

**UNIVERSIDADE FEDERAL DE SANTA CATARINA
CAMPUS FLORIANÓPOLIS**

Vitor Pohlenz

**APLICAÇÃO DE MACHINE LEARNING E FERRAMENTAS
ESTATÍSTICAS PARA MODELAGEM E ANÁLISE DE
FALHAS EM AEROGERADORES**

FLORIANÓPOLIS

2020

Vitor Pohlenz

**Aplicação de Machine Learning e Ferramentas
Estatísticas para Modelagem e Análise de Falhas em
Aerogeradores**

Trabalho de Conclusão de Curso apresentado à Universidade Federal de Santa Catarina como parte dos requisitos necessários para a obtenção do Título de Engenheiro Eletricista.
Orientador: Prof. Edson Cilos Vargas Júnior, Dr.
Coorientador: Prof. Eduardo Luiz Ortiz Batista, Dr.

Universidade Federal de Santa Catarina
Trindade, Florianópolis
Departamento de Engenharia Elétrica

Florianópolis
2020

Ficha Catalográfica

Vitor Pohlenz

Aplicação de Machine Learning e Ferramentas Estatísticas para Modelagem e Análise de Falhas em Aerogeradores - Florianópolis, 2020 - 108 p., 30 cm.

Orientador: Prof. Edson Cilos Vargas Júnior, Dr.

Coorientador: Prof. Eduardo Luiz Ortiz Batista, Dr.

1. Aerogeradores. 2. Operação e Manutenção. 3. *Machine Learning*. 4. *Digital Twin*

I. Universidade Federal de Santa Catarina. Engenharia Elétrica. II. Aplicação de Machine Learning e Ferramentas Estatísticas para Modelagem e Análise de Falhas em Aerogeradores.

Vitor Pohlenz

Aplicação de Machine Learning e Ferramentas Estatísticas para Modelagem e Análise de Falhas em Aerogeradores

Este trabalho foi julgado adequado como parte dos requisitos para obtenção do Título de Bacharel em Engenharia Elétrica e aprovado, em sua forma final, pela Banca Examinadora

Florianópolis, 10 de julho de 2020



Documento assinado digitalmente
Jean Viane Leite
Data: 13/07/2020 11:50:26-0300
CPF: 003.474.909-80

Prof. Jean Viane Leite, Dr.
Coordenador do Curso de Graduação em
Engenharia Elétrica

Banca Examinadora



Documento assinado digitalmente
Edson Cilos Vargas Junior
Data: 10/07/2020 16:10:10-0300
CPF: 088.267.299-13

Prof. Edson Cilos Vargas Júnior, Dr.
Universidade Federal de Santa Catarina
Orientador



Documento assinado digitalmente
Eduardo Luiz Ortiz Batista
Data: 10/07/2020 16:28:39-0300
CPF: 036.521.889-85

Prof. Eduardo Luiz Ortiz Batista, Dr.
Universidade Federal de Santa Catarina
Coorientador



Documento assinado digitalmente
Andreia Zanella
Data: 10/07/2020 16:22:00-0300
CPF: 822.746.650-87

Profa. Andreia Zanella, Dra.
Universidade Federal de Santa Catarina



Documento assinado digitalmente
Richard Demo Souza
Data: 10/07/2020 16:27:21-0300
CPF: 004.267.379-89

Prof. Richard Demo Souza, Dr.
Universidade Federal de Santa Catarina

Dedico este trabalho à todos aqueles que, de alguma forma, contribuíram para a minha formação pessoal e profissional.

Agradecimentos

Agradeço primeiramente à minha família por todo o apoio que recebi durante minha formação acadêmica.

Agradeço especialmente à minha namorada Allana Batista, pelas sugestões de escrita no presente trabalho, sendo uma conselheira editora atenciosa e paciente em todos os meus pedidos de revisão.

Agradeço também ao meu Orientador Edson Cilos Vargas Júnior e meu Coorientador Eduardo Luiz Ortiz Batista, que aceitaram me ajudar nesse desafio, e dedicaram tempo de suas vidas atarefadas para ajudar a confeccionar o presente trabalho. E aos meus amigos, que fossem assuntos acadêmicos ou não, sempre estiveram presentes e me acompanharam em todas as etapas da graduação.

Por fim agradeço à equipe da Empresa Aqtech, em especial ao Diretor Técnico Tiago Kaoru Matsuo e ao Engenheiro Mecânico Marcos Nishioka, os quais sempre foram muito solícitos e atenciosos nas explicações e instruções à respeito do funcionamento de usinas eólicas.

“(...)acredita(-se) que apenas um grande poder pode manter o mal sobre controle, mas não é o que descobri. Descobri que são as pequenas coisas, as tarefas diárias de pessoas comuns que mantêm o mal afastado, simples ações de bondade e amor.”
(Gandalf - O Hobbit)

“O sucesso é ir de fracasso em fracasso sem perder entusiasmo.”
(Winston Churchill)

Resumo

Atualmente a energia eólica demonstra significativa expressividade na matriz elétrica brasileira. Esse tipo de geração de energia exige, todavia, uma grande quantidade de máquinas em operação, uma vez que cada unidade de geração eólica (aerogerador) possui uma capacidade relativamente baixa de geração em comparação, por exemplo, com unidades de geração hidrelétricas. Dessa forma, o grande número de equipamentos gera um desafio para equipes de operação e manutenção dos parques eólicos. Nesse contexto, o presente trabalho tem como foco central a utilização de uma metodologia baseada em ferramentas matemáticas e estatísticas, além de modelos de *Machine Learning*, para modelar o comportamento de componentes dos aerogeradores, bem como estabelecer limites de normalidade para as curvas características (Curvas S) de cada aerogerador. Assim, busca-se desenvolver uma ferramenta de auxílio na identificação de problemas ou comportamentos anormais, ajudando dessa forma as equipes de operação e manutenção no aumento da vida útil e minimização de custos de reparos dos aerogeradores.

Palavras-Chave: 1. Aerogeradores. 2. Operação e Manutenção. 3. *Machine Learning*. 4. *Digital Twin*

Abstract

Nowadays, wind power has become an expressive component of the Brazilian energy matrix. This type of energy source requires a large number of generation units, since typically the generation capacity of an individual wind power generator is relatively lower, especially in comparison with hydro power generators. Thus, the large numbers of machines in operation poses an important challenge for the operation and maintenance teams of the so-called wind farms. In this context, the current work is focused at using a methodology based on statistical and mathematical tools, together with Machine Learning methods, to model the behavior of certain components of the wind generator, as well as establish normality boundaries for its characteristic curve (the S Curve). The final aim is to develop tools for helping the identification of problems and abnormal behavior, which can be of great importance for aiding operation and maintenance teams on increasing the service life of wind generators as well as on minimizing maintenance costs.

Keywords: 1. Wind Generator. 2. operation and maintenance. 3. Machine Learning. 4. Digital Twin

Lista de Figuras

Figura 1 – Disposição da Capacidade Instalada Total Brasileira em 2019 por tipo de fonte, sem incluir a parcela da UHE Itaipu pertencente ao Paraguai.	18
Figura 2 – Desenho esquemático de uma turbina eólica moderna	19
Figura 3 – Representação esquemática típica do desenvolvimento de uma falha mecânica.	23
Figura 4 – Representação esquemática de uma Curva S típica, com suas principais regiões e valores de interesse.	26
Figura 5 – Representação de um <i>Box Plot</i> padrão, onde o primeiro, o segundo e o terceiro quartis mostram onde se encontram 25%, 50% e 75% dos dados. Com a Distância Interquartil (DIQ), é possível ter ideia da dispersão dos dados, e os pontos em azul que estão à $1,5 \cdot \text{DIQ}$ acima do 3º Quartil ou abaixo do 1º Quartil são considerados <i>outliers</i>	30
Figura 6 – Representação de uma função densidade de probabilidade de uma Distribuição Normal $X \sim N(\mu, \sigma^2)$, onde μ é a média e σ é o desvio padrão.	31
Figura 7 – Representação ilustrativa de uma <i>Pipeline</i> simples, contendo os conjuntos de treinamento, teste e validação, o pré processamento, o treinamento dos modelos e sua predição.	34
Figura 8 – Representação esquemática da curva PF	41
Figura 9 – Distribuição dos resíduos das temperaturas de todos os componentes modelados para um aerogerador	46
Figura 10 – Dados medidos e curvas de potência máxima e mínima fornecida pelo fabricante em relação à Curva S para um aerogerador.	47
Figura 11 – Limiares de filtragem de um aerogerador. Indicando os dados brutos, os valores superiores e inferiores do método de filtragem automático (f_{sup} e f_{inf}), os valores de potência acima de zero após a velocidade de cut_{in} mais 1.5 vezes W_{iq} e os valores anteriores a velocidade de cut_{off} mais 1.5 vezes W_{iq} . Sendo W_{iq} a mediana da distância entre os percentis 20% e 80%.	51
Figura 12 – Dados Filtrados, após o procedimento descrito na Subseção 4.3.1.5 para um aerogerador	52
Figura 13 – Limites para Classificação de <i>Outliers</i> e todos os dados (Dados Brutos) de um aerogerador.	54
Figura 14 – Curva S do Aerogerador 001 com o limite inferior (l_{inf}) e superior (l_{sup}).	56
Figura 15 – Pontos da Curva S do Aerogerador 001 entre os limite inferior (l_{inf}) e superior (l_{sup}) para os anos de 2018, 2019 e 2020.	57

Figura 16 – Aero gerador 001. Visualização temporal dos pontos aglomerados entre 1000 kW e 1500 kW nas datas de 12/02/2019 e 05/04/2019 e registro de <i>Log</i> de manutenção preventiva anual em 09/04/2019.	57
Figura 17 – Curva S do Aero gerador 002 com o limite inferior (l_{inf}) e superior (l_{sup}).	58
Figura 18 – Aero gerador 002. Visualização temporal dos pontos aglomerados entre 1000 kW e 1500 kW nas datas de 12/02/2019 e 05/04/2019 e registro de <i>Log</i> de manutenção preventiva anual em 15/04/2019.	59
Figura 19 – Curva S do Aero gerador 003 com o limite inferior (l_{inf}) e superior (l_{sup}).	60
Figura 20 – Aero gerador 003. Visualização temporal dos pontos aglomerados entre 1000 kW e 1500 kW nas datas de 12/02/2019, 05/04/2019 e 16/04/2019 e registro de <i>Log</i> de manutenção preventiva anual em 16/04/2019.	61
Figura 21 – Curva S do Aero gerador 004 com o limite inferior (l_{inf}) e superior (l_{sup}).	62
Figura 22 – Aero gerador 004. Visualização temporal dos pontos aglomerados entre 1000 kW e 1500 kW nas datas de 12/02/2019, 05/04/2019 e 16/04/2019 e registro de <i>Log</i> de manutenção preventiva anual em 17/04/2019.	63
Figura 23 – Curva S do Aero gerador 005 com o limite inferior (l_{inf}) e superior (l_{sup}).	64
Figura 24 – Aero gerador 005. Visualização temporal dos pontos aglomerados entre 1000 kW e 1500 kW nas datas de 12/02/2019, 05/04/2019 e 16/04/2019 e registro de <i>Log</i> de manutenção preventiva anual em 18/04/2019.	65
Figura 25 – Curva S do Aero gerador 006 com o limite inferior (l_{inf}) e superior (l_{sup}).	66
Figura 26 – Aero gerador 006. Visualização temporal dos pontos aglomerados entre 1000 kW e 1500 kW nas datas de 12/02/2019, 05/04/2019 e 16/04/2019 e registro de <i>Log</i> de manutenção preventiva anual em 16/04/2019. Nessa figura foram retirados as observações indicando restrições nas potências de 1000 kW e 1200 kW para melhor visualização dos aglomerados.	67
Figura 27 – Aero gerador 006. Visualização temporal das restrições em 500kW, 1000 kW e 1200 kW, e os registros de <i>Logs</i> em 26/01/2018, 07/12/2018, 16/05/2019 e 03/06/2019.	68
Figura 28 – Curva S do Aero gerador 007 com o limite inferior (l_{inf}) e superior (l_{sup}).	69
Figura 29 – Aero gerador 007. Visualização temporal dos pontos aglomerados entre 1000 kW e 1500 kW nas datas de 12/02/2019, 05/04/2019 e 16/04/2019 e registro de <i>Log</i> de manutenção preventiva anual em 17/04/2019.	70
Figura 30 – Curva S do Aero gerador 008 com o limite inferior (l_{inf}) e superior (l_{sup}).	71
Figura 31 – Aero gerador 008. Visualização temporal dos pontos aglomerados entre 1000 kW e 1500 kW nas datas de 12/02/2019, 05/04/2019 e 16/04/2019 e registro de <i>Log</i> de manutenção preventiva anual em 18/04/2019.	72
Figura 32 – Distribuição dos resíduos do Aero gerador 001.	73

Figura 33 – Curvas com <i>outliers</i> acumulados para o grupo Mancal do Aero gerador 001.	74
Figura 34 – Curvas com <i>outliers</i> acumulados para o grupo <i>Gear Box</i> do Aero gerador 001.	75
Figura 35 – Curvas com <i>outliers</i> acumulados para o grupo Gerador do Aero gerador 001.	76
Figura 36 – Curvas com <i>outliers</i> acumulados para o grupo <i>Pitch</i> do Aero gerador 001.	77
Figura 37 – Distribuição dos resíduos do Aero gerador 002.	78
Figura 38 – Curvas com <i>outliers</i> acumulados para o grupo Mancal do Aero gerador 002.	79
Figura 39 – Curvas com <i>outliers</i> acumulados para o grupo <i>Gear Box</i> do Aero gerador 002.	79
Figura 40 – Curvas com <i>outliers</i> acumulados para o grupo Gerador do Aero gerador 002.	80
Figura 41 – Curvas com <i>outliers</i> acumulados para o grupo <i>Pitch</i> do Aero gerador 002.	81
Figura 42 – Distribuição dos resíduos do Aero gerador 003.	82
Figura 43 – Curvas com <i>outliers</i> acumulados para o grupo Mancal do Aero gerador 003.	83
Figura 44 – Curvas com <i>outliers</i> acumulados para o grupo <i>Gear Box</i> do Aero gerador 003.	84
Figura 45 – Curvas com <i>outliers</i> acumulados para o grupo Gerador do Aero gerador 003.	84
Figura 46 – Curvas com <i>outliers</i> acumulados para o grupo <i>Pitch</i> do Aero gerador 003.	85
Figura 47 – Distribuição dos resíduos do Aero gerador 004.	86
Figura 48 – Curvas com <i>outliers</i> acumulados para o grupo Mancal do Aero gerador 004.	87
Figura 49 – Curvas com <i>outliers</i> acumulados para o grupo <i>Gear Box</i> do Aero gerador 004.	88
Figura 50 – Curvas com <i>outliers</i> acumulados para o grupo Gerador do Aero gerador 004.	88
Figura 51 – Curvas com <i>outliers</i> acumulados para o grupo <i>Pitch</i> do Aero gerador 004.	89
Figura 52 – Distribuição dos resíduos do Aero gerador 005.	90
Figura 53 – Curvas com <i>outliers</i> acumulados para o grupo Mancal do Aero gerador 005.	91
Figura 54 – Curvas com <i>outliers</i> acumulados para o grupo <i>Gear Box</i> do Aero gerador 005.	92
Figura 55 – Curvas com <i>outliers</i> acumulados para o grupo Gerador do Aero gerador 005.	92
Figura 56 – Curvas com <i>outliers</i> acumulados para o grupo <i>Pitch</i> do Aero gerador 005.	93

Figura 57 – Distribuição dos resíduos do Aerogerador 006.	94
Figura 58 – Curvas com <i>outliers</i> acumulados para o grupo Mancal do Aerogerador 006.	95
Figura 59 – Curvas com <i>outliers</i> acumulados para o grupo <i>Gear Box</i> do Aerogerador 006.	96
Figura 60 – Curvas com <i>outliers</i> acumulados para o grupo Gerador do Aerogerador 006.	96
Figura 61 – Curvas com <i>outliers</i> acumulados para o grupo <i>Pitch</i> do Aerogerador 006.	97
Figura 62 – Distribuição dos resíduos do Aerogerador 007.	98
Figura 63 – Curvas com <i>outliers</i> acumulados para o grupo Mancal do Aerogerador 007.	99
Figura 64 – Curvas com <i>outliers</i> acumulados para o grupo <i>Gear Box</i> do Aerogerador 007.	99
Figura 65 – Curvas com <i>outliers</i> acumulados para o grupo Gerador do Aerogerador 007.	100
Figura 66 – Curvas com <i>outliers</i> acumulados para o grupo <i>Pitch</i> do Aerogerador 007.	101
Figura 67 – Distribuição dos resíduos do Aerogerador 008.	102
Figura 68 – Curvas com <i>outliers</i> acumulados para o grupo Mancal do Aerogerador 008.	103
Figura 69 – Curvas com <i>outliers</i> acumulados para o grupo <i>Gear Box</i> do Aerogerador 008.	103
Figura 70 – Curvas com <i>outliers</i> acumulados para o grupo Gerador do Aerogerador 008.	104
Figura 71 – Curvas com <i>outliers</i> acumulados para o grupo <i>Pitch</i> do Aerogerador 008.	105

Lista de Siglas e Abreviaturas

ANEEL	<i>Agência Nacional de Energia Elétrica</i>
CSV	<i>Arquivo de Texto Separado por Vírgulas (Comma Separated Values)</i>
IA	<i>Inteligência Artificial</i>
IDE	<i>Ambiente de Desenvolvimento Integrado (Integrated Development Environment)</i>
LAD	<i>Função Perda Baseada na Mediana dos Erros Absolutos (Least Absolute Deviation)</i>
LOGS	<i>Histórico de Manutenções</i>
ML	<i>Aprendizado de Máquina (Machine Learning)</i>
MSE	<i>Erro quadrático Médio (Mean Square Error)</i>
MW	<i>Mega Watts</i>
PCA	<i>Análise das Componentes Principais (Principal Component Analysis)</i>
RANSAC	<i>Consenso de Amostra Aleatória (Random Sample Consensus)</i>
SCADA	<i>Controle Supervisório e Aquisição de Dados (Supervisory Control And Data Acquisition)</i>
SKLEARN	<i>Sci Kit Learn</i>
SVD	<i>Decomposição em Valores Singulares (Singular Value Decomposition)</i>
UFSC	<i>Universidade Federal de Santa Catarina</i>
UHE	<i>Usine Hidrelétrica</i>

Lista de Símbolos

A_w	Área varrida pelas pás do rotor [m^2]
C_P	Coeficiente de potência
β	Letra grega minúscula beta, representando um ângulo de inclinação
cut_{in}	Valor de velocidade do vento abaixo do qual a turbina não opera
cut_{off}	Valor de velocidade do vento acima do qual a turbina não opera
δ	Mediana da distância interquartil da velocidade do vento dos dados filtrados
DIQ	Distância Interquartil
ϵ	Letra grega minúscula épsilon, representando um resíduo
ϵ	Letra grega minúscula épsilon, representando um vetor de resíduos
$\hat{f}_{inf}, \hat{f}_{sup}$	Modelos de regressão treinados para prever o valor inferior e superior do intervalo de confiança
f_{inf}, f_{sup}	Valores dos limites inferiores e superiores para se realizar a filtragem dos dados
λ	Letra grega minúscula lambda
$\hat{l}_{inf}, \hat{l}_{sup}$	Modelos de regressão treinados para prever o valor inferior e superior do intervalo de confiança
l_{inf}, l_{sup}	Valores dos limites inferiores e superiores para classificação dos outliers da Curva S
μ	Letra grega minúscula mü
m	Número de linhas de uma matriz
ω	Letra grega minúscula ômega, representando a velocidade angular [rad/s]
P	Potência
P_{error}	Valor estimado do erro de medição da potência
ρ	Letra grega minúscula ro
σ	Letra grega minúscula sigma, representando o desvio padrão de alguma variável
θ	Letra grega minúscula teta, representando a matriz coluna dos parâmetros de um modelo
v	Velocidade do vento [m/s]
wiq	Mediana da distância entre os percentis 20% e 80% da velocidade do vento para os dados filtrados
X	Conjunto de pontos unidimensionais
x	Ponto unidimensional
\hat{x}	Mediana de X
$\mathbf{X}_{m \times (n+1)}$	Matriz de m linhas e $n + 1$ colunas contendo as variáveis independentes
$\mathbf{x}^{(i)}$	Linha i da matriz $\mathbf{X}_{m \times (n+1)}$
\mathbf{y}	Matriz coluna de uma variável dependente
$y^{(i)}$	Valor da linha i de \mathbf{y}
$\hat{\mathbf{y}}$	Matriz coluna dos valores preditos de uma variável dependente
	Letra grega minúscula

Sumário

1	INTRODUÇÃO	18
1.1	Objetivos	20
1.2	Objetivos Específicos	20
1.3	Estrutura do Trabalho	20
2	REVISÃO DE BIBLIOGRAFIA	22
2.1	Monitoramento da Temperatura	24
2.2	Curva S	25
2.3	Breve Comentário Sobre <i>Digital Twins</i>	27
3	MÉTODOS MATEMÁTICOS E FERRAMENTAS COMPUTACIONAIS	29
3.1	Diagramas de Caixa e Distância Interquartil	29
3.2	Distribuição Normal	30
3.3	Escalonamento de Atributos	31
3.4	Análise por Componentes Principais	32
3.5	<i>Pipelines e Grid Search</i>	33
3.6	Alguns Modelos de Regressão	35
3.6.1	Regressão Linear	35
3.6.2	RANSAC	37
3.6.3	<i>Gradient Tree Boosting</i>	37
3.7	Análise de Resíduos	39
3.8	Python e R	39
4	METODOLOGIA	41
4.1	Estruturação dos Dados	42
4.2	Modelagem da Temperatura	43
4.2.1	Regressão ou <i>Digital Twin</i>	44
4.2.2	Análise dos Outliers	45
4.3	Limites para Curva S	47
4.3.1	Filtragem	47
4.3.1.1	Estimação da Distância Entre o Percentil 20% e 80%	48
4.3.1.2	Estimativa do Erro de Medição	48
4.3.1.3	Pré Filtragem	49

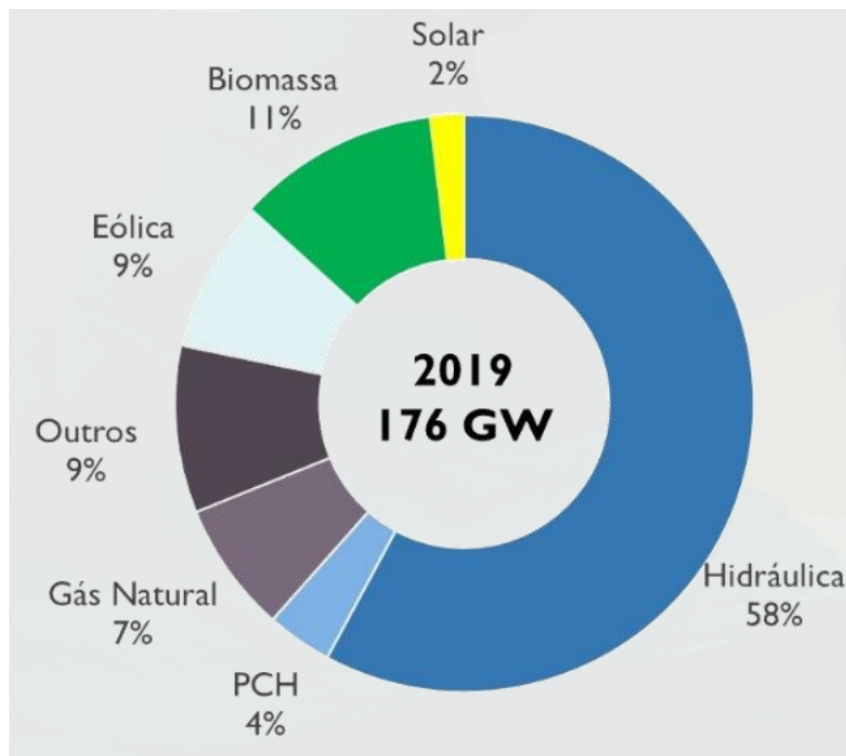
4.3.1.4	Estimação do Intervalo de Regressão	49
4.3.1.5	Filtragem Automática	50
4.3.2	Método para Determinação dos Limites para Curva S	52
4.3.2.1	Estimação da Distância Interquartil	52
4.3.2.2	Estimação do Intervalo de Regressão	52
4.3.2.3	Limites e Outliers	53
5	RESULTADOS	55
5.1	Resultados da Curva S	55
5.1.1	Aerogerador 001	55
5.1.2	Aerogerador 002	58
5.1.3	Aerogerador 003	59
5.1.4	Aerogerador 004	61
5.1.5	Aerogerador 005	63
5.1.6	Aerogerador 006	65
5.1.7	Aerogerador 007	68
5.1.8	Aerogerador 008	70
5.2	Resultados dos <i>Outliers</i> de Temperatura	72
5.2.1	Aerogerador 001	73
5.2.1.1	Mancal	74
5.2.1.2	<i>Gear Box</i>	75
5.2.1.3	Gerador	75
5.2.1.4	<i>Pitch</i>	76
5.2.2	Aerogerador 002	77
5.2.2.1	Mancal	78
5.2.2.2	<i>Gear Box</i>	79
5.2.2.3	Gerador	80
5.2.2.4	<i>Pitch</i>	80
5.2.3	Aerogerador 003	81
5.2.3.1	Mancal	82
5.2.3.2	<i>Gear Box</i>	83
5.2.3.3	Gerador	84
5.2.3.4	<i>Pitch</i>	85
5.2.4	Aerogerador 004	85
5.2.4.1	Mancal	86
5.2.4.2	<i>Gear Box</i>	87
5.2.4.3	Gerador	88
5.2.4.4	<i>Pitch</i>	89

5.2.5	Aerogerador 005	89
5.2.5.1	Mancal	90
5.2.5.2	<i>Gear Box</i>	91
5.2.5.3	Gerador	92
5.2.5.4	<i>Pitch</i>	93
5.2.6	Aerogerador 006	93
5.2.6.1	Mancal	94
5.2.6.2	<i>Gear Box</i>	95
5.2.6.3	Gerador	96
5.2.6.4	<i>Pitch</i>	97
5.2.7	Aerogerador 007	97
5.2.7.1	Mancal	98
5.2.7.2	<i>Gear Box</i>	99
5.2.7.3	Gerador	100
5.2.7.4	<i>Pitch</i>	100
5.2.8	Aerogerador 008	101
5.2.8.1	Mancal	102
5.2.8.2	<i>Gear Box</i>	103
5.2.8.3	Gerador	104
5.2.8.4	<i>Pitch</i>	104
6	CONCLUSÃO	106
	REFERÊNCIAS BIBLIOGRÁFICAS	107

1 Introdução

A matriz energética brasileira é composta majoritariamente por energias renováveis, sendo a principal fonte a geração hidráulica. Uma fonte renovável que tem crescido em importância é a geração eólica, a qual já ocupava 9% da disposição da capacidade instalada no Sistema Energético Brasileiro em 2019 [Ministério de Minas e Energia, 2019], como pode ser observado na Figura 1 divulgada no Plano Decenal de Expansão de Energia 2029 pela Secretaria de Planejamento e Desenvolvimento Energético do Ministério de Minas e Energia.

Figura 1 – Disposição da Capacidade Instalada Total Brasileira em 2019 por tipo de fonte, sem incluir a parcela da UHE Itaipu pertencente ao Paraguai.



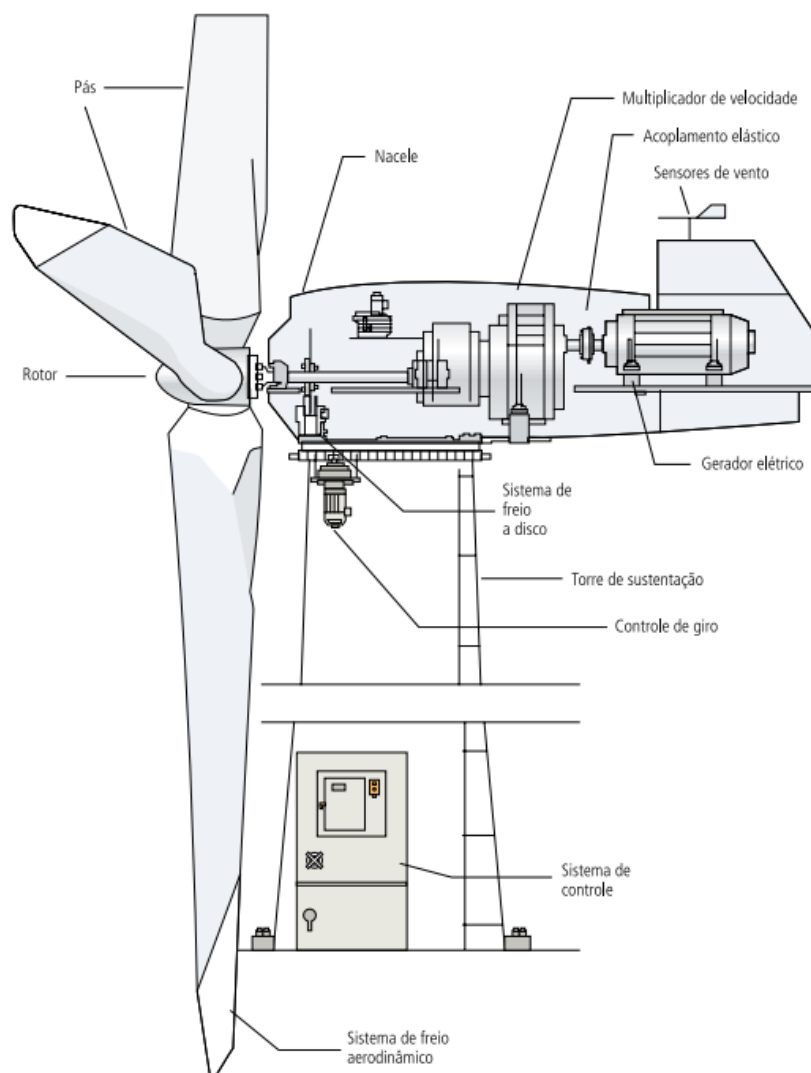
Fonte: [Ministério de Minas e Energia, 2019]

Apesar da capacidade eólica instalada ser menor do que a hidrelétrica, a quantidade de aerogeradores é muito mais expressiva, visto que seriam necessários mais de 953 aerogeradores de 15MW de potência (maior aerogerador previsto para 2024 [Wind Power Monthly, 2020]) para igualar a geração de Itaipu Binacional com seus 20 hidrogeradores de 715MW cada [Itaipu Binacional, 2020]. Com novos projetos envolvendo aerogeradores cada vez maiores, a confiabilidade do aerogerador se torna um dos atributos mais impor-

tantes. Aumentos de confiabilidade podem ser obtidos superando as fraquezas e falhas ocorridas no passado e então formulando novas estratégias e técnicas para minimizar as deficiências encontradas. Neste contexto, cresce o interesse pela implementação de Sistemas de Monitoramento de Condição e Detecção de Falhas (*Condition Monitoring and Fault Detection Systems*) que sejam eficientes, adaptáveis e responsivos [HAMEED et al., 2009].

