O método é apresentado e avaliado experimentalmente em uma bancada de testes que permite a imposição diferentes níveis de falta em um gerador síncrono de polos salientes de 10 kVA conectado à rede elétrica. O método também é validado com a utilização de dados coletados em um gerador de usina de 305 MVA conectado à rede elétrica. O método consiste em monitorar as amplitudes das componentes harmônicas do campo magnético externo obtidas do espectro em frequência (FFT), por meio de um histórico de amplitudes. Assim, quando algumas harmônicas têm suas amplitudes desviadas a ponto de sair de uma zona de controle, o sistema indica a possibilidade de uma falta. A Figura 6 apresenta a evolução de uma falta mecânica ao longo de uma janela de 3 meses observada por meio do incremento na amplitude da 141° harmônica da frequência de giro mecânico do gerador.

Figura 6 - Histórico da amplitude da 141ª componente harmônica da frequência mecânica do gerador durante a evolução de uma falta de vibração mecânica.



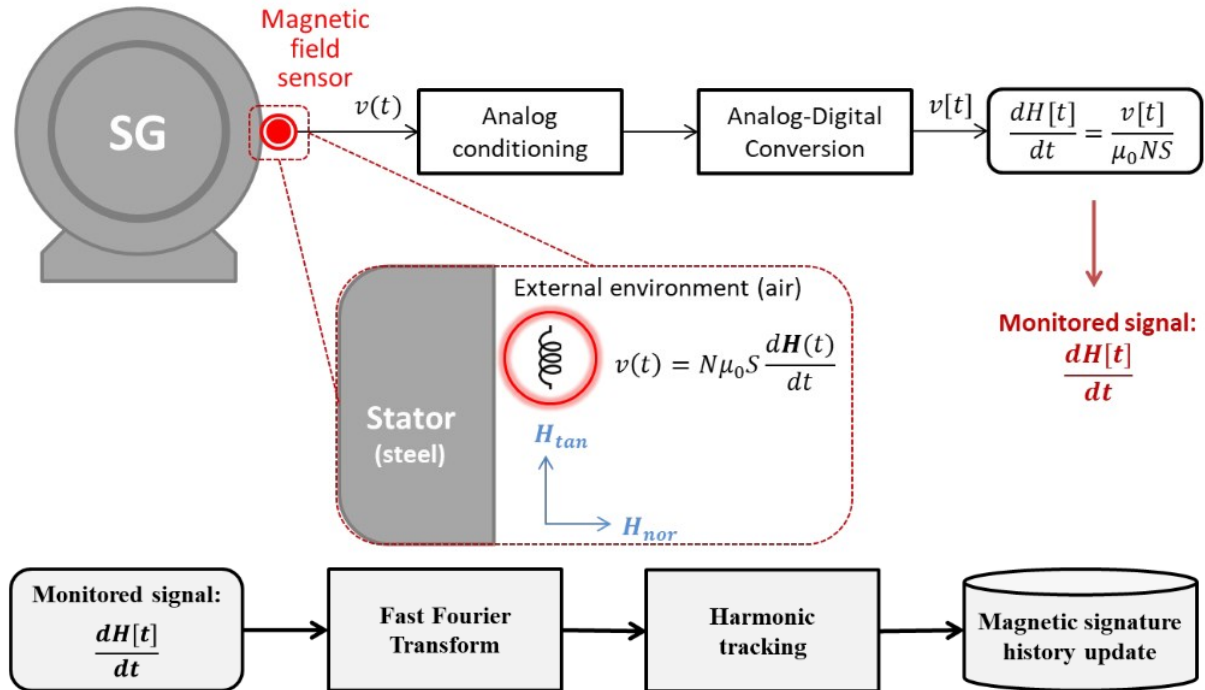
Fonte: retirado de [25].

Também é possível observar, próximo ao final do período de monitoramento, o momento em que é realizada uma manutenção corretiva no gerador. Após a manutenção, a amplitude da harmônica volta ao patamar normal. De acordo com o autor, esta harmônica apresenta maior sensibilidade à falta do que à variação do ponto de operação. Em outras palavras, apesar de a harmônica também ser sensibilizada pelo ponto de operação, as variações causadas pela falta apresentaram maior magnitude. Por fim, os autores concluem que a carga tem um efeito significativo nos históricos que compõem a assinatura magnética do gerador, podendo ocultar o efeito de faltas incipientes que evoluem lentamente ao longo do tempo, dificultando a eficácia da metodologia proposta. Por outro lado, para a falta investigada, os resultados revelaram que a aplicação do método junto com uma estratificação do ponto de operação do gerador (isto é, a avaliação da máquina é realizada dentro de faixas de operação) aumenta a eficiência do método.

No trabalho [26], os autores apresentam um equipamento para monitoração da taxa de variação do campo magnético externo. Os autores afirmam que o campo externo possui características inovadoras com grande potencial de aplicação na manutenção preditiva de geradores síncronos principalmente por ser um sistema de fácil instalação e não invasivo. O equipamento descrito apresenta a capacidade de detectar faltas no estator ou no rotor, tanto de origem elétrica/magnética quanto mecânica. As tecnologias empregadas são detalhadas na proposta deste sistema, incluindo a especificação do sensor de campo magnético, os circuitos de condicionamento de sinais, as técnicas de processamento de dados e os métodos de

detecção de faltas. A Figura 7 apresenta a arquitetura do sistema proposto. Neste trabalho, os autores propõem o termo “metodologia de monitoramento do comportamento da assinatura magnética continuada”.

Figura 7 – Arquitetura do sistema (equipamento e tratamento de dados) proposto para monitoramento do campo externo.



Fonte: retirado e adaptado de [26].

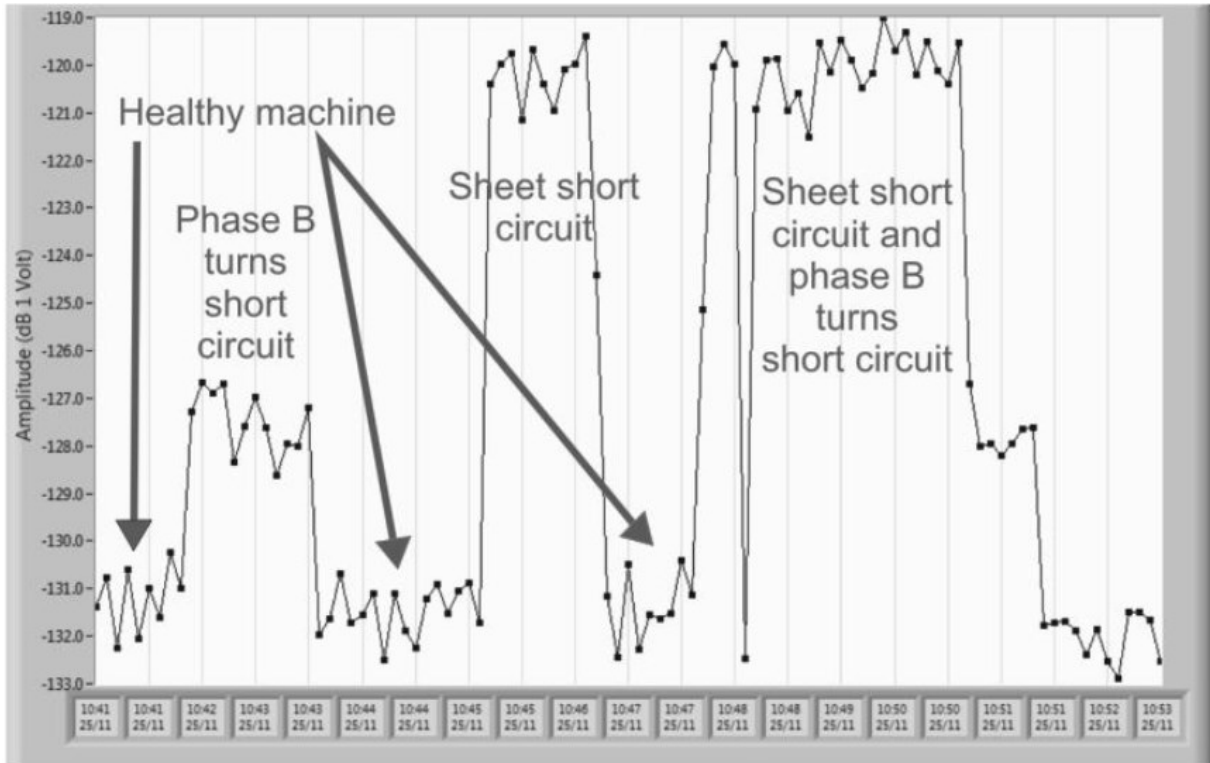
O sistema é composto pelo equipamento junto com o processamento dos dados adquiridos. De maneira geral, os autores propõem o emprego de um sensor indutivo colocado próximo ao gerador para mensurar o campo magnético externo tangencial. O sinal deste sensor é a taxa de variação do campo magnético externo que passa por um tratamento analógico e, por fim, é digitalizado. Os sinais da taxa de variação no tempo do campo magnético externo tangencial são analisados no domínio da frequência empregando a sua FFT. As amplitudes das componentes harmônicas do espectro são coletadas e o vetor de dados que contém essas amplitudes é guardado em um banco de dados. Por fim, o autor conclui que a detecção assertiva de faltas em sua fase inicial permite planejar as ações de manutenção do gerador. Por outro lado, devido ao seu carácter de monitoramento em longos períodos de tempo, o sistema proposto não é adequado para fins de proteção do gerador.

No trabalho [27], o autor apresenta o desenvolvimento de um equipamento para monitoramento do campo magnético externo de um gerador síncrono. Os dados monitorados pelo sistema são as amplitudes das harmônicas oriundas do espectro em

frequência da taxa de variação do campo externo, também chamado em outros trabalhos como assinatura magnética do gerador. Em seu trabalho, os autores apresentam uma análise analítica do comportamento do campo do entreferro do gerador quando existe uma falta de curto-circuito entre espiras em um polo do rotor e valida os resultados analíticos por simulação. Com o equipamento de monitoramento de campo externo desenvolvido, testes em bancada foram realizados para verificar como o sistema reage na presença de faltas elétricas e mecânicas. O tipo de falta mais facilmente detectado é de problemas no rotor. Como resultado dos testes experimentais, o autor apresenta a evolução da harmônica de 862,5 Hz, que corresponde à 69ª harmônica de giro mecânico de um gerador de 8 polos. O gerador empregado é de 10 kVA. Esta harmônica foi escolhida por apresentar a maior sensibilidade para as faltas de curto-circuito entre espiras de um dos polos do rotor (campo do gerador) e curto-circuito entre chapas do estator.

A Figura 8 mostra a evolução da amplitude para a componente harmônica de 862,5 Hz, em função das medições ao longo do tempo com a imposição de faltas. Durante o intervalo de tempo apresentado, o gerador está operando em tensão nominal, 50 Hz e com uma carga resistiva de 70% do seu valor nominal. As faltas incipientes inseridas foram: i) curto-circuito de 10% das voltas de um dos enrolamentos das fases do estator empregando um resistor que limita o valor da corrente de curto-circuito a 7% do seu valor nominal, ii) curto-circuito das chapas (uma parte do núcleo do estator) e iii) curto-circuito de parte da fase do estator e das chapas simultaneamente.

Figura 8 - Evolução da amplitude da componente harmônica de frequência de 862,5 Hz em função das medições ao longo do tempo para o gerador de 8 polos operando a 70% de sua carga nominal, tensão nominal e 50 Hz.



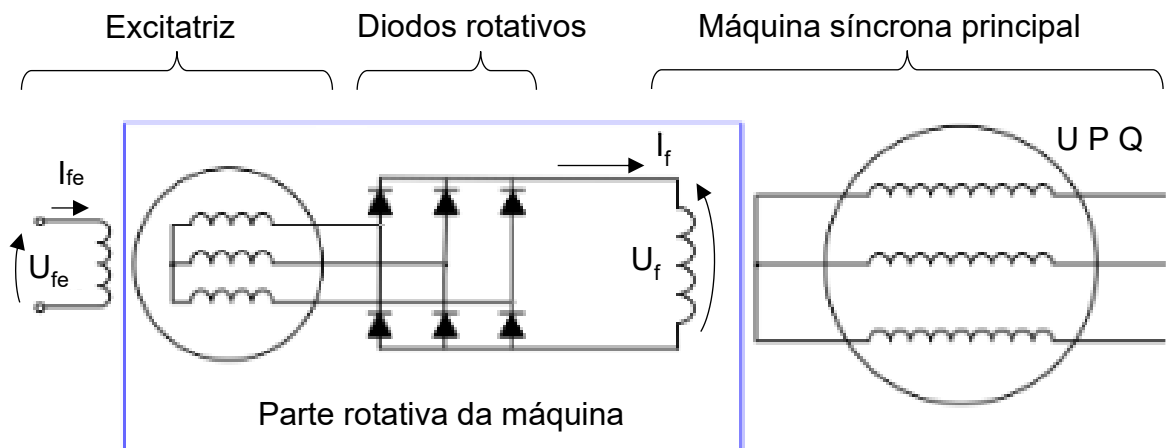
Fonte: retirado de [27].

Na figura, é possível observar que as duas faltas geram diferentes níveis de perturbação na amplitude da harmônica. No entanto, quando as faltas são inseridas simultaneamente, a modificação da amplitude não resulta numa variação distinta em relação ao caso em que apenas as chapas do estator estão na condição de curto-circuito. Com esses resultados, os autores concluem que há a necessidade da avaliação simultaneamente de outras componentes de frequência utilizando, por exemplo, um programa de inteligência artificial ou outras técnicas de monitoramento. Por fim, o autor menciona que o efeito no comportamento do campo magnético externo de todas as faltas deve ser estudado, tanto para os casos com o gerador operando com apenas um tipo de falta, quanto para os casos em que o gerador opera com várias faltas diferentes ocorrendo simultaneamente. Outro grande foco é diferenciar perturbações causadas por variações no ponto de operação causadas por uma falta.

No trabalho [28], os autores focam seu estudo no campo magnético externo gerado pelo sistema de alimentação do campo do gerador sem o uso de escovas. Os sistemas de excitação do campo de geradores síncronos de rotor bobinado se dividem em duas classes: Uso de escovas e anéis, em que uma fonte DC alimenta diretamente

o campo do gerador por meio das escovas; uso de uma excitatriz, em que existe uma pequena máquina síncrona acoplada ao eixo do gerador principal e que serve para gerar um campo DC de maneira a induzir tensão nas bobinas do rotor desta máquina secundária, conforme apresentado pela Figura 9. Por isso, um conjunto de diodos retificadores é necessário para transformar a tensão AC induzida em DC para alimentar o campo da máquina principal. Esses diodos são chamados de diodos rotativos por ficarem no rotor da máquina. A Figura 9 apresenta o esquemático simplificado de uma máquina síncrona que utiliza o sistema de excitação do campo sem escovas.

Figura 9 - Esquemático simplificado de um sistema de alimentação do campo do gerador sem o uso de escovas (*brushless excited synchronous machine*).



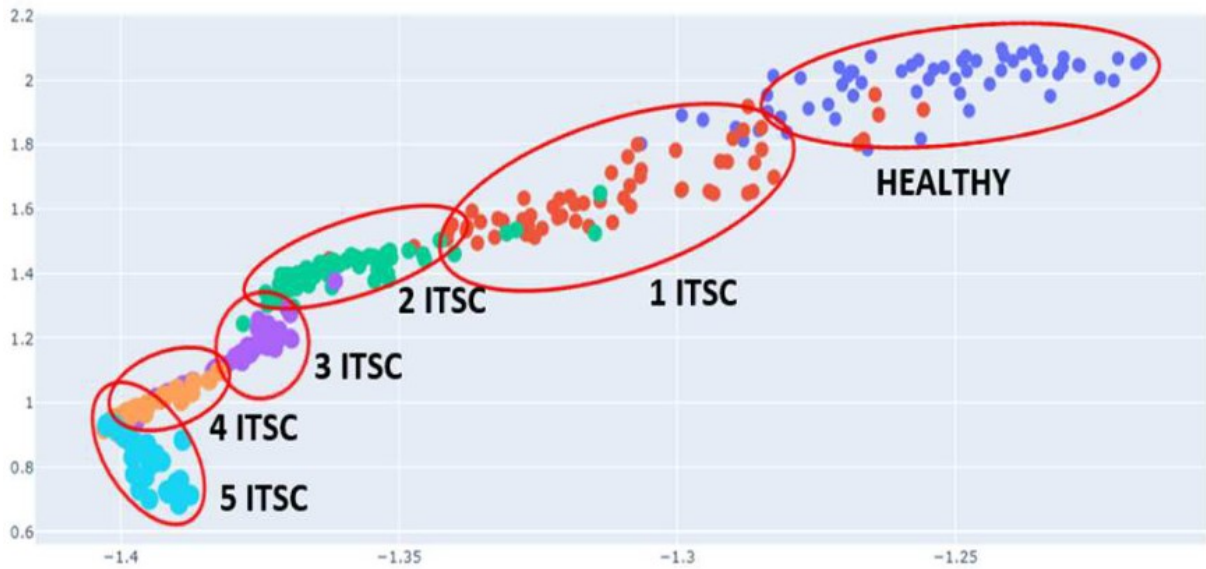
Fonte: retirado e adaptado de [28].

O autor realizou experimentos com o gerador conectado à rede elétrica e em diferentes pontos de operação. Um ponto importante é que o autor apresenta os valores de potência ativa e reativa utilizados em sua análise $\{(0 \text{ kW}, 1 \text{ kvar}), (0 \text{ kW}, -1 \text{ kvar}), (1 \text{ kW}, 1 \text{ kvar}), (1 \text{ kW}, -1 \text{ kvar})\}$. Os experimentos foram realizados para a máquina sadia, com diodos rotativos em curto e com diodos em aberto. Observa-se que em funcionamento normal, as correntes fornecidas pelos enrolamentos da armadura da excitatriz são equilibradas. A harmônica do campo externo que mais se sobressai corresponde à ondulação da tensão da ponte de diodos. Esta frequência é um produto dos impulsos da ponte de diodos (entrada e saída de condução) e da frequência da excitatriz. Por outro lado, em condições de defeito, as correntes fornecidas pelos enrolamentos da armadura da excitatriz são desequilibradas. O efeito direto deste desequilíbrio é que o fluxo parasita da excitatriz muda notavelmente em relação à condição sã. Os resultados mostram que a detecção de um diodo queimado na condição aberto e de um diodo queimado em curto-circuito é realizada pela análise

harmônica do campo magnético externo próximo à excitatriz. Se ocorrer uma falha de um díodo aberto, os harmônicos notáveis são múltiplos da frequência da excitatriz, enquanto se ocorrer uma falha de um díodo em curto-circuito, os harmônicos são o dobro da frequência da excitatriz. Uma conclusão importante apresentada pelo autor é que a posição do sensor indutivo não tem impacto significativo nas análises das harmônicas. Para o caso específico de análise dos diodos rotativos da excitatriz é possível utilizar o campo axial, radial ou uma combinação de ambos. De acordo com o autor, a combinação de ambos (sensor posicionado a 45°) provou ser o posicionamento mais sensato, uma vez que as variações no campo axial e radial são capturadas.

No trabalho [29], os autores apresentam uma metodologia para detectar falta de curto-circuito entre espiras do rotor do gerador síncrono. A metodologia apresentada é baseada em *Variational AutoEncoder* (VAE). O autor utiliza as amplitudes das harmônicas do espectro em frequência do campo magnético externo de um gerador sadio de 325 MVA, 56 polos. Para obter dados com sinal de faltas de curto-circuito entre espiras, o autor utilizou o trabalho de [14] como referência para gerar dados sintéticos de falta para o gerador. Em resumo, os dados sintéticos da máquina em falta são gerados a partir dos dados reais para máquina sadia, em que a amplitude das harmônicas, sensibilizadas pela falta, é multiplicada por uma constante que depende da severidade da falta. Apesar de o autor utilizar alguns pontos de operação diferentes, nenhuma conclusão é apresentada em seu trabalho. Como resultado, ao utilizar o método VAE, o autor consegue agrupar os dados de acordo com a severidade da falta de curto-circuito entre espiras do rotor do gerador, conforme apresentado pela Figura 10. Na Figura 10, o resultado obtido pela aplicação do método VAE no conjunto de dados, ITSC (*inter-turn short circuit*) significa curto-circuito entre espiras e o número de prefixo indica a quantidade de espiras curto-circuitadas. Observa-se que os dados são agrupados de acordo com a severidade da falta. Os dados são apresentados em um chamado espaço latente (*latente space*) de duas dimensões. As dimensões do espaço latente não apresentam interpretação física, servindo apenas para visualização da informação do conjunto de dados. Dados que possuem características próximas são agrupados próximos uns dos outros dentro deste espaço de duas dimensões. O método apresentado pelos autores se mostra capaz de quantificar a severidade de faltas, após um treinamento prévio.

Figura 10 - Resultado da aplicação do método VAE para o conjunto de dados (dados reais de máquina sadia e dados sintéticos de máquina com falta).



Fonte: retirado de [29].

No trabalho [30], os autores apresentam uma análise do campo magnético externo tangencial para monitoramento da saúde do gerador. O trabalho se concentra no monitoramento do campo externo de um gerador de usina de 37 MVA por meio de um sensor indutivo para a componente tangencial do campo. Uma avaliação da amplitude da frequência fundamental mecânica do gerador para a falta de curto-circuito entre espiras do rotor para diferentes pontos de operação é apresentada. Com esse resultado, os autores concluem que, apesar de a falta modificar a amplitude da harmônica, o ponto de operação também possui forte influência na amplitude. Portanto, em análise de faltas de curto-circuito entre espiras do rotor se deve observar pontos de operação de alta potência ativa e reativa.

A Tabela 1 apresenta um resumo das referências selecionadas. Os pontos principais que foram observados dentro dos documentos foram:

1. O tipo de sensor utilizado para monitorar o campo magnético externo.
2. Qual o objetivo do monitoramento.
3. Se estudo apresenta algum comentário ou conclusão sobre a relação entre o ponto de operação e o campo magnético externo. Coluna “P.O.? ”.
4. Se o gerador síncrono monitorado está conectado à rede elétrica. Coluna “Rede?”

Tabela 1 – Lista de referências selecionadas da busca sistemática de literatura.

Autor	Ref.	Tipo de sensor	Objetivo do monitoramento	P.O.?	Rede?
-------	------	----------------	---------------------------	-------	-------

H. Ehya <i>et al.</i>	[13], [14]	Indutivo	Utilizar o campo externo para detectar curto-circuito entre espiras.	Não	Não
H. Ehya <i>et al.</i>	[15]	Indutivo	Utilizar o campo externo para detectar excentricidade estática, dinâmica e conjunta.	Não	Sim
H. Ehya <i>et al.</i>	[16]	Indutivo	Utilizar o campo externo para detectar excentricidade estática, dinâmica.	Não	Não
H. Ehya <i>et al.</i>	[19]	Indutivo / efeito hall	Utilizar o campo externo para detectar curto-circuito entre espiras de rotor.	Não	Sim
H. Ehya <i>et al.</i>	[20]	Indutivo	Utiliza a informação contida no campo externo para diferenciar os tipos de faltas no gerador.	Sim	Sim
H. Ehya <i>et al.</i>	[21]	Indutivo	Utiliza a informação contida no campo externo para inferir sobre a existência de faltas no gerador.	Não	Sim
M. Cuevas <i>et al.</i>	[22]	Indutivo	Utilizar o campo externo para detectar curto-circuito entre espiras de rotor.	Sim	Sim
M. Cuevas <i>et al.</i>	[23], [24]	Indutivo	Utilizar o campo externo para detectar curto-circuito entre espiras de rotor e estator.	Sim	Sim
Luis O. S. Grillo <i>et al.</i>	[25]	Indutivo	Utilizar o campo externo para monitoramento da detecção de faltas no gerador por meio do histórico de harmônicas do campo.	Sim	Sim
Luis O. S. Grillo <i>et al.</i>	[26]	Indutivo	Utilizar o campo externo para monitoramento da saúde do gerador por meio do histórico de harmônicas do campo.	Sim	Sim
H. F. dos Santos <i>et al.</i>	[27]	Indutivo	Utilizar o campo externo para monitorar e coletar dados do campo magnético.	Sim	Sim
P.Tian <i>et al.</i>	[28]	Indutivo	Utilizar o campo externo para monitorar e a saúde dos diodos da excitatriz sem escovas.	Sim	Sim
H. Bechara <i>et al.</i>	[29]	Indutivo	Utilizar o campo externo para detectar curto-circuito entre espiras de rotor.	Não	Sim
R. Romary <i>et al.</i>	[30]	Indutivo	Utilizar o campo externo para detectar curto-circuito entre espiras de rotor.	Sim	Sim

Fonte: elaborado pelo autor.

2.4 CONSIDERAÇÕES FINAIS

A revisão integrativa apresentada neste capítulo demonstrou a crescente relevância do tópico estudado. Diferentes trabalhos apresentam sistemas e técnicas desenvolvidas para monitorar o campo magnético externo. Dentro destes trabalhos vários métodos foram apresentados para extrair informação do campo externo por meio de seu espectro em frequência. De maneira geral, os trabalhos utilizam tanto o campo magnético, indução magnética e a taxa de variação do campo magnético como grandeza monitorada.

O gerador síncrono, assim como as demais máquinas elétricas rotativas, apresenta três componentes de campo externo: axial, tangencial e radial (também é

encontrado na literatura como campo normal). Essas três componentes são estudadas e analisadas dentro da literatura especializada. Indiferente da componente do campo a ser monitorada, os trabalhos mencionam a dificuldade de se posicionar o sensor ao redor do gerador, principalmente quando o gerador está instalado em uma usina. Qualquer variação de relutância no percurso do fluxo magnético altera seu caminho percorrido, e assim, altera a tensão de saída do sensor de campo.

Grande parte dos trabalhos encontrados analisam tanto o campo externo quanto o campo de entreferro. Em geral, as conclusões convergem para o fato de que é possível extrair a mesma informação de ambos os campos com a vantagem de o campo externo ser um método não invasivo.

A maioria dos trabalhos apresentam resultados com faltas aplicadas de maneira individual no gerador. Poucos apresentam e relatam sobre a sobreposição de faltas. Portanto esse é um tópico que ainda carece de informação na literatura.

Apesar de alguns trabalhos mencionarem em suas análises a importância de entender a influência do ponto de operação para a conclusão de seus resultados, nenhuma investigação rigorosa ou análise foi encontrada sobre como o campo magnético externo se comporta com a variação do ponto de operação.

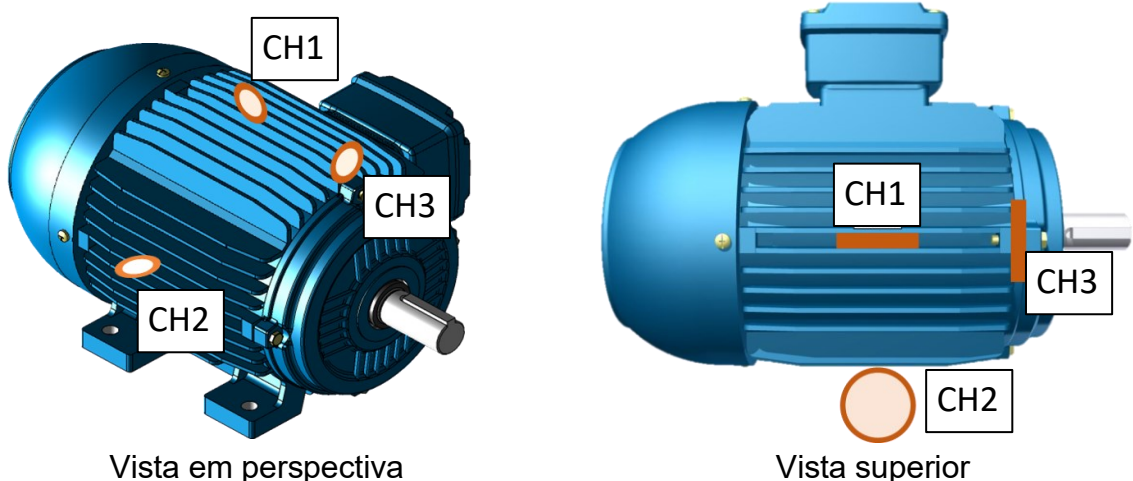
3 TESTES EM GERADOR SÍNCRONO DE POLOS SALIENTES

O primeiro passo para iniciar o entendimento do comportamento do campo magnético externo do gerador em relação ao ponto de operação é realizar um mapeamento com coleta de dados do campo magnético externo e da potência do gerador variando o ponto de operação. Dentro deste trabalho foi proposto um conjunto de testes para obtenção de um conjunto de dados de campo do GSPS de oito polos do grupo gerador especial do Laboratório de Máquinas e Acionamentos Elétricos (LABMAQ) do Departamento de Engenharia Elétrica e Eletrônica da Universidade Federal de Santa Catarina. Este conjunto de dados em mãos possibilita a obtenção de resultados quanto à modelagem do comportamento do campo magnético externo com a variação do ponto de operação.

3.1 OBJETIVO DOS TESTES

Os testes foram propostos de maneira a mapear o campo magnético externo do GSPS em três pontos distintos do exterior da máquina: 1) Campo tangencial no centro do pacote, 2) Campo tangencial no centro do pacote defasado de 90° e 3) Campo axial na cabeça de bobina, conforme observado na Figura 11, respectivamente indicados como CH1, CH2 e CH3.

Figura 11 – Ilustração da disposição dos sensores indutivos ao redor do GSPS.



Fonte: elaborado pelo autor.

Os dados do ponto de operação da máquina devem ser adquiridos simultaneamente com os dados de campo magnético externo para conseguir traçar

um mapa com os eixos de potência ativa, potência reativa e o campo magnético externo.

Outras grandezas, entendidas como grandezas de suporte, também devem ser adquiridas simultaneamente aos dados do campo magnético, que apesar de não serem o foco deste trabalho, podem ser úteis durante a interpretação e estudo do conjunto de dados e para análises futuras. Alguns exemplos de grandezas aqui definidas de suporte são: a corrente de campo, a corrente de armadura, a tensão terminal e a temperatura da máquina. Com os testes realizados, os dados obtidos são agrupados em um conjunto com informação do campo para cada ponto de operação, bem como as grandezas de suporte.

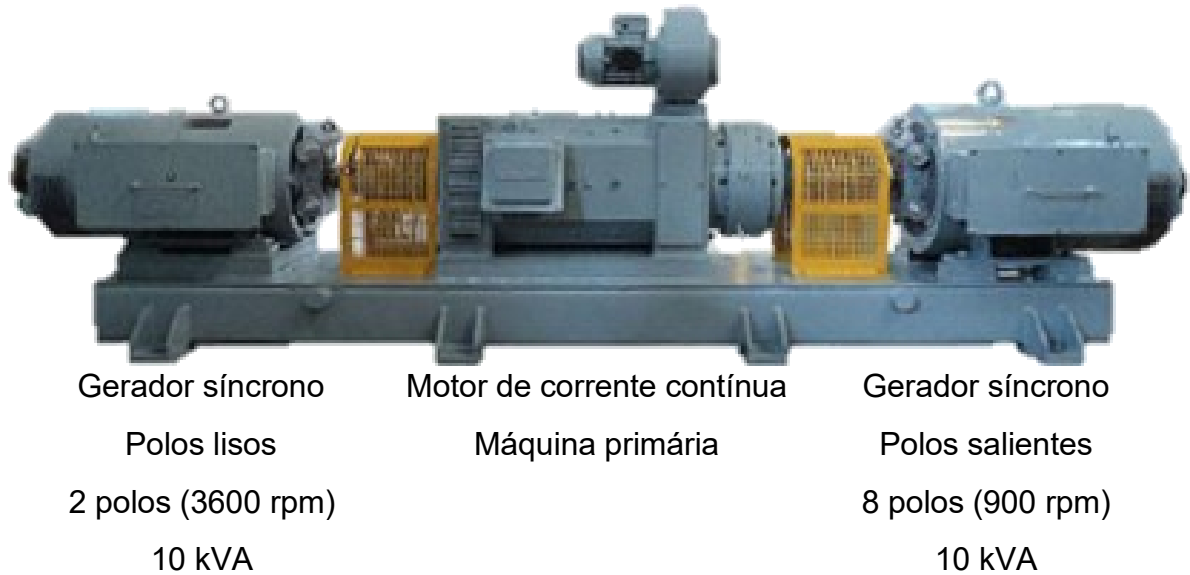
O objetivo dos testes no GSPS é construir um conjunto de dados que possibilite modelar o comportamento do ponto de operação do GSPS a partir de informações contidas no campo magnético externo medido. Ao final dos testes e tratamento dos dados, espera-se obter dois conjuntos de dados: um conjunto de dados que contenha a informação do campo magnético no tempo, representado por formas de onda em uma janela de medição; e um outro conjunto de dados que contenha as informações que não dependam do tempo, as quais serão utilizadas para criar modelos capazes de prever o ponto de operação da máquina em função do campo magnético externo medido.

3.2 A BANCADA DE TESTES

O LABMAQ possui uma bancada com um grupo gerador especial para testes em geradores síncronos desenvolvida especialmente para estudos investigativos de detecção e análise de faltas incipientes.

O grupo gerador é composto por dois geradores síncronos, um de polos salientes e outro de polos lisos. Ambas as máquinas compartilham seus eixos com um motor de corrente contínua que serve como máquina primária. As características desta bancada permitem a possibilidade de inserir faltas elétricas e mecânicas em ambos os geradores, e são apresentadas em detalhes nos trabalhos [3], [5], [11]. A Figura 12 apresenta a bancada com os dois geradores e a máquina primária.

Figura 12 -Grupo gerador especial.



Fonte: elaborado pelo autor.

A bancada possui apenas controle manual, um para controlar a corrente de campo do gerador (regula a tensão terminal do gerador) e outro para controlar a corrente de campo da máquina primária (regula a velocidade do conjunto). A imagem do painel de controle está apresentada na Figura 13. Do painel é possível acompanhar os valores de tensão e corrente em cada fase, bem como a frequência da tensão nos terminais do gerador. O sistema de acionamento do grupo gerador contempla acionamentos sincronizados com a rede elétrica ou desconectado com carga local.

Figura 13 - Painel de controle do grupo gerador especial.

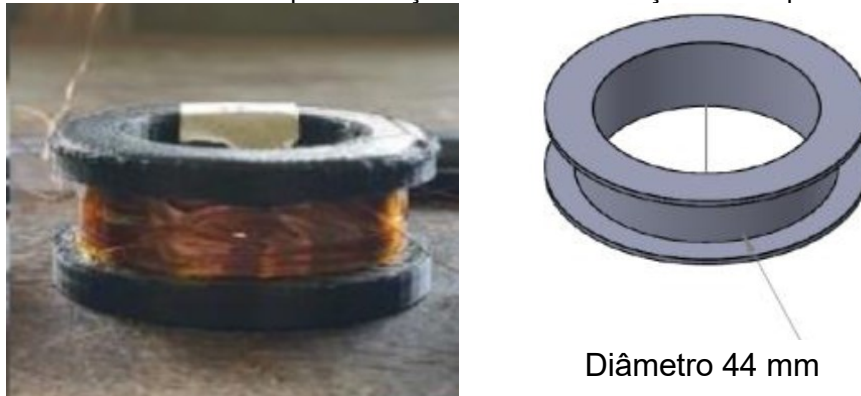


Fonte: elaborado pelo autor.

A instrumentação da bancada fornece maneiras de adquirir os dados de operação do grupo gerador. Um componente da instrumentação é o sistema de monitoramento VibraOne, fornecido pela empresa AQTech Engenharia e Instrumentação S.A., sediada em Florianópolis/SC, que está conectado à bancada e realiza a aquisição e o tratamento de dados de vibração em tempo real com o sistema SCADA [31]. Conectado ao sistema VibraOne está um medidor das grandezas elétricas do gerador. As grandezas elétricas são adquiridas continuamente junto com as grandezas de vibração, formando um único conjunto de dados que pode ser acessado por meio do sistema SCADA.

Para a medição do campo externo, emprega-se um sistema chamado de Unidade de Medição Móvel (UMM), desenvolvido em [11]. A Figura 15 apresenta a UMM e seus sensores conectada a um computador para monitoramento de dados. Ainda nesta figura é apresentado de maneira sucinta o funcionamento da UMM. Quatro sensores podem ser utilizados com esta versão da UMM. O sensor responsável pela medição do campo magnético externo é de formato circular com aproximadamente 4,4 cm de diâmetro [9], [10], conforme apresentado na Figura 14. A UMM possui internamente uma placa de aquisição da National Instruments USB-6361 BNC, conforme visto no canto inferior direito da Figura 15.

Figura 14 - Sensor indutivo utilizado para medição da taxa de variação do campo magnético externo.

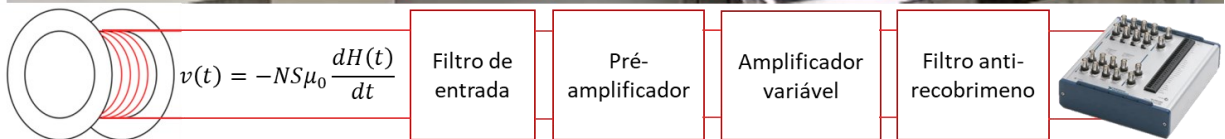
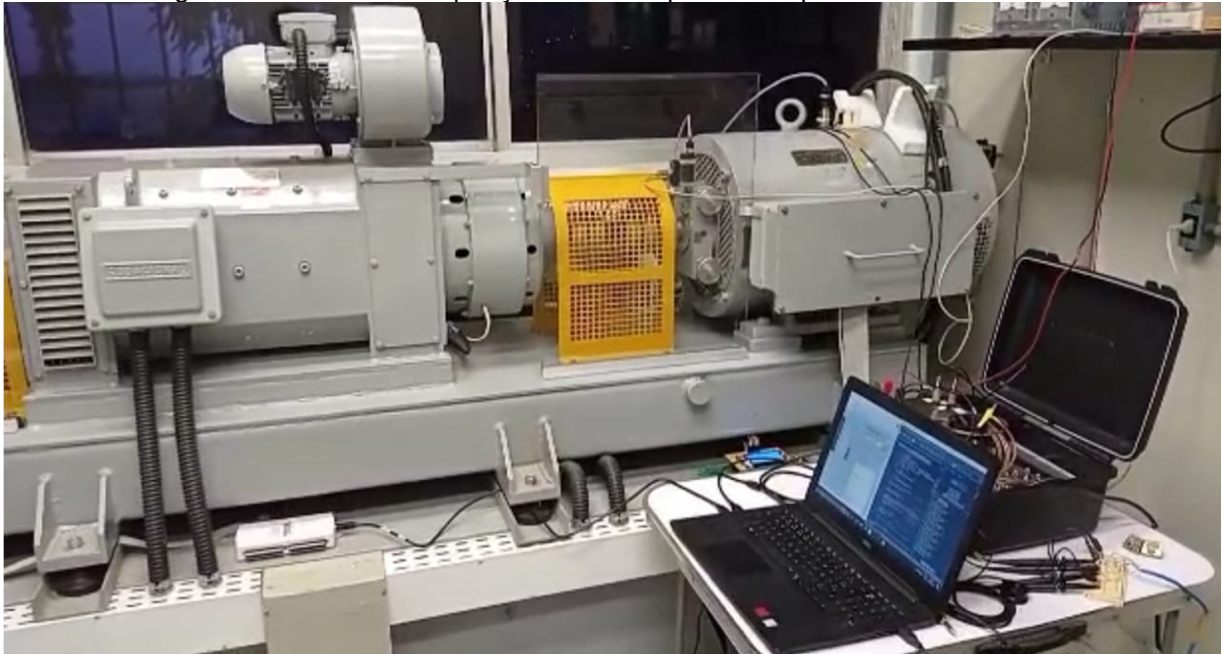


Fonte: retirado e adaptado de [9].

Este sistema possui características semelhantes ao sistema MagAnalyzer, desenvolvido originalmente pela UFSC e, atualmente em conjunto com a empresa AQTech Engenharia e Instrumentação S.A., o qual está em processo de industrialização do lote pioneiro. As análises são realizadas com sinais proporcionais à derivada temporal do campo magnético externo. Ou seja, os sinais dos terminais dos sensores indutivos amplificados e tratados eletronicamente com filtros

adequados. Detalhes podem ser encontrados nos trabalhos de [3], [5], [11] e nos artigos [25], [27].

Figura 15 - Sistema de aquisição de dados para o campo externo do GPS.



Fonte: elaborado pelo autor.

Sabe-se que a forma de onda da tensão induzida nos terminais do sensor é proporcional à taxa de variação do campo magnético que atravessa a área de seção do sensor, à permeabilidade do material ao qual o campo magnético permeia (neste caso o ar) e ao número de espiras do sensor, conforme a equação (1).

$$v(t) = -\mu_0 SN \frac{dH(t)}{dt} \quad (1)$$

Onde:

- t é o instante de tempo.
- $v(t)$ é a tensão induzida nos terminais do sensor indutivo.
- μ_0 é a permeabilidade magnética do vácuo.
- S é a área de seção do sensor indutivo.
- N é o número de espiras do sensor indutivo.
- $H(t)$ é a componente resultante do campo magnético que atravessa a área de seção do sensor indutivo.

A equação (1) relaciona a tensão induzida nos terminais do sensor indutivo com a taxa de variação do campo magnético multiplicado por um fator de escala ($-\mu_0SN$). Observa-se que a natureza do sensor não o permite mensurar campos contínuos, apenas alternados.

Uma limitação prática do sistema durante a realização dos testes é que a aquisição de todos os dados não estão sincronizados e possuem taxas de amostragem diferentes. O sistema do VibraOne adquire os pontos de operação a uma taxa de 1 amostra a cada 10 segundos, o sistema de aquisição de temperatura adquire dados de temperatura a uma taxa de 1 amostra a cada 10 segundos e o sistema de aquisição do campo magnético adquire a uma taxa de 125 000 amostras por segundo em janelas de 5 segundos. Os dados foram sincronizados *a posteriori*.

3.3 PROCEDIMENTOS DE ENSAIOS ADOTADOS

Para criar um conjunto de dados confiável e robusto foi necessário estabelecer métodos a serem utilizados durante a coleta, bem como descrever quais foram os procedimentos adotados para criar o conjunto final de dados. Como o objetivo deste trabalho depende desses dados, tomou-se um cuidado especial com os processos de medição e de tratamento.

Em um primeiro momento, foram levantados dois pontos que estiveram fora do controle do operador durante os testes: 1) a tensão terminal e 2) a temperatura da máquina. Com a máquina conectada à rede elétrica, a tensão terminal do gerador segue a tensão imposta pelo barramento da rede, pois a potência do gerador utilizado neste trabalho é insignificante perante a do sistema elétrico. A tensão terminal ainda é sujeita a leves perturbações na frequência fundamental de 60 Hz e a oscilações na amplitude da tensão terminal causada por adição e retirada de cargas, além de mudança nos taps de regulação de transformadores de subestações próximas. A temperatura do gerador influencia os valores de seus parâmetros internos. Além disso, tendo em vista que este gerador é “pequeno”, variações na temperatura ambiente também podem provocar alterações em sua temperatura de uma maneira rápida o suficiente para a ocorrência perturbações adicionais. Dentre as variações dos parâmetros estão incluídas alterações nos valores das resistências dos enrolamentos de cobre do estator e do rotor, bem como mudança das propriedades magnéticas do núcleo. Conforme a temperatura aumenta, os parâmetros do GSPS e da máquina

primária se modificam, sendo necessário ajustes suaves ocasionais no campo do gerador (potência reativa) e no campo da máquina primária (potência ativa). Desta maneira, a janela de tempo escolhida para coleta de dados não deve ser muito longa de maneira a atenuar possíveis influências das perturbações na tensão terminal, da variação de velocidade e, principalmente, dos efeitos da variação de temperatura.

Antes de iniciar qualquer coleta de dados, o gerador é colocado para operar em um patamar de carga com alta potência ativa e alta potência reativa a fim de aquecer a máquina até seu ponto de regime térmico. Enquanto a máquina aquece, as suas grandezas elétricas e a sua temperatura são monitoradas. Por experiência de testes anteriores de natureza similar, sabe-se que a temperatura da carcaça do GSPS em regime permanente fica na faixa de 38°C a 40°C para uma temperatura ambiente de 35°C. Uma vez que a máquina atinja esse patamar de temperatura, prossegue-se com a imposição do ponto operação desejado e coleta de dados.

O controle da temperatura ambiente foi feito com o controle do fluxo de ar na sala do grupo gerador de tal maneira que ao permitir a troca de ar com o ambiente externo a temperatura caía, e contrariamente interrompendo este fluxo de ar, a temperatura aumente.

Uma vez que a máquina atinge o nível de temperatura estabelecido, segue-se o procedimento de coleta dos dados do maior valor de potência reativa para o menor valor. Para cada ponto de operação a máquina foi regulada para o valor desejado de geração e deixada operar em uma janela de tempo de aproximadamente 1 minuto.

Tomou-se especial cuidado para abranger a maior quantidade possível de pontos de operação dentro da capacidade de geração do GSPS. Os pontos foram escolhidos de maneira a cobrir a maior parte da curva de capacidade do GSPS. Preferencialmente, foram escolhidos pontos de alta potência ativa, de maneira a imitar a operação típica de um gerador de uma usina hidroelétrica. Outra consideração foi o tempo necessário para obter tais dados com o sistema descrito do grupo gerador. Os pontos de operação coletados experimentalmente são apresentados no apêndice B.

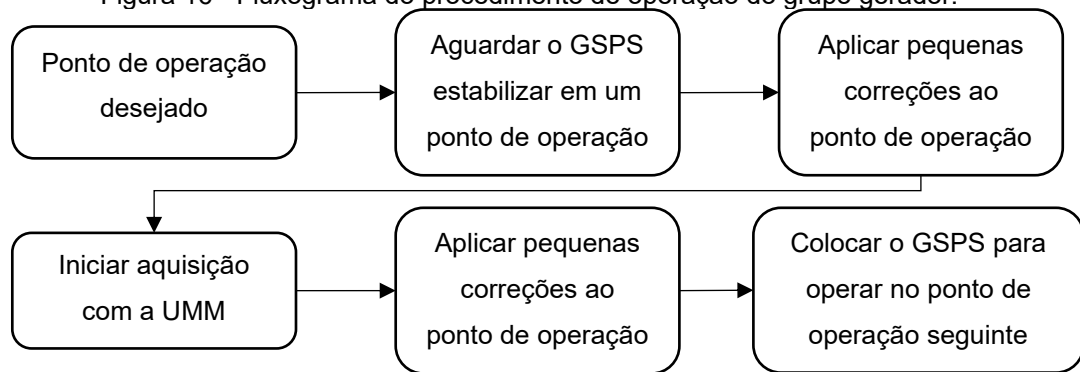
Os valores de potência ativa e reativa junto com as demais grandezas elétricas do gerador foram adquiridos com o VibraOne. O campo magnético externo da máquina foi adquirido com a UMM, e os valores de temperatura com um *dataloger* Yokogawa GP-10.

Após o GSPS atingir a faixa de temperatura de regime permanente, a ordem utilizada para coleta de dados segue os procedimentos descritos a seguir:

1. Máquina é colocada para funcionar no ponto de operação desejado.
2. Espera-se a máquina estabilizar em um ponto de operação próximo ao desejado.
3. Corrige-se o ponto de operação com ajustes finos até que se atinja o ponto de operação próximo suficiente ao desejado. Adotaram-se diferenças máximas de 50 W para a potência ativa e de 50 var de potência reativa em relação aos valores do ponto desejado para o procedimento de ajuste fino.
4. Quando a máquina estabiliza no ponto de operação suficientemente próximo ao desejado, é iniciada a aquisição dos dados de campo com a UMM por pelo menos um minuto. A aquisição de dados do ponto de operação e de temperatura são ininterruptos por todo o tempo em que o GSPS está sob ensaio.
5. Deve-se observar atentamente o ponto de operação atual. Se necessário, aplicam-se ajustes finos no ponto de operação para corrigir desvios causados pela temperatura, ou reinicia-se a aquisição se houver algum degrau de tensão indesejável no barramento ao qual o GSPS está conectado.
6. Após finalizar uma janela de aquisição, a máquina é posta a operar em outro ponto desejável próximo ao ponto anterior, voltando-se ao passo "1".

A Figura 16 apresenta um fluxograma do procedimento adotado para operação do grupo gerador e coleta de dados. Para exemplificar, a Tabela 2 apresenta a ordem na qual foram adquiridos os dados apenas para os pontos de potência ativa mantida em 7 kW.

Figura 16 - Fluxograma do procedimento de operação do grupo gerador.



Fonte: elaborado pelo autor.

Tabela 2 - Lista ordenada dos pontos de operação para o ponto de potência ativa de 7 kW.

Sequência de operação	Ponto de operação		Sequência de operação	Ponto de operação	
	Potência ativa	Potência reativa		Potência ativa	Potência reativa
1	7000 W	600 var	17	7000 W	- 400 var
2	7000 W	5600 var	18	7000 W	- 800 var
3	7000 W	5200 var	19	7000 W	- 1200 var
4	7000 W	4800 var	20	7000 W	- 1600 var
5	7000 W	4400 var	21	7000 W	- 2000 var
6	7000 W	4000 var	22	7000 W	- 2400 var
7	7000 W	3600 var	23	7000 W	- 2800 var
8	7000 W	3200 var	24	7000 W	- 3200 var
9	7000 W	2800 var	25	7000 W	- 3600 var
10	7000 W	2400 var	26	7000 W	- 4000 var
11	7000 W	2000 var	27	7000 W	- 4400 var
12	7000 W	1600 var	28	7000 W	- 4800 var
13	7000 W	1200 var	29	7000 W	- 5200 var
14	7000 W	800 var	30	7000 W	- 5600 var
15	7000 W	400 var	31	7000 W	- 6000 var
16	7000 W	0 var			

Fonte: elaborado pelo autor.

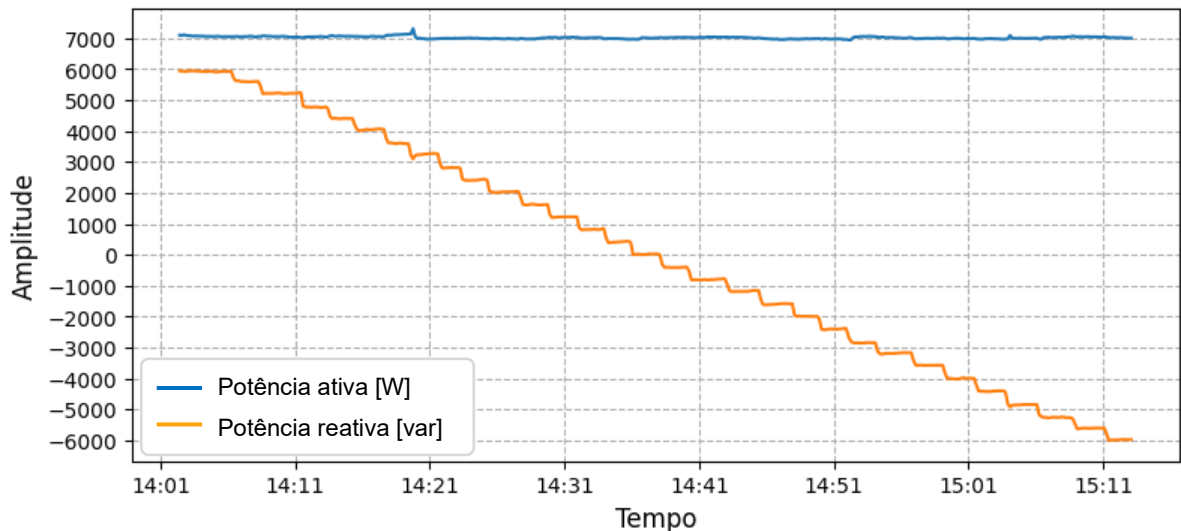
3.3.1 Dados obtidos do sistema de monitoramento VibraOne

Dentre os dados fornecidos pelo VibraOne, foram selecionadas somente as grandezas elétricas de interesse para este trabalho. As grandezas obtidas foram: potência aparente, potência ativa, potência reativa, tensão terminal de armadura (valor médio da tensão de linha das três fases), frequência da tensão terminal (valor de frequência de uma das fases), corrente de armadura (valor médio da corrente de linha das três fases) e corrente de campo.

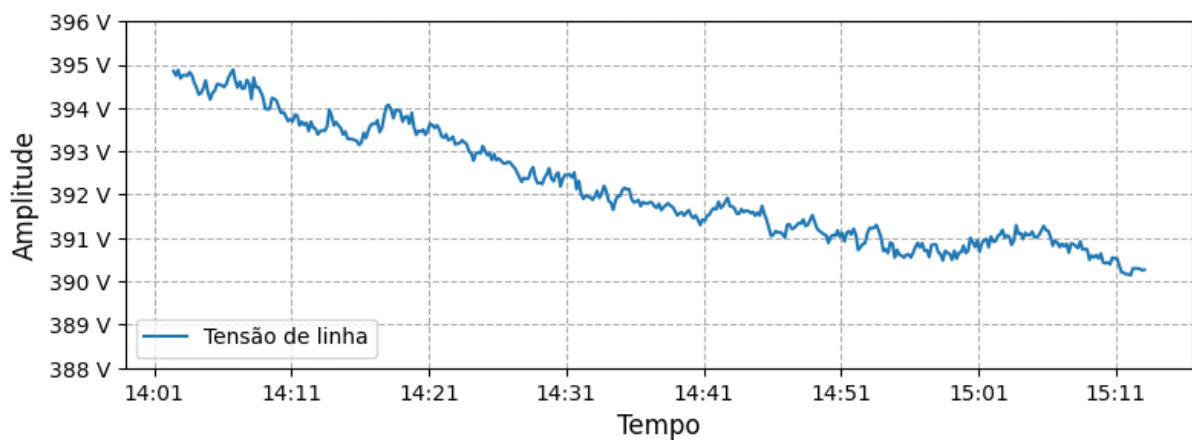
Na Figura 17, as grandezas elétricas obtidas do sistema VibraOne estão representadas para o ponto de potência ativa de 7 kW mantido constante (curva na

cor azul na Figura 17). Com este gráfico se exemplifica o modo como os valores do ponto de operação foram impostos ao GPS variando a potência reativa de 6 kvar até -6 kvar, mantendo constante a potência ativa em 7 kW, diminuindo-se o campo do gerador. Informa-se como curiosidade que este processo de varredura com 31 pontos de potência reativa mantendo a potência ativa constante demorou um pouco mais de uma hora (vide Figura 18 mostrada mais adiante). A Figura 17(a) apresenta a evolução das potências ativa e reativa, a Figura 17(b) a amplitude da tensão terminal, a Figura 17(c) a frequência terminal, a Figura 17(d) a corrente de linha e a Figura 17(e) a corrente de campo. Note que a Figura 17(d) é uma das curvas “V” típicas de máquinas síncronas.

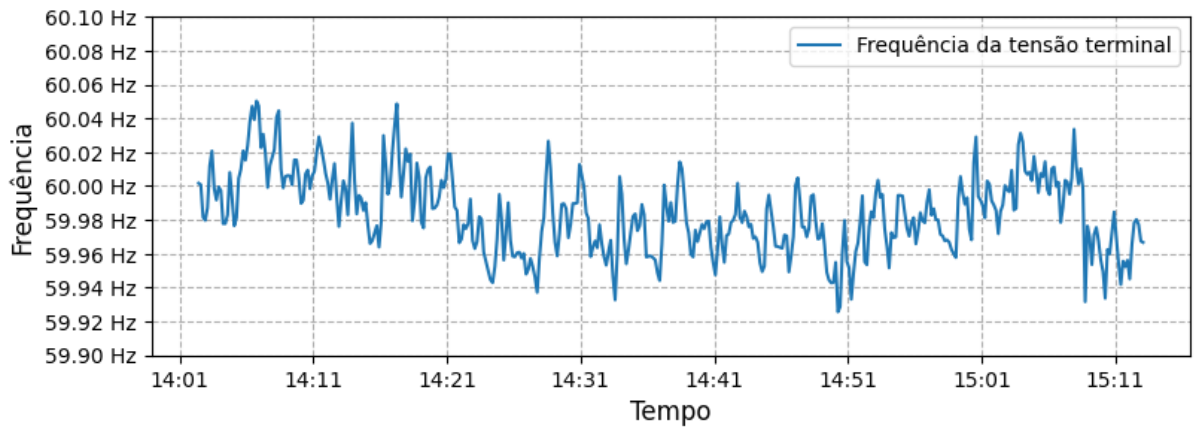
Figura 17 - Representação das grandezas elétricas para a potência ativa de 7 kW com dados obtidos do sistema de monitoramento VibraOne.



