

Francisco Mateus de Oliveira Abrantes

**Técnicas de remoção de outliers na curva de
potência de aerogeradores e parametrização
da curva de operação normal**

Natal – RN

Dezembro de 2022

Francisco Mateus de Oliveira Abrantes

**Técnicas de remoção de outliers na curva de potência
de aerogeradores e parametrização da curva de
operação normal**

Trabalho de Conclusão de Curso de Engenharia de Computação da Universidade Federal do Rio Grande do Norte, apresentado como requisito parcial para a obtenção do grau de Bacharel em Engenharia de Computação

Orientador: Dr. Luiz Affonso Henderson Guedes de Oliveira

Universidade Federal do Rio Grande do Norte – UFRN
Departamento de Engenharia de Computação e Automação – DCA
Curso de Engenharia de Computação

Natal – RN
Dezembro de 2022

Universidade Federal do Rio Grande do Norte - UFRN
Sistema de Bibliotecas - SISBI
Catalogação de Publicação na Fonte. UFRN - Biblioteca Central Zila Mamede

Abrantes, Francisco Mateus de Oliveira.

Técnicas de remoção de outliers na curva de potência de aerogeradores e parametrização da curva de operação normal / Francisco Mateus de Oliveira Abrantes. - 2022.

60 f.: il.

Monografia (graduação) - Universidade Federal do Rio Grande do Norte, Centro de Tecnologia, Curso de Engenharia de Computação. Natal, RN, 2022.

Orientador: Prof. Dr. Luiz Affonso Henderson Guedes de Oliveira.

1. Engenharia de Computação - Monografia. 2. Curva de potência - Monografia. 3. Outliers - Monografia. 4. Operação normal - Monografia. 5. Interpolação - Monografia. 6. Aerogeradores - Monografia. I. Oliveira, Luiz Affonso Henderson Guedes de. II. Título.

RN/UF/BCZM

CDU 004

Francisco Mateus de Oliveira Abrantes

Técnicas de remoção de outliers na curva de potência de aerogeradores e parametrização da curva de operação normal

Trabalho de Conclusão de Curso de Engenharia de Computação da Universidade Federal do Rio Grande do Norte, apresentado como requisito parcial para a obtenção do grau de Bacharel em Engenharia de Computação

Orientador: Dr. Luiz Affonso Henderson Guedes de Oliveira

Trabalho aprovado. Natal – RN, 15 de Dezembro de 2022:

Prof. Dr. Luiz Affonso Henderson Guedes de Oliveira - Orientador
UFRN

Me. Vitor Rodrigues Greati - Examinador externo
UNIVERSITY OF GRONINGEN

Me. Danilo Mikael Costa Barros - Examinador externo
LOGAP

Natal – RN
Dezembro de 2022

AGRADECIMENTOS

A todos os professores da Universidade Federal do Rio Grande do Norte (UFRN) que colaboraram e construíram bases sólidas no meu desenvolvimento e aprendizagem para o crescimento profissional. Seus nomes são inesquecíveis e por isso, dedico-lhes minha profunda admiração e respeito.

A todos aqueles que acreditaram na realização deste trabalho e deram-me forças e estímulo para dar prosseguimento a esta pesquisa e obter sucesso. Em especial, ao meu orientador, o professor Luiz Affonso, ao pessoal da Logap em especial a Danilo Mikael, Joilson Abrantes, Jonathan Batista e Vitor Greati.

Gostaria também de agradecer a minha tia Maria de Fátima Abrantes que me acolheu em sua casa durante todos os anos da graduação. Como também gostaria de agradecer a todos meus colegas de turma, o convívio e as aprendizagens que compartilhamos foram muito enriquecedoras, e com certeza levarei comigo para sempre.

A Deus criador dos céus e da terra, o que me deu a vida. E a minha família que me deu apoio e sustentação para que eu pudesse trilhar esse caminho.

“Desde o começo da civilização, as pessoas tentam entender a ordem fundamental do mundo. Deve haver algo muito especial sobre os limites do universo. E o que pode ser mais especial do que não haver limites? Não deve haver limites para o esforço humano. Somos todos diferentes. Por pior que a vida possa parecer, sempre há algo que podemos fazer em que podemos obter sucesso. Enquanto houver vida, haverá esperança. ”

Stephen Hawking

RESUMO

A curva de potência dos aerogeradores é uma importante forma de avaliar a produção de energia dos aerogeradores, sua composição se dá entre a relação da potência ativa do aerogerador em função da velocidade do vento. Analisar essa curva possibilita avaliar o desempenho do aerogerador. Entretanto, valores anormais nessa curva impossibilitam analisar de forma eficaz o desempenho, além de provocam resultados inválidos nas análises. Esse trabalho se propõe a realizar em uma base de dados real a remoção desses pontos de anomalias com a utilização de métodos já muito difundidos no meio científico para remoção de outliers nas curvas de potência. Além disso, o trabalho propõe que, ao obter a curva sob condições normais, seja feita a parametrização dessa curva, de forma que se possa obter as funções que representam a forma da mesma. Esse processo se dará por meio de interpolações polinomiais das funções que expressam a curva, e com isso será possível obter os coeficientes que possibilitam o armazenamento dessas funções de forma simples. Por fim, com a parametrização dessa curva analisamos o desempenho dos aerogeradores.

Palavras-chaves: Curva de potência; Outliers; Operação normal; Interpolação; Aerogeradores; Desempenho.

ABSTRACT

The power curve is an important way to evaluate the energy production of wind turbines. It expresses the ratio of the active power of the wind turbine as a function of wind speed. Analyzing this curve makes it possible to evaluate the performance of the wind turbine. However, abnormal values on this curve make it impossible to effectively analyze the performance, as they may disturb the analysis, and cause invalid conclusion. This work proposes to remove these anomaly points using methods already widespread in the scientific environment for removing outliers in power curves, with the application of these methods in a real database. In addition, the work proposes the parameterization of the normal curve, so that the functions that represent its shape can be obtained. This process will take place through polynomial interpolations, and with that it will be possible to obtain the coefficients that allow the storage of these functions in a simple way. Finally with the parameterization of this curve, we analyze the performance of wind turbines.

Keywords: Power curve; Outliers; Normal operation; Interpolation; Wind turbines; Performance.

LISTA DE ILUSTRAÇÕES

Figura 1 – Componentes e funcionamento de um aerogerador.	18
Figura 2 – Regiões da curva de potência.	20
Figura 3 – Tipos de outlier na curva de potência.	22
Figura 4 – Representação do gráfico do boxplot na identificação de outliers.	24
Figura 5 – Funcionamento do DBSCAN.	25
Figura 6 – MMQ: Erro da aproximação por uma reta.	27
Figura 7 – Quantidade de valores brutos obtidos do SCADA para cada aerogerador.	29
Figura 8 – Quantidade de valores não nulos obtidos do SCADA para cada aerogerador.	30
Figura 9 – Curva de potência por ano do aerogerador 1.	31
Figura 10 – Curva de potência por ano do aerogerador 2.	32
Figura 11 – Curva de potência por ano do aerogerador 3.	33
Figura 12 – Curva de potência por ano do aerogerador 4.	34
Figura 13 – Quantidade de eventos por aerogerador.	35
Figura 14 – Quantidade de eventos por classificação e por aerogerador.	36
Figura 15 – Quantidade de eventos por nível e por aerogerador.	37
Figura 16 – Amostra com pontos antes da realização da classificação por eventos.	41
Figura 17 – Amostra com pontos rotulados pela classificação por eventos.	41
Figura 18 – Exemplo de identificação de ruído com o DBSCAN.	42
Figura 19 – Método interquartil para seccionamento da curva de potência em intervalos de velocidade do vento.	43
Figura 20 – Visualização do método interquartil como boxplot nos intervalos de velocidade do vento.	44
Figura 21 – Pontos de potência negativa.	45
Figura 22 – Seccionamento da curva de potência normal em intervalos de velocidade do vento.	46
Figura 23 – Seccionamento da curva de potência normal, e identificação dos pontos de máximo e mínimo de potência e do valor de velocidade no início do intervalo.	46
Figura 24 – Ligação dos pontos da curva utilizando retas.	47
Figura 25 – Envelope “S” formado pelas assíntotas usando interpolação polinomial de grau 9.	47
Figura 26 – Obtenção da curva de potência sob condições normais para o aerogerador 1.	49
Figura 27 – Obtenção da curva de potência sob condições normais para o aerogerador 2.	49

Figura 28 – Obtenção da curva de potência sob condições normais para o aerogerador 3.	50
Figura 29 – Obtenção da curva de potência sob condições normais para o aerogerador 4.	50
Figura 30 – Curvas por ano para o aerogerador 1.	54
Figura 31 – Curvas por ano sobrepostas para o aerogerador 1.	55
Figura 32 – Curvas por ano para o aerogerador 2.	55
Figura 33 – Curvas por ano sobrepostas para o aerogerador 2.	56
Figura 34 – Curvas por ano para o aerogerador 3.	56
Figura 35 – Curvas por ano sobrepostas para o aerogerador 3.	57
Figura 36 – Curvas por ano para o aerogerador 4.	58
Figura 37 – Curvas por ano sobrepostas para o aerogerador 4.	58

LISTA DE TABELAS

Tabela 1 – Amostra de dados de potência e velocidade do vento obtidos do SCADA.	29
Tabela 2 – Amostra de dados de eventos.	35
Tabela 3 – Tempo de execução do algoritmo final para remoção de outliers e parametrização da curva sob condições normais.	53

LISTA DE ABREVIATURAS E SIGLAS

ABEEólica	<i>Associação Brasileira de Energia Eólica</i>
DBSCAN	<i>Density-Based Spatial Clustering of Application with Noise</i>
gRPC	<i>Google Remote Procedure Call</i>
MMQ	<i>Método dos Mínimos Quadrados</i>
SCADA	<i>Supervisory Control And Data Acquisition</i>

SUMÁRIO

1	INTRODUÇÃO	14
1.1	Objetivos	15
1.2	Estrutura do Trabalho	15
2	FUNDAMENTAÇÃO TEÓRICA	17
2.1	Energia eólica e aerogeradores	17
2.2	Curva de potência	19
2.3	Eventos associados aos aerogeradores	20
2.4	Outliers	21
2.5	Métodos de remoção de outlier da curva de potência	23
2.5.1	Método interquartil	23
2.5.2	DBSCAM	25
2.6	Interpolação polinomial	26
3	METODOLOGIA	28
3.1	Obtenção dos dados	28
3.2	Dados de potência e velocidade do vento	29
3.3	Dados de eventos	34
3.4	Classificar pontos por eventos	38
3.5	DBSCAN	41
3.6	Método interquartil	42
3.7	Pontos de potência negativa e inválidos	44
3.8	Curva de potência sob condições normais	45
4	RESULTADOS E DISCUSSÕES	48
4.1	Obtenção das curvas de potência sob condições normais para cada aerogerador e análise do tempo de processamento das amostras	48
4.2	Análise de desempenho dos aerogeradores por meio das curvas de potência durante os anos de 2018 à 2021	53
4.2.1	Aerogerador 1	54
4.2.2	Aerogerador 2	55
4.2.3	Aerogerador 3	56
4.2.4	Aerogerador 4	57
5	CONCLUSÃO	59

REFERÊNCIAS	60
--------------------	-----------

1 INTRODUÇÃO

Um dos desafios da humanidade atualmente está na substituição de uma matriz energética altamente poluente, a qual se utiliza de combustíveis fósseis, por uma fonte limpa e renovável. As mudanças climáticas e a crise do petróleo desencadeados pela guerra da Ucrânia e Rússia em 2022 provocaram grandes oscilações dos preços dos combustíveis e também uma baixa disponibilidade no mercado mundial. O desencadeamento desses efeitos trouxeram à tona a necessidade recente de substituir os meios de geração de energia, não sendo mais uma sinalização de promessa futura, e sim uma necessidade urgente. Em decorrência disso, fontes alternativas que utilizam recursos naturais renováveis como o vento, água e luz solar ganharam investimento e espaço. Uma dessas energias que merece destaque é a energia proveniente dos ventos, denominada energia eólica. Seu princípio básico é a utilização da energia cinética proveniente do vento para produção de energia elétrica. Essa fonte de energia, segundo dados obtidos por Ritchie, Roser e Rosado (2022), chegou a corresponder a cerca de 5% de toda a produção de energia no mundo no ano de 2019.

No Brasil, a energia eólica ganhou força nos últimos anos e passou a ser a segunda fonte da matriz energética brasileira com mais de 10% de participação, segundo dados da ABEEÓLICA (2022). Segundo um estudo do mesmo instituto, o nordeste brasileiro é a região que concentra a maior parte do setor de geração de energia eólica do Brasil, e um dos que mais possuem condições ideais para o desenvolvimento dessa atividade, como ventos fortes e constantes. No período chamado de “safra dos ventos”, que se estende de junho até o final do ano, os níveis de produção na região batem recorde, chegando até a atender toda a região durante um dia inteiro e cerca de 20% de todo o país.

A instalação dos complexos eólicos na atualidade se dão em regiões muitas vezes desabitadas e de difícil acesso, como regiões de alta elevação, desertos, até mesmo no mar, o que causa um alto custo para a manutenção. Recentemente, no início do ano de 2022, o governo brasileiro publicou o decreto Nº 10.946 que segundo a ABEEÓLICA (2022), possibilita a concessão de uso de espaços físicos e o aproveitamento dos recursos naturais no mar para a geração de energia elétrica a partir de empreendimentos *offshore*. A partir desse decreto é possível definir novos rumos para a produção de energia eólica *offshore*, segundo a qual as turbinas eólicas são instalados não mais em terra firme (*onshore*) e sim no mar. Devido a essa conquista, se fazem necessários mais investimentos por parte do setor de energia eólica para monitorar e acompanhar o funcionamento normal e detecção de falhas em seus aerogeradores, a fim de diminuir os altos custos com manutenção ou reparo em localidades de difícil acesso.

Nesse sentido, a Ciência de Dados é uma importante ferramenta, visto que os aerogeradores atualmente são equipados com diversos sensores, que fornecem informações de velocidade e direção do vento, posição da nacela, potência, entre outras. Na maioria dos casos, esses dados são captados por um sistema supervisorio SCADA (Supervisory Control And Data Acquisition), muitas vezes em uma resolução de 10 em 10 minutos, o que acarreta em uma quantidade imensa de informação. Trabalhar com essa enorme quantidade de informação é um grande desafio, seja pelo volume de informações ou pela significância dessas informações, visto que muitas vezes elas podem não refletir a realidade, pois não raro as perdas de conexão e degradação dos sensores podem ocasionar dados inválidos, os quais, para aplicações de operação normal, se tornam um empecilho. Diante do exposto, esse trabalho faz a aplicação de técnicas difundidas na literatura para remoção desses dados anormais, além de aplicar uma metodologia para obtenção das curvas normais dos aerogeradores provenientes desses métodos de limpeza. Com isso, obtêm-se as curvas que expressam as características normais de cada máquina, as quais se tornam uma ferramenta para analisar também o desempenho das máquinas com o passar do tempo.

1.1 Objetivos

- **Objetivos gerais:** Os objetivos gerais deste trabalho são obter as curvas de operação normal de aerogeradores, para analisar o desempenho dessas máquinas ao longo do tempo, como também obter uma forma simples de representar e armazenar essas curvas.
- **Objetivos específicos:** Os objetivos específicos deste trabalho são utilizar métodos de remoção de outliers na curva de potência presentes na literatura, detalhando de forma geral o comportamento desses métodos sobre uma base de dados real, como também, a partir dos resultados desses métodos, traçar a curva de potência normal, e assim obter funções que expressem o comportamento dessa curva a fim de fazer análises de desempenho.

1.2 Estrutura do Trabalho

Na sequência, o Capítulo 2 aborda as fundamentações teóricas que compõem o alicerce deste trabalho, como explicações sobre o funcionamento e composição dos aerogeradores, sobre a curva de potência e sobre os métodos de remoção de outliers (método interquartil e DBSCAN), além de tratar dos eventos presentes nos aerogeradores que expressam as condições nas quais essas máquinas se encontram. O Capítulo 3, por sua vez, explica a metodologia aplicada, na obtenção e pré-processamento de dados, além dos princípios das técnicas aplicadas na remoção de outliers. O Capítulo 4 trata dos resultados

obtidos na aplicação dos métodos propostos na literatura, como também da obtenção da curva normal dos aerogeradores. Por fim, o Capítulo 5 traz as principais conclusões e resume as contribuições deste trabalho.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo, são apresentados os conhecimentos necessários para entender como funciona o processo de geração de energia eólica, passando do princípio básico até a constituição dos aerogeradores. Também serão apresentadas formas de analisar a produção de energia por meio da curva de potência, além de técnicas relacionadas a remoção de dados anormais (outliers) dessa mesma curva. Por fim, será explicada a técnica de interpolação utilizada para obter as funções que regem a curva de operação normal do aerogerador.

2.1 Energia eólica e aerogeradores

A energia eólica é a energia obtida por meio do vento, e a origem desse termo se remete à palavra grega Éolo que significa “senhor dos ventos”. Essa forma de obtenção de energia data desde o início da civilização humana, visto que o homem aproveitada a força dos ventos para movimentar embarcações e moinhos para moer grãos, até chegar nos dias atuais, com os grandes aerogeradores. O vento em questão é gerado pela diferença de pressão na qual o ar mais quente sobe e o ar da região de menor pressão desce para ocupar o espaço. Essa energia é capturada pelo aerogerador, transformando energia cinética em energia elétrica.

Um aerogerador é formado por diversos componentes que, juntos fornecem a ele a capacidade de aproveitar a natureza dos ventos para produzir energia. Esses componentes podem ser divididos em Pás, Rotor, Torre, Nacele, Gearbox (caixa multiplicadora), Gerador e Anemômetro. Abaixo são descritas as características de cada componente:

- **Pás:** As pás têm a função de captar o vento e converter sua potência para o centro do rotor. Para isso, elas utilizam princípios aerodinâmicos para maximizar a captação. Sua constituição é feita principalmente por diversos polímeros, como fibra de vidro e poliéster, que as permitem ter uma alta resistência e um baixo peso.
- **Rotor:** O rotor tem a função de fixar as pás e transmitir o movimento de rotação para o eixo de movimento. Seu principal componente é um sistema hidráulico que possibilita o movimento das pás em diferentes posições, a fim de melhorar a força do vento, como também possibilitar a interrupção da turbina por completo, por questões de segurança.
- **Torre:** A torre é o maior elemento do aerogerador, e sua função é de sustentar o rotor e a nacele, como também permitir uma altura ideal para o funcionamento da máquina. Sua constituição varia de implantações em aço e concreto.

- **Nacele:** A nacele é caracterizada por ser um compartimento instalado no alto da torre de sustentação, e pode ser dividida em Sistemas de Controle Eletrônico, Sistema Hidráulico, Sistema de Yaw, Gearbox e Gerador. Por possuir esses diversos componentes a nacele se torna o componente mais pesado do aerogerador.
- **Gearbox:** A gearbox, ou caixa multiplicadora, possui a função de transformar as rotações que as pás transmitem ao eixo de baixa velocidade, para que com isso seja entregue ao eixo de alta velocidade as rotações necessárias para que o gerador possa funcionar.
- **Gerador:** O gerador tem, por sua vez, a função de converter a energia mecânica do eixo em energia elétrica.
- **Anemômetro:** O anemômetro tem a função de medição da velocidade e direção do vento. É o componente cujos dados gerados são usados pelo Sistema de Controle para posicionar a turbina visando melhor captação do vento e maximização da potência.

A Figura 1 abaixo ilustra a constituição dos componentes citados acima e seu princípio de funcionamento:

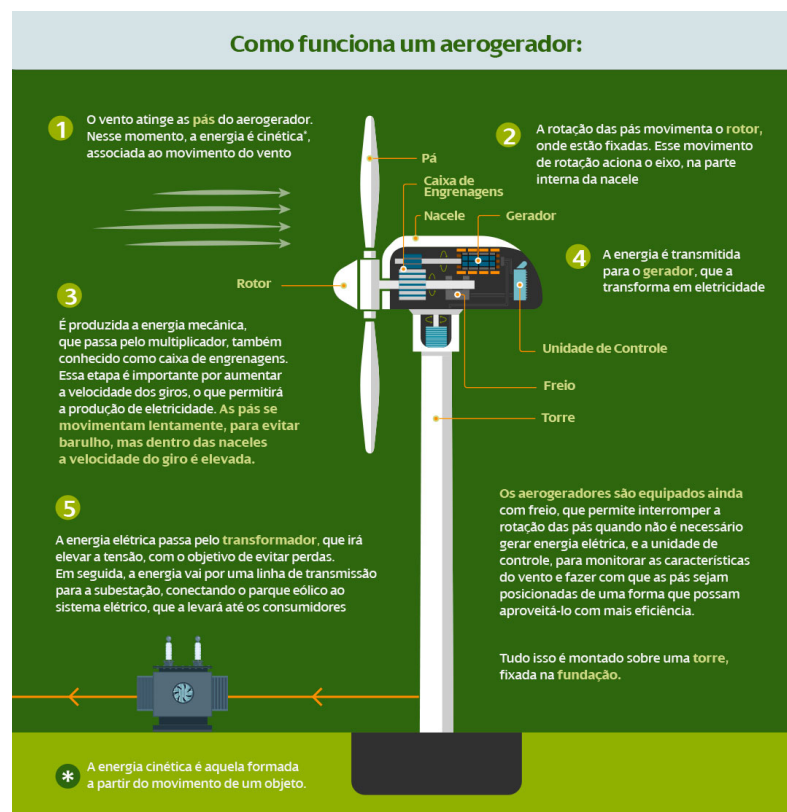


Figura 1 – Componentes e funcionamento de um aerogerador.