Os sistemas de monitoramento em aerogeradores não monitoram apenas as condições operativas da máquina como velocidade, direção do vento e potência de saída. Estes também monitoram componentes mais específicos do aerogerador, como mancais, componentes elétricos, eixos, sistema de controle, dentre outros.

Figura 2 – Desenho esquemático de uma turbina eólica moderna



Fonte: [ANEEL, 2008]

Uma visão esquemática e simplificada de um aerogerador moderno pode ser obser-

vada na Figura 2 e os componentes apresentados são comumente referidos por seus nomes em Inglês, sendo eles:

- Multiplicador de velocidade: *Gear Box*;
- Controle de giro: *Yaw*;
- A ponta ou “nariz” do Rotor: *Hub*;
- O controle da angulação de cada pá é feito por componente chamado *Pitch*.

1.1 Objetivos

O presente trabalho tem como objetivo principal utilizar técnicas de “Aprendizagem de Máquina” ou *Machine Learning* e métodos estatísticos para avaliar o funcionamento de aerogeradores e identificar a presença de anomalias. O interesse na identificação das anomalias é servir de suporte para identificação e predição de diferentes tipos de falhas em aerogeradores, tornando assim mais eficiente o trabalho das equipes de manutenção. As anomalias a serem identificadas serão as relacionadas às temperaturas de diversos componentes da máquina e também à Curva S, sendo esta última uma curva característica da operação de um aerogerador.

1.2 Objetivos Específicos

Os passos específicos para atender o objetivo principal são:

- Modelar o comportamento do aerogerador em relação às temperaturas dos componentes;
- Modelar o comportamento da Curva S para delimitar limites operacionais;
- Identificar as anomalias com base nos modelos aprendidos de Machine Learning e métodos estatísticos;
- Relacionar as anomalias com falhas dos aerogeradores utilizando os registros de manutenção (*Logs*).

1.3 Estrutura do Trabalho

O restante do presente trabalho é estruturado da seguinte forma. No Capítulo 2 é feita uma revisão sobre o monitoramento da condição de aerogeradores e algumas de suas

características. No Capítulo 3 são abordadas alguns métodos matemáticos e ferramentas computacionais. O Capítulo 4 é dedicado à metodologia do trabalho. No Capítulo 5 é apresentado os resultados. Por fim o Capítulo 6 se dedica à conclusão.

Devido a questões de confidencialidade, não serão apresentados ou disponibilizados os códigos ou os dados utilizados no presente trabalho.

2 Revisão de Bibliografia

Após a instalação de uma turbina eólica, a melhor forma para obter a sua disponibilidade máxima é através de boas práticas de operação e manutenção [OROZCO; SHENG; PHILLIPS, 2018]. Nesse contexto, o monitoramento é uma tarefa central.

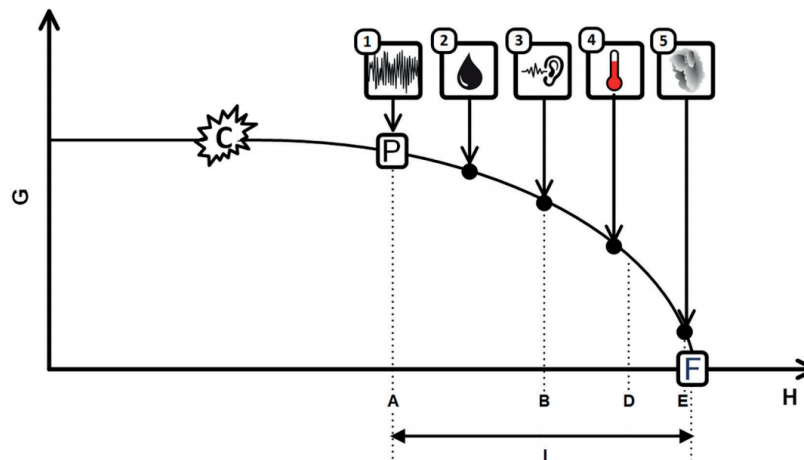
O monitoramento de um aerogerador é tipicamente realizado usando dois tipos de abordagem: uma baseada em modelos físicos e especificações de projeto, a qual envolve o conhecimento de parâmetros de funcionamento da máquina antes de algum tipo de falha começar a se desenvolver, e outra abordagem baseada em modelagem a partir de dados (*data driven models*) [ISO/DIS16079-2, 2019].

A abordagem por modelagem de dados, que pode ser entendida como o uso de dados reais para se criar um modelo, visa principalmente proteger os componentes de interesse e seus sub-componentes, para garantir a maior vida útil possível do aerogerador. Além disso, quando identificada uma falta, frequentemente é possível fazer uma projeção do tempo de uso disponível para a máquina, até que uma manutenção possa ser realizada. Alguns dos componentes tipicamente monitorados em um aerogerador são os pertencentes ao *Drive Train*, como o Mancal Principal, a *Gear Box* quando aplicável, o Gerador (aspectos mecânicos), e os subcomponentes de cada um [ISO/DIS16079-2, 2019].

Dentre os diversos dados que é possível utilizar uma modelagem, os parâmetros operacionais de um aerogerador como a potência, a velocidade do vento e a temperatura dos componentes estão entre os principais descritores dos sintomas que as falhas podem apresentar [ISO/DIS16079-2, 2019].

O monitoramento da temperatura de um aerogerador deve ser condizente com o objetivo de se monitorar seu estado de funcionamento, fornecendo informações que sejam relevantes para operação e manutenção. Uma sugestão de alguns parâmetros a serem monitorados, que podem servir de indicadores de falhas é mostrada na Figura 3 [ISO/DIS16079-2, 2019].

Figura 3 – Representação esquemática típica do desenvolvimento de uma falha mecânica.



Fonte: [ISO/DIS16079-2, 2019]

na qual:

- A: Intervalo de meses;
- B: Intervalo de semanas;
- C: Ponto onde a condição da máquina começa a se modificar;
- D: Intervalo de dias;
- E: Intervalo de minutos;
- F: Ponto onde a falha ocorre;
- G: Condição da máquina;
- H: Tempo;
- I: Intervalo entre a detecção de uma falha em potencial e a ocorrência da falha;
- P: Falha em potencial detectada, aparecimento dos sintomas;
- 1: Vibração detectada por monitoramento de sistemas avançados;
- 2: Partículas de desgaste mecânico podem ser detectadas na análise de óleo;
- 3: Som audível pode ser detectado;
- 4: Aumento da temperatura dos valores absolutos de temperatura é detectado pelos sensores;

- 5: Fumaça é detectada por sensores ou visualmente.

O ponto final no qual ocorre a falha geralmente está associado com altos custos de manutenções, períodos longos de indisponibilidade da máquina e desfalque na geração de energia. Sendo assim quanto antes for identificado um problema, mais tempo se tem para criar um plano de ação que minimize os custos com manutenção e o tempo inoperante da máquina, e por consequência aumente a disponibilidade de geração [ISO/DIS16079-2, 2019].

2.1 Monitoramento da Temperatura

A temperatura dos componentes de um aerogerador é um parâmetro importante que pode servir de auxílio no acompanhamento da sua condição de saúde. Além do que, a temperatura dos componentes em si pode indicar a deterioração devido ao excesso de fricção, folga, má lubrificação ou problemas elétricos (mal contato) dos componentes monitorados [ISO/DIS16079-2, 2019].

Uma das vantagens de se utilizar a temperatura dos componentes para monitoramento da condição do aerogerador, é que não é necessário investimentos em *hardwares* ou instrumentação dedicada como no monitoramento de vibrações e óleo [OROZCO; SHENG; PHILLIPS, 2018]. Uma vez que, a temperatura de diversos componentes é adquirida pelos sistemas de controle supervisorio e aquisição de dados (SCADA). Apesar de alguns sistemas serem mais sofisticados que outros, as técnicas usadas a partir dos registros SCADA podem ser usadas amplamente entre os diversos parques eólicos [ZAHER et al., 2009].

Devido à quantidade de dados adquiridos pelos sistemas SCADA nos parques eólicos, os métodos de baseados em Inteligência Artificial (IA) [HASSOUN et al., 1995] ou *Machine Learning* (ML) [GÉRON, 2019] são fortes candidatos para se modelar as variáveis de interesse, como as temperaturas, e avaliar o estado de saúde de um aerogerador.

Uma forma de abordar o relacionamento da temperatura com a saúde do aerogerador é realizar uma regressão utilizando os dados coletados pelo SCADA. Seja utilizando os valores temperaturas do passado para prever as do futuro com o uso de redes neurais artificiais [ZAHER et al., 2009], ou treinando modelos que usam a potência e a temperatura ambiente para prever a temperatura de cada componente [OROZCO; SHENG; PHILLIPS, 2018]. Essas abordagens visam modelar o comportamento normal dos componentes da máquina, e então detectar quando os componentes apresentam temperaturas anômalas em relação às previstas, relacionando as anomalias com desligamentos devido à super aquecimento [OROZCO; SHENG; PHILLIPS, 2018], ou com falhas dos componentes [ZAHER et al., 2009].

Em função da grande quantidade de modelos de IA ou ML disponíveis atualmente, como por exemplo os disponíveis no pacote computacional *Sci-Kit Learn* [PEDREGOSA et al., 2011], são inúmeras as possibilidades do uso dessas ferramentas para relacionar as temperaturas dos componentes com a saúde do aerogerador. Entretanto, devido aos dados SCADA nem sempre serem rotulados ou à ausência de registros de manutenções ou falhas, esses dados se tornam menos úteis para alguns tipos de algoritmos de ML [OROZCO; SHENG; PHILLIPS, 2018] como os algoritmos de classificação supervisionados [GÉRON, 2019].

Uma abordagem mais detalhada de como utilizar métodos de ML na regressão da temperatura de componentes será apresentada ao longo do Capítulo 4 do presente trabalho.

2.2 Curva S

A curva de potência de saída da máquina, também chamada de Curva S devido a seu formato, é a curva que relaciona a potência fornecida pela máquina com as características do vento. Em teoria, a potência fornecida por uma turbina eólica é regida pela seguinte expressão [SOHONI; GUPTA; NEMA, 2016]:

$$P = \frac{1}{2} \cdot \rho \cdot A_w \cdot C_P(\lambda, \beta) \cdot v^3 \quad (2.1)$$

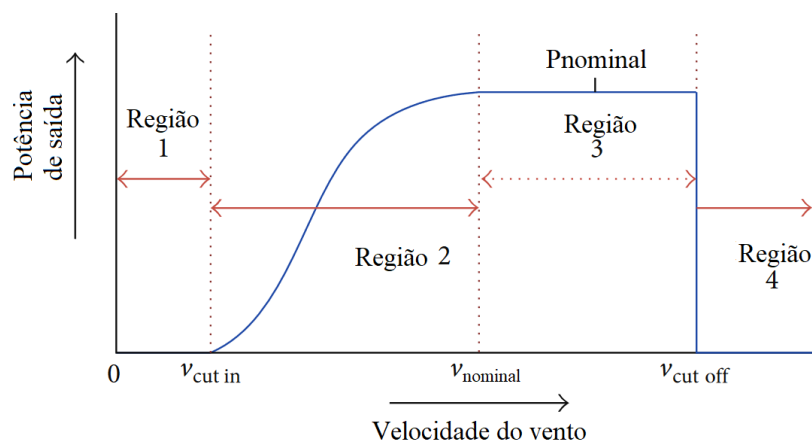
na qual:

- P : Potência de saída da turbina;
- ρ : Densidade do ar [kg/m^3];
- A_w : Área varrida pelas pás do rotor [m^2];
- C_P : Coeficiente de potência, que depende da razão entre velocidade tangencial da pá e a velocidade do vento e do ângulo do *pitch*;
- λ : Razão entre velocidade tangencial da pá e a velocidade do vento. $\lambda = \omega \cdot r/v$ em que ω [rad/s] é a velocidade angular do rotor, r [m] é o raio do centro do rotor até a ponta da pá e v [m/s] é a velocidade do vento;
- β : Ângulo de inclinação do *pitch*;
- v : Velocidade do vento [m/s].

Como pode ser notado a partir de (2.1), a potência depende de diversos fatores, tais como velocidade e direção do vento, densidade do ar (que por sua vez depende da

temperatura, umidade e pressão), entre outros. Assim, verifica-se a complexidade e a dificuldade envolvida em se usar a equação teórica para estimar a potência real de um aerogerador [SOHONI; GUPTA; NEMA, 2016]. Além do mais, muitas vezes valores como densidade do ar e angulação do *pitch*, não estão disponíveis para que se consiga utilizar a curva teórica. Uma representação esquemática de uma Curva S típica, e alguns pontos de interesse são apresentados na Figura 4.

Figura 4 – Representação esquemática de uma Curva S típica, com suas principais regiões e valores de interesse.



Fonte: Adaptado de [SOHONI; GUPTA; NEMA, 2016]

na qual:

- $v_{cut\ in}$: Velocidade mínima do vento para se ter potência de saída;
- $v_{nominal}$: Velocidade do vento a partir da qual a potência atinge seu valor nominal;
- $v_{cut\ off}$: Velocidade máxima do vento permitida para geração, após esse valor o aerogerador é desligado;
- Região 1: Potência de saída é nula devido à velocidades de vento abaixo da de *cut in*;
- Região 2: Potência de saída tem um rápido crescimento até atingir sua potência nominal;
- Região 3: Potência de saída atinge, e se mantém, em seu valor nominal entre $v_{nominal}$ e $v_{cut\ off}$;
- Região 4: A máquina é desligada após o vento ultrapassar $v_{cut\ off}$, permanecendo com potência nula para preservar seus componentes.

Alguns pontos característicos da Curva S, tais como a velocidade de *cut in* e *cut off*, e sua potência nominal são fornecidos pelo fabricante. Muitas vezes, são fornecidas também as curvas características da máquina em tabelas ou gráficos. Entretanto, mesmo com posse desses dados, é difícil relacionar uma expressão analítica com a Curva S real do aerogerador [SOHONI; GUPTA; NEMA, 2016].

Dada a dificuldade de se obter uma expressão analítica para a Curva S efetiva do aerogerador e os possíveis benefícios que podem emergir da sua modelagem prática, existe um grande interesse em se ter um modelo que descreva o comportamento real da geração. Dentre os diversos benefícios que podem ser obtidos de um modelo prático da Curva S estão: a possibilidade de predição da potência futura a ser gerada (para uso no mercado de energia por exemplo), a possibilidade de comparar os valores especificados pelo fabricante com os valores obtidos na operação real, comparação entre turbinas em relação à eficiência, e também a estipulação de valores de referência que podem ser usados para relacionar com falhas ou uma performance abaixo do aceitável [SOHONI; GUPTA; NEMA, 2016].

O uso da Curva S para relacionar com possíveis falhas no aerogerador pode ser feito tendo um modelo de referência para a máquina saudável e então utilizar esse modelo para verificar a operação real da turbina. Assim, caso os valores observados ultrapassem os limites definidos a partir do modelo, um alarme ou alerta pode ser emitido indicando mal funcionamento [HAMEED et al., 2009]. Entretanto essa abordagem traz como desvantagem a necessidade de se saber os períodos em que a máquina estava em bom estado de saúde, requerendo assim algum tipo de rotulação de dados, o que nem sempre está disponível na prática.

2.3 Breve Comentário Sobre *Digital Twins*

Digital Twin pode ser entendida como uma técnica baseada na simulação probabilística, física e escalável de um sistema real, resultando em um espelhamento de alta fidedignidade do comportamento e status do sistema real em tempo hábil. Esta, por sua vez, é criada a partir dos dados históricos proveniente dos sensores e modelos físicos do sistema real [GLAESSGEN; STARGEL, 2012].

Uma *Digital Twin* baseada em dados reais pode reforçar serviços do sistema real, como a manutenção, o monitoramento, a predição da vida útil, entre outros. Além de que, a maioria das aplicações atuais das *Digital Twins* estão na área de prognóstico e gerenciamento da saúde do sistema modelado [TAO et al., 2018a].

Dentro da área de diagnóstico e prognóstico de máquinas e equipamentos, a modelagem de uma *Digital Twin* apresenta, entre outros benefícios, o fácil acesso a uma

cópia virtual fiel do sistema físico real, o qual pode ser de difícil acesso, enriquecimento da informação a partir dos dados reais em conjunto com os virtuais, possibilidade de identificar distúrbios no ambiente real e identificação de falhas ou defeitos em potencial do sistema real. Este último benefício vem da possibilidade de um defeito causar desvios entre o modelo virtual e o sistema real [TAO et al., 2018b].

Quanto à detecção de falhas com auxílio de uma *Digital Twin*, tal aspecto por ser abordado para falhas graduais ou abruptas. Para falhas graduais, os padrões ou sintomas do desenvolvimento de diversos tipos de falhas podem ser aprendidos a partir dos dados históricos reais, seja de um equipamento ou sistema como um todo, sendo possível a identificação e diagnóstico pelo modelo virtual (*Digital Twin*) quando uma nova falha se desenvolver. Para as falhas abruptas, os componentes defeituosos podem ser checados e identificados comparando os parâmetros entre o modelo virtual e o equipamento real [TAO et al., 2018b].

3 Métodos Matemáticos e Ferramentas Computacionais

Diversas ferramentas matemáticas e computacionais podem auxiliar na abordagem de problemas, desde mais simples até mais complexos. Por exemplo, na estatística existem diversas ferramentas que podem auxiliar engenheiros, físicos, químicos entre outros profissionais no desenvolvimento de novos produtos, processos, sistemas ou ainda no melhoramento de sistemas já existentes [MONTGOMERY; RUNGER; HUBELE, 2009].

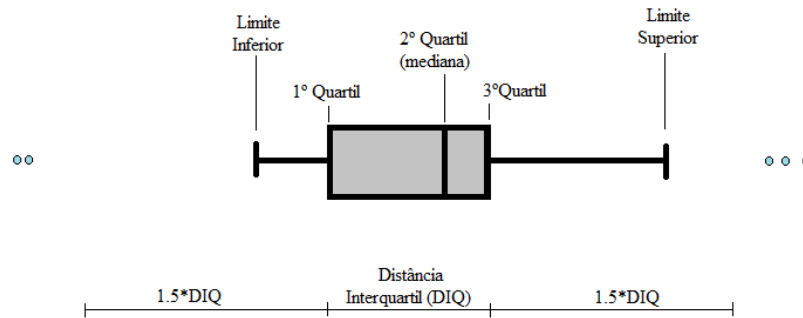
Em relação às ferramentas computacionais, o aprendizado de máquina, mais conhecido como *Machine Learning* (ML), pode ser entendido como a ciência (ou até mesmo a arte) de programar computadores de tal forma que eles aprendam com os dados neles inseridos [GÉRON, 2019]. Uma das grandes vantagens da utilização do ML é a possibilidade de se obter regras e modelos sem que estas sejam explicitamente programadas, em contraste com os casos típicos de programação em que as regras devem ser definidas pelo programador [SAMUEL, 1959]. Essa versatilidade, somada à disponibilidade gratuita de um grande arsenal de ferramentas como os encontrados em bibliotecas computacionais [PEDREGOSA et al., 2011], são um grande incentivo para utilização de *Machine Learning* tanto em problemas que precisem de regras flexíveis e ajustáveis quanto em problemas complexos de difícil modelagem [GÉRON, 2019]. Este capítulo é dedicado à descrição dos métodos matemáticos e ferramentas computacionais utilizados no presente trabalho.

3.1 Diagramas de Caixa e Distância Interquartil

Um *Box Plot* ou Diagrama de Caixas é uma forma gráfica de visualizar facilmente algumas características importantes de um conjunto de dados. Nesse tipo de representação é possível ter de forma rápida uma noção sobre o centro, a dispersão, a simetria e a identificação de valores discrepantes (*outliers*) no conjunto de dados [MONTGOMERY; RUNGER; HUBELE, 2009].

No Diagrama de Caixas é possível visualizar onde se encontram o primeiro, o segundo e o terceiro quartis, ou seja, onde se encontra 25%, 50% e 75% da distribuição dos dados respectivamente. Além disso, o conhecimento da distância interquartil permite ter uma medida de dispersão dos dados, sem que seja influenciada pelos *outliers* [MONTGOMERY; RUNGER; HUBELE, 2009]. Na Figura 5, encontra-se uma representação de um *Box Plot* padrão e suas características de interesse.

Figura 5 – Representação de um *Box Plot* padrão, onde o primeiro, o segundo e o terceiro quartis mostram onde se encontram 25%, 50% e 75% dos dados. Com a Distância Interquartil (DIQ), é possível ter ideia da dispersão dos dados, e os pontos em azul que estão à $1,5 \cdot \text{DIQ}$ acima do 3º Quartil ou abaixo do 1º Quartil são considerados *outliers*.



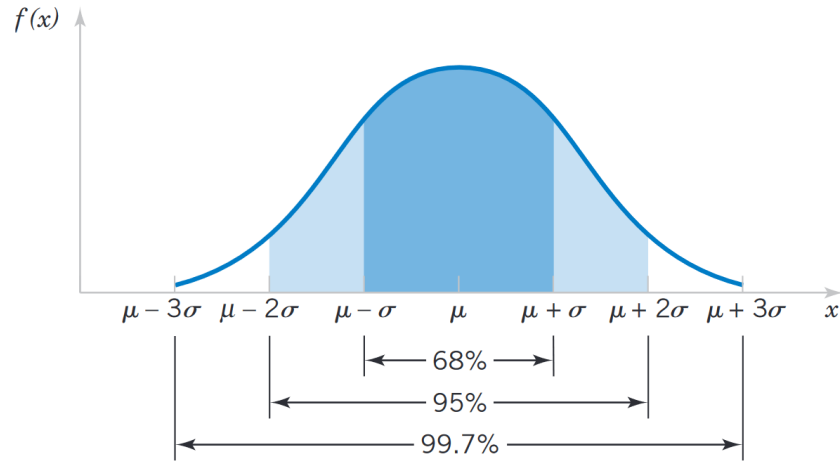
O Diagrama de Caixas explicita os percentis 25%, 50% e 75%. Entretanto pode se ter interesse nos valores de outros percentis dependendo do contexto. De qualquer forma o valor de um percentil sempre representará o valor até o qual temos um certo percentual de dados acumulados. Por exemplo um percentil de 30% significa que até aquele valor temos 30% dos dados acumulados [MONTGOMERY; RUNGER; HUBELE, 2009].

3.2 Distribuição Normal

Possivelmente um dos modelos mais conhecidos para a distribuição de uma variável aleatória é a Distribuição Normal. Esse modelo pode ser completamente caracterizado com apenas dois parâmetros, a média e a variância (ou desvio padrão). Uma forma de identificar que uma variável aleatória X se aproxima, ou pode ser representada, por uma distribuição normal é com a notação $X \sim N(\mu, \sigma^2)$ [MONTGOMERY; RUNGER; HUBELE, 2009].

Uma visualização esquemática da função densidade de probabilidade da Distribuição Normal pode ser vista na Figura 6, sendo que, independentemente de quais forem os parâmetros da distribuição, a área total abaixo da curva sempre terá valor unitário (100% dos dados) [MONTGOMERY; RUNGER; HUBELE, 2009].

Figura 6 – Representação de uma função densidade de probabilidade de uma Distribuição Normal $X \sim N(\mu, \sigma^2)$, onde μ é a média e σ é o desvio padrão.



Fonte: [MONTGOMERY; RUNGER; HUBELE, 2009]

Como pode ser observado na Figura 6, a partir do modelo teórico da Curva Normal, aproximadamente 99.7% de todas as observações se encontram entre 3 desvios à esquerda e à direita da média [MONTGOMERY; RUNGER; HUBELE, 2009]. Assim quando os dados se ajustam a uma distribuição Normal, os valores que se encontram a uma distância maior que 3 desvios padrões à direita ou a esquerda em relação à média, podem ser considerados *outliers*.

3.3 Escalonamento de Atributos

De maneira geral, algoritmos de ML não tem um bom desempenho em dados numéricos que possuem uma diferença de escala muito grande. Assim, uma das etapas mais importantes no tratamento, ou pré processamento, dos dados é o escalonamento de atributos, comumente chamado de *feature scaling* [GÉRON, 2019]. Esse escalonamento é importante devido aos métodos de otimização matemática que os algoritmos executam internamente, como o Gradiente Descendente Estocástico [PLAKHOV; CRUZ, 2004], o qual é sensível à diferença de escala entre as dimensões numéricas.

Algumas das formas de escalonamento mais comuns são o *min max scaling* e *standardization scaling* [GÉRON, 2019]. Entretanto, esses métodos não são robustos a *outliers*. Um escalonamento que é robusto a dados com *outliers* é o *Robust Scaler* [PEDREGOSA et al., 2011].

O *Robust Scaler* é baseado na seguinte transformação aplicada individualmente a

cada atributo do conjunto de dados:

$$x' = \frac{x - \hat{x}}{DIQ} \quad (3.1)$$

onde:

- x é qualquer ponto pertencente ao conjunto de pontos do mesmo atributo X ;
- DIQ é a distância entre o primeiro quartil e o terceiro quartil do conjunto de pontos de X ;
- \hat{x} é a mediana do conjunto de pontos de X .

3.4 Análise por Componentes Principais

Um dos problemas encontrados em *Machine Learning* é a chamada “maldição da dimensionalidade”. Esse problema está associado a conjuntos de dados com muitas dimensões, causando a lentidão no treinamento dos modelos de ML e dificuldade em se obter uma boa solução para o problema [GÉRON, 2019].

Uma das formas mais comuns de se realizar a redução da dimensionalidade é utilizando a técnica da Análise por Componentes Principais também conhecida como *Principal Component Analysis* (PCA). A ideia básica por trás do PCA é fazer a projeção dos dados de um conjunto com dimensão d em um conjunto de dimensão k , sendo $k < d$. A projeção é escolhida de forma a preservar a máxima variância dos dados originais [GÉRON, 2019].

A apresentação rigorosa do funcionamento do PCA [DEISENROTH; FAISAL; ONG, 2020], assim como sua demonstração matemática não serão apresentadas neste trabalho. Será feita apenas uma apresentação descritiva envolvendo os principais conceitos do algoritmo.

Para projeção em uma dimensão menor dos dados, o PCA utiliza a decomposição em valores singulares (SVD) do conjunto de dados $\mathbf{X}_{m \times n}$ para obter seus k componentes principais. Assim, a decomposição é realizada conforme a seguinte expressão [DEISENROTH; FAISAL; ONG, 2020]:

$$\mathbf{X}_{m \times n} = \mathbf{U}_{m \times m} \cdot \mathbf{\Sigma}_{m \times n} \cdot \mathbf{V}_{n \times n}^T \quad (3.2)$$

em que:

- $\mathbf{X}_{m \times n}$ é a matriz com os dados originais;
- $\mathbf{U}_{m \times m}$, $\mathbf{\Sigma}_{m \times n}$ e $\mathbf{V}_{n \times n}$ são as matrizes da decomposição SVD, sendo $\mathbf{V}_{n \times n}$ a matriz que contém os componentes principais.

Após a decomposição apresentada em (3.2), o PCA toma k componentes principais, ou seja $\mathbf{V}_{n \times n}$ se torna $\mathbf{V}_{n \times k}$. Com essa nova quantidade de componentes principais é realizada a projeção de acordo com a seguinte expressão:

$$\mathbf{Xp}_{m \times k} = \mathbf{X}_{m \times n} \cdot \mathbf{V}_{n \times k} \quad (3.3)$$

em que:

- $\mathbf{Xp}_{m \times k}$ é o novo conjunto de dados com menor dimensão;
- $\mathbf{X}_{m \times n}$ é a matriz com os dados originais;
- $\mathbf{V}_{n \times k}$ é a matriz de componentes principais com dimensão reduzida.

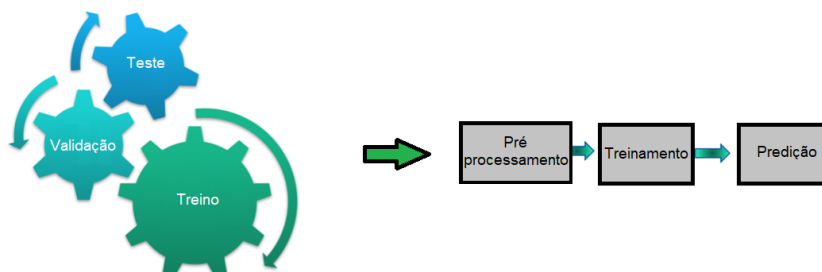
Quanto à escolha do valor de k (a nova dimensão do conjunto de dados), existem algumas abordagens disponíveis na literatura [PEDREGOSA et al., 2011]. Entretanto, uma grande vantagem é obtida usando um método automático de escolha [MINKA, 2001].

A escolha automática do número de dimensões para o PCA envolve uma área mais avançada da matemática que está fora do escopo do presente trabalho e, por esse motivo, não será abordada profundamente. Entretanto, a ideia por trás do método automático de dimensões para o PCA está em unir a preservação da variância máxima com a probabilidade da representação correta dos dados [MINKA, 2001]. Dessa forma, a dimensão é escolhida em conjunto com uma abordagem Bayesiana, que considera as probabilidades condicionais entre as projeções de menor dimensão e o conjunto de dados reais [MINKA, 2001].

3.5 Pipelines e Grid Search

Uma das formas de se organizar um processamento de dados, de tal forma que o mesmo tenha uma sequência lógica e funcional, é organizá-lo em uma *Pipeline*. Em sistemas e *scripts* envolvendo *Machine Learning*, *Pipelines* são muito comuns visto que geralmente existe uma grande quantidade de dados a serem tratados e processados [GÉRON, 2019]. Na Figura 7 é possível ver uma representação ilustrativa de uma *Pipeline* simples.

Figura 7 – Representação ilustrativa de uma *Pipeline* simples, contendo os conjuntos de treinamento, teste e validação, o pré processamento, o treinamento dos modelos e sua predição.