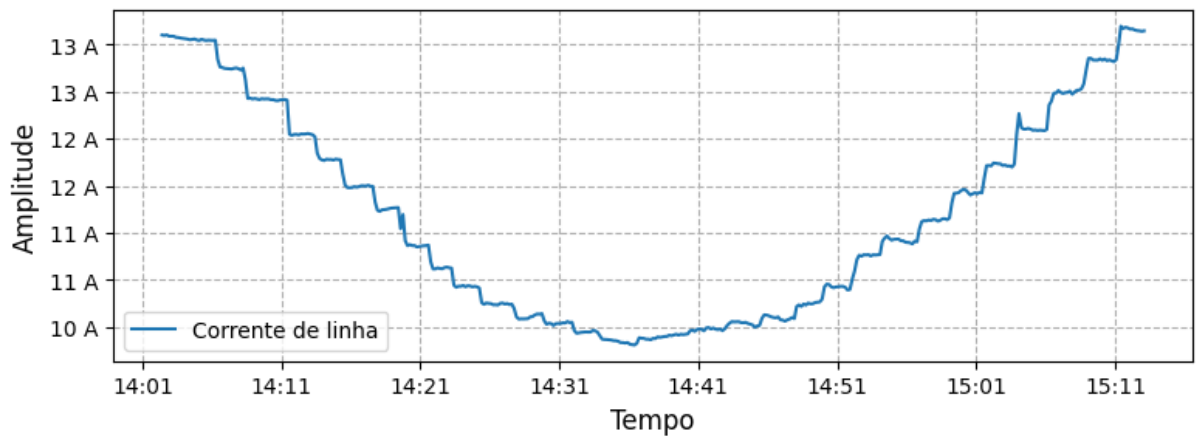
(a)



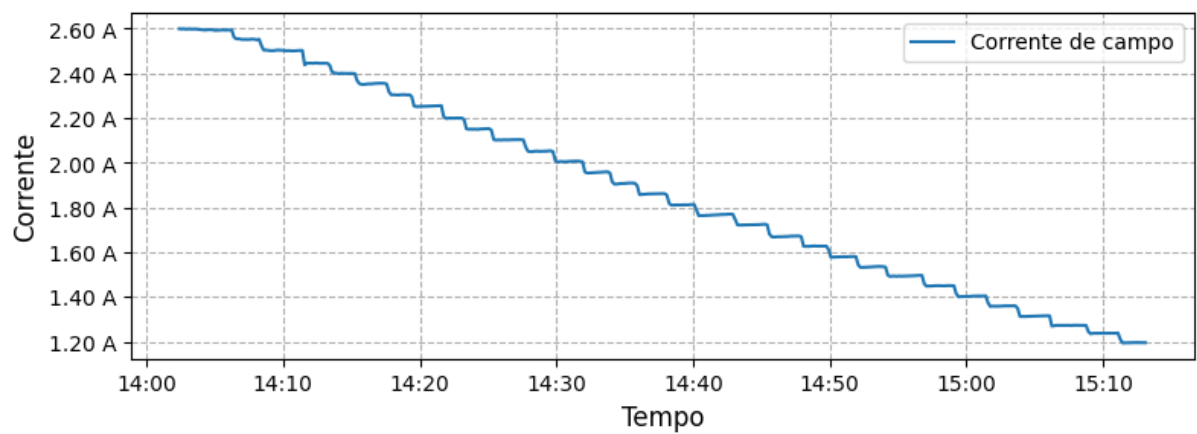
(b)



(c)



(d)



(e)

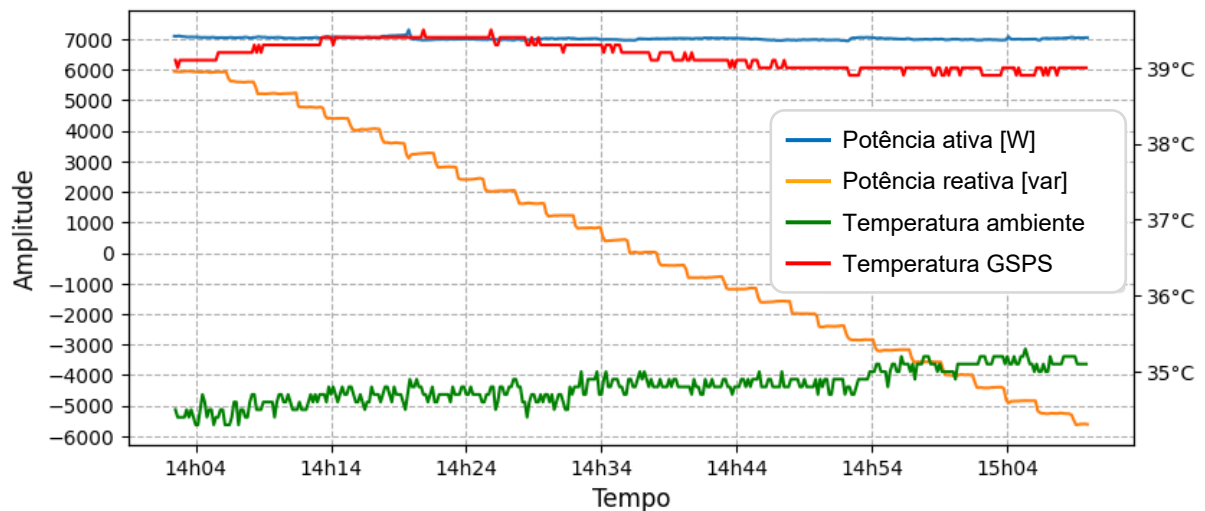
Fonte: elaborado pelo autor.

No Apêndice A - Dados obtidos do sistema de monitoramento VibraOne estão apresentados os valores obtidos em forma de gráficos similares aos da Figura 17.

3.3.2 Dados obtidos do *datalogger* de temperatura

Os tipos de dado obtidos pelo sistema de monitoramento de temperatura foram a temperatura ambiente (termopar colocado sobre uma das prateleiras da bancada, longe da influência do ar expelido pelo gerador) e temperatura da carcaça do gerador (termopar preso com fita encostando na carcaça, na superfície exterior, do GSPS, aproximadamente na metade de seu comprimento axial). Esses dados de temperatura foram adicionados e sincronizados com os dados do ponto de operação. O fato de serem sincronizados não implica em obter valores em um mesmo instante de tempo (perfeitamente sincronizados). Porém, garante-se que os dados obtidos de temperatura dentro de uma janela de aquisição sejam coerentes com os dados de pontos de operação. A Figura 18 apresenta os pontos de operação da Figura 17 adicionando os pontos de temperatura do GSPS e de temperatura ambiente em função tempo do período do ensaio.

Figura 18 – Representação gráfica dos dados de ponto de operação apresentados com dados de temperatura.



Fonte: elaborado pelo autor.

Observa-se nos comportamentos das curvas de temperatura mostrados na Figura 18 que, mesmo com um baixo controle sobre a temperatura, mantém-se a temperatura razoavelmente constante ao longo do ensaio. Detalhadamente, ocorre certo aumento da temperatura do GSPS nos pontos de maior potência reativa, pois a corrente de armadura da máquina possui valores maiores. Conforme o GSPS é colocado para operar em menores patamares de potência reativa, sua temperatura vai diminuindo. Porém, a temperatura volta a aumentar para potências reativas relativamente baixas. Isso pode ser explicado ao observar o gráfico da corrente de

linha da Figura 17. O perfil de temperatura do GPS para os demais ensaios segue o mesmo padrão.

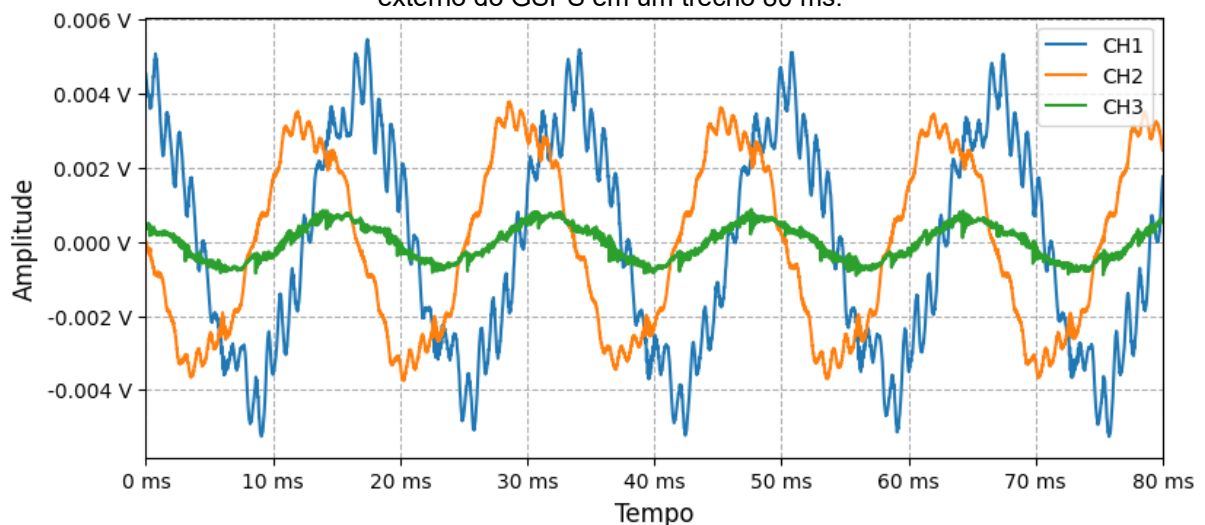
O Apêndice – A Curvas de ponto de operação e temperatura apresenta gráficos similares ao da Figura 18 para os demais pontos de operação.

3.3.3 Dados obtidos da Unidade de Monitoramento Móvel (UMM)

Os dados relativos ao campo magnético obtidos pela UMM naturalmente possuem características de medição e de aquisição diferentes dos valores dos dados das grandezas relacionadas diretamente ao ponto de operação e da grandeza temperatura. Para o campo magnético, é necessário que haja uma taxa de aquisição elevada para obter as informações contidas nas altas frequências. A janela de tempo para das aquisições também é importante para aumentar a resolução espectral da FFT.

A Figura 19 apresenta um trecho de 80 ms de medição do sinal proporcional à taxa de variação no tempo do campo magnético externo do GPS para os três sensores instalados, nominados como CH1, CH2 e CH3 relativamente às posições mostrada na Figura 11. Observa-se a necessidade de um sistema de medição especial devido à baixa amplitude dos sinais lidos, com tensão de pico menores que 6 mV. Para realizar a aquisição de tais níveis de tensão e conseguir uma fidelidade no sinal adquirido se faz uso de filtros analógicos e pré-amplificadores para condicionar o sinal antes de o digitalizar.

Figura 19 - Representação gráfica dos sinais proporcionais à derivada temporal do campo magnético externo do GPS em um trecho 80 ms.



Fonte: elaborado pelo autor.

Além da medição e digitalização das formas de onda no tempo da taxa de variação do campo magnético, a UMM também realiza uma decomposição do sinal na frequência com um rastreamento da harmônica mecânica fundamental e de um número de 266 harmônicas. Esses dados são armazenados em arquivos separados. Há então um arquivo que contém formas de onda no tempo dos sinais dos três sensores com duração de cinco segundos e outro arquivo contendo as amplitudes das harmônicas de cada um dos sinais.

3.4 TRATAMENTO DOS DADOS OBTIDOS

A totalidade dos dados obtidos experimentalmente é separada em dois conjuntos de dados: um conjunto de dados dinâmicos, ou seja, aqueles que têm dependência do tempo, e outro conjunto de dados que independem do tempo. O conjunto dependente do tempo são as formas de onda obtidas para a taxa de variação do campo magnético externo do GPS. As informações contidas nestes dados são extraídas e posteriormente adicionadas ao conjunto de dados independentes do tempo que contém os demais dados de pontos de operação.

Os dados obtidos do sistema de monitoramento VibraOne e do *datalogger* de temperatura precisaram ser selecionados e tratados antes de serem adicionados ao conjunto de dados independentes do tempo. Os pontos de operação de interesse foram obtidos por meio da média dos pontos obtidos após o GPS operar de maneira estável dentro de um patamar de carregamento, conforme apresentado pelos pontos de operação da Figura 17. Uma vez que o GPS estiver estável em um patamar de carregamento, calcula-se a média aritmética dos dados coletados para este ponto de operação durante a janela de aquisição. O cálculo utilizado é representado na equação (2).

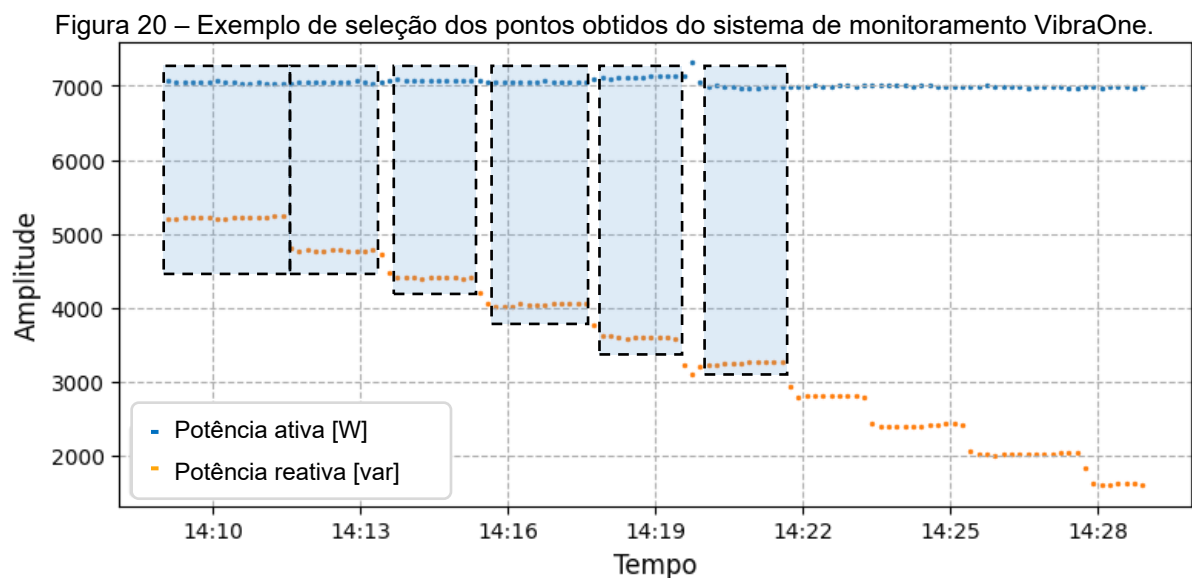
$$\bar{A} = \frac{1}{n} \sum_{i=1}^n a_i \quad (2)$$

Onde:

- \bar{A} é a média da grandeza obtida dos sistemas de monitoramento VibraOne e do *datalogger* de temperatura para o ponto de operação estável.

- a_i é a i ésima amostra selecionada adquirida durante janela de aquisição em operação estável.
- n é o número de amostras adquiridas durante a janela de aquisição e que foram selecionadas para o intervalo de operação estável.

A Figura 20 apresenta o procedimento descrito para obtenção dos dados do ponto de operação descritos no parágrafo anterior. Os blocos em azul com borda tracejada incluem apenas os pontos onde o GPS possui operação em um ponto estável. Pode-se entender esses blocos como o patamar de carregamento para determinado ponto de operação. A Figura 20 exemplifica apenas para os pontos de operação de: (7,0 kW; 5,2 kvar), (7,0 kW; 4,8 kvar), (7,0 kW; 4,4 kvar), (7,0 kW; 4 kvar), (7,0 kW; 3,6 kvar) e (7,0 kW; 3,2 kvar).



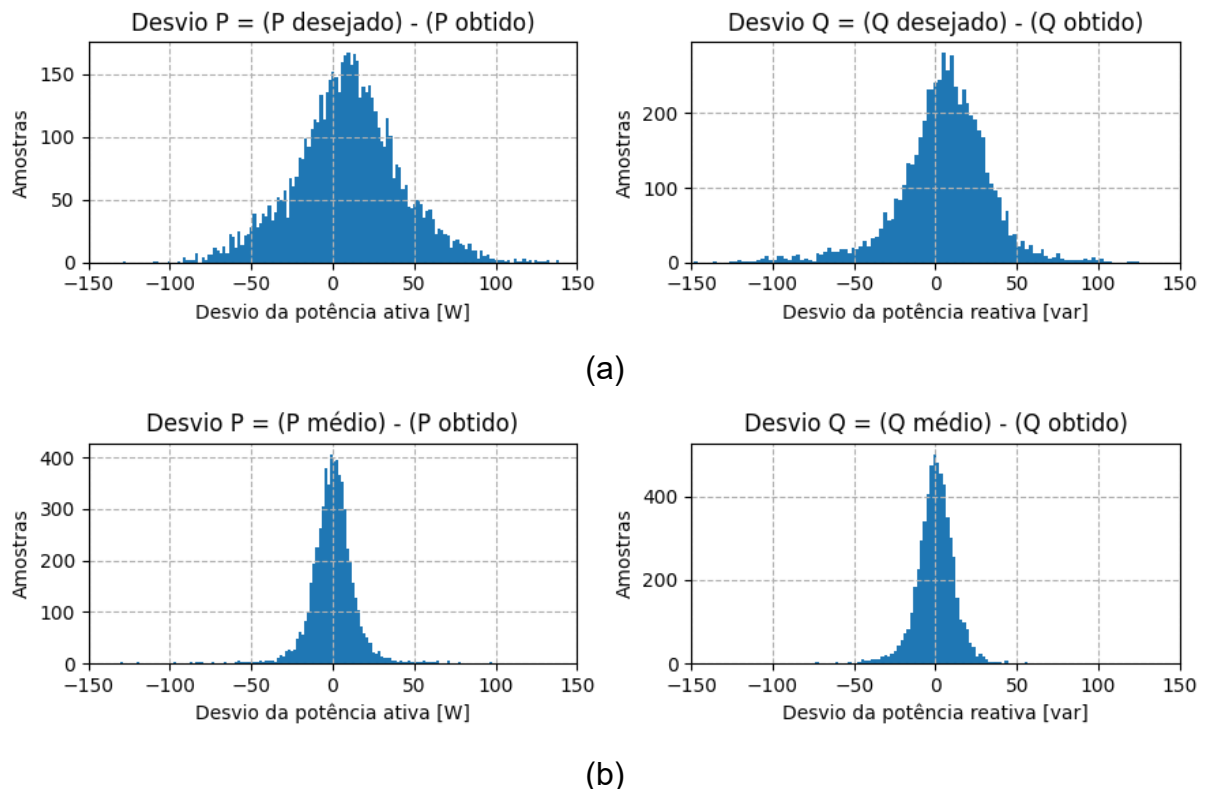
Fonte: elaborado pelo autor.

Entende-se que um ponto de operação estável difere de um ponto de operação mantido constante. Um ponto de operação mantido constante é aquele que mantém a troca de potência com a rede elétrica sem variação no tempo. Ou seja, se um gerador está entregando 7 kW de potência ativa, esses 7 kW serão mantidos constantes independente da janela de tempo da interação gerador-rede. Um ponto de operação estável é aquele que sofre leves alterações (relacionadas a pequenas perturbações) em seu valor devido às variações na rede e das próprias condições de geração da máquina. Dentre as medições obtidas neste trabalho, foram utilizadas aquelas em que o ponto de operação é considerado estável. Quanto maior a janela de aquisição, maior é a influência da rede e da temperatura nos valores do ponto de

operação. Por outro lado, quanto maior a janela de aquisição, melhor a resolução das componentes de baixa frequência obtidas para a análise do campo magnético externo.

Na Figura 20, observa-se a seleção dos dados conforme descrito. Dentro dos blocos azuis estão os dados selecionados. É notável que, apesar de não serem exatamente o mesmo valor, apresentam uma variação relativamente baixa visualmente. A fim de quantificar a variação dos dados de potência, são apresentados na Figura 21 os histogramas dos desvios dos pontos selecionados em relação ao valor teórico desejado (Figura 21(a)) e o desvio dos pontos selecionados para o valor médio dentro do respectivo patamar de carregamento (Figura 21(b)). Observa-se que os histogramas seguem o formato de curva gaussiana. Os desvios calculados em relação ao valor teórico apresentam um valor médio diferente de zero. Isto em parte reforça a afirmação anterior sobre a estabilização em um ponto de operação levemente diferente do ponto desejado.

Figura 21 - Desvio dos dados de potência obtidos.

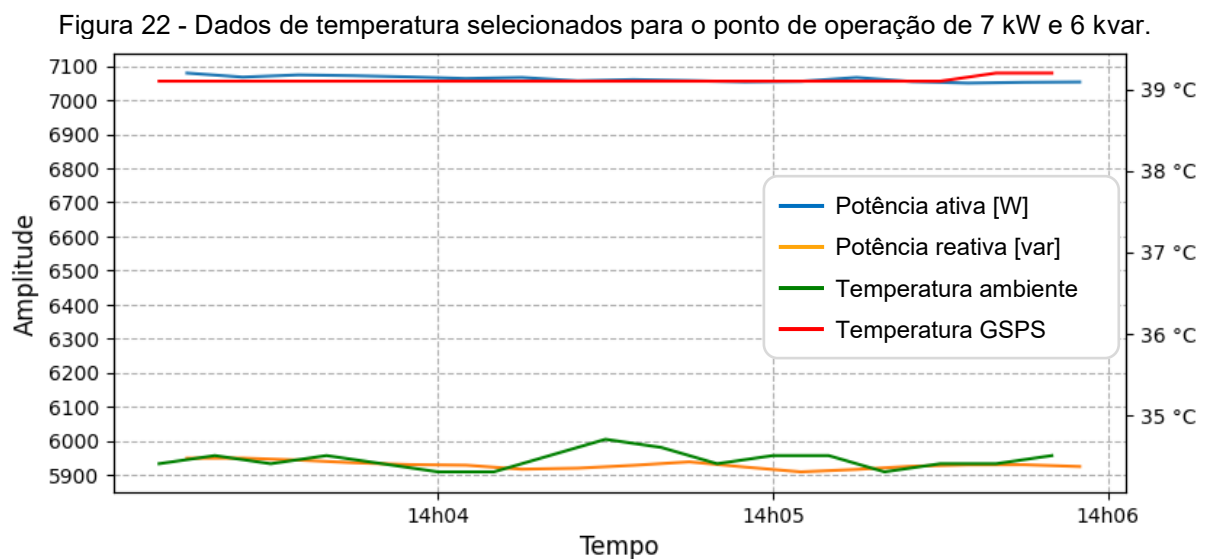


Fonte: elaborado pelo autor.

Foi considerado satisfatório, para a finalidade deste trabalho, utilizar os dados com desvio menor do que 100 W para a potência ativa e menor do que 50 var para a potência reativa relativos à potência média dentro da janela de aquisição. Isto representa um desvio menor que 5% para o ponto de menor potência ativa e 12,5%

para o ponto de menor potência reativa, com exceção de ponto de potência ativa 0 (operação como compensador síncrono).

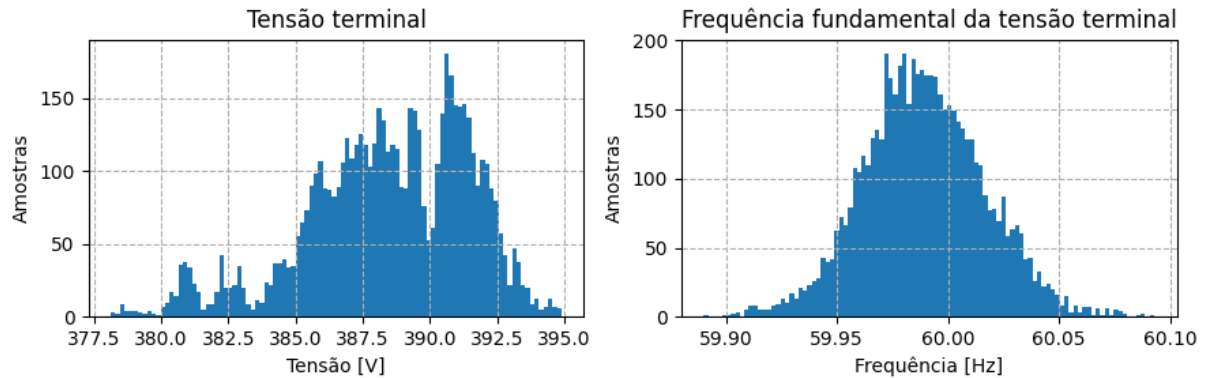
Para a seleção dos pontos de temperatura, e obtenção de seu valor médio dentro da janela de aquisição, foi necessário verificar os instantes de tempo do primeiro e último ponto de potência ativa para os pontos selecionados. Assim, todos os dados de temperatura que estão dentro da janela de tempos selecionada foram utilizados para cálculo da temperatura média da máquina para o ponto de operação em questão.



Fonte: elaborado pelo autor.

Nos dados obtidos pelo equipamento VibraOne se verificou como a tensão terminal da máquina varia, bem como a sua frequência elétrica fundamental. A Figura 23 apresenta os histogramas para os dados de tensão terminal de linha e frequência fundamental da tensão terminal. A variação da tensão terminal apresenta majoritariamente valor superior ao valor teórico esperado de 380 V. Observando a variação da tensão terminal no tempo da Figura 17 (b) para um dos patamares, houve uma variação de aproximadamente 395 V para 390 V. Isso em termos percentuais representa 1,26 % com relação à tensão inicial de 395 V. Esta queda de tensão pode ser associada à injeção de potência reativa na rede elétrica. Observa-se da Figura 17 (a) que os pontos de maior potência reativa apresentam as maiores tensões terminais e conforme a potência reativa injetada na rede diminui, a tensão terminal do gerador também diminui.

Figura 23 - Histogramas da variação da tensão terminal e da frequência da tensão terminal.



Fonte: elaborado pelo autor.

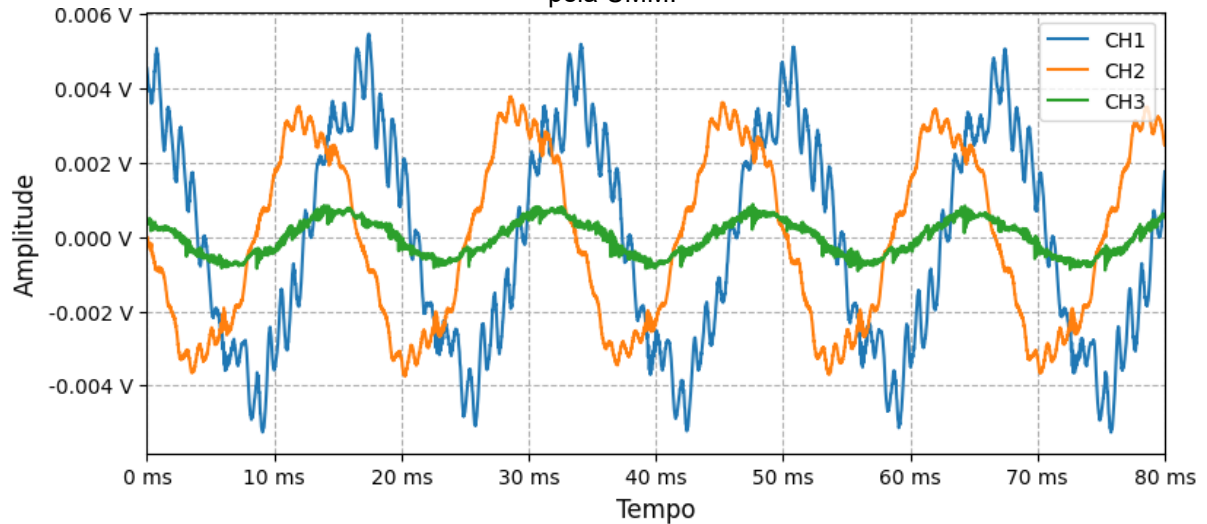
3.4.1 Dados dependentes do tempo

Os dados dependentes do tempo são os dados compostos por formas de onda da taxa de variação do campo magnético externo para cada ponto de operação, dentro da janela de aquisição de cinco segundos. Ou seja, para cada ponto de operação mapeado, foram adquiridos no mínimo 6 arquivos correspondentes às formas de onda da taxa de variação do campo magnético com cinco segundos de duração cada.

A Figura 24 representa um exemplo de dados dependentes do tempo. Nesta figura é apresentado um trecho de um ciclo mecânico (67 ms) do GPS para o campo magnético externo no ponto de operação 7 kW e 6 kvar. As formas de onda são as taxas de variação do campo magnético captadas nos sensores CH1, CH2 e CH3 cuja disposição foi apresentada na Figura 11. Ainda na Figura 24, pode-se observar visualmente nas formas de ondas em todos os sensores que a componente harmônica de frequência fundamental elétrica (60 Hz, 16,7 ms) possui a maior amplitude.

A partir destes dados dependentes do tempo são extraídas as informações para as análises de interesse. Os atributos obtidos destes dados são adicionados ao conjunto de dados independentes do tempo.

Figura 24 - Formas de onda representativas da taxa de variação do campo magnético externo obtidas pela UMM.



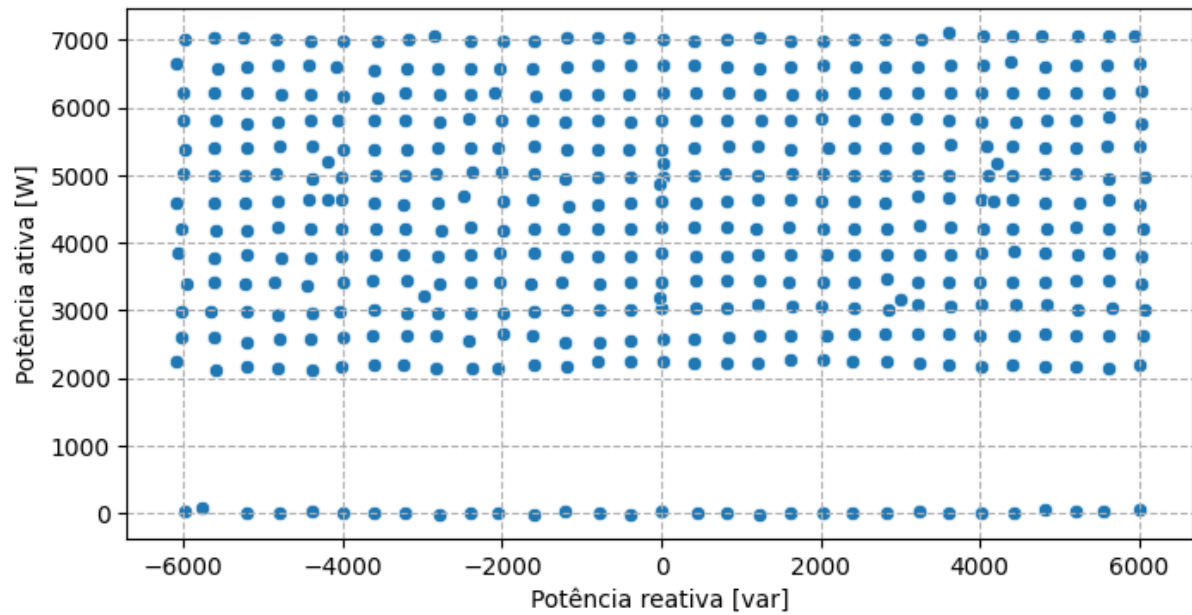
Fonte: elaborado pelo autor.

3.4.2 Dados independentes do tempo

Os dados independentes do tempo são os dados do ponto de operação da máquina juntamente com os dados das grandezas auxiliares, como já mencionado. Para este conjunto dados, a informação do tempo não é de interesse, pois se assume que os pontos obtidos são independentes entre si. Ou seja, a informação contida no ponto de 7 kW e 6 kvar não depende do ponto adquirido imediatamente antes: 7 kW e 5,6 kvar, por exemplo.

Um exemplo de dado independente do tempo é a potência ativa gerada pelo GSPS. Não é de interesse para este trabalho saber qual foi o valor gerado a cada instante de tempo, mas sim um valor médio durante a janela de aquisição, que foi aqui definida com duração de 1 minuto. Assume-se que, para a máquina operando em um ponto de operação estável dentro dos limites da curva de capacidade, as características do campo magnético externo são invariantes dentro do intervalo de aquisição. Por motivo desta suposição, são sempre utilizados os valores médios das grandezas dentro da janela de aquisição. A Figura 25 apresenta os pontos de operação obtidos experimentalmente após o processamento descrito dentro desta seção. Os atributos obtidos dos dados dependentes do tempo para cada ponto de operação são sempre os valores médios deste atributo dentro da janela de aquisição.

Figura 25 - Pontos de operação obtidos experimentalmente.



Fonte: elaborado pelo autor.

3.5 CONSIDERAÇÕES FINAIS

Este capítulo apresentou os procedimentos utilizados para operar o grupo gerador especial e os instrumentos utilizados para realizar a medição e a aquisição dos dados de campo magnético externo e das demais grandezas relacionadas ao ponto de operação. Ainda como escopo deste capítulo, apresentou-se como os dados dos diferentes instrumentos foram agregados para formar os dois conjuntos principais de dados: dados dependentes do tempo e dos dados independentes do tempo. Também, mostrou-se e se analisou brevemente os comportamentos dos dados empregados.

Com os dois conjuntos de dados é possível utilizar a informação para realizar estudos a respeito do campo magnético externo de um gerador síncrono de polos salientes. Caso seja necessário, é possível utilizar o conjunto dos dados dependentes do tempo para extrair informações adicionais sobre o comportamento do campo magnético.

Assume-se que todos os dados obtidos durante os testes são com a máquina operando sem falta. Assim, o conjunto de dados obtidos experimentalmente deve ser representativo de um GSPS sadio.

4 ANÁLISE DOS DADOS DO CAMPO EXTERNO DO GERADOR SÍNCRONO DE POLOS SALIENTES

Durante sua operação, um gerador síncrono produz um campo magnético externo que contém informações específicas. Essas informações contidas na assinatura magnética podem ser avaliadas pelo viés do ponto de operação, como é o caso deste trabalho, e também pelo ponto de vista de monitoramento ao inferir sobre possíveis faltas incipientes. Entende-se que o estudo da influência do ponto de operação sobre o campo magnético externo é complementar aos estudos de faltas incipientes [3], [5], [11].

O estudo e a análise do conjunto de dados independentes do tempo é um passo importante para descobrir relações entre o ponto de operação e o campo magnético externo. As possíveis conclusões desta análise agregam novos conhecimentos ao estado atual dos estudos investigativos sobre o comportamento do campo magnético externo em máquinas elétricas, principalmente em geradores elétricos.

Neste capítulo, será conduzida uma investigação a fim de extrair informações relevantes sobre o ponto de operação do GSPS ensaiado em laboratório. Com os resultados obtidos desta investigação, espera-se encontrar modelos de regressão linear que, quando aplicados ao conjunto de dados independentes do tempo, estime o ponto de operação do GSPS apenas com as informações contidas no campo magnético externo.

4.1 INFORMAÇÕES EXTRAÍDAS DO CAMPO MAGNÉTICO EXTERNO

O capítulo 3 apresentou como foram obtidos os conjuntos de dados dependentes e independentes do tempo. Na presente seção, esses conjuntos serão explorados para verificar quais informações adicionais podem ser utilizadas além das amplitudes das harmônicas dos sinais de tensão advindos dos sensores indutivos.

O valor RMS (*root mean square*) da tensão é fácil de ser obtido a partir dos dados dependentes do tempo. Assim, a métrica foi prontamente extraída dos vetores de tensão dos sensores. Os dados formados pelos valores RMS foram agregados ao conjunto de dados independentes do tempo.

Análises no domínio da frequência por meio da FFT foram também realizadas. Nesta exploração está descrito um método de como encontrar as amplitudes das

harmônicas de interesse. Este método é chamado de rastreamento de harmônicas. Ao fim da investigação dos dados no domínio da frequência é apresentada uma discussão sobre a confiabilidade das componentes harmônicas a serem utilizadas.

Um estudo de correlação dos dados foi realizado com o conjunto de dados independentes do tempo. O estudo apresenta como os coeficientes de correlação foram obtidos e como interpretar seus valores. Três coeficientes diferentes foram utilizados: correlação de Pearson, correlação de Spearman e correlação PhiK. O objetivo do estudo de correlação é entender como o ponto de operação do GPS se relaciona com as grandezas do conjunto de dados independentes do tempo.

4.1.1 Valor RMS

O valor eficaz (*root mean square* – RMS) representa o nível contínuo equivalente de uma forma de onda. O cálculo do valor RMS para cada forma de onda do conjunto dos dados dependentes do tempo foi feito segundo a equação (3). Para cada ponto de operação do conjunto de dados independentes do tempo existem no mínimo seis vetores de tensão obtidos para cada sensor no conjunto de dados dependentes do tempo.

$$RMS(x) = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2} \quad (3)$$

onde:

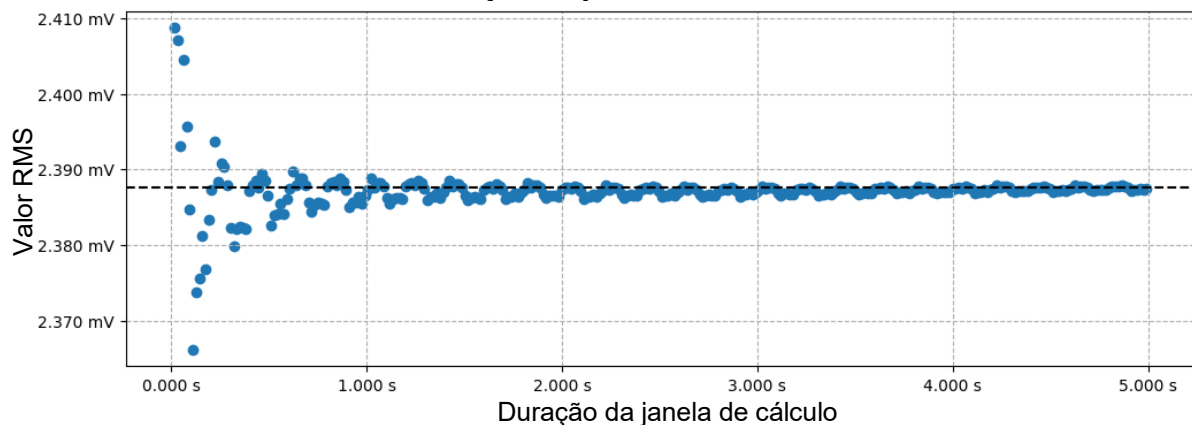
- $RMS(x)$ é o valor RMS calculado para um vetor de tensão x que contém n amostras.
- n é o número de amostras obtidas dentro de vetor de tensão x .
- x_i é a i ésima amostra do vetor de tensão x .

Conforme apresentado no capítulo 3, a janela de tempo para a aquisição da tensão induzida nos sensores de campo magnético foi de um minuto. Porém, o sistema de aquisição utilizado (UMM) adquire os dados em janelas de cinco segundos. Então, dentro dos 60 segundos definidos para a aquisição são realizadas no mínimo 6 aquisições. Necessita-se de um período de tempo extra de aproximadamente 5 segundos para realizar o cálculo da FFT e o rastreamento das amplitudes das harmônicas. Assim a aquisição e o pré-processamento demoram cerca de 10 segundos no equipamento *homemade* utilizado.

O valor RMS foi calculado para janelas de cinco segundos. Entende-se que o cálculo do valor RMS dentro da janela não é o mais adequado. Uma melhor maneira seria calcular o valor RMS em um ciclo completo da menor frequência de interesse, ou em um intervalo de tempo formado por vários ciclos completos.

Como exemplo, a Figura 26 apresenta o comportamento do valor RMS da forma de onda da tensão do sensor com a duração da janela de cálculo. Observa-se que, conforme a janela de tempo utilizada para o cálculo do valor RMS aumenta, há uma convergência para o valor RMS por ciclo de forma de onda, indicado pela linha tracejada. Ainda assim, existe uma pequena oscilação dos valores RMS, a qual é cada vez menos representativa conforme a duração da janela de cálculo aumenta. A oscilação representa uma variação de aproximadamente $\pm 0,04\%$ do o valor RMS por ciclo para a janela de cinco segundos. Essa tendência se mostrou similar nos cálculos dos valores RMS dos três canais.

Figura 26 – Variação do valor RMS da tensão induzida nos sensores de campo para diferentes durações de janela de cálculo.

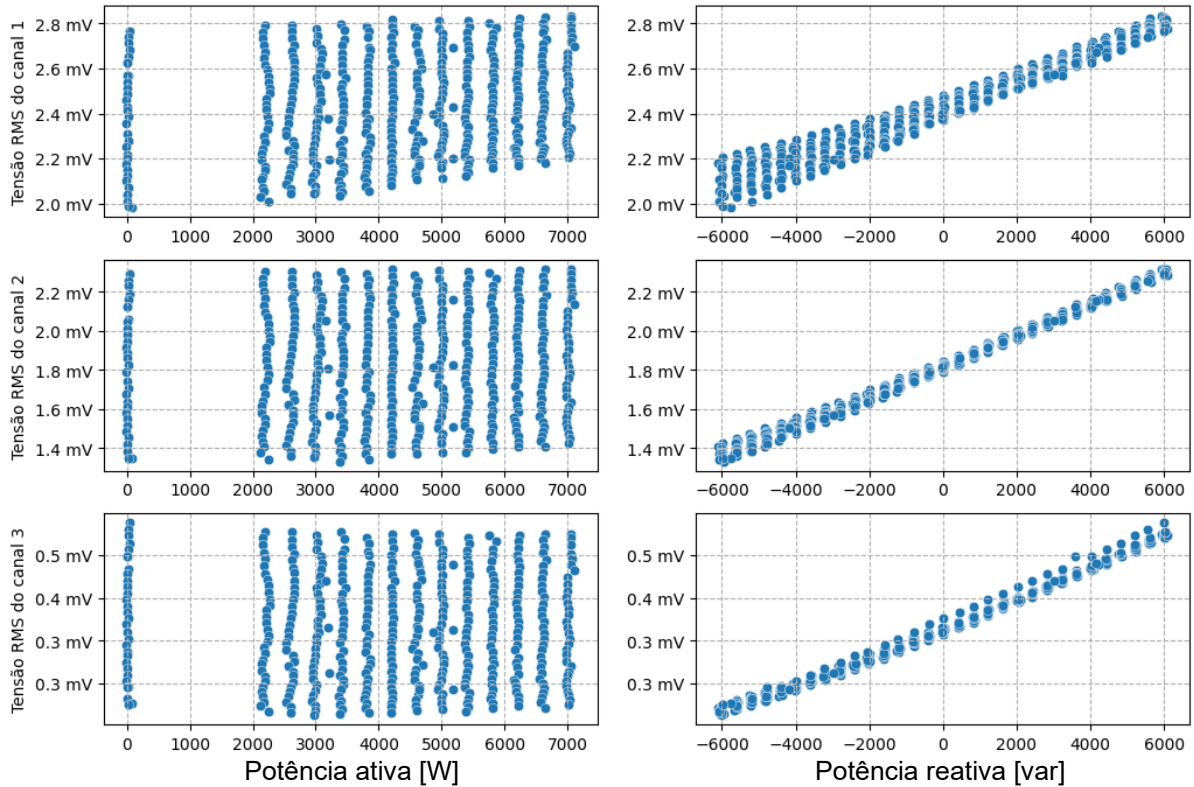


Fonte: elaborado pelo autor.

O valor RMS da tensão induzida no sensor que foi adicionado ao conjunto de dados independentes do tempo é a média dos valores RMS calculados para todos os vetores de tensão obtidos em um mesmo ponto de operação. Entende-se que a utilização do valor médio das aquisições ajuda a mitigar o efeito da oscilação dos valores RMS.

A Figura 27 apresenta os valores RMS dos três canais para cada ponto de operação do GPS. Observa-se uma relação aproximadamente linear entre o valor RMS da tensão induzida nos sensores com a potência reativa. Apenas o canal 1 apresenta uma fraca relação visível entre a potência ativa e o valor RMS da tensão induzida no sensor. Por simplicidade, o valor RMS da tensão induzida ao longo do texto será referido como valor RMS.

Figura 27 - Valor RMS dos vetores de tensão para os três canais, canal 1 (CH1), canal 2 (CH2) e canal 3 (CH3).



Fonte: elaborado pelo autor.

4.1.2 Atributos obtidos do espectro de frequência

Esta seção tem o objetivo de apresentar como as informações contidas no domínio da frequência são extraídas do conjunto de dados dependentes do tempo. As análises feitas são baseadas na transformada discreta de Fourier utilizando a FFT.

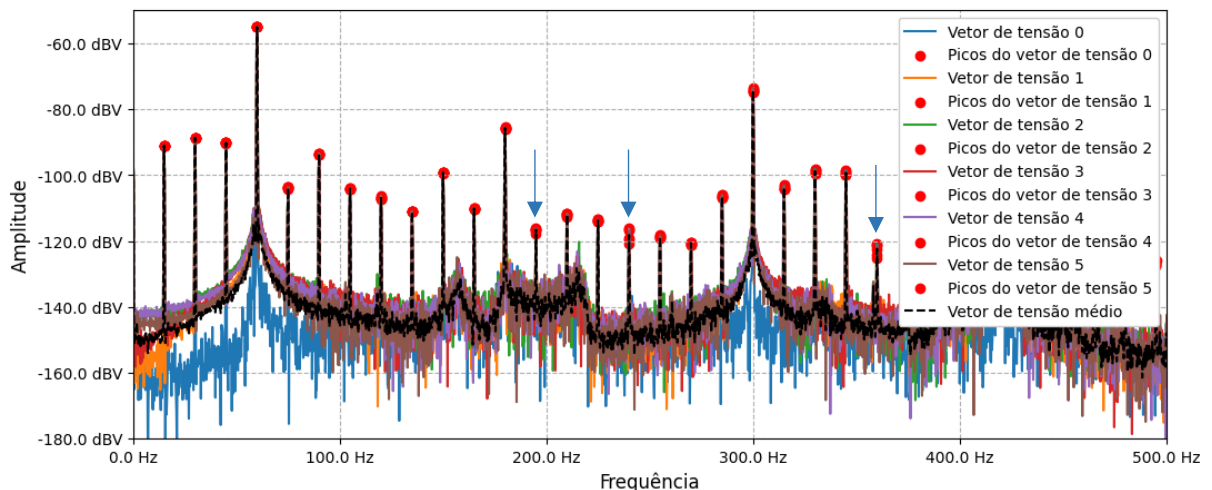
O espectro de frequência foi obtido para cada vetor de tensão do conjunto de dados dependentes do tempo. A partir do espectro de frequência são encontrados as amplitudes das harmônicas múltiplas da frequência fundamental mecânica do gerador (15 Hz). Então, essas amplitudes das harmônicas são adicionadas ao conjunto de dados independentes do tempo, conforme apresentado no capítulo 3. O processo chamado “rastreamento de harmônicas” é feito de maneira automática pelo sistema de aquisição utilizado (UMM).

O algoritmo que faz o rastreamento das harmônicas tem como entrada a frequência fundamental mecânica teórica do GSPS e um espectro de frequência. A partir do espectro de frequência, o algoritmo busca nas proximidades da harmônica múltipla a maior amplitude, e entende que este valor deve ser a amplitude da harmônica de interesse. Há dificuldades no processo de rastreamento por conta do

vazamento espectral e por perturbações na frequência da rede elétrica, principalmente.

A Figura 28 apresenta espectro de frequência até 500 Hz de seis vetores de tensão obtidos do sensor CH1 para um mesmo ponto de operação. As amplitudes das harmônicas são apontadas por pontos vermelhos. Adotou-se a mesma cor (vermelha) em todos os picos dos diferentes vetores de tensão por conveniência. Os gráficos apresentados são resultado da aplicação da FFT com janelamento *Hamming* nos vetores de tensão. Os valores das amplitudes são apresentados em dBV para a observação possível de pequenas variações no espectro. A unidade dBV utiliza 1 V como referência para o cálculo da amplitude em dB. Também é comum se encontrar estes gráficos em dBmV e dB μ V, que utilizam 1 mV e 1 μ V como referência. A experiência prática mostrou que este tipo de rastreamento é eficaz para encontrar os picos no espectro. Apesar deste método de obtenção das amplitudes das harmônicas ser suficientemente bom, ainda é possível notar que existem algumas harmônicas com valor de amplitude relativamente baixo e próximo do ruído de fundo, como também que há harmônicas com grande desvio padrão em sua amplitude, a exemplo das harmônicas indicadas pelas setas azuis.

Figura 28 - Espectro de frequência do vetor de tensão obtido do sensor CH1.

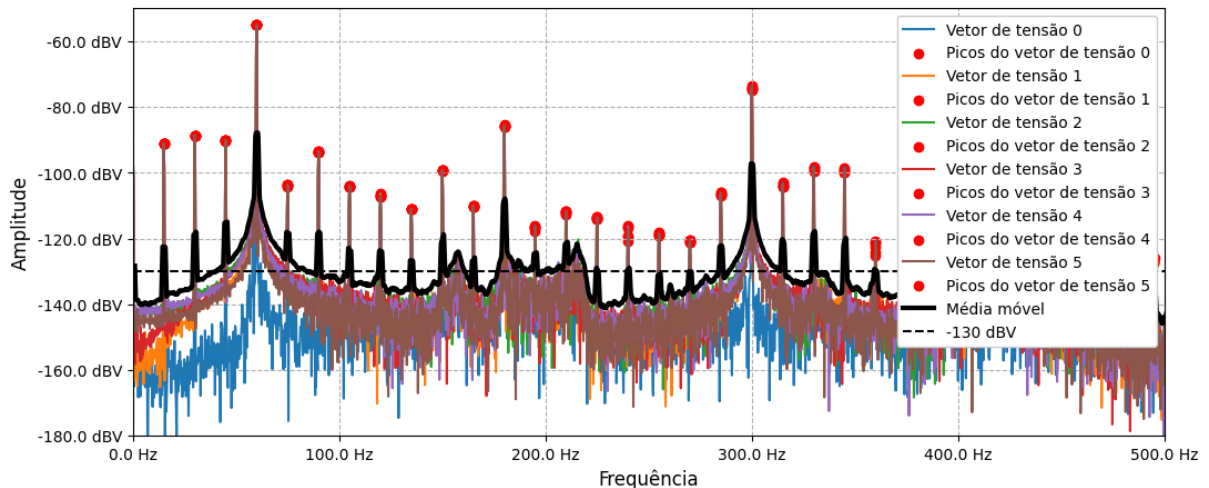


Fonte: elaborado pelo autor.

Pode-se aplicar um filtro para selecionar as harmônicas com valores acima de uma certa amplitude, como exemplificado pela linha tracejada na Figura 29. Neste caso, as harmônicas abaixo da linha tracejada são descartadas. Pode-se verificar que esse filtro é de certa maneira injusto por existirem harmônicas com amplitudes bem definidas abaixo da linha de corte. As harmônicas são ditas bem definidas quando possuem amplitude sobre o ruído de fundo nas proximidades do pico. Por este motivo,

é possível aplicar um filtro de média móvel no espectro do vetor de tensão médio, apresentado pela linha preta na Figura 29, somado em 10 dB. Todas as harmônicas com amplitudes acima desta média móvel podem ser mantidas e as demais descartadas por serem julgadas com amplitude muito próxima ao ruído de base.

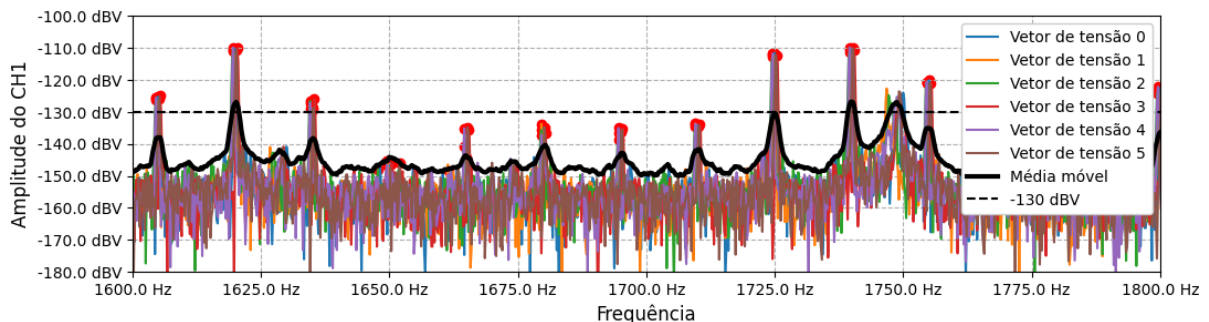
Figura 29 - Espectro de frequência do vetor de tensão obtido do sensor CH1 com filtro de corte por amplitude mínima com aplicação de filtro por média móvel.

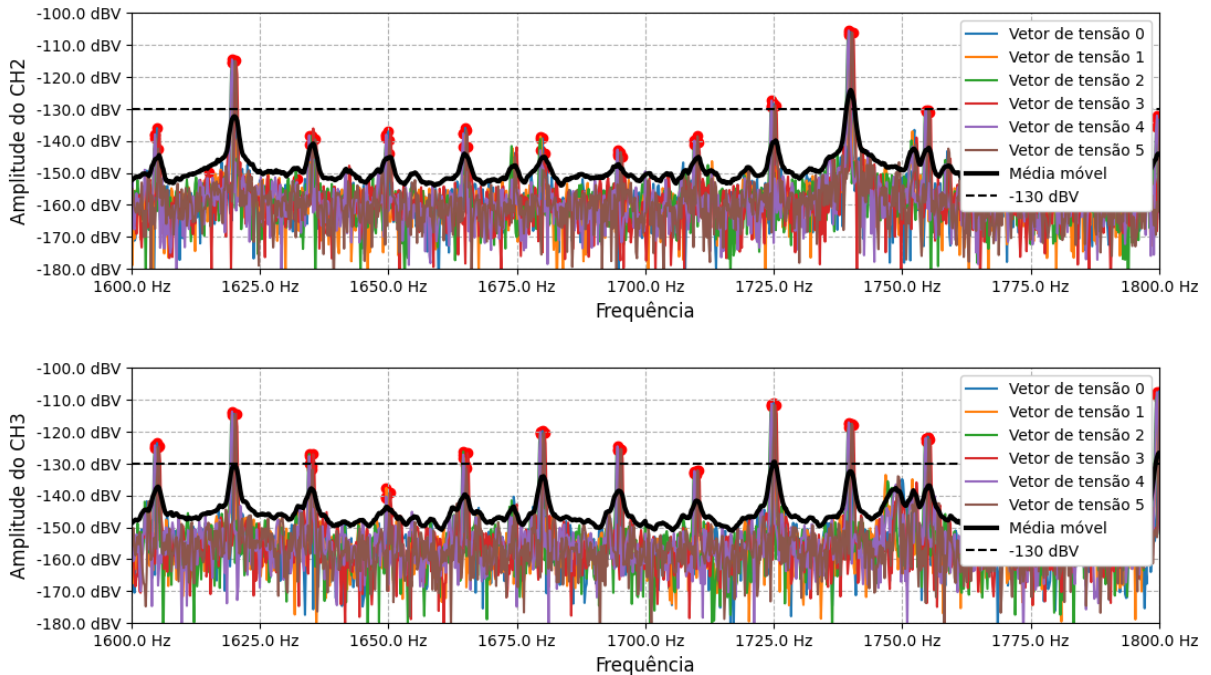


Fonte: elaborado pelo autor.