Fonte: (NEOENERGIA, 2022).

2.2 Curva de potência

Medir e analisar a potência de um aerogerador é um parâmetro chave para a análise de performance e saúde dele. Uma das formas de se obter esse dado é por meio da potência teórica gerada (P) por uma turbina. Essa informação é dada pela seguinte equação, proposta por Sohoni, Gupta e Nema (2016):

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

Os parâmetros descritos na equação são:

P - Potência de saída da turbina eólica (W)

ρ - Densidade do ar (kg/m^3)

A_w - Área coberta pelas pás do rotor (m^2)

C_p - Coeficiente de potência da turbina eólica

λ - Razão de velocidade da ponta da pá

β - Ângulo de inclinação ($^\circ$)

v - Velocidade do vento

Por meio da equação Equação 2.1, é possível verificar uma alta dependência de parâmetros como a velocidade do vento e sua direção, a densidade do ar (em função da temperatura, pressão e umidade), além de parâmetros atrelados à turbina. Existe muita complexidade ao relacionar todos os efeitos desses parâmetros, o que torna um trabalho muito complexo e conseqüentemente difícil de avaliar a potência de saída com base nessas condições. Em contrapartida, a esse problema surge o uso da curva de potência do aerogerador, que é uma forma mais simples e mais viável de estimar e modelar o desempenho, além do fato dela mostrar os dados reais, e não apenas teóricos.

A curva de potência de aerogeradores é dada pela potência em função da velocidade do vento. Sua obtenção se dá de duas formas, segundo Lira, Pinto e Heinzelmann (2019). A primeira forma consiste de medições realizadas em campo e a outra, de simulações computacionais. Aquela é tida como mais representativa, visto que considera todos os parâmetros atrelados ao ambiente na qual o aerogerador se encontra e também fatores operacionais e de fabricação. Já a curva obtida apenas por métodos computacionais, apesar de não levar em conta todos esses parâmetros, pode apresentar boas indicações da produção de energia em função da velocidade do vento. Com isso, segundo o mesmo autor, há uma

incerteza na curva de potência que consiste na representatividade dos resultados da curva de potência relacionada à metodologia de medição da curva.

Segundo Sohoni, Gupta e Nema (2016), a curva de potência de um aerogerador pode ser dividida em quatro regiões. A primeira região corresponde ao período em que a velocidade do vento é menor que um limite mínimo, conhecido como velocidade de cut-in ou velocidade de corte, na qual a potência de saída é zero. A segunda região está localizada entre a velocidade de cut-in e a velocidade nominal (velocidade necessária para começar a produzir). Nessa região ocorre um rápido crescimento da potência produzida. Já a terceira região corresponde ao período de saída constante (nominal) que é produzida até que a velocidade de cut-off seja atingida. Por último, temos a quarta região, que se define como a região em que a velocidade ultrapassa a velocidade de cut-off, na qual o aerogerador é retirado de operação para proteger sua estrutura de condições acima daquelas para as quais ela foi projetada para operar, e portanto, a produção de energia é zero nesta região. A representação dessas quatro regiões podem ser vistas na Figura 2.

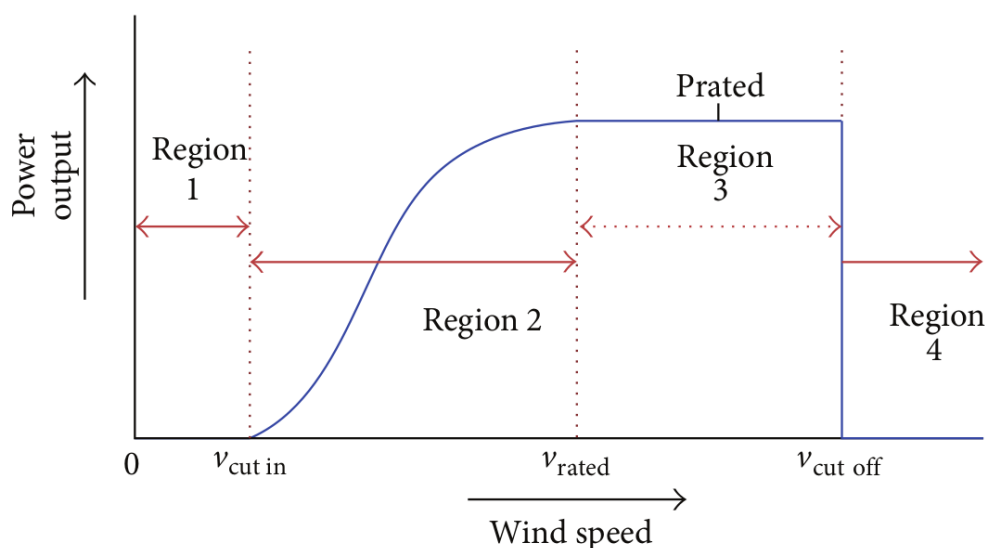


Figura 2 – Regiões da curva de potência.

Fonte: (SOHONI; GUPTA; NEMA, 2016, p. 2).

2.3 Eventos associados aos aerogeradores

Segundo Leahy et al. (2017), os sistemas de eventos de turbinas variam entre os fabricantes, mas geralmente se comportam de maneira semelhante. No nível mais básico, um evento de turbina (também chamado de mensagem de status ou alarme por alguns fabricantes) é gerado para indicar que o estado operacional da turbina mudou. Os fabricantes geralmente atribuem pelo menos três níveis diferente de gravidade aos eventos da turbina:

- **Information:** Os eventos de informação geralmente comunicam mudanças em certas condições operacionais, por exemplo, quando a velocidade do vento é muito baixa para geração, ou quando um interruptor manual foi acionado.
- **Warning:** Os eventos de aviso, por outro lado, são gerados quando o sistema de controle detecta condições operacionais ou variáveis de controle que se aproximam de exceder certos limites.
- **Fault:** Os eventos de falha são gerados quando esses limites são excedidos.

Ainda segundo Leahy et al. (2017), sempre que uma turbina desliga, vários eventos são gerados. Esses eventos podem persistir e permanecer ativos durante o desligamento ou podem ser apagados antes que a turbina volte a funcionar. O intervalo entre o evento ser “limpo” (geralmente como resultado de uma reinicialização remota) e a turbina voltar a funcionar e gerar pode ser de várias horas se a turbina precisar de uma visita ao local para inspeção manual. Além disso, muitos eventos podem ser acionados simultaneamente durante uma parada, o que torna difícil atribuir a parada a um único evento. Isso significa que tentar rotular os dados de acordo com instâncias específicas de alarmes e os horários em que eles estiveram presentes pode levar a uma classificação errônea, sendo necessário agrupar eventos que expressam a mesma característica a fim de classificar de forma mais efetiva o estado.

2.4 Outliers

Segundo Butler, Ringwood e O’Connor (2013), a degradação e precisão dos sensores, bem como falhas de comunicação, podem gerar valores errôneos, mesmo em condições normais. Durante a operação de um parque eólico, pode haver muitos outliers devido ao desligamento da turbina, redução de carga, ruído de comunicação e falha do equipamento. Trabalhar com dados brutos sem uma rotulação e limpeza pode acarretar resultados errôneos.

Existem muitas causas para anomalias de dados, incluindo o tempo de inatividade da turbina, redução do vento, falha do sensor de velocidade do vento, falha do equipamento de comunicação, interferências eletromagnéticas, condições climáticas extremas e pás das turbinas danificadas ou sujas. Segundo Shen, Fu e Zhou (2019) essas anomalias são visíveis na curva de potência, as quais podem ser divididas em quatro tipos: outliers empilhados da curva inferior, da curva intermediária e da curva superior, bem como os outliers dispersos ao redor da curva. A Figura 3 destaca esses tipos de outliers na curva de potência. A seguir, cada tipo será explicado em detalhes.

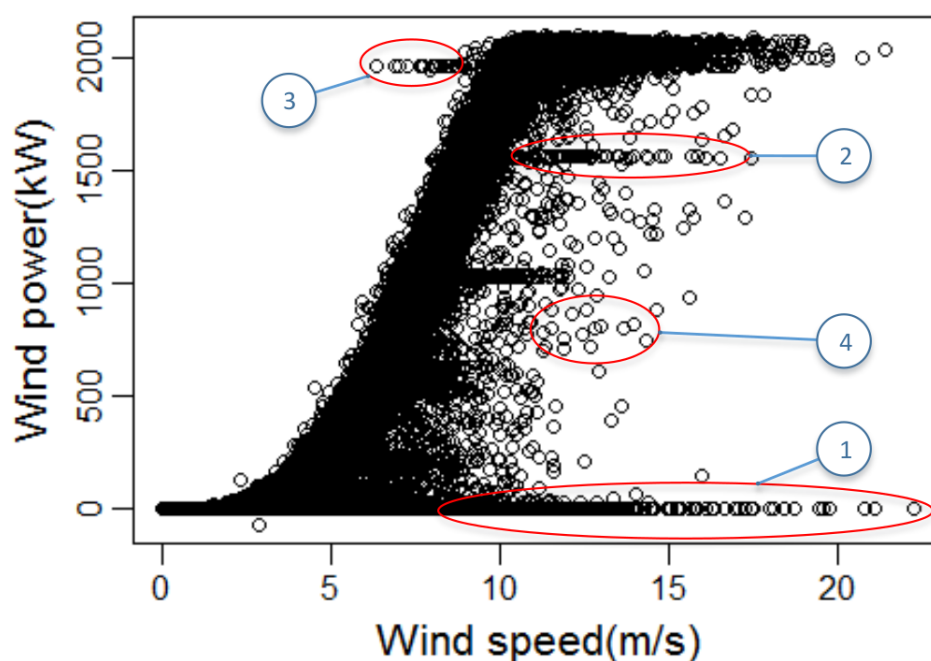


Figura 3 – Tipos de outlier na curva de potência.

Fonte: (SHEN; FU; ZHOU, 2019, p. 2).

1. **Outliers empilhados na curva inferior:** São caracterizados por uma concentração de dados na horizontal da curva de potência. As causas principais desses dados anormais incluem falha no equipamento de comunicação da turbina ou falha no terminal de medição, e manutenção não planejada. Nestes casos, a potência teórica na saída das turbinas é zero. Quando as pás da turbina não giram e o sistema de controle precisa de eletricidade, pode acontecer que a energia eólica seja negativa. Para tanto, os outlier empilhados na parte inferior da curva geralmente flutuam nas proximidades da potência zero e são caracterizados pela concentração de dados.
2. **Outliers empilhados no meio da curva:** Esse tipo de outliers são representados por uma ou mais bandas de dados densos agrupados de forma horizontal, e abaixo do limite inferior da curva de potência. Esse tipo de dados anormais são geralmente resultado da redução da produção ou falha de comunicação, sendo chamados de restrições de potência. A redução da produção é uma série de medidas de controle utilizadas para limitar a potência de saída das turbinas eólicas para mantelas em condição normal de operação. Na operação real de um parque eólico, a redução forçada da produção é muito comum devido à capacidade limitada do atual sistema de energia. Portanto, existem grandes quantidades de outliers nos dados originalmente registrados. Neste caso, os dados de velocidade do vento registram a mudança real, e a potência de saída das turbinas eólicas é mantida em um nível baixo por um longo período. Mesmo que a velocidade do vento exceda a velocidade nominal, a potência de saída é constante em um determinado ponto, inferior à da potência total.

Portanto, os outliers empilhados no meio da curva permanecem inalterados, abaixo da operação total, em intervalos de velocidade do vento relativamente grandes.

3. **Outliers empilhados na curva superior:** Os dados anormais empilhados na curva superior são representados por uma ou mais bandas de dados densos horizontais além do limite superior da curva de potência. A causa de tais outliers é geralmente dada por erros de comunicação ou falha no sensor de velocidade do vento. Esse sensor é frequentemente danificado, resultando em medições incorretas de velocidade do vento. Os outliers empilhados na curva superior manterão a potência constante durante os períodos de programação de baixa velocidade do vento e são caracterizados pela concentração de dados em uma, ou mais faixas horizontais.
4. **Outliers dispersos ao redor da curva:** Os outliers dispersos ao redor da curva são alguns pontos dispersos irregularmente perto da curva de potência. Esses dados anormais são devidos a fatores aleatórios, como ruído de propagação do sinal, falhas de sensor e condições climáticas extremas. Os dados anormais causados por fatores aleatórios irão flutuar nas proximidades dos valores normais, e a flutuação também é aleatória. Portanto, os outliers dispersos ao redor da curva serão distribuídos aleatoriamente além do limite da curva de potência.

2.5 Métodos de remoção de outlier da curva de potência

2.5.1 Método interquartil

Segundo Devore e Silva (2006) os quartis são divisões de um conjunto de dados em quatro partes iguais. As divisões são denominadas Q_1, Q_2 e Q_3 , que são primeiro quartil, segundo quartil e terceiro quartil. O primeiro quartil divide os valores correspondente aos valores 25% mais baixos, dos 75% que são maiores do que eles. O segundo quartil, por sua vez, corresponde aos 50% dos valores menores do que a mediana e 50% são maiores. Já o terceiro quartil divide a parcela correspondente a 75% dos valores mais baixos dos 25% que são maiores do que eles. O segundo quartil é dado pela mediana do conjunto de dados, já o primeiro e terceiro quartil são dados pelas equações abaixo, onde n é o número total de amostras:

$$Q_1 = \frac{n + 1}{4} \quad (2.2)$$

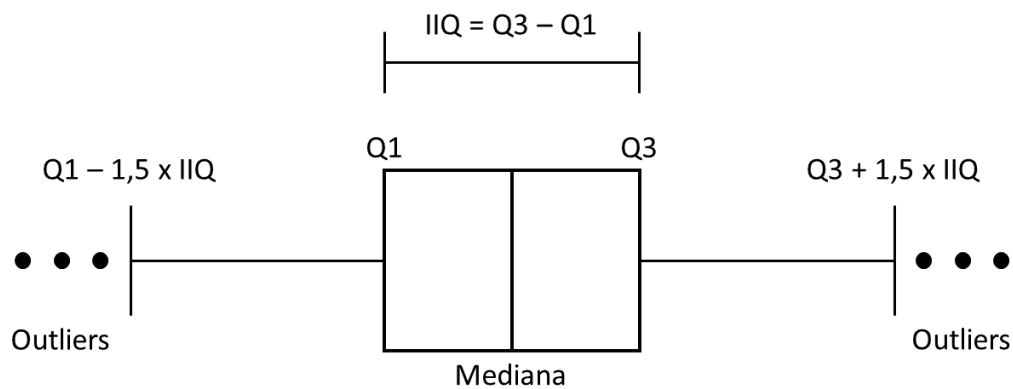
$$Q_3 = \frac{3(n + 1)}{4} \quad (2.3)$$

Além dos três quartis, temos também a amplitude interquartil, comumente chamada de dispersão média, a qual corresponde a diferença entre o terceiro quartil e o primeiro quartil de um conjunto de dados, conforme a equação abaixo:

$$IIQ = Q_3 - Q_1 \quad (2.4)$$

A amplitude interquartil mede a dispersão dos dados que estão entre as 50% das observações centrais. Portanto, não é influenciada por valores extremos.

Uma das aplicações desse método é o diagrama de caixa, comumente chamado de *boxplot*, que é uma ferramenta gráfica que permite a visualização da distribuição dos valores e a discrepância entre eles (outliers), sendo obtidos por meio dos limites inferiores e superiores do boxplot. A representação do boxplot pode ser vista na Figura 4 abaixo.



Box-plot

Figura 4 – Representação do gráfico do boxplot na identificação de outliers.

Fonte: (DIAS, 2022).

O limite de detecção de outliers é construído utilizando o limite interquartil. Sendo assim, os limites inferior e superior de detecção de outlier são dados por:

$$L_{inferior} = Q_1 - 1,5IIQ \quad (2.5)$$

$$L_{superior} = Q_3 + 1,5IIQ \quad (2.6)$$

2.5.2 DBSCAM

Segundo Chauhan (2022) o método do DBSCAN vem do nome *Density-Based Spatial Clustering of Application with Noise* que, traduzido, seria Clusterização Espacial Baseada em Densidade de Aplicações com Ruído. Seu princípio é baseado no agrupamento de dados, sendo esse um dos métodos de clusterização mais difundidos na área de aprendizagem de máquina, juntamente com o K-means. O DBSCAN é atrelado a informações relacionadas à densidade (parte espacial) de agrupamento dos dados e ao ruído atrelado.

Uma das especificidades desse método é que não é necessário estabelecer o número de clusters, em contraste com outros algoritmos de clusterização como o K-means, no qual é necessário definir o número de clusters esperado. Ou seja, não é preciso descobrir o melhor número de clusters, o próprio algoritmo irá fazer essa descoberta.

Em geral, ele é um método que se comporta melhor do que o K-means para dados assimétricos, e também se torna um algoritmo mais rápido. A sua ideia, de maneira geral, é localizar amostras de núcleo de alta densidade e expandir clusters a partir delas. Ele é composto por dois parâmetros principais, o *eps* e o *min_samples*, que são respectivamente, a distância máxima entre duas amostras para que uma seja considerada vizinha da outra e o número de amostras (ou peso total) em uma vizinhança para que um ponto seja considerado como um ponto central. A partir de um conjunto de pontos, o método agrupa pontos próximos entre si a partir de uma medida de distância e um número mínimo de pontos, e marca como outliers os pontos que estão em regiões de baixa densidade. Como pode ser visto na Figura 5 abaixo.

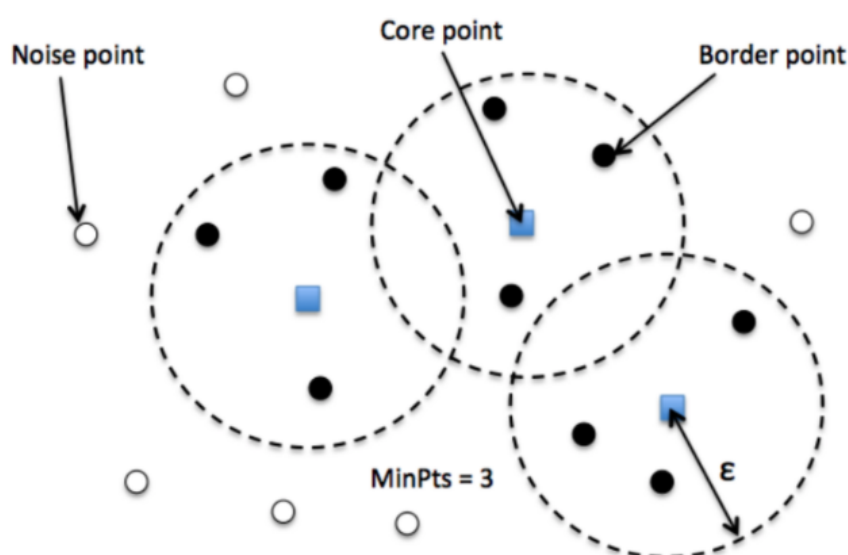


Figura 5 – Funcionamento do DBSCAN.

Fonte: (CHAUHAN, 2022).

Muitos trabalhos na área utilizam o DBSCAN juntamente com o método interquartil para remoção de outlier na curva de potência. Uma das implicações da utilização conjunta

desse métodos é observada pelos autores Yuan, Sun e Ma (2019), que argumentam que o método interquartil tem suas limitações, visto que ele pode identificar correta e efetivamente dados anormais apenas quando a proporção de dados anormais em relação à quantidade total de dados é pequena. Com isso para se utilizar do método interquartil de forma mais eficaz, pode-se usar primeiramente o DBSCAN para identificar e remover a maioria dos dados anormais e depois aplicar o método interquartil.