Fonte: Adaptado de [Datanami, 2018]

As *Pipelines* também podem ter aplicações específicas em algumas partes do *script*. Por exemplo, o pré processamento de dados, incluindo escalonamento, redução de dimensionalidade, entre outros [GÉRON, 2019], pode ser implementado de forma unificada, sendo executado sequencialmente para todo o conjunto de dados [PEDREGOSA et al., 2011].

Além do pré processamento de dados, as *Pipelines* também podem ser construídas de tal forma a se obter um método automatizado para o selecionamento de modelos. Assim, é possível treinar e escolher o modelo que tem a melhor performance, dada uma métrica de desempenho [PEDREGOSA et al., 2011].

Uma forma eficiente de se fazer o selecionamento automático de modelos a partir de uma *Pipeline* é utilizar um método que treine cada modelo em diversas partições do conjunto de dados, mantendo uma partição reservada para avaliação do desempenho, então medir o desempenho de cada modelo na partição que não foi vista no treinamento. Repetindo esse procedimento diversas vezes e com partições escolhidas aleatoriamente, tem-se uma segurança maior que o modelo escolhido com melhor desempenho terá o melhor desempenho fora do conjunto de treinamento. Isso se deve ao fato que a chance de se ter o melhor desempenho devido a uma escolha favorável da partição dos dados decai, visto que os resultados de cada partição para cada modelo podem ser agregados por uma média por exemplo, e ainda avaliado a variação do desempenho devido às partições [PEDREGOSA et al., 2011].

O procedimento descrito para seleção automática do melhor modelo pode ser encontrado no algoritmo *Grid Search* [PEDREGOSA et al., 2011], que por sua vez diminui o risco que o modelo fique enviesado sofrendo *overfitting*, que acontece quando o modelo tem bom desempenho apenas no conjunto de treinamento sem que seja possível utilizá-lo em dados novos, que não foram vistos no treinamento [GÉRON, 2019]. O *Grid Search*

também é muito útil para o selecionamento dos melhores hiper-parâmetros dos modelos de ML [PEDREGOSA et al., 2011], visto que os hiper-parâmetros não são aprendidos pelo modelo a partir dos dados, e devem ser explicitamente definidos para o cada modelo [GÉRON, 2019].

3.6 Alguns Modelos de Regressão

Tipicamente, as bibliotecas computacionais dedicadas a *Machine Learning* apresentam diversos modelos para regressão, desde modelos mais simples como a regressão linear a partir de mínimos quadrados, até modelos mais complexos como redes neurais [PEDREGOSA et al., 2011; GÉRON, 2019]. Inclusive, modelos mais simples como a regressão linear por mínimos quadrados podem ser usados como referência de performance para modelos mais complexos [OROZCO; SHENG; PHILLIPS, 2018]. Sendo assim, uma noção sobre o funcionamento de alguns modelos é interessante para se obter um melhor desempenho e utilização dos mesmos em um problema real [GÉRON, 2019].

3.6.1 Regressão Linear

A coleção de ferramentas estatísticas usadas para modelar a relação entre variáveis é conhecida como análise de regressão. A regressão linear consiste na modelagem de um relacionamento linear entre uma variável dependente e uma, ou mais, variáveis independentes [MONTGOMERY; RUNGER; HUBELE, 2009].

De forma matricial a expressão que descreve a regressão linear para um conjunto de m pontos contendo n variáveis independentes é dada por [GÉRON, 2019]:

$$\hat{\mathbf{y}} = \mathbf{X}_{m \times (n+1)} \cdot \boldsymbol{\theta} \quad (3.4)$$

onde:

- $\hat{\mathbf{y}}$ é a matriz coluna dos valores preditos a partir de $\mathbf{X}_{m \times (n+1)}$;
- m é o número de linhas, que representam as unidades observadas;
- n é o número de colunas ou variáveis dos dados;
- $\mathbf{X}_{m \times (n+1)}$ é composto pelas colunas $[\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$, sendo que a coluna \mathbf{x}_0 possui todos os elementos com valor igual a 1, e as outras colunas $[\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$ possuem os n vetores das variáveis independentes, variáveis essas também chamadas de *features*;

- $\boldsymbol{\theta}$ é o vetor (matriz coluna) dos parâmetros do modelo, contendo os pesos $\theta_1, \theta_2, \dots, \theta_n$ a serem multiplicados pelas *features* e θ_0 que será somado para cada ponto. θ_0 é o *bias* ou intercepto do modelo.

Os parâmetros do modelo, ou seja os valores de $\boldsymbol{\theta}$, podem ser calculados de forma fechada definindo uma função custo (*loss*) para o modelo e então minimizando essa função. Uma das funções custo mais utilizadas é o erro quadrático médio. O erro quadrático médio, ou *Mean Square Error* (MSE), para o modelo linear apresentado em (3.4) pode ser expressado da seguinte forma [GÉRON, 2019]:

$$MSE(\boldsymbol{\theta}) = \frac{1}{m} \cdot \sum_{i=1}^m (\mathbf{x}^{(i)} \cdot \boldsymbol{\theta} - y^{(i)})^2 \quad (3.5)$$

onde:

- m é o número de linhas, ou número de pontos a ser calculado o erro;
- $\mathbf{x}^{(i)}$ é a linha (ou ponto) das variáveis independentes (com a coluna de valores unitários);
- $y^{(i)}$ é o valor da variável dependente referente ao valor de $\mathbf{x}^{(i)}$.

Definida a função custo, $\boldsymbol{\theta}$ pode ser obtido a partir da minimização de tal função. A solução fechada para tal minimização é dada por

$$\boldsymbol{\theta} = \left(\mathbf{X}_{m \times (n+1)}^T \cdot \mathbf{X}_{m \times (n+1)} \right)^{-1} \cdot \mathbf{X}_{m \times (n+1)}^T \cdot \mathbf{y} \quad (3.6)$$

onde:

- $\mathbf{X}_{m \times (n+1)}$ é a matriz contendo as *features* e a coluna de valor unitário como descrito em (3.4);
- \mathbf{y} é a matriz coluna dos valores da variável dependente contida nos dados.

A Expressão (3.6) também é conhecida como Equação Normal. O problema de se utilizar essa forma fechada para encontrar $\boldsymbol{\theta}$ está na operação de inversão de matriz, em função do custo computacional envolvido e da eventual inexistência da inversa a ser calculada. Sendo assim o tempo e o custo computacional para se calcular $\boldsymbol{\theta}$ usando (3.6) é alto para conjuntos de dados que possuam muitas (e.g., milhares) de *features* [GÉRON, 2019].

Uma forma de se obter $\boldsymbol{\theta}$ sem ter que calcular as inversas das matrizes é utilizar um algoritmo que minimize a função custo iterativamente visando encontrar o $\boldsymbol{\theta}$ ótimo. Para o caso do Erro Quadrático Médio definido em (3.5), assim como para diversas outras funções

custo que são convexas, o algoritmo do Gradiente Descendente [PLAKHOV; CRUZ, 2004] é uma boa opção. Tal algoritmo busca, por meio iterativo, o valor mínimo da função com base no cálculo do gradiente de cada ponto em relação à função custo [GÉRON, 2019].

3.6.2 RANSAC

A Regressão Linear com função custo MSE é pouco robusta a dados que apresentam anomalias (*outliers*). Na tentativa de contornar esse problema, algumas bibliotecas oferecem modelos como o *Random Sample Consensus* (RANSAC), o qual permite a realização de uma Regressão Linear que seja robusta a *outliers* [PEDREGOSA et al., 2011].

O RANSAC pode ser entendido como um Consenso de Amostras Aleatórias, visto que o algoritmo pega amostras aleatórias do conjunto de dados total, e então identifica quais são os *outliers* para não utilizá-los na estimação dos parâmetros do modelo. As amostras contendo apenas os pontos que não são *outliers* são chamadas de consenso para o modelo em questão [FISCHLER; BOLLES, 1981].

A ideia por trás do RANSAC é servir de auxílio aos modelos de regressão, sendo que o RANSAC em conjunto com os modelos de regressão ajuda na obtenção de parâmetros que não tenham sido influenciados por *outliers*.

O algoritmo funciona de forma iterativa, onde uma amostra aleatória é retirada dos dados e então os parâmetros de um modelo M1 são calculados. Se o erro de M1 estiver abaixo de uma tolerância especificada, utiliza-se M1 juntamente com uma margem (por exemplo dentro de dois desvios padrões de distância) para criar o conjunto S1*. Os pontos de S1* considerados sem *outliers* são utilizados para estimar o modelo de regressão final. Caso o erro de M1 esteja acima da tolerância, o processo é repetido para k tentativas. Se após k tentativas não se obter sucesso ou o algoritmo terminar com um erro, é utilizado o conjunto S1* que apresentou menor erro dentre todas as amostras [FISCHLER; BOLLES, 1981].

3.6.3 Gradient Tree Boosting

O *Gradient Tree Boosting* é um método de *Ensemble* (literalmente “conjunto” em português) que utiliza como preditor base as Árvores de Decisão [GÉRON, 2019]. De forma simplificada, as Árvores de Decisão fazem suas previsões calculando médias da variável dependente a partir de intervalos das variáveis independentes. Esses intervalos são selecionados de tal forma a minimizar a função custo. Assim, primeiramente são selecionados quais são os melhores intervalos das variáveis independentes e então as médias da variável dependente são calculadas em cada intervalo [GÉRON, 2019; PEDREGOSA et al., 2011].

Por sua vez o *Gradient Tree Boosting* agrega as predições de diversas Árvores de Decisão em apenas uma predição (por esse motivo ele é um método de *Ensemble*). Essas Árvores são treinadas de forma sequencial sendo que cada árvore é treinada nos resíduos de sua predecessora [GÉRON, 2019].

A função custo utilizada pelo *Gradient Tree Boosting* é na verdade a função custo que as Árvores de Decisão que o compõe utilizam. Para tarefas de regressão é comum utilizar como função custo o MSE. Entretanto isso pode fazer com que o modelo não seja tão robusto em relação aos *outliers* [PEDREGOSA et al., 2011]. Uma forma de se contornar tal problema é utilizando outra função custo para o *Gradient Tree Boosting*, como por exemplo a mediana do erro absoluto [PEDREGOSA et al., 2011; FRIEDMAN, 2001].

A função custo baseada no erro absoluto também é conhecida por *Least Absolute Deviation* (LAD) [FRIEDMAN, 2001], e quando esta é utilizada no *Gradient Tree Boosting* visa minimizar a mediana do erro absoluto em suas Árvores de Decisão (estimador base) [PEDREGOSA et al., 2011]. A representação genérica da função custo baseada no erro absoluto é dada por

$$L = \text{mediana} \left\{ \frac{y^{(i)} - F_{j-1}(\mathbf{x}^{(i)})}{h(\mathbf{x}^{(i)}; \mathbf{a}_j)} \right\}_1^m \quad (3.7)$$

em que:

- $y^{(i)}$ é o valor da variável dependente na linha i ;
- $\mathbf{x}^{(i)}$ é a linha (ou ponto) das variáveis independentes;
- $F_{j-1}(\mathbf{x}^{(i)})$ é a predição da Árvore de regressão antecessora;
- $h(\mathbf{x}^{(i)}; \mathbf{a}_j)$ é a mediana da variável dependente considerando os intervalos das variáveis independentes de cada preditor.

Outra função custo que pode ser utilizada no *Gradient Tree Boosting* é em relação ao percentil do erro. A minimização da função perda LAD pode ser interpretada como a minimização do percentil 50% do erro absoluto [FRIEDMAN, 2001] e leva a predição da mediana para cada sub conjunto dos dados pelas Árvores de Decisão [PEDREGOSA et al., 2011]. Mas além disso, é possível utilizar uma função perda para que o modelo faça a predição de qualquer percentil, com isso é possível utilizar o *Gradient Tree Boosting* para realizar a predição de intervalos de regressão [PEDREGOSA et al., 2011], visto que os intervalos de regressão são baseados na distribuição dos dados do conjunto.

3.7 Análise de Resíduos

A partir do momento em que usamos algum modelo de regressão para modelar o comportamento de um conjunto de dados, estamos fazendo algumas suposições sobre tais dados. Por exemplo, para conseguirmos estimar os parâmetros de um modelo, geralmente assumimos que os erros são não correlacionados e que se comportam como uma distribuição normal com variância constante [MONTGOMERY; RUNGER; HUBELE, 2009].

A análise dos resíduos visa de maneira geral checar se o modelo de regressão é adequado para um determinado conjunto de dados. Os resíduos podem ser escritos da forma $\epsilon_i = y_i - \hat{y}_i$, em que $i = 1, 2, \dots, m$, y_i é o o valor real e \hat{y}_i é o valor predito do modelo para cada observação [MONTGOMERY; RUNGER; HUBELE, 2009]. A seguinte expressão apresenta os resíduos com uma forma de representação vetorizada e mais compacta:

$$\boldsymbol{\epsilon} = \mathbf{y} - \hat{\mathbf{y}} \quad (3.8)$$

onde:

- $\boldsymbol{\epsilon}$ é a matriz coluna de resíduos;
- \mathbf{y} é a matriz coluna com os valores reais da variável a ser predita pelo modelo de regressão;
- $\hat{\mathbf{y}}$ é a matriz coluna com as predições feita pelo modelo.

Quando o modelo de regressão está adequado, seus resíduos devem ser (ou pelo menos se aproximar) de uma variável aleatória normal de média zero e variância constante, da forma $\boldsymbol{\epsilon} \sim N(0, \sigma_\epsilon^2)$. Por outro lado, quando os resíduos apresentarem algum padrão que não seja aleatório com distribuição normal, é provável que este modelo não seja adequado para a modelagem correta do conjunto de dados [MONTGOMERY; RUNGER; HUBELE, 2009].

3.8 Python e R

Atualmente existem diversas linguagens computacionais que servem de ferramentas auxiliares para desempenhar uma infinidade de tarefas. Na área de Análise de Dados ou Ciência de Dados (*Data Science*), duas linguagens muito utilizadas e com uma grande variedade de bibliotecas são R e Python. [WICKHAM; GROLEMUND, 2016; VANDERPLAS, 2016].

Cada uma dessas linguagens possuem forças e fraquezas diferentes. R pode se destacar por suas bibliotecas com uma grande variedade de ferramentas estatísticas (entre

muitas outras) e suas funções vetorizadas [WICKHAM; GROLEMUND, 2016], enquanto Python por suas bibliotecas de *Machine Learning* (entre muitas outras também) e por sua agilidade com operações iterativas [VANDERPLAS, 2016]. Entretanto muitas vezes é possível realizar o mesmo tipo de implementação tanto em uma linguagem quanto em outra.

Ambas as linguagens possuem *softwares* gratuitos e *open source* para se desenvolver códigos de forma colaborativa, um exemplo para R é o Rstudio [RStudio Team, 2015], e um exemplo para Python é o Google Colab [BISONG, 2019].

Sendo assim o uso dessas ferramentas computacionais é de grande ajuda na área de *Data Science*, sendo usando as linguagens de forma individual ou então em conjunto, utilizando os melhores aspectos de cada uma [WICKHAM; GROLEMUND, 2016; VANDERPLAS, 2016].

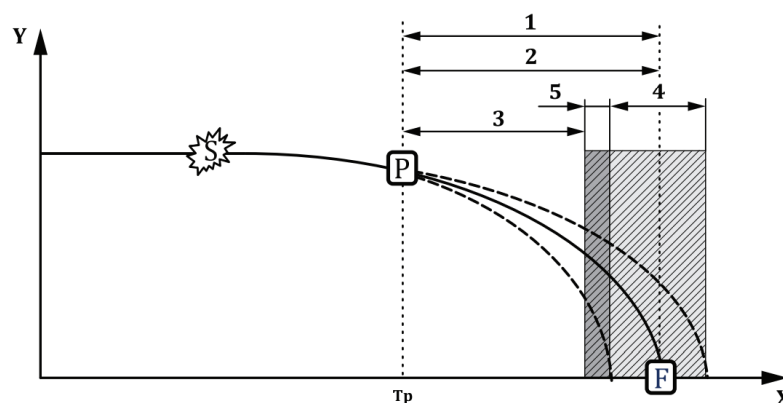
4 Metodologia

Segundo a ISO 16079-1, os passos entre uma falta (problema) ocorrer e essa falta ocasionar uma falha são:

- Uma falta ocorre;
- A falta é percebida;
- A falta é identificada;
- Faz se uma predição do tempo de vida útil do ativo até que uma falha ocorra.

Esses passos podem ser visualizados em uma curva conhecida como PF, conforme pode ser visto na Figura 8.

Figura 8 – Representação esquemática da curva PF



Fonte: [ISO16079-1, 2017]

Nessa figura, tem-se:

- S: Ponto onde a falta começa a se desenvolver, ou onde um erro ocorre;
- P: Ponto onde é percebido que alguma falta ocorreu e então é identificado qual tipo de falta e qual componente foi afetado;
- F: Ponto onde a falha ocorre, e o ativo para de funcionar como deveria;
- X: Tempo;
- Y: Condição do ativo;

- 1: Intervalo de P até F;
- 2: Tempo estimado até que a falha ocorra;
- 3: Tempo de vida útil estimado;
- 4: Intervalo de confiança da estimativa;
- 5: Margem de risco.

Seguindo a ideia que essas são as etapas que levam a uma falha, junto com a inexistência de dados rotulados e uma falta de conhecimento sobre o comportamento dos aerogeradores em relação ao seu estado de saúde, decidiu-se por abordar a segunda etapa da curva PF no presente trabalho, sendo ela a etapa da percepção que alguma falta (problema) ocorreu.

Todo o desenvolvimento foi realizado de forma à identificar anomalias no aerogerador devido às suas condições operativas, a fim de tentar relacionar as anomalias com futuras falhas que tenham ocorrido após as mesmas.

Do início do desenvolvimento até a análise das anomalias, o trabalho percorreu as seguintes etapas:

- Estruturação dos dados;
- Divisão em duas classes de modelagem;
- Análise das anomalias para cada tipo de modelagem.

4.1 Estruturação dos Dados

Os dados utilizados são provenientes de um sistema SCADA (Controle Supervisório e Aquisição de Dados). Para ter acesso aos mesmos foi necessário fazer a exportação manual em diversos arquivos “.csv” providenciados pelo dono do ativo.

A massa de arquivos era então processada e filtrada usando a linguagem de programação R com a IDE Rstudio [RStudio Team, 2015], e estruturada em um formato tabular de dados, centralizados em apenas um arquivo “.csv”. As medidas eram registradas de dez em dez minutos para todas as variáveis, fazendo com que o arquivo final tivesse em torno de dez milhões de linhas (instâncias).

Esse arquivo estruturado então foi utilizado como o *dataset* para os algoritmos de ML (*Machine Learning*) e modelagem do comportamento do aerogerador, sendo tal *dataset* dividido em três conjuntos: treinamento, validação e teste.

O problema de detecção de anomalias ou identificação de um aerogerador saudável pode ser encarado como um problema de classificação. Porém, devido à falta de dados rotulados e ao interesse de se ter um modelo computacional que simule o comportamento da máquina física e possa identificar falhas potenciais no equipamento [TAO et al., 2018b], foi decidido utilizar algoritmos de regressão para se obter um (ou mais) modelo(s) de predição para cada variável de interesse. Após isso, utilizou-se técnicas estatísticas em conjunto com as predições dos modelos para classificar os dados como anomalias (*outliers*) ou não.

Todas as modelagens foram realizadas utilizando a plataforma Google Colab [BI-SONG, 2019] em linguagem Python e com auxílio da biblioteca Sklearn [PEDREGOSA et al., 2011].

4.2 Modelagem da Temperatura

Para modelagem da temperatura, primeiramente foram selecionadas as variáveis do registro SCADA que não eram temperaturas e a temperatura ambiente. Tais atributos foram escolhidos como variáveis independentes, tendo como variável dependente as temperaturas de diversos componentes. Para cada temperatura de cada componente foi desenvolvido um modelo individual.

As variáveis disponíveis do sistema SCADA utilizadas como independentes foram:

- Posição da Nacelle [graus];
- Potência de saída [kW];
- Velocidade do rotor [rpm];
- Velocidade do vento [m/s];
- Temperatura ambiente [°C].

Essas variáveis passaram por um pré-processamento, utilizando-se PCA com o selecionamento automático do número de componentes principais (“*mle*” no Sklearn) [MINKA, 2001; PEDREGOSA et al., 2011]. Optou-se por essa abordagem pois assim o algoritmo pode ser aplicado em diversas máquinas que possam ter mais variáveis nos registros SCADA, sem que seja feita uma escolha de *features* específica para cada uma.

Em conjunto com o PCA, foi aplicado o escalonamento das variáveis com escalonador robusto a anomalias (*Robust Scaler*) para melhorar a performance dos algoritmos de regressão.

4.2.1 Regressão ou *Digital Twin*

Após o pré-processamento dos dados foi criado um algoritmo que faz o treinamento e selecionamento automático dos modelos candidatos de *Machine Learning*, sendo facilmente escalável e paralelizado.

Pelo fato da implementação usar um Pipeline em conjunto com uma função de selecionamento de modelos, o número de modelos candidatos pode ser escalado facilmente, sendo acrescentados novos candidatos ou retirados os existentes. O algoritmo também seleciona candidatos de treinamento paralelizado ou sequencial, como modelos inseridos no *Grid Search* ou *Gradient Tree Boosting*.

Os modelos candidatos à regressão são provenientes da biblioteca Sklearn [PEDRE-GOSA et al., 2011], sendo todos eles e/ou suas respectivas funções custo (*loss*) escolhidas de tal forma que cada modelo se tornasse mais robusto ao enviesamento devido à possível presença de anomalias.

Atualmente os modelos candidatos são:

- Regressão Linear;
- Regressão RANSAC (Random Sample Consensus);
- Gradient Tree Boosting Regression.

O algoritmo com todos os candidatos foi executado para cada um dos componentes que se desejava modelar o comportamento da temperatura e então o melhor candidato foi automaticamente selecionado para cada componente.

Dez componentes foram modelados em relação à temperatura, sendo eles:

- Mancal A;
- Mancal B;
- Mancal da Gear Box;
- Gear Box;
- Mancal 1 do Gerador;
- Mancal 2 do Gerador;
- Mancal do Eixo Principal;
- Motor do Pitch 1;
- Motor do Pitch 2;

- Motor do Pitch 3.

Após todos os componentes terem seus respectivos modelos selecionados e treinados, os modelos foram salvos para futuramente serem utilizados sem que seja necessário realizar um novo treinamento. Os modelos salvos ficam disponíveis inclusive para outros sistemas e/ou aplicativos utilizarem, porém no trabalho em questão os modelos são salvos para posteriormente serem carregados e realizarem a predição da temperatura de cada componente.

A temperatura predita por sua vez é comparada com a temperatura medida do componente, e os resíduos são salvos para posterior análise de anomalias, sendo os resíduos obtidos a partir da seguinte expressão:

$$\epsilon_k = \mathbf{y}_k - \hat{\mathbf{y}}_k \quad (4.1)$$

onde:

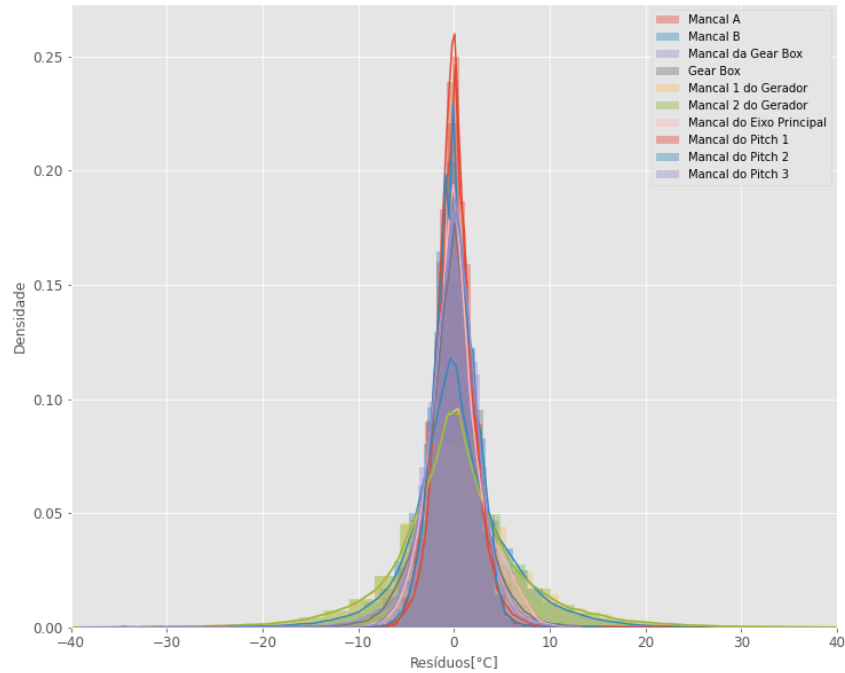
- k é o componente em questão, $k = 1, 2, \dots, 10$;
- \mathbf{y} é o valor medido;
- $\hat{\mathbf{y}}$ é o valor predito.

Os resíduos foram então salvos em um arquivo “csv” para serem analisados e posteriormente identificadas as anomalias.

4.2.2 Análise dos Outliers

Com os resíduos computados para cada componente, eles foram analisados visualmente para verificar se os mesmos seguiam uma distribuição que se assemelha com a normal, para então ser calculado a média $\bar{\epsilon}_k$ e o desvio padrão σ_k de cada componente. As distribuições de todos os componentes de um aerogerador são apresentadas na Figura 9.

Figura 9 – Distribuição dos resíduos das temperaturas de todos os componentes modelados para um aerogerador



Fonte: Elaborado utilizando Google Colab [BISONG, 2019]

Após calculados os parâmetros (média e desvio padrão) dos resíduos, foi utilizado o critério ou “*rule of thumb*” de que qualquer ponto que estiver mais do que três desvios padrões acima da média será considerado uma anomalia. Ou seja:

$$\begin{cases} \epsilon \text{ é um anomalia,} & \text{se } (\epsilon - \bar{\epsilon}_k) > 3 \cdot \sigma_k, \forall \epsilon \in \epsilon_k \\ \epsilon \text{ não é uma anomalia,} & \text{caso contrário} \end{cases} \quad (4.2)$$

onde:

- ϵ é um ponto pertencente aos resíduos do componente k ;
- ϵ_k são todos os resíduos de um componente k ;
- k é o componente em questão, $k = 1, 2, \dots, 10$.

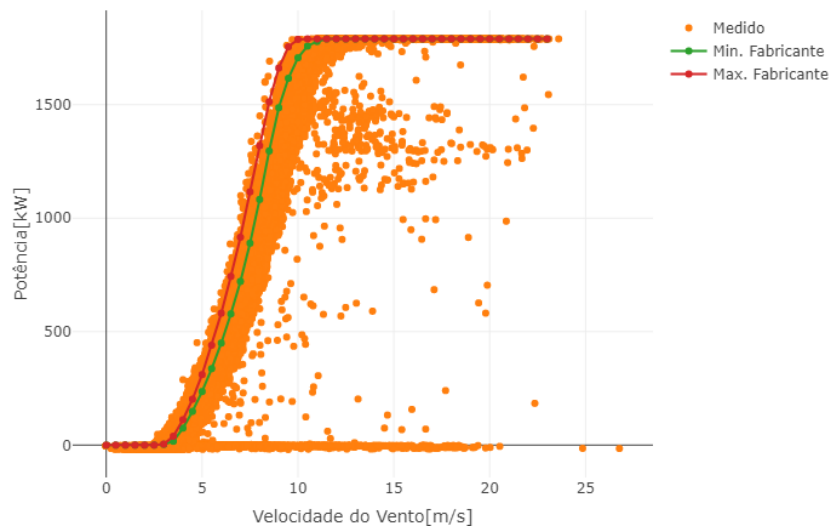
Além do critério dos três desvios padrões acima da média, um outro critério foi adotado, o qual leva em consideração se o ponto anterior (em ordem temporal) também foi considerado uma anomalia. Assim, um ponto será considerado uma anomalia se o seu ponto anterior também o for.

Com as anomalias identificadas, foram computadas as curvas de anomalias ou *outliers* acumulados para cada componente a fim de se relacionar os acúmulos de *outliers* com possíveis falhas.

4.3 Limites para Curva S

Como mencionado anteriormente, a Curva S é uma curva característica da geração eólica. Assim, essa parte do trabalho tem como objetivo encontrar limites operacionais com base nos próprios dados do aerogerador, visto que as curvas fornecidas pelos fabricantes nem sempre são adequadas para se utilizar como limites operacionais. Isso pode ser observado na Figura 10, onde 82,48% dos dados se encontram abaixo do mínimo das curvas fornecidas pelo fabricante.

Figura 10 – Dados medidos e curvas de potência máxima e mínima fornecida pelo fabricante em relação à Curva S para um aerogerador.



Fonte: Elaborado utilizando Rstudio [RStudio Team, 2015]

Primeiramente, tentou-se realizar uma abordagem similar à encontrada na literatura [SOHONI; GUPTA; NEMA, 2016], a qual envolve o treinamento de algoritmos de regressão e então o uso de uma métrica como 3 desvios padrões nos resíduos para classificação dos *outliers*. Porém, devido à grande quantia de anomalias presentes, o próprio desvio padrão é influenciado, fazendo com que os limites fossem visivelmente ruins. Os aglomerados de anomalias podem ser verificados na Figura 10, como por exemplo para velocidade do vento acima de 10 m/s e potência entre 1000kW e 1500kW.

4.3.1 Filtragem

Devido ao problema da grande quantia de anomalias, foi decidido realizar primeiramente uma filtragem nos dados para então definir os limites para Curva S. Para realizar a filtragem optou-se por utilizar o modelo de regressão *Gradient Tree Boosting Regression* (disponibilizado pela biblioteca Sklearn [PEDREGOSA et al., 2011]) para predição de

intervalos de regressão para a potência, em conjunto com a mediana da distância entre os percentis 20% e 80% para velocidade do vento, os dados de projeto disponibilizados pelo fabricante e uma estimativa do erro de medição.

4.3.1.1 Estimação da Distância Entre o Percentil 20% e 80%

Como pode ser visto na Figura 10, existem certos intervalos de velocidade do vento para os quais a potência permanece constante, podendo ser interpretado como uma “largura” da curva para cada valor de potência medida. Ainda na Figura 10 é possível ver que essa “largura” é variável, sendo consideravelmente alterada para potências próximas de zero a potências próximas da nominal, que nesse caso é de 1790kW.