Para exemplificar a possível aplicação dos filtros, encontrou-se no espectro uma harmônica que possui nível abaixo da média móvel somada de 10 dB. A harmônica de 1650 Hz possui um nível abaixo da média móvel no canal 1, porém é bem definida pelo critério da média móvel para os canais 2 e 3, conforme apresentado na Figura 30. Assim, a harmônica de 1650 Hz do canal 1 seria descartada e mantida para os canais 2 e 3. Também é possível notar na Figura 30 que cortar harmônicas baseada em um filtro de amplitude mínima resulta na perda da informação contida nas harmônicas de 1665, 1680, 1695 e 1710 Hz para o canal 1, as harmônicas em 1605, 1635, 1695, 1650, 1665, 1680, 1695 e 1710 Hz para o canal 2, e as harmônicas em 1650 e 1710 Hz para o canal 3. A legenda dos picos dos vetores de tensão foi omitida por simplicidade.

Figura 30 - Aplicação do filtro de média móvel para a harmônica de 1650 Hz.

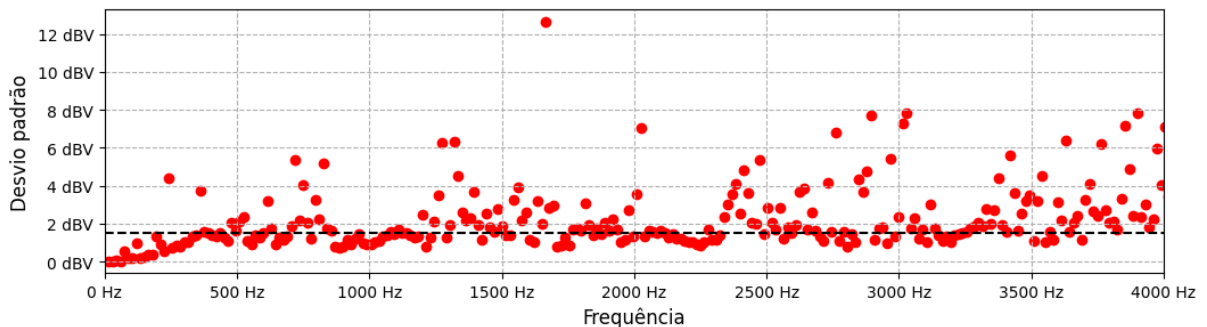




Fonte: elaborado pelo autor.

Em outro filtro que pode ser empregado, utiliza-se o desvio padrão entre os 6 vetores de tensão para o mesmo ponto e mesmo sensor, descartando as harmônicas cujas amplitudes possuem desvio padrão maior que um determinado valor. A Figura 31 apresenta o desvio padrão para as harmônicas do vetor de tensão do sensor CH1 em dBV. A linha tracejada significa um possível limite de corte em 1,8 dBV, por exemplo.

Figura 31 - Desvio padrão, em dBV, das amplitudes das harmônicas filtradas para CH1



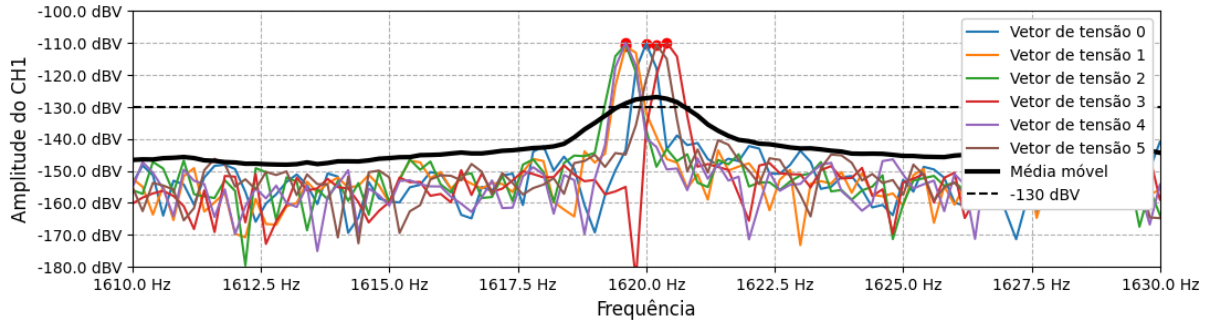
Fonte: elaborado pelo autor.

Conforme observado durante os experimentos e apresentado na Figura 23, a frequência da rede elétrica não é fixa em 60 Hz. Portanto, é esperado que existam flutuações na frequência fundamental do GPS e suas harmônicas.

Utilizando a mesma informação apresentada no gráfico do canal 1 da Figura 30, pode-se diminuir os limites da escala horizontal com o intuito mostrar em detalhe a harmônica de 1620 Hz. Observa-se na Figura 32 o fenômeno da flutuação da

frequência da harmônica fundamental por meio da variação do valor da harmônica nominal de 1620 Hz. Apesar de os picos selecionados estarem em frequências diferentes de 1620 Hz, no conjunto de dados independentes do tempo foi adicionado o valor médio das harmônicas filtradas nas proximidades de 1620 Hz e considerado como sendo o pico da harmônica de 1620 Hz para o ponto de operação no qual os vetores de tensão foram obtidos.

Figura 32 - Flutuação da frequência fundamental observada na harmônica de 1620 Hz.



Fonte: elaborado pelo autor.

Outra consideração sobre as oscilações da frequência fundamental é sobre a amplitude verificada na raia estipulada para a frequência durante o cálculo da FFT. Sabe-se que, se a frequência possui um pequeno desvio, como exemplo de 60 para 60,1 Hz, a frequência real fica entre duas raias da FFT calculada (em uma aquisição com período de 5 segundos, o espaçamento entre as raias de frequência da FFT são de 0,2 Hz). Neste contexto também ocorre o fenômeno de vazamento espectral próximo à amplitude da harmônica. Uma maneira de diminuir a influência do vazamento espectral seria aumentar o tempo de aquisição para 10 segundos, utilizando janelamento adequado, assim a distância entre as raias da FFT ficaria de 0,1 Hz e também utilizar a técnica de *zero-padding*. A utilização desta técnica se mostrou eficiente para corrigir a amplitude da harmônica e melhorar a resolução de espectral da FFT. A técnica de *zero-padding* consiste em adicionar zeros ao final do vetor de tensão. Pode-se entender o *zero-padding* como interpolar a FFT ao aumentar artificialmente o tempo de duração do vetor de tensão [32].

4.1.2.1 Considerações finais sobre os atributos de frequência

Nesta seção foi apresentado como foram calculadas as amplitudes das componentes harmônicas dos vetores de tensão presentes no conjunto de dados independentes do tempo.

Observando as imagens da seção 4.1.2, principalmente a Figura 29, sobre o comportamento das amplitudes das harmônicas (FFTs com janelamento Hamming, transformando os níveis de tensão para dB utilizando 1 V como referência (dBV)), um olhar rigoroso não permite afirmar que o nível de ruído de fundo é linear. Apesar disso, para os resultados obtidos, observa-se que as harmônicas com amplitude maior do que -120 dBV possuem, em geral, picos bem definidos até 5 kHz. Indo além, pode-se traçar uma linha em - 130 dBV e observar que a maior parte das harmônicas acima desta amplitude são bem definidas, com exceção das harmônicas em torno de 60 Hz e as harmônicas em torno de 1020 Hz.

Ao observar partes do espectro com frequências mais altas, como é o caso da Figura 30, pode-se considerar que o ruído de fundo apresenta um comportamento estável com um nível aproximado de - 150 dBV. Assim, tem-se que a maior parte das harmônicas possuem picos bem definidos quando suas amplitudes estão acima de - 140 dBV. Ainda assim, existem algumas harmônicas que estão no mesmo nível do ruído de fundo e, portanto, estudos futuros precisam ser feitos para verificar sua confiabilidade.

Não cabe a este trabalho realizar um estudo de confiabilidade, nem determinar critérios de escolha para as harmônicas. Alguns possíveis filtros foram propostos e devem ser investigados criteriosamente. Na ausência aqui de conclusões sobre a aplicação destes filtros e sobre a confiabilidade das harmônicas, todas as harmônicas obtidas foram adicionadas ao conjunto de dados. Sugere-se, portanto, como trabalho futuro avaliar a necessidade da implementação de filtros adequados.

Com o conjunto de dados independentes do tempo, pode-se extrair ainda mais informações para estimar o ponto de operação, tanto da potência ativa quanto da potência reativa. Esta informação poderia estar relacionada com as fases entre as harmônicas de diferentes pontos de operação. Entende-se que existem complicações para a extração da informação da fase das harmônicas e, portanto, não serão abordadas neste trabalho. Sugere-se, como trabalho futuro, investigar e utilizar técnicas de processamento de dados para obter estas informações de fase e, uma vez que os atributos de fase forem obtidos, utilizar os passos descritos nas próximas seções para avaliar seus resultados.

4.1.3 Correlação entre potência elétrica e campo magnético externo

A possibilidade de estimar o ponto de operação a partir do campo externo carrega intrinsicamente a suposição de que o campo externo contenha informações suficientes para realizar tal tarefa. Assim, um estudo de correlação foi realizado com a finalidade de explorar as relações entre os dados de campo e os pontos de operação, empregando o conjunto dos dados independentes do tempo. Busca-se entender como os dados se comportam com relação às grandezas de interesses. Neste estudo foram utilizadas as correlações de Pearson, Spearman e Phik [33].

O APÊNDICE D – Análise e aplicação dos cálculos de correlação, apresenta uma análise sobre as correlações de Pearson, Spearman e Phik, além da aplicação de cada um dos cálculos de correlação com exemplos e demonstrações.

Conforme indicado em [34], bons atributos para um modelo de regressão podem ser selecionados empregando o estudo de correlação. Um grupo adequado de atributos é composto por atributos altamente correlacionados com a variável alvo, porém não correlacionados entre si. Ou seja, um atributo é útil se ele for correlacionado com a variável alvo e, caso contrário, o atributo é dito irrelevante.

Dois problemas que podem ser encontrados no conjunto de dados independentes do tempo são: i) a correlação espúria e ii) multicolinearidade.

A correlação espúria indica uma relação aparente e mensurável entre duas variáveis, porém sem efeito causal. A correlação espúria pode vir mascarada como uma correlação a uma terceira variável não levada em consideração em primeiro momento [35], [36]. Visto o funcionamento do GPS, assume-se que as variações das amplitudes das harmônicas da taxa de variação do campo externo sejam causadas por variações do ponto de operação.

A multicolinearidade é uma condição definida como a relação de dependência entre duas variáveis supostamente independentes. Uma alta correlação entre os atributos a serem utilizados por um modelo é um indicativo da existência deste fenômeno. Assim, entende-se que caso houver multicolinearidade dentro de um conjunto de dados, as informações contidas nas variáveis correlacionadas entre si serão redundantes. Portanto, a informação que estas variáveis inserem no modelo de regressão linear causará instabilidades nos coeficientes calculados [37], [38], [39]. Em [40] é reforçado o efeito da multicolinearidade sobre a estabilidade dos coeficientes do modelo de regressão. Ainda em [40] é sugerido utilizar algum modelo de regressão

que penalize grandes variações nos coeficientes. Dentre os modelos sugeridos pelo autor estão os modelos com penalização para coeficientes com altos valores.

Uma maneira de verificar relações entre os dados é por meio de um gráfico de pontos. Na Figura 33 estão apresentados os dados usuais do GSPS colocados em gráficos de pontos com as potências ativa e reativa

Na Figura 33 (a), observa-se que a tensão terminal é influenciada pela potência entregue à rede elétrica pelo GSPS. Como a potência elétrica flui do gerador para a rede é esperado um aumento da tensão no barramento de conexão do gerador.

Na Figura 33 (b), observa-se que a corrente de linha apresenta uma relação com as potências ativa e reativa. As curvas em “V” são observadas para a corrente de linha com relação a potência reativa. Uma dessas curvas foi apresentada na Figura 17(d). A Figura 33 (b) também mostra uma família de curvas características em diversos valores de potência ativa. A corrente de campo apresenta uma relação direta com a potência reativa pelo próprio funcionamento do GSPS.

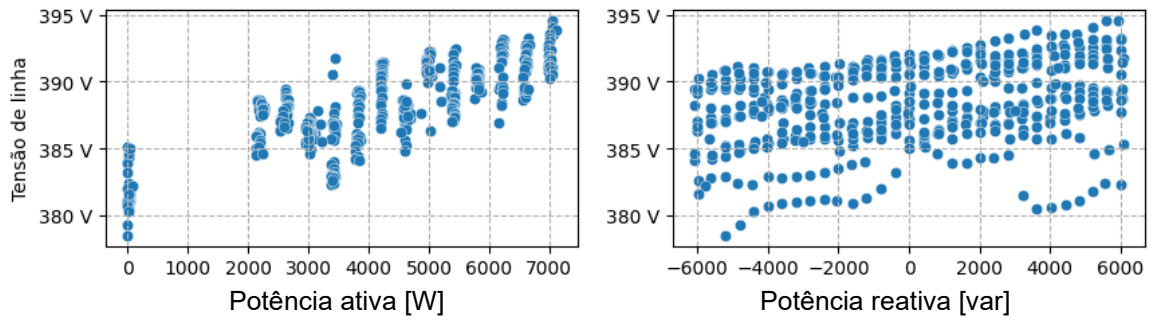
A Figura 33 (c) apresenta a variação das potências ativa e reativa com relação à corrente de campo do GSPS.

A Figura 33 (d) apresenta a variação da frequência da componente fundamental elétrica da tensão de linha do GSPS. Não é possível visualizar alguma relação entre a frequência da tensão e as potências ativa e reativa injetadas na rede pelo GSPS.

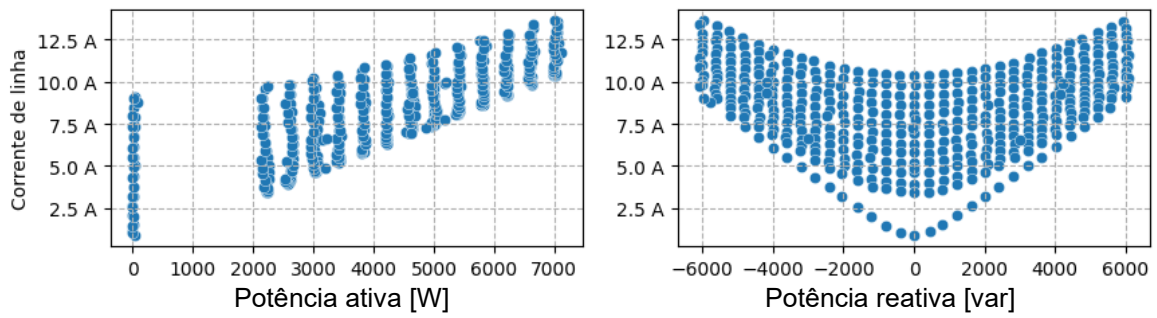
A Figura 33 (e) apresenta a temperatura ambiente quando foram obtidas as grandezas relacionadas aos pontos de operação do GSPS. A temperatura ambiente está relacionada com a temperatura do GSPS devido à maneira como o calor é dissipado na máquina. Observa-se que a maioria dos pontos de operação foram obtidos dentro da faixa de temperatura ambiente entre 33 e 36°C.

A Figura 33 (f) apresenta a temperatura da carcaça do GSPS quando foram obtidas as grandezas relacionadas aos pontos de operação do GSPS. Sabe-se que o ponto de operação causa variações de temperatura do GSPS, pois, ao aumentar a potência aparente fornecida pelo gerador à rede, há uma maior corrente elétrica nos enrolamentos do estator e rotor do GSPS e, portanto, resulta em mais energia sendo dissipada em forma de calor.

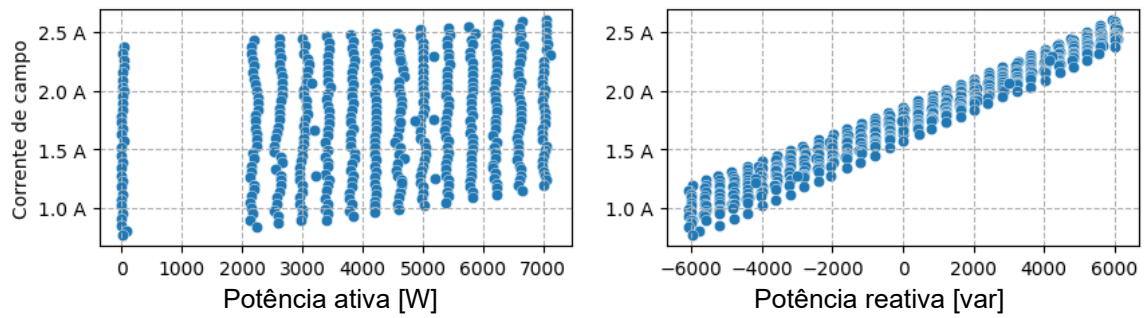
Figura 33 - Dados gerais do GPS comparados com os dados de potência ativa e reativa.



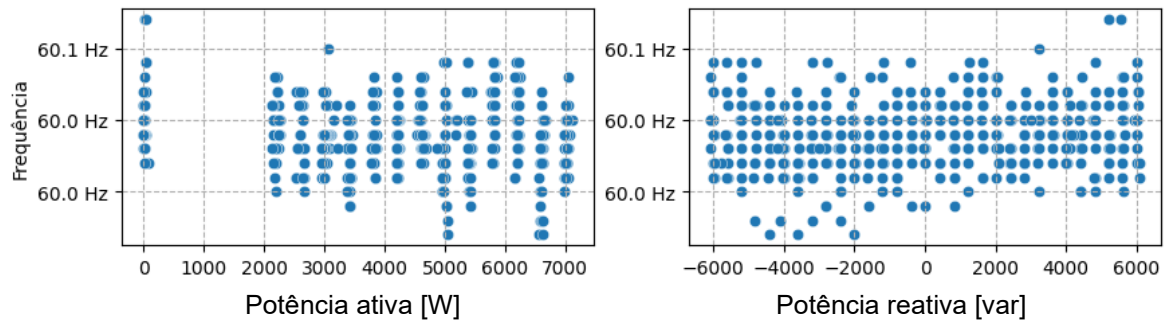
(a)



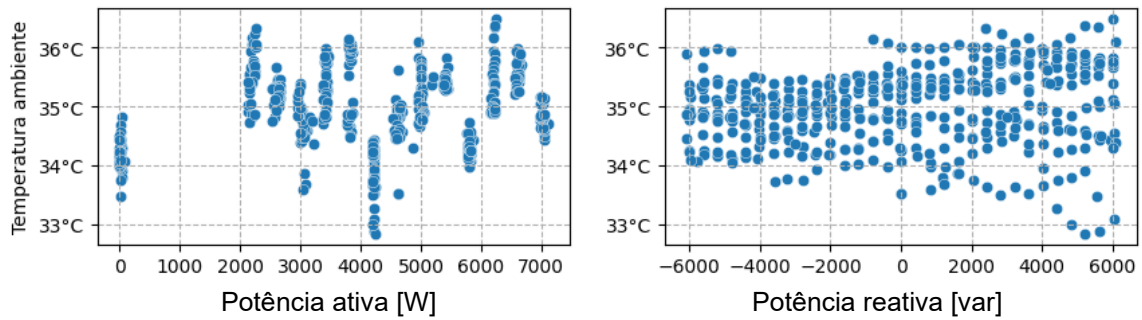
(b)



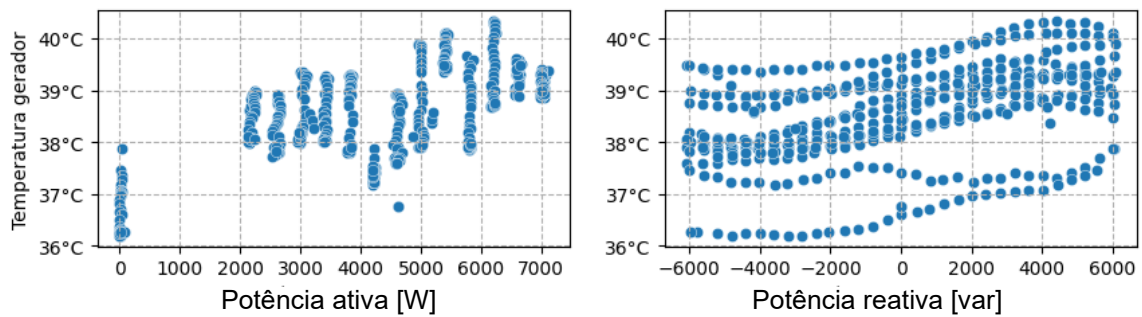
(c)



(d)



(e)



(f)

Fonte: elaborado pelo autor.

4.1.3.1 Correlação de Pearson

A correlação de Pearson é utilizada para avaliar a relação de linearidade entre duas variáveis. O coeficiente de correlação ρ (ρ de Pearson) pode variar de -1 (indicando uma forte correlação negativa) a 1 (indicando uma forte correlação positiva). Uma correlação com valor igual a zero indica a inexistência de relação linear entre as variáveis. A definição da correlação de Pearson é a covariância entre duas variáveis dividido pelo produto de seus desvios padrão [34], [41], [42], [43], [44]. A equação (4) apresenta como é calculado o coeficiente ρ de Pearson com as equações (5) (covariância entre duas variáveis) e (6) (desvio padrão) como equações auxiliares.

$$\rho_{x,y} = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y} \quad (4)$$

$$\text{cov}(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n} \quad (5)$$

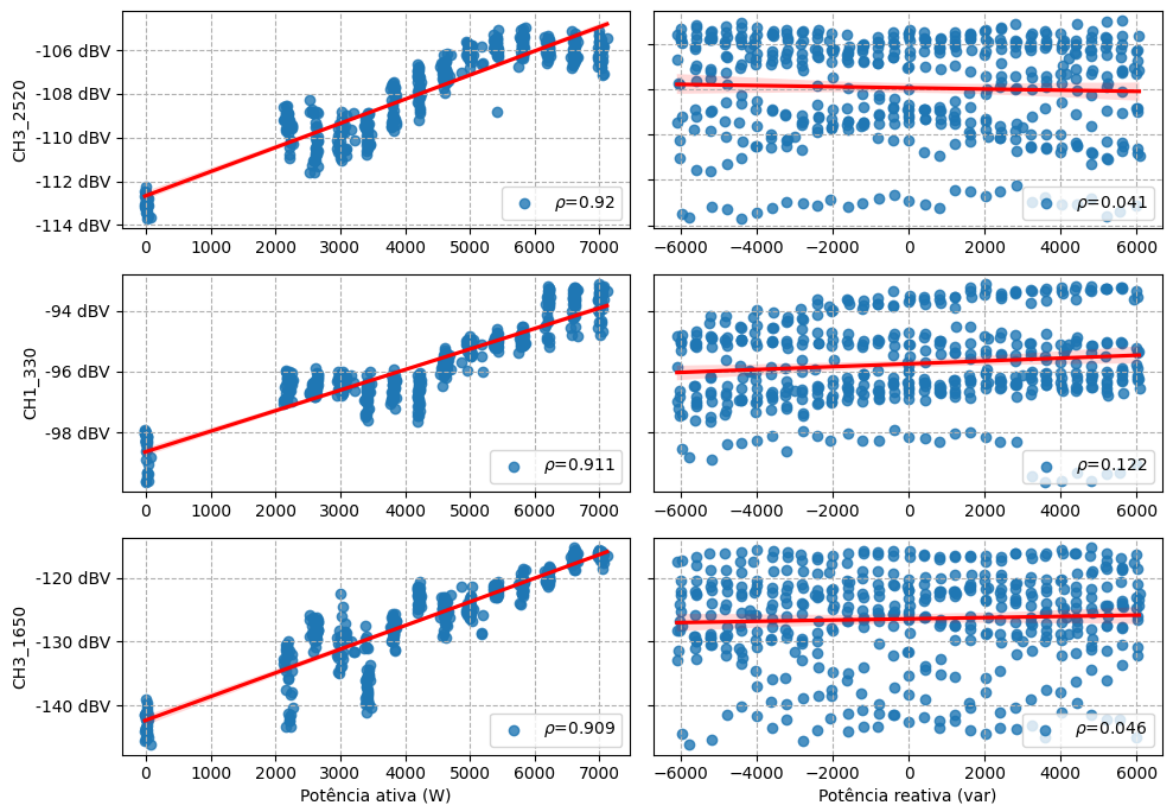
$$\sigma_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} \quad (6)$$

Onde:

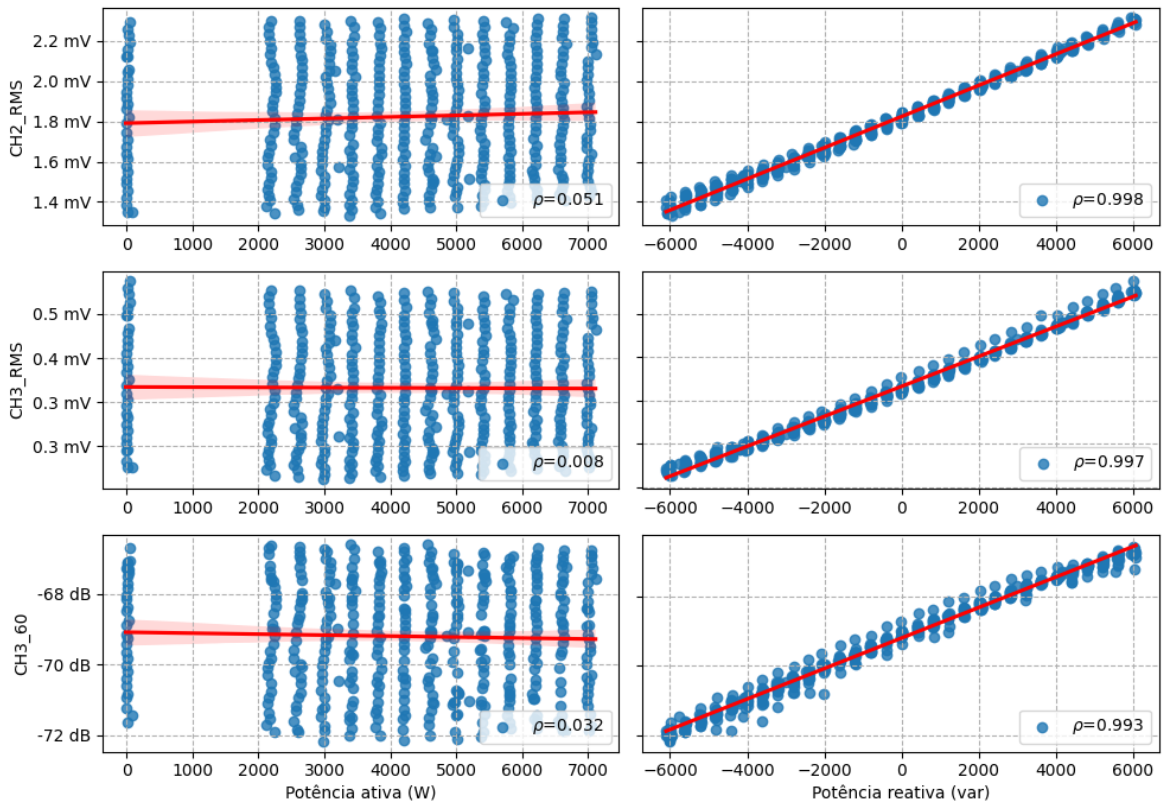
- $\rho_{x,y}$ é o coeficiente de correlação de Pearson entre as variáveis x e y .
- σ_x é o desvio padrão da variável x . σ_y é o desvio padrão da variável y .
- \bar{x} é o valor médio de x , \bar{y} é o valor médio de y .
- x_i é o i ésimo elemento de x , y_i é o i ésimo elemento de y .
- n é o número total de amostras.

Os resultados obtidos por meio do cálculo da correlação de Pearson para o conjunto de dados independentes do tempo foram ranqueados em ordem decrescente (do atributo de maior correlação para o de menor). A Figura 34 (a) apresenta os três atributos mais correlacionados com a potência ativa e a Figura 34 (b) os três atributos mais correlacionados com a potência reativa. As linhas vermelhas indicam a regressão pelo método dos mínimos quadrados e o coeficiente de correlação está contido em cada gráfico. A Figura 35 apresenta as correlações entre as variáveis mais correlacionadas com as potências ativa e reativa, respectivamente pelo método de Pearson.

Figura 34 - Atributos mais relacionadas com a potência ativa e reativa de acordo com a correlação de Pearson.



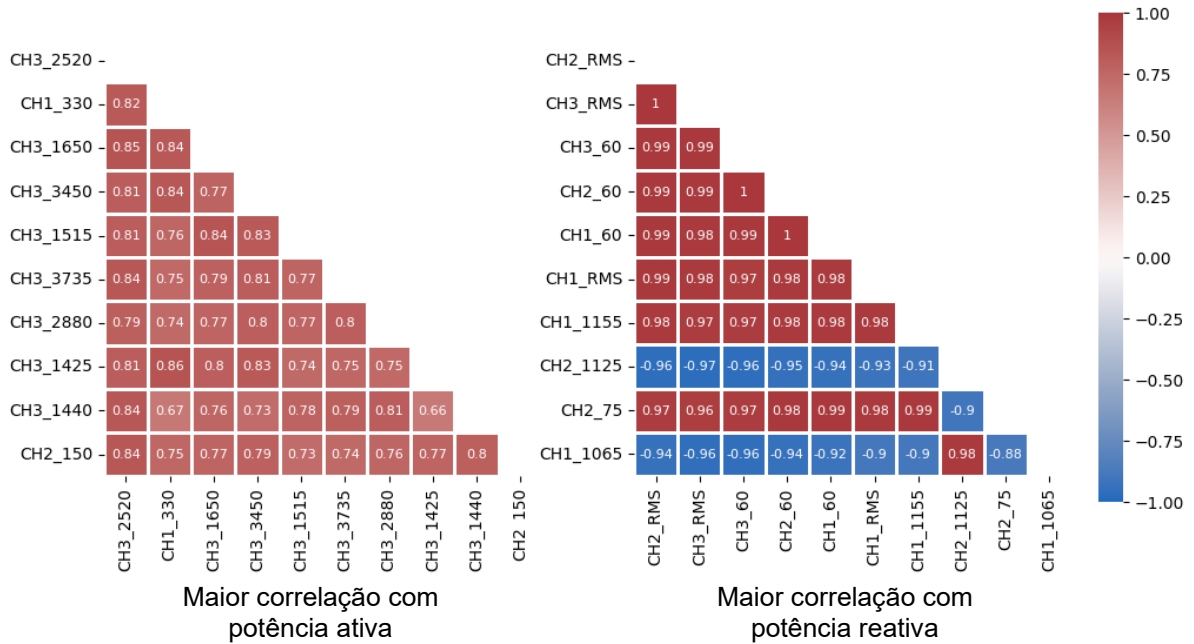
(a)



(b)

Fonte: elaborado pelo autor.

Figura 35 - Correlações entre atributos mais correlacionados com as potências pelo método de Pearson.



Fonte: elaborado pelo autor.

4.1.3.2 Correlação de Spearman

A correlação de Spearman é utilizada para avaliar a relação entre duas variáveis. A diferença para a correlação de Pearson reside no fato de que a correlação de Spearman indica relações monotônicas, i.e., duas variáveis aumentam ou decrescem de valor conjuntamente, mas não necessariamente com a mesma taxa. Em outras palavras, em vez de trabalhar diretamente com as amostras de dados, como é o caso da correlação de Pearson, a correlação de Spearman opera nas classificações relativas dos valores dos dados.

O coeficiente de correlação ρ (ρ de Spearman) pode variar de -1 (indicando uma forte correlação negativa) a 1 (indicando uma forte correlação positiva). Uma correlação com valor igual a zero indica a inexistência de relação monotônica entre as variáveis. Esta correlação também é encontrada como correlação de *rankings* de Spearman justamente por seu cálculo ser feito após se ordenar os dados das variáveis [34], [41], [42], [43] e [44]. A equação (7) apresenta como é calculado para duas variáveis x e y o coeficiente ρ de Spearman [45].

$$\rho = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \quad (7)$$

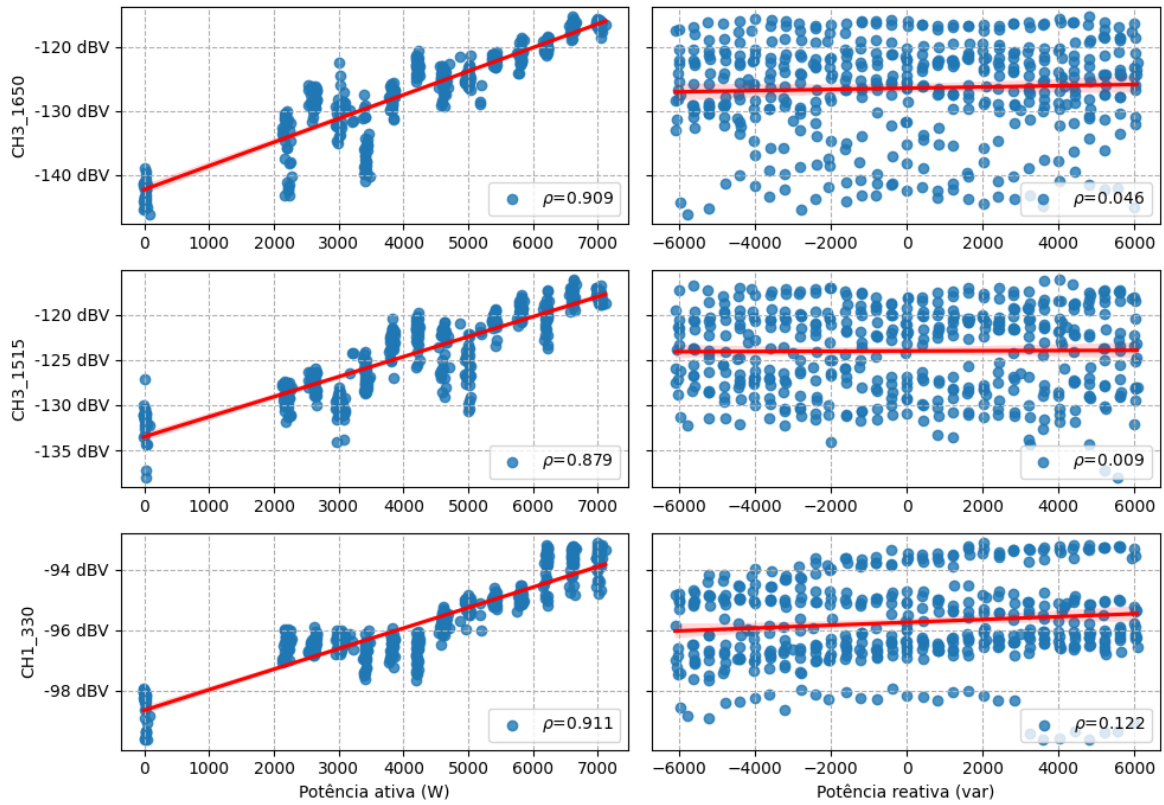
onde:

- d_i é a diferença entre $rank(x_i)$ e $rank(y_i)$, sendo definido como $d_i = (rank(x_i) - rank(y_i))$
- x_i é o i ésimo elemento de x , y_i é o i ésimo elemento de y .
- n é o número total de amostras.

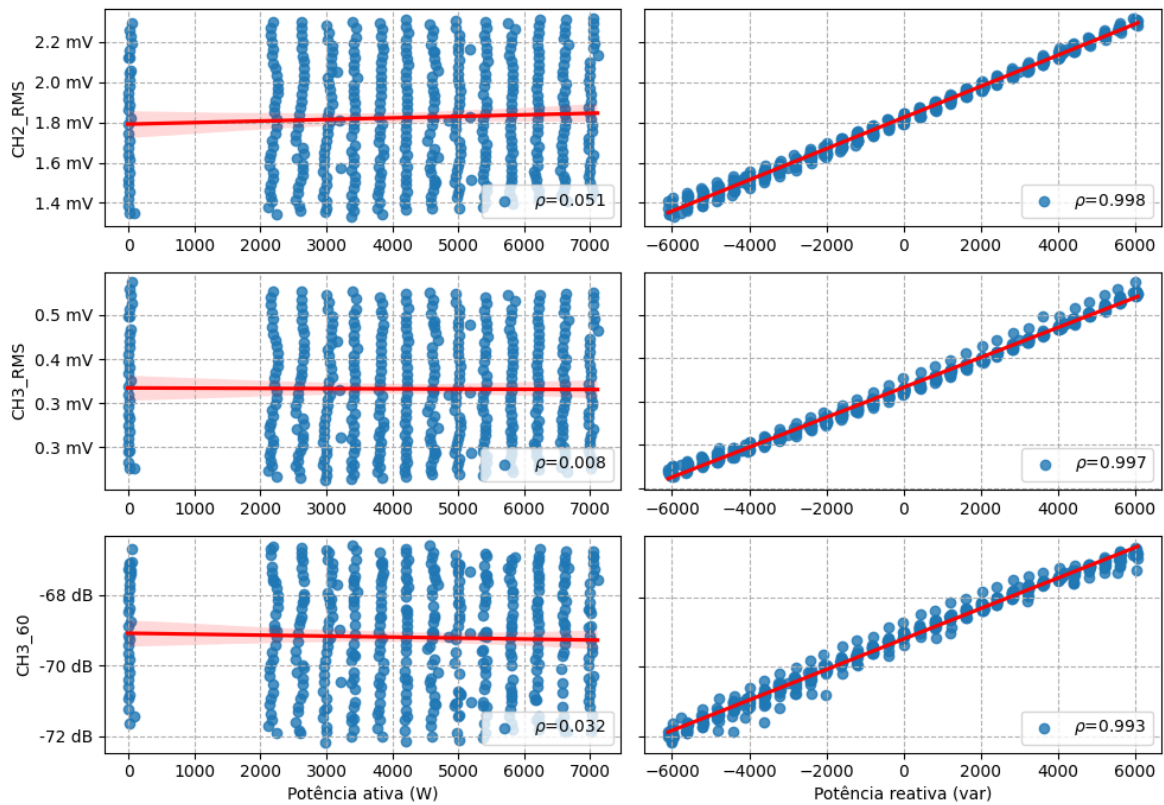
Outra interpretação para calcular o ρ de Spearman é atribuir um *rank* a cada variável e calcular a correlação de Pearson com o *ranking* dos dados ao invés de utilizar os próprios dados [33].

Os resultados obtidos com o cálculo da correlação de Spearman para o conjunto de dados independentes do tempo foram ranqueados em ordem decrescente. A Figura 36 (a) apresenta os três atributos mais correlacionados com a potência ativa e a Figura 36 (b) os três atributos mais correlacionados com a potência reativa. As linhas vermelhas indicam a regressão pelo método dos mínimos quadrados e o coeficiente de correlação está contido em cada gráfico. A Figura 37 apresenta as correlações entre as variáveis mais correlacionadas com as potências ativa e reativa, respectivamente pelo método de Spearman.

Figura 36 - Atributos mais relacionadas com a potência ativa e reativa de acordo com a correlação de Spearman.



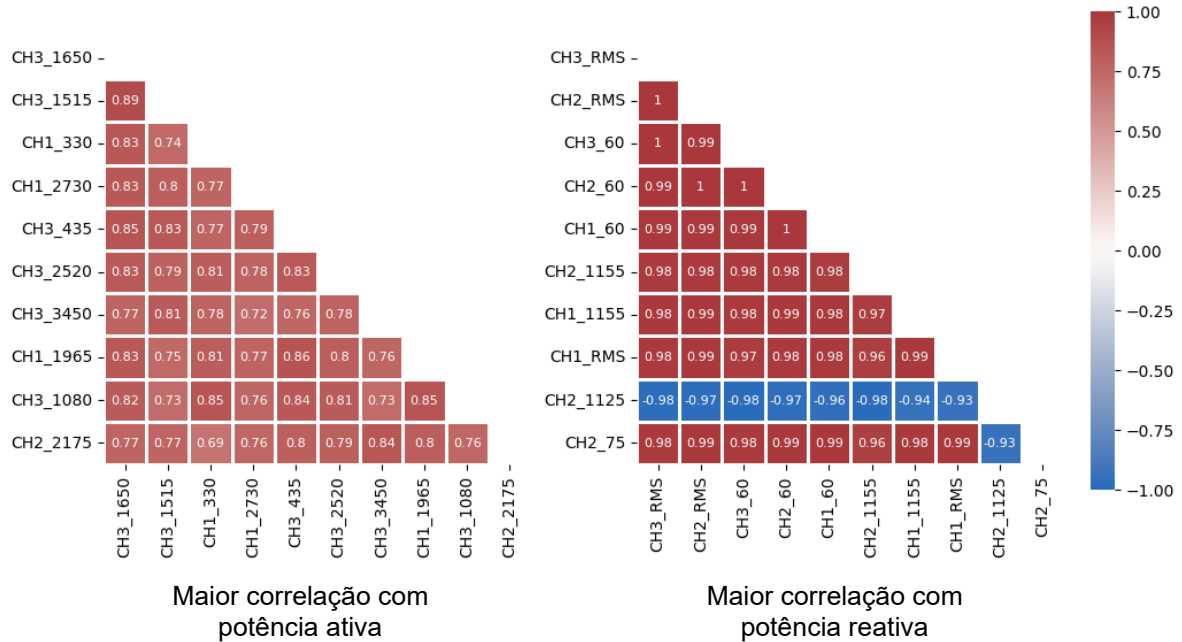
(a)



(b)

Fonte: elaborado pelo autor.

Figura 37 - Correlações entre atributos mais correlacionados com as potências pelo método de Spearman.



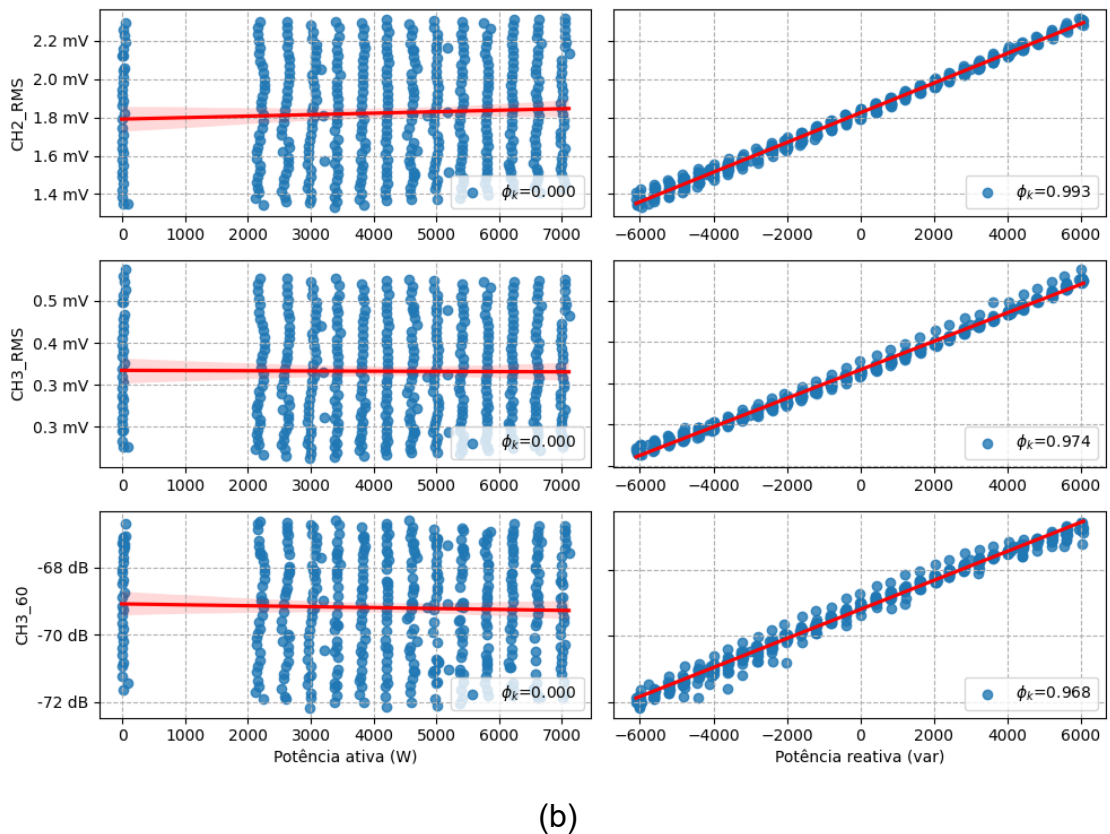
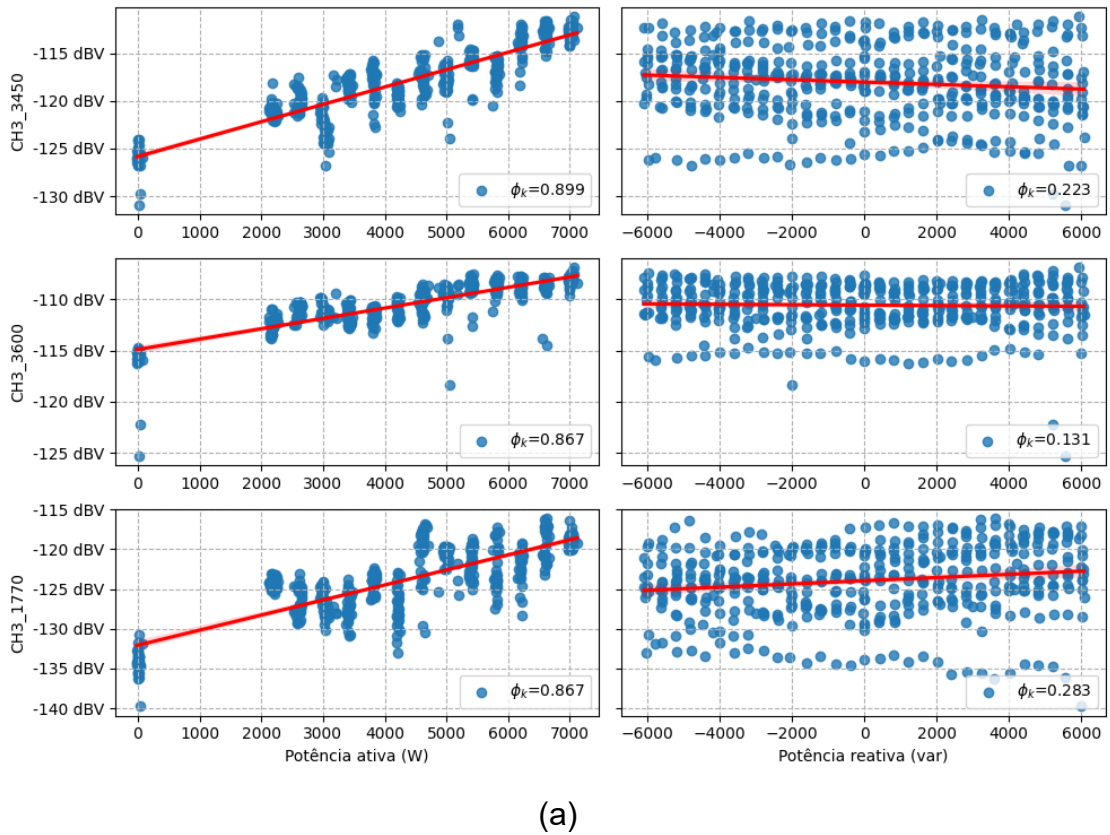
Fonte: elaborado pelo autor.

4.1.3.3 Correlação PhiK

O coeficiente de correlação PhiK (ϕ_K) foi desenvolvido para trazer a simplicidade da análise da correlação de Pearson, estendendo o seu uso para relações não-lineares entre variáveis, bem como quantificar relações entre variáveis numéricas e categóricas [34], [41], [42], [43], [44] e [33]. O coeficiente de correlação ϕ_K pode variar de 0 (indicando a inexistência de relação entre as duas variáveis) a 1 (indicando uma forte relação). O procedimento de cálculo do coeficiente ϕ_K é apresentado em detalhes em [33].

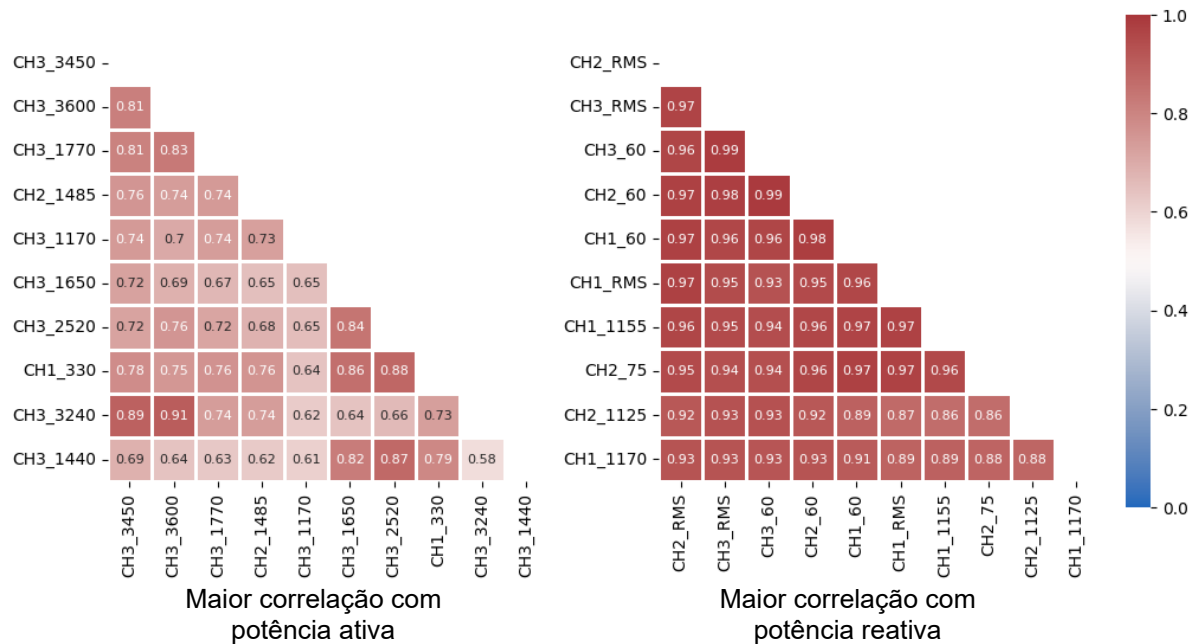
A correlação de PhiK é útil em encontrar relações não lineares entre os dados. Portanto, serve como informação complementar às correlações de Pearson e Spearman. A Figura 38 (a) apresenta os três atributos mais correlacionados com a potência ativa e a Figura 38 (b) os três atributos mais correlacionados com a potência reativa. As linhas vermelhas indicam a regressão pelo método dos mínimos quadrados e o coeficiente de correlação está contido em cada gráfico. A Figura 39 apresenta as correlações entre as variáveis mais correlacionadas com as potências ativa e reativa, respectivamente pelo método de PhiK.

Figura 38 - Atributos mais relacionadas com a potência ativa e reativa de acordo com a correlação de PhiK.



Fonte: elaborado pelo autor.

Figura 39 - Correlações entre atributos mais correlacionados com as potências pelo método de PhiK.



Fonte: elaborado pelo autor.

4.1.3.4 Considerações sobre a correlação dos dados

Com o estudo de correlação foi possível concluir que os dados do conjunto independente do tempo contêm informações relevantes sobre o ponto de operação do GSPS. Observou-se que, em geral, os atributos mais correlacionados com a potência ativa possuem baixa relação com a potência reativa e os atributos mais correlacionados com a potência reativa possuem baixa relação com a potência ativa, conforme apresentado nas figuras Figura 34, Figura 36 e Figura 38.

Os atributos que apresentam alta correlação com a potência ativa são bons candidatos para serem utilizados em um modelo de regressão linear para estimar a potência ativa do ponto de operação desta máquina. A mesma lógica pode ser aplicada para os atributos que apresentam alta correlação com a potência reativa.

Conforme apresentado nas figuras Figura 35, Figura 37 e Figura 39 foi possível observar que os atributos que apresentam alta correlação com apenas uma das potências, apresentam alta correlação entre si. Esta alta correlação entre atributos é um forte indicativo de multicolinearidade. Assim, entende-se que a informação contida nestes atributos tende a ser redundante. Ao adicionar esses atributos à um modelo que estima alguma potência, os coeficientes do modelo tenderão a ser instáveis.

Para todos os coeficientes de correlação calculados foi possível observar que os atributos mais correlacionados com a potência ativa apresentam uma relação geralmente não linear em contraste com os atributos mais correlacionados com a potência reativa, que são próximos de uma relação linear.

O valor RMS do campo contém uma forte correlação com a potência reativa, independentemente da posição do sensor. A relação entre as variáveis é claramente linear, o que explica o coeficiente de Pearson estar próximo da unidade. Também foi possível observar pela Figura 27 que o valor RMS do campo não apresenta relação significativa com a potência ativa para esta máquina ensaiada.

Outro ponto a ser considerado dentro da análise de correlação é a amplitude dos atributos. Observa-se que a maior parte dos atributos mais correlacionados com a potência ativa apresentam amplitudes relativamente baixas, geralmente entre - 90 dBV e - 100 dBV.

Nas tabelas Tabela 3 e Tabela 4 são apresentados os 15 atributos com o maior coeficiente de correlação para cada tipo de correlação. Os valores apresentados são absolutos. Assim, também estão nesta lista os módulos dos atributos que são fortemente relacionados negativamente com as potências. Com os valores apresentados na Tabela 3 e na Tabela 4, pode-se analisar de maneira simplificada o posicionamento dos sensores indutivos ao redor do GSPS. Para a potência ativa, o canal 3 (sensor posicionado na cabeça de bobina) foi o posicionamento que apresentou maior correlação de modo geral. Para a potência reativa, os canais 1 e 2 foram os posicionamentos que apresentaram maior correlação, além do valor RMS e a harmônica elétrica de 60 Hz dos três canais, que apresentaram coeficientes de correlação próximos à 1 para os três tipos de correlação. Então, entende-se que a maior parte da informação sobre a potência reativa é encontrada do valor RMS lido por um sensor instalado de maneira arbitrária ao redor do GSPS.

Tabela 3 - Comparação entre os coeficientes de correlação para a potência ativa.

Rank	Pearson		Spearman		PhiK	
	Atributo	Coeficiente de correlação	Atributo	Coeficiente de correlação	Atributo	Coeficiente de correlação
1	CH3_2520	0,920	CH3_1650	0,923	CH3_3450	0,899
2	CH1_330	0,910	CH3_1515	0,892	CH3_3600	0,867
3	CH3_1650	0,901	CH1_330	0,889	CH3_1770	0,867
4	CH3_3450	0,891	CH1_2730	0,886	CH2_1485	0,848
5	CH3_1515	0,879	CH3_435	0,884	CH3_1170	0,844
6	CH3_3735	0,875	CH3_2520	0,878	CH3_1650	0,834
7	CH3_2880	0,874	CH3_3450	0,875	CH3_2520	0,830
8	CH3_1425	0,863	CH1_1965	0,860	CH1_330	0,826
9	CH3_1440	0,860	CH3_1080	0,849	CH3_3240	0,821
10	CH2_150	0,858	CH2_2175	0,847	CH3_1440	0,819
11	CH2_120	0,854	CH3_1440	0,847	CH3_1245	0,814
12	CH1_375	0,848	CH3_3735	0,846	CH2_3900	0,812
13	CH1_915	0,844	CH2_150	0,845	CH3_1575	0,809
14	CH3_3600	0,838	CH3_3600	0,839	CH3_2880	0,805
15	CH3_585	0,833	CH1_765	0,837	CH3_195	0,798

Fonte: elaborado pelo autor

Tabela 4 - Comparação entre os coeficientes de correlação para a potência reativa.