2.6 Interpolação polinomial

Uma interpolação polinomial é a busca de uma função polinomial que se ajusta melhor a um dado conjunto de pontos. A função buscada, portanto, tem a seguinte forma:

$$p_n(x) = \alpha_0 + \alpha_1 x + \dots + \alpha_n x^n \quad (2.7)$$

Queremos dessa forma, obter os coeficientes $\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_n$ tais que:

$$p_n(x_k) = \alpha_0 + \alpha_1 x_k + \dots + \alpha_n x_k^n = y_k, \quad \forall k = 0, 1, \dots, n \quad (2.8)$$

A equação 2.8 corresponde a um sistema linear com $n + 1$ equações e $n + 1$ incógnitas. Essa equação na forma matricial é denominada de matriz de Vandermonde que é demonstrada abaixo:

$$V\alpha = y \quad (2.9)$$

$$V = \begin{bmatrix} 1 & x_0 & x_0^2 & \cdots & x_0^n \\ 1 & x_1 & x_1^2 & \cdots & x_1^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^n \end{bmatrix}, \quad \alpha = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}, \quad y = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_n \end{bmatrix} \quad (2.10)$$

$$\begin{bmatrix} 1 & x_0 & x_0^2 & \cdots & x_0^n \\ 1 & x_1 & x_1^2 & \cdots & x_1^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^n \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_n \end{bmatrix} \quad (2.11)$$

Com base na equação 2.11, a resolução da forma de Vandermonde requer a solução de um sistema linear $n \times n$, em que n denota o grau do polinômio interpolador.

Neste trabalho, foi utilizado o Método dos Mínimos Quadrados (MMQ), que possibilita a obtenção da função de aproximação de um conjunto de pontos segundo um

critério de minimização de uma medida de erro pré-definida. A Equação 2.12 abaixo representa a função polinomial que será aproximada, na qual n é o grau do polinômio e e o erro.

$$y = \alpha_0 + \alpha_1 x + \dots + \alpha_n x^n + e \quad (2.12)$$

A medida de erro adotada é a de erro quadrático médio, que é representado pela distância entre os pontos experimentais e os pontos correspondentes da função aproximada. Esse erro pode ser visualizado na Figura 6, onde é feita uma aproximação por uma reta:

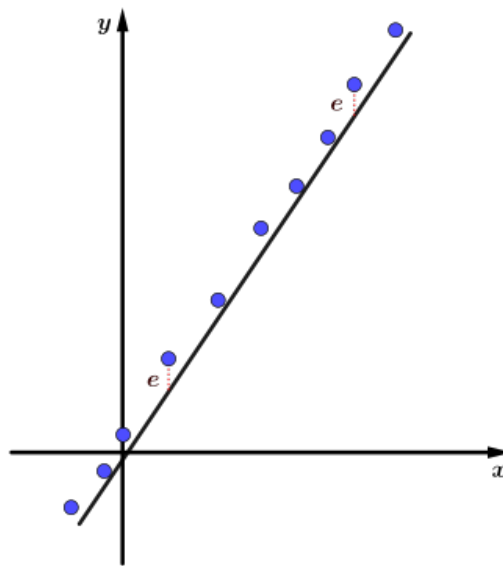


Figura 6 – MMQ: Erro da aproximação por uma reta.

Fonte: (NUMÉRICA, 2022).

Percebe-se que no MMQ não é necessário a função passar por todos os pontos. Portanto, a função buscada é aquela pela qual temos a minimização desse erro. O ajuste por MMQ é realizado de maneira a minimizar a soma dos quadrados dos erros para todos os pontos, ou seja, é a diferença entre y_{medido} da função original e o $y_{aproximado}$ calculado para a aproximação. Com isso a soma total dos erros é dado pela equação Equação 2.13 abaixo:

$$\begin{aligned} S_r &= \sum_{i=1}^n e_i^2 \\ &= \sum_{i=1}^n (y_{i,medido} - y_{i,aproximado})^2 \end{aligned} \quad (2.13)$$

3 METODOLOGIA

O presente capítulo se refere à metodologia aplicada neste trabalho, desde a obtenção dos dados até as etapas de pré-processamento e utilização de métodos de remoção de outliers na curva de potência, e por fim a obtenção das funções que representam a curva normal. A primeira seção será destinada à explicação do processo de obtenção dos dados brutos. A segunda seção irá tratar da avaliação dos valores de velocidade do vento e potência obtidos, como também da remoção de intervalos com valores faltantes dessas variáveis, além de realizar a apresentação das curvas de potência para cada aerogerador com os dados obtidos. A terceira seção irá tratar da obtenção dos dados de eventos dos aerogeradores selecionados na primeira seção. Nela serão explicados quais os principais campos presentes e como estão divididos os eventos. A quarta seção tratará da ideia de rotular os dados de potência e velocidade do vento usando os dados de eventos, para se obter os valores em que o aerogerador se encontrava apenas em condições normais. A quinta e sexta seção irão tratar dos métodos de remoção de outliers nas curvas de potência, respectivamente, do método do DBSCAN e do Interquartil. A sétima seção tratará da remoção de pontos de potência negativa ou que apresentam um valor positivo e estão no intervalo antes da velocidade de corte. Por fim, a oitava seção tratará da obtenção da curva normal do aerogerador e das funções que compõem essa curva.

3.1 Obtenção dos dados

A obtenção dos dados referentes às diversas variáveis presentes nos aerogeradores durante um período de tempo se deu por meio da conexão gRPC com um historiador que faz o rastreamento e armazenamento dos dados coletados pelo sistema supervisor SCADA. A obtenção foi feita utilizando a linguagem *Python* (v3.10.8), ao qual foram feitas requisições para o historiador utilizando *tags* contendo a identificação do aerogerador, a variável de interesse, e o período de interesse. Foram coletados dados para 4 aerogeradores entre os anos de 2018 a 2021, sendo esses dados de valores de potência ativa e velocidade do vento, além de dados de eventos desses aerogeradores. Os dados obtidos foram agrupados utilizando o pacote *Pandas* (v1.5.1) em diferentes *dataframes* contendo as informações referentes a cada aerogerador, e posteriormente exportados como arquivos *.parquet*. A escolha desse tipo de arquivo se deu pelo fato de que ele utiliza uma compactação dos dados e preserva a tipagem dos campos do *dataframe*.

3.2 Dados de potência e velocidade do vento

Os dados coletados referentes à potência e velocidade do vento foram obtidos aos pares durante o mesmo intervalo de tempo. Um exemplo da amostra desses dados para um determinado aerogerador pode ser visto na Tabela 1 abaixo:

Tempo	Velocidade do vento (m/s)	Potência ativa (kW)
2018-01-01 00:00:00	7,7	574,6
2018-01-01 00:10:00	6,4	390,6
2018-01-01 00:20:00	6,8	419,2
2018-01-01 00:30:00	6,6	366,2
2018-01-01 00:40:00	5,4	189,8

Tabela 1 – Amostra de dados de potência e velocidade do vento obtidos do SCADA.

Fonte: Elaborado pelo autor.

Ao todo foram obtidas 21040 instâncias de dados, onde cada instância é composta pelo instante da obtenção da amostra e os valores de velocidade do vento e potência ativa cada aerogerador durante 4 anos. Outro ponto a ser destacado é que todos esses dados estão com um intervalo de 10 minutos, que é a resolução utilizada para obtenção dos dados e são uma forma de agregação feita pelo SCADA. A Figura 7 apresenta a proporção da quantidade de dados por aerogerador e por ano.

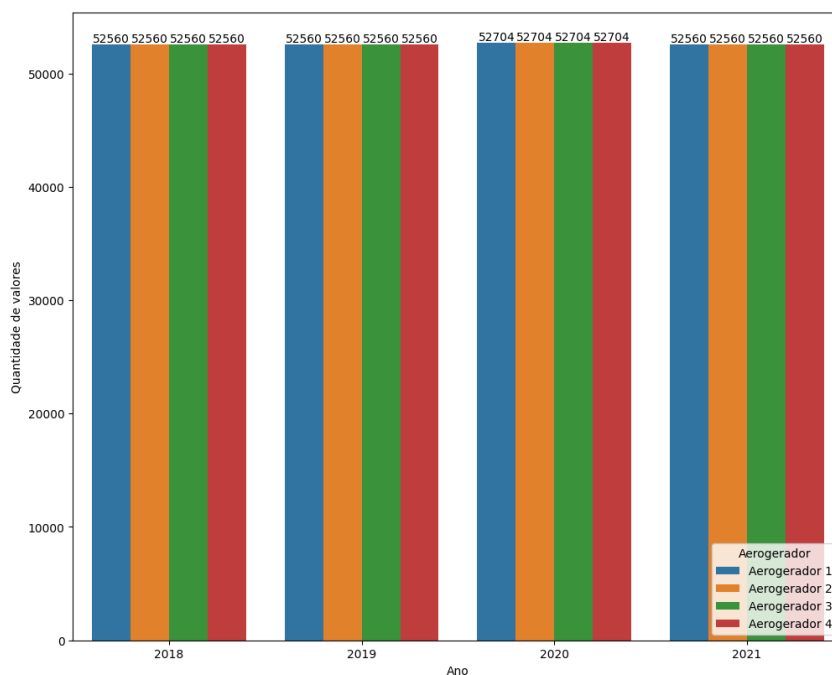


Figura 7 – Quantidade de valores brutos obtidos do SCADA para cada aerogerador.

Fonte: Elaborado pelo autor.

Posteriormente, foi analisada a consistência desses dados a fim de remover valores

nulos, foram observados períodos sem informações dos sensores, os quais faltantes foram removidos, ficando apenas dados que teriam informações das duas variáveis de interesse (potência e velocidade do vento). A Figura 8 mostra a quantidade de dados não nulos distribuídos para cada aerogerador durante os anos de 2018 a 2021.

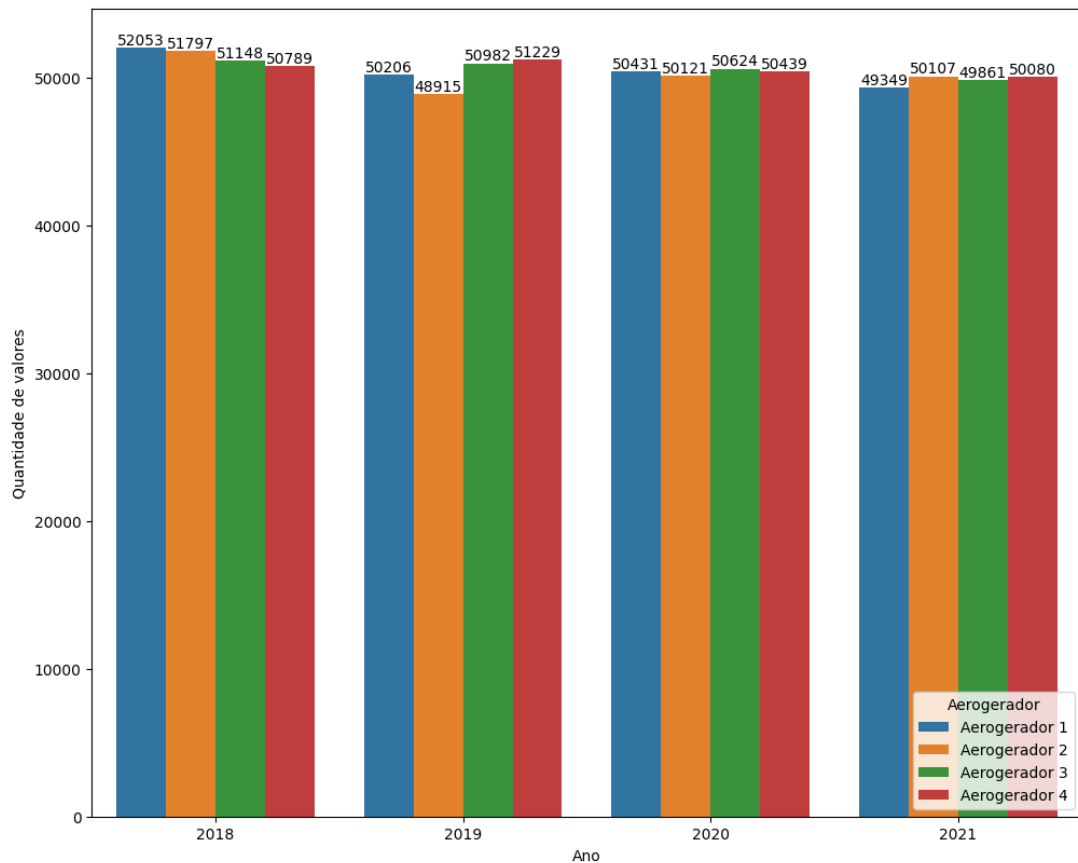


Figura 8 – Quantidade de valores não nulos obtidos do SCADA para cada aerogerador.

Fonte: Elaborado pelo autor.

De posse dos dados pré-processados, foram obtidas quatro curvas de potência para cada aerogerador, referentes a cada ano do período de amostragem dos dados. As figuras 9,10,11 e 12 a seguir são referentes às curvas obtidas para os 4 aerogeradores analisados nesse trabalho, e a plotagem desses curvas foram feitas utilizando as bibliotecas *Seaborn* (v0.12.1) e *Matplotlib* (v3.6.2).

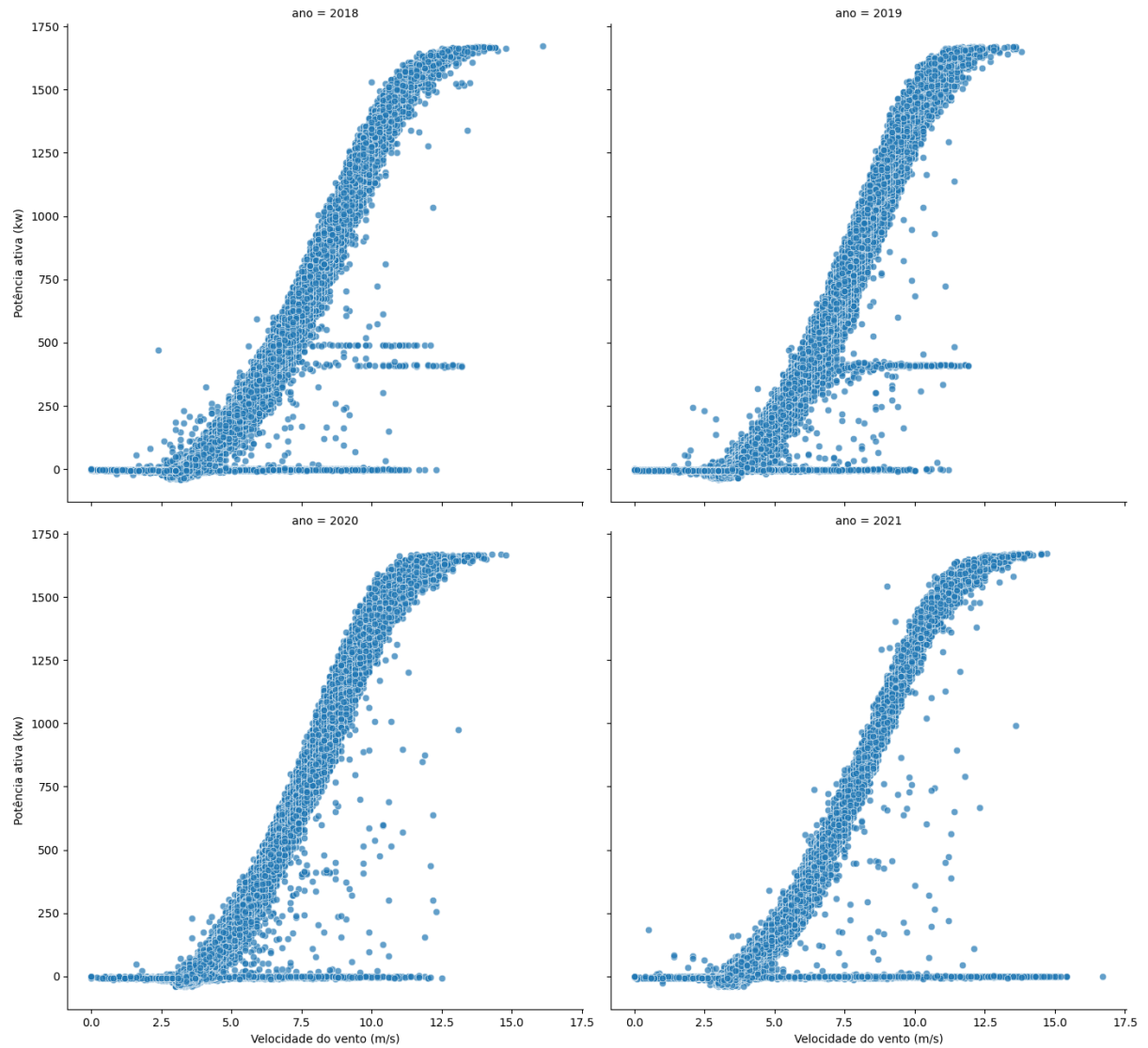


Figura 9 – Curva de potência por ano do aerogerador 1.

Fonte: Elaborado pelo autor.

Na Figura 9, que corresponde às curvas de potência do aerogerador 1, é possível identificar a presença dos outliers apresentados na fundamentação teórica deste trabalho, entre os quais podemos destacar pontos de restrição de potência bem nítidos nos anos de 2018 e 2019, como também outliers empilhados na curva inferior em todos os períodos e outliers dispersos ao redor da curva.

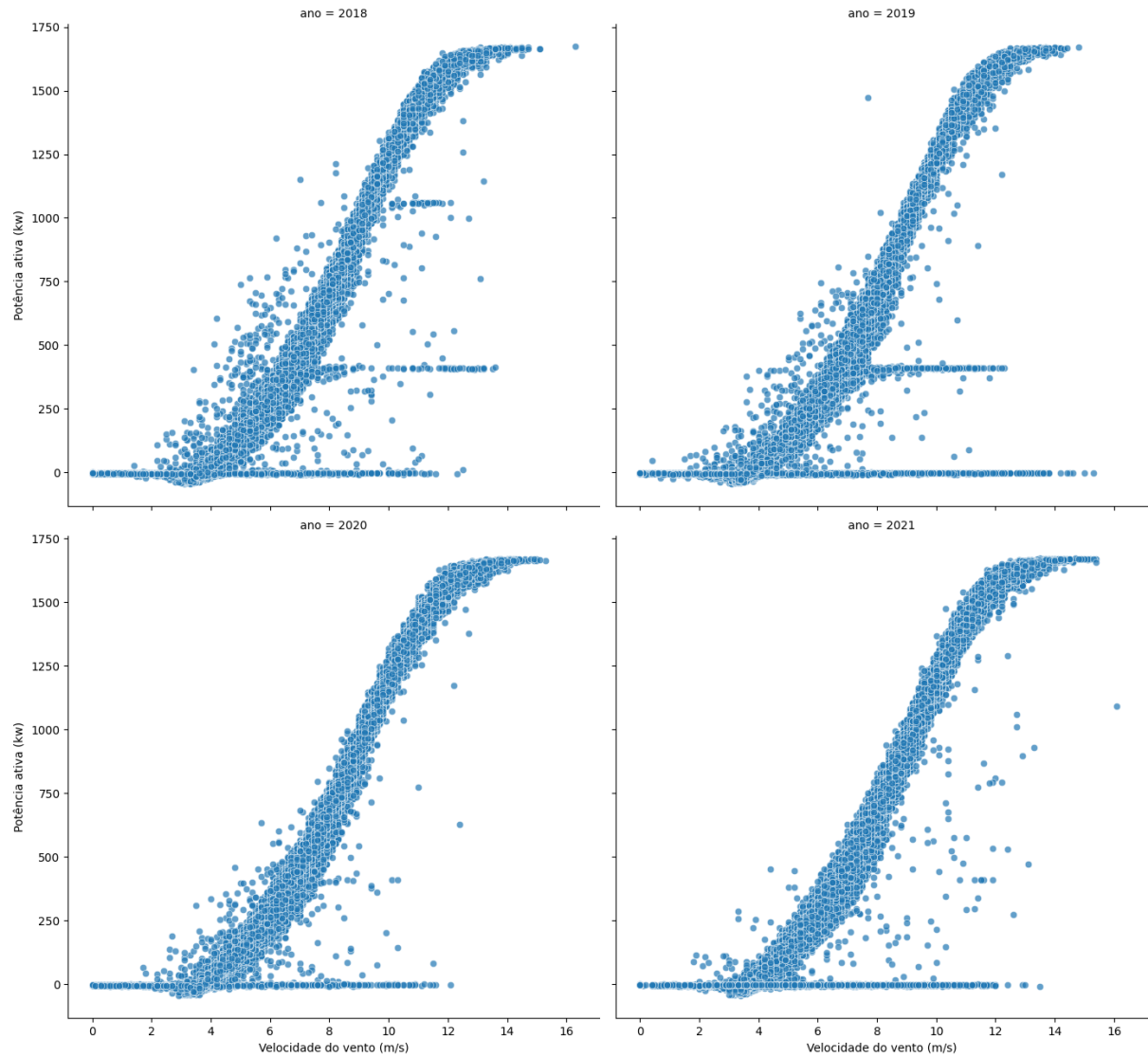


Figura 10 – Curva de potência por ano do aerogerador 2.