Para estimar essa “largura” foi definido como uma *rule of thumb* a divisão da potência em 300 intervalos, fazendo com que cada intervalo tenha uma amplitude de menos de 0.5% em relação a potência nominal, uma vez que

$$\frac{1790 - 0}{300} \cdot \frac{1}{1790} = 0.00333... \simeq 0.33\%.$$

A partir dos intervalos onde a potência foi considerada aproximadamente constante por ter uma amplitude máxima de 0.33% em relação à potência nominal, foi computada a distância entre os percentis 20% e 80% da velocidade do vento e então tomada a mediana das distâncias.

A distância entre os percentis 20% e 80% foi propositalmente utilizada diferentemente do que é encontrada na literatura. Normalmente se utiliza a distância entre os percentis 25% e 75% (primeiro e terceiro quartil), como por exemplo na construção de Diagramas de Caixa (*Box Plots*) [MONTGOMERY; RUNGER; HUBELE, 2009]. Porém, devido ao fato dessa etapa ser uma etapa de filtragem e não da determinação dos limites em si, foi utilizada essa distância maior entre os percentis.

4.3.1.2 Estimativa do Erro de Medição

Para estimar o erro de medição, utilizou-se a diferença entre o valor máximo e mínimo da potência para as velocidades do vento abaixo da velocidade de *cut in* menos a mediana da distância entre os percentis 20% e 80%, como mostrado na seguinte expressão:

$$P_{error} = \max P(v < cut_in) - \min P(v < cut_in) \quad (4.3)$$

em que:

- P_{error} é o valor estimado do erro de medição;
- max e min são os operadores de máximo e mínimo respectivamente;

- P é a potência em kW;
- cut_in é o valor da velocidade do vento para qual a potência deveria ser zero segundo a especificação do projeto do aerogerador, nesse caso $cut_in = 3$ [m/s].

4.3.1.3 Pré Filtragem

Uma pré filtragem foi realizada nos dados brutos para ajudar no desempenho dos modelos de predição do intervalo de regressão para potência. Essa pré filtragem foi feita de acordo com a seguinte expressão:

$$\begin{cases} P(v > cut_in + 1.5 \cdot wiq) > 0, \\ P(v > 15) > 1790 - P_{error}/2, \\ v < cut_off + 1.5 \cdot wiq \end{cases} \quad (4.4)$$

na qual:

- P é a potência em kW;
- cut_in é o valor da velocidade do vento para qual a potência deveria ser zero segundo a especificação do projeto do aerogerador, nesse caso $cut_in = 3$ [m/s];
- wiq é a mediana da distância entre os percentis 20% e 80% da velocidade do vento;
- P_{error} é o valor estimado do erro de medição da potência;
- cut_off é o valor da velocidade do vento a partir da qual o aerogerador deve ser desligado segundo a especificação do projeto do aerogerador, nesse caso $cut_off = 23$ [m/s].

4.3.1.4 Estimação do Intervalo de Regressão

O intervalo de regressão pode ser entendido como a predição de dois percentis dos dados. Por exemplo, se quisermos estabelecer entre quais valores encontramos 90% dos dados centrais podemos estimar o percentil 5% e o percentil 95%.

O valor dos percentis para os intervalos de regressão foram definidos empiricamente. A escolha foi feita observando o comportamento dos intervalos de regressão para diversos valores, aumento gradativamente de 5% até 95% e verificando visualmente para qual valor os limiares paravam de ser influenciados pelas anomalias.

A partir dessa análise visual, foi selecionado um intervalo de regressão de 20%. Tal valor a priori pode parecer pequeno, porém além do modelo para predição do intervalo de regressão da potência, também foi utilizada a mediana da distância entre os percentis 20% e 80%, resultando em limiares maiores para filtragem dos dados.

4.3.1.5 Filtragem Automática

Utilizando conjuntamente os modelos de predição do intervalo de regressão para potência, a mediana da distância entre os percentis 20% e 80% da velocidade do vento e a estimativa do erro de medição, foi criado um método automático para filtragem, sem que seja necessária inserção manual dos parâmetros.

O método se baseia em criar um valor de filtragem inferior e superior definidos da seguinte forma:

$$\begin{cases} f_{inf} = \hat{f}_{inf}(v - 1.5 \cdot wiq) - P_{error}/2 \\ f_{sup} = \hat{f}_{sup}(v + 1.5 \cdot wiq) + P_{error}/2 \end{cases} \quad (4.5)$$

em que:

- f_{inf} e f_{sup} são os valores limites inferiores e superiores para se realizar a filtragem dos dados;
- \hat{f}_{inf} e \hat{f}_{sup} são os modelos de regressão (*Gradient Tree Boosting*) treinados para predizer o valor inferior e superior do intervalo de regressão (no caso de 20%);
- v é a velocidade do vento;
- wiq é a mediana da distância entre os percentis 20% e 80% da velocidade do vento;
- P_{error} é o valor estimado do erro de medição.

Tendo os valores inferiores e superiores estabelecidos para filtragem, os dados foram filtrados respeitando estes, assim como as velocidades de cut_in e cut_off , de acordo com a seguinte expressão:

$$\begin{cases} f_{inf} \leq P \leq f_{sup}, \\ P(v > cut_in + 1.5 \cdot wiq) > 0, \\ v < cut_off + 1.5 \cdot wiq \end{cases} \quad (4.6)$$

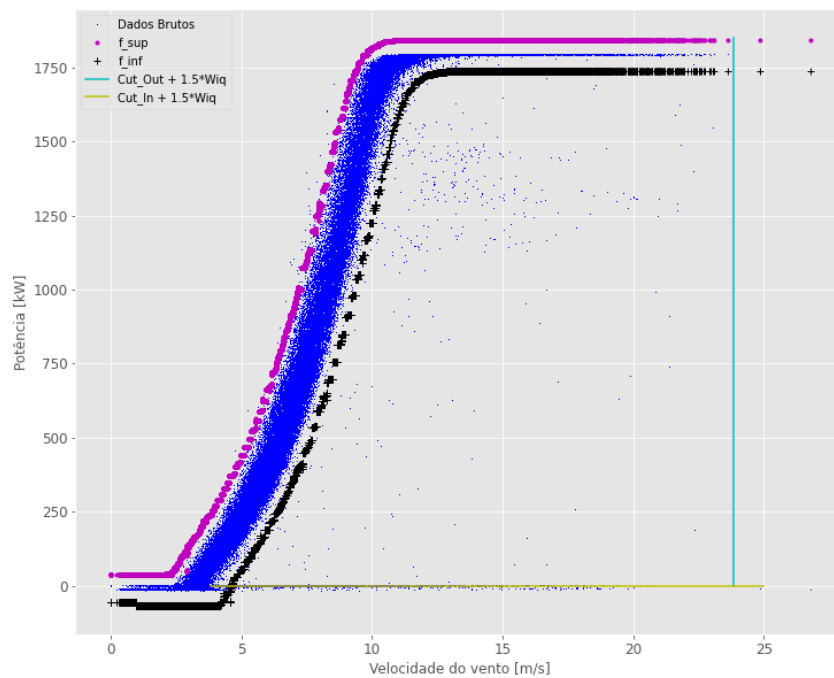
onde:

- f_{inf} e f_{sup} são os valores limites inferiores e superiores para se realizar a filtragem dos dados;
- cut_in é o valor da velocidade do vento para qual a potência deveria ser zero segundo a especificação do projeto do aerogerador, nesse caso $cut_in = 3$ [m/s];
- wiq é a mediana da distância entre os percentis 20% e 80% da velocidade do vento;

- cut_off é o valor da velocidade do vento a partir da qual o aerogerador deve ser desligado segundo a especificação do projeto do aerogerador, nesse caso $cut_off = 23$ [m/s].

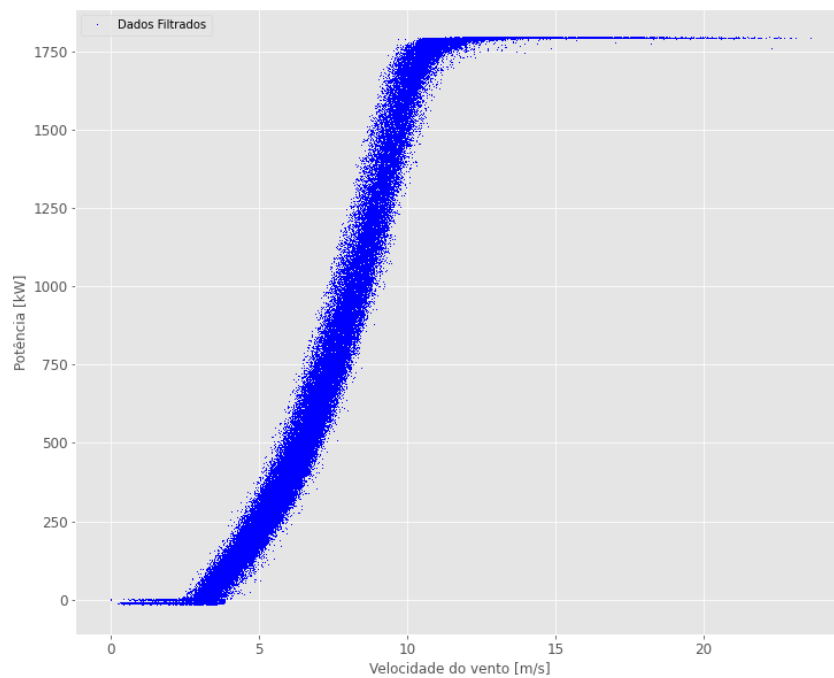
Os limiares para filtragem assim como os dados filtrados podem ser observados nas Figuras 11 e 12.

Figura 11 – Limiares de filtragem de um aerogerador. Indicando os dados brutos, os valores superiores e inferiores do método de filtragem automático (f_sup e f_inf), os valores de potência acima de zero após a velocidade de cut_in mais 1.5 vezes W_{iq} e os valores anteriores a velocidade de cut_off mais 1.5 vezes W_{iq} . Sendo W_{iq} a mediana da distância entre os percentis 20% e 80%.



Fonte: Elaborado utilizando Google Colab [BISONG, 2019]

Figura 12 – Dados Filtrados, após o procedimento descrito na Subseção 4.3.1.5 para um aerogerador



Fonte: Elaborado utilizando Google Colab [BISONG, 2019]

4.3.2 Método para Determinação dos Limites para Curva S

Os passos para determinação dos limites para Curva S foram basicamente os mesmos utilizados para obtenção da filtragem automática dos dados. A única diferença é que os dados utilizados foram os dados **filtrados**.

4.3.2.1 Estimação da Distância Interquartil

A estimação da distância interquartil (distância entre o percentil 25% e 75% também conhecidos como primeiro e terceiro quartil) seguiu os mesmos passos da Subseção 4.3.1.1, dividindo-se a potência dos dados filtrados em 300 intervalos e então calculada a mediana da distância interquartil da velocidade do vento de todos os intervalos. A distância interquartil aqui é a mesma presente na literatura como por exemplo na construção de Diagramas de Caixa (*Box Plots*) [MONTGOMERY; RUNGER; HUBELE, 2009].

4.3.2.2 Estimação do Intervalo de Regressão

A estimação do intervalo de regressão seguiu os mesmos passos da Subseção 4.3.1.4 porém usando um percentual de 50%. Novamente foi treinado o modelo Gradient Tree

Boosting Regression para predição do limite inferior e superior do intervalo de regressão para potência.

4.3.2.3 Limites e Outliers

Tendo os modelos para predição do intervalo de regressão treinados com os dados filtrados, a mediana da distância interquartil para velocidade do vento e a estimativa do erro de medição computados, os limites foram definidos similarmente ao método de filtragem da Subseção 4.3.1.5, conforme apresentado na seguinte expressão:

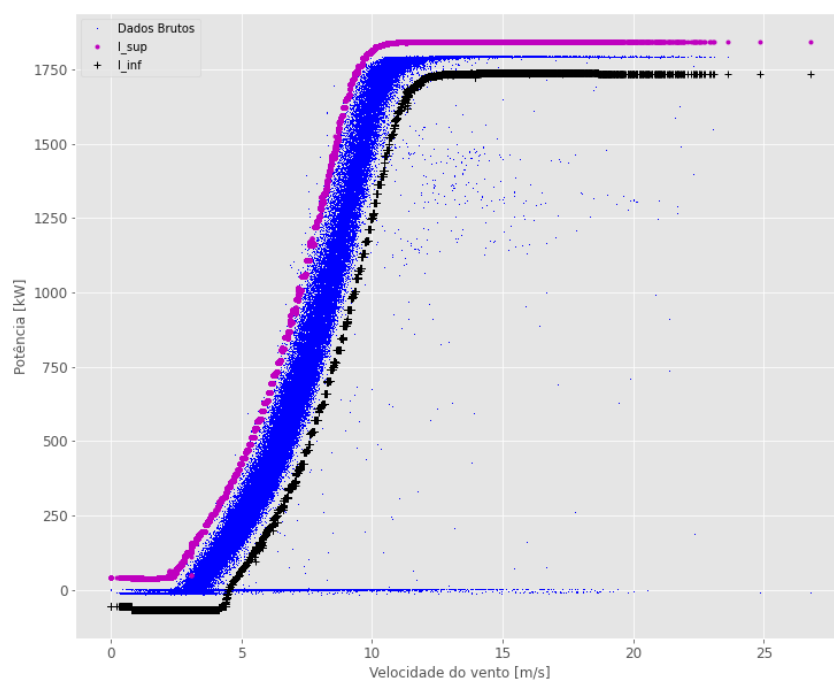
$$\begin{cases} l_{inf} = \hat{l}_{inf}(v - 1.5 \cdot \delta) - P_{error}/2 \\ l_{sup} = \hat{l}_{sup}(v + 1.5 \cdot \delta) + P_{error}/2 \end{cases} \quad (4.7)$$

onde:

- l_{inf} e l_{sup} são os valores limites inferiores e superiores para classificação dos *outliers* da Curva S;
- \hat{l}_{inf} e \hat{l}_{sup} são os modelos de regressão (*Gradient Tree Boosting*) treinados para prever o valor inferior e superior do intervalo de regressão dos dados filtrados (no caso de 50%);
- v é a velocidade do vento;
- δ é a mediana da distância interquartil da velocidade do vento dos dados filtrados;
- P_{error} é o valor estimado do erro de medição.

Todo e qualquer ponto que se encontrasse fora dos limites $l_{inf} \leq P(v) \leq l_{sup}$ foi considerado um *outlier*. Os valores limites junto com todos os dados do aerogerador podem ser vistos na Figura 13

Figura 13 – Limites para Classificação de *Outliers* e todos os dados (Dados Brutos) de um aerogerador.



Fonte: Elaborado utilizando Google Colab [BISONG, 2019]

5 Resultados

A metodologia descrita no Capítulo 4 foi aplicada em oito aerogeradores do mesmo modelo e do mesmo fabricante para um parque eólico real. Os resultados em relação às Curvas S e aos *outliers* de todos os componentes para todos os aerogeradores serão apresentados ao longo deste capítulo.

Uma ressalva é feita em relação aos componentes:

- Mancal do Pitch 1;
- Mancal do Pitch 2;
- Mancal do Pitch 3;
- Mancal do Eixo Principal.

Esses componentes, devido a um problema de aquisição de dados, possuem aproximadamente 22% a menos de observações em relação ao restante dos componentes. Mesmo assim, devido à metodologia e a uma implementação robusta a *missing data*, foi possível modelar esses componentes. Porém, é possível que esse volume a menos de dados reflita no ajuste dos modelos e consequentemente no comportamento dos resíduos.

5.1 Resultados da Curva S

Uma pequena observação em relação à Curva S de todas as máquinas é que potência igual ou próxima a zero para velocidades do vento acima de 3 [m/s] não significam necessariamente defeito. Em períodos de manutenção, por exemplo, a máquina fica inoperante temporariamente e não gera potência de saída mesmo tendo velocidade de vento suficiente.

5.1.1 Aerogerador 001

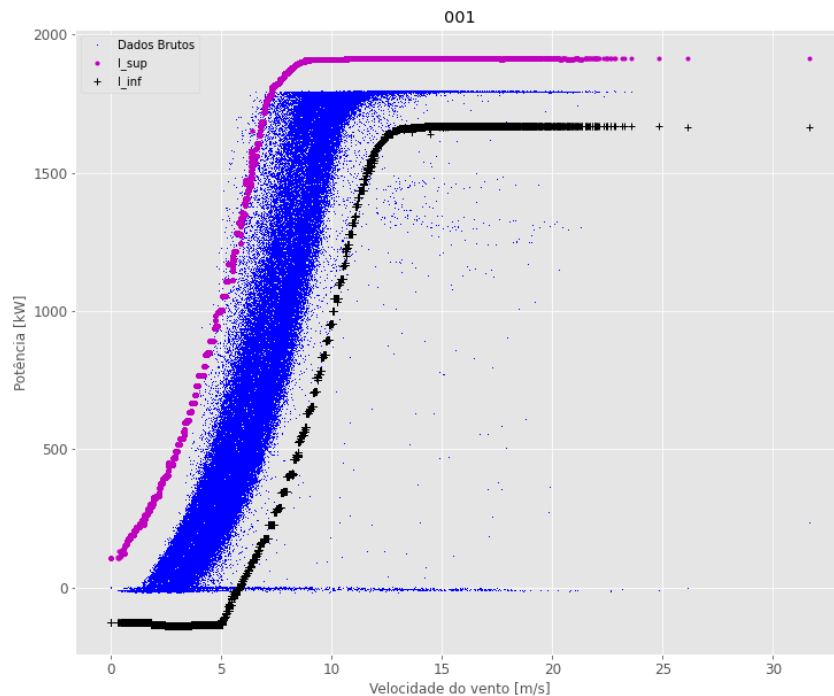
O Aerogerador 001 apresentou um comportamento diferente dos demais. A “largura” da curva no intervalo de velocidade do vento entre 3 [m/s] e 11 [m/s] é expressiva, como pode ser observado na Figura 14.

Em relação a essa “largura”, foi identificado o ano em que cada ponto entre os limites l_{inf} e l_{sup} ocorreu. Com isso foi verificado que entre 2018 e 2020 esse aerogerador foi perdendo seu desempenho. Visto que para 2018 a mesma velocidade do vento gerava

mais potência que para os anos seguintes na região abaixo da potência nominal. Esse comportamento pode ser visto na Figura 15.

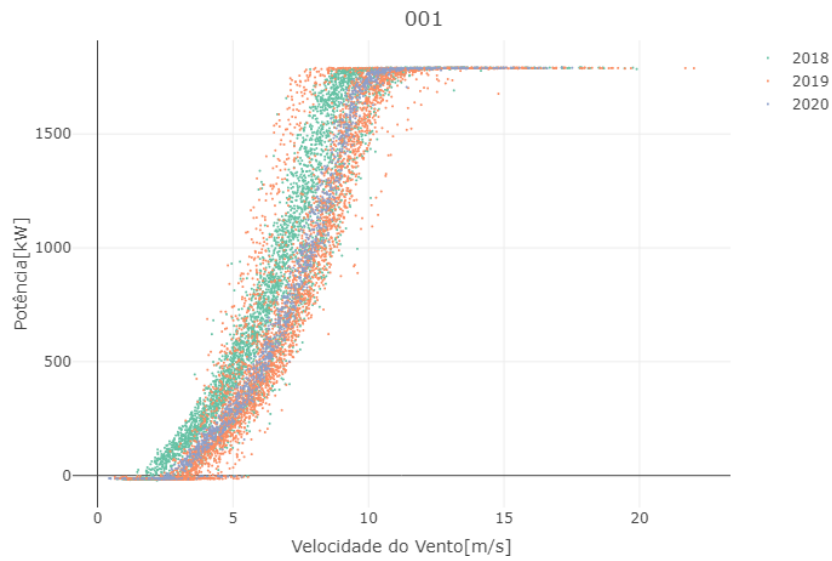
Analisando a região abaixo de l_{inf} na Figura 14, é possível ver uma aglomeração de pontos entre 1000 [kW] e 1500 [kW]. Para essa região, foi verificado em qual período de tempo se concentrava essa aglomeração, e então verificado a relação com os *Logs*. É possível verificar na Figura 16 que os pontos se concentram em 12/02/2019 e 05/04/2019 e que após o *Log* em 09/04/2019, o qual indica que foi realizada uma manutenção preventiva anual no aerogerador, essas aglomerações não voltaram a ocorrer.

Figura 14 – Curva S do Aerogerador 001 com o limite inferior (l_{inf}) e superior (l_{sup}).



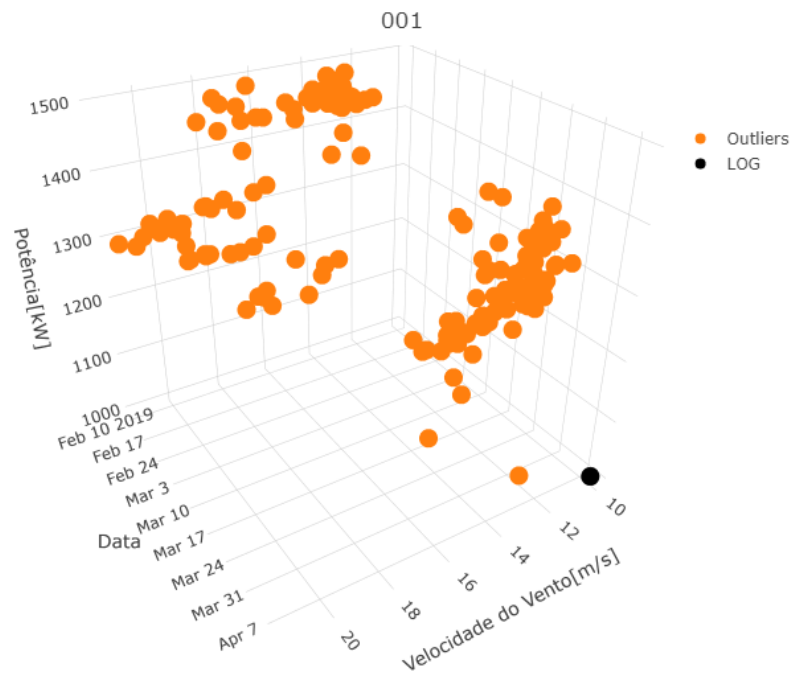
Fonte: Elaborado utilizando Google Colab [BISONG, 2019]

Figura 15 – Pontos da Curva S do Aerogerador 001 entre os limite inferior (l_{inf}) e superior (l_{sup}) para os anos de 2018, 2019 e 2020.



Fonte: Elaborado utilizando Rstudio [RStudio Team, 2015]

Figura 16 – Aerogerador 001. Visualização temporal dos pontos aglomerados entre 1000 kW e 1500 kW nas datas de 12/02/2019 e 05/04/2019 e registro de *Log* de manutenção preventiva anual em 09/04/2019.



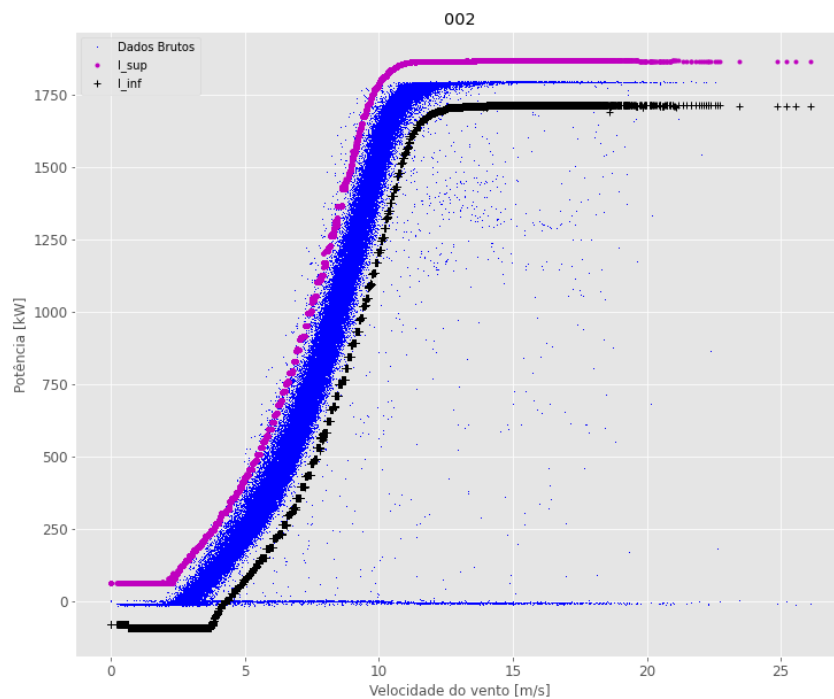
Fonte: Elaborado utilizando Rstudio [RStudio Team, 2015]

5.1.2 Aerogerador 002

O Aerogerador apresenta curvas mais “estritas”. Entretanto, como pode ser observado na Figura 17, ele apresenta vários pontos dispersos abaixo do limite inferior.

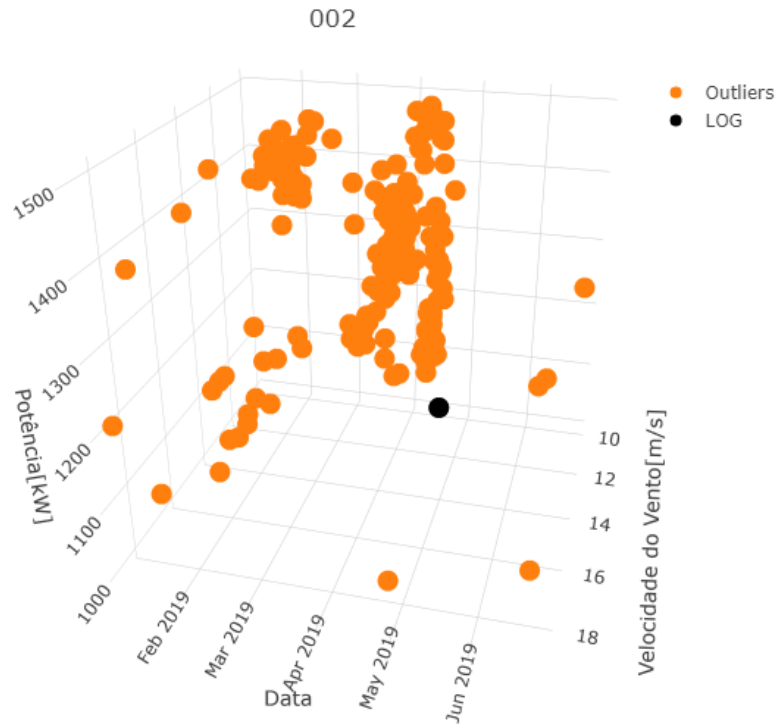
Em relação aos pontos abaixo do limite inferior, foi analisado temporalmente se existiam aglomerações em alguma região temporal e de potência. Coincidentemente, foram encontradas aglomerações de pontos entre 1000 kW e 1500 kW nas datas de 12/02/2019 e 05/04/2019, e após o *Log* em 15/04/2019, o qual indica a realização de uma manutenção preventiva anual no aerogerador, essas aglomerações não voltaram a ocorrer. Entretanto ainda ocorreram pontos dispersos nessa faixa de potência após a manutenção. Todas essas informações podem ser observadas na Figura 18.

Figura 17 – Curva S do Aerogerador 002 com o limite inferior (l_{inf}) e superior (l_{sup}).



Fonte: Elaborado utilizando Google Colab [BISONG, 2019]

Figura 18 – Aero gerador 002. Visualização temporal dos pontos aglomerados entre 1000 kW e 1500 kW nas datas de 12/02/2019 e 05/04/2019 e registro de *Log* de manutenção preventiva anual em 15/04/2019.



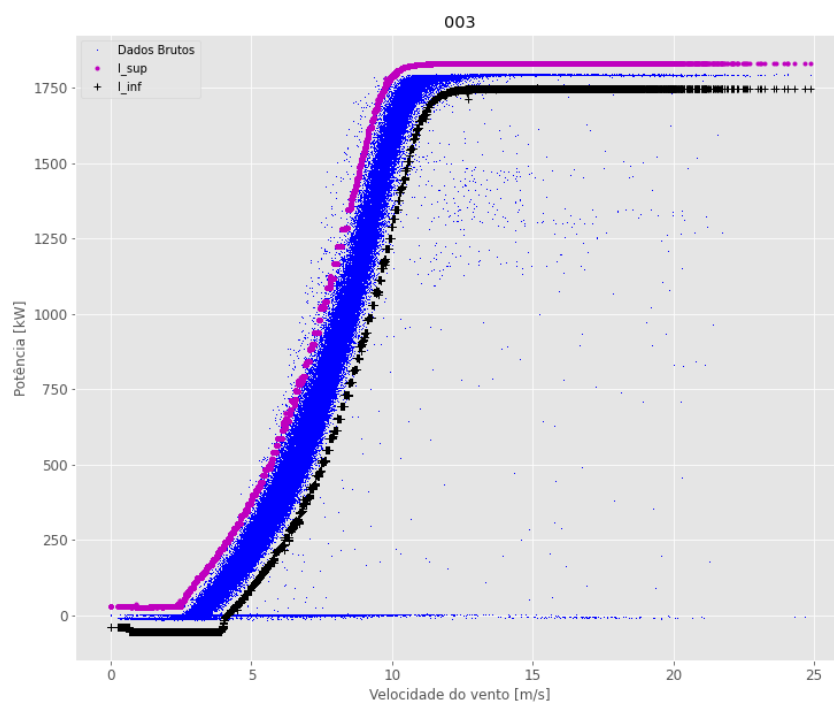
Fonte: Elaborado utilizando Rstudio [RStudio Team, 2015]

5.1.3 Aero gerador 003

Esse aerogerador apresenta uma Curva S relativamente mais “estreita” comparado aos outros e menos pontos abaixo da limite inferior se comparado com o Aero gerador 002, como pode ser visto na Figura 19.