Rank	Pearson		Spearman		PhiK	
	Atributo	Coeficiente de correlação	Atributo	Coeficiente de correlação	Atributo	Coeficiente de correlação
1	CH2_RMS	0,997	CH3_RMS	0,997	CH2_RMS	0,993
2	CH3_RMS	0,997	CH2_RMS	0,997	CH3_RMS	0,974
3	CH3_60	0,993	CH3_60	0,994	CH3_60	0,968
4	CH2_60	0,992	CH2_60	0,994	CH2_60	0,967
5	CH1_60	0,986	CH1_60	0,988	CH1_60	0,958
6	CH1_RMS	0,980	CH2_1155	0,987	CH1_RMS	0,955
7	CH1_1155	0,978	CH1_1155	0,984	CH1_1155	0,950
8	CH2_1125	0,971	CH1_RMS	0,981	CH2_75	0,939
9	CH2_75	0,967	CH2_1125	0,980	CH2_1125	0,939
10	CH1_1065	0,957	CH2_75	0,977	CH1_1170	0,934
11	CH1_15	0,957	CH3_1035	0,966	CH1_15	0,921
12	CH3_1035	0,943	CH1_15	0,964	CH3_1035	0,915
13	CH1_1020	0,937	CH1_1065	0,960	CH1_1065	0,901
14	CH1_2235	0,933	CH1_2235	0,945	CH2_195	0,886
15	CH2_1020	0,930	CH2_1035	0,945	CH2_2190	0,880

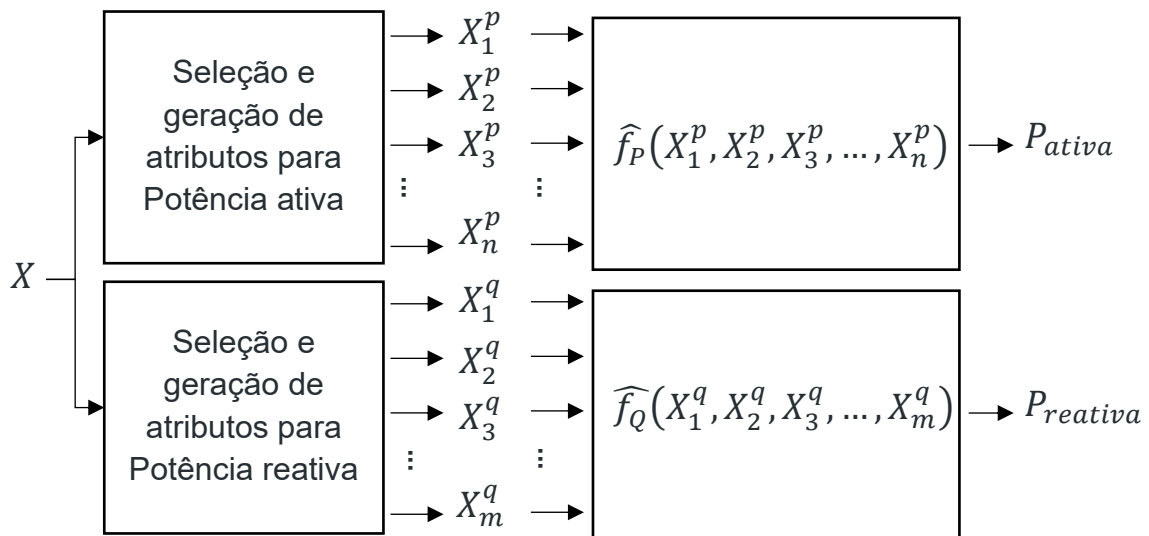
Fonte: elaborado pelo autor.

4.2 MODELAGEM DOS DADOS DE CAMPO

Esta seção tem por objetivo apresentar os procedimentos adotados para gerar os modelos de regressão linear que estimem o ponto de operação do GPS. De maneira simplificada, deseja-se encontrar uma função cujos coeficientes são ajustados com parte do conjunto de dados, durante uma fase de treinamento. Posteriormente, os dados restantes são utilizados para testar e quantificar, com alguma métrica de desempenho, a qualidade com que esta função representa o conjunto total de dados.

Conforme verificado no estudo de correlação, dentro do conjunto de dados existem atributos que possuem maior correlação com a potência ativa, assim como existem atributos que possuem maior correlação com a potência reativa. De maneira geral, observou-se que quando um atributo possui alta correlação com um tipo de potência, este mesmo atributo possui baixa correlação com o outro tipo de potência. Portanto, para encontrar uma função que estime o ponto de operação por meio do campo magnético externo, optou-se por gerar dois modelos distintos: um para a potência ativa e um para a potência reativa, conforme apresentado pelo esquema na Figura 40. O conjunto de dados independentes do tempo, representado pelos atributos X , contém as amplitudes das harmônicas e o valor RMS dos vetores de tensão dos três canais. Os atributos X_i^p e X_j^q , onde $i=1, 2, 3, \dots, n$ e $j=1, 2, 3, \dots, m$, são os selecionados para serem utilizados nos modelos \hat{f}_p e \hat{f}_q respectivamente, sendo \hat{f}_p a função que estima a potência ativa (P_{ativa}) e \hat{f}_q a função que estima a potência reativa ($P_{reativa}$).

Figura 40 – Escolha dos modelos a ser utilizado.



Fonte: elaborado pelo autor.

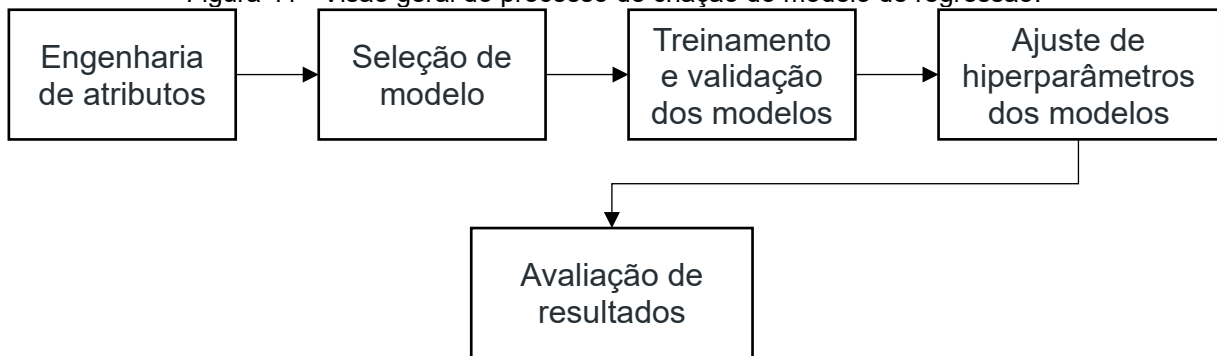
A metodologia utilizada para obter um modelo capaz de estimar o ponto de operação do GPS a partir do conjunto de dados independentes do tempo está apresentada na Figura 41. Essa metodologia é baseada no trabalho [46] que também visa resolver um problema de regressão para estimação de duas variáveis.

Dentro do bloco Engenharia de atributos, apresentado na Figura 41, são aplicadas técnicas para escolher os atributos que contêm a maior quantidade de informação do conjunto de dados. Também são utilizadas técnicas para gerar novos atributos não encontrados originalmente no conjunto de dados. Estes novos atributos devem capturar a informação contida em um conjunto de alta dimensão e trazer para uma dimensão menor. Em seguida, os dados são divididos em conjuntos de treino e teste. A partir desse ponto, os dados de teste são deixados de lado e somente serão utilizados novamente ao final do processo de criação do modelo de regressão.

Dentro da seleção de modelo, listam-se os diferentes modelos para um sistema de várias entradas e uma única saída.

As etapas de treinamento, validação e ajuste de hiperparâmetros são realizadas em sequência com o conjunto de dados separado para treino. Hiperparâmetros são parâmetros do modelo que não fazem parte do objetivo de minimização da função custo da regressão linear. Cada modelo recebe como entrada os atributos selecionados pelo passo anterior e seus hiperparâmetros são ajustados de maneira a cobrir uma grade de atributos, aplicando a técnica de *Grid Search* [47] em combinação com uma validação cruzada *k-fold* [48]. As técnicas *Grid Search* e *k-fold* serão apresentadas adiante. Após encontrados os hiperparâmetros ótimos, na etapa de ajuste dos hiperparâmetros, o modelo será utilizado para prever os dados do conjunto de dados para teste. Essa técnica é chamada de *hold-out* por manter fora do conjunto de treino uma parte do conjunto de dados para uma avaliação final de desempenho do modelo.

Figura 41 - Visão geral do processo de criação do modelo de regressão.



Fonte: Adaptado de [46].

A Figura 42 apresenta o fluxo do processamento de dados utilizado neste trabalho. Primeiramente, os dados do conjunto de atributos X (apresentado na Figura 40 e composto por amplitudes das harmônicas e valores RMS) são padronizados. Em seguida, são gerados novos atributos com as técnicas *Principal Component Analysis* (PCA), *t-distributed Stochastic Neighbor Embedding* (t-SNE) e *Uniform Manifold Approximation and Projection* (UMAP).

O conjunto formado pelos atributos originais padronizados e os atributos gerados com aplicação de PCA, t-SNE e UMAP, é dividido em um conjunto de treino e um conjunto de teste. O conjunto de treino é utilizado para a validação cruzada e para encontrar os hiperparâmetros que resultam nas melhores métricas de desempenho. O conjunto de teste é separado e será utilizado somente ao final para avaliar os melhores modelos encontrados.

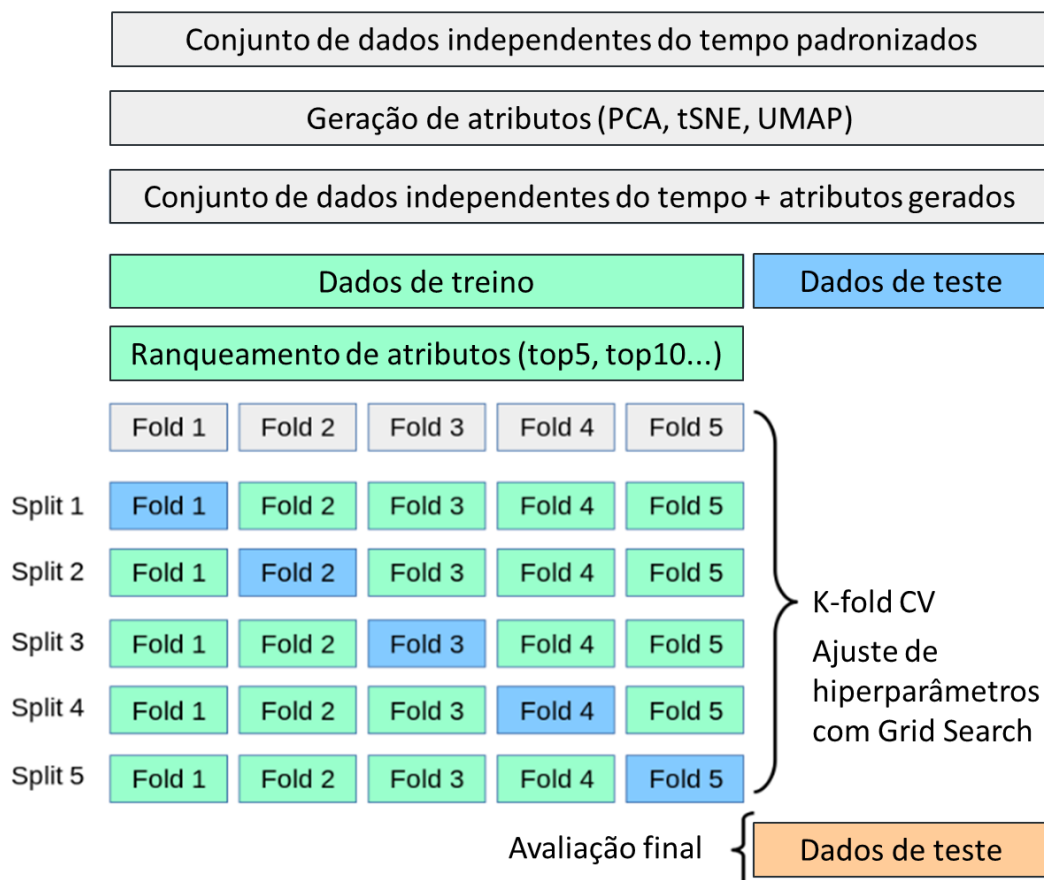
Os atributos do conjunto de treino foram avaliados com a biblioteca *Selective* [49] e correlação para identificar quais atributos possuem maior relevância. Desta maneira é possível ranquear os atributos. Com os atributos ranqueados, aplica-se a técnica de validação cruzada *k-fold* para determinar a quantidade de atributos e ajustar os hiperparâmetros de cada modelo. Por fim, são escolhidos os modelos que obtiveram as melhores métricas de desempenho dentro da validação cruzada *k-fold* para serem aplicados ao conjunto de teste. O modelo selecionado foi aquele que apresentou a melhor média das métricas de desempenho entre o treino e o teste.

A técnica de validação cruzada *k-fold* consiste em repartir o conjunto de dados em k partes (ou divisões), onde uma dessas partes é utilizada para validação (blocos azuis na Figura 42 dentro do conjunto de treino) e o restante para treinamento do modelo (blocos verdes na Figura 42 dentro do conjunto de treino). O treino e a validação devem se repetir k vezes até que se esgote o conjunto de dados. As k vezes

que o treinamento ocorre não são sobrepostas, i.e., o treinamento sempre ocorre com conjuntos diferentes [48]. Ao final, é possível calcular as médias e os desvios padrão para as métricas de desempenho escolhidas.

A técnica de *Grid Search* é utilizada para ajustar os hiperparâmetros de um modelo. Geralmente é utilizada em conjunto com a validação cruzada *k-fold* por este ser um método robusto para extrair a métrica de desempenho desejada e avaliar os hiperparâmetros escolhidos. Neste sentido, para a aplicação da técnica de *Grid Search* é construído um conjunto de hiperparâmetros, onde cada hiperparâmetro possui um grupo de candidatos. Como exemplo, em uma regressão linear com penalização L2 (*Ridge*), existe o hiperparâmetro *alpha* que pode variar entre 10^{-3} e 100. Assim, um grupo de candidatos possíveis para *alpha* é dado por: $[10^{-3}, 10^{-2}, 10^{-1}, 1, 2, 5, 15, 30, 60, 100]$. Com esse grupo de candidatos para *alpha* é possível aplicar uma validação cruzada *k-fold* por candidato e obter suas métricas de desempenho. Ao final, o candidato que fornece o melhor desempenho é escolhido como hiperparâmetro *alpha*.

Figura 42 – Fluxo de processamento de dados: validação cruzada *k-fold* com teste final *hold-out*.



Fonte: Adaptado de [48].

Além da validação cruzada, existe um teste final que é feito com um conjunto de dados separados no início do processo de modelagem. A aplicação da técnica de *hold-out* é a parte final do procedimento apresentado no diagrama da Figura 42. Um ponto importante a ser observado, é que a seleção de atributos deve ser feita após o conjunto ser separado em teste e validação a fim de evitar que a informação do conjunto de teste “vaze” para dentro do conjunto de treino. Optou-se por deixar 20% do conjunto total como dados de teste e os 80% restantes para a validação cruzada e ajuste de hiperparâmetros. Isso representa 89 observações (pontos de operação) para teste e 355 observações para a validação cruzada.

4.2.1 Engenharia de atributos

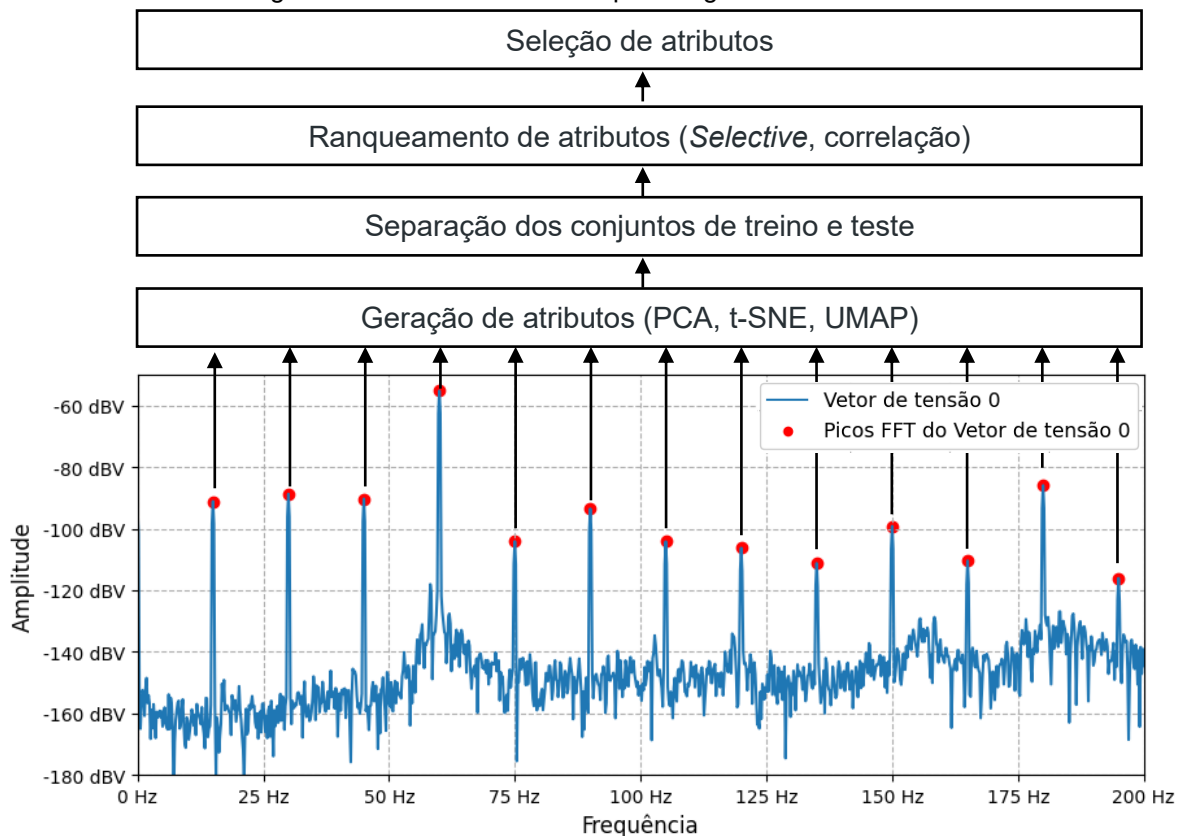
O desempenho do modelo de regressão depende de bons atributos selecionados para que durante a fase de treinamento consiga capturar a maior quantidade de informação disponível e, desta forma, atingir um bom desempenho de generalização. Conforme discutido anteriormente, o conjunto de dados independentes do tempo possui o problema de alta dimensionalidade. São 804 atributos (amplitudes de harmônicas e valores RMS) em 444 observações (pontos de operação). Na literatura especializada, este tipo de conjunto de dados é conhecido como “ $p \gg n$ ”, onde p é o número de atributos (*predictors*) e n o número de observações (*samples*) [50]. O resultado esperado ao se utilizar tal conjunto de dados é um comportamento inesperado dos modelos, visto que a maioria dos modelos conhecidos assumem $p \ll n$. Ao utilizar o conjunto total de dados em um modelo de regressão linear é esperado um *overfit*, i.e., dentro do conjunto separado para treino, o modelo terá um bom desempenho. Porém, ao utilizar este modelo no conjunto separado para teste, o seu desempenho será ruim. Além disso, utilizar atributos irrelevantes ou que adicionem informação redundante resulta em um pior desempenho do modelo de regressão linear [50].

Conjuntos de dados de dimensão elevada são difíceis de serem visualizados. Para auxiliar na visualização da estrutura de um conjunto de dados, a dimensão deve ser reduzida. Algumas técnicas, tais como PCA, t-SNE e UMAP, diminuem a dimensão do conjunto de dados. Tais técnicas, que serão apresentadas adiante, possibilitam obter a maior parte da informação com poucos atributos gerados a partir do conjunto original, facilitando assim a sua visualização em duas ou três dimensões [51], [52], [53], [54], [55].

A escolha dos atributos que apresentam o maior coeficiente de correlação de Pearson é um bom ponto de partida para modelos que utilizam a linearidade entre duas variáveis. Pode-se dizer que atributos com alta correlação de PhiK serão bons para modelos que utilizam relações não lineares entre as variáveis. Apesar do estudo de correlação indicar quais atributos apresentam forte correlação com o ponto de operação, foram utilizadas outras técnicas, além da correlação, para escolher os atributos do modelo de regressão.

A Figura 43 apresenta uma simplificação de como foi implementado o processo de engenharia de atributos. As técnicas de PCA, t-SNE e UMAP foram utilizadas com os 804 atributos para capturar a informação com poucos atributos. Esses novos atributos foram adicionados ao conjunto de dados. Após a separação em conjuntos de treino e teste, um ranqueamento foi realizado utilizando a correlação dos dados e a biblioteca *Selective* [49] no conjunto de treino. Por fim a quantidade de atributos foi selecionada como um hiperparâmetro durante a validação cruzada utilizando o conjunto de dados separado para treino durante as etapas de treinamento e validação e ajuste de hiperparâmetros.

Figura 43 – Processo adotado para engenharia de atributos.



Fonte: elaborado pelo autor.

4.2.1.1 PCA – Principal Componente Analysis

Considerada por [56] como sendo a técnica mais popular de redução de dimensionalidade, o PCA é uma técnica de álgebra linear utilizada para criar uma projeção do conjunto de dados originais em um subespaço de menor dimensão. Os dados originais são linearmente projetados em um espaço de menor dimensão de forma que a variância seja máxima. A implementação do PCA é descrita de maneira prática e em detalhes em [57]. Neste trabalho foi utilizada a implementação do PCA na biblioteca *scikit learn* [58].

Antes de aplicar a técnica de PCA ao conjunto de dados é preciso padronizar ou normalizar os dados. Este passo inicial é considerado uma boa prática conforme apresentado por [56], [57], [59] por restringir as variações e as ordens de grandeza dos dados. Atribui-se o termo padronização ao processo de mapear os dados para média igual a zero e desvio padrão unitário. A padronização dos dados é calculada a partir da Equação (8) [60] e é aplicada quando os dados possuem distribuição gaussiana, ou muito próxima a uma gaussiana.

$$Z_{Xstandard} = \frac{(X - \bar{X})}{\sigma_x} \quad (8)$$

onde:

- $Z_{Xstandard}$ é um atributo padronizado para média zero e desvio padrão unitário.
- X é atributo original.
- \bar{X} é a média do atributo original.
- σ_x é o desvio padrão do atributo original.

Uma alternativa para a padronização é a normalização dos dados. A normalização geralmente é preferida quando a distribuição dos dados segue um padrão uniforme. A normalização é calculada a partir da Equação (9) [60].

$$Z_{Xnorm} = \frac{X - X_{min}}{X_{max} - X_{min}} (t_{max} - t_{min}) + t_{min} \quad (9)$$

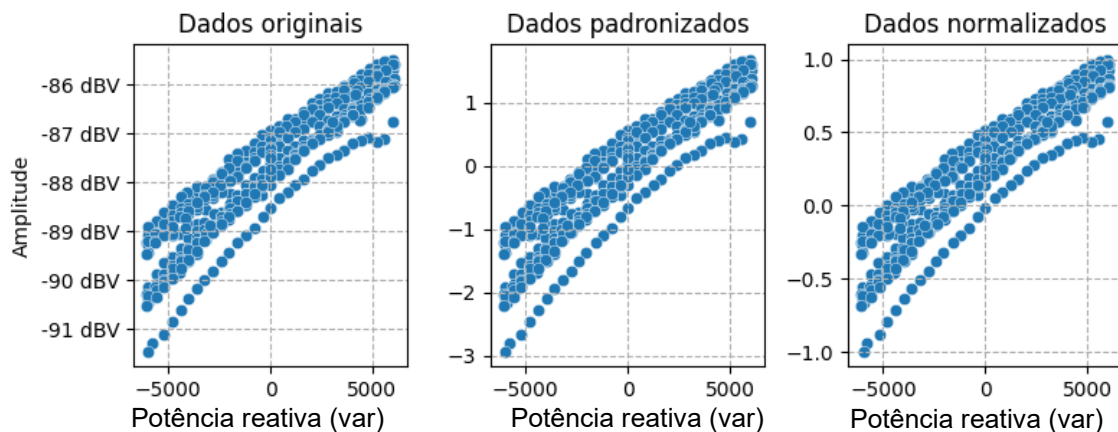
onde:

- Z_{Xnorm} é um atributo normalizado para o *range* $[t_{min}; t_{max}]$.
- X é atributo original.
- X_{min} é o valor mínimo de X .
- X_{max} é o valor máximo de X .

- t_{max} é novo valor máximo desejado para o remapeamento de X .

No caso da normalização, t_{min} e t_{max} podem ser substituídos por -1 e 1, respectivamente, resultando na normalização com remapeamento para [-1;1]. A Figura 44 apresenta a comparação dos dados padronizados e normalizados com os dados originais para a harmônica de 45 Hz do canal 1, sendo o eixo x a potência reativa. Observa-se que a forma dos dados permanece constante, porém os valores do eixo y são mapeados para novas escalas.

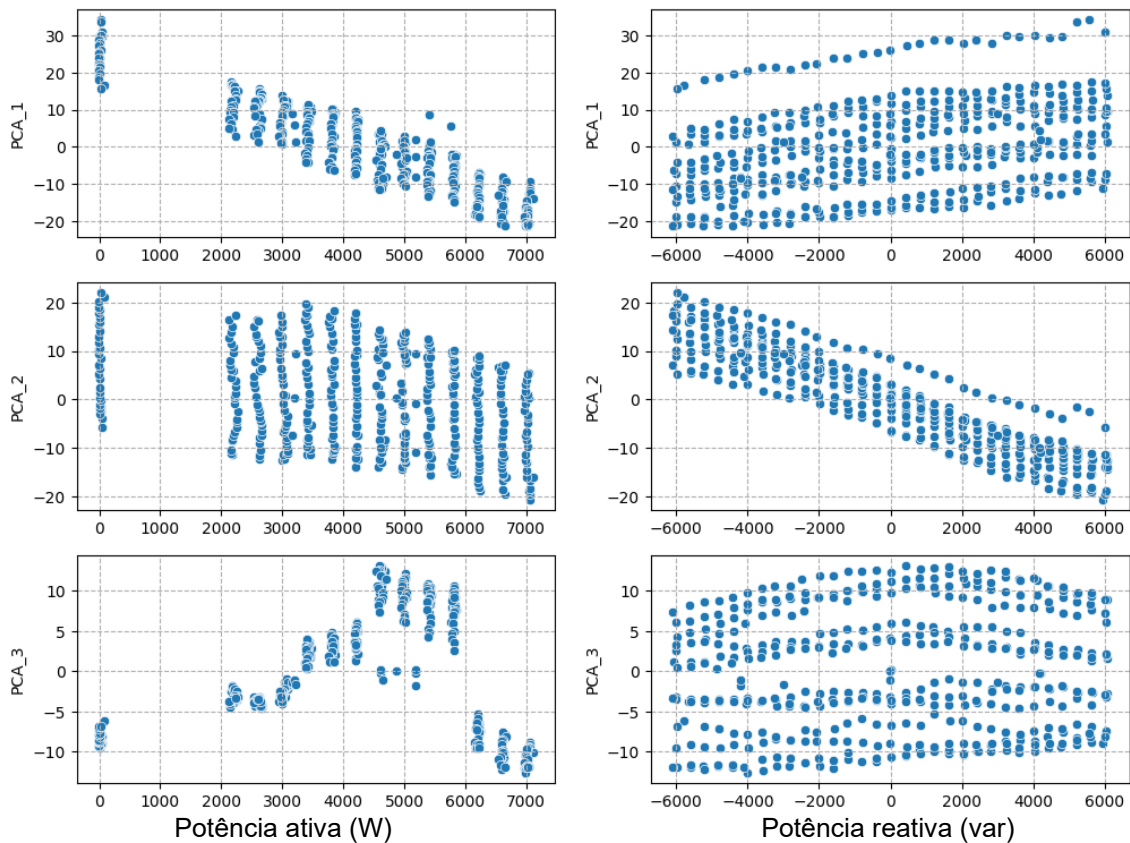
Figura 44 - Dados originais, padronizados e normalizados do canal 1 para a harmônica de 45 Hz.



Fonte: elaborado pelo autor.

A Figura 45 apresenta os três primeiros componentes principais do conjunto de dados, sendo PCA_1 o primeiro componente, PCA_2 o segundo componente e PCA_3 o terceiro componente. Os três atributos apresentados conseguem explicar aproximadamente 43% da variância do conjunto de dados. Para ser possível cobrir 70% da variância são necessários 19 componentes, e para cobrir 90%, 66 componentes.

Figura 45 - Exemplo dos três primeiros componentes principais dos dados padronizados.



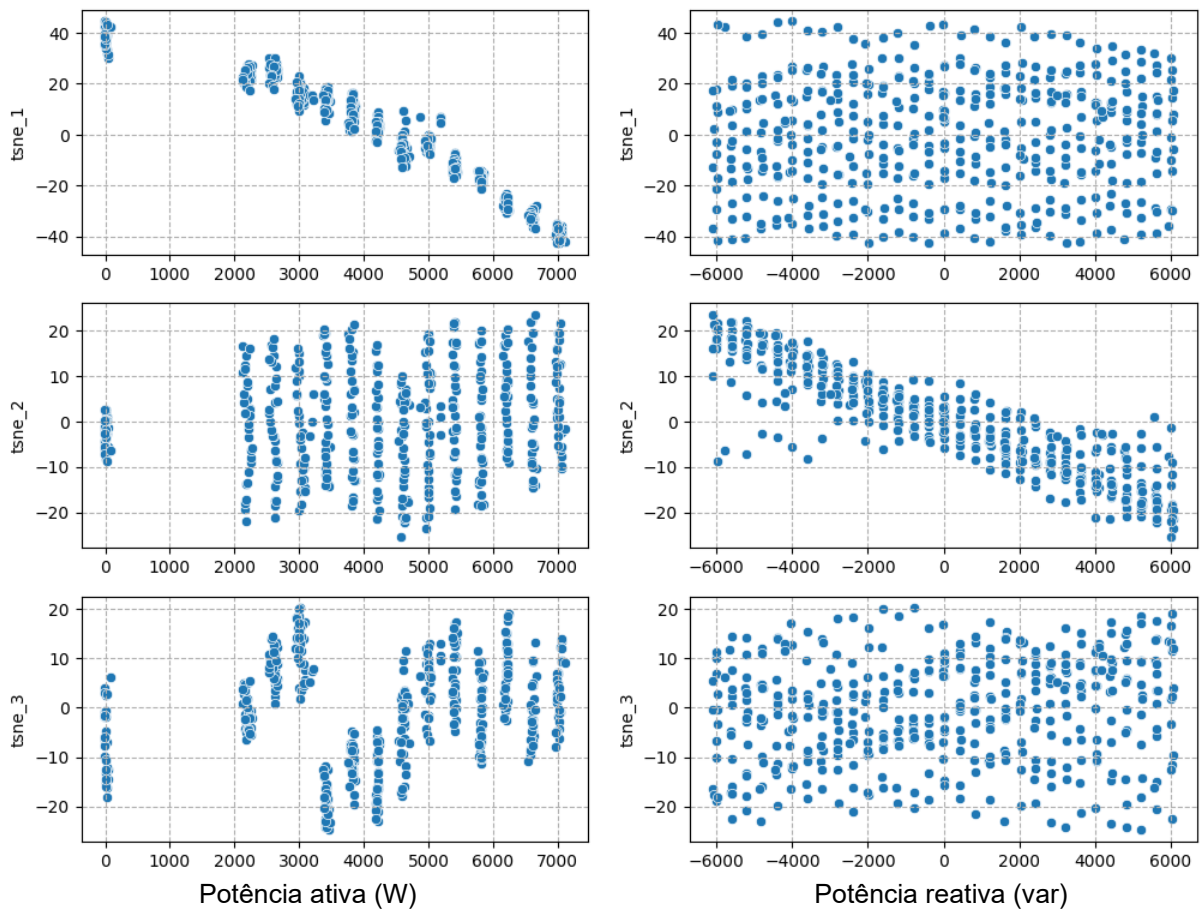
Fonte: elaborado pelo autor.

4.2.1.2 *t*-SNE – *t*-distributed Stochastic Neighbor Embedding

A técnica t-SNE converte as afinidades dos pontos de dados em probabilidades. As afinidades no espaço original são representadas por probabilidades conjuntas gaussianas e as afinidades no espaço incorporado são representadas por distribuições t de *Student*. Isso permite que o t-SNE seja particularmente sensível à estrutura local e tenha algumas outras vantagens em relação às técnicas existentes como, por exemplo, revelar estruturas e relações existentes em escalas diferentes em um único mapa [53], [54], [61], [62]. Outro detalhe importante sobre o t-SNE é que o algoritmo possui uma variação estocástica. Por otimizar uma função custo não convexa, os atributos resultantes da aplicação do método t-SNE são dependentes da condição inicial do algoritmo. O estado aleatório inicial pode ser travado por uma variável chamada *random_state*. Ao atribuir um valor inteiro à essa variável é assegurada a reprodutibilidade do algoritmo por todas as vezes que o algoritmo for aplicado. Outra variável de entrada do t-SNE é a perplexidade (*perplexity*) que pode ser interpretada como um foco que o algoritmo tem para variações locais ou globais no conjunto de dados. Recomenda-se que o valor de

perplexity esteja entre 5 e 50 [63]. Dentro do algoritmo, ainda existe, por padrão, uma aproximação pelo método de Barnes-Hut que permite que sejam escolhidos apenas três componentes [64], [65]. Assim como foi feito para o PCA, antes de aplicar a técnica t-SNE, os dados foram padronizados. A Figura 46 apresenta os três componentes gerados com t-SNE a partir do conjunto de dados original, sendo t-SNE_1 o primeiro componente, t-SNE_2 o segundo componente e t-SNE_3 o terceiro componente.

Figura 46 – Atributos gerados com t-SNE.



Fonte: elaborado pelo autor.

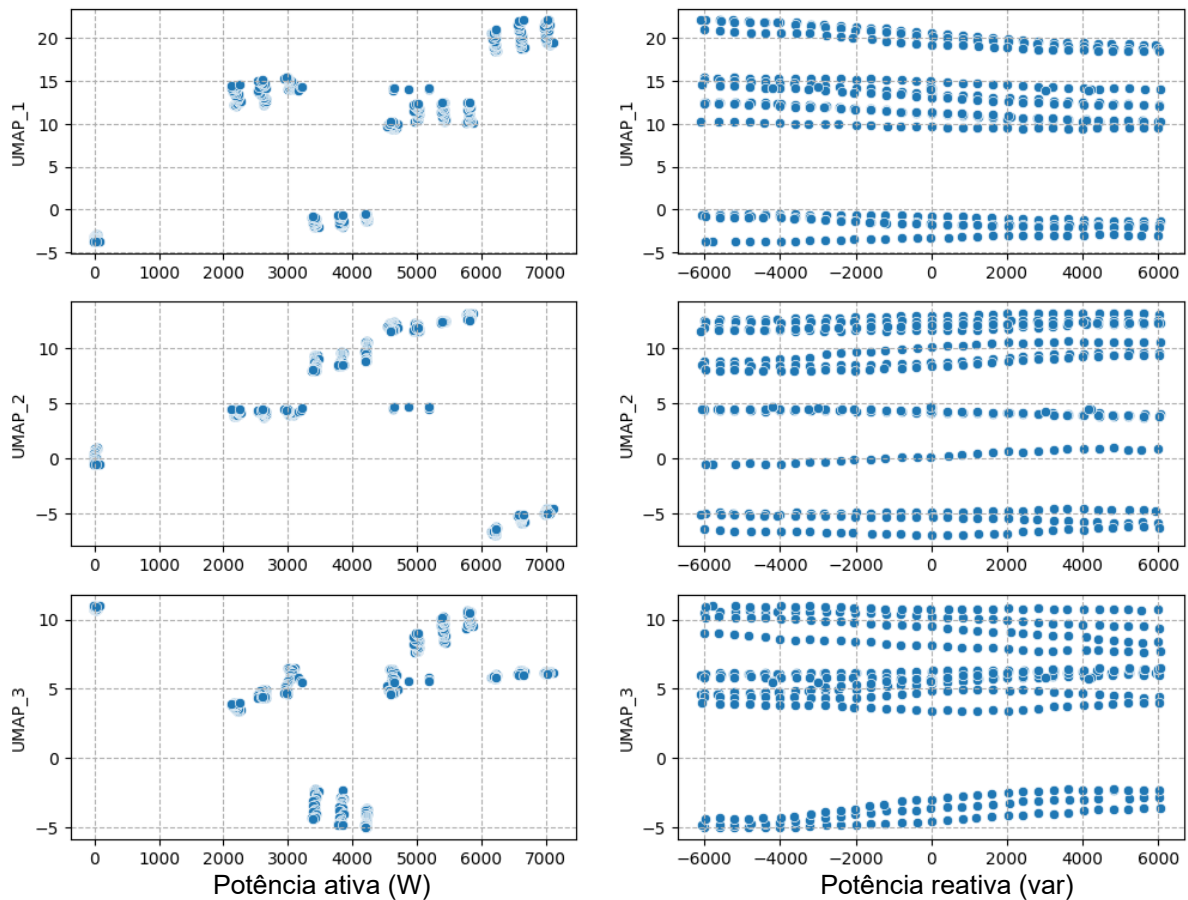
4.2.1.3 UMAP – Uniform Manifold Approximation and Projection

A técnica UMAP é utilizada neste trabalho como uma alternativa ao t-SNE para tentar capturar as não linearidades nos dados em um subespaço de menor dimensão que o conjunto original. Esta técnica apresenta uma melhor preservação das características do conjunto de dados e um menor custo computacional quando comparada ao t-SNE [66], [67].

As técnicas de *manifold learning* são uma abordagem para a redução não linear da dimensionalidade. Os algoritmos para essa tarefa são baseados na ideia de que a dimensionalidade de muitos conjuntos de dados é apenas artificialmente alta. Essas técnicas são usadas para criar uma projeção de baixa dimensão a partir de conjunto dados de alta dimensão, geralmente para fins de visualização. A projeção é utilizada para criar uma representação de baixa dimensão do conjunto de dados e, ao mesmo tempo, preservar suas características globais. Por isso, tais técnicas serão utilizadas para produzir atributos em baixa dimensão que representem o conjunto de dados independentes do tempo.

Assim como o t-SNE, o UMAP também possui uma natureza estocástica. O estado aleatório inicial pode ser travado por uma variável chamada *random_state*. Todos os parâmetros de entrada da biblioteca utilizada que implementa o método UMAP foram mantidos em seus valores padrões. Optou-se por gerar apenas três componentes para manter a quantidade de componentes geradas com t-SNE. Assim como foi feito para as técnicas PCA e t-SNE, antes de aplicar a técnica UMAP, os dados foram padronizados. A Figura 47 apresenta o comportamento dos três componentes gerados com UMAP a partir do conjunto de dados original.

Figura 47 - Atributos gerados com UMAP.



Fonte: elaborado pelo autor.

4.2.1.4 Ranqueamento de atributos

Uma dificuldade conhecida sobre a escolha de atributos é que, se feita manualmente, pode consumir muito tempo além de a escolha depender do conhecimento e da experiência de quem os estará selecionando. Existem técnicas que permitem uma escolha automatizada dos atributos. Estas técnicas podem escolher os n atributos que mais contribuem para a predição da variável alvo.

Para realizar a tarefa de selecionar os atributos automaticamente, empregou-se a biblioteca *Selective* [49]. Esta biblioteca foi desenvolvida pelo centro de excelência em inteligência artificial da *Fidelity* investimentos e deve ser utilizada com a linguagem de programação Python. No momento da redação deste trabalho, a biblioteca é de livre acesso [49]. A biblioteca *Selective* utiliza diversos métodos de seleção supervisionados e não-supervisionados para problemas de regressão e de classificação, incluindo as correlações de Pearson e Spearman. A Tabela 5 apresenta os métodos e seus respectivos seletores. Cada seletor designa ao atributo uma

pontuação de importância. Ao final do processo, as pontuações de cada atributo são somadas. Assim uma lista é gerada com os atributos listados em ordem decrescente de importância.

Tabela 5 - Métodos utilizados para o ranqueamento de atributos.

Método	Seletor
Variância por atributo	Seleciona os atributos baseado no cálculo da variância. Caso o atributo possua variância maior que um valor.
Correlação	Correlação de Pearson
	Correlação de Spearman
Análise estatística	Teste F para regressão utilizando ANOVA
	Informação mútua
Métodos lineares	Sem regularização
	Com regularização <i>Ridge</i>
	Com regularização <i>Lasso</i>
Métodos baseados em florestas	Arvore de decisão
	<i>Random Forest Regressor</i>
	<i>XGBoost Regressor</i>
	<i>LightGB Regressor</i>
	<i>AdaBoost Regressor</i>
	<i>CatBoost Regressor</i>

Fonte: adaptado de [49].

Os dez atributos mais importantes selecionados por *Selective* para as potências ativa e reativa são apresentados na Tabela 6. Optou-se também por implementar manualmente um ranqueamento baseado nas correlações de Pearson, Spearman e PhiK [34]. Os dez atributos mais importantes selecionados por ranqueamento de correlação para as potências ativa e reativa também são apresentados na Tabela 6.

Tabela 6 - Atributos ranqueados em ordem decrescente por *Selective* e por correlação.

Rank	Selective		Correlação	
	Atributos para potência ativa	Atributos para potência reativa	Atributos para potência ativa	Atributos para potência reativa
1	UMAP_1	CH3_RMS	tsne_1	CH3_RMS
2	CH3_1440	CH2_RMS	CH3_3450	CH2_RMS
3	CH3_435	CH3_1155	CH3_1650	CH3_60
4	CH1_330	CH2_1155	CH3_2520	CH2_60
5	CH1_765	CH1_2205	CH1_330	CH1_60
6	CH1_1965	CH2_1125	PCA_1	CH1_RMS
7	CH3_3555	CH1_60	CH3_1515	CH1_1155
8	CH3_1650	CH2_60	CH3_1440	CH2_1125
9	UMAP_3	CH1_1140	CH3_2880	CH2_75
10	CH1_3375	CH3_60	CH3_3735	CH2_1155

Fonte: elaborado pelo autor.

Dentro do estudo de correlação foi observado, por meio dos dados contidos na Tabela 3, que a informação proveniente do canal 3 possuía de modo geral maior correlação com a potência ativa. A tendência do canal 3 se manteve para o ranqueamento feito ao ponderar os três tipos de correlação. Em contraste, dentro dos atributos selecionados pela biblioteca *Selective*, existe uma significativa parcela de harmônicas do canal 1 que, ao se analisar os 20 atributos mais importantes, nota-se uma distribuição quase homogênea entre o canal 1 (6 contagens) e o canal 3 (7 contagens).

Para a potência reativa, observando o ranqueamento da biblioteca *Selective* e correlação na Tabela 6, observa-se que o valor RMS do campo contém informação relevante sobre a potência reativa, especialmente nos canais 2 e 3, bem como a amplitude da harmônica de 60Hz para os três canais.

4.2.1.5 Seleção de atributos

Para entender melhor como os atributos ranqueados devem se comportar ao serem utilizados em uma tarefa de regressão, uma breve análise foi feita para tentar determinar uma quantidade aceitável de atributos a serem utilizados nos modelos. De maneira automática, várias combinações de atributos foram utilizadas com o modelo de regressão linear (método dos mínimos quadrados) com a finalidade de verificar com alguma métrica de desempenho a quantidade de atributos a serem utilizados durante o treinamento dos modelos e produzir um modelo de referência. O método utilizado para verificar o número de atributos a ser utilizado foi composto por uma validação cruzada *k-fold* com 5 divisões. As métricas escolhidas para a validação cruzada foram o coeficiente de determinação R^2 , o erro médio absoluto (MAE – *Mean absolute error*) e a raiz do erro médio quadrático (RMSE – *Root mean squared error*). As métricas são dadas pelas Equações (10), (11) e (12), respectivamente do coeficiente de determinação R^2 , MAE e RMSE. Os valores de MAE e RMSE foram calculados em Watts e var.

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (10)$$

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \quad (11)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (12)$$

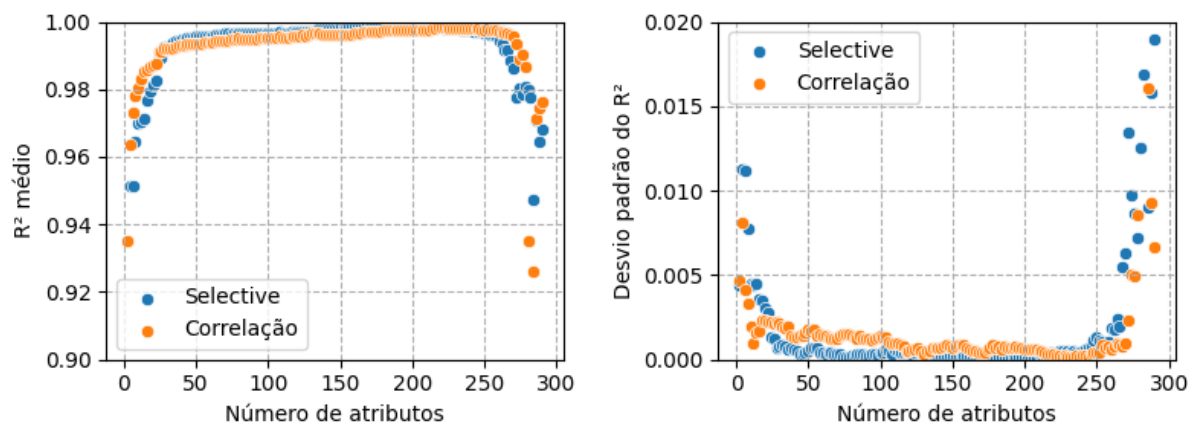
onde:

- y_i é o i ésimo elemento da variável alvo.
- \hat{y}_i é o i ésimo elemento previsto pelo modelo.
- n é o número total de observações.

O coeficiente de determinação R^2 indica a qualidade com que um modelo de regressão estima os dados reais. O erro médio absoluto (MAE) indica a média absoluta com que o modelo de regressão erra ao estimar os dados reais, com a mesma unidade de grandeza dos dados reais. A raiz do erro médio quadrático (RMSE) também indica um valor médio de erro com que o modelo de regressão erra ao estimar os dados reais, porém, o RMSE é sensível à grandes variações. Portanto, se os resíduos da regressão possuírem alta variância o RMSE tenderá a ter valores maiores que o MAE, com a mesma unidade de grandeza dos dados reais.

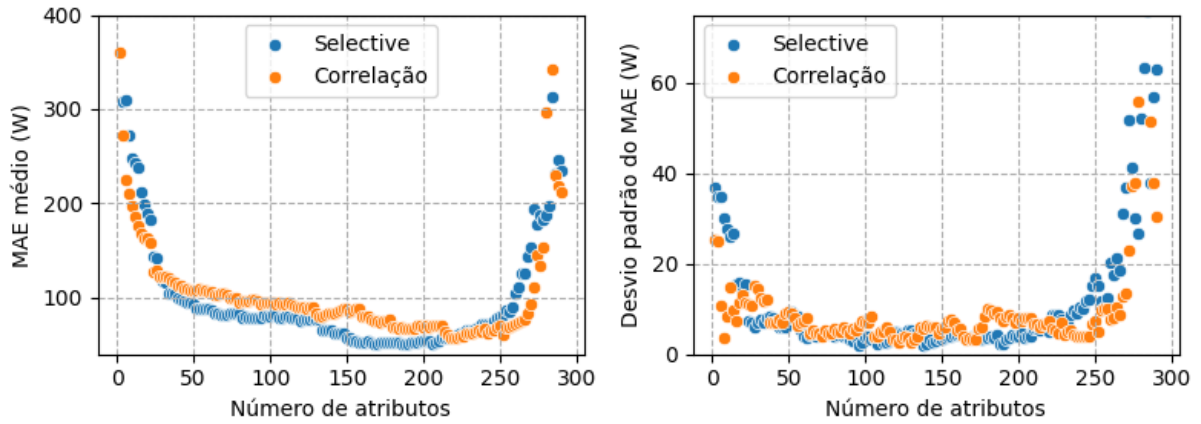
Após cada validação cruzada, um novo atributo é inserido e o processo repetido até o total de 300 atributos. Da Figura 48 até a Figura 53 estão os gráficos do comportamento das métricas geradas durante a validação cruzada dos modelos de regressão linear para atributos ranqueados por *Selective* e correlação.

Figura 48 - Curva do R^2 médio em cada validação cruzada para os atributos ranqueados por *Selective* e correlação para a potência ativa.



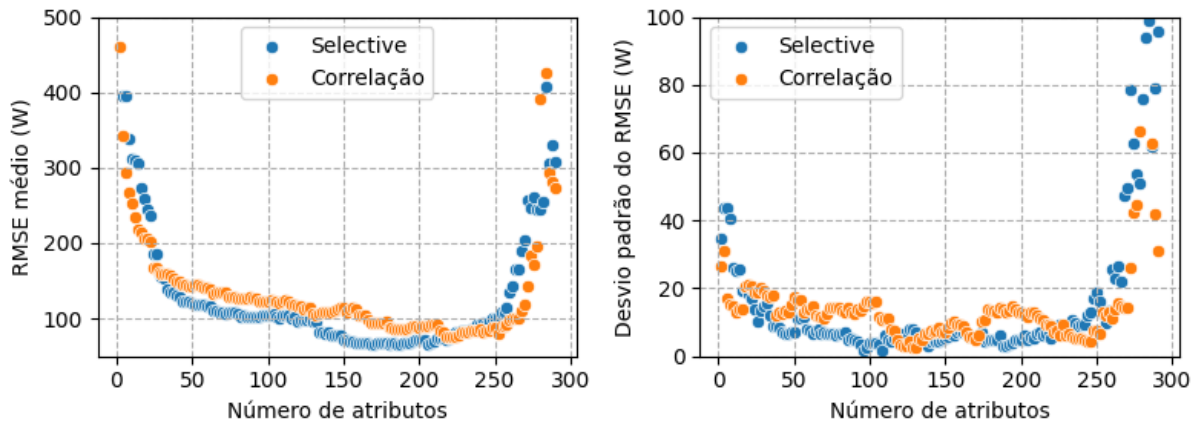
Fonte: elaborado pelo autor.

Figura 49 - Curva do MAE médio em cada validação cruzada para os atributos ranqueados por *Selective* e correlação para a potência ativa.



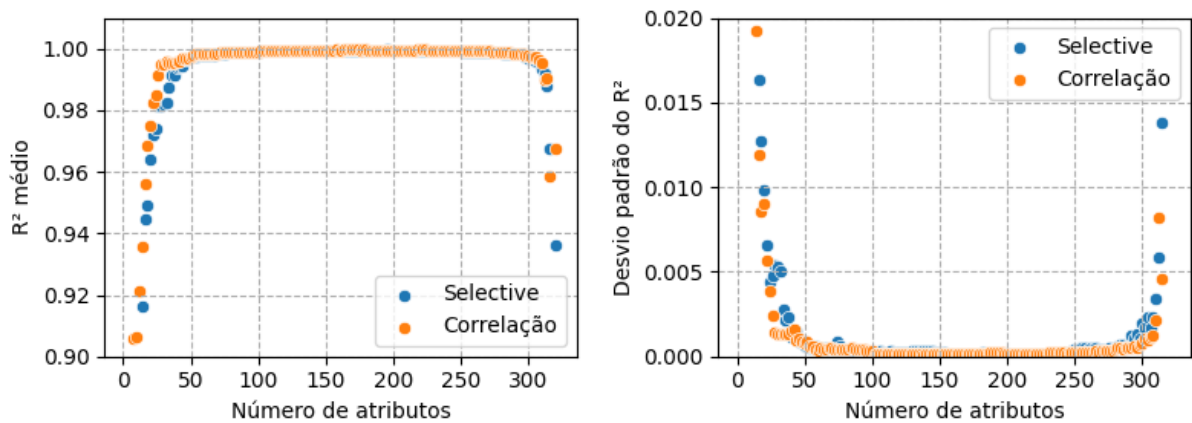
Fonte: elaborado pelo autor.

Figura 50 - Curva do RMSE médio em cada validação cruzada para os atributos ranqueados por *Selective* e correlação para a potência ativa.



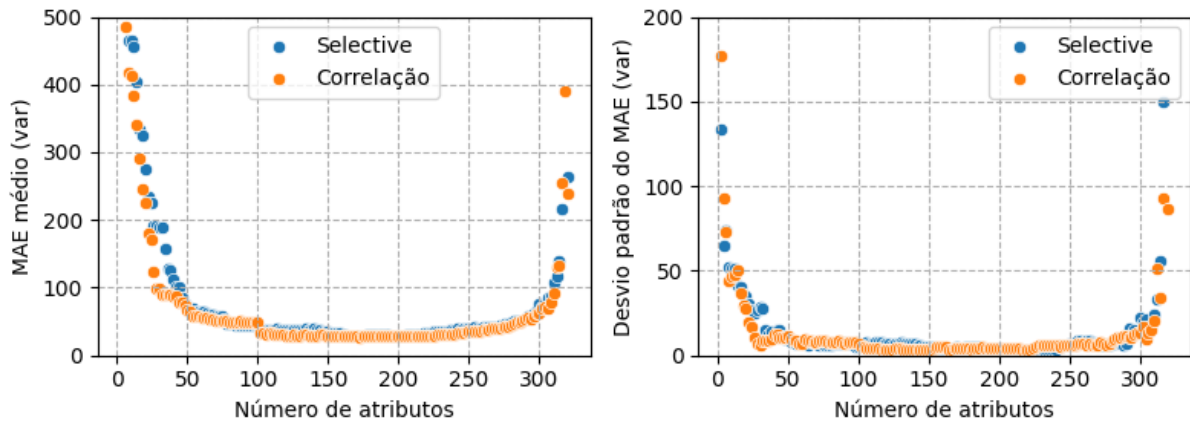
Fonte: elaborado pelo autor.

Figura 51 - Curva do R^2 médio em cada validação cruzada para os atributos ranqueados por *Selective* e correlação para a potência reativa.



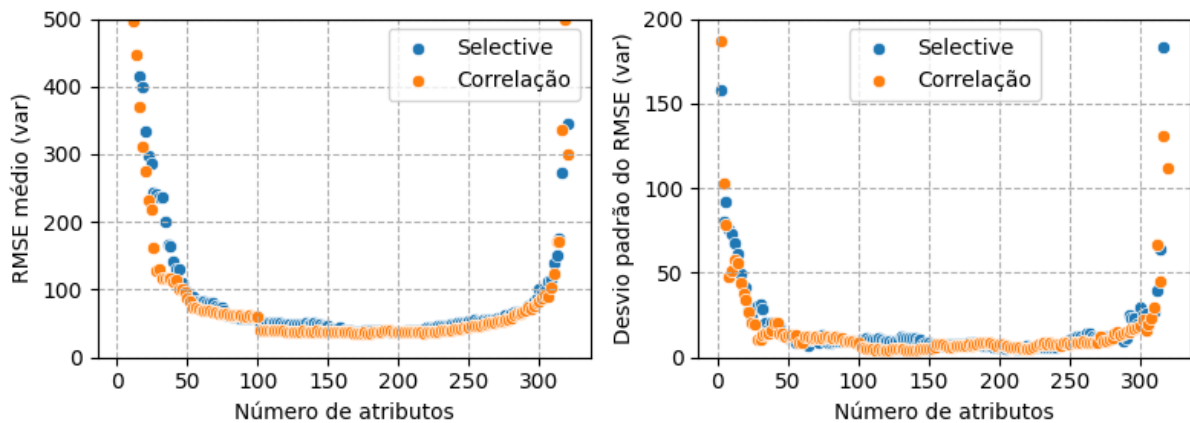
Fonte: elaborado pelo autor.