Fonte: Elaborado pelo autor.

Na curva de potência do aerogerador 2 presente na Figura 10, é possível notar também uma grande concentração de outliers e, assim como no aerogerador 1, ele apresenta, nos anos de 2018 e 2019, pontos com restrição de potência, e outliers empilhados na curva inferior, além de que a concentração de outliers dispersos é maior do que no aerogerador 1.

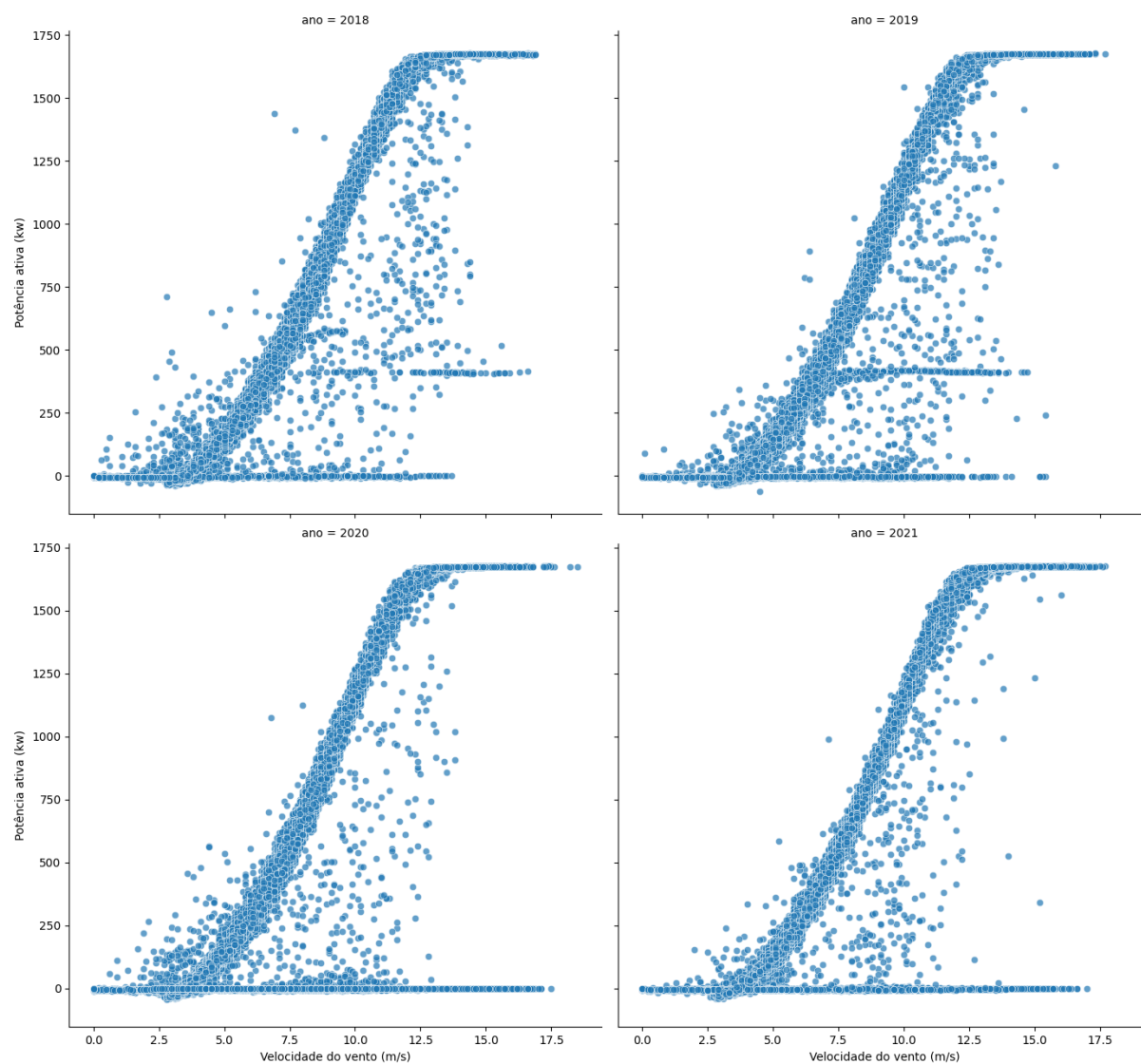


Figura 11 – Curva de potência por ano do aerogerador 3.

Fonte: Elaborado pelo autor.

Nas curvas apresentadas para o aerogerador 3 presente na Figura 11, é possível notar os mesmos tipos de outliers dos aerogeradores citados anteriormente, mas acontecendo uma concentração maior de outliers dispersos.

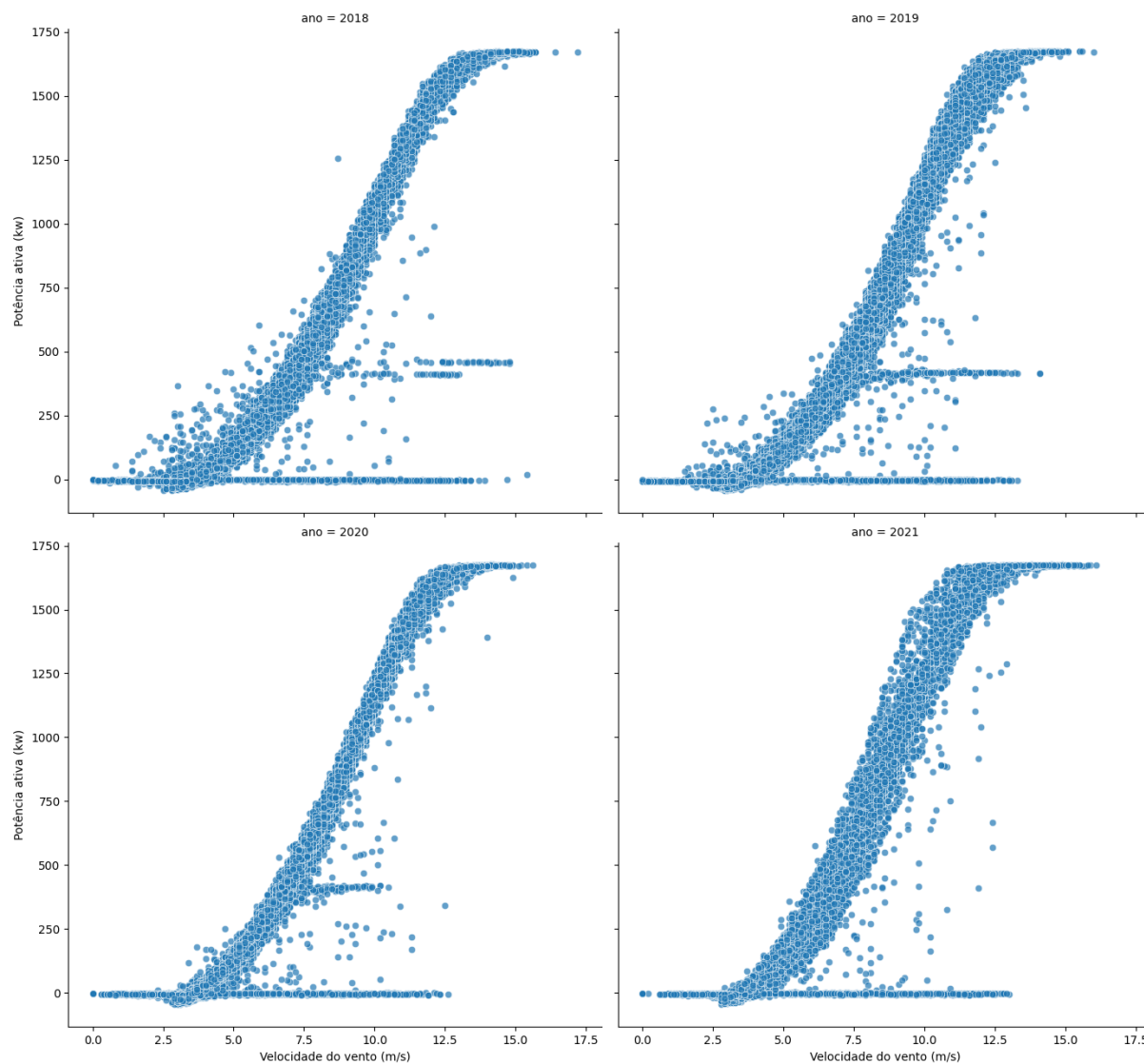


Figura 12 – Curva de potência por ano do aerogerador 4.

Fonte: Elaborado pelo autor.

Já para as curvas do aerogerador 4 presente na Figura 12, é possível verificar as mesmas ocorrências de outliers, sendo que esse aerogerador apresenta alguns fatores interessantes para a curva referente ao ano de 2021. Nela, é possível ver um alargamento da curva de potência em relação aos anos anteriores, como também a presença de restrição de potência para a curva do ano de 2020, algo que não ocorreu nos aerogeradores anteriores, que possuem essas restrições apenas nos anos de 2018 e 2019.

3.3 Dados de eventos

Foram obtidos dados de eventos relacionados aos 4 aerogeradores citados anteriormente durante os anos de 2018 a 2021. Os eventos obtidos apresentam informações como período inicial e final do evento, a sua duração, o código do evento, nome do evento,

período da última modificação, sua classificação e nível. Neste trabalho, foram utilizadas apenas as informações de código, início e fim do evento, nome do evento, classificação e nível. Uma amostra dessas informações pode ser vista na Tabela 2 abaixo.

Código	Inicia em	Finaliza em	Nome	Classificação	Nível
499	2018-01-19 11:22:05	2018-01-19 11:28:31	Unbalanced pitch - Generation	Manufacturer	ERROR
219	2018-01-23 15:53:01	2018-01-23 16:03:03	Wind sensor fault	Manufacturer	ERROR
0	2018-01-19 11:28:31	2018-01-19 11:28:31	Turbine OK	Normal operation	INFO
1	2018-01-29 18:12:45	2018-01-29 18:13:57	Power up	Normal operation	INFO
604	2018-01-29 16:51:18	2018-01-29 18:13:57	Remote stop	Owner	WARN
12	2018-01-29 16:53:01	2018-01-29 18:13:57	Grid drop voltage L1	Utility	ERROR
14	2018-02-04 01:39:42	2018-02-04 06:41:50	Grid drop voltage L3	Utility	ERROR
207	2018-02-12 04:07:39	2018-02-12 04:28:46	Long untwist CCW	Environmental	WARN
206	2018-12-24 02:50:55	2018-12-24 03:09:09	Long untwist CW	Environmental	WARN

Tabela 2 – Amostra de dados de eventos.

Fonte: Elaborado pelo autor.

Ao todo, foram 9606 eventos coletados para todos os aerogeradores. O aerogerador 3 foi o que mais apresenta eventos durante esse período, com cerca de 3930 eventos. Já o aerogerador 4 foi o que menos apresentou eventos, com cerca de 1441 eventos ao todo. Os outros dois aerogeradores restantes apresentaram juntos cerca de 2100 eventos. A Figura 13 mostra a quantidade de eventos por aerogerador durante o período analisado.

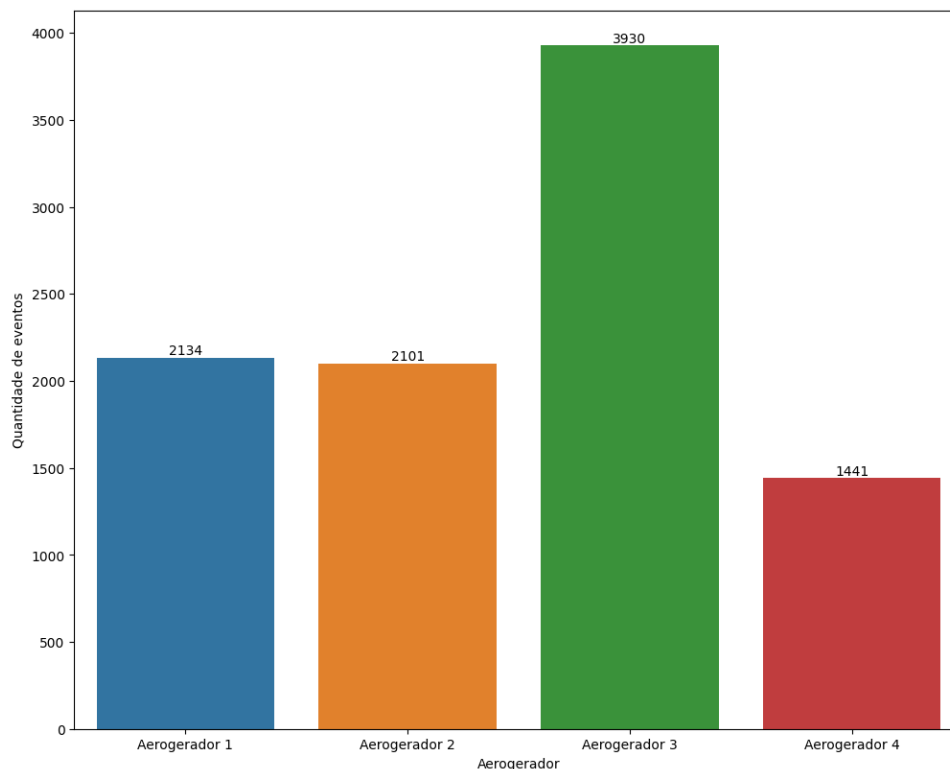


Figura 13 – Quantidade de eventos por aerogerador.

Fonte: Elaborado pelo autor.

Os modelos de aerogeradores analisados foram da Vestas. Este fabricante adota 5

categorias de classificação para os eventos, que são: *Manufacturer*, *Normal operation*, *Owner*, *Utility* e *Environmental*. Os eventos de *Manufacturer*, *Utility*, *Owner* e *Environmental* são eventos que influenciam na produção do aerogerador, representado, por exemplo, uma manutenção ou problemas em algum componente, ou até mesmo a parada remota do aerogerador. Já eventos de *Normal operation* são para sinalização de que o aerogerador está funcionando de acordo com a normalidade. A Figura 15 abaixo mostra a quantidade de eventos por classificação e por aerogerador para o período analisado.

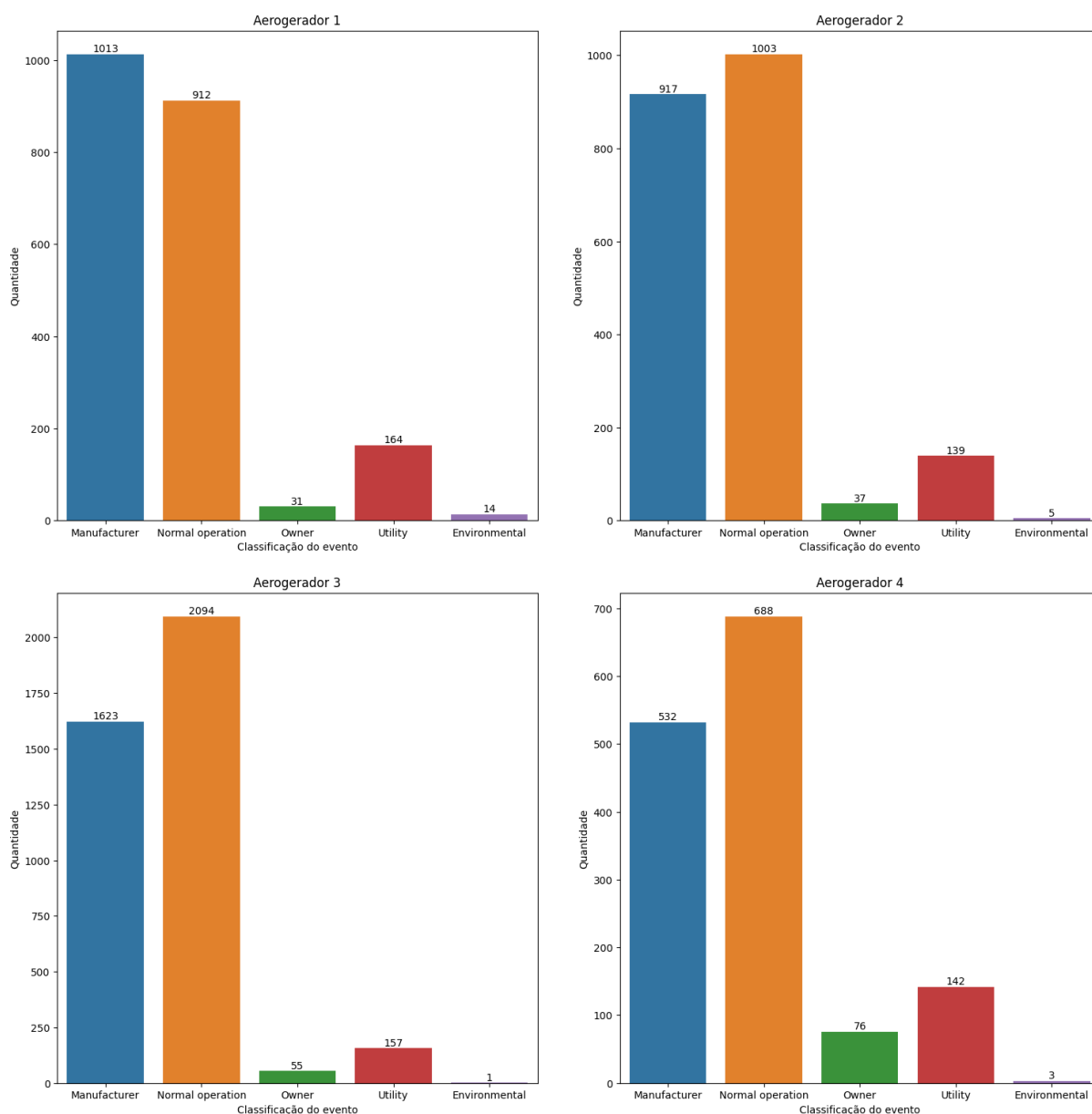


Figura 14 – Quantidade de eventos por classificação e por aerogerador.

Fonte: Elaborado pelo autor.

Outra característica dos eventos está relacionada ao seu nível, que foram divididos em *Error*, *Info* e *Warn*. Como foi explicado na fundamentação teórica os eventos de *Info* estão relacionados a informações de funcionamento dos componentes do aerogerador,

sinalizando mudanças de certas condições operacionais. Já eventos de *Warn* são para quando o sistema de controle detecta condições que se aproximam de exceder um determinado limite, próximo a ocorrer algo mais grave. E, por fim, os eventos do tipo *Error* são quando problemas são encontrados, ou, em outras palavras são emitidos quando o sistema detecta que algumas condições passaram do limite estabelecido. A quantidade de eventos por níveis para cada aerogerador pode ser visto na Figura 15 abaixo.

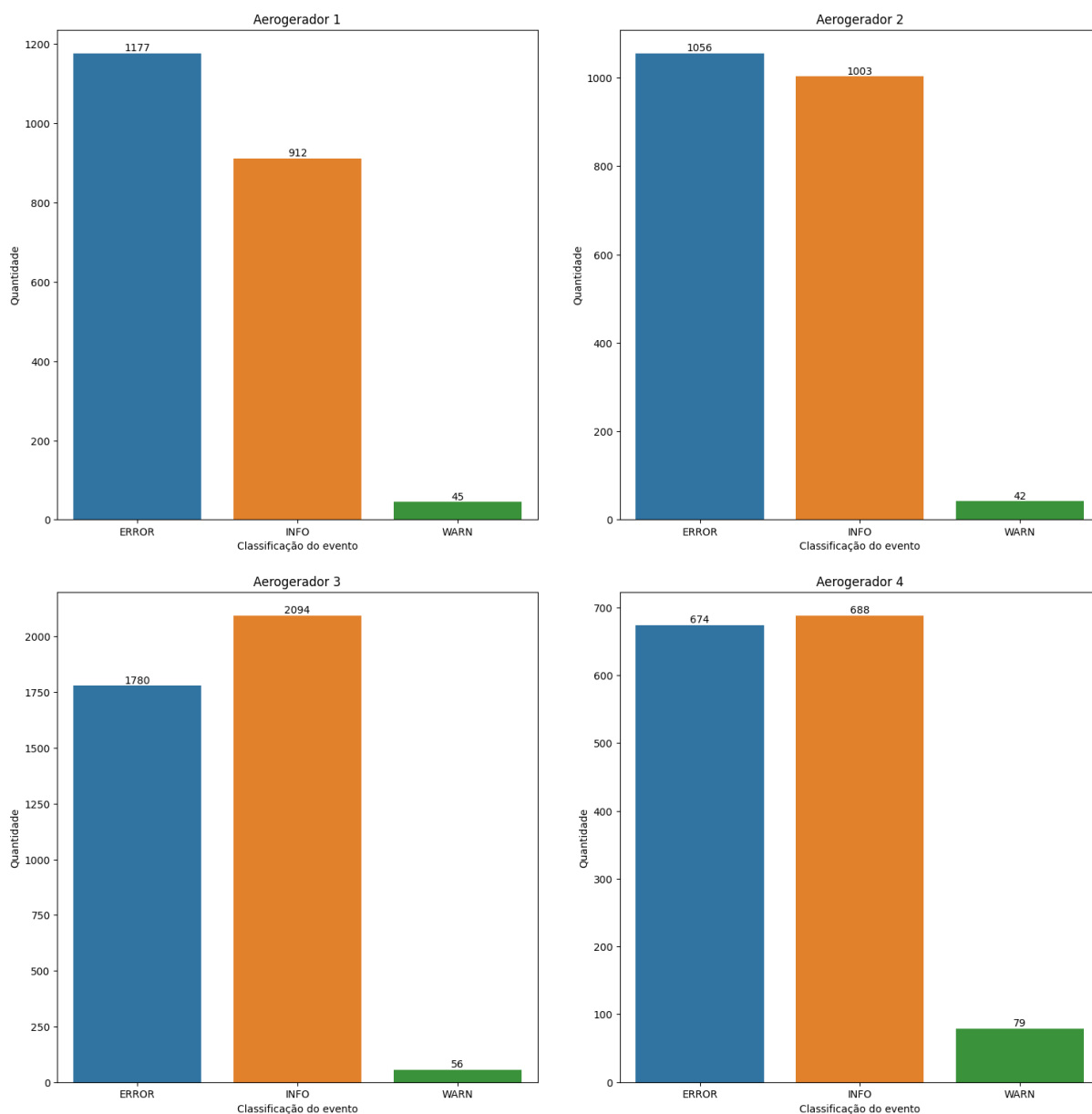


Figura 15 – Quantidade de eventos por nível e por aerogerador.