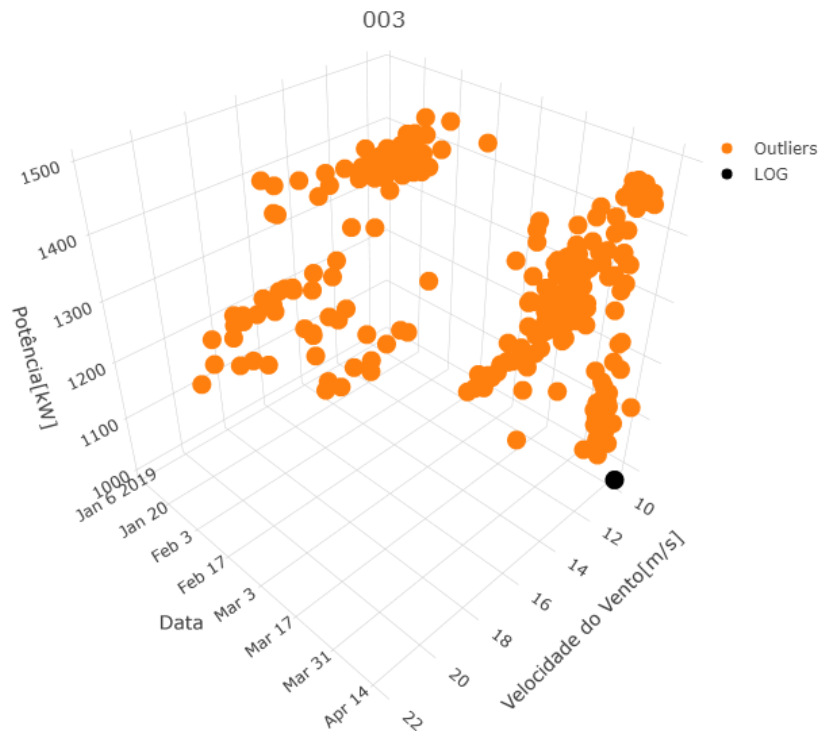
Novamente, como pode ser visto na Figura 20, foi encontrada uma aglomeração de pontos entre 1000 kW e 1500 kW nas datas 12/02/2019, 05/04/2019 e 16/04/2019. Essas aglomerações não apareceram mais depois da manutenção preventiva realizada em 16/04/2019.

Figura 19 – Curva S do Aerogerador 003 com o limite inferior (l_{inf}) e superior (l_{sup}).



Fonte: Elaborado utilizando Google Colab [BISONG, 2019]

Figura 20 – Aerogerador 003. Visualização temporal dos pontos aglomerados entre 1000 kW e 1500 kW nas datas de 12/02/2019, 05/04/2019 e 16/04/2019 e registro de *Log* de manutenção preventiva anual em 16/04/2019.



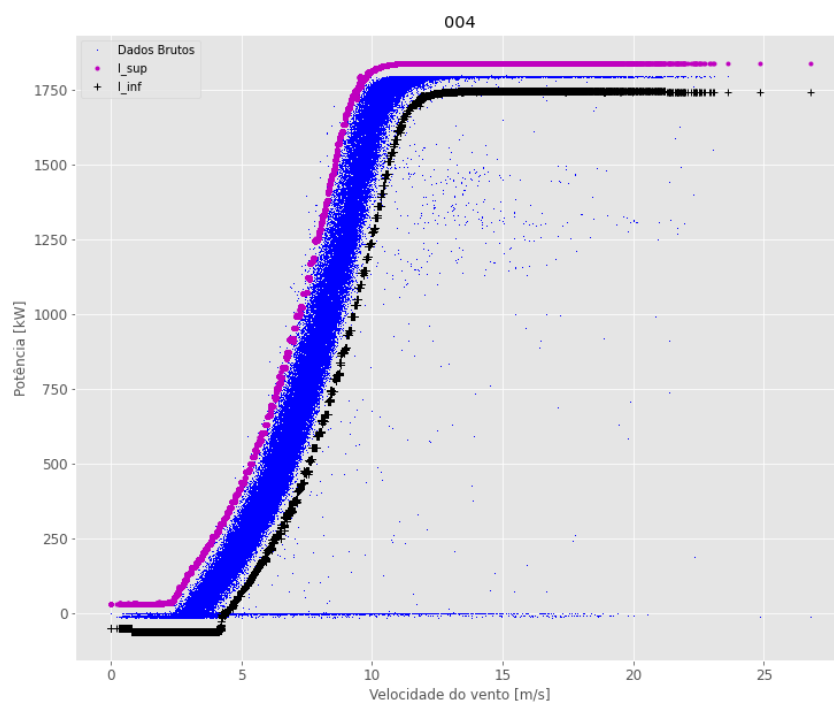
Fonte: Elaborado utilizando Rstudio [RStudio Team, 2015]

5.1.4 Aerogerador 004

O comportamento do aerogerador 004 é muito similar ao do aerogerador 003 e também apresenta uma aglomeração de pontos entre 1000 kW e 1500 kW como pode ser visto na Figura 21.

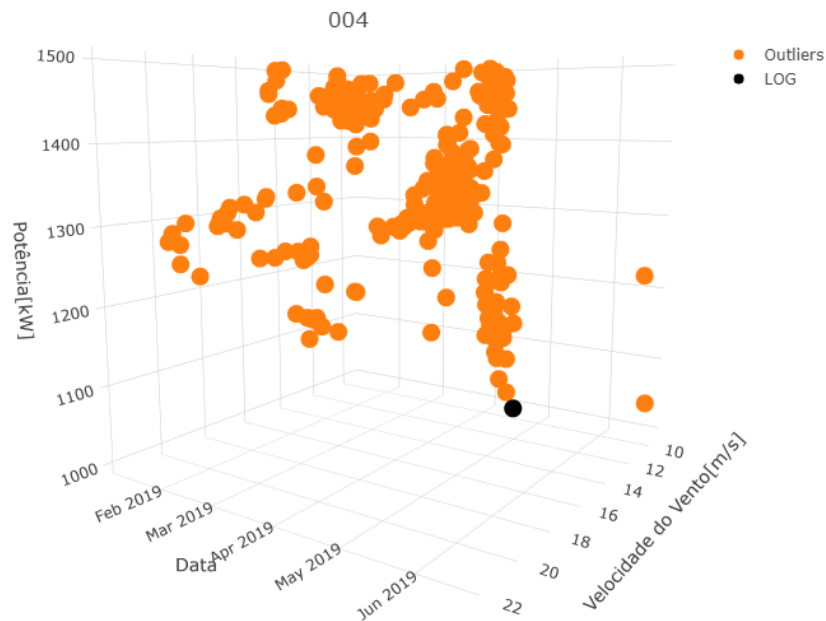
Novamente para o Aerogerador 004 foi encontrada uma aglomeração de pontos entre 1000 kW e 1500 kW nas datas 12/02/2019, 05/04/2019 e 16/04/2019. No dia 17/04/2019 consta um *Log* de manutenção preventiva anual. A partir dessa data não apareceram mais aglomerações. Entretanto alguns pontos dispersos continuam aparecendo, como pode ser visto na Figura 22.

Figura 21 – Curva S do Aerogerador 004 com o limite inferior (l_{inf}) e superior (l_{sup}).



Fonte: Elaborado utilizando Google Colab [BISONG, 2019]

Figura 22 – Aerogerador 004. Visualização temporal dos pontos aglomerados entre 1000 kW e 1500 kW nas datas de 12/02/2019, 05/04/2019 e 16/04/2019 e registro de *Log* de manutenção preventiva anual em 17/04/2019.



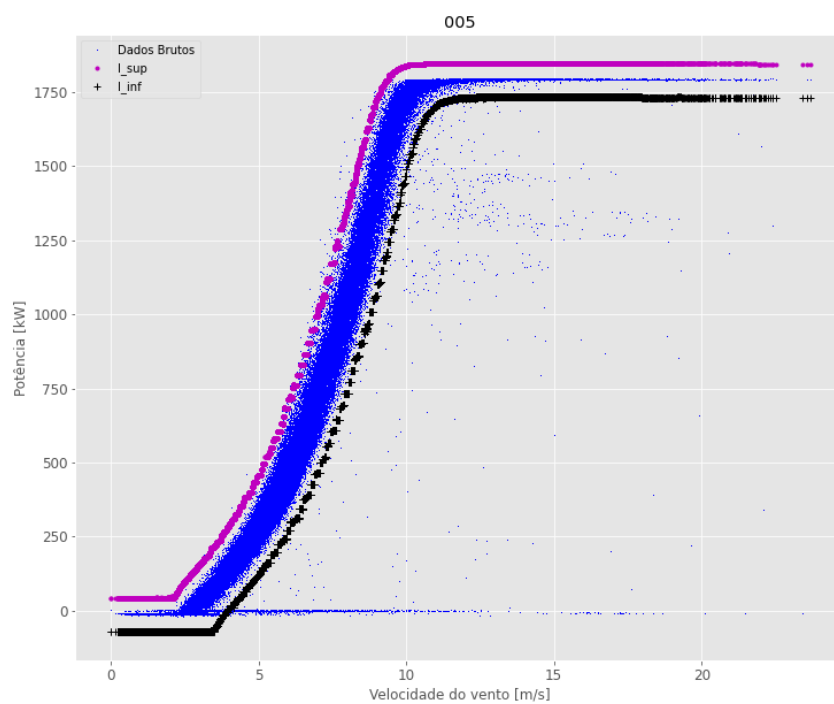
Fonte: Elaborado utilizando Rstudio [RStudio Team, 2015]

5.1.5 Aerogerador 005

Esse aerogerador apresenta um comportamento similar ao dos aerogeradores 002, 003 e 004 com aglomeração de pontos entre 1000 kW e 1500 kW como pode ser visto na Figura 23.

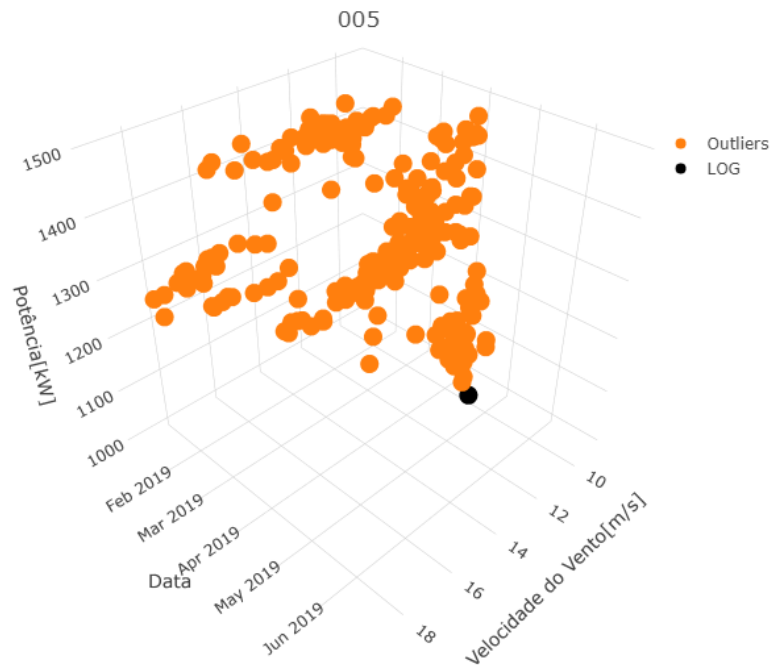
Tais aglomerações se encontram nas datas 12/02/2019, 05/04/2019 e 16/04/2019. No dia 18/04/2019 consta um *Log* de manutenção preventiva anual e a partir dessa data não apareceram mais aglomerações. Entretanto, alguns pontos dispersos continuam aparecendo, como pode ser visto na Figura 24.

Figura 23 – Curva S do Aerogerador 005 com o limite inferior (l_{inf}) e superior (l_{sup}).



Fonte: Elaborado utilizando Google Colab [BISONG, 2019]

Figura 24 – Aerogerador 005. Visualização temporal dos pontos aglomerados entre 1000 kW e 1500 kW nas datas de 12/02/2019, 05/04/2019 e 16/04/2019 e registro de *Log* de manutenção preventiva anual em 18/04/2019.



Fonte: Elaborado utilizando Rstudio [RStudio Team, 2015]

5.1.6 Aerogerador 006

O Aerogerador apresenta um comportamento diferente dos demais, como é possível visualizar na Figura 25, existem “linhas” nas quais a potência permanece constante nos valores de 500 kW, 1000 kW e 1200 kW, mesmo com a variação da velocidade do vento. Isso indica que a máquina está sob alguma restrição e que não pode gerar mais potência mesmo com disponibilidade do vento.

Além das observações indicando restrições, é possível ver novamente o aglomerado de pontos entre 1000 kW e 1500 kW, esse comportamento pode ser observado na Figura 26. Nessa Figura em específico foram retirados os valores (linhas) indicando restrições nas potências de 1000 kW e 1200 kW para melhor visualização dos aglomerados nas datas de 12/02/2019, 05/04/2019 e 16/04/2019. Novamente esses aglomerados não ocorreram mais após o *Log* de manutenção preventiva no dia 16/04/2019, apesar de pontos dispersos continuarem aparecendo.

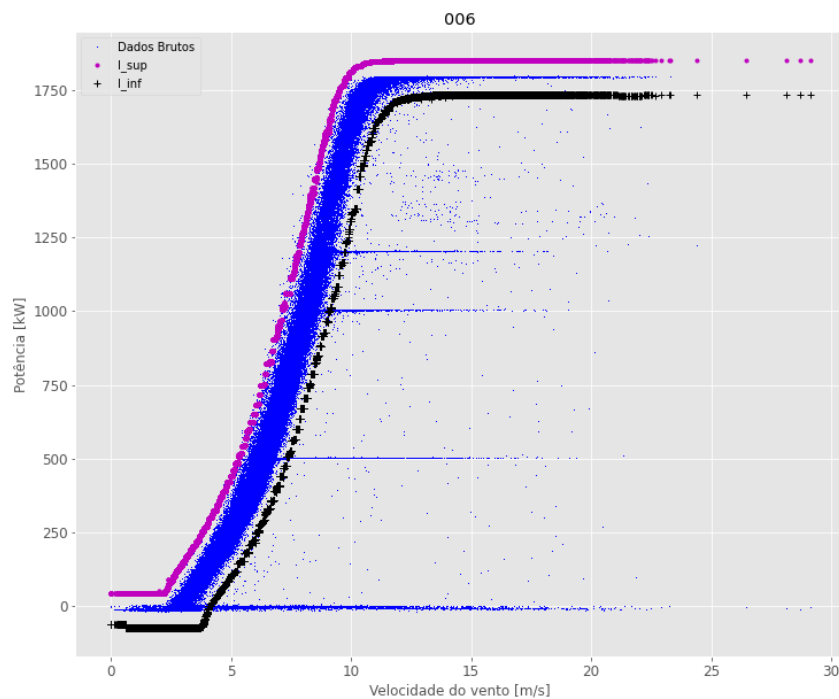
Quanto ao comportamento devido às restrições da máquina, é possível ver na Figura 27 os seguintes aspectos:

- Em 24/01/2018 a máquina se encontrava em restrição de 1200 kW de potência e en-

tão no dia 26/01/2018 consta um *Log* de reparo, onde foi substituído um acumulador de pressão de um freio hidráulico desse aerogerador;

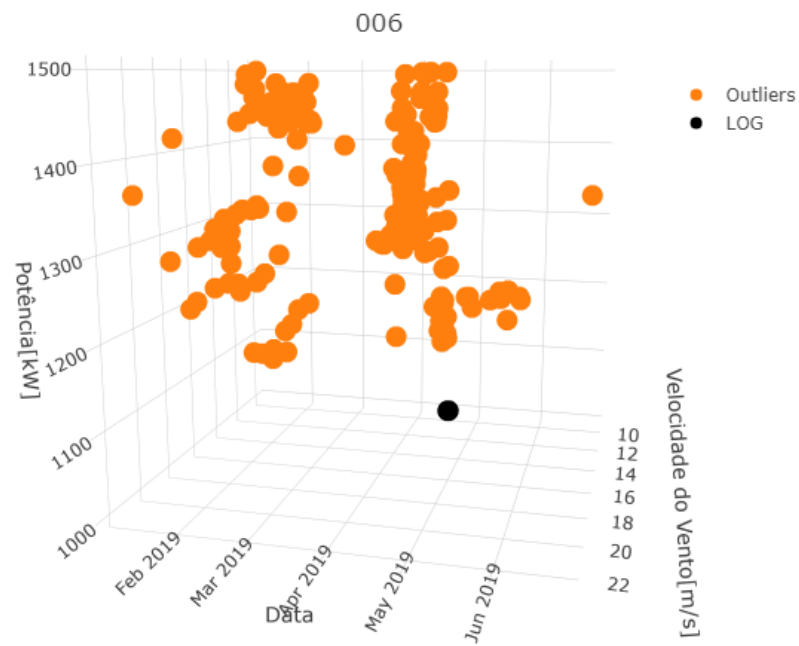
- Entre os dias 13/11/2018 a 24/11/2018, a máquina se encontrava em restrição de 1000 kW de potência, e entre os dias 28/11/2018 e 06/12/2018 se encontrava em restrição de 500 kW. Então no dia 07/12/2018 consta um *Log* de reparo, a partir do qual o aerogerador ficou inoperante até dia 12/12/2018 devido à necessidade de substituição da *Gearbox*, pois esta apresentou defeito;
- Entre os dias 26/04/2019 a 21/05/2019, a máquina se encontrava em restrição de 1200 kW de potência. No dia 16/05/2019 consta um *Log* **explicitando** que a máquina está em condição de restrição devido a componentes que estão descalibrados. No dia 03/06/2019 consta um *Log* de reparo onde foi substituído o freio hidráulico desse aerogerador.

Figura 25 – Curva S do Aerogerador 006 com o limite inferior (l_{inf}) e superior (l_{sup}).



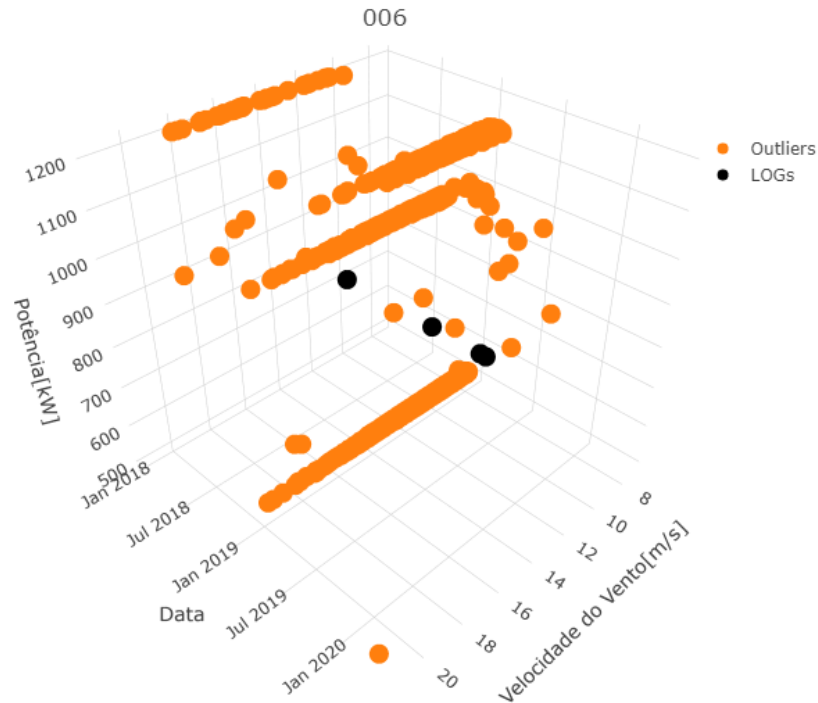
Fonte: Elaborado utilizando Google Colab [BISONG, 2019]

Figura 26 – Aerogerador 006. Visualização temporal dos pontos aglomerados entre 1000 kW e 1500 kW nas datas de 12/02/2019, 05/04/2019 e 16/04/2019 e registro de *Log* de manutenção preventiva anual em 16/04/2019. Nessa figura foram retirados as observações indicando restrições nas potências de 1000 kW e 1200 kW para melhor visualização dos aglomerados.



Fonte: Elaborado utilizando Rstudio [RStudio Team, 2015]

Figura 27 – Aerogerador 006. Visualização temporal das restrições em 500kW, 1000 kW e 1200 kW, e os registros de *Logs* em 26/01/2018, 07/12/2018, 16/05/2019 e 03/06/2019.



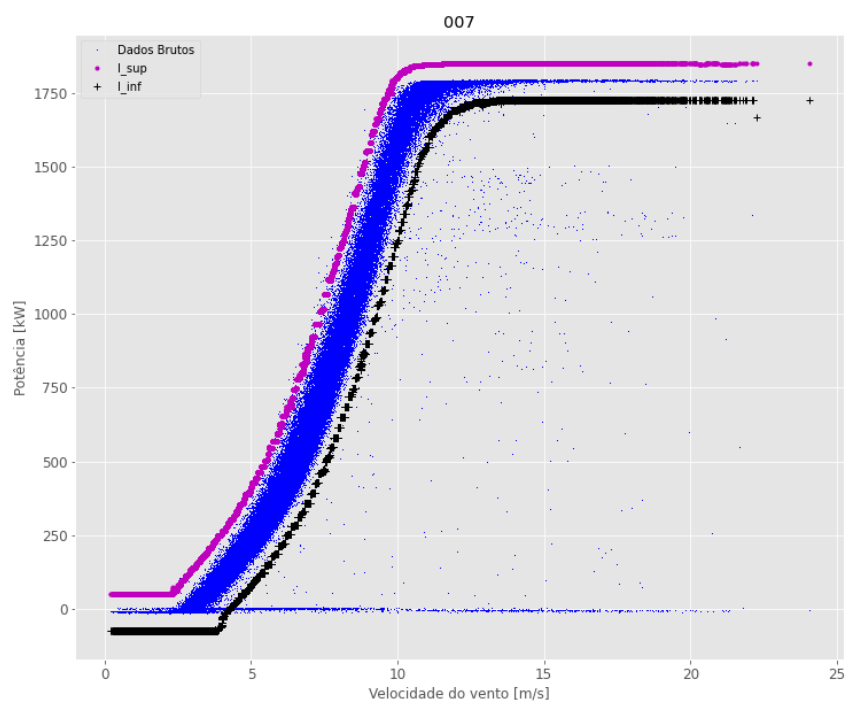
Fonte: Elaborado utilizando Rstudio [RStudio Team, 2015]

5.1.7 Aerogerador 007

O Aerogerador 007 apresenta um comportamento similar aos outros, com aglomerações de pontos entre 1000 kW e 1500 kW, como pode ser observado na Figura 28.

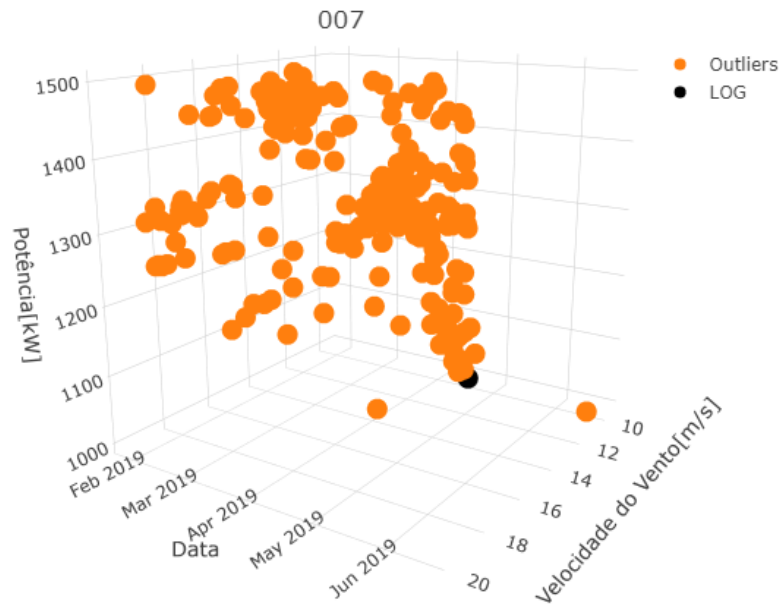
Como pode ser observado na Figura 29, essas aglomerações ocorreram nas mesmas datas dos aerogeradores anteriores, nos dias 12/02/2019, 05/04/2019 e 16/04/2019 e não foi mais observado esse comportamento após o *Log* de manutenção preventiva do dia 17/04/2019.

Figura 28 – Curva S do Aerogerador 007 com o limite inferior (l_{inf}) e superior (l_{sup}).



Fonte: Elaborado utilizando Google Colab [BISONG, 2019]

Figura 29 – Aerogerador 007. Visualização temporal dos pontos aglomerados entre 1000 kW e 1500 kW nas datas de 12/02/2019, 05/04/2019 e 16/04/2019 e registro de *Log* de manutenção preventiva anual em 17/04/2019.



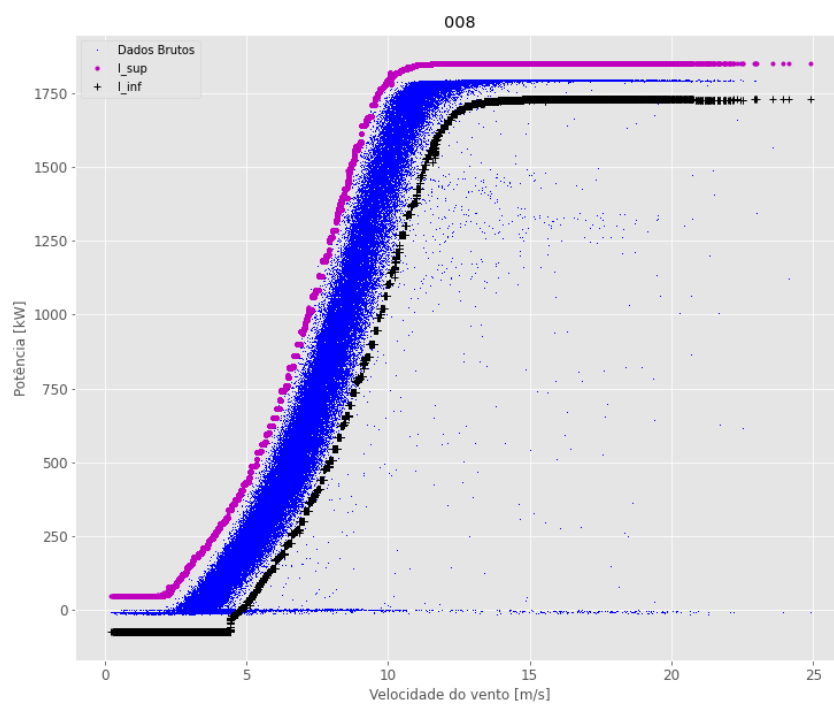
Fonte: Elaborado utilizando Rstudio [RStudio Team, 2015]

5.1.8 Aerogerador 008

O Aerogerador 008 apresenta um comportamento similar a todos os outros, com aglomerações de pontos entre 1000 kW e 1500 kW como pode ser observado na Figura 30.

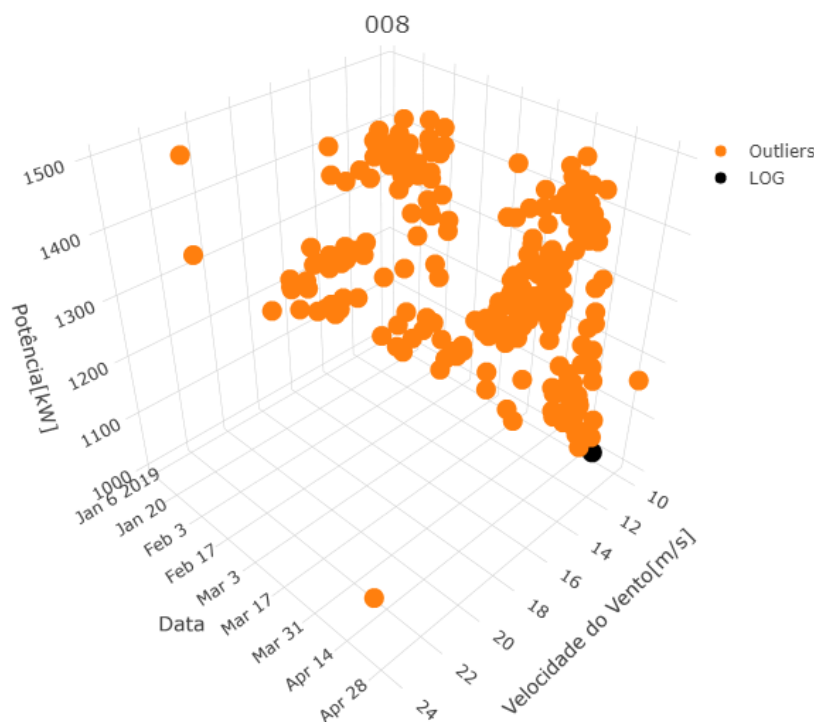
Novamente, essas aglomerações ocorreram nas mesmas datas dos aerogeradores anteriores, nos dias 12/02/2019, 05/04/2019 e 16/04/2019 e não foi mais observado esse comportamento após o *Log* de manutenção preventiva do dia 18/04/2019, como pode ser observado na Figura 31.

Figura 30 – Curva S do Aerogerador 008 com o limite inferior (l_{inf}) e superior (l_{sup}).



Fonte: Elaborado utilizando Google Colab [BISONG, 2019]

Figura 31 – Aerogerador 008. Visualização temporal dos pontos aglomerados entre 1000 kW e 1500 kW nas datas de 12/02/2019, 05/04/2019 e 16/04/2019 e registro de *Log* de manutenção preventiva anual em 18/04/2019.



Fonte: Elaborado utilizando Rstudio [RStudio Team, 2015]

5.2 Resultados dos *Outliers* de Temperatura

Como descrito na Seção 4.2, após utilizados os critérios para determinação dos *outliers*, foi computada a curva cumulativa de *outliers* para cada componente. Essas curvas serão apresentadas ao longo desta seção, sendo verificadas as variações abruptas do número de *outliers* e relacionadas essas variações com *Logs* posteriores de manutenção.

As curvas cumulativas serão apresentadas para os dez componentes de cada um dos oito aerogeradores, lembrando que os Mancais dos *Pitches* 1, 2 e 3 assim como o Mancal do Eixo Principal possuem em torno de 22% de dados a menos que os outros componentes. Sendo assim esses componentes estão em “desvantagem” em relação aos outros no que diz respeito a resultados mais robustos, e até em relação a diferenciações dos resíduos de uma distribuição normal.