Figura 52 - Curva do MAE médio em cada validação cruzada para os atributos ranqueados por *Selective* e correlação para a potência reativa.



Fonte: elaborado pelo autor.

Figura 53 - Curva do RMSE médio em cada validação cruzada para os atributos ranqueados por *Selective* e correlação para a potência reativa.



Fonte: elaborado pelo autor.

Conforme esperado, os modelos para estimar a potência reativa apresentam um bom desempenho mesmo com poucos atributos. Com menos de 50 atributos o modelo de mínimos quadrados é capaz de atingir um R^2 próximo da unidade. Observa-se ainda que, mesmo com apenas 25 atributos, o modelo atinge um R^2 superior a 0,98.

Pelos resultados apresentados na Figura 53, observa-se que o modelo precisa de pelo menos 50 atributos para desviar menos de 100 var em média. Isto representa 1% da potência nominal do gerador. Pelos resultados da Figura 50, para a potência ativa, precisa-se de no mínimo 75 atributos para obter um desvio médio aproximado de 100 W (1% da potência nominal do gerador).

Ao continuar adicionando atributos ao modelo, espera-se que instabilidades comecem a surgir em seus coeficientes e o modelo comece a ficar proficiente em

estimar os dados do conjunto de treino e perca a capacidade de generalização, causando um *overfit*. Um sintoma de *overfit* é quando o erro de validação do modelo começa a aumentar conforme a complexidade do modelo aumenta.

As curvas geradas, da Figura 48 até a Figura 53, para análise de R^2 , MAE e RMSE variando de acordo com o número de atributos do modelo servem de referência para a escolha do número de atributos a serem utilizados para a seleção do modelo.

A análise feita para os atributos ranqueados por *Selective* e por correlação demonstram as mesmas características quanto ao desempenho do modelo de regressão linear para os diferentes números de atributos. Não obstante, observa-se que as curvas de desempenho dos atributos ranqueados por *Selective* apresenta melhores métricas para a potência ativa a partir de aproximadamente 25 atributos, conforme apresentado nas Figuras Figura 48, Figura 49 e Figura 50. Para a potência reativa, o ranqueamento por *Selective* e correlação apresentaram métricas de desempenho similares para a mesma quantidade de atributos.

Das curvas de desempenho, observa-se que o melhor desempenho possível para o modelo de regressão por método dos mínimos quadrados para a potência ativa acontece quando o modelo tem aproximadamente 200 atributos de entrada. A partir de 200 atributos, o desempenho do modelo é deteriorada. A mesma afirmação é válida para a potência reativa, com a diferença que aproximadamente 100 atributos bastam para o modelo atingir seu melhor desempenho.

As figuras apresentadas são um tanto peculiares quanto aos seus resultados. Entende-se que é coerente o comportamento das curvas das métricas: de um alto valor para os erros MAE e RMSE, com diminuição até aproximadamente 200 atributos para a potência ativa e apenas 100 atributos para a potência reativa. Após este pico, o modelo apresenta apenas degradação das métricas quanto mais atributos são inseridos. A interpretação para o coeficiente de determinação R^2 segue a mesma lógica.

O apêndice C – visualização de importância de atributos apresenta como os coeficientes do modelo de regressão *baseline* se comportam com diferentes números de atributos de entrada: 5, 15, 30 e 60 atributos. Neste estudo, foram utilizados os atributos ranqueados por *Selective* e atributos ranqueados por correlação.

Pôde ser observado nos resultados apresentados nas figuras do apêndice C como o valor dos coeficientes varia com o aumento do número de atributos que o

modelo linear recebe. A importância do atributo está diretamente relacionada com a magnitude do coeficiente, quanto maior o coeficiente do atributo, maior será seu impacto ao estimar um ponto de operação com o modelo de regressão.

Observou-se que, de maneira geral, a importância dos atributos é mais bem distribuída para a potência ativa. Acredita-se ser resultado da baixa correlação entre os atributos originais, conforme visto no estudo de correlação. Os atributos gerados artificialmente apresentam, de maneira geral, grande relevância quando comparado aos atributos originais. Isso pode ser confirmado observando UMAP_2, UMAP_3 e tsne_1 para os gráficos de importância para a potência ativa na Figura 82 por exemplo.

Durante o desenvolvimento do estudo de importância dos coeficientes foi observado que para a potência reativa, remover alguns dos atributos como CH3_RMS, CH2_60, CH3_60, CH1_1005, CH1_1020 ajuda a diminuir a variação dos atributos mais importantes como CH2_RMS e CH1_RMS, visto que esses atributos são altamente correlacionados entre si.

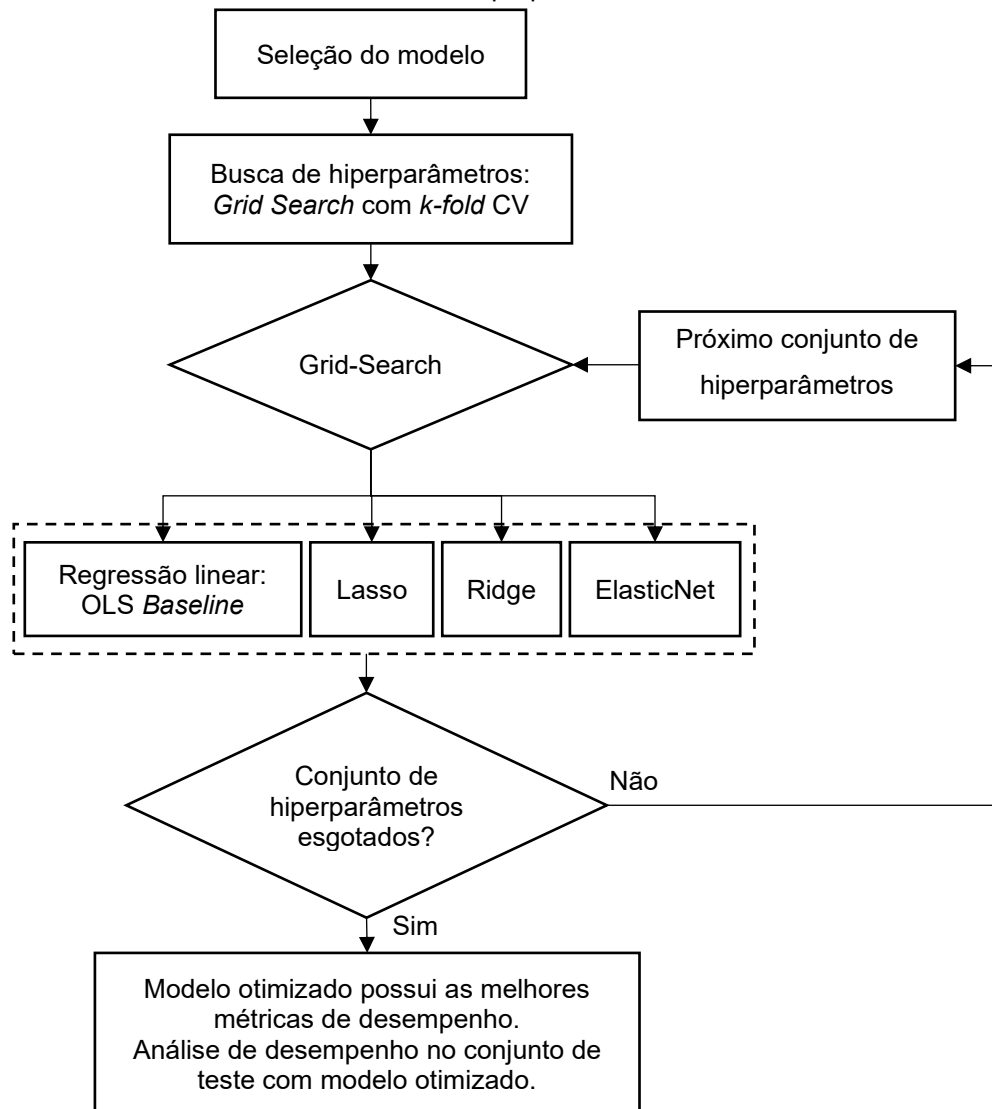
Conforme observado nas figuras Figura 86 e Figura 94, apenas poucos coeficientes possuem uma elevada magnitude. Isso pode ser explicado pelo estudo de correlação, onde foi verificado a alta correlação entre os atributos preditores da potência reativa. Os atributos correlacionados entre si carregam informação redundante e ao minimizar a função custo da regressão linear, apenas um desses atributos é visto como importante para o modelo de regressão.

4.2.2 Seleção do modelo, treinamento, validação e ajuste de hiperparâmetros

A seleção dos modelos de regressão foi feita por meio da avaliação das métricas de desempenho de cada modelo para o conjunto separado para o treino, conforme o fluxo de processamento de dados apresentado na Figura 42.

Antes de serem avaliados no conjunto de teste, os modelos foram configurados e passaram por uma validação cruzada *k-fold* para encontrar a combinação de hiperparâmetros que resultou no melhor desempenho. O método utilizado para ajustar os hiperparâmetros dos modelos foi o *Grid Search*, onde cada rodada de validação cruzada aplicava os hiperparâmetros de uma lista de possíveis candidatos.

Figura 54 - Diagrama em blocos do processo de seleção do modelo, treinamento, validação e ajuste de hiperparâmetros.



Fonte: Adaptado de [46].

Após encontrado o conjunto de hiperparâmetros que resultaram no maior R^2 , os modelos passam por uma análise de desempenho. Para a análise de desempenho, os modelos foram treinados com todos os dados do conjunto de treino (dados utilizados para a validação cruzada *k-fold*) e testado no conjunto separado para o teste. Uma representação em forma de diagrama de blocos que resume a seleção do modelo, treinamento, validação e ajuste de hiperparâmetros é apresentada na Figura 54.

Para cada modelo avaliado com validação cruzada foi gerada uma tabela com indicadores de desempenho. Esta tabela apresenta os resultados com o valor médio da métrica (\bar{x}) atribuído seu desvio padrão (σ_x), conforme a Equação (13).

$$\text{Métrica} = \bar{x} \pm \sigma_x \quad (13)$$

O estudo de correlação mostrou que relação observada entre a potência reativa e o valor RMS do campo é aproximadamente linear, assim como para as amplitudes da harmônica de 60 Hz provenientes dos três canais. Para a potência ativa, apesar de possuir um alto ρ de Pearson (próximos de 0,8), a relação não é claramente linear. Após aplicar PCA e t-SNE ao conjunto de dados, foi possível obter dois atributos, PCA_1 e tsne_1 que apresentam, visualmente, uma relação linear com a potência ativa. O *ranking* de correlação confirma o pressuposto, visto que o atributo tsne_1 ocupa a primeira posição e PCA_1 a sexta.

Por este motivo, foi decidido utilizar modelos de regressão que exploram a relação linear entre variável dependente e independente. Assim, quatro modelos serão abordados: regressão linear simples pelo método dos mínimos quadrados, regressão linear com penalização L1 (aplicação da norma L1 no vetor de coeficientes da regressão), regressão linear com penalização L2 (aplicação da norma L2 no vetor de coeficientes da regressão) e regressão linear com penalização L1 e L2 (aplicação da normas L1 e L2 no vetor de coeficientes da regressão).

4.2.2.1 Modelo de regressão linear Baseline

O modelo de regressão linear utilizado pelo método dos mínimos quadrados (OLS – *Ordinary Least Square*) serve como modelo de referência (também chamado *baseline*). O desempenho deste modelo, variando o número de atributos de entrada, tanto para a potência ativa quanto para a potência reativa, foi apresentado dentro da seção de seleção de atributos nas Figura 48 e Figura 53.

A função hipótese da regressão linear é dada pela equação (14).

$$\hat{y} = f(\mathbf{x}) = w_0 + w_1x_1 + w_2x_2 + \dots + w_mx_m \quad (14)$$

onde:

- \hat{y} é o valor previsto para $f(\mathbf{x})$.
- $w_0, w_1, w_2, \dots, w_m$ são os coeficientes do modelo.
- $x_0, x_1, x_2, \dots, x_m$ são os atributos do modelo.
- m é o número total de atributos de entrada do modelo.

Se $m=1$, apenas um atributo, a equação (14) representa a equação de uma reta, onde w_0 é o coeficiente linear (também chamado de *bias* ou *intercept term*) e w_1 é o coeficiente angular.

Em notação vetorial é possível expressar a equação (14) pela equação (15).

$$\hat{y} = f(\mathbf{x}) = w_0 + [w_1 \quad \cdots \quad w_m] \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} = b + \mathbf{w}^T \mathbf{x} \quad (15)$$

onde:

- b é o coeficiente linear (*bias*).
- $\mathbf{w} = [w_1 \quad \cdots \quad w_m]^T$ é o vetor de coeficientes do modelo.
- $\mathbf{x} = [x_1 \quad \cdots \quad x_m]^T$ é o vetor de atributos.

É matematicamente conveniente zerar o coeficiente linear ($b=0$) e adicionar um atributo constante de valor 1 como primeiro elemento do vetor \mathbf{x} , resultando na equação (16).

$$\hat{y} = f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} \quad (16)$$

onde:

- $\mathbf{w} = [w_0 \quad w_1 \quad \cdots \quad w_m]^T$ é o vetor de coeficientes do modelo.
- $\mathbf{x} = [1 \quad x_1 \quad \cdots \quad x_m]^T$ é o vetor de atributos.

A função hipótese utilizada é parametrizada por \mathbf{w} . Os coeficientes devem ser escolhidos de maneira a minimizar a função custo definida pela equação (17).

$$J(\mathbf{w}) = J(f(\mathbf{x})) = \sum_{i=1}^n L(f(\mathbf{x}^{(i)}), y^{(i)}) = \sum_{i=1}^n L(\mathbf{w}^T \mathbf{x}^{(i)}, y^{(i)}) \quad (17)$$

onde:

- $J(\mathbf{w})$ é a função custo parametrizada por \mathbf{w} .
- n é o número total de amostras em cada atributo.
- $y^{(i)}$ é a i ésima amostra de y . $\mathbf{x}^{(i)}$ é a i ésima amostra de \mathbf{x} .
- $L(\mathbf{w}^T \mathbf{x}^{(i)}, y^{(i)}) = L(\hat{y}^{(i)}, y^{(i)})$ é a função perda a ser utilizada na função custo.

Ao utilizar o erro quadrático como função perda, $L(\hat{y}^{(i)}, y^{(i)}) = (\hat{y} - y)^2$, obtém-se a equação (18). Assim, ao minimizar o erro quadrático dos resíduos, obtém-

se os valores dos coeficientes da reta com o menor erro quadrático. A regressão linear que minimiza o erro quadrático é conhecida como o estimador dos mínimos quadrados, ou OLS – *Ordinary Least Square estimator*.

$$J(\mathbf{w}) = \sum_{i=1}^n (\mathbf{w}^T \mathbf{x}^{(i)} - y^{(i)})^2 \quad (18)$$

Como neste caso a função custo é convexa, ela possui um valor mínimo. Portanto, para encontrar os valores de \mathbf{w} que fazem $J(\mathbf{w})$ ser mínimo, é preciso derivar a função custo e igualar a zero, $\frac{\partial J(\mathbf{w})}{\partial w_j} = 0$. De maneira resumida, o modelo de regressão linear minimiza a soma do quadrado dos resíduos, conforme a equação (19).

$$\min_{\mathbf{w}} J(\mathbf{w}) = \min_{\mathbf{w}} \sum_{i=1}^n (\mathbf{w}^T \mathbf{x}^{(i)} - y^{(i)})^2 \quad (19)$$

A implementação do modelo de regressão linear utilizada é a da biblioteca *scikit-learn* [68]. A função minimizada pela biblioteca *scikit-learn* do Python para o modelo de regressão linear é apresentada pela equação (20).

$$\min_{\mathbf{w}} \|\mathbf{X}\mathbf{w} - \mathbf{y}\|_2^2 \quad (20)$$

onde:

- $\mathbf{w} = [w_0 \ w_1 \ \dots \ w_m]^T$ é o vetor de coeficientes do modelo.
- $\mathbf{y} = [y^{(1)} \ y^{(2)} \ \dots \ y^{(n)}]^T$ é o vetor de rótulos do modelo.
- $\mathbf{X} = \begin{bmatrix} 1 & x_1^{(1)} & x_2^{(1)} & \dots & x_m^{(1)} \\ 1 & x_1^{(2)} & x_2^{(2)} & \dots & x_m^{(2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_1^{(n)} & x_2^{(n)} & \dots & x_m^{(n)} \end{bmatrix}$ é o vetor de atributos.
- n é o número total de amostras em cada atributo.

O treinamento do modelo passou por uma validação cruzada *k-fold* com 5 divisões e diferentes combinações de atributos foram utilizadas. Os atributos utilizados são da lista de atributos ranqueados por *Selective* e por correlação. Por não possuir nenhum hiperparâmetro a ser ajustado, o modelo de regressão por mínimos quadrados não precisou de *Grid Search*.

Utilizou-se do modelo de regressão linear para encontrar o número de atributos a serem utilizados durante a seleção dos demais modelos dentro da etapa de seleção de modelos.

As tabelas Tabela 7 e Tabela 8 apresentam as métricas de desempenho para o modelo *baseline* ao estimar as potências ativa e reativa com diferentes quantidades de atributos.

Tabela 7 - Métricas de performance para o modelo de regressão linear baseline no conjunto de validação para a potência ativa.

Nº Atr.	RMSE [W]	MAE [W]	R^2
5	411 ±29	332 ±27	0,9476 ±0,0091
10	270 ±18	206 ±13	0,9776 ±0,0021
15	219 ±20	173 ±16	0,9851 ±0,0024
30	188 ±20	141 ±8,2	0,9891 ±0,0015
60	111 ±3,2	85 ±3,3	0,9962 ±0,0004
120	89 ±4,6	72 ±3,9	0,9975 ±0,0003
200	83 ± 9,6	64 ±8,4	0,9978 ±0,0005

Fonte: elaborado pelo autor.

Tabela 8 - Métricas de performance para o modelo de regressão linear baseline no conjunto de validação para a potência reativa.

Nº Atr.	RMSE [var]	MAE [var]	R^2
5	137 ±5,9	110 ±5,5	0,9985 ±0,0001
10	108 ±4,2	89 ±2,6	0,9991 ±0,0001
15	72 ±4,5	58 ±3,18	0,9996 ±0,0001
30	68 ±6	54 ±2,4	0,9996 ±0,0001
60	40 ±3	31 ±2,6	0,9999 ±0,0000
120	35 ±5,1	26 ±2,8	0,9999 ±0,0000
200	38 ± 4,1	29 ±6,2	0,9999 ±0,0000

Fonte: elaborado pelo autor.

4.2.2.2 Modelo de regressão com penalização L1 (Lasso regression)

O modelo de regressão linear com penalização L1 também é chamado de *Lasso regression - Least Absolute Shrinkage and Selection Operator*. O *Lasso* é um modelo linear que estima coeficientes esparsos. Ele é útil em alguns contextos devido à sua tendência de preferir soluções com menos coeficientes diferentes de zero, reduzindo efetivamente o número de atributos dos quais uma determinada solução depende. Dentro da função de minimização existe um termo adicionado, $\alpha\|\mathbf{w}\|_1$, que serve como uma penalidade aos coeficientes.

A função minimizada pela biblioteca *scikit-learn* para o modelo de regressão *Lasso* é dada pela equação (21).

$$\min_{\mathbf{w}} \frac{1}{2n} \|\mathbf{X}\mathbf{w} - \mathbf{y}\|_2^2 + \alpha \|\mathbf{w}\|_1 \quad (21)$$

onde:

- $\mathbf{w} = [w_0 \ w_1 \ \dots \ w_m]^T$ é o vetor de coeficientes do modelo.
- $\mathbf{y} = [y^{(1)} \ y^{(2)} \ \dots \ y^{(n)}]^T$ é o vetor de rótulos do modelo.
- $\mathbf{X} = \begin{bmatrix} 1 & x_1^{(1)} & x_2^{(1)} & \dots & x_m^{(1)} \\ 1 & x_1^{(2)} & x_2^{(2)} & \dots & x_m^{(2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_1^{(n)} & x_2^{(n)} & \dots & x_m^{(n)} \end{bmatrix}$ é o vetor de atributos.
- n é o número total de amostras em cada atributo.
- α é o hiperparâmetro *alpha* utilizado para a regularização L1.

As tabelas Tabela 9 e Tabela 10 apresentam as métricas de desempenho para o modelo *Lasso* ao estimar as potências ativa e reativa. Essas tabelas apresentam as métricas para cada quantidade de atributos e hiperparâmetro *alpha*.

Tabela 9 - Métricas de performance para o modelo *Lasso* no conjunto de validação para a potência ativa.

Nº Atr.	RMSE [W]	MAE [W]	R^2	Hiperparâmetros	
5	411 ±29	332 ±27	0,9476 ±0,0091	<i>alpha</i>	1e-5
10	269 ±17	206 ±14	0,9777 ±0,0021	<i>alpha</i>	5
15	219 ±20	171 ±17	0,9851 ±0,0023	<i>alpha</i>	5
30	188 ±20	145 ±14	0,9891 ±0,0015	<i>alpha</i>	0,1
60	107 ±6,2	83 ±5,5	0,9965 ±0,0005	<i>alpha</i>	1
120	94 ±6,4	74 ±3,8	0,9973 ±0,0003	<i>alpha</i>	1
200	80 ± 6,9	62 ±3,2	0,9980 ±0,0003	<i>alpha</i>	1

Fonte: elaborado pelo autor.

Tabela 10 - Métricas de performance para o modelo *Lasso* no conjunto de validação para a potência reativa.

Nº Atr.	RMSE [var]	MAE [var]	R^2	Hiperparâmetros	
5	133 ±5,2	107 ±5,1	0,9986 ±0,0002	<i>alpha</i>	3
10	119 ±8,4	97 ±5,5	0,9989 ±0,0001	<i>alpha</i>	1e-5
15	79 ±3,3	64 ±2,6	0,9995 ±0,0000	<i>alpha</i>	1
30	71 ±4,2	57 ±2,2	0,9996 ±0,0001	<i>alpha</i>	1
60	62 ±5,1	48 ±3,8	0,9997 ±0,0001	<i>alpha</i>	3
120	58 ±3,6	46 ±3,2	0,9997 ±0,0001	<i>alpha</i>	3

200	63 ± 5	49 ±4,2	0,9997 ±0,0001	<i>alpha</i>	3
-----	--------	---------	----------------	--------------	---

Fonte: elaborado pelo autor.

4.2.2.3 Modelo de regressão com penalização L2 (*Ridge regression*)

O modelo de regressão linear com penalização L2 também é chamado de *Ridge regression*. A regressão *Ridge*, assim como a regressão *Lasso*, impõe uma penalidade a magnitude dos coeficientes. Dentro da função de minimização existe um termo adicionado, $\alpha\|\mathbf{w}\|_2^2$, que serve como uma penalidade quadrática aos coeficientes.

A função minimizada pela biblioteca *scikit-learn* para o modelo de regressão *Ridge* é dada pela equação (22).

$$\min_{\mathbf{w}} \|X\mathbf{w} - \mathbf{y}\|_2^2 + \alpha\|\mathbf{w}\|_2^2 \quad (22)$$

onde:

- $\mathbf{w} = [w_0 \ w_1 \ \dots \ w_m]^T$ é o vetor de coeficientes do modelo.
- $\mathbf{y} = [y^{(1)} \ y^{(2)} \ \dots \ y^{(n)}]^T$ é o vetor de rótulos do modelo.
- $X = \begin{bmatrix} 1 & x_1^{(1)} & x_2^{(1)} & \dots & x_m^{(1)} \\ 1 & x_1^{(2)} & x_2^{(2)} & \dots & x_m^{(2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_1^{(n)} & x_2^{(n)} & \dots & x_m^{(n)} \end{bmatrix}$ é o vetor de atributos.
- α é o hiperparâmetro *alpha* utilizado para a regularização L2.

As tabelas Tabela 11 e Tabela 12 apresentam as métricas de desempenho para o modelo *Ridge* ao estimar as potências ativa e reativa. Essas tabelas apresentam as métricas para cada quantidade de atributos e hiperparâmetro *alpha*.

Tabela 11 - Métricas de performance para o modelo *Ridge* no conjunto de validação para a potência ativa.

Nº Atr.	RMSE [W]	MAE [W]	R^2	Hiperparâmetros	
5	411 ±28	332 ±27	0,9476 ±0,0090	<i>alpha</i>	1
10	270 ±17	206 ±12	0,9776 ±0,0020	<i>alpha</i>	1
15	219 ±20	173 ±18	0,9849 ±0,0023	<i>alpha</i>	5
30	188 ±20	144 ±15	0,9892 ±0,0014	<i>alpha</i>	3
60	108 ±3,9	85 ±4,1	0,9964 ±0,0005	<i>alpha</i>	3
120	90 ±4,6	73 ±4,2	0,9975 ±0,0003	<i>alpha</i>	1
200	67 ±3,5	52 ±3,1	0,9986 ±0,0001	<i>alpha</i>	3

Fonte: elaborado pelo autor.

Tabela 12 - Métricas de performance para o modelo *Ridge* no conjunto de validação para a potência reativa.

Nº Atr.	RMSE [var]	MAE [var]	R^2	Hiperparâmetros	
5	133 ±5,6	107 ±5,4	0,9986 ±0,0001	<i>alpha</i>	1e-2
10	108 ±8,4	88 ±2,1	0,9990 ±0,0001	<i>alpha</i>	1e-5
15	72 ±5,1	58 ±3,3	0,9996 ±0,0000	<i>alpha</i>	1e-5
30	67 ±4,8	53 ±2,1	0,9996 ±0,0001	<i>alpha</i>	1e-2
60	39 ±1,4	30 ±1,1	0,9999 ±0,0000	<i>alpha</i>	1e-5
120	33 ±2,1	26 ±1,6	0,9999 ±0,0000	<i>alpha</i>	1e-5
200	39 ± 5,8	30 ±3,5	0,9999 ±0,0000	<i>alpha</i>	1e-3

Fonte: elaborado pelo autor.

4.2.2.4 Modelo de regressão com penalização L1 e L2 (*Elastic net*)

O modelo de regressão linear com penalização L1 e L2 também é chamado de *Elastic net*. Essa combinação permite o aprendizado de um modelo esparso em que poucos pesos são diferentes de zero, como o *Lasso*, e ainda mantém as propriedades de regularização do *Ridge*. Dentro da função de minimização existem dois termos adicionados, $\alpha\rho\|\mathbf{w}\|_1$ e $0,5 * \alpha(1 - \rho)\|\mathbf{w}\|_2^2$, que combinam as penalidades L1 e L2 para captar as características de *Lasso* e *Ridge* em mesmo modelo.

A função minimizada pela biblioteca *scikit-learn* para o modelo de regressão *Elastic net* é dada pela equação (23).

$$\min_{\mathbf{w}} \frac{1}{2n} \|X\mathbf{w} - \mathbf{y}\|_2^2 + \alpha\rho\|\mathbf{w}\|_1 + \frac{\alpha(1 - \rho)}{2} \|\mathbf{w}\|_2^2 \quad (23)$$

onde:

- $\mathbf{w} = [w_0 \ w_1 \ \dots \ w_m]^T$ é o vetor de coeficientes do modelo.
- $\mathbf{y} = [y^{(1)} \ y^{(2)} \ \dots \ y^{(n)}]^T$ é o vetor de rótulos do modelo.
- $X = \begin{bmatrix} 1 & x_1^{(1)} & x_2^{(1)} & \dots & x_m^{(1)} \\ 1 & x_1^{(2)} & x_2^{(2)} & \dots & x_m^{(2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_1^{(n)} & x_2^{(n)} & \dots & x_m^{(n)} \end{bmatrix}$ é o vetor de atributos.
- n é o número total de amostras em cada atributo.
- α é o hiperparâmetro *alpha* utilizado para a regularização.
- ρ é o hiperparâmetro *l1_ratio* utilizado para a definir o peso dado a regularização L1 e L2. $\rho = 1$ significa que a regularização será feita apenas com L1, $\rho = 0$ significa que a regularização será feita apenas com L2.

As tabelas Tabela 13 e Tabela 14 apresentam as métricas de desempenho para o modelo *Elastic net* ao estimar as potências ativa e reativa. Essas tabelas apresentam as métricas para cada quantidade de atributos e hiperparâmetros *alpha* e *l1_ratio*.

Tabela 13 - Métricas de performance para o modelo *Elastic net* no conjunto de validação para a potência ativa.

Nº Atr.	RMSE [W]	MAE [W]	R^2	Hiperparâmetros	
5	410 ±19	332 ±20	0,9482 ±0,0054	<i>alpha</i>	0,01
				<i>l1_ratio</i>	0,6
10	273±10	209 ±8	0,9771 ±0,0023	<i>alpha</i>	0,01
				<i>l1_ratio</i>	0,6
15	222 ±20	173 ±23	0,9848 ±0,0028	<i>alpha</i>	0,1
				<i>l1_ratio</i>	0,7
30	188 ±14	145 ±12	0,9891 ±0,0013	<i>alpha</i>	0,01
				<i>l1_ratio</i>	0,3
60	112 ±6	86 ±3,7	0,9964 ±0,0004	<i>alpha</i>	0,01
				<i>l1_ratio</i>	0,3
120	97 ±9	77 ±7,4	0,9971 ±0,0005	<i>alpha</i>	0,1
				<i>l1_ratio</i>	0,7
200	75 ±10	58 ±6,3	0,9982 ±0,0005	<i>alpha</i>	0,1
				<i>l1_ratio</i>	0,5

Fonte: elaborado pelo autor.

Tabela 14 - Métricas de performance para o modelo *Elastic net* no conjunto de validação para a potência reativa.

Nº Atr.	RMSE [var]	MAE [var]	R^2	Hiperparâmetros	
5	131 ±8,2	106 ±6,7	0,9986 ±0,002	<i>alpha</i>	0,001
				<i>l1_ratio</i>	0,7
10	116±8,6	94 ±6,9	0,9989 ±0,0002	<i>alpha</i>	0,001
				<i>l1_ratio</i>	0,5
15	74 ±5,6	61 ±9996	0,9848 ±0,0001	<i>alpha</i>	0,001
				<i>l1_ratio</i>	0,3
30	73 ±4	56 ±1,9	0,9996 ±0,0001	<i>alpha</i>	0,001
				<i>l1_ratio</i>	0,7
60	47 ±2,1	36 ±0,8	0,9998 ±0,0000	<i>alpha</i>	0,001
				<i>l1_ratio</i>	0,3
120	46 ±0,9	36 ±0,8	0,9998 ±0,0000	<i>alpha</i>	0,001

				<i>l1_ratio</i>	0,4
200	59 ±3,2	46 ±2,8	0,9997 ±0,0000	<i>alpha</i>	0,001
				<i>l1_ratio</i>	0,4

Fonte: elaborado pelo autor.

4.2.3 Análise de desempenho no conjunto de teste

As métricas de desempenho dos modelos lineares apresentam concordância entre si durante a validação, sem grandes diferenças. Todos os modelos analisados são capazes de atingir um coeficiente de determinação, R^2 , maior que 0,99 para a potência reativa com ao menos 5 atributos. Para a potência ativa, os mesmos modelos atingem o coeficiente de determinação próximos de 0,94 com 5 atributos. São necessários ao menos 60 atributos para os modelos estimarem a potência ativa com R^2 maior que 0,99.

As melhores combinações de cada modelo (número de atributos e hiperparâmetros otimizados) foram aplicados ao conjunto separado inicialmente para testes em uma avaliação final.

As tabelas 15 até tabela 18 apresentam as métricas de desempenho para cada um dos modelos avaliados (*OLS*, *Lasso*, *Ridge* e *Elastic Net*) no conjunto de teste. Observa-se que cada tabela apresenta o modelo com seus hiperparâmetros e número de atributos ótimos obtidos durante a validação cruzada *k-fold*. Nessas tabelas, “Alvo” indica qual a potência estimada: “P” para a potência ativa e “Q” para potência reativa.

As imagens apresentadas da Figura 55 até a Figura 62 apresentam os gráficos da aderência dos pontos de potência estimados vs. os pontos de potência medidos, bem como o comportamento dos resíduos em função da potência estimada.

A interpretação dos resultados dos modelos apresentados entre as figuras Figura 55 e Figura 62 mostra nos gráficos da esquerda a capacidade dos modelos de estimarem o ponto de operação. Um modelo que consiga capturar toda a informação do conjunto de dados seria apresentado com todos os pontos azuis sobre a reta tracejada. Neste caso, o modelo ideal apresenta uma relação 1 para 1 dos dados estimados para os dados medidos.

Nos gráficos da direita são apresentados os resíduos do modelo de regressão ao estimar os pontos de operação do conjunto de teste. Um modelo que consiga capturar toda a informação do conjunto de dados apresentará resíduos sem relação

com as potências ativa e reativa. O esperado é que os resíduos apresentem uma distribuição normal em torno de zero.

As tabelas de desempenho dos modelos apresentam desempenho com R^2 superior a 0,99 para todos os modelos, independentemente da potência avaliada.

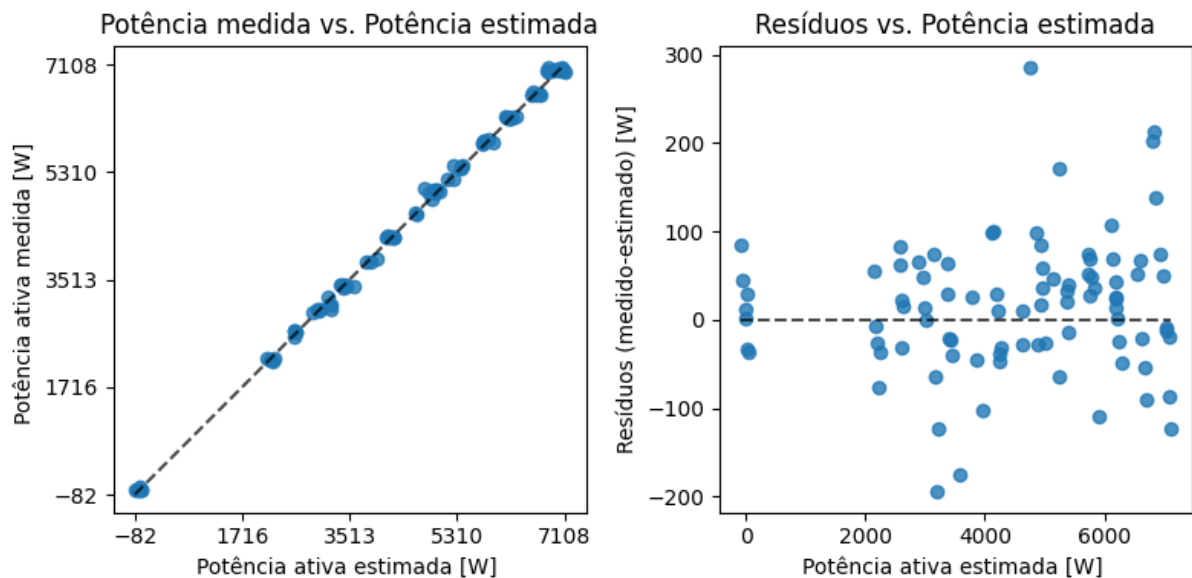
O melhor desempenho de RMSE foi obtido com o modelo *baseline* para a potência ativa (87,7 W) e com os modelos *baseline* e Ridge para a potência reativa (47,1 var). O melhor desempenho de MAE foi obtido com o modelo *baseline* para a potência ativa (62,5 W) e com os modelos *baseline* e Ridge para a potência reativa (32,3 var).

Tabela 15 - Métricas de performance para o modelo de regressão linear *baseline* no conjunto de teste.

Alvo	Nº Atr.	RMSE	MAE	R^2
P	200	87,7 W	62,5 W	0,9980
Q	120	47,1 var	32,3 var	0,9998

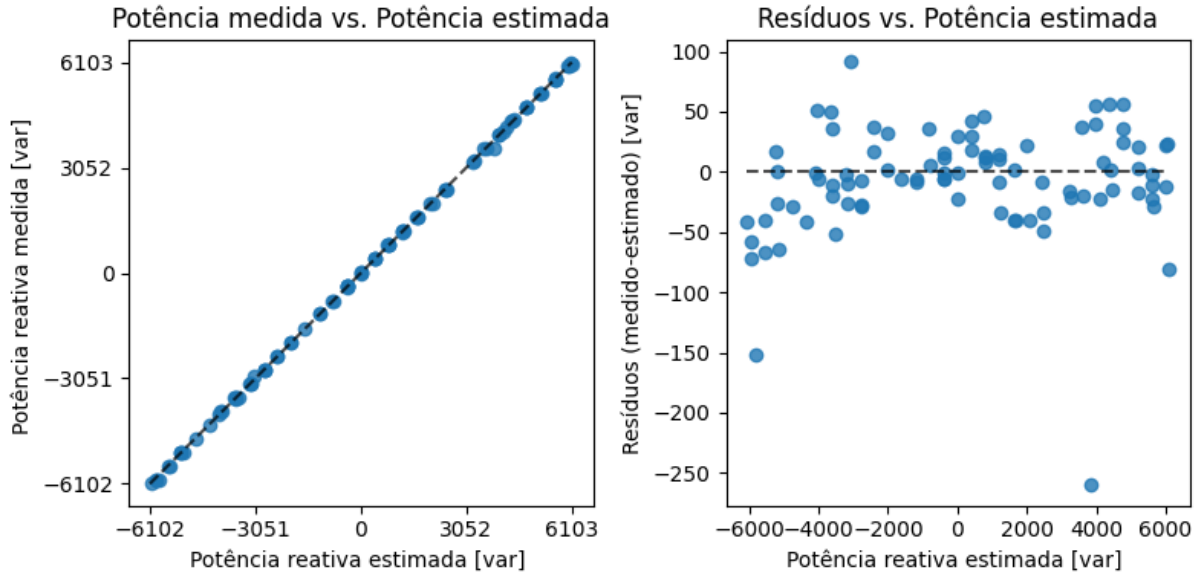
Fonte: elaborado pelo autor.

Figura 55 - Resultado para o modelo de regressão linear *baseline* no conjunto de teste para a potência ativa.



Fonte: elaborado pelo autor.

Figura 56 - Resultado para o modelo de regressão linear *baseline* no conjunto de teste para a potência reativa.



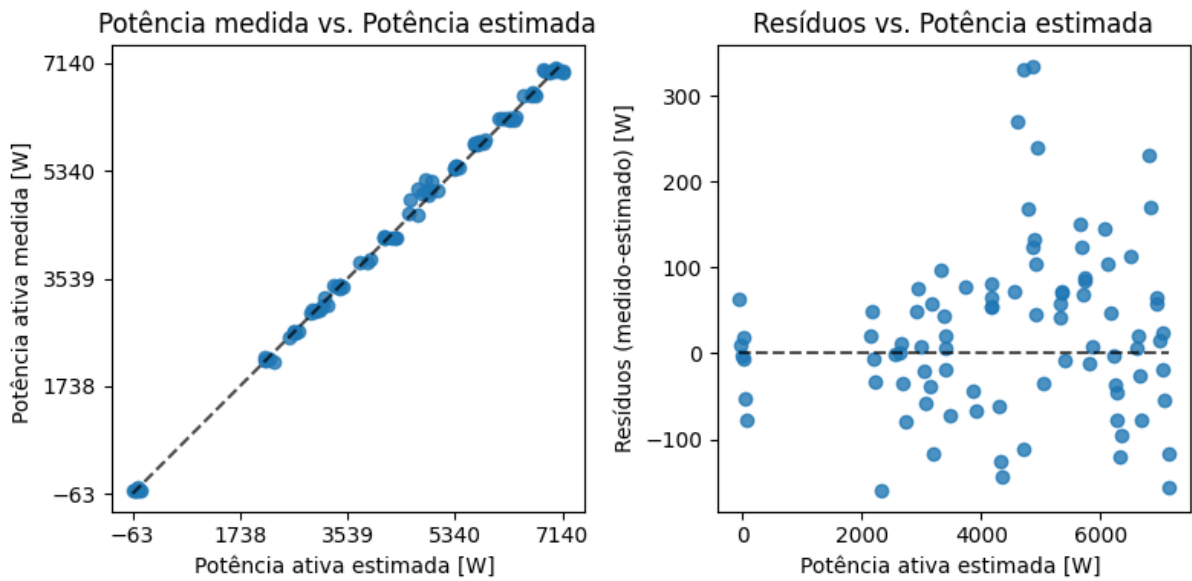
Fonte: elaborado pelo autor.

Tabela 16 - Métricas de performance para o modelo *Lasso* no conjunto de teste.

Alvo	Nº Atr.	RMSE	MAE	R^2	Hiperparâmetros	
P	200	101,6 W	75,0 W	0,9973	<i>alpha</i>	1
Q	120	78,6 var	55,7 var	0,9995	<i>alpha</i>	3

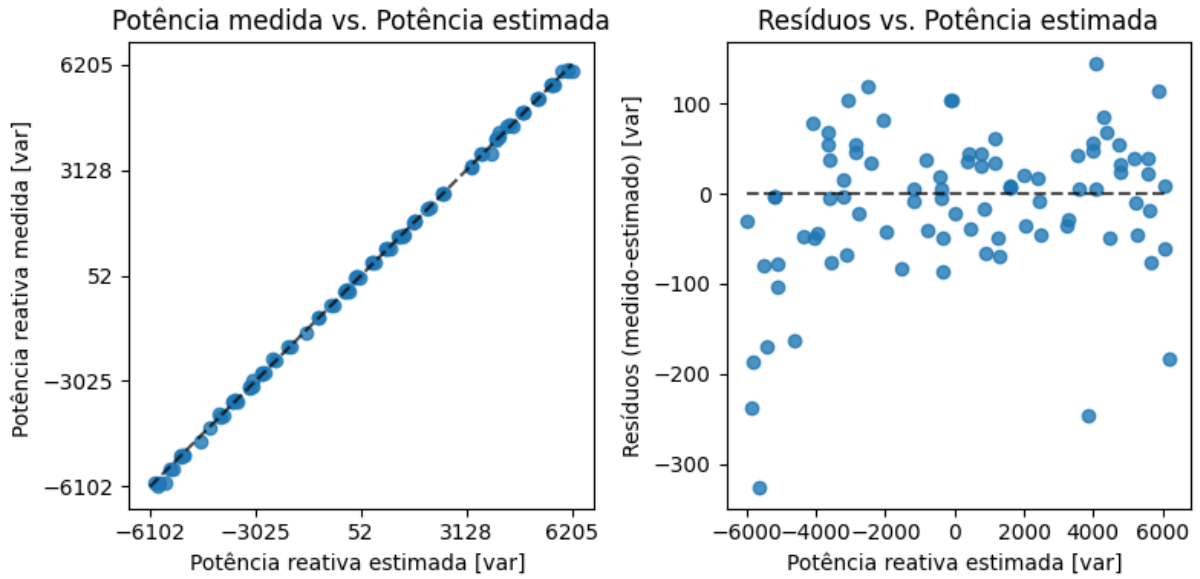
Fonte: elaborado pelo autor.

Figura 57 - Resultado para o modelo *Lasso* no conjunto de teste para a potência ativa.



Fonte: elaborado pelo autor.

Figura 58 - Resultado para o modelo *Lasso* no conjunto de teste para a potência reativa.



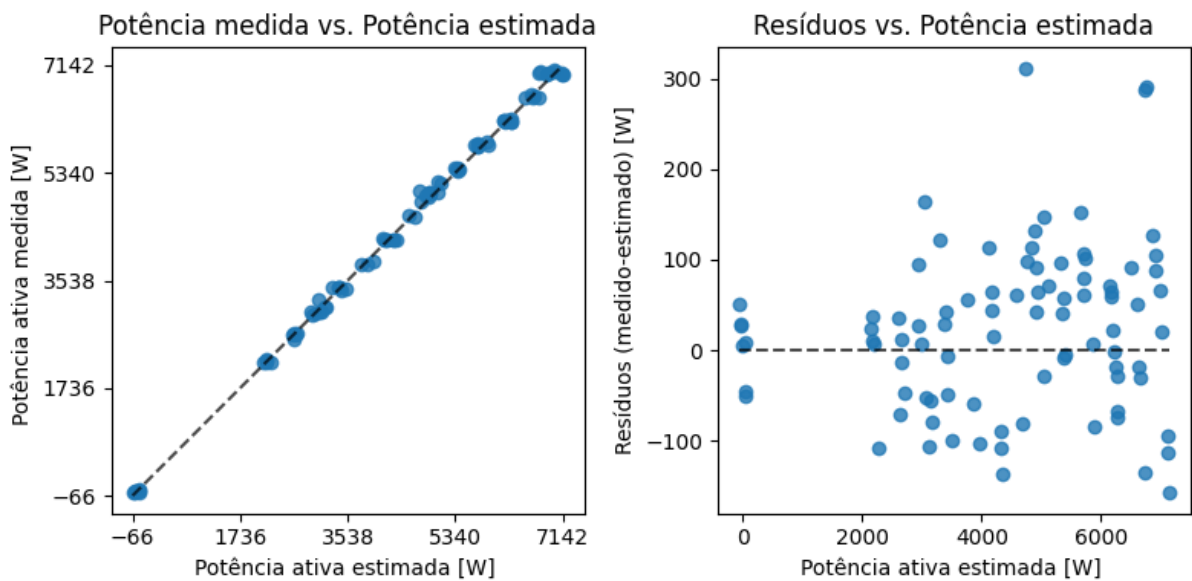
Fonte: elaborado pelo autor.

Tabela 17 - Métricas de performance para o modelo *Ridge* no conjunto de teste.

Alvo	Nº Atr.	RMSE	MAE	R^2	Hiperparâmetros	
P	200	91,6 W	65,9 W	0,9978	<i>alpha</i>	3
Q	120	47,1 var	32,3 var	0,9998	<i>alpha</i>	1e-5

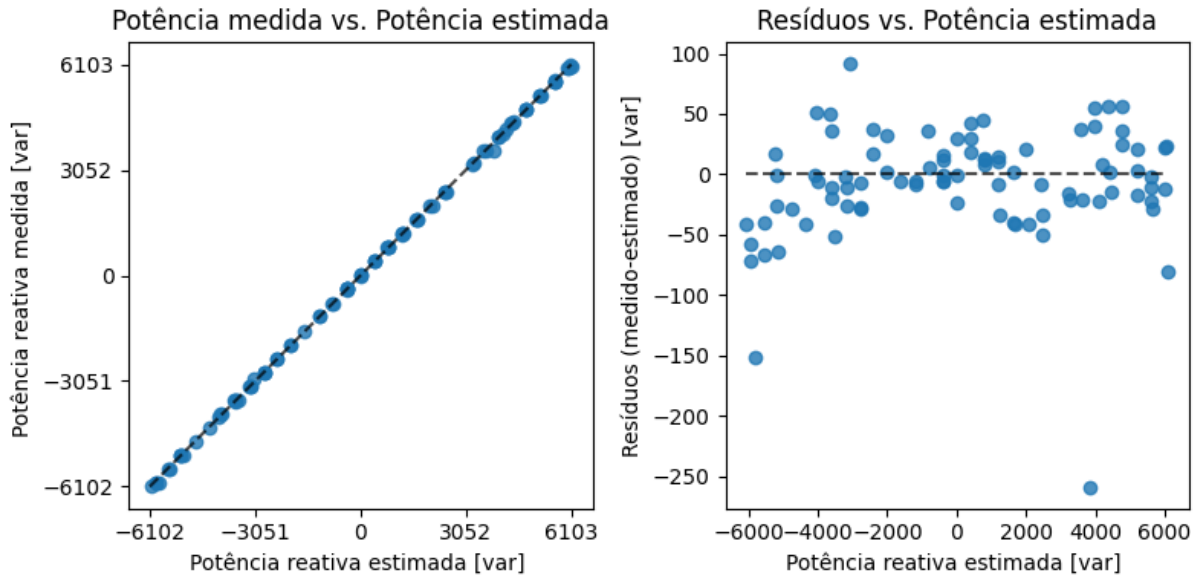
Fonte: elaborado pelo autor.

Figura 59 - Resultado para o modelo *Ridge* no conjunto de teste para a potência ativa.



Fonte: elaborado pelo autor.

Figura 60 - Resultado para o modelo *Ridge* no conjunto de teste para a potência reativa.



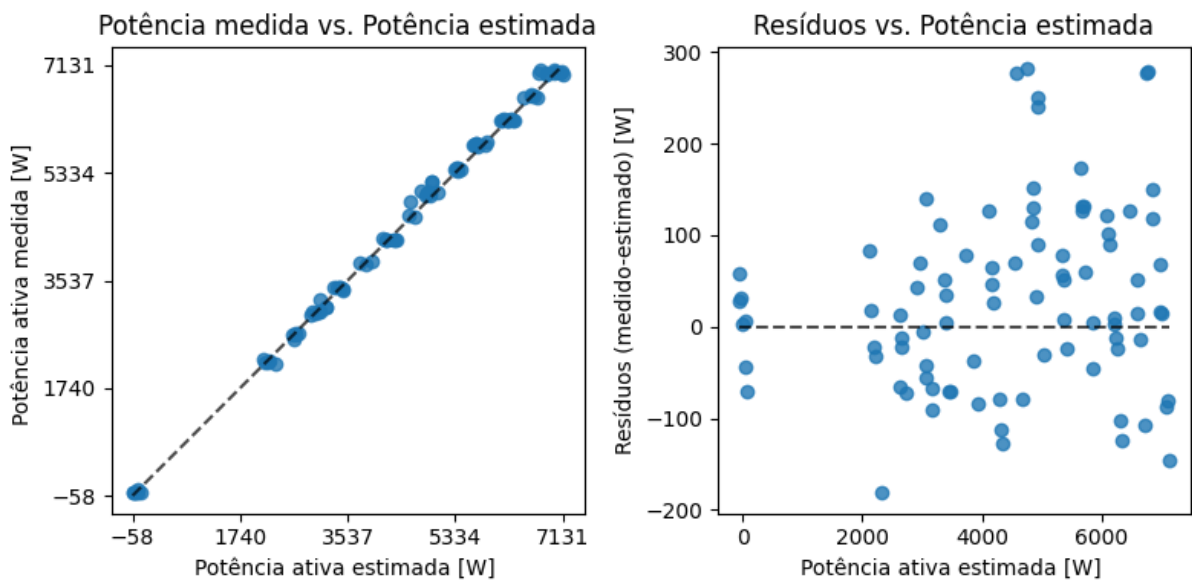
Fonte: elaborado pelo autor.

Tabela 18 - Métricas de performance para o modelo *Elastic net* no conjunto de teste.

Alvo	Nº Atr.	RMSE	MAE	R^2	Hiperparâmetros	
					α	$l1_ratio$
P	200	104,2 W	76,5 W	0,9972	α	0,1
					$l1_ratio$	0,5
Q	120	59,5 var	41 var	0,9997	α	0,001
					$l1_ratio$	0,4

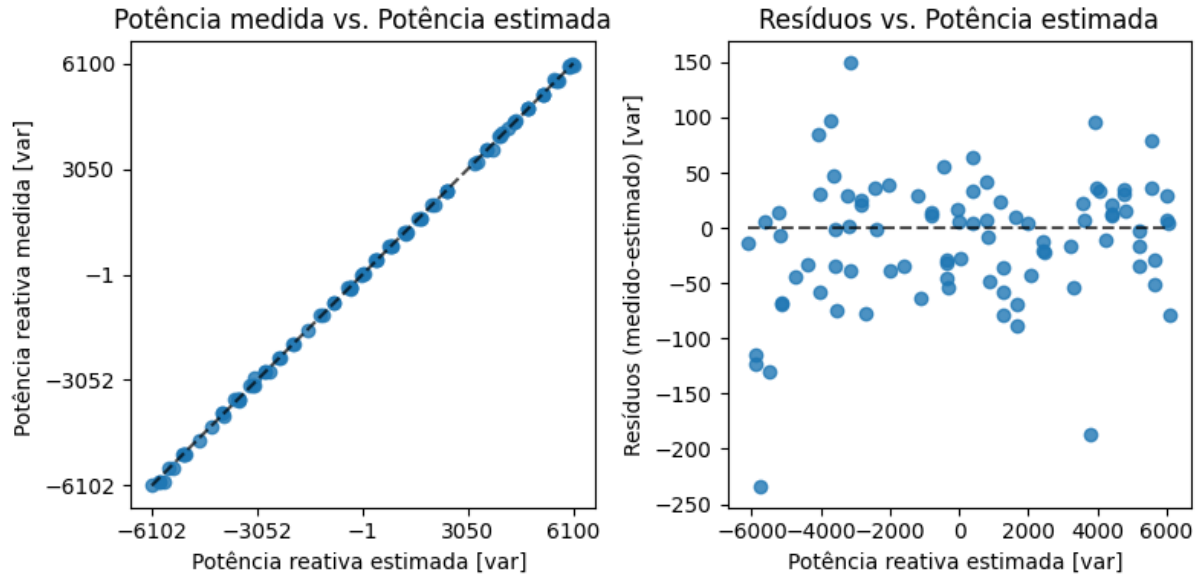
Fonte: elaborado pelo autor.

Figura 61 - Resultado para o modelo *Elastic net* no conjunto de teste para a potência ativa.



Fonte: elaborado pelo autor.

Figura 62 - Resultado para o modelo *Elastic net* no conjunto de teste para a potência reativa.



Fonte: elaborado pelo autor.

4.3 BREVE ANÁLISE E CONSUSÕES DOS RESULTADOS ATINGIDOS

O trabalho desenvolvido neste capítulo apresentou como as informações do conjunto de dados dependentes do tempo foram extraídas e inseridas no conjunto de dados independentes do tempo. A primeira parte do capítulo descreveu como o valor RMS e as amplitudes das harmônicas foram obtidas. Também foi apresentada uma análise de correlação dos dados a fim de encontrar relações entre o ponto de operação e as harmônicas do campo.

Uma conclusão importante da primeira parte é que, para os atributos mais correlacionados com a potência reativa, existe uma clara relação linear entre os dados de entrada do modelo e os dados de saída. Mostrou-se que um modelo de regressão linear é suficiente para modelar o conjunto de dados independentes do tempo. Isto foi comprovado com os resultados dos modelos lineares aplicados. Apesar de a potência ativa não apresentar uma relação nitidamente linear que possa ser representada por uma equação de primeiro grau, foi possível obter bons resultados (alto R^2 e baixos erros RMSE e MAE) com os modelos selecionados. A potência reativa apresenta uma relação claramente linear com seus atributos. No entanto, é observado que os atributos também estão fortemente relacionados entre si. Como não é possível diferenciar as contribuições individuais de cada atributo para a estimativa da potência

reativa, utilizaram-se modelos com termos de regularização na função custo, tais como *Lasso*, *Ridge* e *Elastic net*, de maneira a reduzir o efeito da multicolinearidade.

Ao gerar novos atributos a partir do conjunto de dados com as técnicas de PCA, t-SNE e UMAP foi possível obter uma representação em baixa dimensão que explica a variação nos dados. Neste sentido, o atributo PCA_1 apresentou alta correlação de Pearson ρ com a potência ativa, o que indica que este atributo recuperou uma relação linear entre o conjunto de dados e a potência ativa. Da mesma forma, o atributo PCA_2 recuperou uma relação linear entre o conjunto de dados e a potência reativa. Ambos os atributos gerados apresentam alta correlação de Pearson com um dos tipos de potência e praticamente nenhuma relação com o outro tipo, conforme observado no conjunto de dados pelo estudo de correlação.

Para a geração de atributos por t-SNE é possível observar a mesma tendência que ocorreu com PCA. Ao observar os atributos tsne_1 e tsne_2 nos resultados apresentados na Figura 46 é possível chegar a mesma conclusão que observando os atributos PCA_1 e PCA_2 nos resultados apresentados na Figura 45.