Fonte: Elaborado pelo autor.

3.4 Classificar pontos por eventos

A partir das informações de velocidade do vento e potência, é possível obter a curva de potência de um aerogerador, e assim observar o seu comportamento. Porém apenas esses dados não nos fornecem informações fidedignas, pois não sabemos em que condições os componentes e o aerogerador estavam se comportando, se nesse período de tempo ele estava sob manutenção, ou até mesmo se havia algum problema em um componente. Essas informações são relevantes e nos permitem definir que os dados foram obtidos ou não em condições normais.

Com base nos eventos e nos dados de potência e velocidade, podemos relacionar o período do evento ao período em que os dados das variáveis do aerogerador foram obtidas. O período pré-evento e pós-evento, muitas vezes devido à resolução em que os dados são obtidos, não expressam de forma correta o estado da máquina. Os pontos nesse período são chamados de pontos de transição. Eles expressam o estado da máquina um pouco antes e depois do evento, e por meio deles é possível expressar de forma mais completa o estado da máquina durante todo o intervalo do evento.

Uma característica que deve ser levada em conta nessa rotulação é que muitas vezes vários eventos são lançados ao mesmo tempo. Isso é um empecilho visto que às vezes não é possível definir qual evento em si ocasionou por exemplo uma parada no aerogerador, ou nesse mesmo período vários eventos que representam o mesmo estado são gerados com uma curta duração. Unir esses eventos que apresentam o mesmo nível ou efeito, ou que estão divididos em pequenos intervalos, é uma opção para melhorar a identificação do estado atual da máquina. Com base nessa premissa, o primeiro objetivo a ser realizado é filtrar eventos que coincidem com o período analisado e que expressam uma característica em comum. Para esse trabalho, serão utilizados apenas os eventos que causam influência na produção do aerogerador e conseqüentemente nas variáveis de potência e velocidade do evento. Com isso, serão selecionados eventos das classes de *Manufacturer*, *Owner*, *Utility* e *Environmental*, como também, a fim de se eliminarem redundâncias, será feito a união de eventos que se intersectam.

O Algoritmo 1, em pseudocódigo abaixo foi baseado, no algoritmo disponível em GeeksforGeeks (2022) e tem complexidade $O(n \log(n))$. Sua função é unir intervalos que se intersectam, e sua lógica foi aplicada para unir eventos que se intersectam.

Algoritmo 1: Unir eventos que se intersectam**Dados:** Lista de *intervalos* contendo o início e fim do intervalo**Saída:** Lista de intervalos mesclados

```

1 se tamanho(intervalos) = 0 então
2   | resultado ← lista vazia;
3 senão
4   | intervalos ← ordenarDeFormaCrescentePeloInicioDoPeriodo(intervalos);
5   | resultado ← intervalos[0];
6   para indice=1 até tamanho(intervalos) faça
7     | se intercepta(intervalos[indice], ultimoValorLista(resultado)) então
8       | inicio ← ultimoValorLista(resultado).inicio;
9       | fim ← maximo(intervalos[indice].fim,ultimoValorLista(resultado).fim);
10      | ultimoValorLista(resultado) ← Intervalo(inicio = inicio,fim = fim);
11      senão
12        | Inlui em resultados o valor de intervalos[indice]
13      fim
14   fim
15 fim
16 retorna resultados

```

Após obter e unir eventos que se intersectam, é necessário rotular os pontos com base na incidência dentro do evento, e também rotular os pontos pré e pós-evento. Também verificaremos se um ponto em questão é ou não de parada, e para isso verificaremos se a potência é menor que 0,01 e se a velocidade do vento é maior que a velocidade de corte (V_{cut-in}), a partir da qual o aerogerador começa a produzir. Para definir os estados de pré e pós-evento, utilizamos a resolução de aquisição dos dados, que é de 10 minutos para determinar esses estados. O Algoritmo 2 abaixo representa a lógica aplicada na rotulação dos pontos por meio dos eventos, no qual os pontos que coincidem dentro do período do evento, e não são de parada, serão classificados como “NA TRANSICAO DO EVENTO”. Já os pontos que coincidem no período de pré e pós evento serão classificados como de “TRANSICAO”.

Algoritmo 2: Classificar pontos

Dados: *dataset* contendo campos de tempo e valores de potência e velocidade do vento nesses instantes e uma lista de *periodos* contendo o início e fim dos eventos do resultado do processamento do Algoritmo 1

Saída: *dataset* com um novo campo com o tipo do ponto

```

1  cutin ← 4.0;
2  espacamento_em_minutos ← 10;
3  dataset ← criaNovoCampo(nome = “Tipo”, valor = “NORMAL”);
4  para indice_dataset=0 até tamanho(dataset) faça
5      para indice=0 até tamanho(periodos) faça
6          inicio ← periodos[indice].inicio;
7          inicio_menos_10min ← subtrai10Minutos(inicio);
8          fim ← periodos[indice].fim;
9          fim_mais_10min ← adiciona10Minutos(fim);
10         tempo ← dataset[indice_dataset].tempo;
11         velocidade ← dataset[indice_dataset].velocidade_vento;
12         potencia ← dataset[indice_dataset].potencia_ativa;
13         criterio_parada ← (velocidade ≥ cutin) & (potencia ≤ 0.00);
14         selecao_delta_inicio ← ((tempo ≥ inicio_menos_10min) & (tempo ≤ inicio));
15         selecao_delta_fim ← ((tempo ≥ fim) & (tempo ≤ fim_mais_10min));
16         selecao_dentro_do_periodo_do_evento ← (tempo ≥ inicio) & (tempo ≤ fim);
17         selecao_do_tipo_parada ← selecao_dentro_do_periodo_do_evento & criterio_parada;
18         selecao_tipo_transicao_no_periodo_do_evento ←
19             selecao_dentro_do_periodo_do_evento & ~ criterio_parada;
20         selecao_tipo_transicao ← selecao_delta_inicio | selecao_delta_fim;
21         se selecao_tipo_transicao_no_periodo_do_evento então
22             dataset[indice_dataset].Tipo ← “NA TRANSICAO DO EVENTO”;
23         pare
24         senão se selecao_tipo_transicao então
25             dataset[indice_dataset].Tipo ← “TRANSICAO”;
26         pare
27         senão se selecao_do_tipo_parada então
28             dataset[indice_dataset].Tipo ← “PARADA”;
29         pare
30     fim
31 retorna dataset

```

Para exemplificar o processo de classificação dos pontos usando os eventos as figuras 16, 17 seguintes mostram respectivamente os pontos antes da rotulação e pós rotulação.

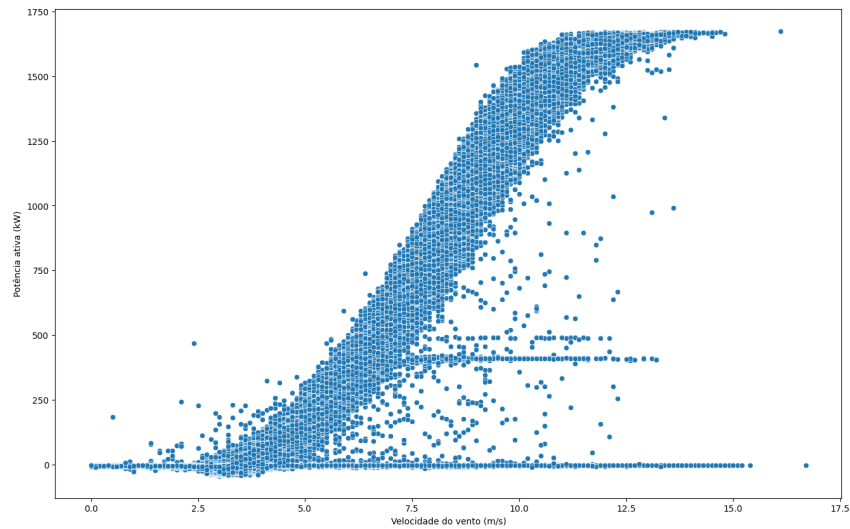


Figura 16 – Amostra com pontos antes da realização da classificação por eventos.

Fonte: Elaborado pelo autor.

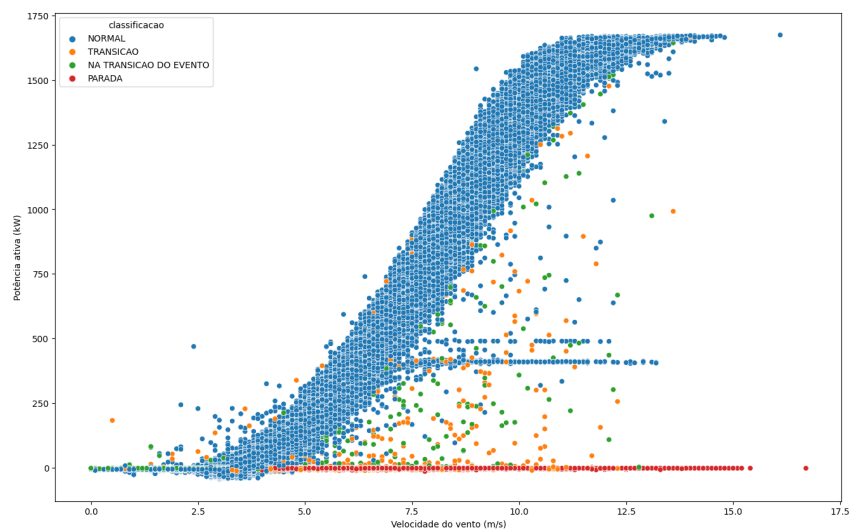


Figura 17 – Amostra com pontos rotulados pela classificação por eventos.

Fonte: Elaborado pelo autor.

Com base na Figura 17, é possível identificar que os outliers presentes a baixo da curva de potência foram classificados como pontos de parada e os outliers dispersos ao redor da curva foram classificados como pontos de transição. Com isso, vários outlier podem já ser removidos utilizando-se esse método.

3.5 DBSCAN

O DBSCAN é um método não supervisionado de clusterização, sendo muito usado para remover outliers em curvas de potência dada a sua natureza de agrupar pontos por densidade e com isso identificar ruídos; no caso da curva de potência, esses serão os

outliers dispersos ao redor da curva. Para aplicar esse método foi utilizado a biblioteca do *scikitlearn* (v1.1.3), e os parâmetros utilizados foram $eps = 1.5$ e $min_samples = 5$, que correspondem respectivamente à distância máxima entre duas amostras para que uma seja considerada vizinha da outra, e o número de amostras (ou peso total) em uma vizinhança para que um ponto seja considerado como um ponto central. Esses valores de parâmetros foram os que melhor se aplicaram a todo o conjunto dos 4 aerogeradores, e que conseguiram contemplar toda a curva de potência como um cluster e os outliers dispersos como ruído. A Figura 18 abaixo mostra o uso do método em uma instância dos dados, na qual é possível verificar a rotulação de pontos da curva como um cluster só (rotulação 0) e os pontos que não pertence ao cluster como ruído (rotulação -1).

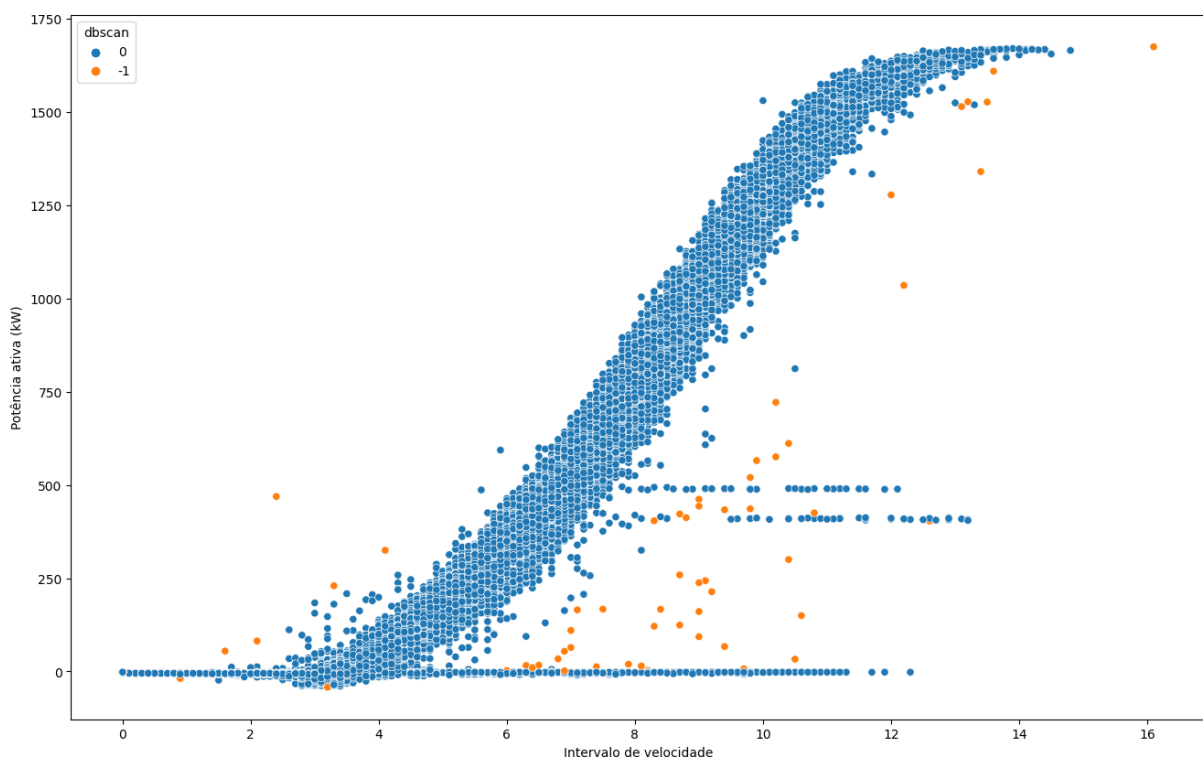


Figura 18 – Exemplo de identificação de ruído com o DBSCAN.

Fonte: Elaborado pelo autor.

3.6 Método interquartil

O método interquartil foi utilizado para identificar e remover de forma mais profunda os diversos tipos de outliers presentes na curva de potência. Para usá-lo, deve-se primeiro ordenar os valores de velocidade do vento e seccionar a curva de potência em intervalos de velocidade. Neste trabalho, foi utilizado o seccionamento da curva em intervalos que começam do menor valor de velocidade da curva (V_{min}) até o maior valor de velocidade (V_{max}), com um espaçamento de 0,5 entre cada intervalo. A Figura 19 abaixo mostra esse seccionamento da curva em intervalos de velocidade.

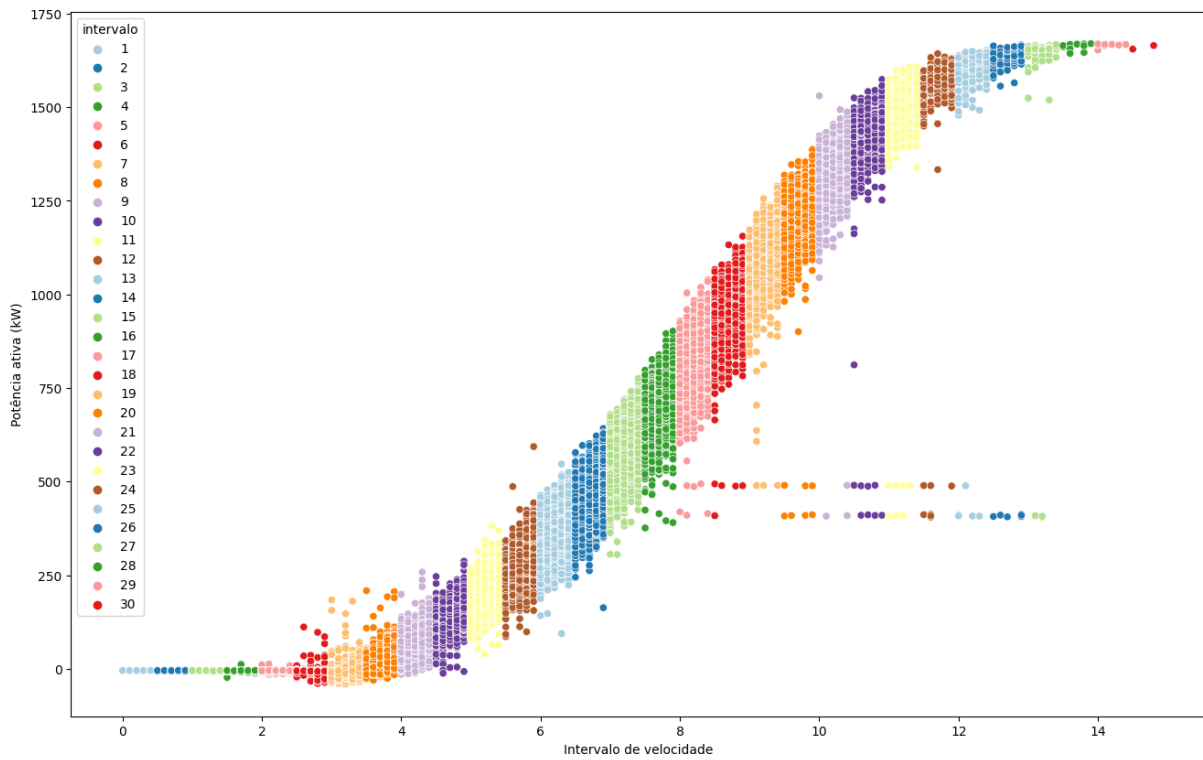


Figura 19 – Método interquartil para seccionamento da curva de potência em intervalos de velocidade do vento.

Fonte: Elaborado pelo autor.

Com a curva seccionada aplicamos os limites interquartil para cada intervalo levando em conta a variável de potência, e com isso obtemos os pontos dentro dos limites interquartis como pontos normais e os pontos fora desses limites como outliers. Uma representação mais fácil de entender seria aplicar o gráfico de boxplot para cada intervalo e com isso visualizar os pontos que estão fora dos limites interquartis (outliers), a Figura 20 abaixo mostra esse representação do método interquartil com boxplots em cada intervalo.