Para diminuir o número de gráficos a serem analisados e facilitar as visualizações, as curvas de *outliers* acumulados serão apresentadas em grupos, reunidos da seguinte forma:

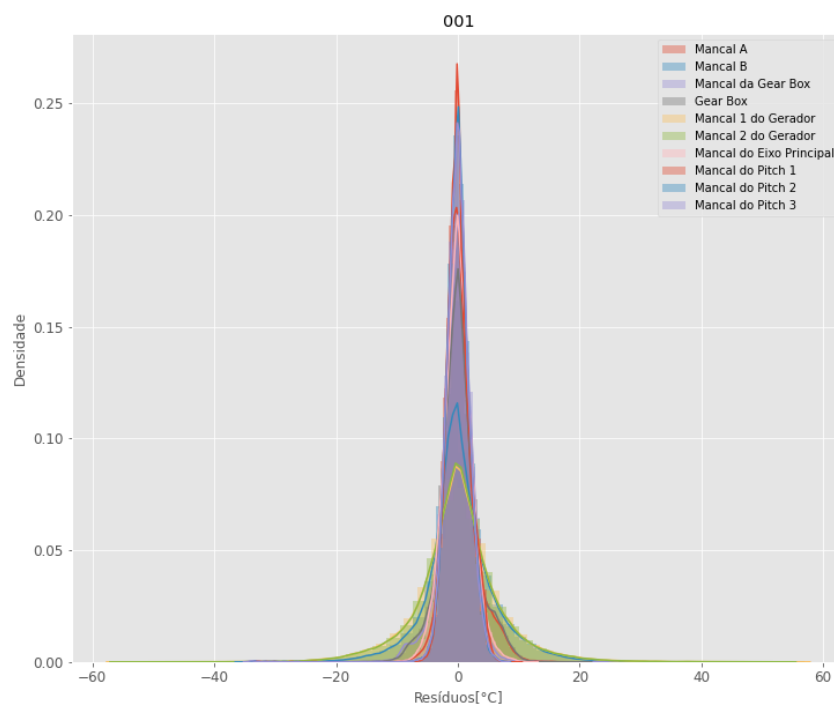
- Mancais: Mancais A, B e do Eixo Principal;
- *Gear Box*: Mancais da Gear Box e Gear Box;
- Gerador: Mancais 1 e 2 do Gerador;
- *Pitch*: Mancais dos *Pitches* 1, 2 e 3.

Como comentado adiante neste capítulo, alguns resíduos apresentam valores muito altos, chegando a 800°C de erro. Esses erros tão altos se devem aos valores reais que realmente estão na casa das centenas de °C de temperatura. Para impedir que valores assim influenciem a média e o desvio padrão na identificação dos *outliers*, esses parâmetros foram calculados apenas para resíduos abaixo de 60°C. Esse valor foi estipulado através de uma heurística vinda da observação da distribuição dos resíduos de todos os componentes para todas as máquinas.

5.2.1 Aero gerador 001

O Aero gerador 001 é um dos quais apresenta resíduos com valores acima de 800°C. Entretanto, para melhor visualização da distribuição dos resíduos, o gráfico foi truncado entre -60°C e 60°C para todos os componentes. Como pode ser observado na Figura 32.

Figura 32 – Distribuição dos resíduos do Aero gerador 001.



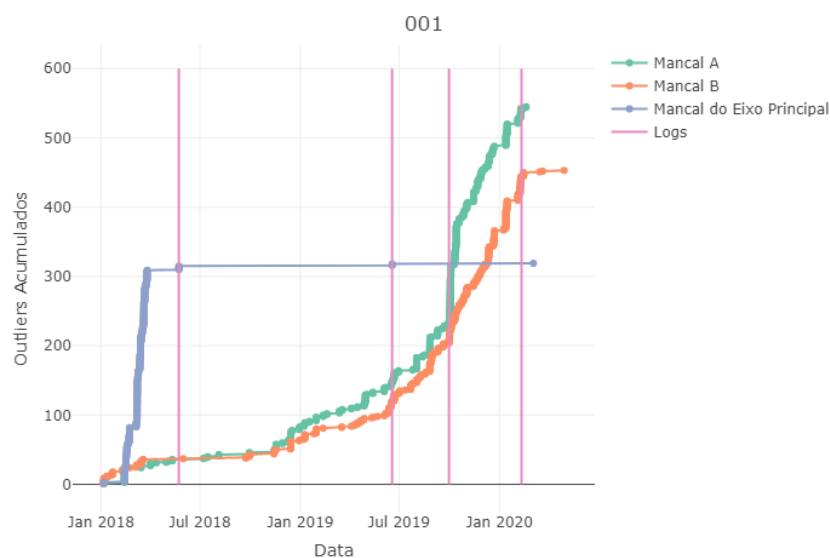
Fonte: Elaborado utilizando Google Colab [BISONG, 2019]

5.2.1.1 Mancal

Na Figura 33 é possível observar os seguintes aspectos:

- Mancal Principal:
 - Aumento abrupto dos *outliers* até o dia 26/03/2018. Então no dia 23/05/2018 consta *Log* de manutenção **específica** nesse mancal, incluindo aplicação de graxa pois o time de manutenção percebeu que o componente apresentava temperaturas elevadas;
 - Dia 18/06/2019 consta *Log* de substituição do sensor de temperatura desse componente, devido a um defeito no sensor que estava registrando temperaturas mais elevadas que as reais.
- Mancais A e B:
 - Dia 30/09/2019 sistema SCADA da máquina indica sobreaquecimento do rotor, coincidindo com o aumento abrupto de *outliers* nesses mancais;
 - Dia 10/02/2020 consta *Log* **explicitando** manutenção nesses mancais, incluindo purga de graxa, pois os mesmos vinham apresentando altas temperaturas segundo o sistema SCADA. Após essa manutenção os componentes não apresentaram mais *outliers*;

Figura 33 – Curvas com *outliers* acumulados para o grupo Mancal do Aerogerador 001.



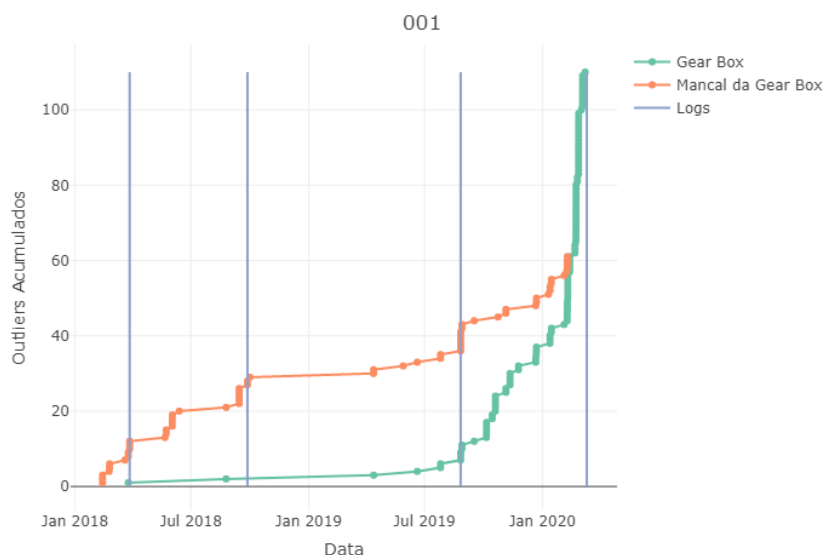
Fonte: Elaborado utilizando Rstudio [RStudio Team, 2015]

5.2.1.2 Gear Box

Na figura 34 é possível observar os seguintes aspectos:

- Mancal da *Gear Box*:
 - Dia 27/03/2018 consta *Log* de manutenção preventiva na máquina, assim como no dia 27/09/2018. Após essa data a máquina não apresentou mais *outliers* até 12/05/2019.
- Gear Box:
 - Dia 26/08/2019 consta *Log* de manutenção geral, sendo trocados os filtros de óleo da *Gear box* e completado o nível do óleo;
 - No dia 10/03/2020 consta *Log* **explicitando** troca de uma válvula termostática da *Gear Box*, coincidindo com o aumento abrupto de *outliers* desde 03/02/2020. Após essa manutenção esse componente não apresentou mais *outliers*.

Figura 34 – Curvas com *outliers* acumulados para o grupo *Gear Box* do Aerogerador 001.



Fonte: Elaborado utilizando Rstudio [RStudio Team, 2015]

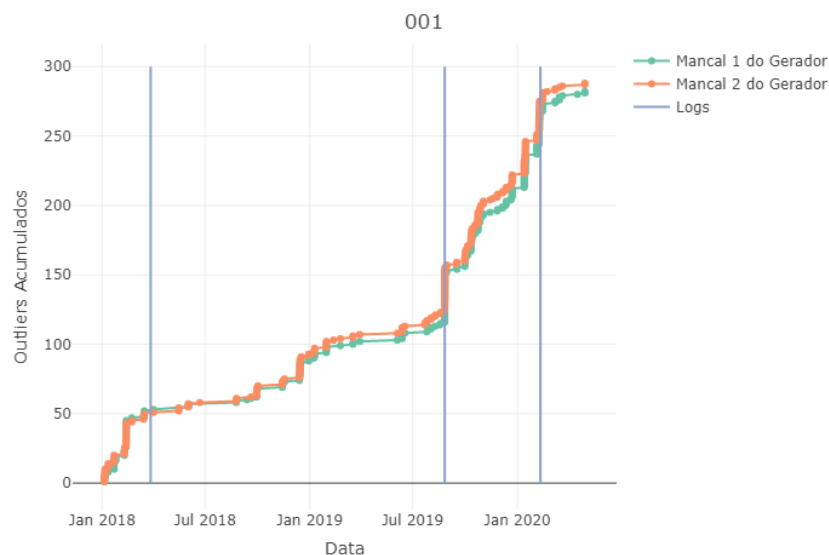
5.2.1.3 Gerador

Na Figura 35 é possível observar os seguintes aspectos:

- Mancais 1 e 2 do Gerador:

- Dia 27/09/2018 consta *Log* de manutenção preventiva na máquina. Após essa data a máquina não apresentou aumento excessivo de *outliers* até 26/08/2019;
- Dia 26/08/2019 consta *Log* de manutenção geral nesse aerogerador, incluindo trocas de filtros, óleo e lubrificação para diversos componentes;
- No dia 10/02/2020 consta *Log especificando* a lubrificação dos mancais do gerador para evitar sobre aquecimentos que vinham acontecendo. Coincidindo subidas abruptas de ambas as curvas desde 30/09/2019.

Figura 35 – Curvas com *outliers* acumulados para o grupo Gerador do Aerogerador 001.

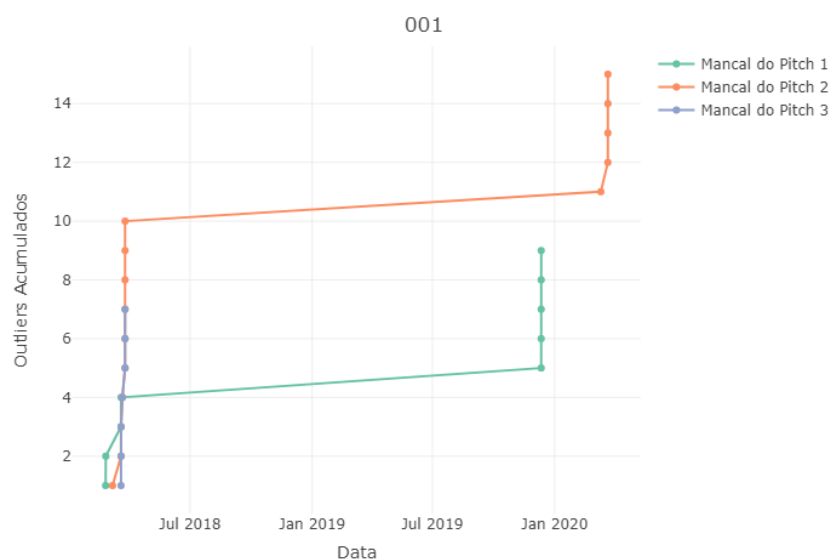


Fonte: Elaborado utilizando Rstudio [RStudio Team, 2015]

5.2.1.4 *Pitch*

Para esse Aerogerador não constam *Logs* indicando manutenções específicas nos *Pitches*, apenas nas baterias que fornecem alimentação a eles. Na Figura 36, encontram-se os *outliers* acumulados para todos os *Pitches*, com variações abruptas em 27/03/2018, 11/12/2019 e 21/03/2020. Entretanto valor máximo acumulado foi igual a 15, sendo muito inferior ao valor máximo de qualquer outro componente do Aerogerador 001.

Figura 36 – Curvas com *outliers* acumulados para o grupo *Pitch* do Aerogerador 001.

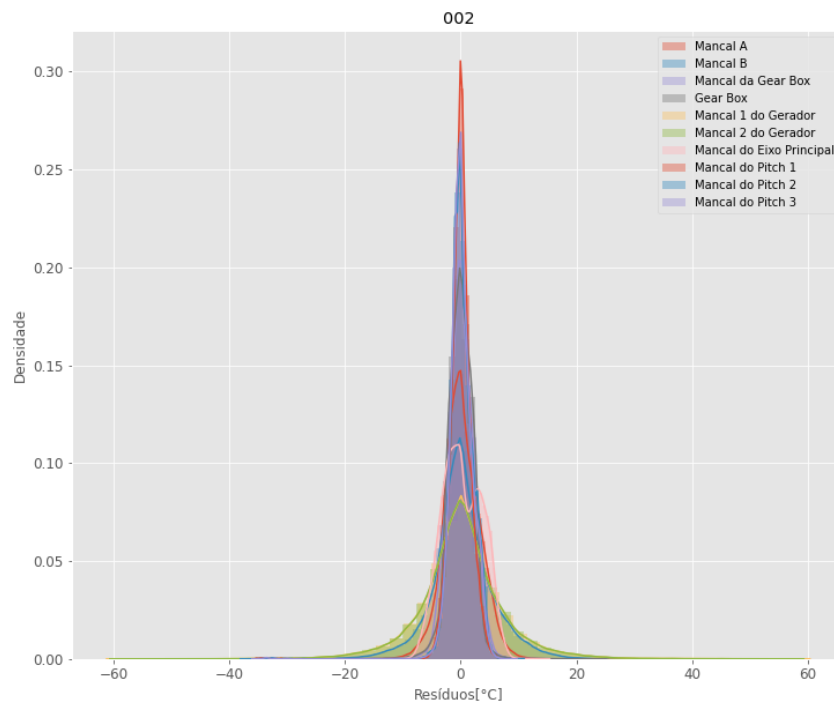


Fonte: Elaborado utilizando Rstudio [RStudio Team, 2015]

5.2.2 Aerogerador 002

A distribuição dos resíduos do Aerogerador 002 se encontra Figura 37, na qual é possível perceber que não existe a presença de resíduos com valores acima de 60 °C.

Figura 37 – Distribuição dos resíduos do Aerogerador 002.



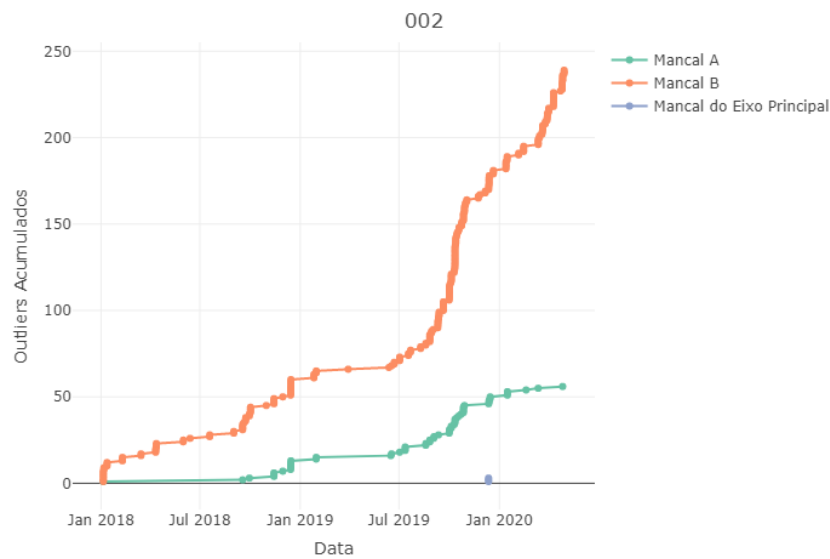
Fonte: Elaborado utilizando Google Colab [BISONG, 2019]

5.2.2.1 Mancais

Na Figura 37 é possível observar as curvas cumulativas de *outliers* para os Mancais A e B e para o Mancais do Eixo principal. A curva do Mancais A não apresenta variações abruptas, o Mancais do Eixo principal praticamente não possui *outliers* e o Mancais B apresenta um comportamento crescente com acréscimo na velocidade de crescimento a partir de 22/06/2019.

Não foi possível relacionar *Logs* de falhas para nenhum desses componentes, o que seria esperado para o Mancais A e Mancais do Eixo principal a partir de suas curvas. Para o Mancais B é possível perceber que o valor acumulado continua crescendo até o último registro que se teve acesso, no dia 29/04/2020.

Figura 38 – Curvas com *outliers* acumulados para o grupo Mancal do Aerogerador 002.

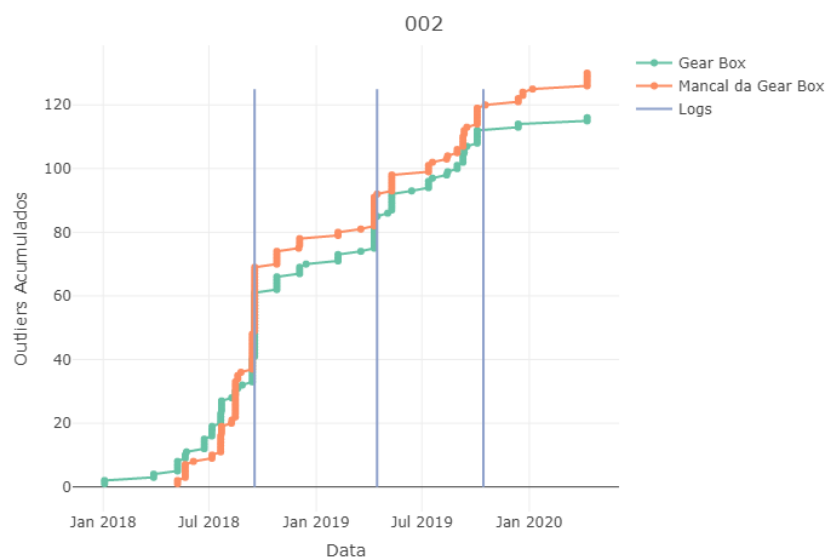


Fonte: Elaborado utilizando Rstudio [RStudio Team, 2015]

5.2.2.2 Gear Box

Na Figura 39 é possível observar aumentos abruptos as curvas do Mancal da *Gear Box* e da *Gear Box* que coincidem com 3 *Logs* de manutenção preventiva, nos dias de 17/09/2018, 15/04/2019 e 14/10/2019. Após cada *Log* as curvas abaixam sua velocidade de crescimento por alguns períodos de tempo.

Figura 39 – Curvas com *outliers* acumulados para o grupo *Gear Box* do Aerogerador 002.



Fonte: Elaborado utilizando Rstudio [RStudio Team, 2015]

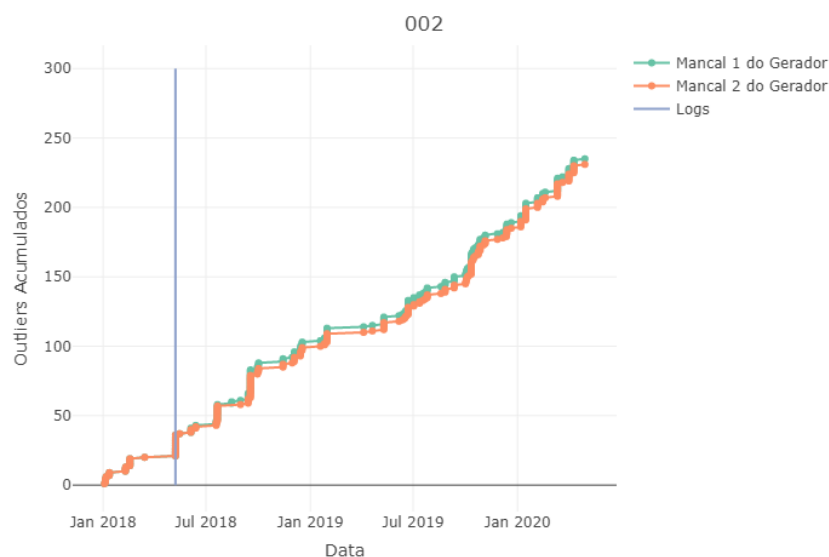
5.2.2.3 Gerador

Os Mancais 1 e 2 do Gerador não apresentaram grandes variações se comparados com os outros componentes do mesmo aerogerador, nem sequer foram encontrados nos *Logs* para essa máquina registros de defeitos nesses componentes.

Uma ressalva é feita para um *Log* que consta a substituição das escovas do gerador devido ao desgaste. Entretanto, esse é um componente elétrico do gerador que não foi modelado, e não um componente mecânico como os Mancais 1 e 2 para os quais foram modeladas as temperaturas.

Tanto o *Log* constando a troca das escovas no dia 08/05/2018 quanto as curvas cumulativas se encontram na Figura 40.

Figura 40 – Curvas com *outliers* acumulados para o grupo Gerador do Aerogerador 002.

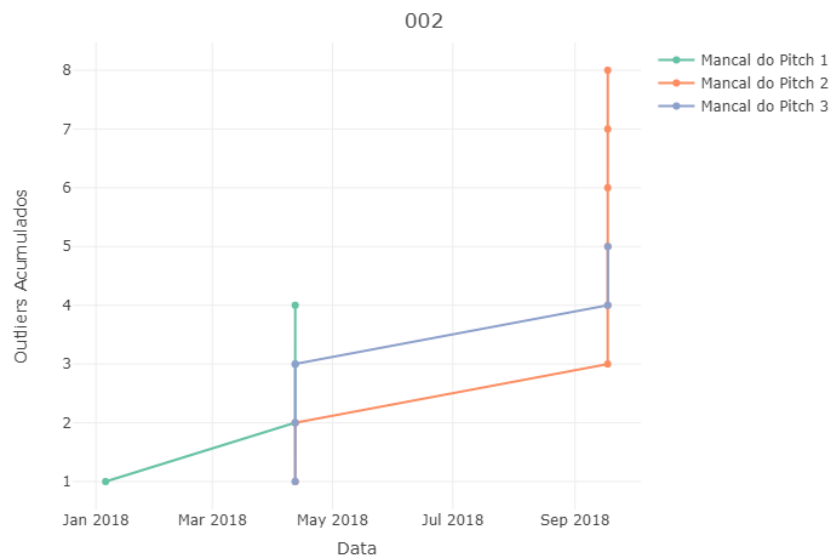


Fonte: Elaborado utilizando Rstudio [RStudio Team, 2015]

5.2.2.4 Pitch

Não constam *Logs* de manutenção nos *Pitches*, apenas a substituição de suas baterias, que não são componentes que foram modelados em relação à temperatura. As curvas desses componentes se mostram condizentes ao fato de não ter registros de defeitos, sendo que ocorrem poucos *outliers*, presentes apenas em 2018 e possui valor máximo acumulado igual a 8.

Figura 41 – Curvas com *outliers* acumulados para o grupo *Pitch* do Aerogerador 002.

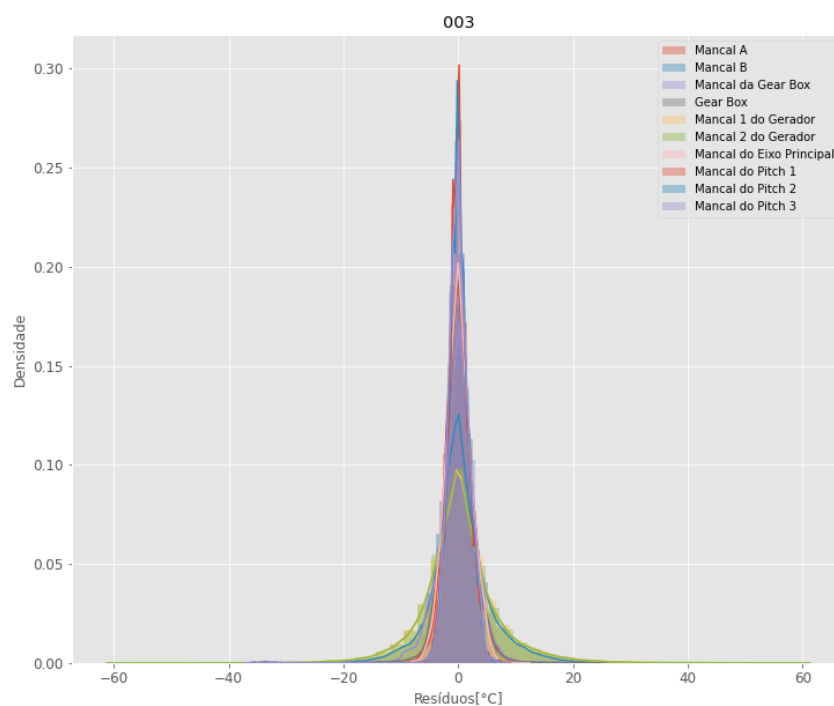


Fonte: Elaborado utilizando Rstudio [RStudio Team, 2015]

5.2.3 Aerogerador 003

O Aerogerador 003 também apresenta componentes com resíduos na casa dos 800°C, esses componentes são o Mancal da *Gear Box* e a *Gear Box*. A distribuição dos resíduos desse aerogerador pode ser observado na Figura 42.

Figura 42 – Distribuição dos resíduos do Aero gerador 003.

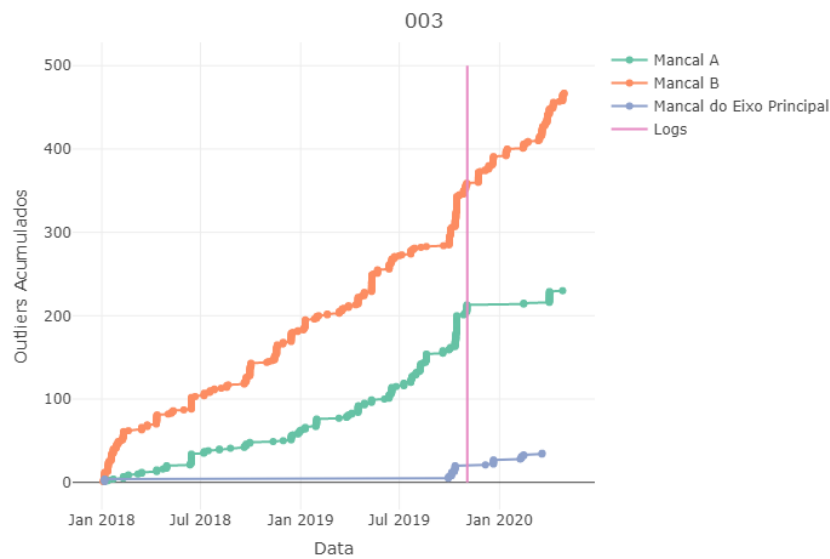


Fonte: Elaborado utilizando Google Colab [BISONG, 2019]

5.2.3.1 Mancais

Não foram encontrados *Logs* de defeitos nesses componentes. Como se pode ver na Figura 43, o único *Log* que se conseguiu relacionar a esses componentes se encontra no dia 02/11/2019, e se refere a uma limpeza de sensores de rotação da máquina.

Figura 43 – Curvas com *outliers* acumulados para o grupo Mancal do Aerogerador 003.



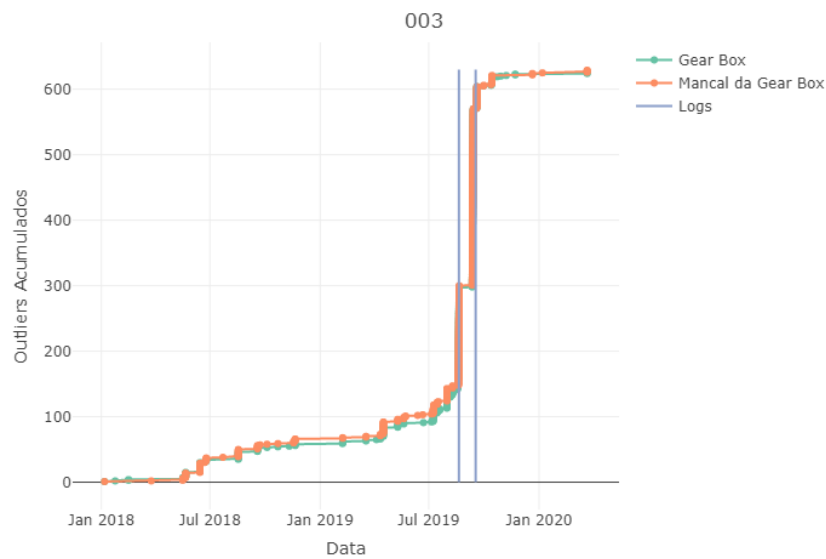
Fonte: Elaborado utilizando Rstudio [RStudio Team, 2015]

5.2.3.2 Gear Box

As curva de *outliers* acumulados tanto do Mancal da *Gear Box* quanto da *Gear Box* em si desse aerogerador apresentaram uma subida abrupta no final do dia 18/08/2019. Após isso a *Gear Box* teve que ser substituída entre 20/08/2019 e 17/09/2019. Segundo os *Logs* da substituição dessa *Gear Box* o compartimento onde fica o óleo para sua lubrificação e refrigeração foi danificado.

As curvas da *Gear Box* e do Mancal da *Gear Box* bem como os *Logs* de reparo se encontram na Figura 44.

Figura 44 – Curvas com *outliers* acumulados para o grupo *Gear Box* do Aerogerador 003.

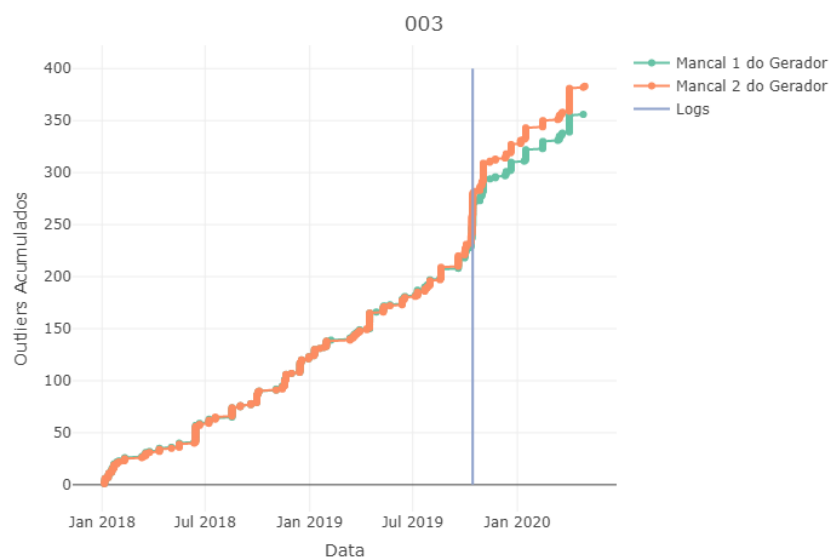


Fonte: Elaborado utilizando Rstudio [RStudio Team, 2015]

5.2.3.3 Gerador

Os Mancais 1 e 2 do Aerogerador 003 não apresentaram defeitos. A única variação abrupta das curvas ocorre na mesma data do *Log* em 14/10/2019 como pode ser observado na Figura 45, onde foi realizada a limpeza dos sensores relacionados à velocidade de rotação da máquina.