Os coeficientes gerados por UMAP apresentaram significativa divergência de PCA e t-SNE. De todos os atributos gerados, nenhum apresenta relação aparente com a potência reativa. Porém, apresentam uma relação não linear com a potência ativa. Isso fica evidente ao se observar o comportamento dos dados ranqueados por *Selective*. Os métodos não lineares utilizados para o ranqueamento, dentro da biblioteca *Selective*, observaram uma forte relação entre UMAP_1 e UMAP_2 com a potência ativa.

Durante a seleção de atributos, um experimento foi realizado utilizando o modelo de regressão linear pelo método dos mínimos quadrados para verificar as métricas de desempenho MAE, RMSE e R^2 com diferentes quantidades de atributos. Foi observado que para estimar a potência ativa são necessários mais atributos do que para estimar a potência reativa.

Extrair as harmônicas do campo externo significa segmentar a informação que ele carrega em valores discretos. Sendo assim, ao adicionar mais harmônicas, simplesmente se adiciona mais “peças ao quebra cabeça”. O detalhe é que nem todas as peças servem para completar o quebra-cabeça. O desafio é filtrar no conjunto de harmônicas quais delas são necessárias para estimar a potência desejada. Neste sentido o ranqueamento pelos métodos da correlação e com a biblioteca *Selective* se

mostraram eficazes em selecionar quais são as harmônicas relevantes para cada potência.

A segunda parte do capítulo, modelagem dos dados de campo, apresentou a metodologia aplicada para estimar a potência ativa e reativa de um GSPS a partir do conjunto de dados independentes do tempo. Nesta segunda parte estão descritos todos os passos e decisões tomadas até chegar na avaliação de desempenho final do modelo.

Para a validação e o teste dos modelos, foi observado que ao se utilizar os atributos ranqueados por *Selective*, os modelos obtiveram melhor desempenho quando comparado ao desempenho utilizando os atributos ranqueados por correlação. Por este motivo, dentro da seleção de modelo e posteriormente na análise de desempenho, apenas os atributos ranqueados por *Selective* foram utilizados.

Todo o fluxo do processamento de dados foi automatizado com biblioteca *scikit-learn*. Como resultado obtido da modelagem do ponto de operação do GSPS com os dados de campo externo se conclui que é possível estimar o ponto de operação com modelos de natureza linear.

Foi observado nos resultados que o modelo com melhor desempenho na validação não necessariamente possui o melhor desempenho no teste. Como exemplo, o modelo que apresentou o melhor desempenho para a potência ativa no conjunto de testes foi o próprio modelo linear *baseline* com $R^2=0,9980$, RMSE=87,7 W, MAE=87,7 W e, durante a validação, o modelo *Ridge* apresentou o melhor desempenho com $R^2=0,9985$, RMSE=70,7 W e MAE=55,1 W.

As tabelas Tabela 19 e Tabela 20 apresentam o resumo das métricas de desempenho dos modelos analisados: RMSE, MAE e R^2 . Os resultados são apresentados tanto para a validação quanto para o teste. As métricas de desempenho da validação são apresentadas como “média \pm desvio padrão” devido aos *k-folds*. As métricas de desempenho do teste são apenas os valores obtidos durante o teste. A média das métricas entre validação e teste considera o pior caso ao utilizar como métrica de validação “média + desvio padrão”. Em negrito são apresentados os modelos que apresentaram as melhores médias das métricas de desempenho entre a validação e teste.

Das tabelas Tabela 19 e Tabela 20 é possível observar que para os dois casos o modelo de regressão com penalização L2 (*Ridge*) apresentou o melhor desempenho para estimar tanto a potência ativa quanto a potência reativa do GSPS. Para a

potência ativa, o modelo *Ridge* foi ajustado com $\alpha=3$ e 200 atributos. Para a potência reativa, $\alpha=10^{-5}$ e 120 atributos.

Tabela 19 - Análise de resultados dos modelos de regressão para a potência ativa.

Modelo	Nº Atr.	RMSE	MAE	R^2	
<i>Baseline</i>	Validação	200	83 ± 9,6 W	64 ± 8,4 W	0,9978 ± 0,0005
	Teste	200	87,7 W	62,5 W	0,9980
	Média	n/a	90,15 W	67,45 W	0,9976
<i>Lasso</i>	Validação	200	80 ± 6,9 W	62 ± 3,2 W	0,9980 ± 0,0003
	Teste	200	101,6 W	75,0 W	0,9973
	Média	n/a	94,25 W	70,1 W	0,9975
<i>Ridge</i>	Validação	200	67 ± 3,5 W	52 ± 3,1 W	0,9986 ± 0,0001
	Teste	200	91,6 W	65,9 W	0,9978
	Média	n/a	81,05 W	60,5 W	0,9982
<i>Elastic net</i>	Validação	200	75 ± 10 W	58 ± 6,3 W	0,9982 ± 0,0005
	Teste	200	104,2 W	76,5 W	0,9972
	Média	n/a	94,6 W	70,4 W	0,9975

Fonte: elaborado pelo autor.

Tabela 20 - Análise de resultados dos modelos de regressão para a potência reativa.

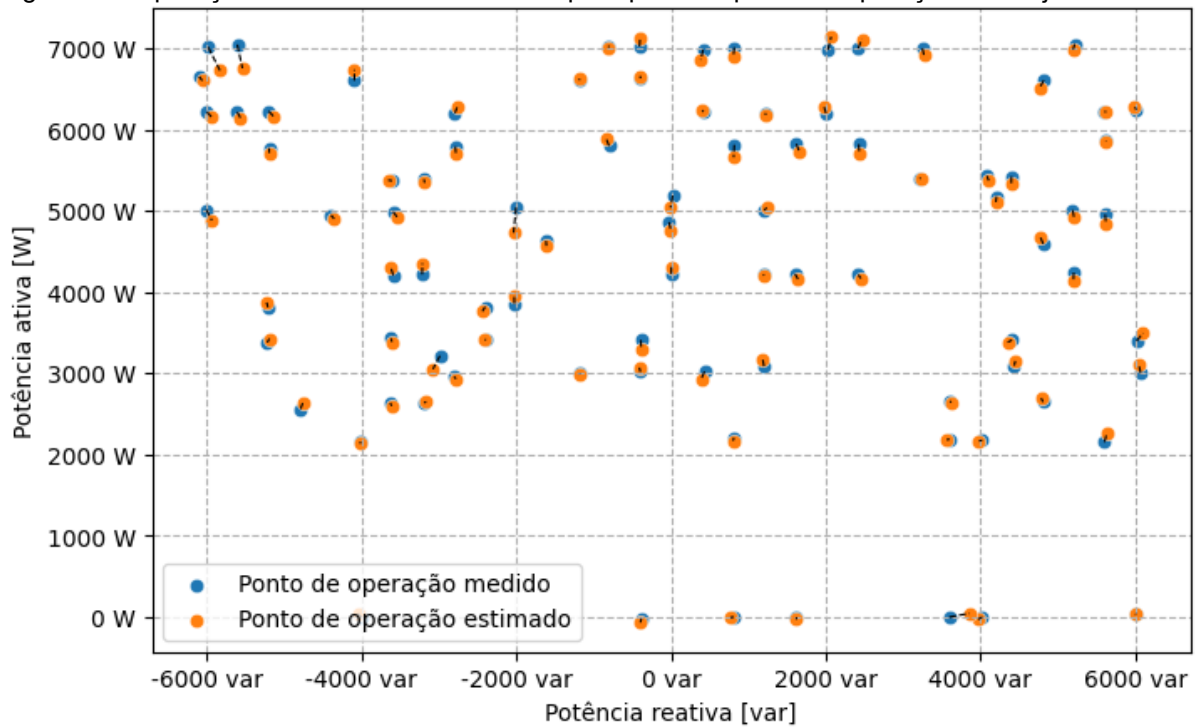
Modelo	Nº Atr.	RMSE	MAE	R^2	
<i>Baseline</i>	Validação	120	35 ± 5,1 var	26 ± 2,8 var	0,9999 ± 0,0000
	Teste	120	47,1 var	32,3 var	0,9998
	Média	n/a	43,6 var	30,55 var	0,9998
<i>Lasso</i>	Validação	120	58 ± 3,6 var	46 ± 3,2 var	0,9997 ± 0,0001
	Teste	120	78,6 var	55,7 var	0,9995
	Média	n/a	70,1 var	52,45 var	0,9995
<i>Ridge</i>	Validação	120	33 ± 2,1 var	26 ± 1,6 var	0,9999 ± 0,0000
	Teste	120	47,1 var	32,3 var	0,9998
	Média	n/a	41,1 var	29,95 var	0,9998
<i>Elastic net</i>	Validação	120	46 ± 0,9 var	36 ± 0,8 var	0,9998 ± 0,0000
	Teste	120	59,5 var	41 var	0,9997
	Média	n/a	53,2 var	38,9 var	0,9997

Fonte: elaborado pelo autor.

Como o modelo *Ridge* apresentou o melhor resultado no treinamento e teste, ele foi utilizado para estimar o ponto de operação e apresentar o resultado em um

mapa do ponto de operação O resultado final é apresentado na Figura 63, onde está presente o mapa do ponto de operação de ensaios com o GPS, onde os pontos medidos estão na cor azul e os pontos estimados pelo modelo final na cor laranja. As linhas pretas tracejadas indicam a distância entre o ponto medido e o ponto estimado. A maior distância é de 326 VA para o ponto teórico de 7 kW e -6 kvar, sendo 7015,21 W e -5984,63 var para o ponto medido e 6726,9 W e -5832,7 var para o ponto estimado.

Figura 63 - Aplicação do modelo final escolhido para prever o ponto de operação do conjunto de teste



Fonte: elaborado pelo autor.

5 CONCLUSÃO

Este trabalho desenvolveu uma metodologia e procedimentos para se estimar o ponto de operação de geradores síncronos conectados à rede elétrica por meio da medição de campos magnéticos externos à máquina. O trabalho empregou dados coletados em um gerador síncrono de polos salientes de uma bancada destinada ao estudo do comportamento de campos magnéticos externos quando há a presença de faltas incipientes na máquina. Neste contexto da detecção de faltas em estágio inicial por meio do campo magnético externo, a revisão integrativa de literatura realizada demonstrou a necessidade de considerar variações no ponto de operação de geradores síncronos conectados à rede elétrica quando analisar o comportamento do campo magnético. De maneira geral, os trabalhos encontrados apresentaram métodos de como monitorar o campo externo e verificar harmônicas alteradas por alguma falta elétrica ou mecânica. Os trabalhos que consideram o ponto de operação em suas análises, relatam que as alterações causadas no campo externo pelo ponto de operação devem ser levadas em consideração em qualquer sistema de monitoramento desta natureza. A exceção se faz presente em trabalhos cujos métodos de detecção de faltas se dizem ser independentes do ponto de operação. Pela experiência e desenvolvimentos na pesquisa no GRUCAD, incluindo análises realizadas em gerador de grande porte de uma usina hidrelétrica, o ponto de operação provoca alterações significativas no campo magnético externo da máquina. Tais alterações podem ser maiores do que aquelas provocadas por faltas incipientes. Um estudo desenvolvido no GRUCAD apresentou um conjunto de diretrizes para monitorar a saúde de geradores síncronos sob o ponto de vista de perceber alterações no campo magnético externo e inferir sobre a existência de alguma falta que já esteja ocorrendo na máquina ou ainda em estado inicial. Dentro deste conjunto de diretrizes se destaca a necessidade do estudo e da caracterização do ponto de operação.

Esta dissertação contribui com um estudo e modelagem do ponto de operação do gerador síncrono de polos salientes utilizando apenas a informação contida no campo magnético externo da máquina. A metodologia desenvolvida pode ser aperfeiçoada, adaptada e agregada a sistemas de detecção de faltas de máquinas por meio do campo magnético externo.

Os desenvolvimentos realizados, de maneira original, mostraram que é possível estimar o ponto de operação de um gerador de polos salientes por meio de

informações contidas em seu campo magnético externo, mesmo quando está em operação conectado à rede elétrica. O trabalho foi desenvolvido de maneira a compreender as etapas entendidas como essenciais, sendo elas o estudo do estado da arte por meio de uma revisão integrativa sobre o tema. Experimentos especiais foram realizados para obtenção de um conjunto de dados confiável, onde os procedimentos adotados estão documentados neste manuscrito. Este conjunto de dados foi tratado e, como resultado direto dos experimentos, foram obtidos dois conjuntos de dados: um conjunto de dados independentes do tempo e um conjunto de dados dependentes do tempo. Uma parcela dos dados coletados foi analisada e utilizada como treinamento para gerar um modelo capaz de estimar o ponto de operação do gerador apenas com informações do campo magnético externo. Com o conjunto de dados independentes do tempo, desenvolveram-se diferentes modelos de regressão capazes de estimar o ponto de operação do GSPS empregado com uma diferença de até 105 W para a potência ativa e 78 var para a potência reativa. Estes valores foram os piores casos de estimação, e representam diferenças de 1,05% na estimativa da potência ativa e de 0,78% na estimativa da potência reativa em relação à potência nominal. Estes modelos ainda podem ser aprimorados e ter seus desempenhos aperfeiçoados. Os resultados obtidos foram considerados satisfatórios, indicando a capacidade do método proposto.

O estudo realizado sobre a caracterização do ponto de operação sob o ponto de vista do campo magnético externo serve como base para avanços futuros no grupo de pesquisa. Pela natureza deste trabalho ser restrito ao GSPS sem faltas, ou seja, os dados foram colhidos com a máquina sadia, acredita-se que os modelos gerados neste trabalho podem ajudar a construção de uma árvore de decisão para verificar se há ou não uma falta incipiente no gerador. A informação da variação do campo magnético externo com o ponto de operação pode ser utilizada para avaliar se uma alteração detectada em uma harmônica foi causada por variações no ponto de operação ou por alguma falta.

Este trabalho apresenta suas limitações devido à complexidade do assunto estudado e à quantidade de variáveis a serem levadas em consideração durante as análises. Portanto é de fundamental importância entender as limitações das análises e modelos propostos neste trabalho.

Uma limitação dos modelos gerados é que todos os passos até a geração final dos modelos foram feitos com base nos dados da taxa de variação do campo

magnético externo, além de considerar todo o processamento eletrônico até a placa de aquisição. Em outras palavras, os modelos gerados ao final deste trabalho são associados à localização dos sensores (CH1, CH2 e CH3), aos sensores em si (aspectos construtivos de cada sensor), ao sistema eletrônico de condicionamento de sinais e ao sistema eletrônico de aquisição de dados. Portanto, caso houver alguma mudança em qualquer um desses itens, os modelos aqui gerados não possuirão a desempenho apresentada.

Dentro da análise dos atributos em frequência, considerou-se que todas as harmônicas rastreadas são aptas para integrarem o conjunto de dados utilizado na modelagem. Apenas as amplitudes das harmônicas foram utilizadas para a modelagem. Não foram avaliados modelos gerados apenas com os atributos oriundos exclusivamente de PCA, t-SNE ou UMAP.

As seguintes sugestões ou perspectivas de continuidade de trabalhos futuros são elencadas:

- Testes de validação dos modelos gerados neste trabalho: Utilizar os modelos gerados neste trabalho para estimar o ponto de operação do gerador. Recomenda-se adicionar uma rotina de cálculo para estimar o campo externo a partir do modelo gerado dentro do sistema de monitoramento de campo externo. Caso o sistema eletrônico de coleta de dados e monitoramento seja diferente do utilizado neste trabalho, deve-se colher novamente os dados e gerar modelos seguindo os passos descritos.
- Utilizar o mesmo método aplicado neste trabalho com outras formas de extrair atributos do conjunto original de dados: pode-se explorar a técnica PHATE ou *autoencoders* e verificar se é possível encontrar outras relações entre os dados de campo e as potências.
- Estudar a confiabilidade das harmônicas do campo magnético externo: Analisar a variação das harmônicas do campo e definir regras/filtros que utilizem a amplitude e o desvio padrão das harmônicas de campo. Espera-se como resultado deste trabalho um direcionamento sobre quais devem ser as harmônicas utilizadas para as análises de ponto de operação e detecção de faltas. Desta maneira, pode-se quantificar a confiança com que o modelo realiza as estimativas.
- A avaliação por meio das fases das componentes harmônicas se mostrou complexa e infrutífera em primeiro momento. Não obstante, a informação contida nas fases pode complementar as avaliações por meio das amplitudes das

harmônicas. Recomenda-se uma investigação detalhada de como o ponto de operação pode impactar as fases das harmônicas do campo externo, e se caso for constatado desvios de fase por parte das harmônicas, esta informação poderá ser agregada ao conjunto de dados independentes do tempo. Ainda dentro deste estudo das fases das harmônicas, podem ser realizados a aplicação dos filtros estudados e desenvolvidos em uma continuidade da investigação.

REFERÊNCIAS

- [1] M. de M. e E. Empresa de Pesquisa Energética, “Anuário Estatístico de Energia Elétrica 2023”, Rio de Janeiro, 2023.
- [2] M. Galea, P. Giangrande, V. Madonna, e G. Buticchi, “Reliability-Oriented Design of Electrical Machines: The Design Process for Machines’ Insulation Systems MUST Evolve”, *IEEE Industrial Electronics Magazine*, vol. 14, nº 1, p. 20–28, mar. 2020, doi: 10.1109/MIE.2019.2947688.
- [3] H. F. D. Santos, “Detecção e avaliação de faltas em geradores síncronos por meio de históricos de harmônicos do campo magnético externo”, Tese de doutorado, Universidade Federal de Santa Catarina, Florianópolis, 2021.
- [4] M. Rigoni, “Desenvolvimento e aplicação de sensores de campo magnético por indução”, Dissertação de mestrado, Universidade Federal de Santa Catarina, Florianópolis, 2009.
- [5] M. Rigoni, “Desenvolvimento de um sistema de detecção e avaliação de faltas em geradores síncronos por meio do campo magnético externo”, Tese de doutorado, Universidade Federal de Santa Catarina, Florianópolis, 2014.
- [6] H. F. D. Santos, “Desenvolvimento de um equipamento para monitoração de geradores síncronos através do campo magnético externo”, Dissertação de mestrado, Universidade Federal de Santa Catarina, Florianópolis, 2016.
- [7] M. Hoegen, “Estratégias de detecção de alterações no comportamento do campo magnético externo de geradores síncronos”, Dissertação de mestrado, Universidade Federal de Santa Catarina, Florianópolis, 2021.
- [8] D. G. Aurich, “Modelagem analítica para simulação de interações entre grandezas eletromagnéticas em geradores síncronos”, Monografia, Universidade Federal de Santa Catarina, Florianópolis, 2022.
- [9] L. B. Antunes, “Estudo de sensores aplicados na detecção de faltas em geradores síncronos”, Monografia, Universidade Federal de Santa Catarina, Florianópolis, 2021.
- [10] G. F. dos Santos *et al.*, “Estudo de aspectos construtivos de sensores de campo magnético por indução”, *Anais: 19 Simpósio Brasileiro de Micro-ondas e Optoeletrônica, 14 Congresso Brasileiro de Eletromagnetismo*, p. 506–510, 2020.
- [11] L. B. Antunes, “Estratégias para detecção de faltas incipientes em geradores síncronos através do fluxo magnético no entreferro”, Dissertação de mestrado, Universidade Federal de Santa Catarina, Florianópolis, 2023.
- [12] H. A. Ferenhof e R. F. Fernandes, “Desmistificando a revisão de literatura como base para redação científica: Método SSF”, *Revista ABC*, vol. 21, nº 3, p. 550–560, 2016.
- [13] H. Ehya, A. Nysveen, e R. Nilssen, “Pattern Recognition of Inter-Turn Short Circuit Fault in Wound Field Synchronous Generator via Stray Flux Monitoring”, em *2020 International Conference on Electrical Machines (ICEM)*, Gothenburg, Sweden: IEEE, ago. 2020, p. 2631–2636. doi: 10.1109/ICEM49940.2020.9270986.
- [14] H. Ehya e A. Nysveen, “Pattern Recognition of Interturn Short Circuit Fault in a Synchronous Generator Using Magnetic Flux”, *IEEE Trans. on Ind. Applicat.*, vol. 57, nº 4, p. 3573–3581, jul. 2021, doi: 10.1109/TIA.2021.3072881.
- [15] H. Ehya, A. Nysveen, e J. A. Antonino-Daviu, “Static, Dynamic and Mixed Eccentricity Faults Detection of Synchronous Generators based on Advanced Pattern Recognition Algorithm”, em *IEEE Int. Symp. Diagn. Electr. Mach., Power*

- Electron. Drives, SDEMPED*, Institute of Electrical and Electronics Engineers Inc., 2021, p. 173–179. doi: 10.1109/SDEMPED51010.2021.9605488.
- [16] H. Ehya, A. Nysveen, R. Nilssen, e Y. Liu, “Static and dynamic eccentricity fault diagnosis of large salient pole synchronous generators by means of external magnetic field”, *IET Electr Power Appl*, vol. 15, n° 7, p. 890–902, 2021, doi: 10.1049/elp2.12068.
- [17] C. E. Shannon, “A Mathematical Theory of Communication”, *Bell System Technical Journal*, vol. 27, n° 3, p. 379–423, jul. 1948, doi: 10.1002/j.1538-7305.1948.tb01338.x.
- [18] H. Ehya, A. Nysveen, e T. N. Skreien, “Performance Evaluation of Signal Processing Tools Used for Fault Detection of Hydrogenerators Operating in Noisy Environments”, *IEEE Trans Ind Appl*, vol. 57, n° 4, p. 3654–3665, 2021, doi: 10.1109/TIA.2021.3078136.
- [19] H. Ehya, A. Nysveen, J. A. Antonino-Daviu, e B. Akin, “Inter-turn Short Circuit Fault Identification of Salient Pole Synchronous Generators by Descriptive Paradigm”, em *IEEE Energy Convers. Congr. Expo., ECCE - Proc.*, Institute of Electrical and Electronics Engineers Inc., 2021, p. 4246–4253. doi: 10.1109/ECCE47101.2021.9595993.
- [20] H. Ehya, A. Nysveen, e J. A. Antonino-Daviu, “Advanced Fault Detection of Synchronous Generators Using Stray Magnetic Field”, *IEEE Trans Ind Electron*, vol. 69, n° 11, p. 11675–11685, 2022, doi: 10.1109/TIE.2021.3118363.
- [21] H. Ehya, A. Nysveen, B. Akin, e K. N. Gyftakis, “An Expeditious Algorithm for Identification and Classification of Rotor Faults in Salient Pole Synchronous Generators”, *IEEE Trans Ind Electron*, vol. 71, n° 2, p. 2008–2017, 2024, doi: 10.1109/TIE.2023.3253954.
- [22] M. Cuevas, R. Romary, J.-P. Lecoite, e T. Jacq, “Non-Invasive Detection of Rotor Short-Circuit Fault in Synchronous Machines by Analysis of Stray Magnetic Field and Frame Vibrations”, *IEEE Trans. Magn.*, vol. 52, n° 7, p. 1–4, 2016, doi: 10.1109/TMAG.2016.2514406.
- [23] M. Cuevas, R. Romary, J.-P. Lecoite, F. Morganti, e T. Jacq, “Noninvasive Detection of Winding Short-Circuit Faults in Salient Pole Synchronous Machine with Squirrel-Cage Damper”, *IEEE Trans Ind Appl*, vol. 54, n° 6, p. 5988–5997, 2018, doi: 10.1109/TIA.2018.2861860.
- [24] M. Cuevas, R. Romary, J.-P. Lecoite, F. Morganti, e T. Jacq, “Non-invasive detection of winding short-circuit faults in salient-pole synchronous machine”, em *IEEE Int. Electric Mach. Drives Conf., IEMDC*, Institute of Electrical and Electronics Engineers Inc., 2017. doi: 10.1109/IEMDC.2017.8002102.
- [25] L. O. S. Grillo, C. A. C. Wengerkievicz, N. J. Batistela, P. Kuo-Peng, e L. M. D. Freitas, “A Method for Statistical Processing of Magnetic Field Sensor Signals for Non-Invasive Condition Monitoring of Synchronous Generators”, *Sensors*, vol. 22, n° 22, 2022, doi: 10.3390/s22228631.
- [26] L. O. S. Grillo *et al.*, “Equipment for Monitoring Synchronous Generators Condition through External Magnetic Field Waveforms”, *J. Microw. Optoelectron. Electromagn. Appl.*, vol. 21, n° 4, p. 527–548, dez. 2022, doi: 10.1590/2179-10742022v21i4268102.
- [27] H. F. Dos Santos *et al.*, “Non-invasive monitoring system of synchronous generator using external field”, *J. Microwaves Optoelectron. Electromag. Appl.*, vol. 16, n° 1, p. 70–89, 2017, doi: 10.1590/2179-10742017v16i1872.
- [28] P. Tian, J. M. Guerrero, C. A. Platero, S. B. Lee, K. N. Gyftakis, e J. Antonino-Daviu, “Condition Monitoring of Rotating Diodes in Synchronous Machines

- Through the Exciter Stray Flux Analysis”, em *IEEE Trans Ind Appl*, Institute of Electrical and Electronics Engineers Inc., 2023, p. 3175–3185. doi: 10.1109/TIA.2023.3234934.
- [29] H. Bechara, R. Zemouri, A. Tahan, B. Kedjar, A. Merkhof, e K. Al-Haddad, “Non-Invasive Anomaly Diagnosis for Hydro Electrical Generators Rotor Inter-Turn Short-Circuit Detection Using Stray Flux and the VAE”, em *Int. Conf. Electr. Mach., ICEM*, Institute of Electrical and Electronics Engineers Inc., 2022, p. 1855–1861. doi: 10.1109/ICEM51905.2022.9910797.
- [30] R. Romary, R. Pusca, e T. Jacq, “Identification of sensitive feature in the stray magnetic field to detect rotor short-circuit fault in synchronous generators”, em *Proc. IEEE Int. Symp. Diagn. Electr. Mach., Power Electron. Drives, SDEMPED*, Zarri L. e Lee S.B., Orgs., Institute of Electrical and Electronics Engineers Inc., 2023, p. 123–128. doi: 10.1109/SDEMPED54949.2023.10271496.
- [31] “AQtech”, AQTech. Acesso em: 20 de julho de 2023. [Online]. Disponível em: <https://www.aqtech.com/hidro>
- [32] MATLAB, “Amplitude Estimation and Zero Padding - MATLAB & Simulink”. Acesso em: 24 de fevereiro de 2024. [Online]. Disponível em: <https://www.mathworks.com/help/signal/ug/amplitude-estimation-and-zero-padding.html>
- [33] M. Baak, R. Koopman, H. Snoek, e S. Klous, “A new correlation coefficient between categorical, ordinal and interval variables with Pearson characteristics”, *Computational Statistics & Data Analysis*, vol. 152, p. 107043, dez. 2020, doi: 10.1016/j.csda.2020.107043.
- [34] D. G. Pascual, *Artificial Intelligence Tools DECISION SUPPORT SYSTEMS IN CONDITION MONITORING AND DIAGNOSIS*, 1º ed. Boca Raton, FL, 2015.
- [35] E. Alencar, “O que é a correlação espúria”, Medium. Acesso em: 13 de fevereiro de 2024. [Online]. Disponível em: <https://euleralencar.medium.com/o-que-%C3%A9-a-correla%C3%A7%C3%A3o-esp%C3%BAria-819b77482764>
- [36] “Spurious Correlations”. Acesso em: 13 de fevereiro de 2024. [Online]. Disponível em: <https://www.tylervigen.com/spurious-correlations>
- [37] J. I. Daoud, “Multicollinearity and Regression Analysis”, *J. Phys.: Conf. Ser.*, vol. 949, p. 012009, dez. 2017, doi: 10.1088/1742-6596/949/1/012009.
- [38] “What is multicollinearity and how to remove it? | by Sharoon Saxena | Analytics Vidhya | Medium”. Acesso em: 13 de fevereiro de 2024. [Online]. Disponível em: <https://medium.com/analytics-vidhya/what-is-multicollinearity-and-how-to-remove-it-413c419de2f>
- [39] G. S. R, “Why Is Multicollinearity A Problem?”, Medium. Acesso em: 13 de fevereiro de 2024. [Online]. Disponível em: <https://medium.com/@gowthamsr37/why-multicollinearity-is-a-problem-4c273d1d1f2e>
- [40] D. C. Montgomery e G. C. Runger, *Applied statistics and probability for engineers*, 3rd ed. New York: Wiley, 2003.
- [41] J. Ramzai, “Clearly explained: Pearson V/S Spearman Correlation Coefficient”, Medium. Acesso em: 24 de fevereiro de 2024. [Online]. Disponível em: <https://towardsdatascience.com/clearly-explained-pearson-v-s-spearman-correlation-coefficient-ada2f473b8>
- [42] J. Brownlee, “How to Calculate Correlation Between Variables in Python”, *MachineLearningMastery.com*. Acesso em: 24 de fevereiro de 2024. [Online]. Disponível em: <https://machinelearningmastery.com/how-to-use-correlation-to-understand-the-relationship-between-variables/>

- [43] R. Python, “NumPy, SciPy, and pandas: Correlation With Python – Real Python”. Acesso em: 24 de fevereiro de 2024. [Online]. Disponível em: <https://realpython.com/numpy-scipy-pandas-correlation-python/>
- [44] K. R. Chan, “How to calculate Pearson, Spearman and Phik correlation between variables using Python”, Omics Diary. Acesso em: 24 de fevereiro de 2024. [Online]. Disponível em: <https://medium.com/omics-diary/how-to-calculate-pearson-spearman-and-phik-correlation-between-variables-using-python-3853513ca9c2>
- [45] G. W. Corder e D. I. Foreman, *Nonparametric statistics: a step-by-step approach*, Second edition. Hoboken, New Jersey: Wiley, 2014.
- [46] L. Dos Santos Coelho, H. V. Hultmann Ayala, e V. Cocco Mariani, “CO and NOx emissions prediction in gas turbine using a novel modeling pipeline based on the combination of deep forest regressor and feature engineering”, *Fuel*, vol. 355, p. 129366, jan. 2024, doi: 10.1016/j.fuel.2023.129366.
- [47] “3.2. Tuning the hyper-parameters of an estimator”, scikit-learn. Acesso em: 13 de fevereiro de 2024. [Online]. Disponível em: https://scikit-learn/stable/modules/grid_search.html
- [48] “3.1. Cross-validation: evaluating estimator performance”, scikit-learn. Acesso em: 13 de fevereiro de 2024. [Online]. Disponível em: https://scikit-learn/stable/modules/cross_validation.html
- [49] “fidelity/selective”. Fidelity Investments, 22 de janeiro de 2024. Acesso em: 13 de fevereiro de 2024. [Online]. Disponível em: <https://github.com/fidelity/selective>
- [50] J. Brownlee, “How to Handle Big-p, Little-n ($p \gg n$) in Machine Learning”, MachineLearningMastery.com. Acesso em: 13 de fevereiro de 2024. [Online]. Disponível em: <https://machinelearningmastery.com/how-to-handle-big-p-little-n-p-n-in-machine-learning/>
- [51] J. Brownlee, “Introduction to Dimensionality Reduction for Machine Learning”, MachineLearningMastery.com. Acesso em: 13 de fevereiro de 2024. [Online]. Disponível em: <https://machinelearningmastery.com/dimensionality-reduction-for-machine-learning/>
- [52] “6.5. Unsupervised dimensionality reduction”, scikit-learn. Acesso em: 13 de fevereiro de 2024. [Online]. Disponível em: https://scikit-learn/stable/modules/unsupervised_reduction.html
- [53] J. Brownlee, “6 Dimensionality Reduction Algorithms With Python”, MachineLearningMastery.com. Acesso em: 13 de fevereiro de 2024. [Online]. Disponível em: <https://machinelearningmastery.com/dimensionality-reduction-algorithms-with-python/>
- [54] “2.2. Manifold learning”, scikit-learn. Acesso em: 13 de fevereiro de 2024. [Online]. Disponível em: <https://scikit-learn/stable/modules/manifold.html>
- [55] “2.5. Decomposing signals in components (matrix factorization problems)”, scikit-learn. Acesso em: 13 de fevereiro de 2024. [Online]. Disponível em: <https://scikit-learn/stable/modules/decomposition.html>
- [56] J. Brownlee, “Principal Component Analysis for Dimensionality Reduction in Python”, MachineLearningMastery.com. Acesso em: 13 de fevereiro de 2024. [Online]. Disponível em: <https://machinelearningmastery.com/principal-components-analysis-for-dimensionality-reduction-in-python/>
- [57] J. Brownlee, “How to Calculate Principal Component Analysis (PCA) from Scratch in Python”, MachineLearningMastery.com. Acesso em: 13 de fevereiro de 2024. [Online]. Disponível em: <https://machinelearningmastery.com/calculate-principal-component-analysis-scratch-python/>

- [58] “sklearn.decomposition.PCA”, scikit-learn. Acesso em: 13 de fevereiro de 2024. [Online]. Disponível em: <https://scikit-learn/stable/modules/generated/sklearn.decomposition.PCA.html>
- [59] “Importance of Feature Scaling”, scikit-learn. Acesso em: 13 de fevereiro de 2024. [Online]. Disponível em: https://scikit-learn/stable/auto_examples/preprocessing/plot_scaling_importance.html
- [60] “6.3. Preprocessing data”, scikit-learn. Acesso em: 13 de fevereiro de 2024. [Online]. Disponível em: <https://scikit-learn/stable/modules/preprocessing.html>
- [61] “sklearn.manifold.TSNE”, scikit-learn. Acesso em: 13 de fevereiro de 2024. [Online]. Disponível em: <https://scikit-learn/stable/modules/generated/sklearn.manifold.TSNE.html>
- [62] “t-SNE”, Laurens van der Maaten. Acesso em: 13 de fevereiro de 2024. [Online]. Disponível em: <https://lvdmaaten.github.io/tsne/>
- [63] M. Wattenberg, F. Viégas, e I. Johnson, “How to Use t-SNE Effectively”, *Distill*, vol. 1, nº 10, p. e2, out. 2016, doi: 10.23915/distill.00002.
- [64] L. van der Maaten, “Barnes-Hut-SNE”. arXiv, 8 de março de 2013. Acesso em: 13 de fevereiro de 2024. [Online]. Disponível em: <http://arxiv.org/abs/1301.3342>
- [65] L. van der Maaten e G. Hinton, “vandermaaten08a.pdf”, *Journal of Machine Learning Research*, Netherlands, p. 27, 2008.
- [66] L. McInnes, J. Healy, e J. Melville, “UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction”. arXiv, 17 de setembro de 2020. Acesso em: 13 de fevereiro de 2024. [Online]. Disponível em: <http://arxiv.org/abs/1802.03426>
- [67] “UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction — umap 0.5 documentation”. Acesso em: 13 de fevereiro de 2024. [Online]. Disponível em: <https://umap-learn.readthedocs.io/en/latest/>
- [68] “1.1. Linear Models”, scikit-learn. Acesso em: 25 de fevereiro de 2024. [Online]. Disponível em: https://scikit-learn/stable/modules/linear_model.html
- [69] “Diagrama de caixa”, *Wikipédia, a enciclopédia livre*. 22 de março de 2023. Acesso em: 19 de fevereiro de 2024. [Online]. Disponível em: https://pt.wikipedia.org/w/index.php?title=Diagrama_de_caixa&oldid=65539971

APÊNDICE A – Dados obtidos do sistema de monitoramento e dados de temperatura

Neste apêndice estão apresentados todos os dados brutos obtidos do sistema de monitoramento VibraOne em conjunto com os dados do monitoramento de temperatura. Observa-se que os dados apresentados entre a Figura 64 e a Figura 76 não possuem nenhum tratamento. Os filtros aplicados para obter o ponto de operação em cada patamar de carregamento estão descritos no capítulo 3. As figuras apresentam o ponto de operação representado pela potência ativa (curva em azul) e a potência reativa (curva em amarelo). A temperatura do gerador (curva em vermelho) indica a temperatura externa do gerador utilizado para a coleta de dados. A temperatura ambiente ao qual o gerador esteve submetido durante os testes está apresentada na curva verde.

A Figura 77 apresenta os dez pontos de operação extra que são intermediários aos demais pontos de operação. Os pontos de operação extra que foram coletados são: (3200 W, 3000 var), (3200 W, 0 var), (3200 W, -3000 var), (4600 W, 4200 var), (4600 W, 0 var), (4600 W, -4200 var), (4800 W, 0 var), (5200 W, 4200 var), (5200 W, 0 var), (5200 W, -4200 var). A Figura 78 apresenta os dados coletado para o gerador operando como compensador síncrono.

Figura 64 - Varredura dos pontos de operação para a potência ativa de 7 kW.

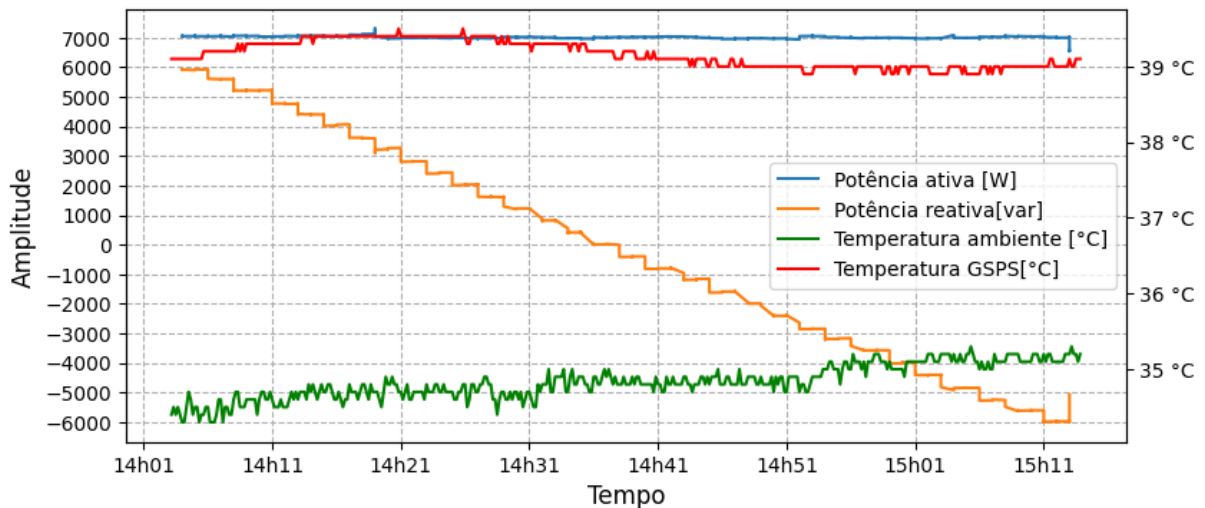


Figura 65 - Varredura dos pontos de operação para a potência ativa de 6,6 kW.

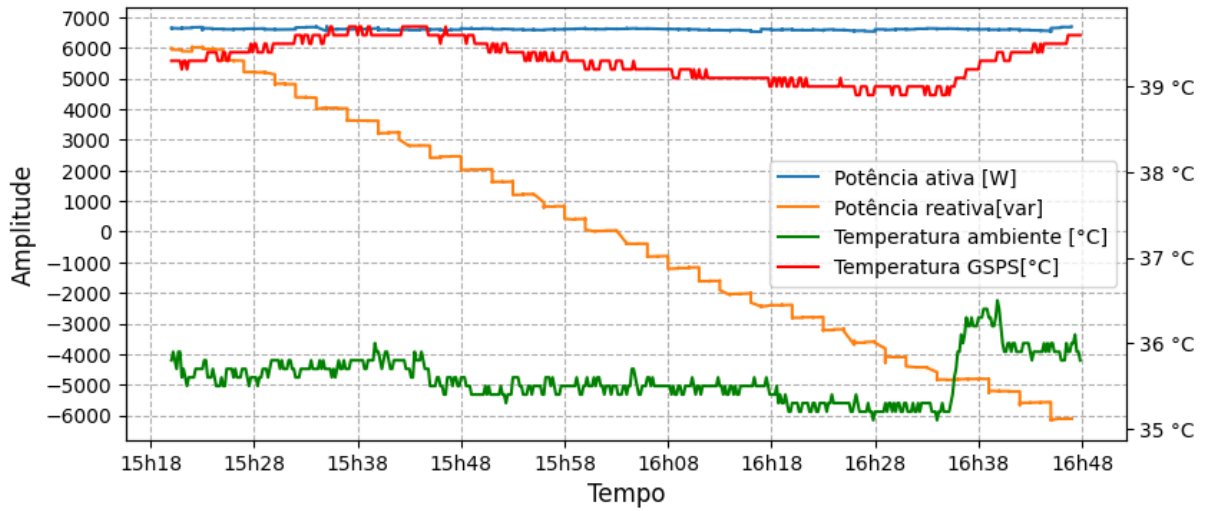


Figura 66 - Varredura dos pontos de operação para a potência ativa de 6,2 kW.

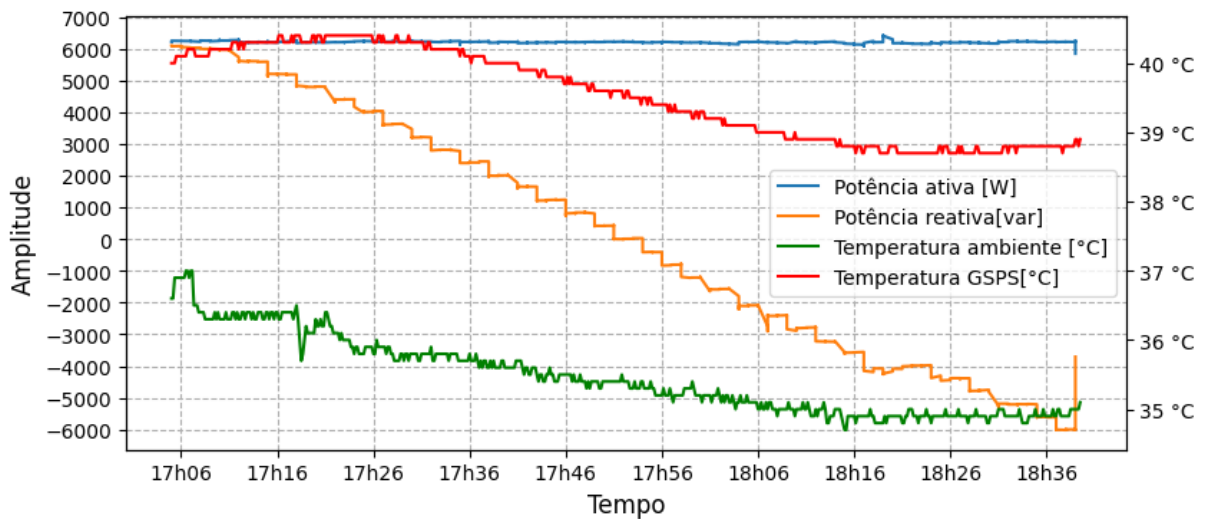


Figura 67 - Varredura dos pontos de operação para a potência ativa de 5,8 kW.

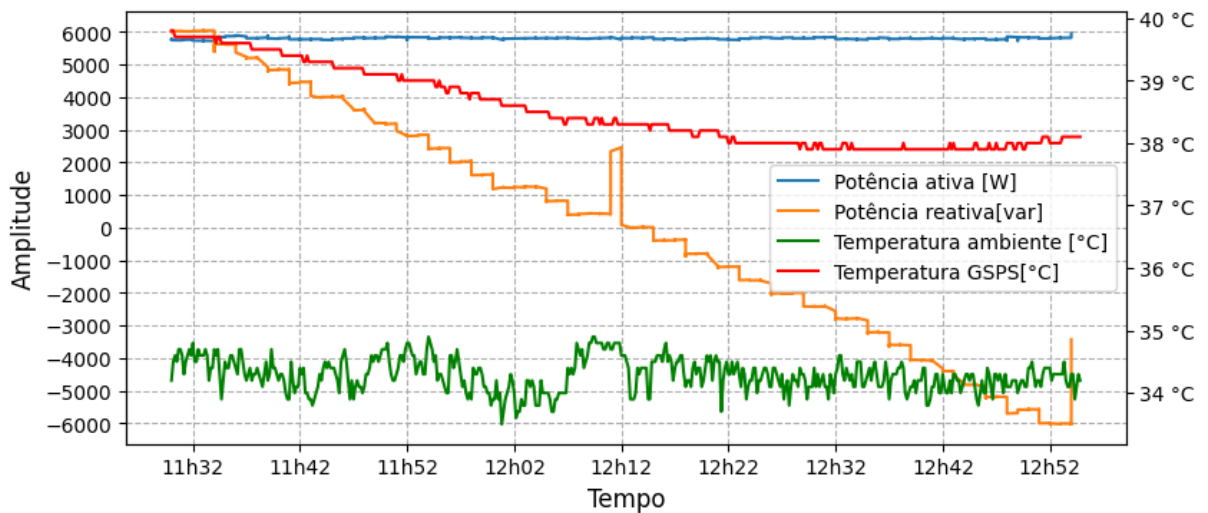


Figura 68 - Varredura dos pontos de operação para a potência ativa de 5,4 kW.

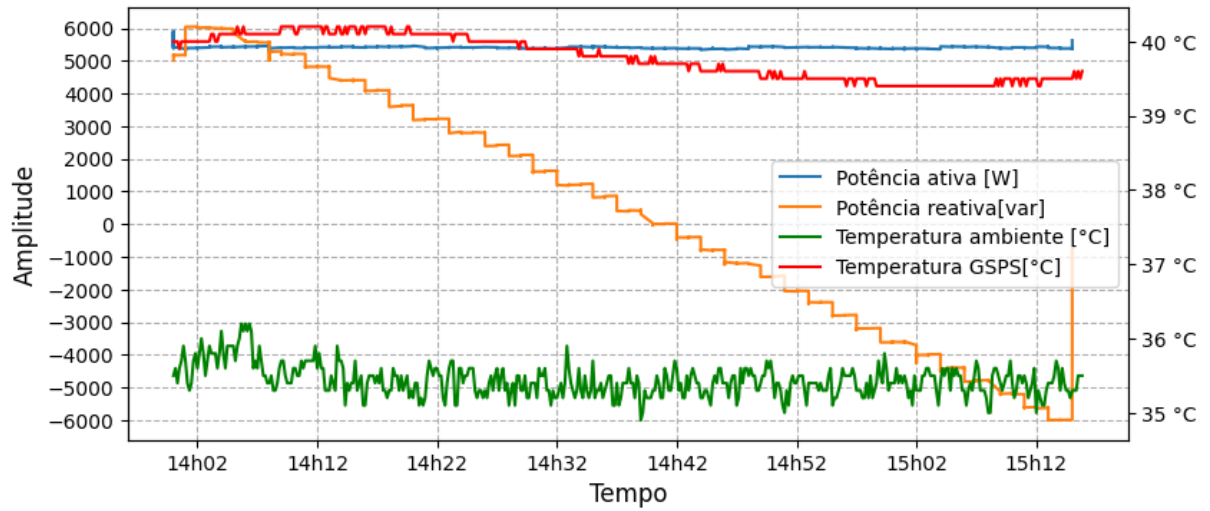


Figura 69 - Varredura dos pontos de operação para a potência ativa de 5 kW.

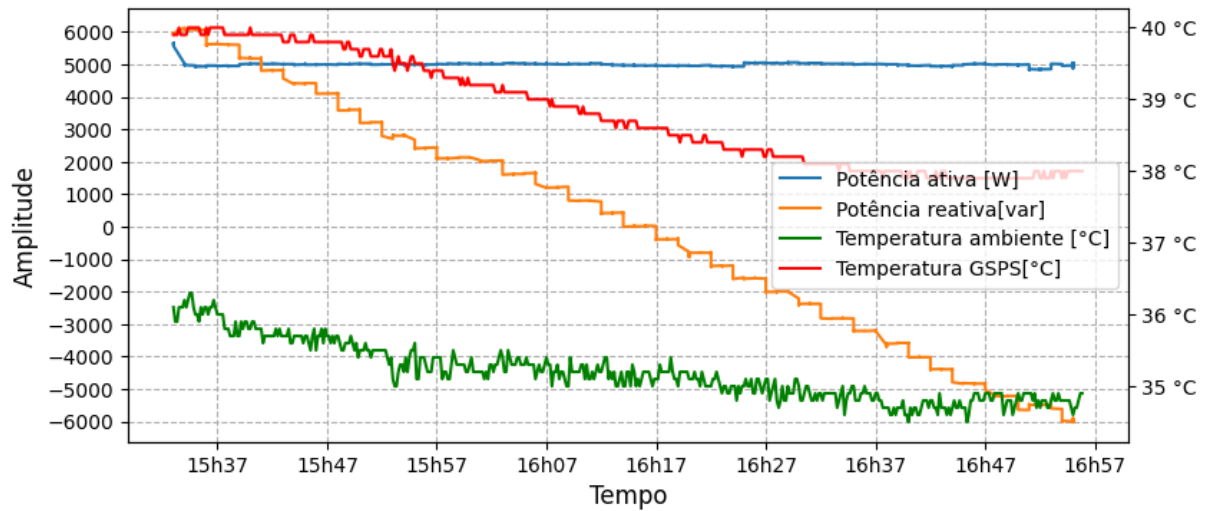


Figura 70 - Varredura dos pontos de operação para a potência ativa de 4,6 kW.

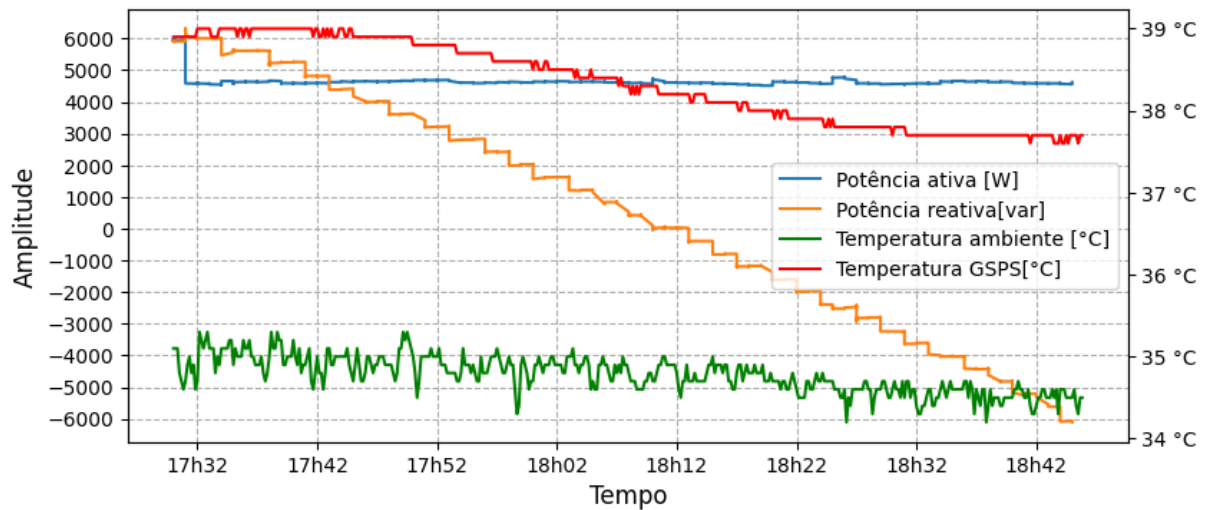


Figura 71 - Varredura dos pontos de operação para a potência ativa de 4,2 kW.

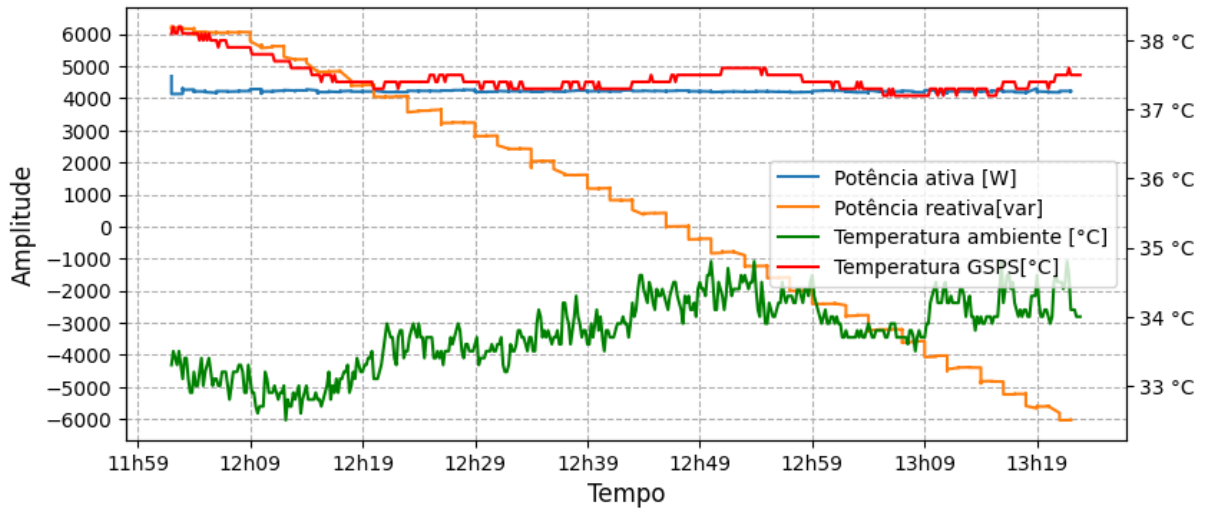


Figura 72 - Varredura dos pontos de operação para a potência ativa de 3,8 kW.

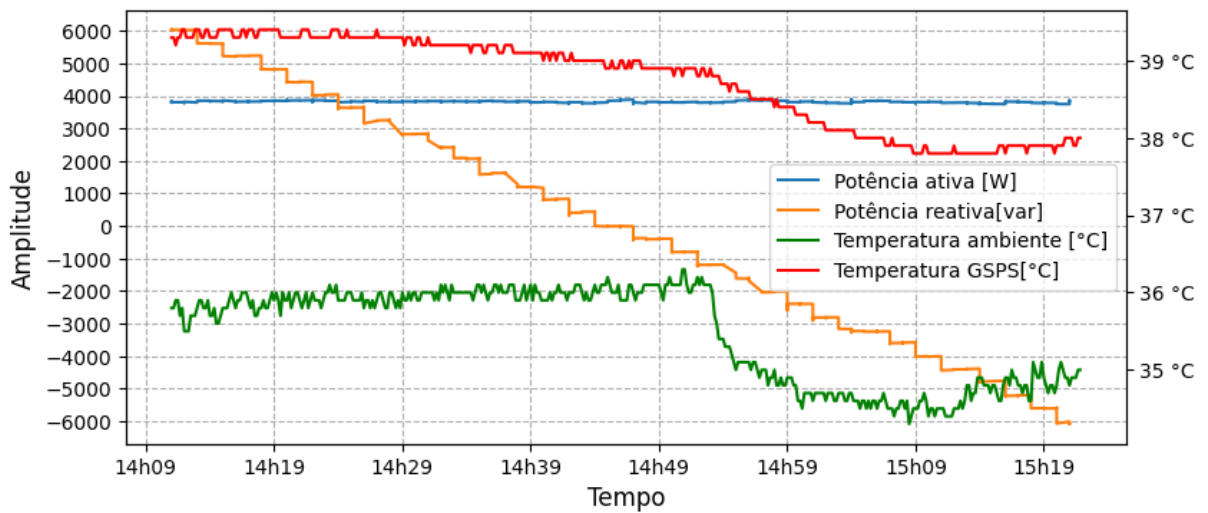


Figura 73 - Varredura dos pontos de operação para a potência ativa de 3,4 kW.