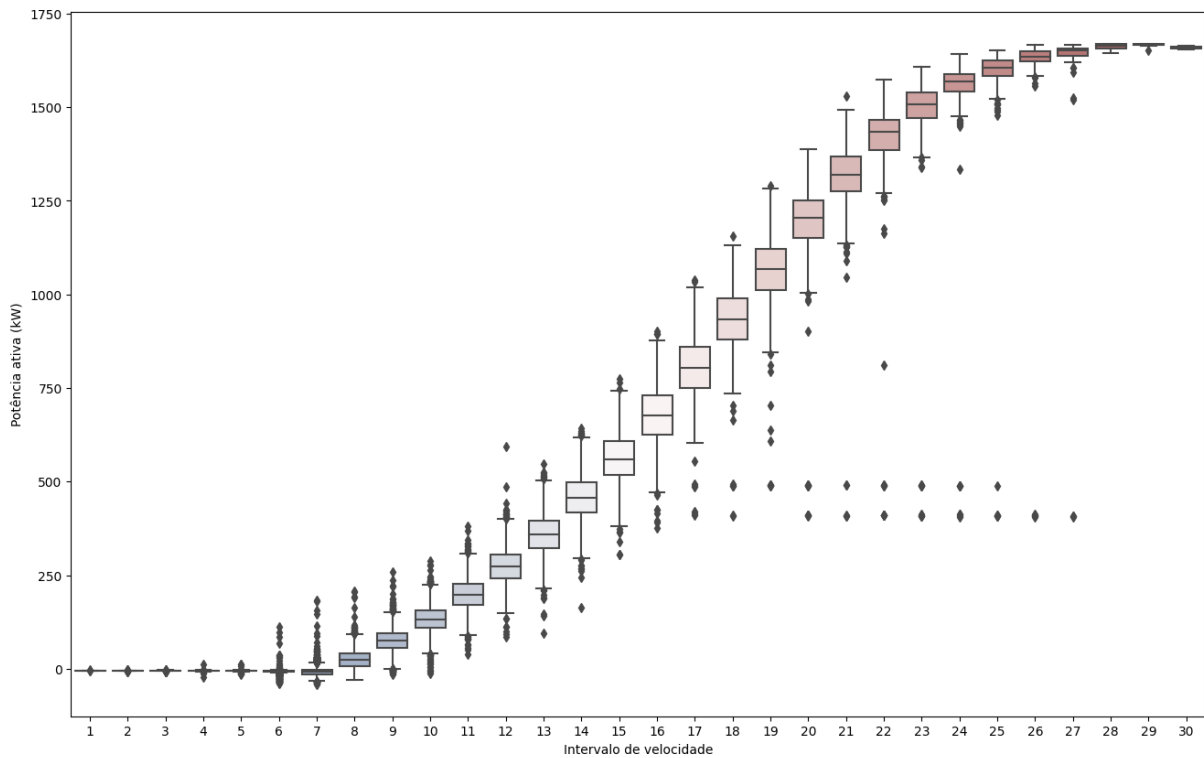


Figura 20 – Visualização do método interquartil como boxplot nos intervalos de velocidade do vento.

Fonte: Elaborado pelo autor.

3.7 Pontos de potência negativa e inválidos

Após a remoção de outliers, temos a curva de potência com uma forma mais compacta em forma de “S”. Porém, nesse estado ela apresenta alguns erros associados, que se referem a valores com potência negativa e valores de potência maior que 0 com velocidade abaixo de cut-in. Os valores de potência negativa ocorrem normalmente quando as pás da turbina não giram e o sistema de controle precisa de eletricidade. Já os valores de potência inválidas podem ser ocasionados por falhas dos sensores ou no processo de transmissão, já que só há produção de energia quando a velocidade de cut-in é atingida. A Figura 21 abaixo mostra a identificação desses pontos. Para que se tenha a curva de potência com condições normais, é necessário remover esses pontos e com isso obter a curva normal.

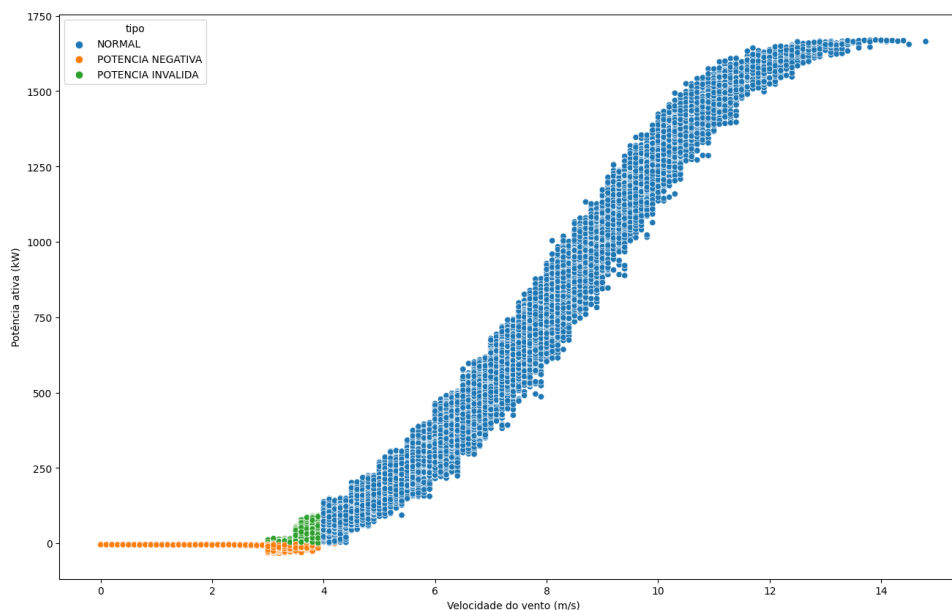


Figura 21 – Pontos de potência negativa.

Fonte: Elaborado pelo autor.

3.8 Curva de potência sob condições normais

Com a remoção dos outliers e pontos de potência negativa, temos a curva de potência com valores sob condições normais. Com base nesses valores, podemos identificar o envelope “S” da curva formada por duas assíntotas, que correspondem às partes superior e inferior da curva. O principal objetivo deste trabalho é justamente obter uma representação dessa curva e comparar como ela se comporta ao longo do tempo, permitindo comparar o desempenho dos aerogeradores. Podemos obter polinômios que parametrizem as duas assíntotas possibilitando o armazenamento de forma simples da curva por meio de seus coeficientes. Para implementar a parametrização da curva e obter os polinômios interpoladores das duas assíntotas, é necessário novamente dividir a curva de potência em intervalos de velocidade do vento e, em cada intervalo, obter os valores de potência mínima e máxima naquele intervalo. Com esses valores e o valor da velocidade do vento no início do intervalo, podemos obter os pontos que compõem as duas assíntotas. Pode acontecer que alguns intervalos não possuam pontos e com isso os pontos correspondentes teriam valor nulo. Para mitigar isso, ao final da obtenção dos pontos, os valores nulos são substituídos utilizando uma interpolação simples pelo próprio pacote do *Pandas* (*interpolate*) usando o método linear. A Figura 22 mostra o seccionamento em intervalos de velocidade do vento que levam em consideração os mesmos critérios do seccionamento do método interquartil. Já a Figura 23 mostra a marcação dos pontos de máximo e mínimo de potência com o valor de velocidade do vento no início do intervalo.

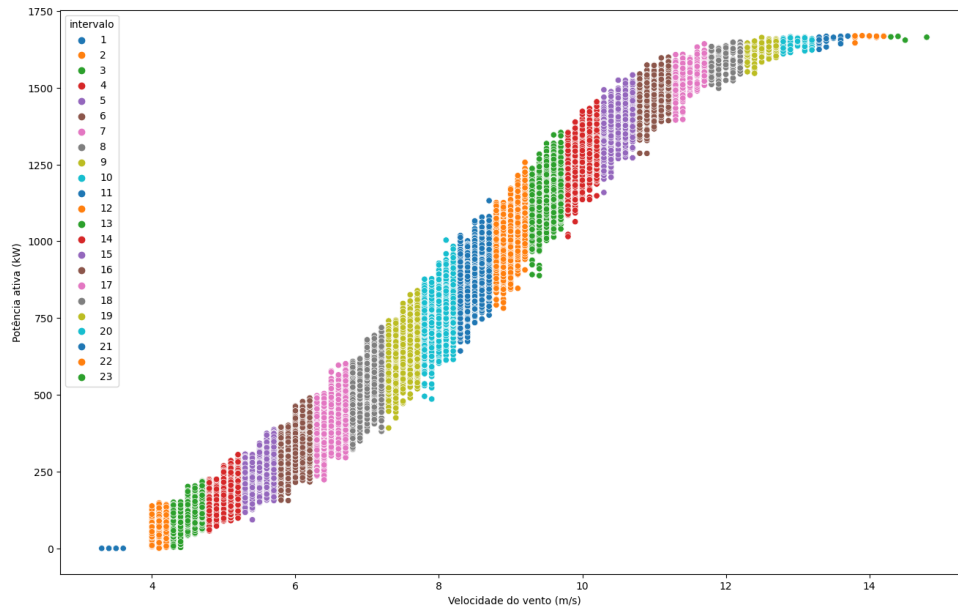


Figura 22 – Seccionamento da curva de potência normal em intervalos de velocidade do vento.

Fonte: Elaborado pelo autor.

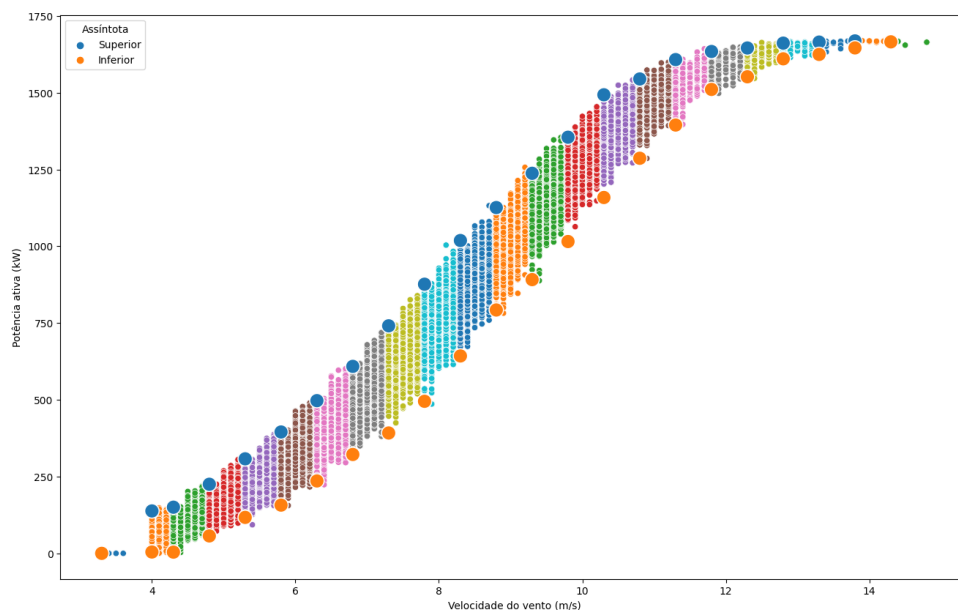


Figura 23 – Seccionamento da curva de potência normal, e identificação dos pontos de máximo e mínimo de potência e do valor de velocidade no início do intervalo.

Fonte: Elaborado pelo autor.

A Figura 24 mostra a ligação dos pontos obtidos usando retas, pela qual podemos identificar o envelope “S” da curva de potência. Porém, ao fazer isso, não temos uma aproximação tão suave. Para melhorar essa suavização e também obter uma forma simples de armazenar as funções que compõem as assíntotas, fazemos uma interpolação dos pontos usando o método de interpolação do MMQ para aproximar uma função polinomial de grau

9. Para isso, foi utilizada a função *polyfit* da biblioteca *NumPy* (v1.23.4). A Figura 25 mostra o resultado da aproximação da função polinomial de grau 9 para o conjunto de pontos das assíntotas.

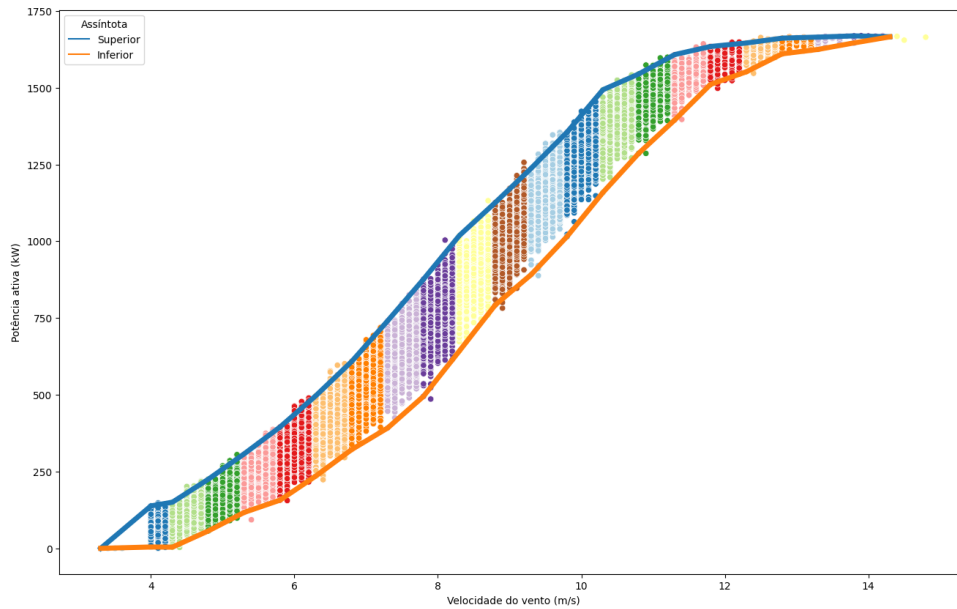


Figura 24 – Ligação dos pontos da curva utilizando retas.

Fonte: Elaborado pelo autor.

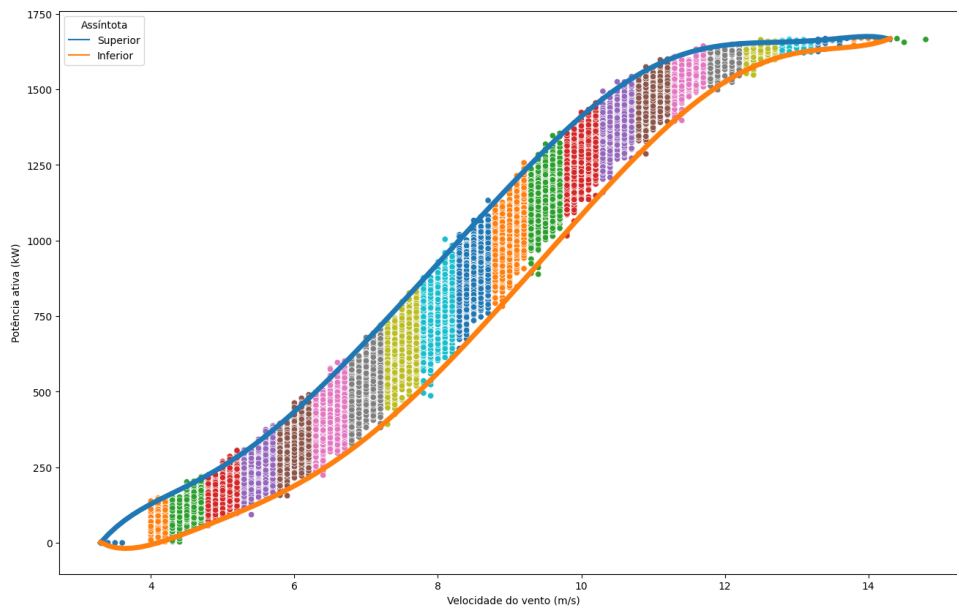


Figura 25 – Envelope “S” formado pelas assíntotas usando interpolação polinomial de grau 9.

Fonte: Elaborado pelo autor.

4 RESULTADOS E DISCUSSÕES

O presente capítulo será dividido em duas seções. Na primeira, serão mostrados os processamentos dos dados dos 4 aerogeradores que levam à obtenção da curva de potência sob condições normais, como também serão apresentados uma breve análise dos resultados obtidos para cada aerogerador, e os polinômios interpoladores obtidos para cada curva de cada aerogerador para todo o período de 4 anos de amostras. Como também será exposto o tempo de processamento de cada aerogerador, e será feita uma breve análise desses tempos. Na segunda seção serão mostradas as curvas obtidas por ano para cada aerogerador e serão analisadas as questões de desempenho com o passar dos anos para cada amostra.

4.1 Obtenção das curvas de potência sob condições normais para cada aerogerador e análise do tempo de processamento das amostras

Com base no que foi demonstrado na metodologia do trabalho, foi feito inicialmente o pré-processamento dos dados para 4 aerogeradores durante os períodos de 2018 a 2021. Com esses dados, foram utilizados inicialmente os eventos de cada aerogerador para classificar os pontos de velocidade e potência no período como valores normais. Depois de realizar essa primeira filtragem, os dados passaram pelos métodos de remoção de outliers do DBSCAN e posteriormente pelo interquartil, para que com isso fosse feita uma limpeza mais profunda dos dados, removendo assim a maioria dos outliers. Depois desses processos de remoção de outliers, os dados foram novamente processados a fim de remover pontos de potência negativa e pontos com valores de potência inválidos com base na velocidade de corte. Após todos esses processamentos, as curvas de potência com apenas pontos normais foram obtidas e, com isso, foi feita a obtenção das assíntotas que as descrevem, assim seguindo todo o procedimento discutido na metodologia. As Figuras 26, 27, 28 e 29 a seguir mostram os passos utilizados para obter a curva de potência em condições normais para cada aerogerador.

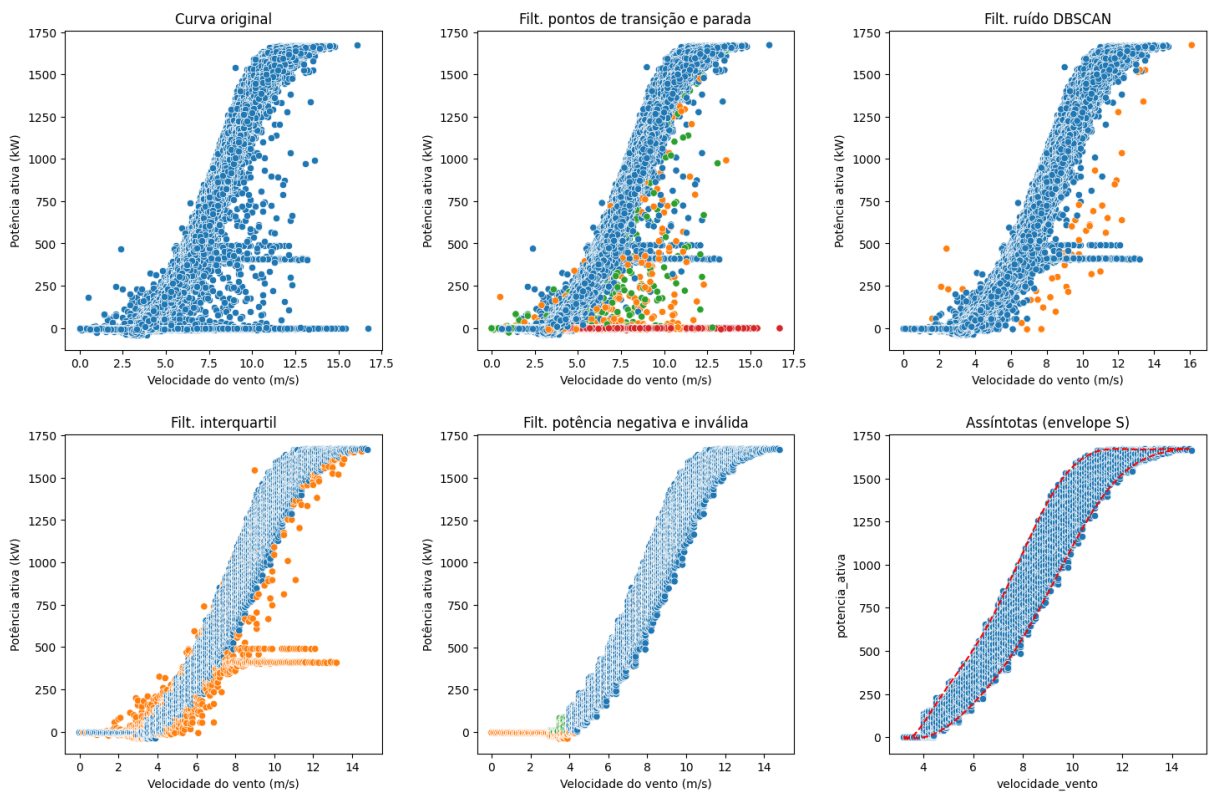


Figura 26 – Obtenção da curva de potência sob condições normais para o aerogerador 1.

Fonte: Elaborado pelo autor.