Figura 45 – Curvas com *outliers* acumulados para o grupo Gerador do Aerogerador 003.

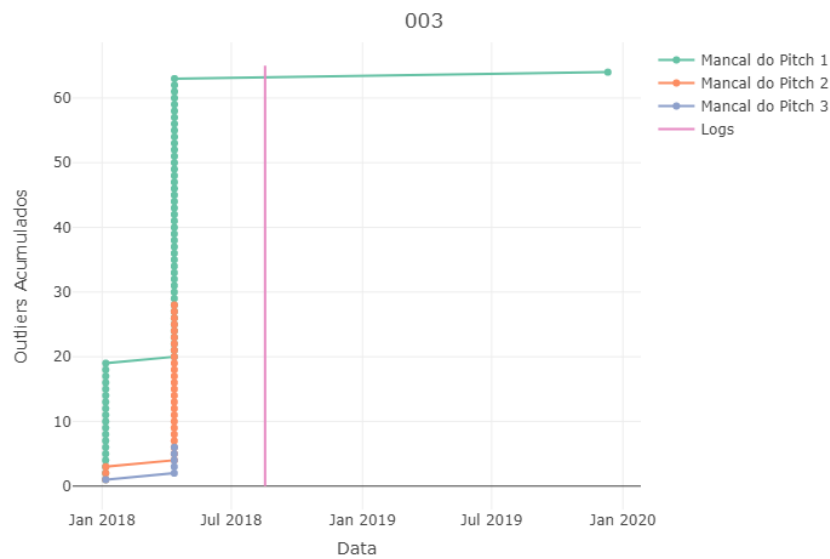


Fonte: Elaborado utilizando Rstudio [RStudio Team, 2015]

5.2.3.4 Pitch

Não constam *Logs* de defeito dos *Pitches* para esse aerogerador, entretanto existe um *Log* que consta a coleta de graxa dos 3 *Pitches* para análise no dia 17/08/2018, essa coleta ocorre após uma subida das curvas dos componentes em 12/04/2018, como pode ser observado na Figura 46.

Figura 46 – Curvas com *outliers* acumulados para o grupo *Pitch* do Aerogerador 003.

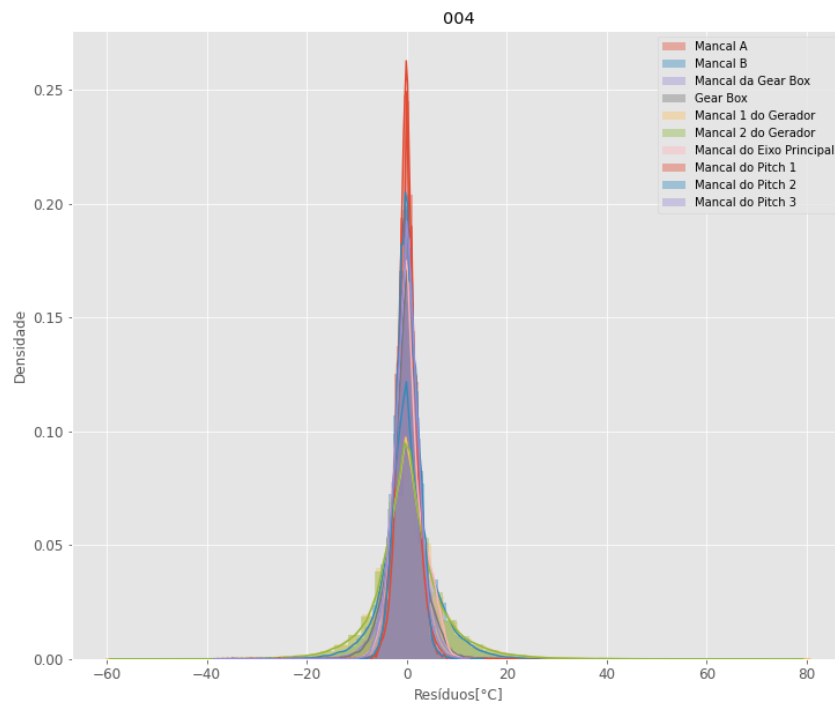


Fonte: Elaborado utilizando Rstudio [RStudio Team, 2015]

5.2.4 Aerogerador 004

A distribuição dos resíduos do Aerogerador 004 se encontra Figura 47, na qual é possível perceber que não existe a presença de resíduos com valores acima de 80 °C.

Figura 47 – Distribuição dos resíduos do Aerogerador 004.

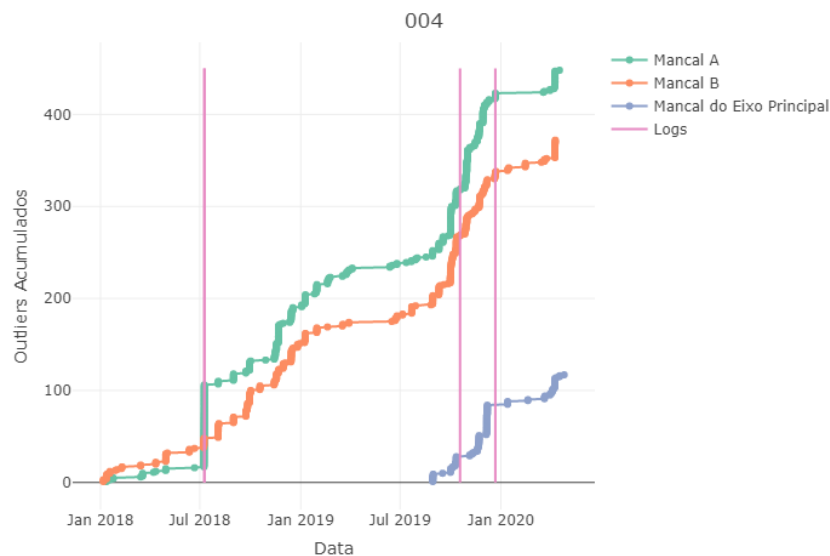


Fonte: Elaborado utilizando Google Colab [BISONG, 2019]

5.2.4.1 Mancal

Na Figura 48 é possível visualizar a subida abrupta do Mancal A em 09/07/2018 que coincide com um *Log* de manutenção na mesma data. Este *Log* indica substituição das pastilhas de freio do rotor. Nos dias 18/10/2019 e 21/12/2019 constam alarmes do sistema SCADA por alta temperatura de óleo, que coincide com os picos das subidas dos 3 componentes.

Figura 48 – Curvas com *outliers* acumulados para o grupo Mancal do Aerogerador 004.



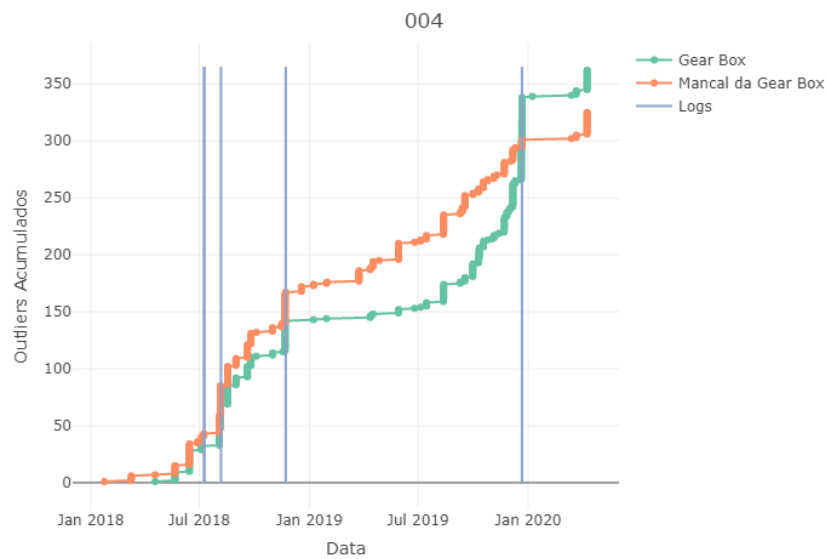
Fonte: Elaborado utilizando Rstudio [RStudio Team, 2015]

5.2.4.2 Gear Box

Na Figura 49 é possível observar os seguintes aspectos:

- No dia 09/07/2019 a primeira subida abrupta das curvas coincide com um *Log* de substituição das pastilhas do freio do rotor;
- No dia 06/08/2018 consta um *Log* de reparo em peças específicas da *Gear Box*, sendo que é possível observar uma subida abrupta a partir do dia 03/08/2018;
- No dia 22/11/2018 foi realizada uma manutenção geral na máquina, coincidindo com outra subida nas curvas. Após isso a curva da *Gear Box* permaneceu com baixo crescimento até 12/08/2019;
- No dia 21/12/2019 consta um *Log* que o aerogerador apresentou falha por alta temperatura no óleo da *Gear Box*, sendo possível perceber na Figura 49 o rápido crescimento dessa curva logo antes dessa data.

Figura 49 – Curvas com *outliers* acumulados para o grupo *Gear Box* do Aerogerador 004.

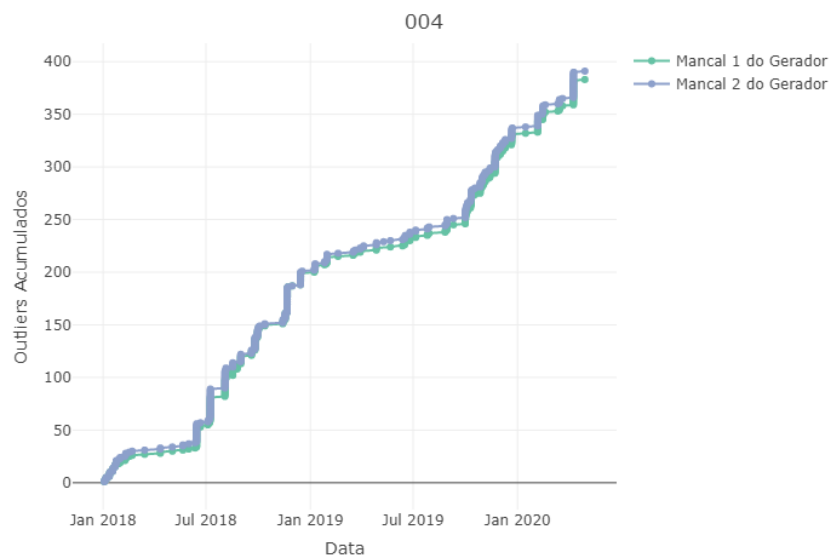


Fonte: Elaborado utilizando Rstudio [RStudio Team, 2015]

5.2.4.3 Gerador

Não foram encontrados *Logs* que indicassem que ocorreu defeitos nos Mancais 1 e 2 do gerador para o Aerogerador 004. Suas curvas de *outliers* acumulados se encontram na Figura 50.

Figura 50 – Curvas com *outliers* acumulados para o grupo Gerador do Aerogerador 004.



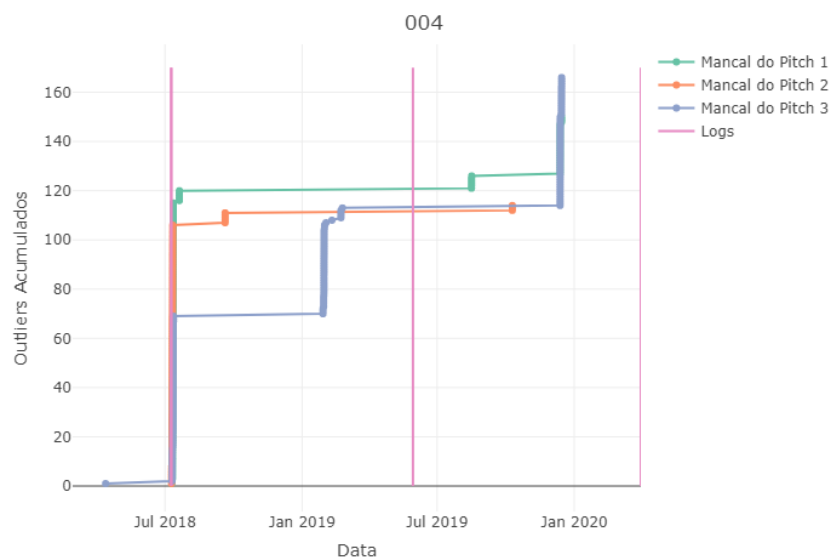
Fonte: Elaborado utilizando Rstudio [RStudio Team, 2015]

5.2.4.4 Pitch

Na Figura 51 é possível observar os seguintes aspectos:

- No dia 09/07/2019 a primeira subida abrupta das curvas coincide com um *Log* de substituição das pastilhas do freio do rotor;
- No dia 29/05/2019 consta um *Log* **especificando** que foi trocado ventilador de refrigeração do *Pitch* 3, visto que este estava aquecendo mais que os outros. É possível perceber que apenas o *Pitch* 3 possui uma subida abrupta no dia 29/01/2019, indicando várias ocorrências de *outliers* para apenas esse componente nesse período;
- No 13/12/2019 novamente o *Pitch* 3 apresentou uma subida abrupta, então no dia 30/03/2020 foram realizadas manutenções em todos os *Pitches*.

Figura 51 – Curvas com *outliers* acumulados para o grupo *Pitch* do Aerogerador 004.

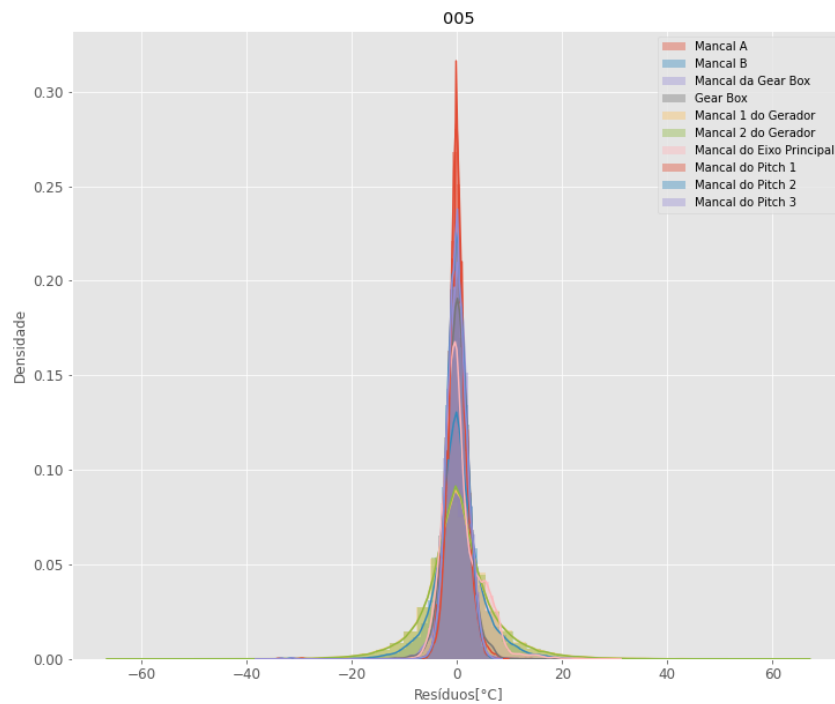


Fonte: Elaborado utilizando Rstudio [RStudio Team, 2015]

5.2.5 Aerogerador 005

A distribuição dos resíduos do Aerogerador 005 se encontra Figura 52, na qual é possível perceber que não existe a presença de resíduos com valores acima de 65°C.

Figura 52 – Distribuição dos resíduos do Aero gerador 005.

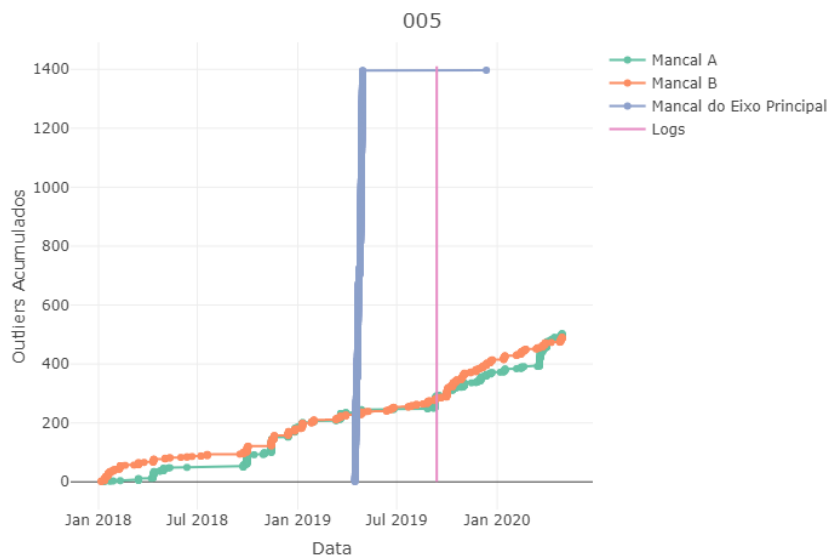


Fonte: Elaborado utilizando Google Colab [BISONG, 2019]

5.2.5.1 Mancais

Não foi encontrado nenhum registro de defeito nos componentes do grupo Mancais para o Aero gerador 005. O único *Log* que pode ser relacionado foi na data de 12/09/2019 onde foi realizada a troca de todo o óleo do Mancais do Eixo Principal, como pode ser observado na Figura 53.

Figura 53 – Curvas com *outliers* acumulados para o grupo Mancal do Aerogerador 005.

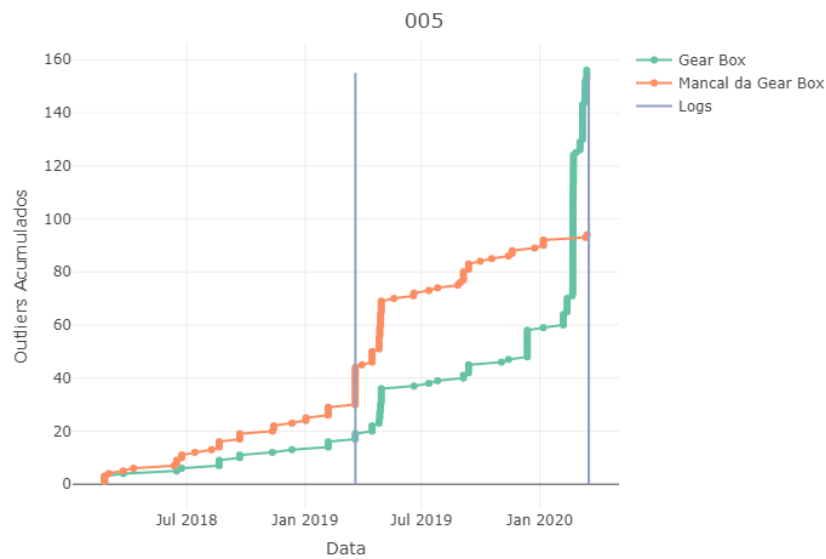


Fonte: Elaborado utilizando Rstudio [RStudio Team, 2015]

5.2.5.2 Gear Box

Para a *Gear Box* e o Mancal da *Gear Box* foi possível relacionar dois *Logs*. O primeiro em 20/03/2019 no qual foi realizada uma manutenção geral na máquina e coincide com uma subida na curva do Mancal da *Gear Box*. O segundo *Log* é no dia 17/03/2020 no qual foi realizado diversas manutenções na *Gear Box*, sendo possível verificar na Figura 54 que a subida na curva da *Gear Box* para após essa manutenção.

Figura 54 – Curvas com *outliers* acumulados para o grupo *Gear Box* do Aerogerador 005.

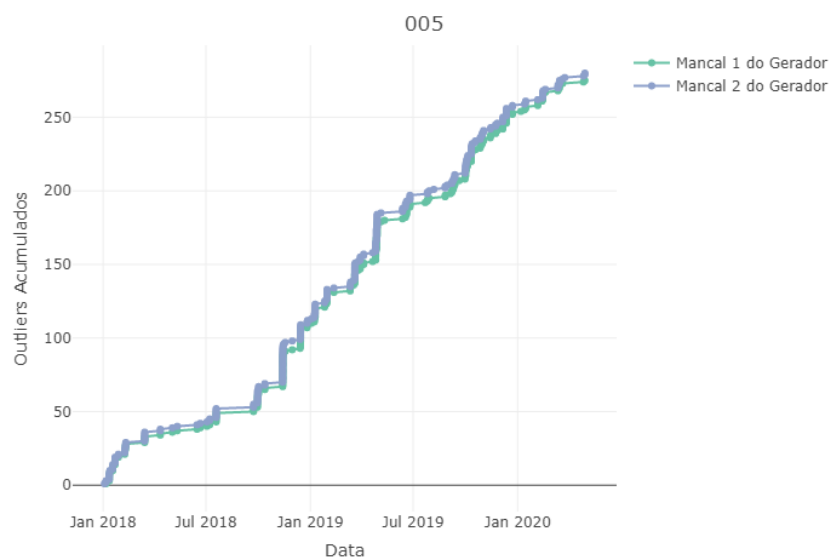


Fonte: Elaborado utilizando Rstudio [RStudio Team, 2015]

5.2.5.3 Gerador

Não foram encontrados *Logs* de defeitos a serem relacionados aos Mancais 1 e 2 do Gerador, as curvas de *outliers* acumulados para esses componentes podem ser observadas na Figura 55.

Figura 55 – Curvas com *outliers* acumulados para o grupo Gerador do Aerogerador 005.

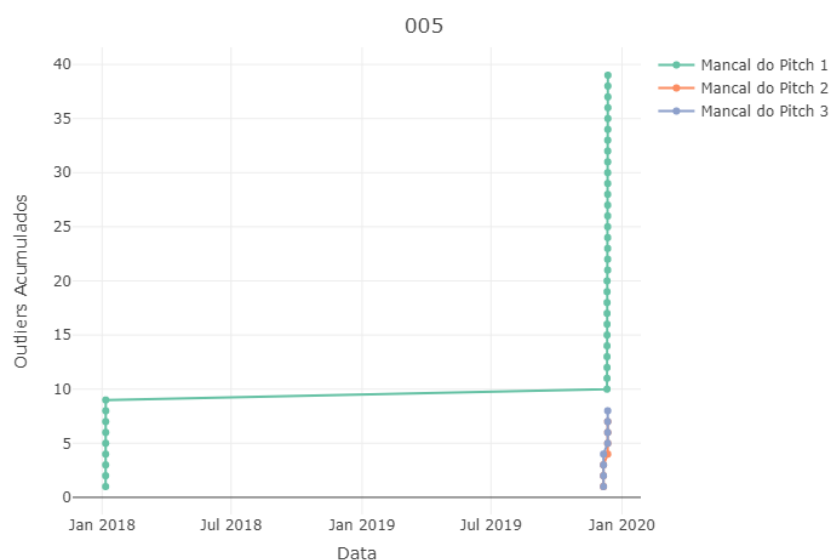


Fonte: Elaborado utilizando Rstudio [RStudio Team, 2015]

5.2.5.4 Pitch

Não foram encontrados *Logs* relacionados aos *Pitches* para esse aerogerador, mesmo assim suas curvas cumulativas são apresentadas na Figura 56.

Figura 56 – Curvas com *outliers* acumulados para o grupo *Pitch* do Aerogerador 005.

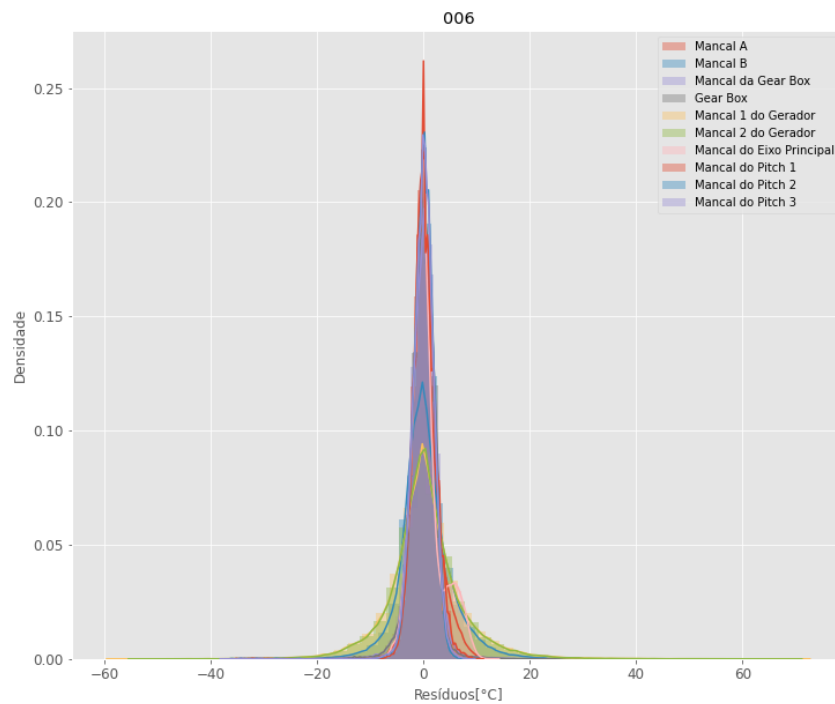


Fonte: Elaborado utilizando Rstudio [RStudio Team, 2015]

5.2.6 Aerogerador 006

A distribuição dos resíduos do Aerogerador 006 é apresentada na Figura 57, onde o valor máximo dos resíduos é de 72°C.

Figura 57 – Distribuição dos resíduos do Aeroogerador 006.



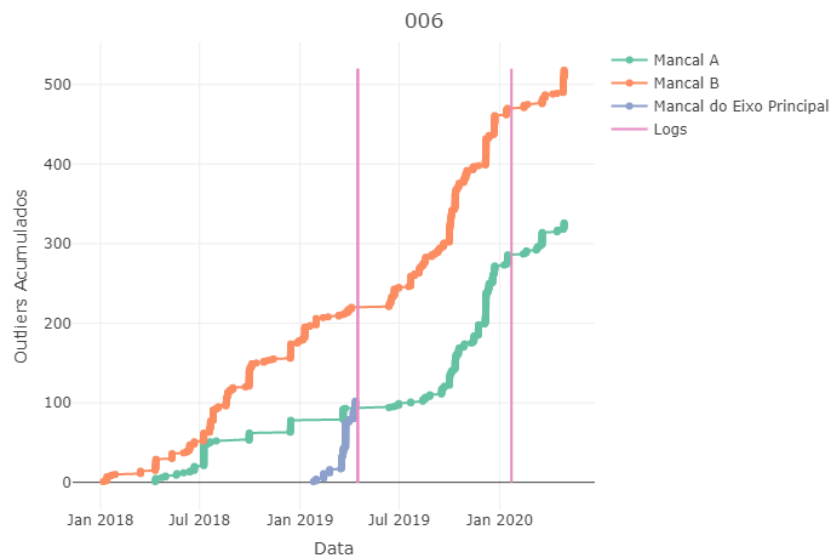
Fonte: Elaborado utilizando Google Colab [BISONG, 2019]

5.2.6.1 Mancal

Na Figura 58 é possível observar uma subida abrupta para o Mancal do Eixo Principal a partir do dia 16/03/2019, após isso no dia 16/04/2019 consta um *Log* de reparo no sistema hidráulico de freio, após esse *Log* tanto a curva do Mancal do Eixo Principal quanto do Mancal B permanecem constantes até 12/06/2019.

Além disso na Figura 58 é possível perceber que a subida das curvas do Mancal A e do Mancal B diminuem o crescimento após o *Log* de manutenção geral em 22/01/2020.

Figura 58 – Curvas com *outliers* acumulados para o grupo Mancal do Aerogerador 006.



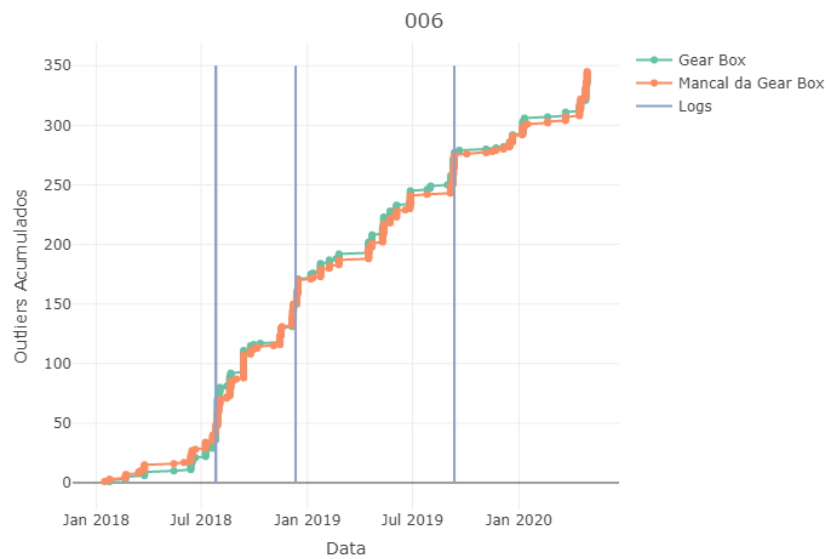
Fonte: Elaborado utilizando Rstudio [RStudio Team, 2015]

5.2.6.2 Gear Box

Esse aerogerador apresenta três *Logs* que são relacionados às curvas do Mancal da *Gear Box* e da *Gear Box*. Que podem ser verificados na Figura 59.

- Dia 26/07/2018, *Log* que foi constatado folga no rolamento do eixo rápido da *Gear Box* coincidindo com a primeira subida abrupta das curvas;
- Dia 11/12/2018, foi realizado um reparo na *Gear Box* pois esta apresentava defeito, coincidindo com a subida das curvas a partir de 05/12/2018;
- Dia 11/09/2019, *Log* de manutenção geral preventiva, após o qual as curvas diminuem o seu crescimento.

Figura 59 – Curvas com *outliers* acumulados para o grupo *Gear Box* do Aerogerador 006.