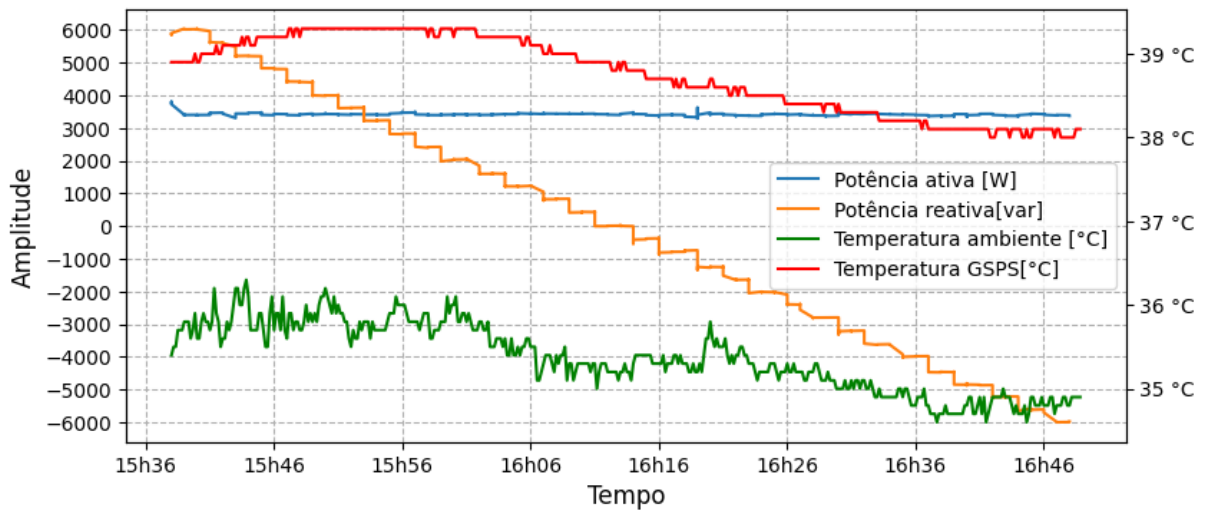


Figura 74 - Varredura dos pontos de operação para a potência ativa de 3 kW.

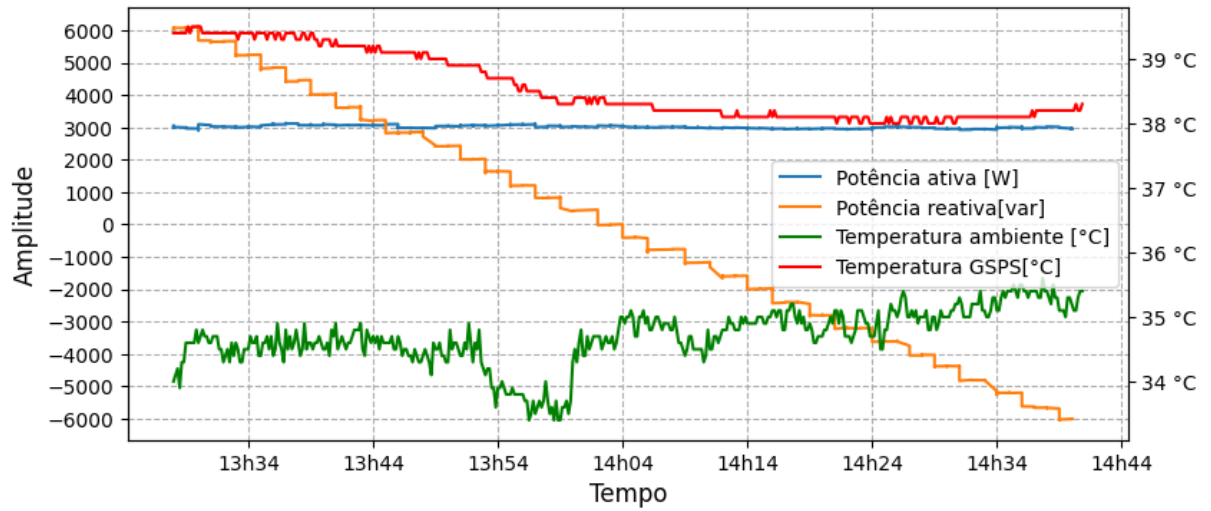


Figura 75 - Varredura dos pontos de operação para a potência ativa de 2,6 kW.

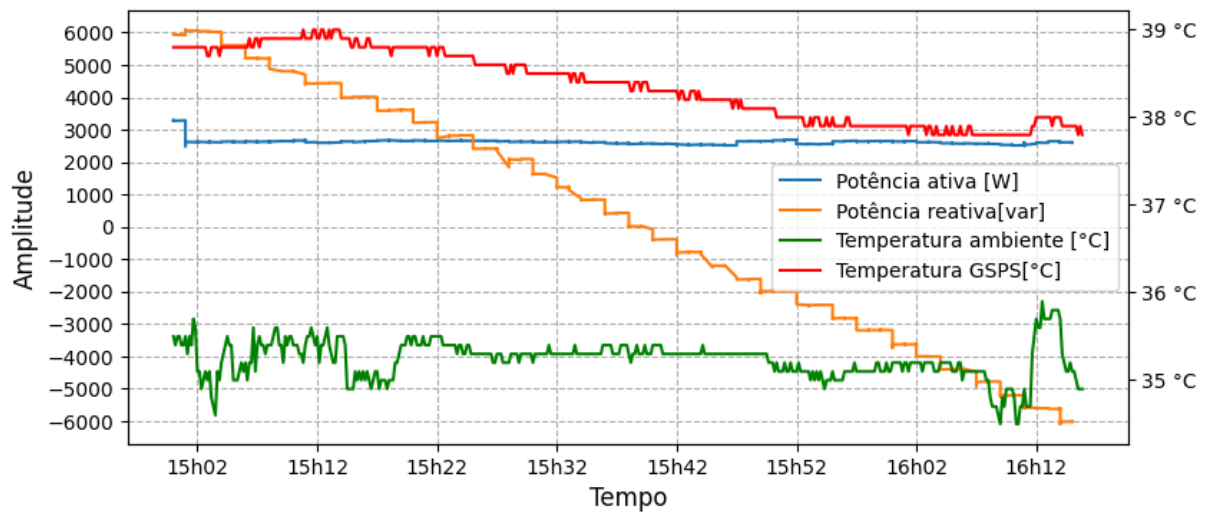


Figura 76 - Varredura dos pontos de operação para a potência ativa de 2,2 kW.

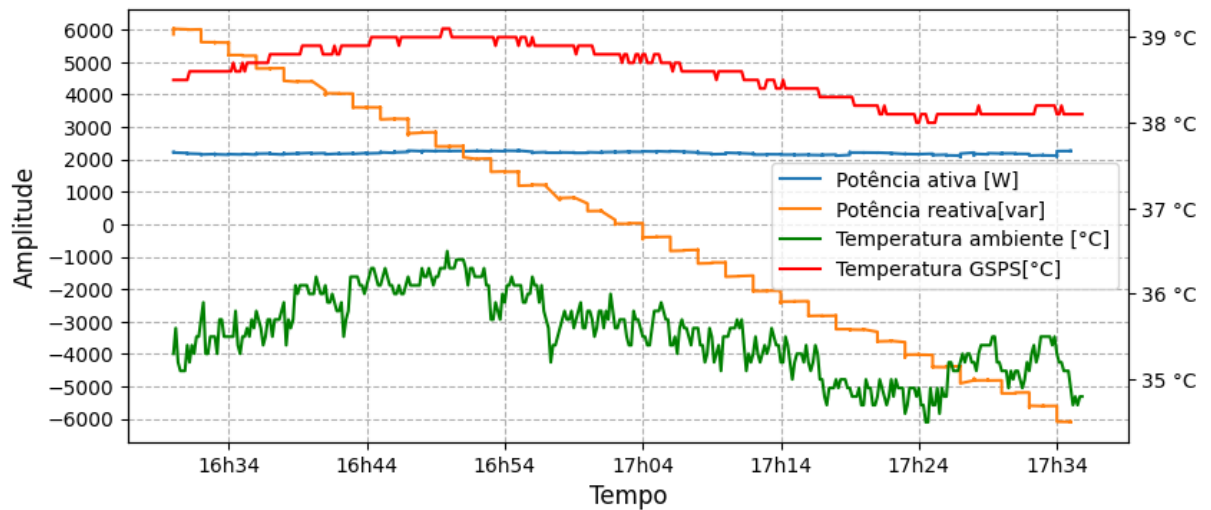


Figura 77 - Varredura de pontos de operação extra.

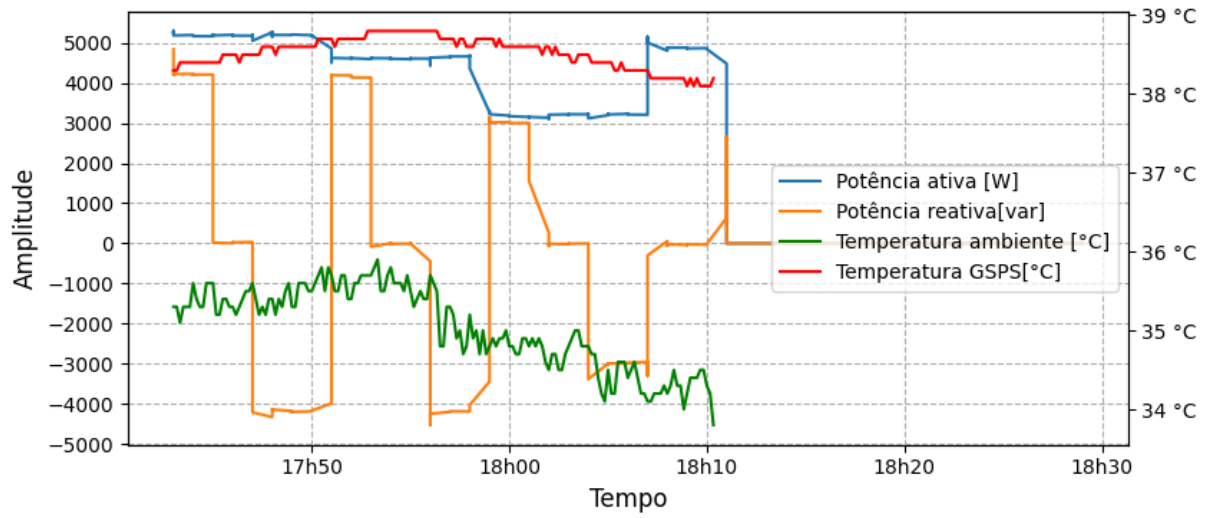
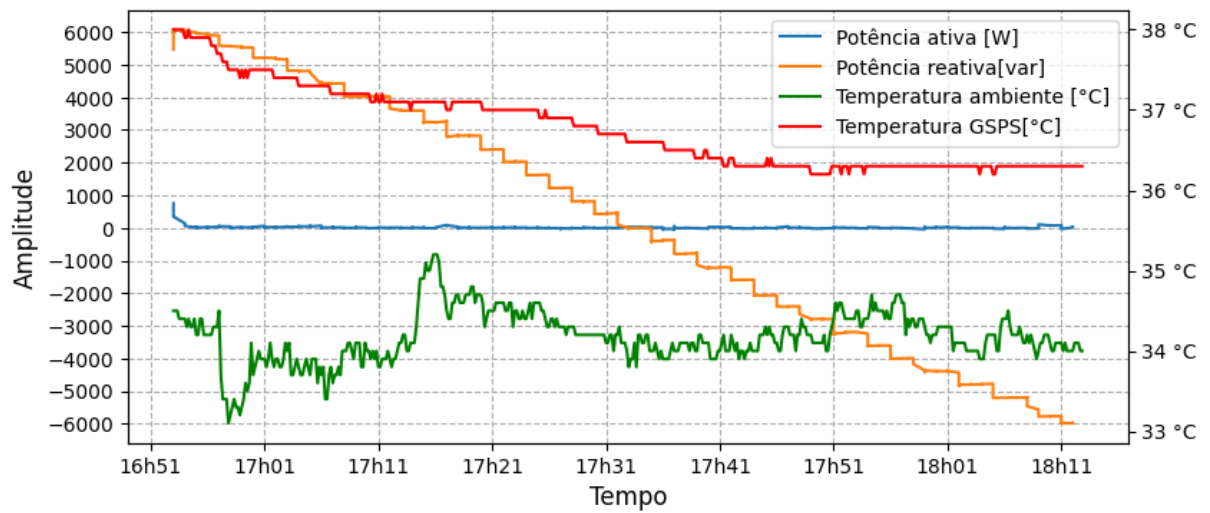


Figura 78 - Varredura dos pontos de operação para a potência ativa de 0 kW. Gerador operando como compensador síncrono.



APÊNDICE B – Pontos de operação

Tabela 21 - Pontos de operação utilizados.

Sequência de operação	Ponto de operação		Sequência de operação	Ponto de operação	
	Potência ativa	Potência reativa		Potência ativa	Potência reativa
1	7000 W	6000 var	223	4200 W	4000 var
2	7000 W	5600 var	224	4200 W	3600 var
3	7000 W	5200 var	225	4200 W	3200 var
4	7000 W	4800 var	226	4200 W	2800 var
5	7000 W	4400 var	227	4200 W	2400 var
6	7000 W	4000 var	228	4200 W	2000 var
7	7000 W	3600 var	229	4200 W	1600 var
8	7000 W	3200 var	230	4200 W	1200 var
9	7000 W	2800 var	231	4200 W	800 var
10	7000 W	2400 var	232	4200 W	400 var
11	7000 W	2000 var	233	4200 W	0 var
12	7000 W	1600 var	234	4200 W	- 400 var
13	7000 W	1200 var	235	4200 W	- 800 var
14	7000 W	800 var	236	4200 W	- 1200 var
15	7000 W	400 var	237	4200 W	- 1600 var
16	7000 W	0 var	238	4200 W	- 2000 var
17	7000 W	- 400 var	239	4200 W	- 2400 var
18	7000 W	- 800 var	240	4200 W	- 2800 var
19	7000 W	- 1200 var	241	4200 W	- 3200 var
20	7000 W	- 1600 var	242	4200 W	- 3600 var
21	7000 W	- 2000 var	243	4200 W	- 4000 var
22	7000 W	- 2400 var	244	4200 W	- 4400 var
23	7000 W	- 2800 var	245	4200 W	- 4800 var
24	7000 W	- 3200 var	246	4200 W	- 5200 var
25	7000 W	- 3600 var	247	4200 W	- 5600 var
26	7000 W	- 4000 var	248	4200 W	- 6000 var
27	7000 W	- 4400 var	249	3800 W	6000 var
28	7000 W	- 4800 var	250	3800 W	5600 var
29	7000 W	- 5200 var	251	3800 W	5200 var
30	7000 W	- 5600 var	252	3800 W	4800 var
31	7000 W	- 6000 var	253	3800 W	4400 var
32	6600 W	6000 var	254	3800 W	4000 var
33	6600 W	5600 var	255	3800 W	3600 var
34	6600 W	5200 var	256	3800 W	3200 var
35	6600 W	4800 var	257	3800 W	2800 var
36	6600 W	4400 var	258	3800 W	2400 var
37	6600 W	4000 var	259	3800 W	2000 var
38	6600 W	3600 var	260	3800 W	1600 var
39	6600 W	3200 var	261	3800 W	1200 var
40	6600 W	2800 var	262	3800 W	800 var
41	6600 W	2400 var	263	3800 W	400 var
42	6600 W	2000 var	264	3800 W	0 var

43	6600 W	1600 var	265	3800 W	- 400 var
44	6600 W	1200 var	266	3800 W	- 800 var
45	6600 W	800 var	267	3800 W	- 1200 var
46	6600 W	400 var	268	3800 W	- 1600 var
47	6600 W	0 var	269	3800 W	- 2000 var
48	6600 W	- 400 var	270	3800 W	- 2400 var
49	6600 W	- 800 var	271	3800 W	- 2800 var
50	6600 W	- 1200 var	272	3800 W	- 3200 var
51	6600 W	- 1600 var	273	3800 W	- 3600 var
52	6600 W	- 2000 var	274	3800 W	- 4000 var
53	6600 W	- 2400 var	275	3800 W	- 4400 var
54	6600 W	- 2800 var	276	3800 W	- 4800 var
55	6600 W	- 3200 var	277	3800 W	- 5200 var
56	6600 W	- 3600 var	278	3800 W	- 5600 var
57	6600 W	- 4000 var	279	3800 W	- 6000 var
58	6600 W	- 4400 var	280	3400 W	6000 var
59	6600 W	- 4800 var	281	3400 W	5600 var
60	6600 W	- 5200 var	282	3400 W	5200 var
61	6600 W	- 5600 var	283	3400 W	4800 var
62	6600 W	- 6000 var	284	3400 W	4400 var
63	6200 W	6000 var	285	3400 W	4000 var
64	6200 W	5600 var	286	3400 W	3600 var
65	6200 W	5200 var	287	3400 W	3200 var
66	6200 W	4800 var	288	3400 W	2800 var
67	6200 W	4400 var	289	3400 W	2400 var
68	6200 W	4000 var	290	3400 W	2000 var
69	6200 W	3600 var	291	3400 W	1600 var
70	6200 W	3200 var	292	3400 W	1200 var
71	6200 W	2800 var	293	3400 W	800 var
72	6200 W	2400 var	294	3400 W	400 var
73	6200 W	2000 var	295	3400 W	0 var
74	6200 W	1600 var	296	3400 W	- 400 var
75	6200 W	1200 var	297	3400 W	- 800 var
76	6200 W	800 var	298	3400 W	- 1200 var
77	6200 W	400 var	299	3400 W	- 1600 var
78	6200 W	0 var	300	3400 W	- 2000 var
79	6200 W	- 400 var	301	3400 W	- 2400 var
80	6200 W	- 800 var	302	3400 W	- 2800 var
81	6200 W	- 1200 var	303	3400 W	- 3200 var
82	6200 W	- 1600 var	304	3400 W	- 3600 var
83	6200 W	- 2000 var	305	3400 W	- 4000 var
84	6200 W	- 2400 var	306	3400 W	- 4400 var
85	6200 W	- 2800 var	307	3400 W	- 4800 var
86	6200 W	- 3200 var	308	3400 W	- 5200 var
87	6200 W	- 3600 var	309	3400 W	- 5600 var
88	6200 W	- 4000 var	310	3400 W	- 6000 var
89	6200 W	- 4400 var	311	3000 W	6000 var
90	6200 W	- 4800 var	312	3000 W	5600 var

91	6200 W	- 5200 var	313	3000 W	5200 var
92	6200 W	- 5600 var	314	3000 W	4800 var
93	6200 W	- 6000 var	315	3000 W	4400 var
94	5800 W	6000 var	316	3000 W	4000 var
95	5800 W	5600 var	317	3000 W	3600 var
96	5800 W	5200 var	318	3000 W	3200 var
97	5800 W	4800 var	319	3000 W	2800 var
98	5800 W	4400 var	320	3000 W	2400 var
99	5800 W	4000 var	321	3000 W	2000 var
100	5800 W	3600 var	322	3000 W	1600 var
101	5800 W	3200 var	323	3000 W	1200 var
102	5800 W	2800 var	324	3000 W	800 var
103	5800 W	2400 var	325	3000 W	400 var
104	5800 W	2000 var	326	3000 W	0 var
105	5800 W	1600 var	327	3000 W	- 400 var
106	5800 W	1200 var	328	3000 W	- 800 var
107	5800 W	800 var	329	3000 W	- 1200 var
108	5800 W	400 var	330	3000 W	- 1600 var
109	5800 W	0 var	331	3000 W	- 2000 var
110	5800 W	- 400 var	332	3000 W	- 2400 var
111	5800 W	- 800 var	333	3000 W	- 2800 var
112	5800 W	- 1200 var	334	3000 W	- 3200 var
113	5800 W	- 1600 var	335	3000 W	- 3600 var
114	5800 W	- 2000 var	336	3000 W	- 4000 var
115	5800 W	- 2400 var	337	3000 W	- 4400 var
116	5800 W	- 2800 var	338	3000 W	- 4800 var
117	5800 W	- 3200 var	339	3000 W	- 5200 var
118	5800 W	- 3600 var	340	3000 W	- 5600 var
119	5800 W	- 4000 var	341	3000 W	- 6000 var
120	5800 W	- 4400 var	342	2600 W	6000 var
121	5800 W	- 4800 var	343	2600 W	5600 var
122	5800 W	- 5200 var	344	2600 W	5200 var
123	5800 W	- 5600 var	345	2600 W	4800 var
124	5800 W	- 6000 var	346	2600 W	4400 var
125	5400 W	6000 var	347	2600 W	4000 var
126	5400 W	5600 var	348	2600 W	3600 var
127	5400 W	5200 var	349	2600 W	3200 var
128	5400 W	4800 var	350	2600 W	2800 var
129	5400 W	4400 var	351	2600 W	2400 var
130	5400 W	4000 var	352	2600 W	2000 var
131	5400 W	3600 var	353	2600 W	1600 var
132	5400 W	3200 var	354	2600 W	1200 var
133	5400 W	2800 var	355	2600 W	800 var
134	5400 W	2400 var	356	2600 W	400 var
135	5400 W	2000 var	357	2600 W	0 var
136	5400 W	1600 var	358	2600 W	- 400 var
137	5400 W	1200 var	359	2600 W	- 800 var
138	5400 W	800 var	360	2600 W	- 1200 var

139	5400 W	400 var	361	2600 W	- 1600 var
140	5400 W	0 var	362	2600 W	- 2000 var
141	5400 W	- 400 var	363	2600 W	- 2400 var
142	5400 W	- 800 var	364	2600 W	- 2800 var
143	5400 W	- 1200 var	365	2600 W	- 3200 var
144	5400 W	- 1600 var	366	2600 W	- 3600 var
145	5400 W	- 2000 var	367	2600 W	- 4000 var
146	5400 W	- 2400 var	368	2600 W	- 4400 var
147	5400 W	- 2800 var	369	2600 W	- 4800 var
148	5400 W	- 3200 var	370	2600 W	- 5200 var
149	5400 W	- 3600 var	371	2600 W	- 5600 var
150	5400 W	- 4000 var	372	2600 W	- 6000 var
151	5400 W	- 4400 var	373	2200 W	6000 var
152	5400 W	- 4800 var	374	2200 W	5600 var
153	5400 W	- 5200 var	375	2200 W	5200 var
154	5400 W	- 5600 var	376	2200 W	4800 var
155	5400 W	- 6000 var	377	2200 W	4400 var
156	5000 W	6000 var	378	2200 W	4000 var
157	5000 W	5600 var	379	2200 W	3600 var
158	5000 W	5200 var	380	2200 W	3200 var
159	5000 W	4800 var	381	2200 W	2800 var
160	5000 W	4400 var	382	2200 W	2400 var
161	5000 W	4000 var	383	2200 W	2000 var
162	5000 W	3600 var	384	2200 W	1600 var
163	5000 W	3200 var	385	2200 W	1200 var
164	5000 W	2800 var	386	2200 W	800 var
165	5000 W	2400 var	387	2200 W	400 var
166	5000 W	2000 var	388	2200 W	0 var
167	5000 W	1600 var	389	2200 W	- 400 var
168	5000 W	1200 var	390	2200 W	- 800 var
169	5000 W	800 var	391	2200 W	- 1200 var
170	5000 W	400 var	392	2200 W	- 1600 var
171	5000 W	0 var	393	2200 W	- 2000 var
172	5000 W	- 400 var	394	2200 W	- 2400 var
173	5000 W	- 800 var	395	2200 W	- 2800 var
174	5000 W	- 1200 var	396	2200 W	- 3200 var
175	5000 W	- 1600 var	397	2200 W	- 3600 var
176	5000 W	- 2000 var	398	2200 W	- 4000 var
177	5000 W	- 2400 var	399	2200 W	- 4400 var
178	5000 W	- 2800 var	401	2200 W	- 4800 var
179	5000 W	- 3200 var	402	2200 W	- 5200 var
180	5000 W	- 3600 var	403	2200 W	- 5600 var
181	5000 W	- 4000 var	404	2200 W	- 6000 var
182	5000 W	- 4400 var	405	0 W	6000 var
183	5000 W	- 4800 var	406	0 W	5600 var
184	5000 W	- 5200 var	407	0 W	5200 var
185	5000 W	- 5600 var	408	0 W	4800 var
186	5000 W	- 6000 var	409	0 W	4400 var

187	4600 W	6000 var	410	0 W	4000 var
188	4600 W	5600 var	411	0 W	3600 var
189	4600 W	5200 var	412	0 W	3200 var
190	4600 W	4800 var	413	0 W	2800 var
191	4600 W	4400 var	414	0 W	2400 var
192	4600 W	4000 var	415	0 W	2000 var
193	4600 W	3600 var	416	0 W	1600 var
194	4600 W	3200 var	417	0 W	1200 var
195	4600 W	2800 var	418	0 W	800 var
196	4600 W	2400 var	419	0 W	400 var
197	4600 W	2000 var	420	0 W	0 var
198	4600 W	1600 var	421	0 W	- 400 var
199	4600 W	1200 var	422	0 W	- 800 var
200	4600 W	800 var	423	0 W	- 1200 var
201	4600 W	400 var	424	0 W	- 1600 var
202	4600 W	0 var	425	0 W	- 2000 var
203	4600 W	- 400 var	426	0 W	- 2400 var
204	4600 W	- 800 var	427	0 W	- 2800 var
205	4600 W	- 1200 var	428	0 W	- 3200 var
206	4600 W	- 1600 var	429	0 W	- 3600 var
207	4600 W	- 2000 var	430	0 W	- 4000 var
208	4600 W	- 2400 var	431	0 W	- 4400 var
209	4600 W	- 2800 var	432	0 W	- 4800 var
210	4600 W	- 3200 var	433	0 W	- 5200 var
211	4600 W	- 3600 var	434	0 W	- 5600 var
212	4600 W	- 4000 var	435	0 W	- 6000 var
213	4600 W	- 4400 var	436	3200 W	3000 var
214	4600 W	- 4800 var	437	3200 W	0 var
215	4600 W	- 5200 var	438	3200 W	- 3000 var
216	4600 W	- 5600 var	439	4600 W	4200 var
217	4600 W	- 6000 var	440	4600 W	0 var
218	4200 W	6000 var	441	4600 W	- 4200 var
219	4200 W	5600 var	442	4800 W	0 var
220	4200 W	5200 var	443	5200 W	4200 var
221	4200 W	4800 var	444	5200 W	0 var
222	4200 W	4400 var	445	5200 W	- 4200 var

APÊNDICE C – Visualização da importância de atributos

O estudo de verificação de importância de atributos tem como objetivo apresentar, de maneira visual, como os coeficientes da regressão linear variam com o aumento do número de atributos. O modelo utilizado para este estudo foi a regressão linear pelo método dos mínimos quadrados

As figuras das importâncias dos coeficientes são uma boa avaliação para entender melhor o modelo de regressão *baseline* e como seus coeficientes podem variar por multicolinearidade. O número de atributos escolhido para visualizar os coeficientes da regressão foram os valores discretos 5, 15, 30 e 60 atributos.

As figuras de importância de atributos apresentam dois gráficos. O gráfico da esquerda é uma visualização de maneira quantitativa dos valores dos coeficientes do modelo de regressão *baseline* para diferentes quantidades de atributos de entrada. O gráfico da direita é um *box-plot*, [69], que apresenta o valor médio do coeficiente adicionado de sua variabilidade. O valor médio do coeficiente tende a ser o mesmo valor apresentado no gráfico da esquerda. A variabilidade do coeficiente foi obtida por uma validação cruzada *k-fold* com 5 divisões. Em cada rodada de validação, os coeficientes da regressão foram salvos e, portanto, a visualização da variação pôde ser obtida.

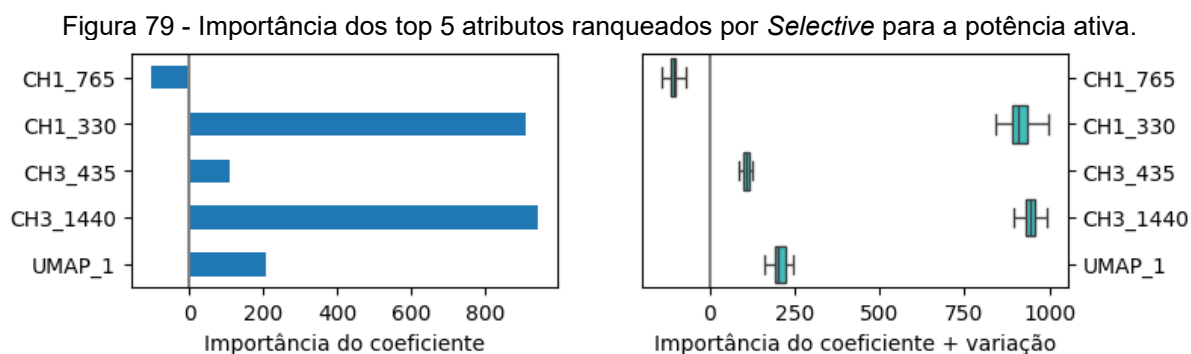


Figura 80 - Importância dos top 15 atributos ranqueados por *Selective* para a potência ativa.

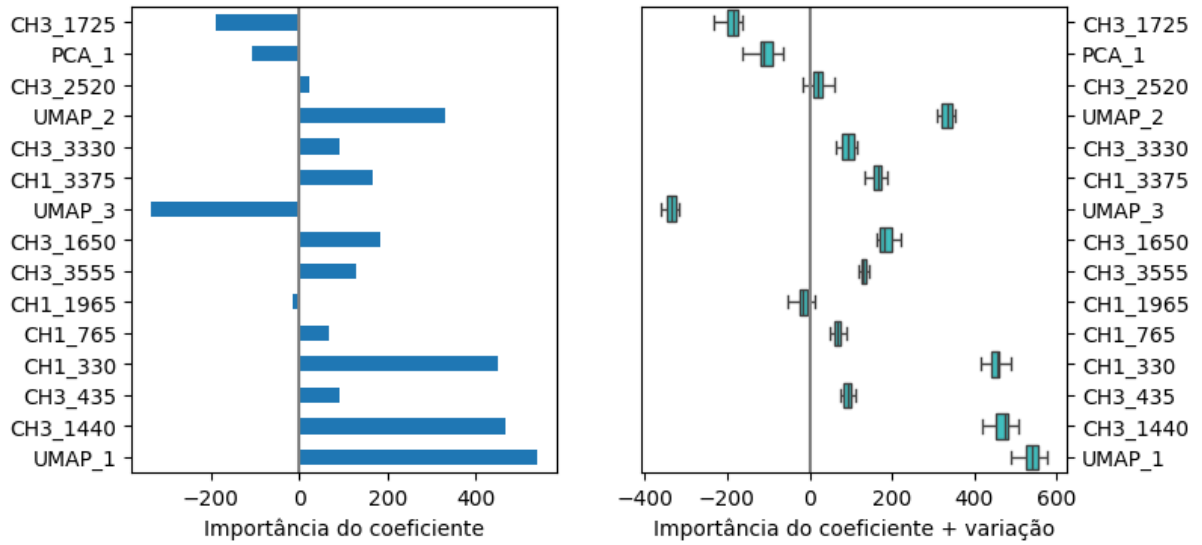
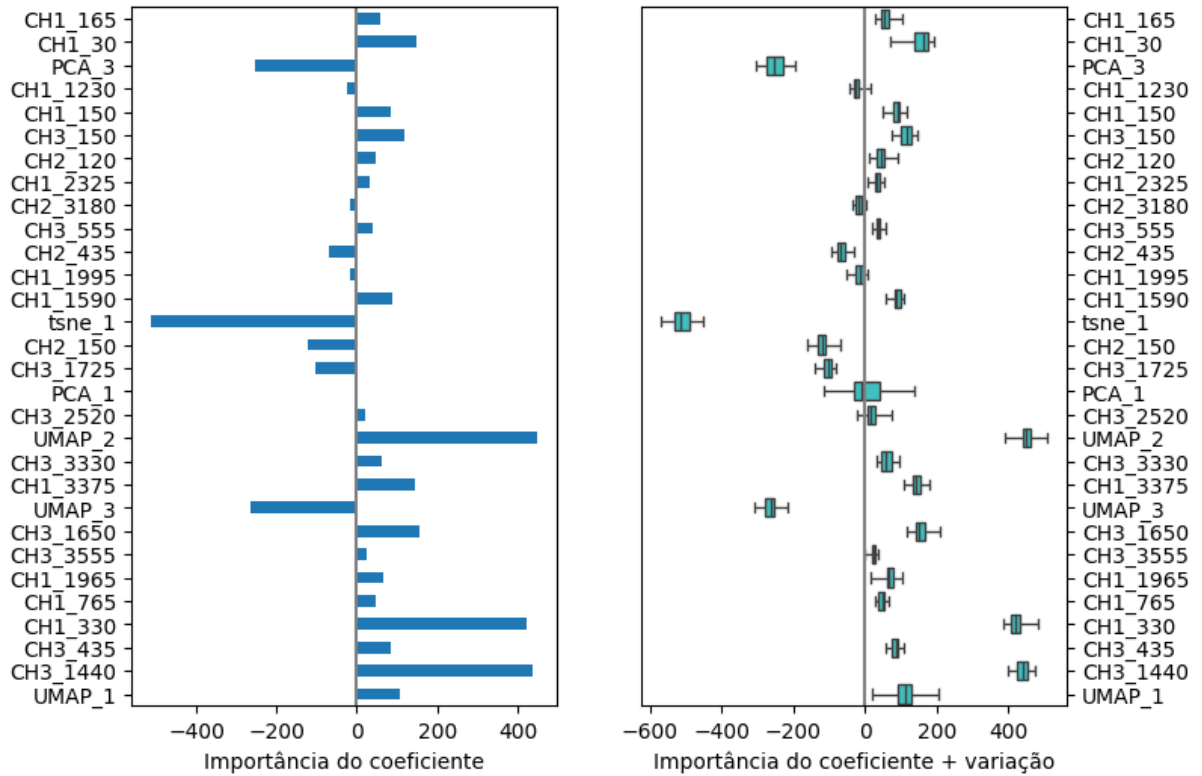
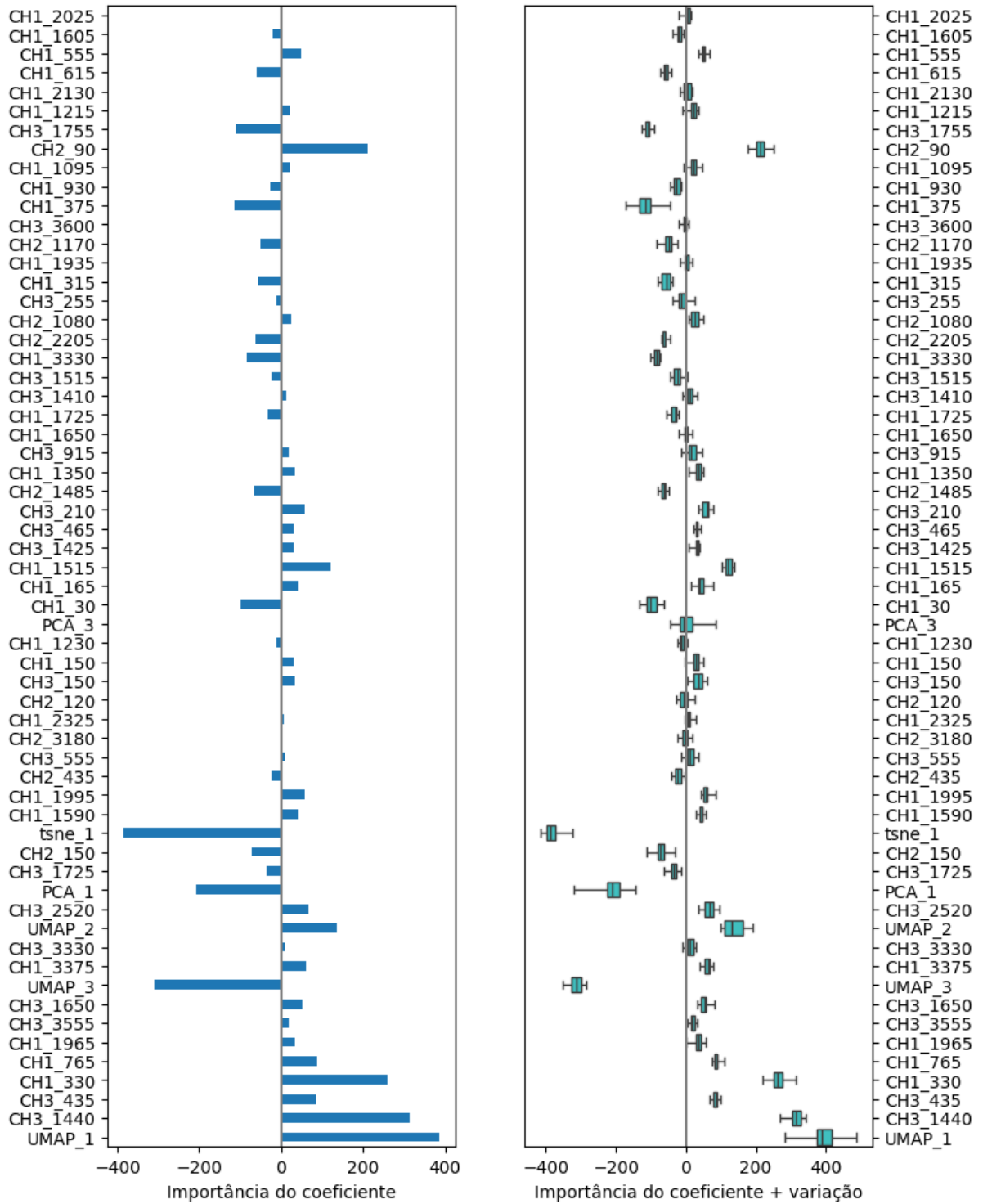


Figura 81 - Importância dos top 30 atributos ranqueados por *Selective* para a potência ativa.



Fonte: elaborado pelo autor.

Figura 82 - Importância dos top 60 atributos ranqueados por *Selective* para a potência ativa.



Fonte: elaborado pelo autor.

Figura 83 - Importância dos top 5 atributos ranqueados por *Selective* para a potência reativa.

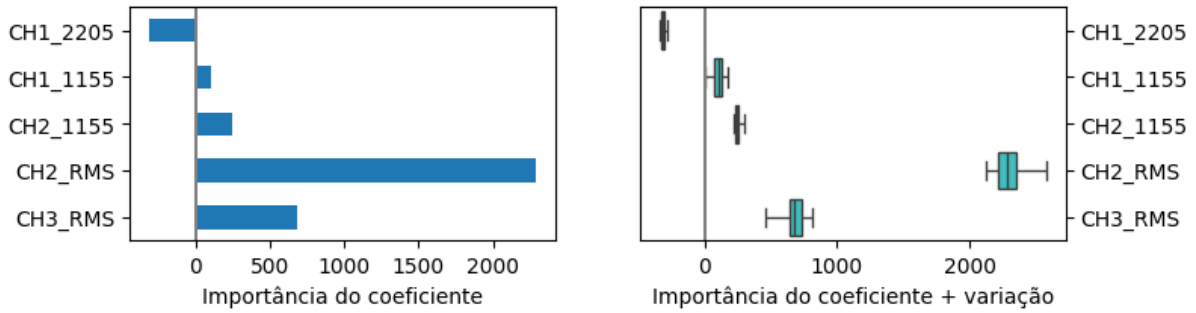


Figura 84 - Importância dos top 15 atributos ranqueados por *Selective* para a potência reativa.

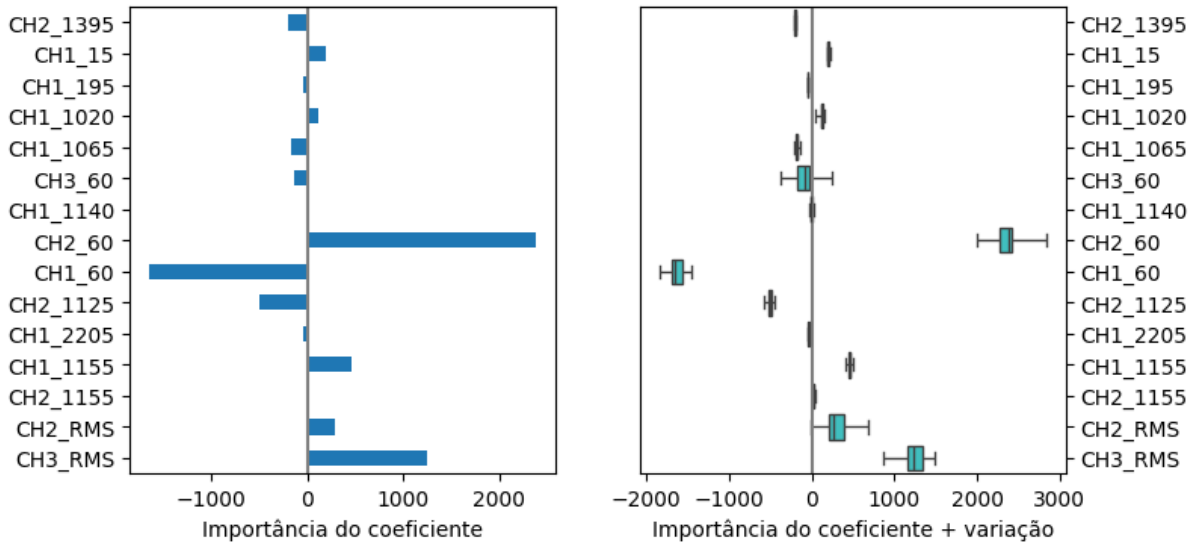
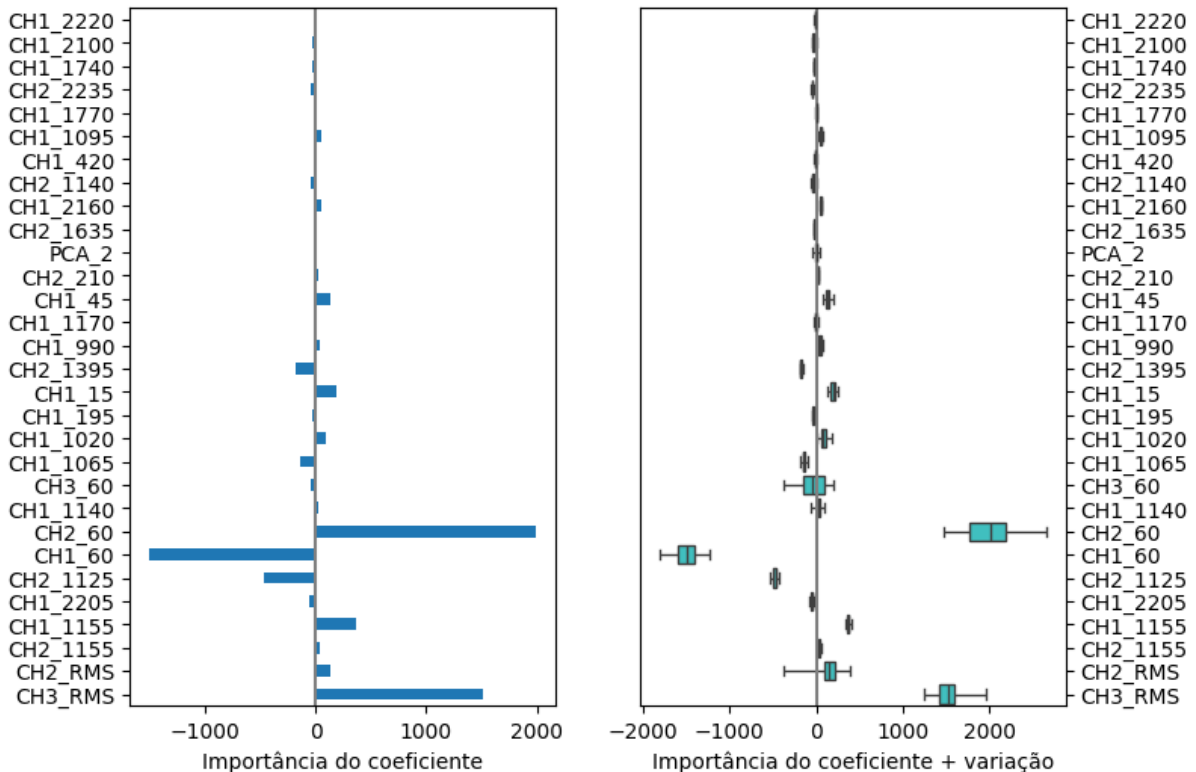
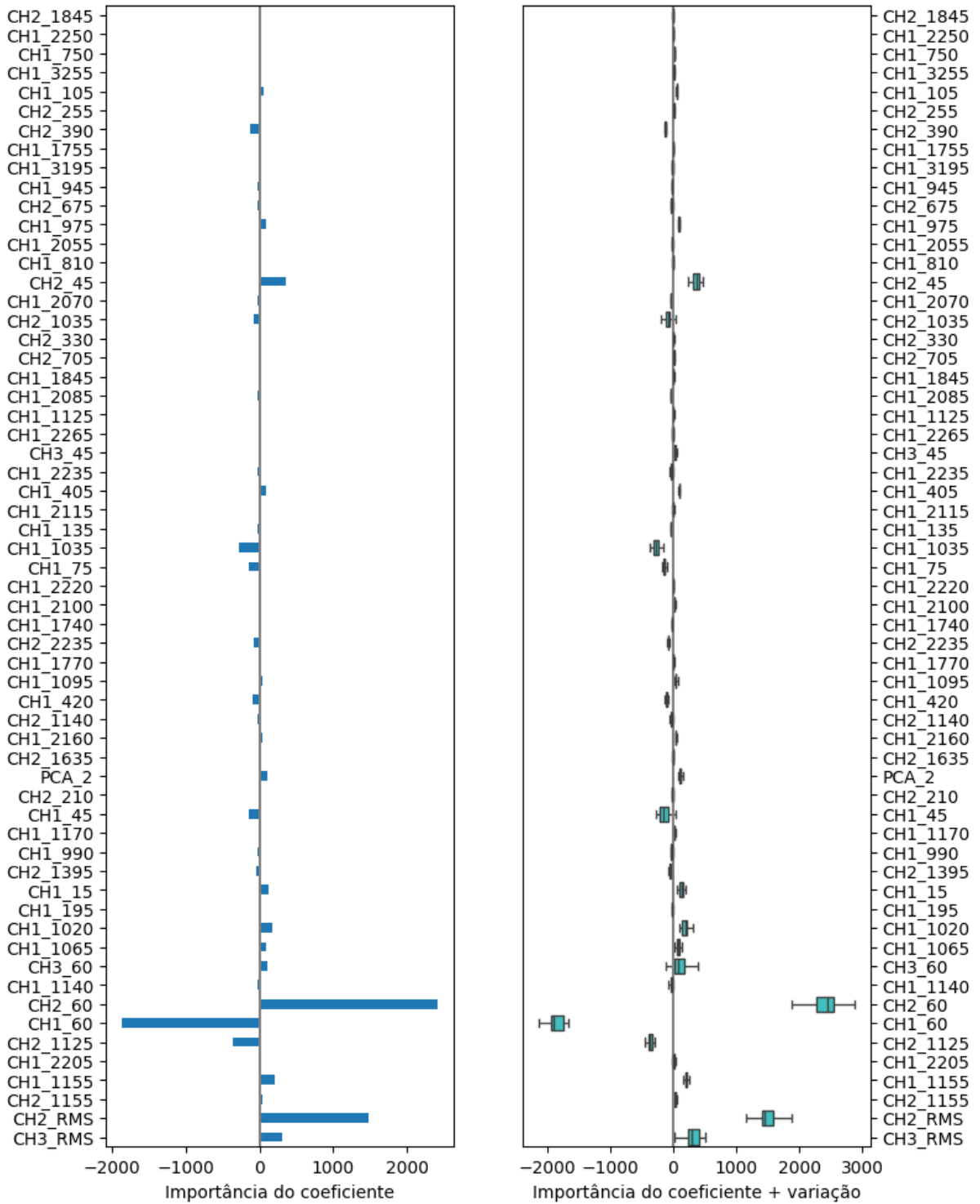


Figura 85 - Importância dos top 30 atributos ranqueados por *Selective* para a potência reativa.



Fonte: elaborado pelo autor.

Figura 86 - Importância dos top 60 atributos ranqueados por *Selective* para a potência reativa.



Fonte: elaborado pelo autor.

Figura 87 - Importância dos top 5 atributos ranqueados por correlação para a potência ativa.

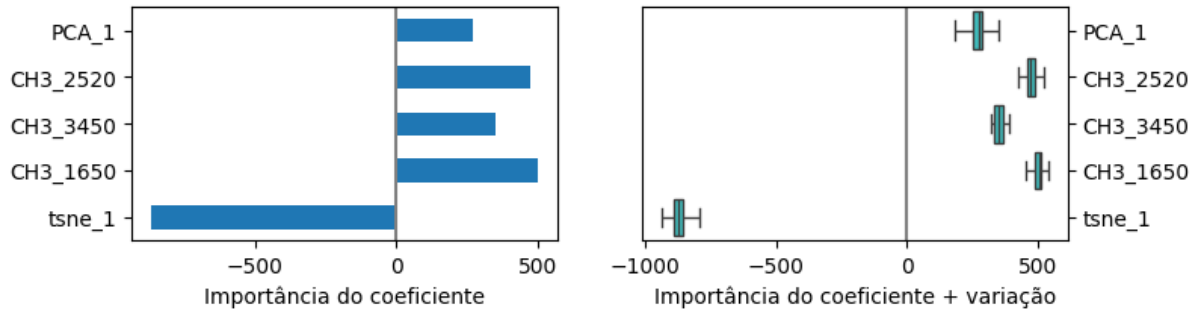


Figura 88 - Importância dos top 15 atributos ranqueados por correlação para a potência ativa.

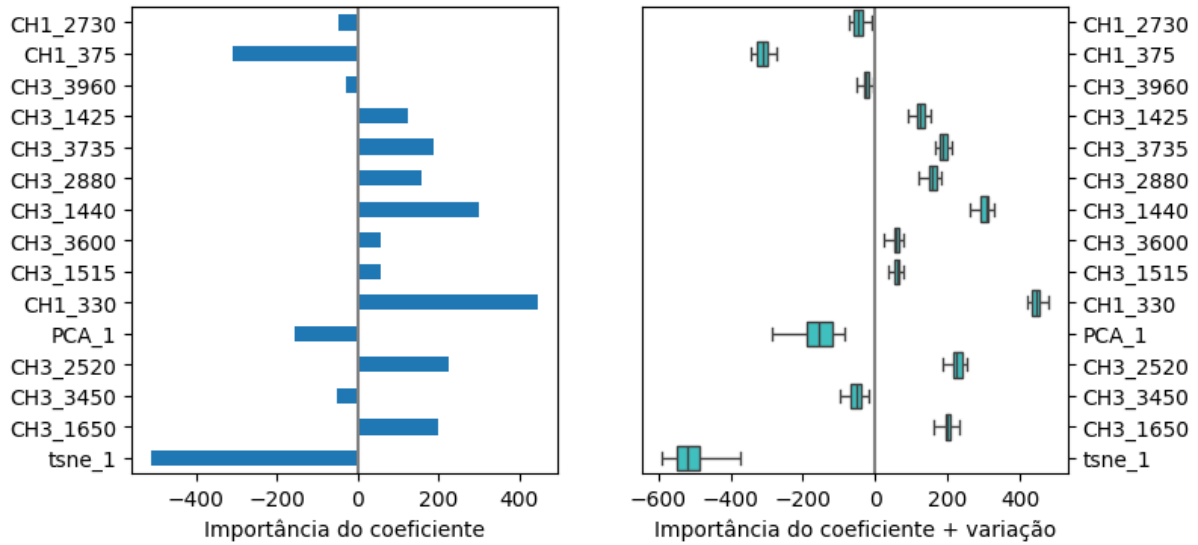
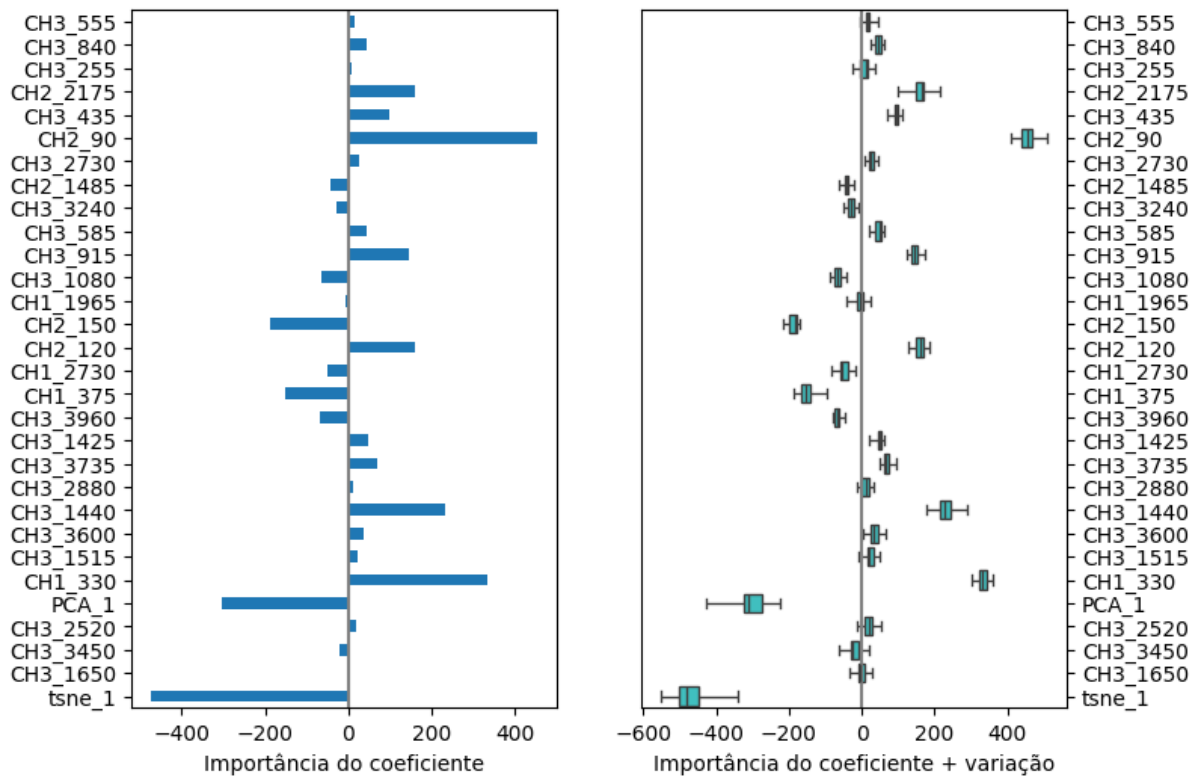
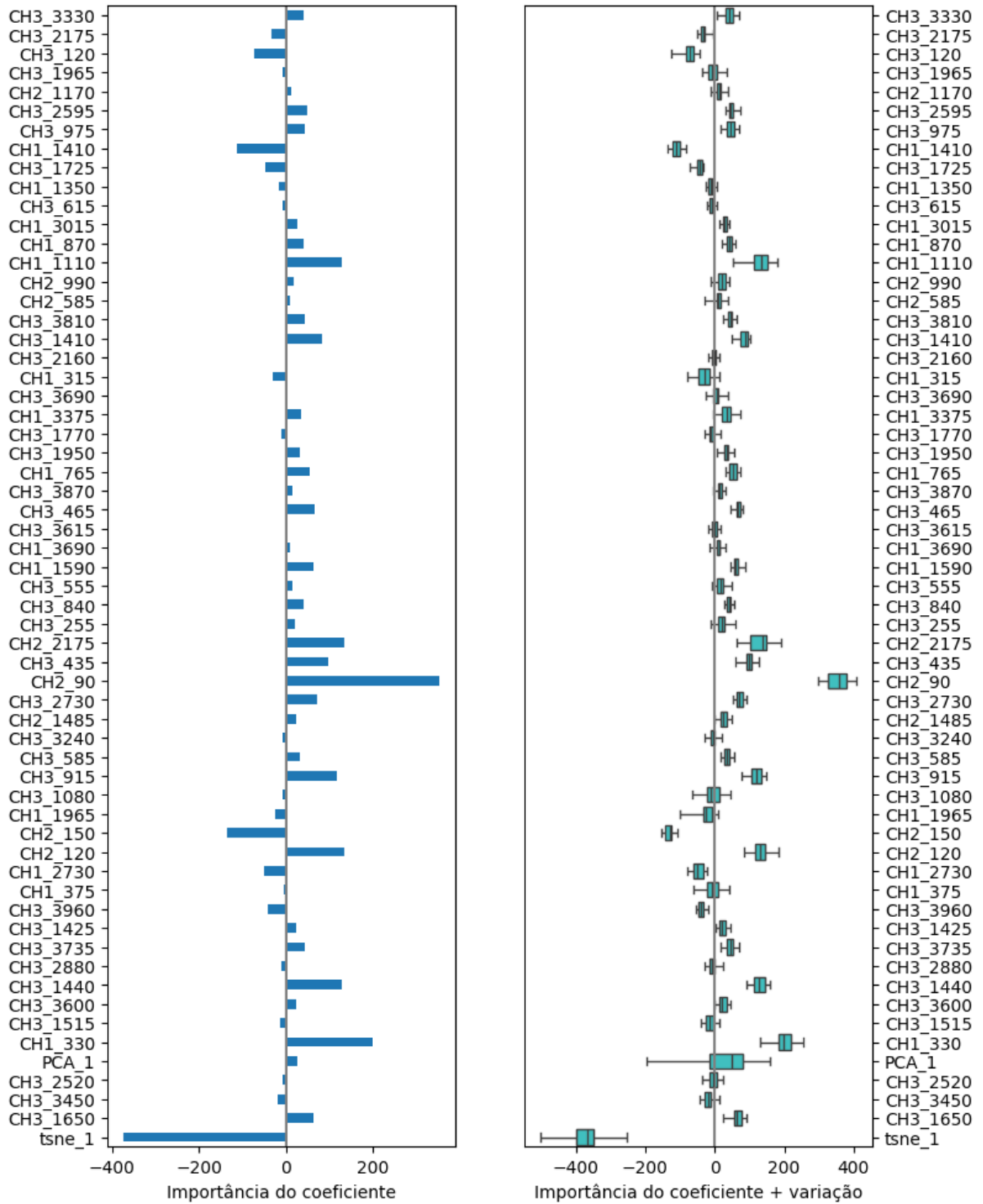


Figura 89 - Importância dos top 30 atributos ranqueados por correlação para a potência ativa.