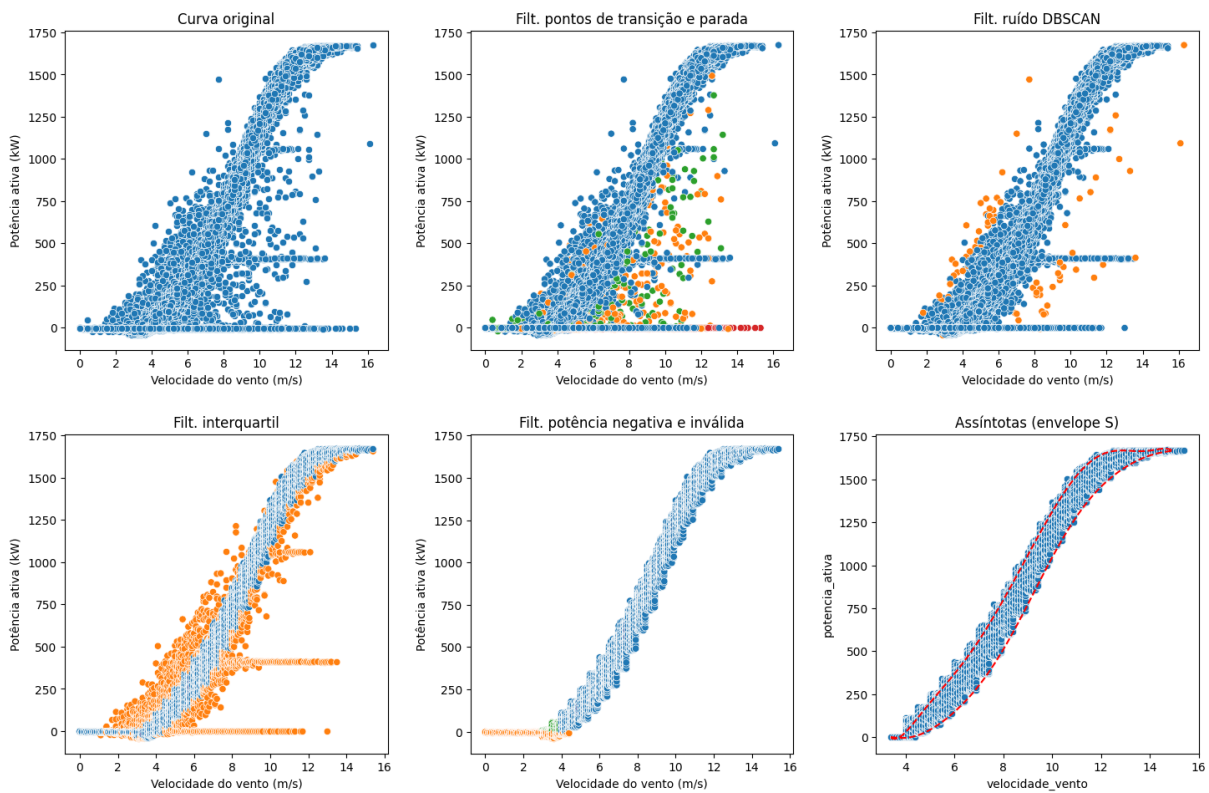


Figura 27 – Obtenção da curva de potência sob condições normais para o aerogerador 2.

Fonte: Elaborado pelo autor.

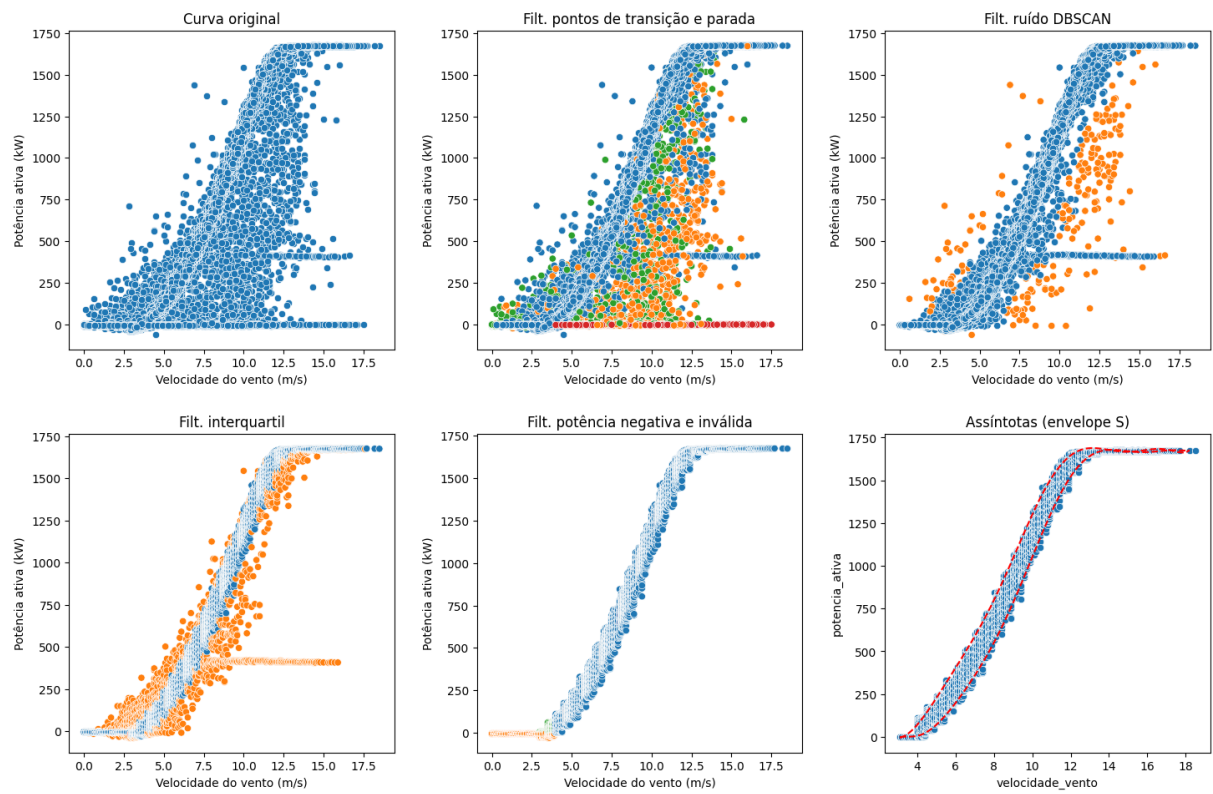


Figura 28 – Obtenção da curva de potência sob condições normais para o aerogerador 3.

Fonte: Elaborado pelo autor.

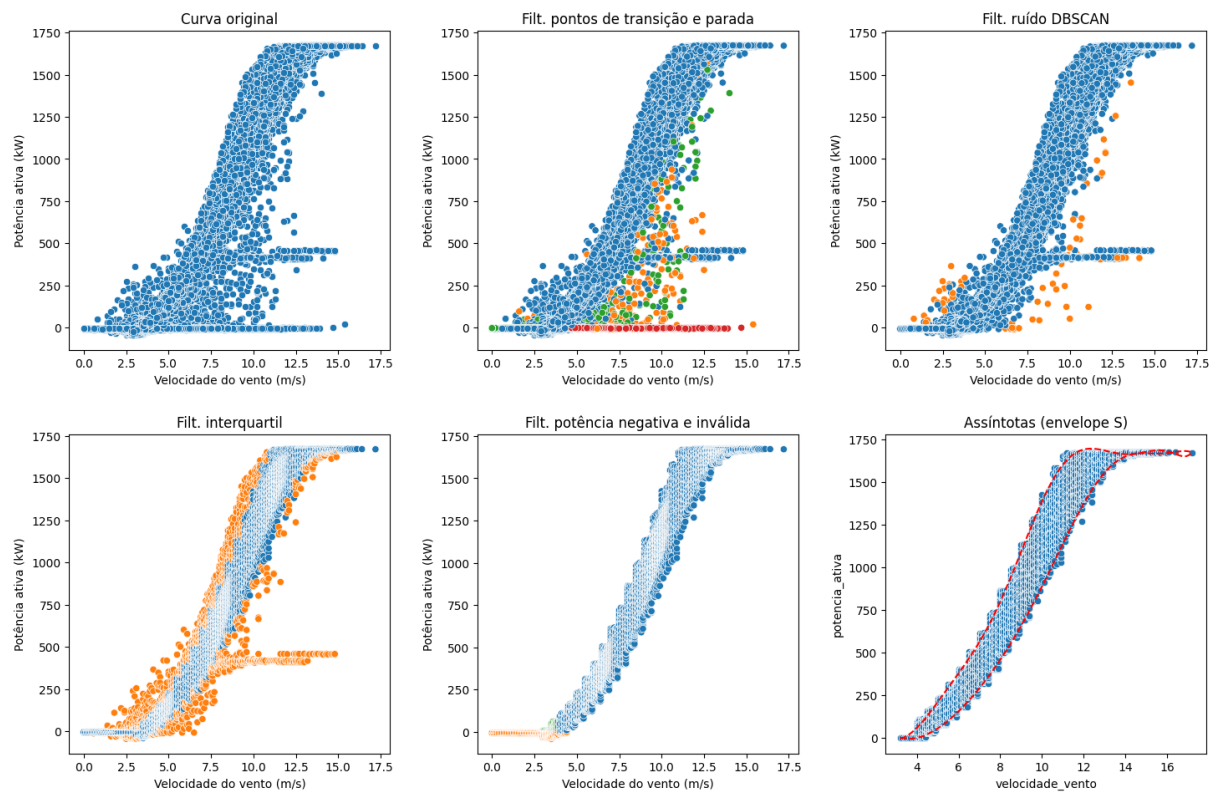


Figura 29 – Obtenção da curva de potência sob condições normais para o aerogerador 4.

Fonte: Elaborado pelo autor.

No início do processamento dos dados do aerogerador 1 (Figura 26), é possível identificar outliers concentrados abaixo da curva de potência, pontos de restrição de potência, e uma quantidade razoável de outliers dispersos ao redor da curva. O processamento para remoção desses outliers obteve êxito para esse aerogerador, pois praticamente todos os outliers foram removidos e a curva em forma de “S” foi obtida. Com isso, é possível obter as assíntotas superior e inferior que compõem a curva. Na curva obtida, é possível identificar um alargamento da área da curva, algo que se reflete no deslocamento da curva durante os anos, no qual esse deslocamento ocorre quando o aerogerador perde ou ganha desempenho, sendo um deslocamento a esquerda uma ganho de desempenho, visto que o aerogerador consegue produzir mais com um valor menor de velocidade do vento, já o deslocamento a direita o aerogerador perde desempenho ao produzir menos com uma velocidade de vento maior. Os dois polinômios interpoladores obtidos para o aerogerador 1 podem ser vistos nas Equações 4.1 e 4.2 abaixo, correspondendo às assíntotas superior e inferior, respectivamente.

$$\begin{aligned}
 P_{Superior}(x) = & -2,0 \cdot 10^{-5}x^9 + 0,00283x^8 - 0,14476x^7 + 3,93577x^6 \\
 & - 63,70966x^5 + 640,96121x^4 - 4023,15694x^3 + 15252,80173x^2 \\
 & - 31629,29055x + 27259,07385
 \end{aligned} \quad (4.1)$$

$$\begin{aligned}
 P_{Inferior}(x) = & 2,0 \cdot 10^{-5}x^9 - 0,00236x^8 + 0,09584x^7 - 2,14123x^6 \\
 & + 29,01647x^5 - 247,19329x^4 + 1322,0339x^3 - 4243,09654x^2 \\
 & + 7357,97419x - 5236,44914
 \end{aligned} \quad (4.2)$$

Para o aerogerador 2 (Figura 26), é possível verificar a existência de vários outliers dispostos na curva, sendo os mais significativos na parte inicial, além de uma quantidade alta de outliers dispersos. O processamento também obteve êxito em remover todos os outliers presentes, e com isso foi possível obter uma aproximação da curva sob condições normais. A curva resultante apresentou um comportamento uniforme em relação à dispersão dos pontos normais, o que leva a crer que esse aerogerador obteve um desempenho quase que constante com o passar do tempo. Os polinômios interpoladores que compõem as assíntotas da curva desse aerogerador podem ser vistos nas Equações 4.3 e 4.4 abaixo, correspondendo às assíntotas superior e inferior, respectivamente.

$$\begin{aligned}
P_{Superior}(x) = & -0,00021x^9 + 0,01757x^8 - 0,62932x^7 + 12,89566x^6 \\
& - 166,48196x^5 + 1401,95636x^4 - 7682,64323x^3 + 26356,57552x^2 \\
& - 51110,13389x + 42424,34952
\end{aligned} \quad (4.3)$$

$$\begin{aligned}
P_{Inferior}(x) = & 8,0 \cdot 10^{-5}x^9 - 0,00679x^8 + 0,24495x^7 - 4,97074x^6 \\
& + 62,28151x^5 - 498,56118x^4 + 2546,10342x^3 - 7963,97236x^2 \\
& + 13778,69859x - 10019,88541
\end{aligned} \quad (4.4)$$

Para o aerogerador 3 (Figura 28), é possível visualizar uma grande quantidade de outliers dispersos ao redor da curva. Como se nota, o processamento obteve êxito em remover de forma efetiva a maioria dos outliers. A curva obtida apresenta uma uniformidade até a região na qual a potência máxima é obtida, e depois disso ela permanece constante com o aumento da velocidade. Os polinômios interpoladores que correspondem aos pontos das assíntotas da curva desse aerogerador podem ser vistos nas Equações 4.5 e 4.6 abaixo.

$$\begin{aligned}
P_{Superior}(x) = & 3,0 \cdot 10^{-5}x^9 - 0,00267x^8 + 0,10185x^7 - 2,14182x^6 \\
& + 27,15834x^5 - 213,53673x^4 + 1029,8501x^3 - 2877,16204x^2 \\
& + 4188,39099x - 2433,28959
\end{aligned} \quad (4.5)$$

$$\begin{aligned}
P_{Inferior}(x) = & 3,0 \cdot 10^{-5}x^9 - 0,00267x^8 + 0,11008x^7 - 2,53522x^6 \\
& + 35,74038x^5 - 318,87255x^4 + 1795,34177x^3 - 6110,4332x^2 \\
& + 11368,15048x - 8820,81171
\end{aligned} \quad (4.6)$$

No processamento dos dados do aerogerador 4 (Figura 28), é possível notar uma grande quantidade de outliers no início da curva e um alargamento de sua área, além de outliers dispersos e restrições de potência que se concentram abaixo da curva em altas densidades. No processamento, esses outliers foram removidos de forma satisfatória, possibilitando assim obter a curva normal. Na curva normal obtida, observa-se um comportamento parecido com o do aerogerador 1, no qual há a presença de um alargamento da área da curva, mas de forma mais compacta. Os polinômios das assíntotas que forma essa curva também foram obtidas para esse aerogerador. Eles podem ser vistos nas Equações 4.7 e 4.8 abaixo.

$$\begin{aligned}
P_{Superior}(x) = & 8,0 \cdot 10^{-5}x^9 - 0,00704x^8 + 0,25888x^7 - 5,28318x^6 \\
& + 65,5612x^5 - 510,38432x^4 + 2480,29945x^3 - 7202,24272x^2 \\
& + 11337,9096x - 7434,69413
\end{aligned} \quad (4.7)$$

$$\begin{aligned}
P_{Inferior}(x) = & -2,0 \cdot 10^{-5}x^9 + 0,00153x^8 - 0,05325x^7 + 1,03908x^6 \\
& - 12,5645x^5 + 98,30156x^4 - 502,75232x^3 + 1660,6035x^2 \\
& - 3254,07058x + 2843,69793
\end{aligned} \quad (4.8)$$

O algoritmo foi executado de forma sequencial, em um notebook com processador core i5, 8ª geração, 16gb de memória RAM, 240gb de SSD e sistema operacional Ubuntu 22.04.1, o tempo de execução do processamento e obtenção das assíntotas das curvas de potência em condições normais para cada aerogerador para todo o período de dados é mostrado na Tabela 3 abaixo.

Aerogerador	Quantidade de dados	Quantidade de eventos	Tempo (s)
Aerogerador 1	202039	2134	5,62882
Aerogerador 2	200940	2101	5,00753
Aerogerador 3	202615	3930	9,02868
Aerogerador 4	202537	1441	3,94999

Tabela 3 – Tempo de execução do algoritmo final para remoção de outliers e parametrização da curva sob condições normais.

Fonte: Elaborado pelo autor.

É possível identificar que a quantidade de dados de potência e velocidade do vento para ambos os aerogeradores foram relativamente próximas. O que mais se diferenciou foi a quantidade de eventos, e foi essa quantidade de eventos que interferiu no tempo de execução para cada aerogerador, visto que o aerogerador 3 que possuía a maior quantidade de eventos levou quase o dobro do tempo que os demais. Isso se refere ao método de classificar os pontos pelos eventos. De modo geral, o tempo de execução foi relativamente rápido dado a quantidade de dados, e a complexidade assintótica do algoritmo é estimada em $O(n^2)$.

4.2 Análise de desempenho dos aerogeradores por meio das curvas de potência durante os anos de 2018 à 2021

Com a obtenção da curva de potência sob condições normais para cada aerogerador durante todos os anos, é possível fazer uma comparação do seu desempenho levando em

conta o deslocamento da curva com o passar dos anos. Um deslocamento a esquerda corresponde a um melhor desempenho do aerogerador, visto que ele conseguiu obter um ganho de potência com velocidades de vento menores, já um deslocamento para a direita ele perde desempenho ao obter um valor de potência menor para velocidades de vento maiores do que se comparado com um deslocamento a esquerda. As seções seguintes mostram as curvas de potência para cada aerogerador e fazem essa análise por meio da sobreposição das curvas para o período de 4 anos.

4.2.1 Aerogerador 1

As curvas obtidas para o aerogerador 1 para cada ano da base de dados pode ser visto na Figura 30 abaixo. É possível ver que as curvas mesmo em anos distintos se assemelham entre si, sendo possível destacar também uma pequena variação com o passar dos anos na área de cada curva.

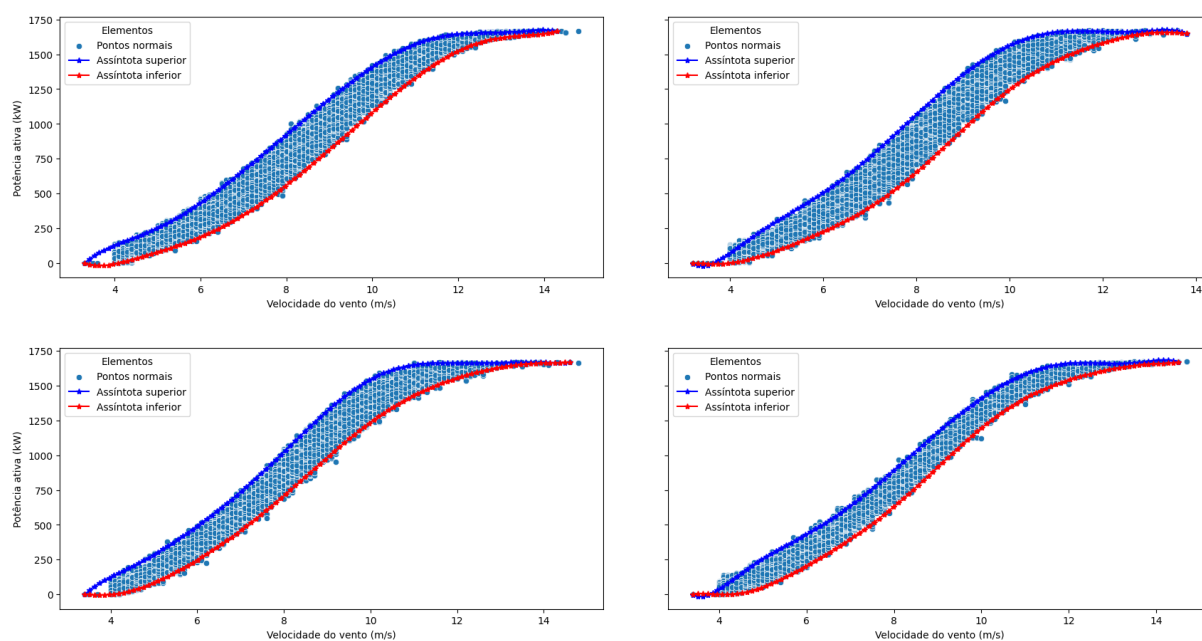


Figura 30 – Curvas por ano para o aerogerador 1.

Fonte: Elaborado pelo autor.

Fazendo a sobreposição dessas curvas (Figura 31), é possível identificar que durante o ano de 2018 o aerogerador não teve um desempenho tão bom como nos anos posteriores. Pois esse ano acaba ficando mais deslocado à direita do que todas as outras curvas, ocasionando uma perda de desempenho. Já a curva obtida para o ano de 2019 se apresenta de forma mais deslocada à esquerda, o que significa um desempenho melhor do que as outras curvas dos anos posteriores. Mesmo assim, as curvas dos anos posteriores apresentam um comportamento parecido com a de 2019.