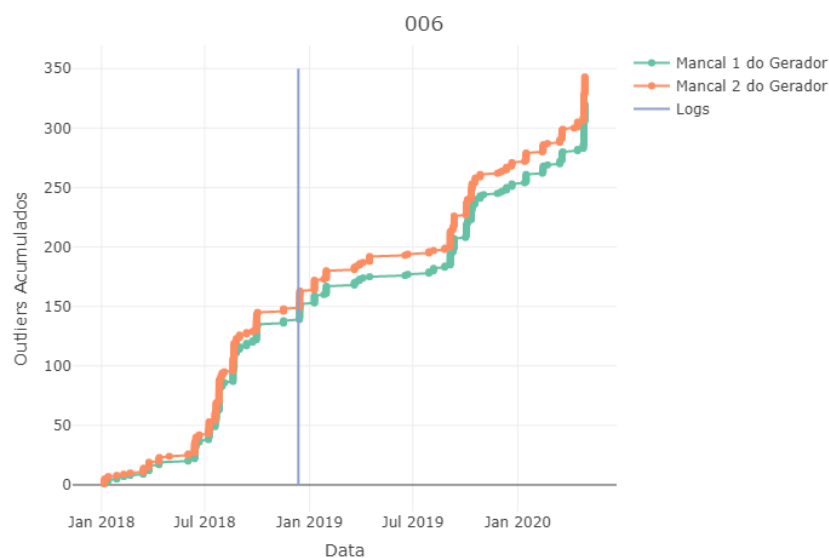


Fonte: Elaborado utilizando Rstudio [RStudio Team, 2015]

5.2.6.3 Gerador

Para os Mancais 1 e 2 do Gerador foi possível relacionar apenas um *Log* com as curvas de *outliers* acumulados, o *Log* no dia 12/12/2018 consta a realização no alinhamento do gerador, como pode ser visto na Figura 60.

Figura 60 – Curvas com *outliers* acumulados para o grupo Gerador do Aerogerador 006.

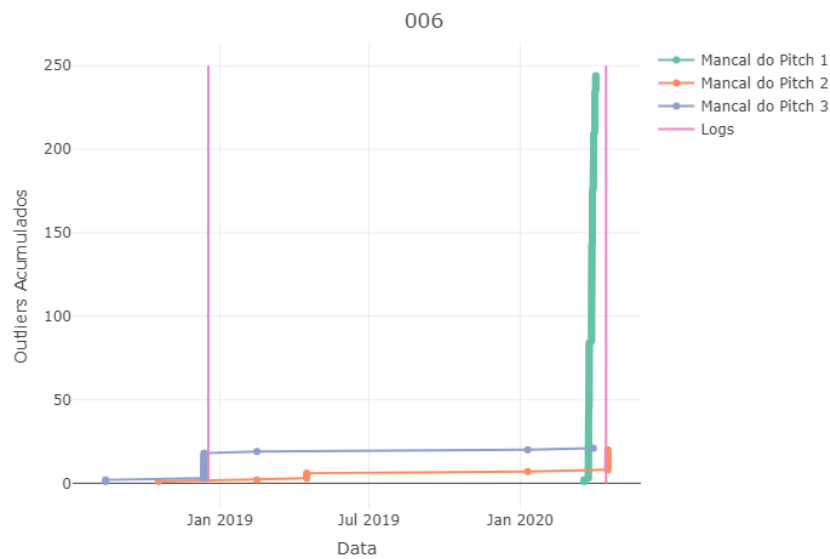


Fonte: Elaborado utilizando Rstudio [RStudio Team, 2015]

5.2.6.4 Pitch

Em relação aos *Pitches*, como pode ser observado na Figura 61, no dia 17/12/2018 foi consta um *Log* manutenção em todos os *Pitches* sendo que no dia 12/12/2018 o Mancal do Pitch 3 teve uma subida abrupta. Após isso no dia 18/03/2020 o Mancal do *Pitch* 1 teve uma subida abrupta, então no dia 14/04/2020 consta um *Log* de manutenção geral.

Figura 61 – Curvas com *outliers* acumulados para o grupo *Pitch* do Aerogerador 006.

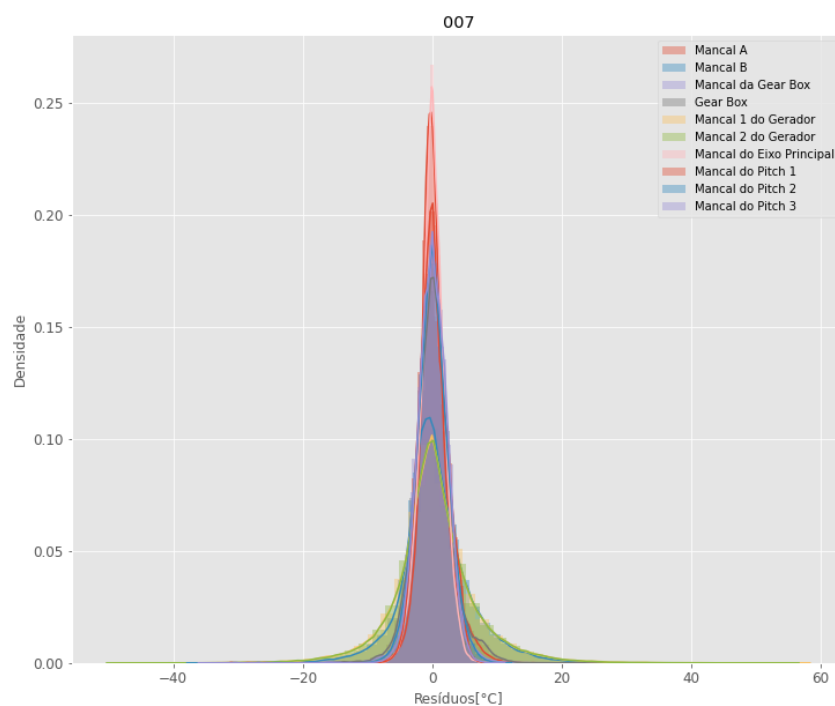


Fonte: Elaborado utilizando Rstudio [RStudio Team, 2015]

5.2.7 Aerogerador 007

É possível visualizar a distribuição dos resíduos para o Aerogerador 007 na Figura 62, sendo que o valor máximo dos resíduos não ultrapassa 60°C.

Figura 62 – Distribuição dos resíduos do Aerogerador 007.

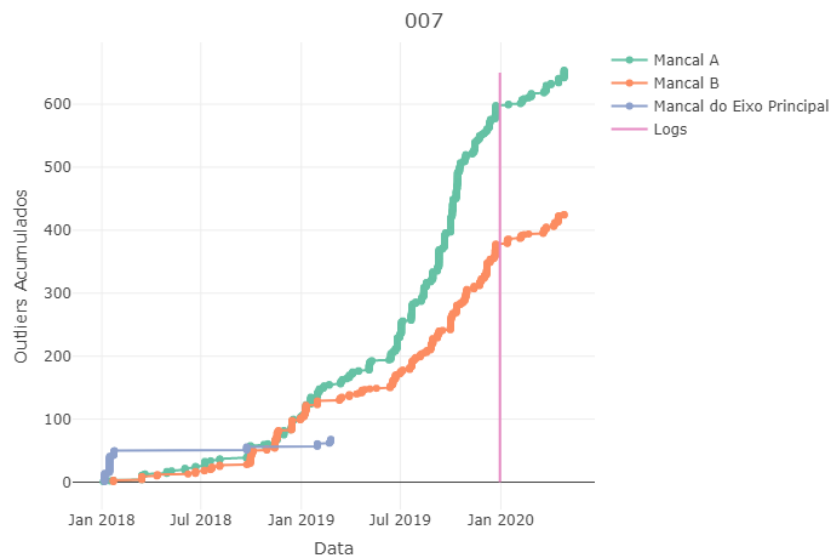


Fonte: Elaborado utilizando Google Colab [BISONG, 2019]

5.2.7.1 Mancal

Para os componentes do Grupo Mancal não foi possível relacionar nenhum *Log* de manutenção nesses componentes, o único *Log* relacionado foi do alarme de um motor de segurança no dia 30/12/2019 como pode ser visto na Figura 63.

Figura 63 – Curvas com *outliers* acumulados para o grupo Mancal do Aerogerador 007.

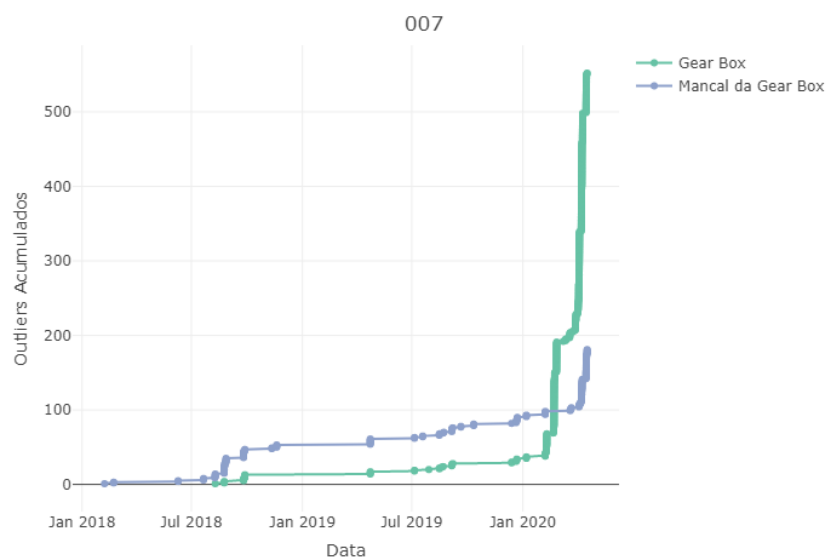


Fonte: Elaborado utilizando Rstudio [RStudio Team, 2015]

5.2.7.2 Gear Box

Não foi possível relacionar a *Gear Box* ou o Mancal da *Gear Box* com os *Logs* de falhas. Entretanto é possível ver que as curvas cumulativas continuam a crescer até o final do período em que se tem dados, como pode ser observado na Figura 64.

Figura 64 – Curvas com *outliers* acumulados para o grupo *Gear Box* do Aerogerador 007.

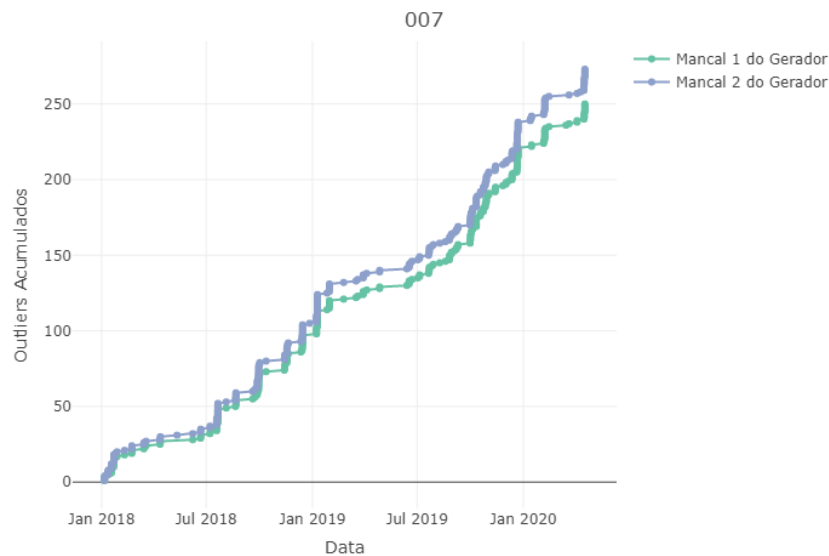


Fonte: Elaborado utilizando Rstudio [RStudio Team, 2015]

5.2.7.3 Gerador

Os Mancais 1 e 2 do Gerador do Aerogerador 007 não apresentaram defeito. Em suas curvas apresentadas na Figura 65 é possível observar que não existem subidas abruptas, em comparação com os outros componentes dessa mesma máquina.

Figura 65 – Curvas com *outliers* acumulados para o grupo Gerador do Aerogerador 007.

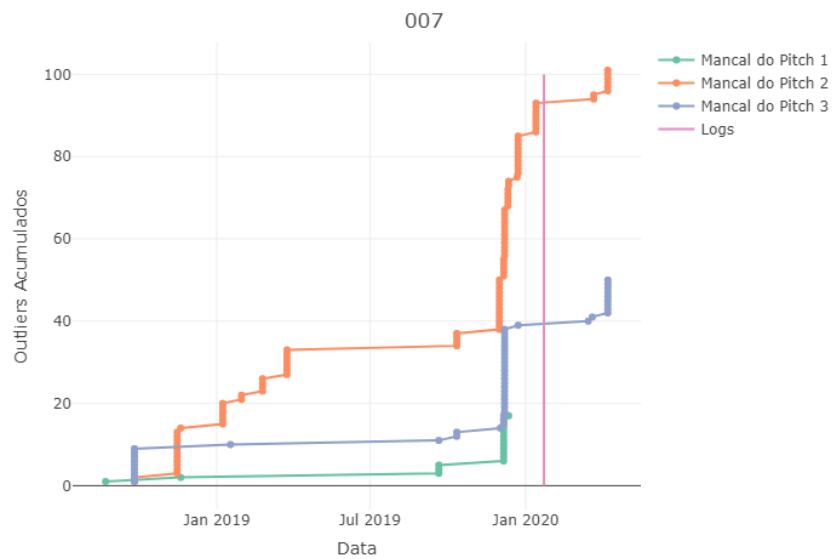


Fonte: Elaborado utilizando Rstudio [RStudio Team, 2015]

5.2.7.4 Pitch

Não foram encontrados *Logs* de defeitos nos *Pitches* para essa máquina, a única relação entre as curvas desses componentes foi com um *Log* de manutenção geral em 22/01/2020, após o qual as curvas não apresentaram mais subidas abruptas, como pode ser visualizado na Figura 66.

Figura 66 – Curvas com *outliers* acumulados para o grupo *Pitch* do Aerogerador 007.

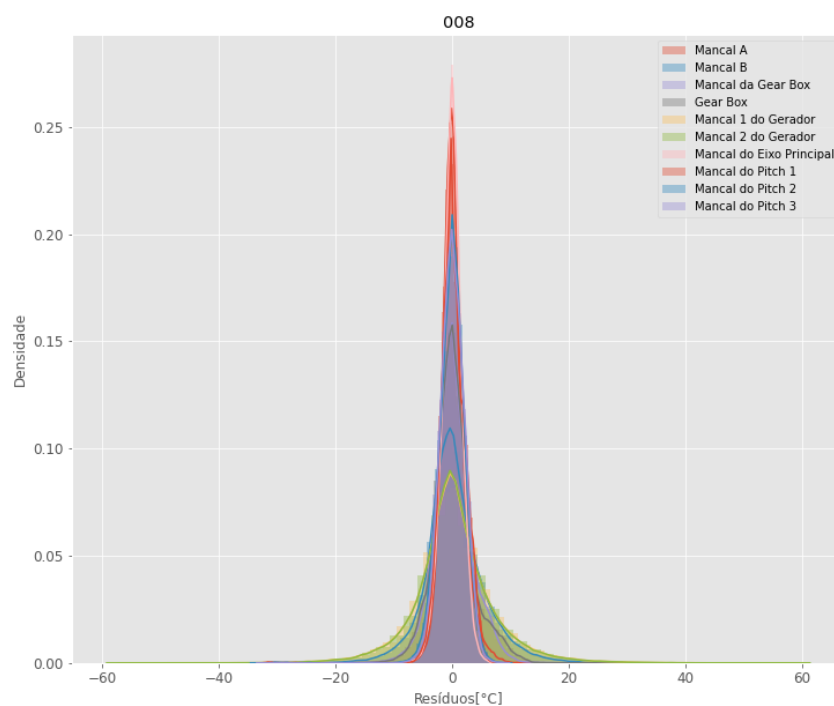


Fonte: Elaborado utilizando Rstudio [RStudio Team, 2015]

5.2.8 Aerogerador 008

Os valores dos resíduos para o Aerogerador 008 se encontram entre -59°C e 78°C . Entretanto, para melhor visualização da distribuição dos resíduos, o gráfico foi truncado entre -60°C e 60°C para todos os componentes. Como pode ser observado na Figura 67.

Figura 67 – Distribuição dos resíduos do Aerogerador 008.

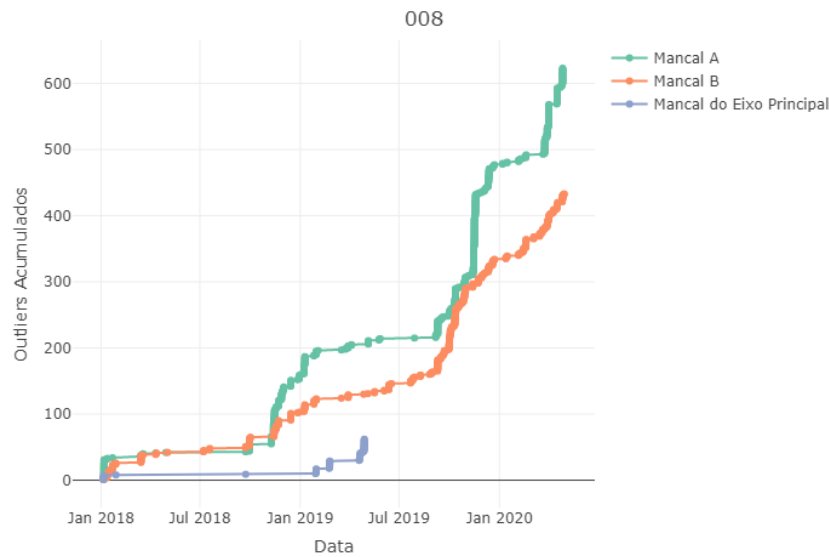


Fonte: Elaborado utilizando Google Colab [BISONG, 2019]

5.2.8.1 Mancal

Não foi possível relacionar os registros de *Logs* com defeitos nos componentes do Grupo Mancal, mesmo que suas curvas apresentem variações abruptas, como é possível verificar na Figura 68.

Figura 68 – Curvas com *outliers* acumulados para o grupo Mancal do Aerogerador 008.

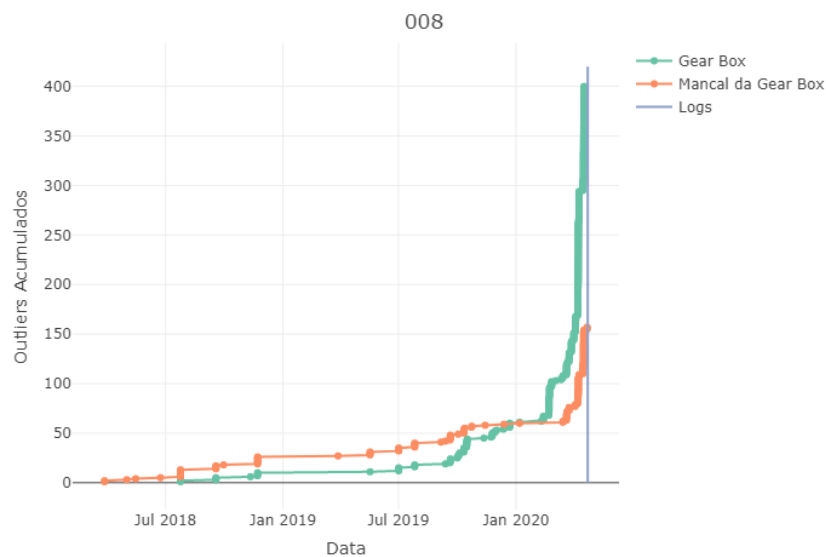


Fonte: Elaborado utilizando Rstudio [RStudio Team, 2015]

5.2.8.2 Gear Box

Quanto à *Gear Box* e ao Mancal da *Gear Box* para o Aerogerador 008, é possível verificar um aumento abrupto nos *outliers* acumulados a partir do dia 20/03/2020, sendo que no dia 22/04/2020 consta um *Log* de reparo das mangueiras de uma bomba de refrigeração.

Figura 69 – Curvas com *outliers* acumulados para o grupo *Gear Box* do Aerogerador 008.

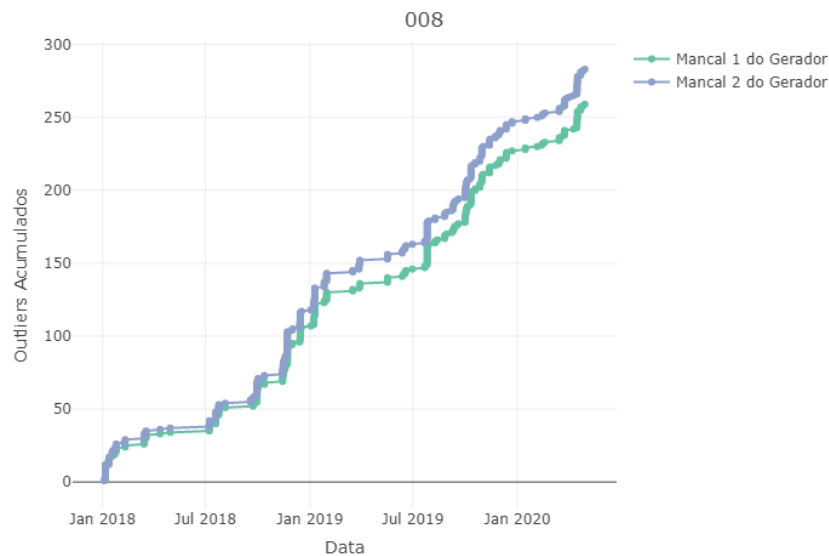


Fonte: Elaborado utilizando Rstudio [RStudio Team, 2015]

5.2.8.3 Gerador

Não foram encontrados *Logs* de defeitos nos Mancais 1 e 2 do Gerador para essa máquina, assim como suas curvas não apresentam subidas abruptas quando comparadas com outros componentes do mesmo aerogerador, como pode ser observado na Figura 70.

Figura 70 – Curvas com *outliers* acumulados para o grupo Gerador do Aerogerador 008.

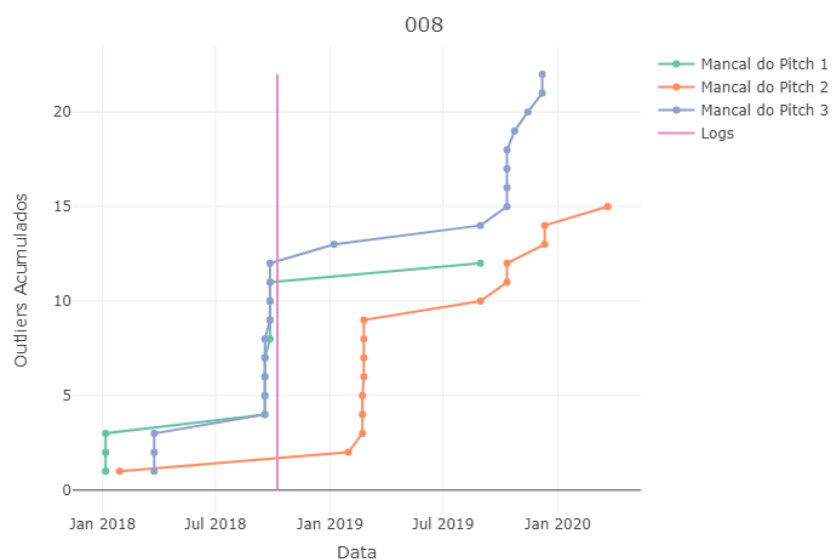


Fonte: Elaborado utilizando Rstudio [RStudio Team, 2015]

5.2.8.4 Pitch

Nesse aerogerador foi possível relacionar apenas um *Log* descrevendo falhas nos *Pitches*, esse *Log* é do dia 08/10/2018 como pode ser observado na Figura 71, e explicita a falha de comunicação e mal funcionamento desses componentes que foram reparados. Antes desse *Log* os Mancais dos *Pitches* 1 e 3 apresentaram uma subida abrupta nos *outliers* acumulados a partir do dia 18/09/2020. Após esse período outras subidas abruptas ocorreram, entretanto não consta mais registros nos *Logs* sobre esses componentes.

Figura 71 – Curvas com *outliers* acumulados para o grupo *Pitch* do Aerogerador 008.



Fonte: Elaborado utilizando Rstudio [RStudio Team, 2015]

6 Conclusão

O presente trabalho mostrou potencial para ser utilizado no monitoramento de aerogeradores, visto que com base nos *outliers* das temperaturas e Curvas S foi possível ver relações entre essas anomalias e alguns defeitos da máquina. Além disso, em grande parte dos casos, quando não se apresentavam comportamentos discrepantes nos componentes não constavam defeitos para esses mesmos componentes. Outro ponto é que geralmente quando existiam comportamentos muito discrepantes em um componente, como subidas íngremes nas curvas cumulativas, os *Logs* indicavam alguma manutenção sendo realizada posteriormente, o que nos leva a supor que esse comportamento possa servir de preditor antes que um defeito ocorra nos componentes do aerogerador. Sendo assim, um próximo passo que pode ser tomado é a utilização em maior escala dessa metodologia, inclusive para máquinas de outros fabricantes. Entretanto, para tal, seria interessante aprimorar alguns pontos da metodologia proposta.

Alguns pontos podem ser aprimorados para melhorar a presente metodologia, como o aumento de números de candidatos nos modelos de regressão, a fim de se encontrar modelos que melhor se adéquem aos dados. Outra melhoria seria o desenvolvimento de um algoritmo automático para relacionar as curvas de *outliers* acumulados com os *Logs* de defeitos nos aerogeradores, assim como o uso de novas *features* como medidas de vibrações dos componentes de cada máquina, para melhor descrever comportamento dos aerogeradores. Além disso, pode ser vantajoso fazer alguma análise referente à existência de sazonalidades nos comportamentos dos componentes.

Outra melhoria que pode ser realizada é utilizar o conhecimento da equipe de manutenção do parque eólico para verificar, e talvez até desconsiderar, registros de *Logs* em que a própria manutenção influencie na quantia de *outliers*. Além de que, com base no conhecimento prático desses profissionais, pode ser possível melhorar a análise para relacionar os sobreaquecimentos com falhas.

Essas melhorias ficam de sugestão para trabalhos futuros, em conjunto com a possibilidade de desenvolver um sistema automático de monitoramento para os aerogeradores.

Referências Bibliográficas

ANEEL, A. Atlas de energia elétrica do brasil. *Brasília*, 2008.

BISONG, E. *Building Machine Learning and Deep Learning Models on Google Cloud Platform*. [S.l.]: Springer, 2019.

Datanami. *How to Build a Better Machine Learning Pipeline*. 2018. Disponível em: <<https://www.datanami.com/2018/09/05/how-to-build-a-better-machine-learning-pipeline/>>. Acesso em: 18 de jun. de 2020.

DEISENROTH, M. P.; FAISAL, A. A.; ONG, C. S. *Mathematics for machine learning*. [S.l.]: Cambridge University Press, 2020.

FISCHLER, M. A.; BOLLES, R. C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, ACM New York, NY, USA, v. 24, n. 6, p. 381–395, 1981.

FRIEDMAN, J. H. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, JSTOR, p. 1189–1232, 2001.

GÉRON, A. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. [S.l.]: O'Reilly Media, 2019.

GLAESSGEN, E.; STARGEL, D. The digital twin paradigm for future nasa and us air force vehicles. In: *53rd AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics and materials conference 20th AIAA/ASME/AHS adaptive structures conference 14th AIAA*. [S.l.: s.n.], 2012. p. 1818.

HAMEED, Z. et al. Condition monitoring and fault detection of wind turbines and related algorithms: A review. *Renewable and Sustainable energy reviews*, Elsevier, v. 13, n. 1, p. 1–39, 2009.

HASSOUN, M. H. et al. *Fundamentals of artificial neural networks*. [S.l.]: MIT press, 1995.

ISO16079-1. *Condition monitoring and diagnostics of wind turbines – Part 1: General guidelines*. Geneva, Suíça, 2017.

ISO/DIS16079-2. *Condition monitoring and diagnostics of wind turbines – Part 2: Monitoring the drive train*. Geneva, Suíça, 2019.

Itaipu Binacional. *Unidades Geradoras*. 2020. Disponível em: <<https://www.itaipu.gov.br/energia/unidades-geradoras>>. Acesso em: 23 de mai. de 2020.

Ministério de Minas e Energia. *Plano Decenal de Expansão de Energia 2029*. 2019. Disponível em: <<http://www.mme.gov.br/documents/78404/0/Apresentaç~ao+PDE+2029+-+REIVE+BARROS/e6df4e97-f42f-9214-e24f-e122b4e0863d>>. Acesso em: 23 de mai. de 2020.

- MINKA, T. P. Automatic choice of dimensionality for pca. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2001. p. 598–604.
- MONTGOMERY, D. C.; RUNGER, G. C.; HUBELE, N. F. *Engineering statistics*. [S.l.]: John Wiley & Sons, 2009.
- OROZCO, R.; SHENG, S.; PHILLIPS, C. Diagnostic models for wind turbine gearbox components using scada time series data. In: IEEE. *2018 IEEE International Conference on Prognostics and Health Management (ICPHM)*. [S.l.], 2018. p. 1–9.
- PEDREGOSA, F. et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, v. 12, p. 2825–2830, 2011.
- PLAKHOV, A.; CRUZ, P. A stochastic approximation algorithm with step-size adaptation. *Journal of Mathematical Sciences*, v. 120, n. 1, 2004.
- RStudio Team. *RStudio: Integrated Development Environment for R*. Boston, MA, 2015. Disponível em: <<http://www.rstudio.com/>>.
- SAMUEL, A. L. Some studies in machine learning using the game of checkers. *IBM Journal of research and development*, IBM, v. 3, n. 3, p. 210–229, 1959.
- SOHONI, V.; GUPTA, S.; NEMA, R. A critical review on wind turbine power curve modelling techniques and their applications in wind based energy systems. *Journal of Energy*, Hindawi, v. 2016, 2016.
- TAO, F. et al. Digital twin in industry: State-of-the-art. *IEEE Transactions on Industrial Informatics*, IEEE, v. 15, n. 4, p. 2405–2415, 2018.
- TAO, F. et al. Digital twin driven prognostics and health management for complex equipment. *CIRP Annals*, Elsevier, v. 67, n. 1, p. 169–172, 2018.
- VANDERPLAS, J. *Python data science handbook: Essential tools for working with data*. [S.l.]: "O'Reilly Media, Inc.", 2016.
- WICKHAM, H.; GROLEMUND, G. *R for data science: import, tidy, transform, visualize, and model data*. [S.l.]: "O'Reilly Media, Inc.", 2016.
- Wind Power Monthly. *SGRE raises offshore bar with 14MW+ turbine and 222m rotor*. 2020. Disponível em: <<https://www.windpowermonthly.com/article/1683570/sgre-upped-offshore-stakes-14mw+-turbine-222m-rotor>>. Acesso em: 23 de mai. de 2020.
- ZAHER, A. et al. Online wind turbine fault detection through automated scada data analysis. *Wind Energy: An International Journal for Progress and Applications in Wind Power Conversion Technology*, Wiley Online Library, v. 12, n. 6, p. 574–593, 2009.