Fonte: elaborado pelo autor.

Figura 90 - Importância dos top 60 atributos ranqueados por correlação para a potência ativa.



Fonte: elaborado pelo autor.

Figura 91 - Importância dos top 5 atributos ranqueados por correlação para a potência reativa.

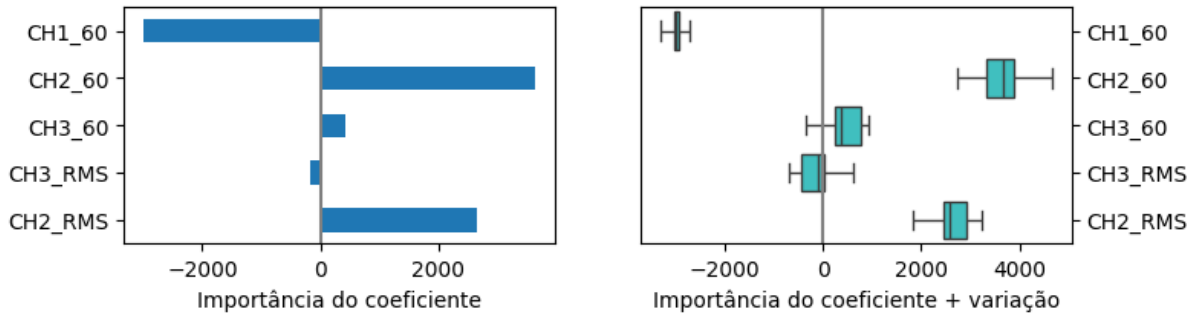


Figura 92 - Importância dos top 15 atributos ranqueados por correlação para a potência reativa.

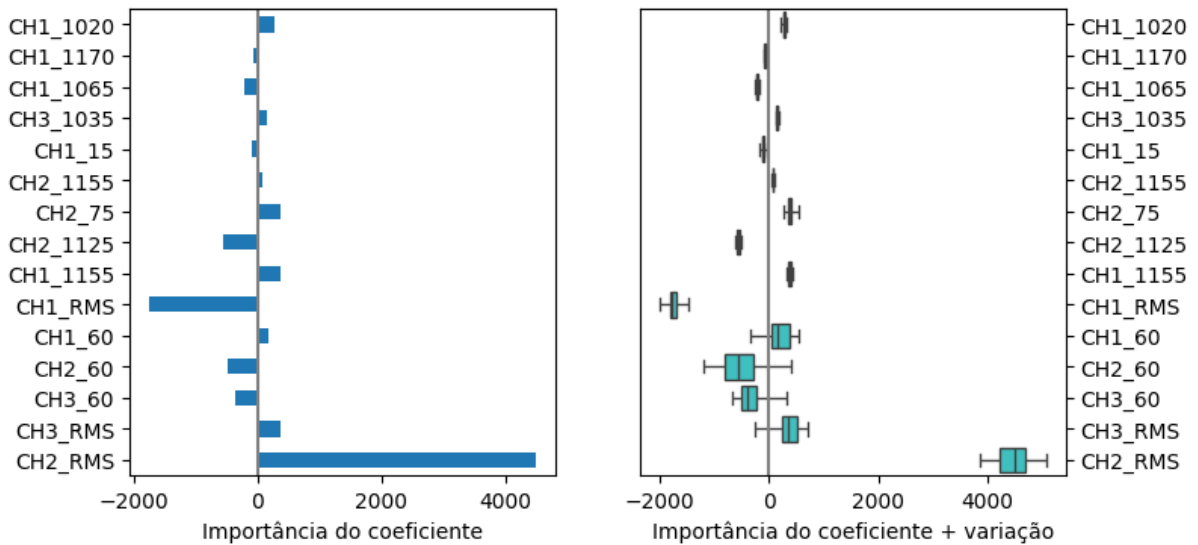
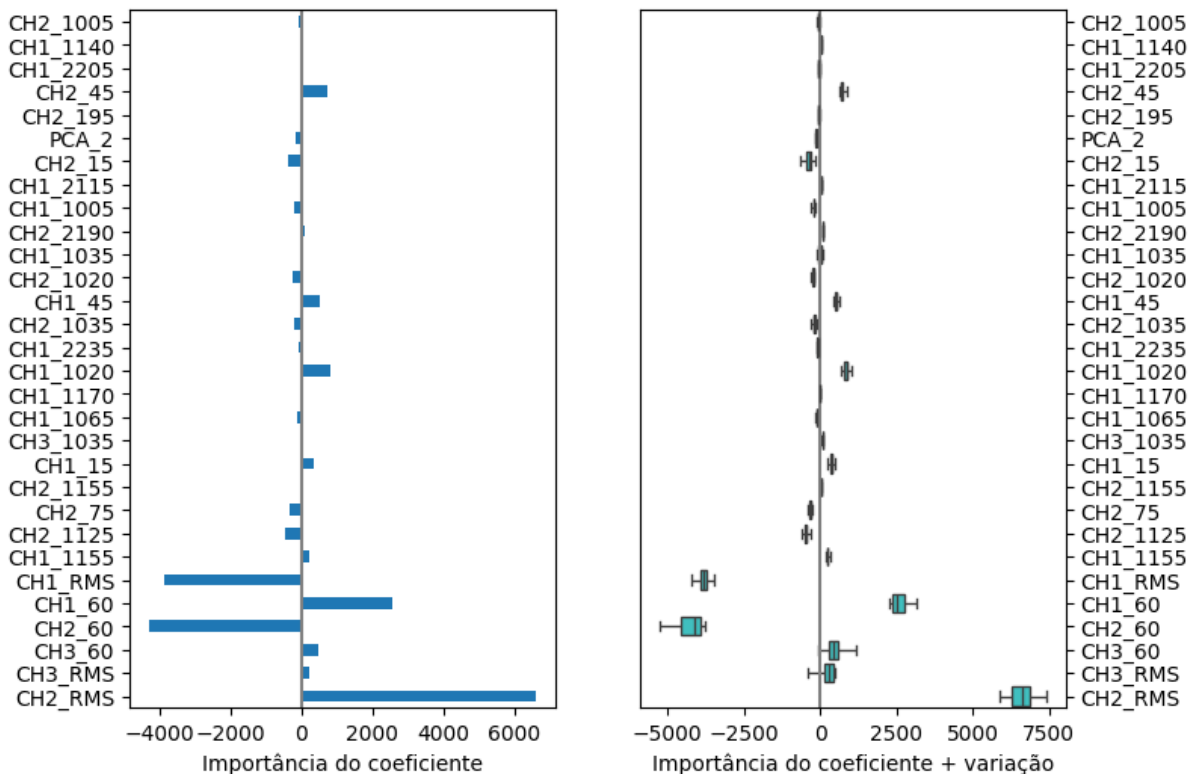
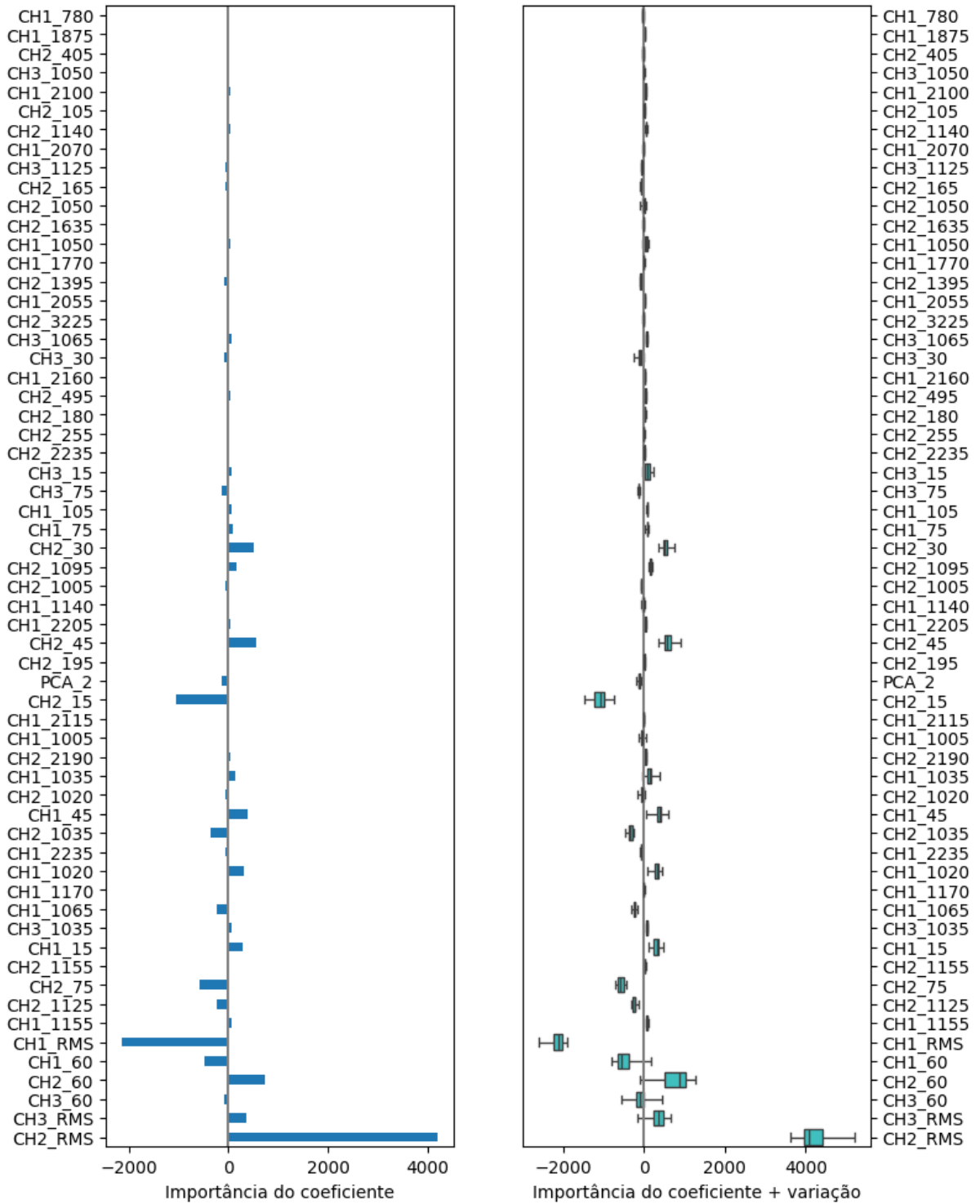


Figura 93 - Importância dos top 30 atributos ranqueados por correlação para a potência reativa.



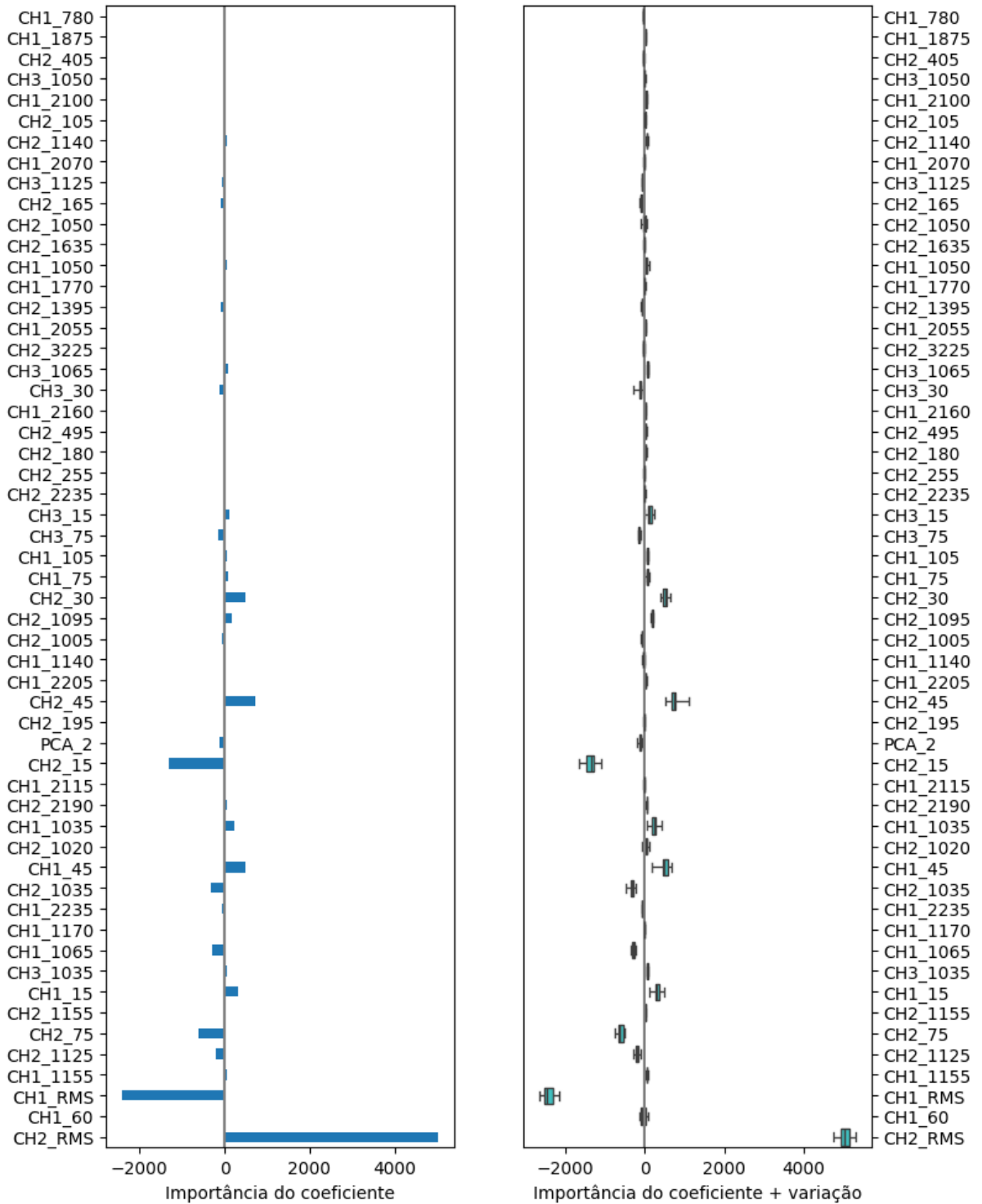
Fonte: elaborado pelo autor.

Figura 94 - Importância dos top 60 atributos ranqueados por correlação para a potência reativa.



Fonte: elaborado pelo autor.

Figura 95 - Importância dos top 60 atributos ranqueados por correlação para a potência reativa com atributos removidos.



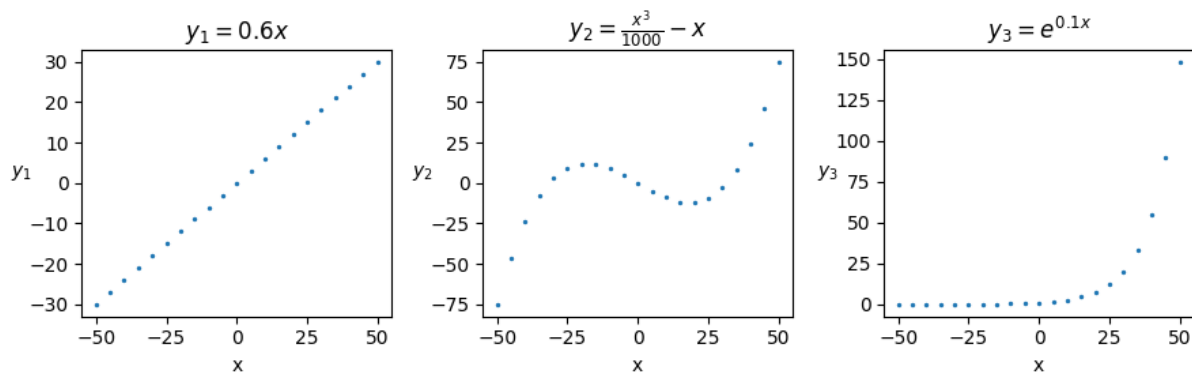
Fonte: elaborado pelo autor.

APÊNDICE D – Análise e aplicação dos cálculos de correlação

Neste apêndice é apresentado como são calculados os coeficientes de correlação de Pearson e Spearman. Também são apresentados resultados da aplicação da biblioteca PhiK sobre o mesmo conjunto de dados. Ao final, são apresentadas observações e análises sobre a aplicação de cada tipo de correlação.

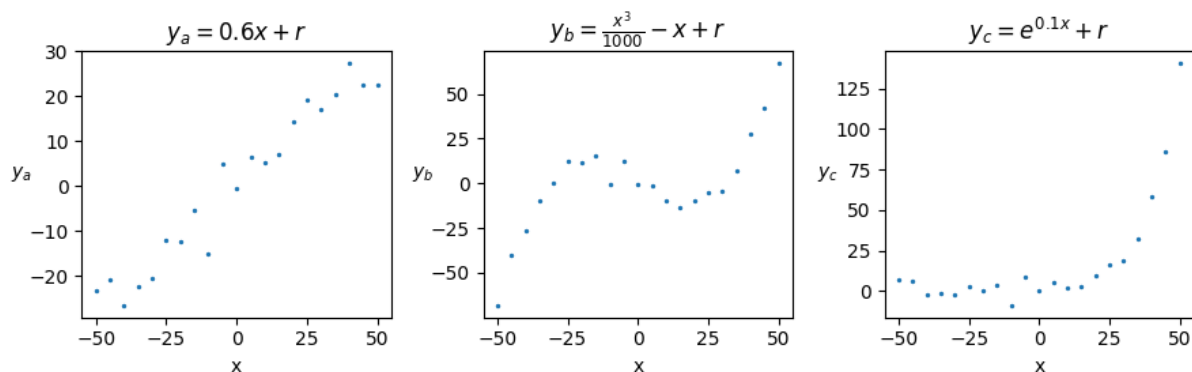
Inicialmente, o conjunto de dados genérico a ser utilizado é definido. Para exemplificar a aplicação dos cálculos de correlação, serão utilizadas aqui três funções com diferentes comportamentos: uma função linear de primeira ordem, uma função polinomial de terceira ordem e uma função exponencial que apresenta crescimento monotônico. As funções foram escolhidas arbitrariamente para fim de estudo. As três funções são apresentadas na Figura 96. A Figura 97 apresenta as mesmas três funções com adição de ruído à variável dependente. O ruído é amostrado de uma distribuição gaussiana de média zero e desvio padrão 5 e indicado nas equações da Figura 97 como r .

Figura 96 - Funções escolhidas para estudo de correlação.



Fonte: elaborado pelo autor.

Figura 97 - Funções escolhidas para estudo de correlação com adição de ruído.



Fonte: elaborado pelo autor.

A Tabela 22 apresenta os valores numéricos dos dados das figuras Figura 96 e Figura 97 em ordem crescente da variável independente x .

Tabela 22 - Valores numéricos dos dados para estudo de correlação.

Índice	x	y_1	y_2	y_3	y_a	y_b	y_c
1	-50,000	-30,000	-75,000	0,007	-23,201	-68,201	6,805
2	-45,000	-27,000	-46,125	0,011	-20,876	-40,001	6,135
3	-40,000	-24,000	-24,000	0,018	-26,552	-26,552	-2,533
4	-35,000	-21,000	-7,875	0,030	-22,490	-9,365	-1,460
5	-30,000	-18,000	3,000	0,050	-20,637	0,363	-2,587
6	-25,000	-15,000	9,375	0,082	-12,151	12,224	2,931
7	-20,000	-12,000	12,000	0,135	-12,280	11,720	-0,145
8	-15,000	-9,000	11,625	0,223	-5,266	15,359	3,958
9	-10,000	-6,000	9,000	0,368	-15,237	-0,237	-8,869
10	-5,000	-3,000	4,875	0,607	4,833	12,708	8,439
11	0,000	0,000	0,000	1,000	-0,482	-0,482	0,518
12	5,000	3,000	-4,875	1,649	6,402	-1,473	5,051
13	10,000	6,000	-9,000	2,718	5,317	-9,683	2,035
14	15,000	9,000	-11,625	4,482	7,105	-13,520	2,586
15	20,000	12,000	-12,000	7,389	14,316	-9,684	9,705
16	25,000	15,000	-9,375	12,182	19,123	-5,252	16,305
17	30,000	18,000	-3,000	20,086	16,987	-4,013	19,073
18	35,000	21,000	7,875	33,115	20,236	7,111	32,352
19	40,000	24,000	24,000	54,598	27,428	27,428	58,027
20	45,000	27,000	46,125	90,017	22,648	41,773	85,665
21	50,000	30,000	75,000	148,413	22,428	67,428	140,841

Fonte: elaborado pelo autor.

O coeficiente de correlação ρ de Pearson é calculado com a equação (4) sendo as equações (5) (covariância entre duas variáveis) e (6) (desvio padrão) como equações auxiliares.

Primeiramente, calculam-se os valores médios de cada variável dependente e independente, apresentados em (5) como \bar{x} e \bar{y} e o desvio padrão destas mesmas variáveis, apresentadas em (4) como σ_x e σ_y . Os valores médios de cada variável apresentada estão listados na Tabela 23. Os desvios padrão de cada variável apresentada estão listados na Tabela 24.

Tabela 23 - Valor médio das variáveis para cálculo de correlação de Pearson.

n	\bar{x}	\bar{y}_1	\bar{y}_2	\bar{y}_3	\bar{y}_a	\bar{y}_b	\bar{y}_c
21	0	0	0	17,96	0,363	0,363	18,32

Fonte: elaborado pelo autor.

Tabela 24 - Desvios padrão das variáveis para cálculo de correlação de Pearson.

σ_x	σ_{y_1}	σ_{y_2}	σ_{y_3}	σ_{y_a}	σ_{y_b}	σ_{y_c}
31,02	18,61	29,78	37,46	17,81	27,6	35,75

Fonte: elaborado pelo autor.

O cálculo da covariância entre a variável independente e as variáveis dependentes é apresentado em (24), (25), (26), (27), (28), e (29)

$$\begin{aligned} cov(x, y_1) &= \frac{\sum_{i=1}^n (x_{(i)} - \bar{x})(y_{1(i)} - \bar{y}_1)}{n} \\ &= \frac{\sum_{i=1}^{21} (x_{(i)} - 0)(y_{1(i)} - 0)}{21} = 550 \end{aligned} \quad (24)$$

$$cov(x, y_2) = \frac{\sum_{i=1}^{21} (x_{(i)} - 0)(y_{2(i)} - 0)}{21} = 591,25 \quad (25)$$

$$cov(x, y_3) = \frac{\sum_{i=1}^{21} (x_{(i)} - 0)(y_{3(i)} - 17,96)}{21} = 759,66 \quad (26)$$

$$cov(x, y_a) = \frac{\sum_{i=1}^{21} (x_{(i)} - 0)(y_{a(i)} - 0,363)}{21} = 511,24 \quad (27)$$

$$cov(x, y_b) = \frac{\sum_{i=1}^{21} (x_{(i)} - 0)(y_{b(i)} - 0,363)}{21} = 552,49 \quad (28)$$

$$cov(x, y_c) = \frac{\sum_{i=1}^{21} (x_{(i)} - 0)(y_{c(i)} - 18,32)}{21} = 720,9 \quad (29)$$

Os cálculos dos coeficientes de correlação ρ de Pearson entre a variável independente e as variáveis dependentes são apresentados em (30), (31), (32), (33), (34), e (35).

$$\rho_{x, y_1} = \frac{cov(x, y_1)}{\sigma_x \sigma_{y_1}} = \frac{550}{31,02 * 18,61} = 1,0 \quad (30)$$

$$\rho_{x,y_2} = \frac{cov(x, y_2)}{\sigma_x \sigma_{y_2}} = \frac{591,25}{31,02 * 29,78} = 0,672 \quad (31)$$

$$\rho_{x,y_3} = \frac{cov(x, y_3)}{\sigma_x \sigma_{y_3}} = \frac{759,66}{31,02 * 37,46} = 0,686 \quad (32)$$

$$\rho_{x,y_a} = \frac{cov(x, y_a)}{\sigma_x \sigma_{y_a}} = \frac{511,24}{31,02 * 17,81} = 0,971 \quad (33)$$

$$\rho_{x,y_b} = \frac{cov(x, y_b)}{\sigma_x \sigma_{y_b}} = \frac{552,49}{31,02 * 27,6} = 0,677 \quad (34)$$

$$\rho_{x,y_c} = \frac{cov(x, y_c)}{\sigma_x \sigma_{y_c}} = \frac{720,9}{31,02 * 35,75} = 0,682 \quad (35)$$

O coeficiente de correlação ρ de Spearman é calculado com a equação (7). Primeiramente, calculam-se os *ranks* das variáveis independente e dependentes. valores médios de cada variável dependente e independente. A diferença entre os *ranks* de duas variáveis, $d_i = (rank(x_i) - rank(y_i))$, é utilizada para calcular o coeficiente de correlação ρ de Spearman. Os *ranks* nada mais são do que o índice dos valores das variáveis quando ordenados do menor valor para o maior valor. As tabelas Tabela 25 apresenta os *ranks* das variáveis

Tabela 25 – Cálculo dos *ranks* das variáveis.

Índice	x	$rank(x_{(i)})$	y_1	$rank(y_{1(i)})$	y_2	$rank(y_{2(i)})$	y_3	$rank(y_{3(i)})$
1	-50	1	-30	1	-75,000	1	0,007	1
2	-45	2	-27	2	-46,125	2	0,011	2
3	-40	3	-24	3	-24,000	3	0,018	3
4	-35	4	-21	4	-7,875	8	0,030	4
5	-30	5	-18	5	3,000	12	0,050	5

6	-25	6	-15	6	9,375	16	0,082	6
7	-20	7	-12	7	12,000	18	0,135	7
8	-15	8	-9	8	11,625	17	0,223	8
9	-10	9	-6	9	9,000	15	0,368	9
10	-5	10	-3	10	4,875	13	0,607	10
11	0	11	0	11	0,000	11	1,000	11
12	5	12	3	12	-4,875	9	1,649	12
13	10	13	6	13	-9,000	7	2,718	13
14	15	14	9	14	-11,625	5	4,482	14
15	20	15	12	15	-12,000	4	7,389	15
16	25	16	15	16	-9,375	6	12,182	16
17	30	17	18	17	-3,000	10	20,086	17
18	35	18	21	18	7,875	14	33,115	18
19	40	19	24	19	24,000	19	54,598	19
20	45	20	27	20	46,125	20	90,017	20
21	50	21	30	21	75,000	21	148,413	21

Fonte: elaborado pelo autor.

Tabela 26 – Continuação do cálculo dos *ranks* das variáveis.

Índice	y_a	$rank(y_{a(i)})$	y_b	$rank(y_{b(i)})$	y_c	$rank(y_{c(i)})$
1	-23,201	2	-68,201	1	6,805	13
2	-20,876	4	-40,001	2	6,135	12
3	-26,552	1	-26,552	3	-2,533	3
4	-22,490	3	-9,365	7	-1,460	4
5	-20,637	5	0,363	13	-2,587	2
6	-12,151	8	12,224	16	2,931	9
7	-12,280	7	11,720	15	-0,145	5
8	-5,266	9	15,359	18	3,958	10
9	-15,237	6	-0,237	12	-8,869	1
10	4,833	11	12,708	17	8,439	14
11	-0,482	10	-0,482	11	0,518	6
12	6,402	13	-1,473	10	5,051	11
13	5,317	12	-9,683	6	2,035	7
14	7,105	14	-13,520	4	2,586	8
15	14,316	15	-9,684	5	9,705	15
16	19,123	17	-5,252	8	16,305	16
17	16,987	16	-4,013	9	19,073	17
18	20,236	18	7,111	14	32,352	18
19	27,428	21	27,428	19	58,027	19

20	22,648	20	41,773	20	85,665	20
21	22,428	19	67,428	21	140,841	21

Fonte: elaborado pelo autor.

O coeficiente de correlação ρ de Spearman pode ser calculado de duas maneiras: 1) utilizar os ranks das variáveis para calcular o coeficiente de correlação ρ de Pearson, isso resultará no coeficiente de correlação ρ de Spearman; 2) utilizar a equação (7) para calcular o coeficiente de correlação ρ de Spearman.

Para finalidade de estudo, os cálculos para os coeficientes de correlação ρ de Spearman entre as variáveis serão feitos pelos dois métodos.

A abordagem apresentada pelo método um consiste, inicialmente, em calcular a covariância entre a variável independente e uma variável dependente, conforme apresentado por (36). Nota-se que foram escolhidas para a demonstração apenas as variáveis x e y_1 . Tanto as médias dos *ranks* e os desvios padrão dos *ranks* de todas as variáveis são iguais a 11 e 6,20, respectivamente.

$$\begin{aligned}
 & cov(rank(x), rank(y_1)) \\
 &= \frac{\sum_{i=1}^n (rank(x_{(i)}) - \overline{rank(x)})(rank(y_{1(i)}) - \overline{rank(y_1)})}{n} \\
 &= \frac{\sum_{i=1}^{21} (rank(x_{(i)}) - 11)(rank(y_{1(i)}) - 11)}{21} = 36,66
 \end{aligned} \tag{36}$$

O cálculo do coeficiente de correlação ρ de Pearson utilizando os *ranks* da variável independente e da variável dependente é apresentado em (37). Os valores contidos abaixo da segunda casa decimal foram omitidos para simplicidade de notação.

$$\rho_{x,y_1} = \frac{cov(rank(x), rank(y_1))}{\sigma_x \sigma_y} = \frac{36,66}{6,20 * 6,20} = 1,0 \tag{37}$$

A abordagem apresentada pelo método dois é a aplicação direta da equação (7). Ao inserir os valores em (7), a equação resultante é apresentada em (38). Observa-se que sempre que os *ranks* das variáveis forem iguais o termo da soma será

zerado. Isto é um reflexo da correlação de Spearman indicar força da relação monotônica entre as variáveis.

$$\begin{aligned}\rho_{x,y_1} &= 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} = 1 - \frac{6 \sum_{i=1}^{21} (\text{rank}(x_{(i)}) - \text{rank}(y_{1(i)}))^2}{21(21^2 - 1)} \\ &= 1 - \frac{0}{21(21^2 - 1)} = 1,0\end{aligned}\quad (38)$$

Ao utilizar as variáveis x e y_2 para o cálculo da correlação de Spearman, espera-se obter um valor menor que 1,0 pois a curva de y_2 não apresenta relação monotônica com x , conforme observado na Figura 96. O procedimento de cálculo apresentado com os dois métodos para x e y_1 , também foi aplicado para x e y_2 . As equações (39), (40) demonstram o cálculo para o método 1 e (41) para o método 2.

$$\begin{aligned}& \text{cov}(\text{rank}(x), \text{rank}(y_2)) \\ &= \frac{\sum_{i=1}^n (\text{rank}(x_{(i)}) - \overline{\text{rank}(x)})(\text{rank}(y_{2(i)}) - \overline{\text{rank}(y_2)})}{n} \\ &= \frac{\sum_{i=1}^{21} (\text{rank}(x_{(i)}) - 11)(\text{rank}(y_{2(i)}) - 11)}{21} = 17,04\end{aligned}\quad (39)$$

$$\rho_{x,y_2} = \frac{\text{cov}(\text{rank}(x), \text{rank}(y_2))}{\sigma_x \sigma_y} = \frac{17,04}{6,20 * 6,20} = 0,465\quad (40)$$

$$\begin{aligned}\rho_{x,y_2} &= 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} = 1 - \frac{6 \sum_{i=1}^{21} (\text{rank}(x_{(i)}) - \text{rank}(y_{2(i)}))^2}{21(21^2 - 1)} \\ &= 1 - \frac{6 * 824}{21(21^2 - 1)} = 0,465\end{aligned}\quad (41)$$

O coeficiente de correlação PhiK (ϕ_K) [33] possui um cálculo não trivial. Em resumo, o valor do coeficiente é obtido após uma série de pré-processamentos para preparar o conjunto de dados:

- Montar uma tabela de contingência com o conjunto de dados. No caso de variáveis numéricas é preciso estratificar os dados para montar a tabela. A referência [33] sugere estratificar os dados igualmente em 10 intervalos.
- A tabela de contingência com as N observações contém r linhas e k colunas.
- Realizar o teste estatístico de contingência χ^2 pelo método de Pearson.
- Interpreta-se o resultado do teste de contingência χ^2 como sendo resultado de duas variáveis normalmente distribuídas.
 - Se $\chi^2 < \chi^2_{ped}$, onde χ^2_{ped} é calculado como um limite mínimo para relação entre variáveis, o coeficiente ρ é igual a zero.
 - Se $\chi^2 \geq \chi^2_{ped}$, com os valores N, r e k da tabela de contingência, deve-se inverter a função $\chi^2_{b.n.}(\rho, N, r, k)$ e resolver numericamente para ρ para o intervalo [0,1].
- A solução encontrada para ρ é o coeficiente ϕ_K .

Aqui é apresentado apenas um resumo de como proceder para calcular o coeficiente de correlação ϕ_K . Para maiores detalhes, deve-se consultar a referência [33], onde é apresentado o procedimento de cálculo adotado e como calcular cada variável ou coeficiente apresentado neste resumo.

Felizmente, os mesmos autores da referência [33] também disponibilizam uma biblioteca em Python que implementa toda a rotina de cálculo do coeficiente de correlação ϕ_K .

Por motivos de simplicidade, utilizou-se apenas a biblioteca em Python com a implementação da rotina de cálculo para ϕ_K .

De maneira a resumir os resultados dos coeficientes de correlação, são apresentados nas figuras Figura 98 e Figura 99 os três coeficientes de correlação: ρ_p para a correlação de Pearson, ρ_s para a correlação de Spearman e ϕ_K para a correlação PhiK utilizadas nas funções apresentadas nas figuras Figura 96 e Figura 97. Foi necessário utilizar um número maior de amostras para que $\chi^2 \geq \chi^2_{ped}$ dentro do cálculo da correlação de ϕ_K . Foram utilizadas 201 amostras para o cálculo dos coeficientes de correlação

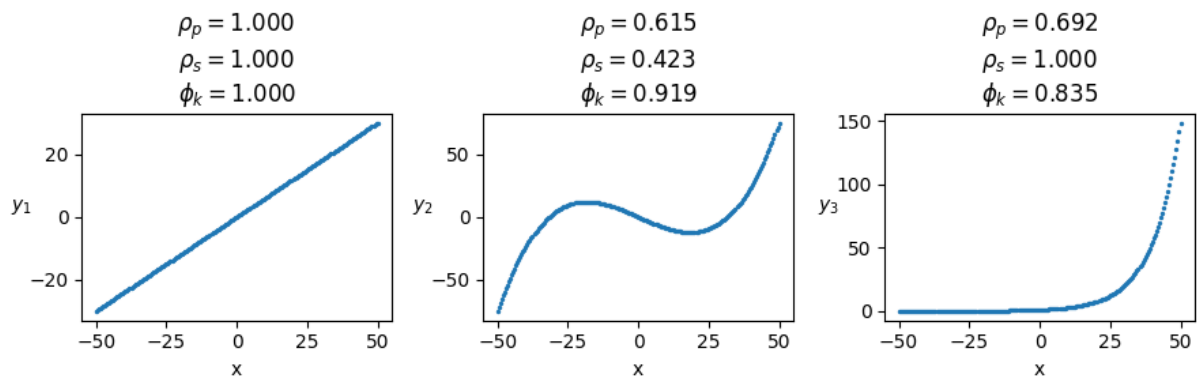
É possível observar nas figuras Figura 98 e Figura 99 que a adição de ruído impacta o valor dos coeficientes de correlação. Para a função linear (y_1 e y_a), todas

as correlações apresentam valor elevado, o que indica a existência de uma forte relação destas variáveis com a variável independente x .

Para a função polinomial de terceiro grau (y_2 e y_b), que não é linear, a correlação de Pearson e Spearman apresentam baixos valores. No entanto o coeficiente ϕ_K consegue capturar a relação dos dados apresentando um alto valor de correlação. Isto é um indicativo que o coeficiente ϕ_K consegue capturar as relações não lineares entre variáveis.

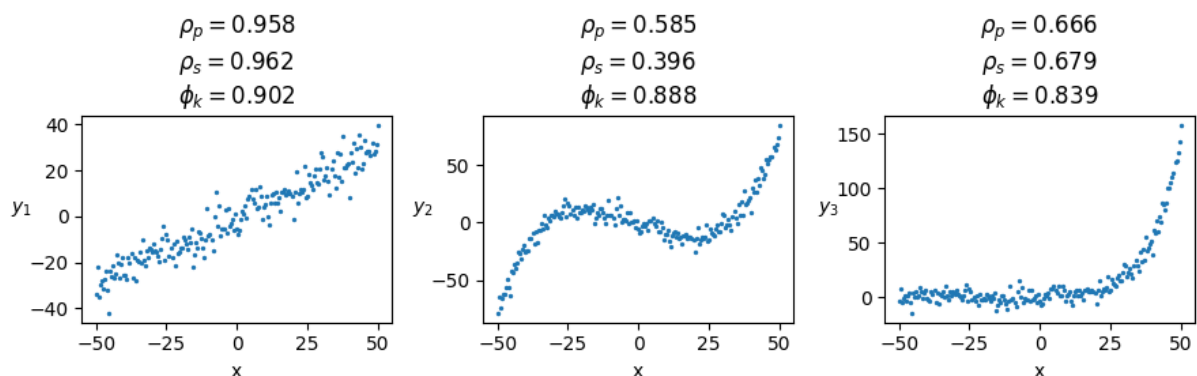
Para a função exponencial (y_3 e y_c), o coeficiente de Spearman apresenta o maior valor por capturar a relação monotônica e não linear de y_3 . Para y_c , a correlação de Spearman não consegue capturar a relação monotônica e não linear com tanta eficácia. Isto pode ser motivado pelo fato de que a maior parte da função exponencial está abaixo do nível d ruído somado à variável dependente, conforme observado na Figura 99 para os valores de $-50 < x < 25$. Em todo caso, o coeficiente ϕ_K também consegue capturar as relações não lineares entre variáveis com um resultado muito próximo do caso sem ruído.

Figura 98 - Coeficientes de correlação para funções sem adição de ruído.



Fonte: elaborado pelo autor.

Figura 99 - Coeficientes de correlação para funções com adição de ruído.



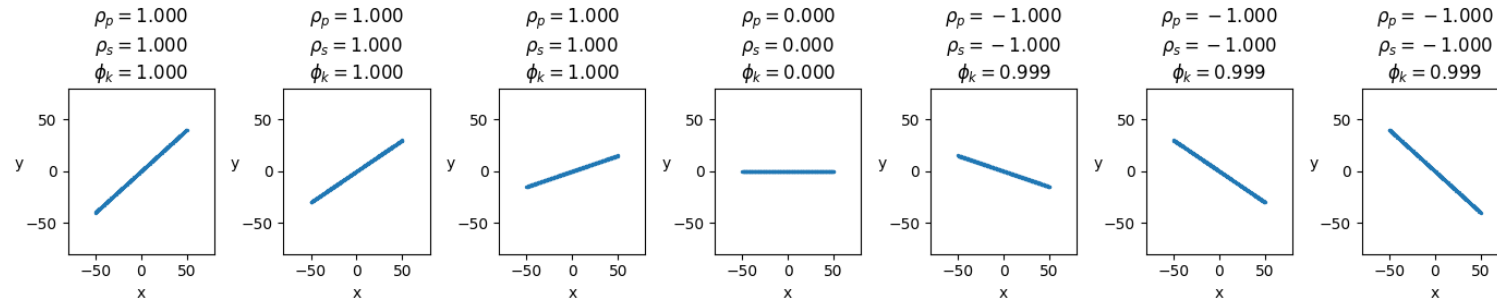
Fonte: elaborado pelo autor.

Os resultados observados nas figuras Figura 98 e Figura 99 são úteis para ter um primeiro contato com os coeficientes. Com a finalidade de entender o comportamento dos três tipos de correlação, um conjunto de diferentes funções com diferentes características foi avaliado e é apresentado na Figura 100. Em todos os casos foi adicionado ruído à variável dependente e/ou independente.

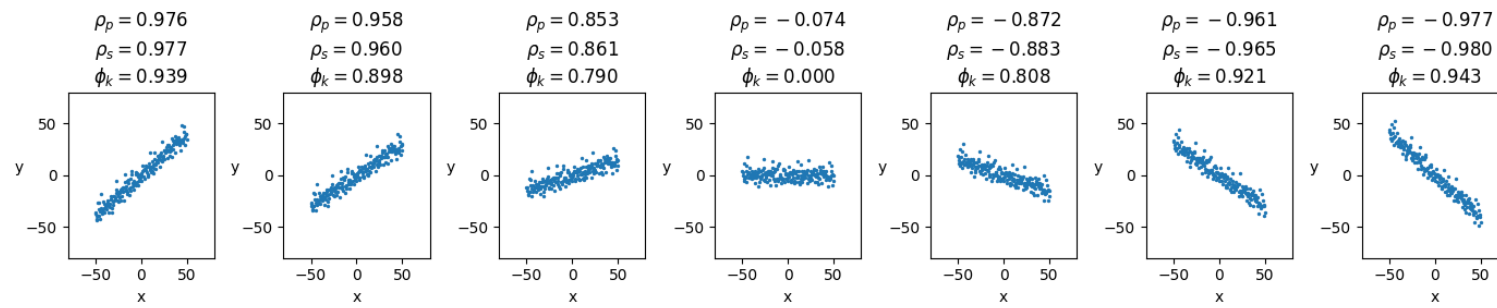
Comentários sobre a Figura 100:

1. A variável independente (x) é a mesma para todas as imagens.
2. O ruído adicionado é amostrado de uma distribuição gaussiana de média zero e desvio padrão 5.
3. O ruído foi adicionado às funções de três diferentes maneiras: i) somente adição de ruído na variável dependente, ii) somente adição de ruído na variável independente, iii) adição de ruído nas variáveis dependente e independente.
4. Ao todo, são apresentadas 14 funções. Sete destas funções são lineares com diferentes coeficientes angulares. As outras sete funções são funções não-lineares com e sem relação monotônica.
5. Os coeficientes de correlação estão indicados como: ρ_p - Correlação de Pearson, ρ_s - Correlação de Spearman, ϕ_K - Correlação PhiK.
6. As imagens da Figura 100 são apresentadas em ordem alfabética:
 - a. Funções lineares (que também são monotônicas) sem adição de ruído.
 - b. Funções lineares com adição de ruído à variável dependente (y).
 - c. Funções lineares com adição de ruído à variável independente (x).
 - d. Funções lineares com adição de ruído à variável dependente (y) e independente (x).
 - e. Funções não-lineares que podem ou não ser monotônicas sem adição de ruído.
 - f. Funções não-lineares com adição de ruído à variável dependente (y).
 - g. Funções não-lineares com adição de ruído à variável independente (x).
 - h. Funções não-lineares com adição de ruído à variável dependente (y) e independente (x).

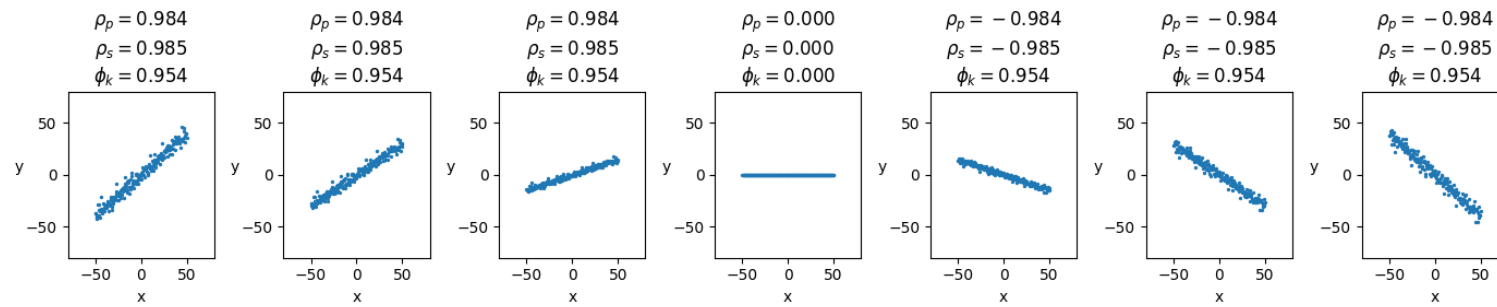
Figura 100 - Cálculo dos coeficientes de correlação para diferentes funções.



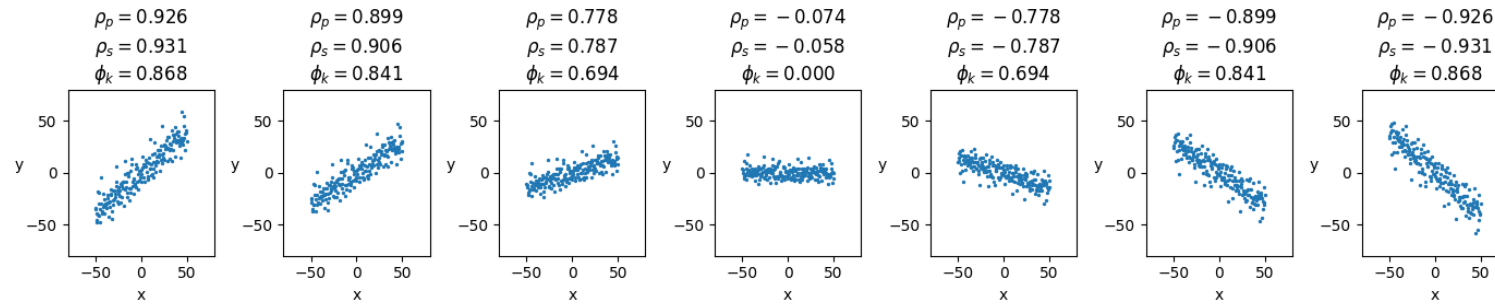
(a)



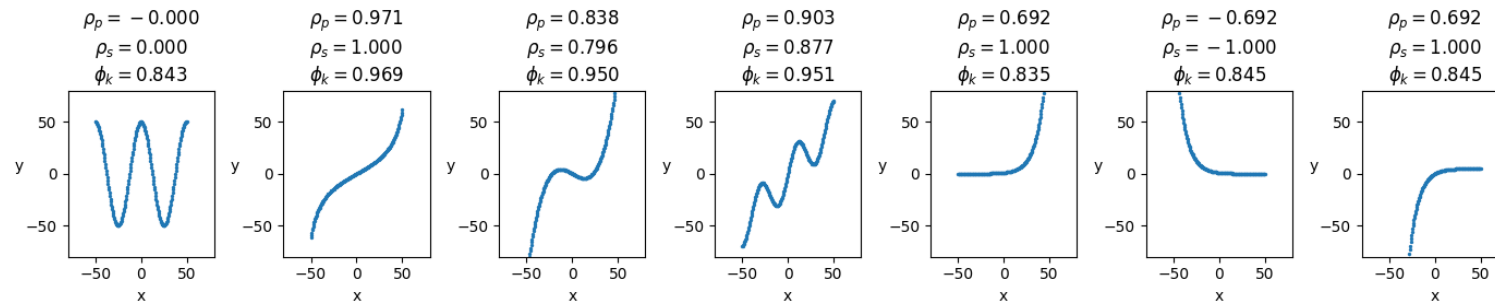
(b)



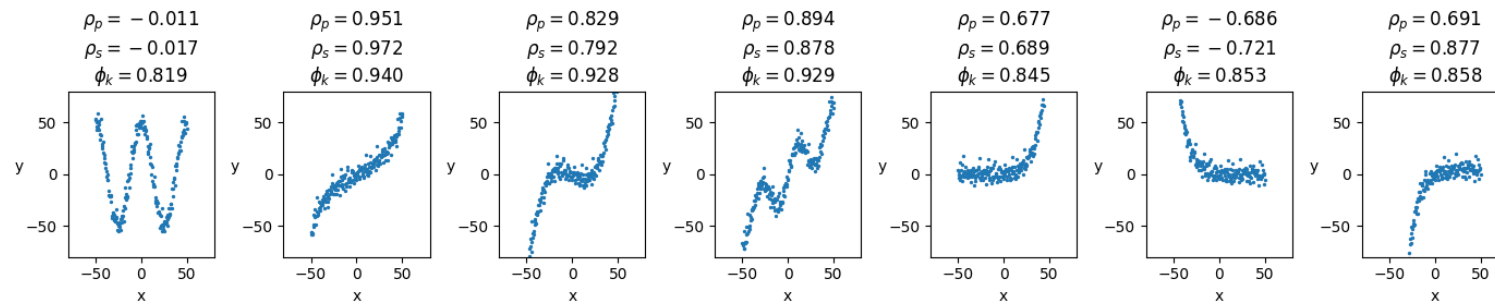
(c)



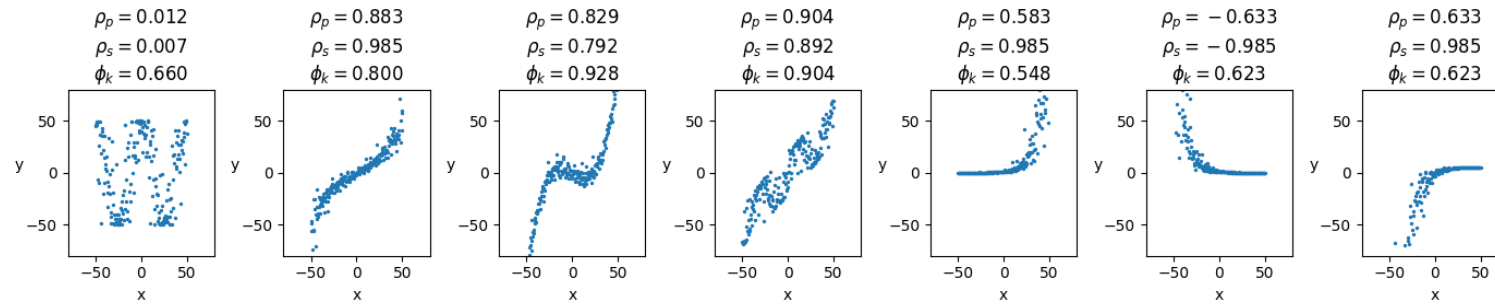
(d)



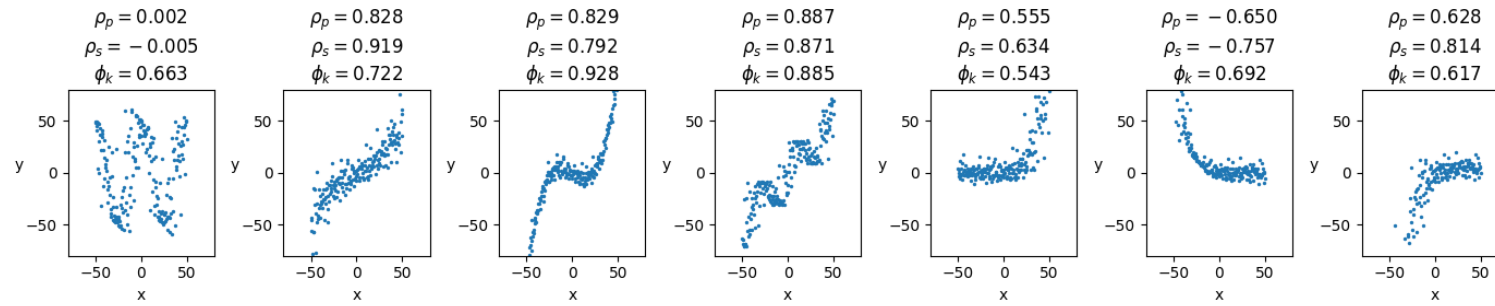
(e)



(f)



(g)



(h)

Uma função linear sempre é monotônica pois sempre cresce ou decresce na mesma direção e seus *ranks* são sempre alinhados.

Uma função monotônica pode ou não ser linear, como é o exemplo das funções exponenciais e trigonométrica ($\tan(x)$) apresentadas na Figura 100 (e).

Uma função não-linear pode ou não ser monotônica.

Com essas observações, pode-se notar na Figura 100 que os casos de relação linear (Figura 100 (a), (b), (c) e (d)) sempre apresentam altos coeficientes ρ_p , ρ_s e ϕ_K . Os três coeficientes conseguem capturar a relação de linearidade entre as variáveis.

Para o caso da Figura 100 (e), (f), (g) e (h), que apresentam relações não-lineares entre as variáveis, o coeficiente ϕ_K sempre apresenta um valor elevado. Fato que indica a existência de uma intensa relação entre as variáveis, sendo esta relação linear ou não.

Para os casos exponenciais da Figura 100 (e), o coeficiente ρ_s atinge seu valor máximo (1) para as duas funções exponenciais que crescem com o aumento da variável independente (x) e mínimo para a função exponencial que decresce com o aumento da variável independente (x). Este é um indicativo da relação monotônica positiva e negativa respectivamente.

Para o caso trigonométrico que apresenta dois ciclos de um cosseno na Figura 100 (e), ambos os coeficientes ρ_p e ρ_s não indicam relação entre as variáveis, porém, o coeficiente ϕ_K indica alta relação entre as variáveis. Neste caso, a interpretação é que existe uma relação não-linear entre as variáveis.

Ao observar e analisar os gráficos da Figura 100, é possível generalizar a análise de correlação com alguns pontos:

1. O coeficiente de correlação ρ_p indica a intensidade da relação linear entre variáveis.
2. O coeficiente de correlação ρ_s indica a intensidade da relação monotônica entre variáveis.
3. O coeficiente de correlação ϕ_K indica a intensidade da relação não-linear entre variáveis, podendo também indicar a intensidade da relação linear entre variáveis.