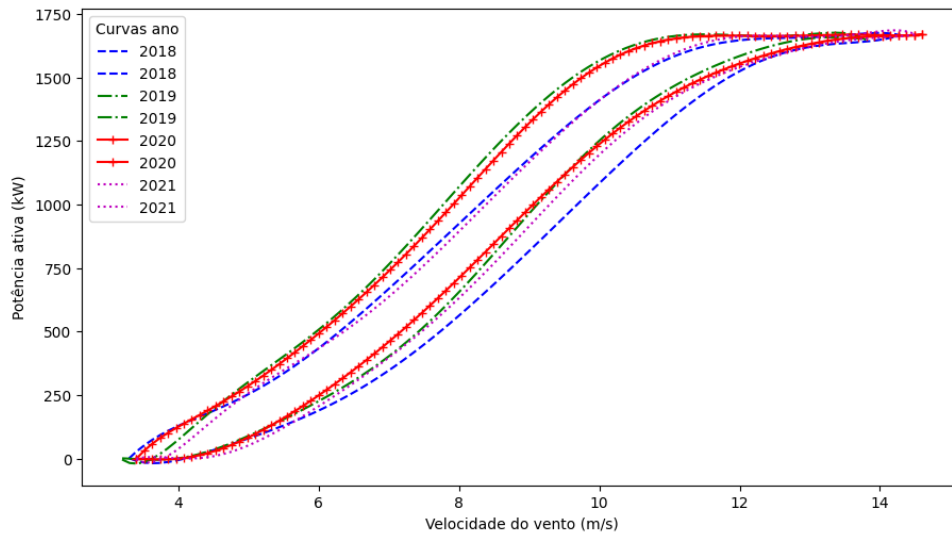


Figura 31 – Curvas por ano sobrepostas para o aerogerador 1.

Fonte: Elaborado pelo autor.

4.2.2 Aerogerador 2

Nas curvas obtidas para o aerogerador 2 (Figura 32), é possível identificar uma maior consistência nos últimos anos, e de forma geral elas se comportam de forma semelhante com o passar do tempo.

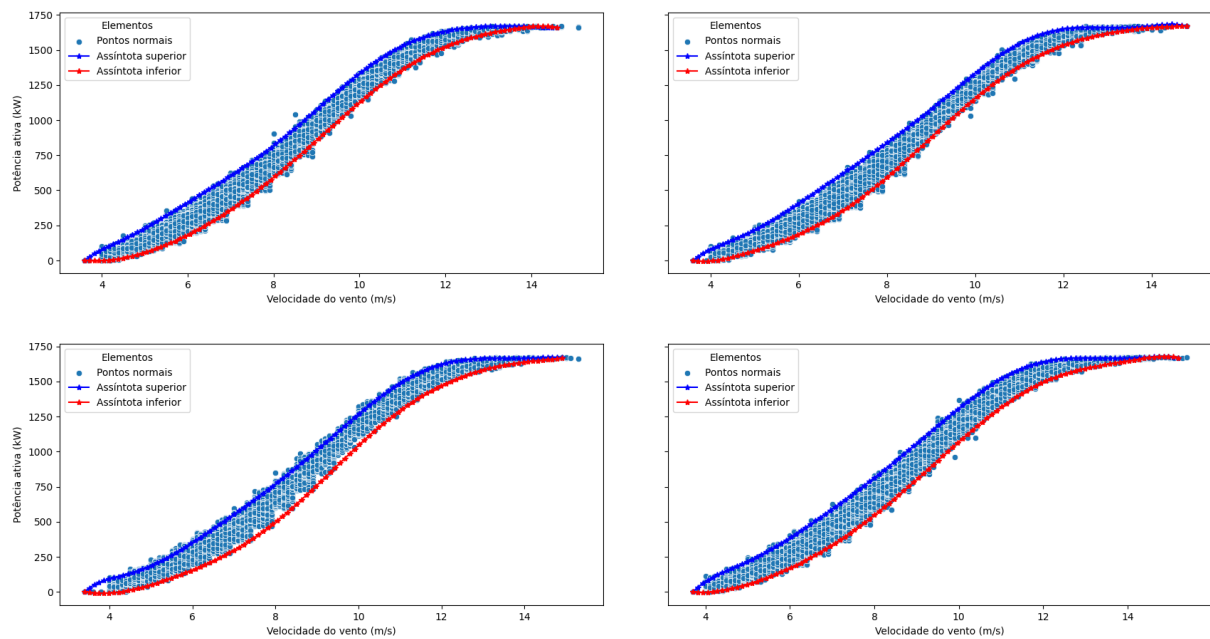


Figura 32 – Curvas por ano para o aerogerador 2.

Fonte: Elaborado pelo autor.

Ao fazer a sobreposição dessas curvas (Figura 33), é possível identificar que esse aerogerador tem um desempenho semelhante com o passar dos anos, notando-se apenas uma pequena variação do deslocamento das curvas. De certa forma observa-se que o

aerogerador perdeu um pouco de desempenho com o passar do tempo, visto que de 2018 a 2021 as curvas vão se deslocando para a direita.

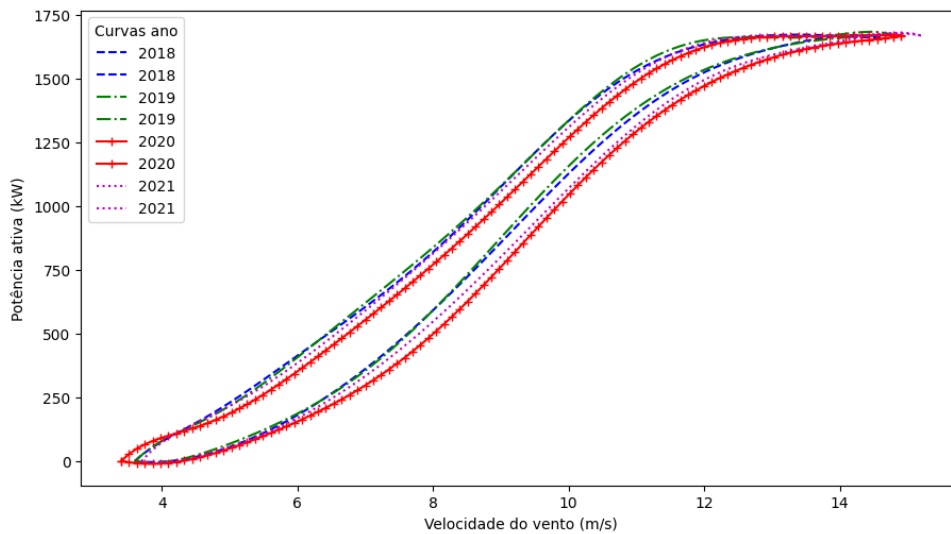


Figura 33 – Curvas por ano sobrepostas para o aerogerador 2.

Fonte: Elaborado pelo autor.

4.2.3 Aerogerador 3

Ao verificar as curvas obtidas para o aerogerador 3 (Figura 34), nota-se que este aerogerador se distingue dos demais ao apresentar uma simetria muito grande da curva de potência, mesmo com o passar dos anos, além de possuir uma alta constância no período final de todas as curvas.

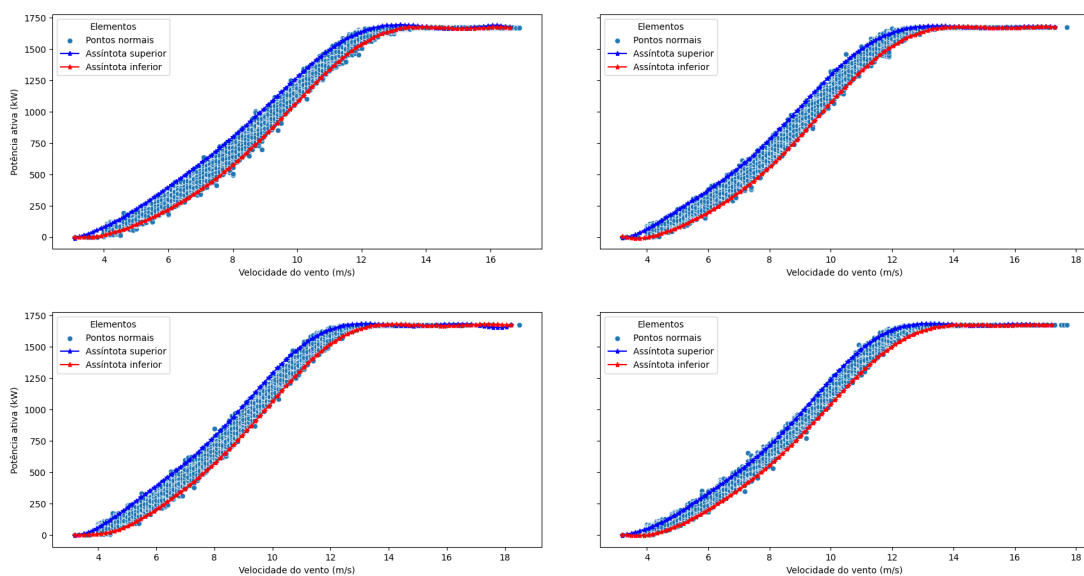


Figura 34 – Curvas por ano para o aerogerador 3.

Fonte: Elaborado pelo autor.

Ao verificar a sobreposição dessas curvas (Figura 35), é possível avaliar que o aerogerador tem uma variação mínima no deslocamento de suas curvas com o passar do tempo, e que com isso seu desempenho não mudou muito com o passar dos anos. Nos últimos anos nota-se que houve apenas um pequeno deslocamento mais à esquerda, o que pode indicar que o aerogerador melhorou um pouco seu desempenho ao longo dos anos.

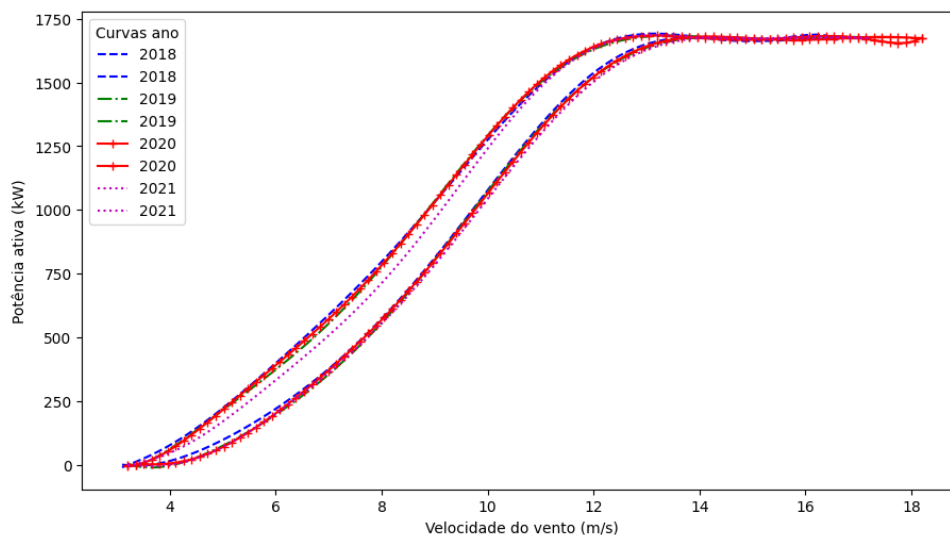


Figura 35 – Curvas por ano sobrepostas para o aerogerador 3.

Fonte: Elaborado pelo autor.

4.2.4 Aerogerador 4

Ao verificar as curvas obtidas para cada ano do aerogerador 4 (Figura 36), é possível verificar uma certa similaridade do ano de 2018 com o ano de 2020, e do ano de 2019 com 2021, sendo que no primeiro caso a curva de potência é mais compacta. Já para o segundo caso, ela apresenta um alargamento maior de sua área, além de que nos dois casos acontece um longo período constante para todos os anos no final das curvas.

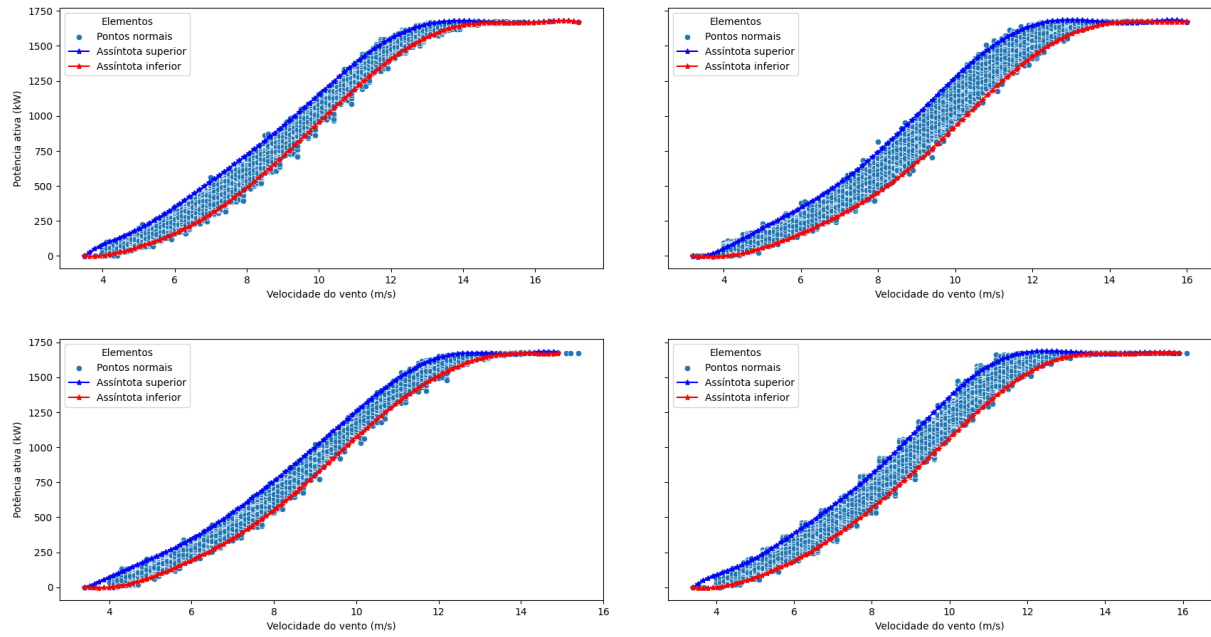


Figura 36 – Curvas por ano para o aerogerador 4.

Fonte: Elaborado pelo autor.

Ao fazer a sobreposição dessas curvas (Figura 37), é possível identificar que esse aerogerador teve um melhor desempenho nos anos de 2019 e 2021, em que existe um deslocamento mais à esquerda, e nos anos de 2018 e 2020 acontece um deslocamento mais à direita, no qual o aerogerador perdeu um pouco de desempenho.

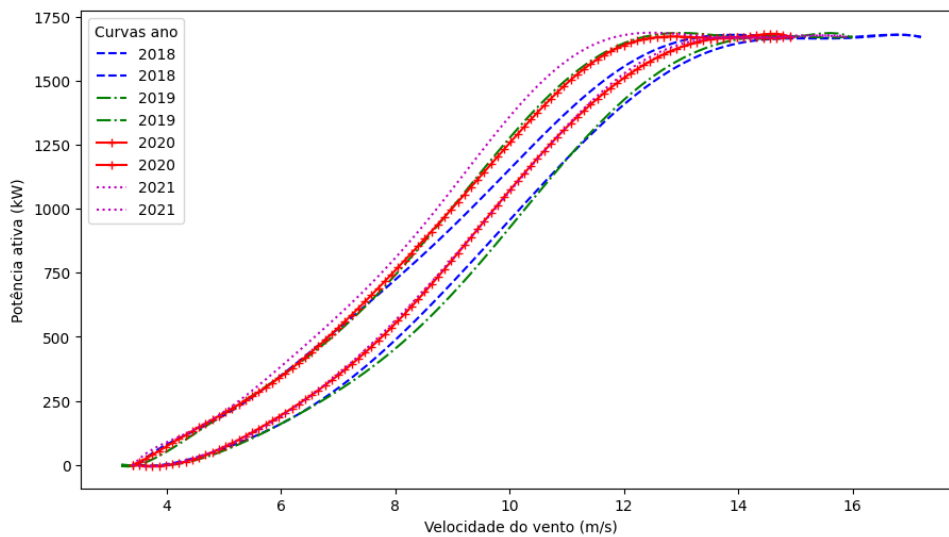


Figura 37 – Curvas por ano sobrepostas para o aerogerador 4.

Fonte: Elaborado pelo autor.

5 CONCLUSÃO

Neste trabalho, obtiveram-se dados reais de aerogeradores para um período de 4 anos, correspondendo aos anos de 2018 a 2021. Dessas informações, foram obtidos dados das variáveis de velocidade do vento e potência ativa, como também eventos relacionados a cada aerogerador. Com base nesses, dados foram feitas análises preliminares de consistência e pré-processamento, a fim de remover períodos faltantes, e pré-visualizar as curvas de potência para cada aerogerador. Analisar a curva de potência desses aerogeradores foi um dos principais enfoques deste trabalho, e com isso foram realizados os processamentos dos dados presentes nessas curvas, a fim de remover dados anormais e obter a curva sob condições normais. Com isso, foi possível obter a curva normal para parametrização e analisar o desempenho dos aerogeradores sob o enfoque dessa parametrização.

A análise realizada neste trabalho possibilitou a aplicação de métodos de remoção de outliers difundidos no meio científico, a fim de obter pontos de operação normal dos aerogeradores. Os métodos assim aplicados foram de rotulação de dados de produção dos aerogeradores pelos eventos envolvidos no período, como também a aplicação dos métodos do DBSCAN e interquartil. Com os resultados obtidos, foi atingido o principal objetivo do trabalho, obtenção da curva de potência sob condições normais, parametrizada por meio de interpolação polinomial em termo das duas funções que compõem as assíntotas da curva.

Com a parametrização da curva normal, foi possível fazer o armazenamento dos coeficientes dos polinômios obtidos de forma simples, as quais poderão ser aplicado por exemplo no armazenamento em um banco de dados, e com isso ser aplicados a sistemas que fazem o monitoramento de aerogeradores. Essa parametrização pode ser de bom uso para analisar o desempenho dos aerogeradores com o passar do tempo, se tornando uma ótima ferramenta de avaliação.

Por fim, este trabalho conseguiu de forma eficaz obter e parametrizar as curvas de potência sob condições normais de aerogeradores, além de obter as funções que a compõem, sendo assim uma porta de entrada para diversos outros trabalhos a serem realizados, visto que com a obtenção dessas curvas é possível futuramente utilizar como um classificador, utilização em alarmes de monitoramento, como também poderão ser utilizado outros métodos de ajuste para as funções, como aplicação da função logística.

REFERÊNCIAS

- ABEEÓLICA. *O Setor*. 2022. Disponível em: <https://abeeolica.org.br/energia-eolica/o-setor/>. Acesso em: 22 de novembro 2022.
- BUTLER, S.; RINGWOOD, J.; O'CONNOR, F. Exploiting scada system data for wind turbine performance monitoring. In: . [S.l.: s.n.], 2013. p. 389–394.
- CHAUHAN, N. S. *DBSCAN Clustering Algorithm in Machine Learning*. 2022. Disponível em: <https://www.kdnuggets.com/2020/04/dbscan-clustering-algorithm-machine-learning.html>. Acesso em: 22 de novembro 2022.
- DEVORE, J.; SILVA, J. da. *Probabilidade e estatística: para engenharia e ciências*. [S.l.]: Pioneira Thomson Learning, 2006. ISBN 9788522109241.
- DIAS, R. F. *Identificando e Removendo Outliers*. 2022. Disponível em: <https://cursos.alura.com.br/course/introducao-python-pandas/task/41002>. Acesso em: 22 de agosto 2022.
- GEEKSFORGEEEKS. *Merge Overlapping Intervals*. 2022. Disponível em: <https://www.geeksforgeeks.org/merging-intervals/>. Acesso em: 21 de outubro 2022.
- LEAHY, K. et al. Automatically identifying and predicting unplanned wind turbine stoppages using scada and alarms system data: Case study and results. *Journal of Physics: Conference Series*, IOP Publishing, v. 926, n. 1, p. 012011, nov 2017.
- LIRA, A.; PINTO, M. de O.; HEINZELMANN, B. *Energia eólica: Princípios e operação*. [S.l.]: Saraiva Educação S.A., 2019. ISBN 9788536532219.
- NEOENERGIA. *VOCÊ SABE COMO FUNCIONA UM AEROGERADOR?* 2022. Disponível em: <https://www.neoenergia.com/pt-br/te-interessa/meio-ambiente/Paginas/como-funciona-um-aerogerador.aspx>. Acesso em: 11 de novembro 2022.
- NUMÉRICA, C. *MMQ - Introdução*. 2022. Disponível em: <https://cn.ect.ufrn.br/index.php?r=conteudo%2Fmmq>. Acesso em: 25 de novembro 2022.
- RITCHIE, H.; ROSER, M.; ROSADO, P. Energy. *Our World in Data*, 2022. Disponível em: <https://ourworldindata.org/energy>. Acesso em: 21 de outubro 2022.
- SHEN, X.; FU, X.; ZHOU, C. A combined algorithm for cleaning abnormal data of wind turbine power curve based on change point grouping algorithm and quartile algorithm. *IEEE Transactions on Sustainable Energy*, v. 10, n. 1, p. 46–54, 2019.
- 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*, v. 2016, p. 1–18, 01 2016.
- YUAN, T.; SUN, Z.; MA, S. Gearbox fault prediction of wind turbines based on a stacking model and change-point detection. *Energies*, v. 12, p. 4224, 11 2019.