

UNIVERSIDADE FEDERAL DO MARANHÃO
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

Hellen Dianne Pereira de Souza

*DENSIFICAÇÃO DE DADOS BASEADA EM MEAN
SHIFT APLICADA A PREVISÃO DE GERAÇÃO DE
ENERGIA EÓLICA*

São Luís

2022

Hellen Dianne Pereira de Souza

*DENSIFICAÇÃO DE DADOS BASEADA EM MEAN
SHIFT APLICADA A PREVISÃO DE GERAÇÃO DE
ENERGIA EÓLICA*

Dissertação apresentada ao Programa de Pós-graduação em Engenharia Elétrica da UFMA, como requisito para a obtenção do grau de MESTRE em Engenharia Elétrica.

Shigeaki Leite de Lima, Dr.

(Orientador)

Denisson Queiroz Oliveira, Dr.

(Coorientador)

São Luís

2022

Souza, Hellen Dianne Pereira de Souza

DENSIFICAÇÃO DE DADOS BASEADA EM MEAN SHIFT
APLICADA A PREVISÃO DE GERAÇÃO DE ENERGIA EÓLICA

/ Hellen Dianne Pereira de Souza Souza - 2022

69f.

Orientador: Shigeaki Leite de Lima

Co-orientador: Denisson Queiroz Oliveira

Impresso por computador (fotocópia)

Dissertação (Graduação em Engenharia Elétrica) - Universidade
Federal do Maranhão, curso de graduação em Engenharia Elétrica,
São Luís, 2022.

1.Gestão de Ativos em Engenharia. 2. Custos 3.Confiabilidade. I.
Lima, Shigeaki Leite de ,orientador. II. Título.

CDU 621.314.1

Hellen Dianne Pereira de Souza

*DENSIFICAÇÃO DE DADOS BASEADA EM MEAN
SHIFT APLICADA A PREVISÃO DE GERAÇÃO DE
ENERGIA EÓLICA*

Dissertação apresentada ao Programa de Pós-graduação em Engenharia Elétrica da UFMA, como requisito para a obtenção do grau de MESTRE em Engenharia Elétrica.

Aprovado em 25 de fevereiro de 2022

BANCA EXAMINADORA

Shigeaki Leite de Lima, Dr.

(Orientador)

Denisson Queiroz Oliveira, Dr.

(Coorientador)

Oswaldo Ronald Saavedra Mendez, Dr.

(Examinador Interno)

André Nunes de Souza, Dr.

(Examinadora Externo)

UNESP-Bauru

Aos meus pais, irmão e namorada.

Aos amigos, pelo apoio e companheirismo.

Resumo

O Brasil nos últimos 10 anos tem apresentado um elevado potencial na produção de energia eólica, exibindo a cada ano um crescimento significativo com instalações de novos parques e aumento da potência instalada. Em vista disso, é importante a realização de análises técnicas que permitam viabilizar a geração de energia eólica, com as medições dos fatores que influenciam na velocidade do vento e que podem comprometer a geração de energia. Nesse sentido, muitos modelos de previsão de velocidade de vento foram desenvolvidos, modelos como os *Numeric Weather Prediction (NWP)*, que fazem o monitoramento das condições atmosféricas, produzindo informações que permitem realizar previsões de geração. Outros como os modelos estatísticos e os híbridos fazem as análises com base em dados de medição de uma série histórica. A previsão de geração de energia eólica é usada por muitas empresas de energia com a finalidade de estimar com segurança o recurso eólico e desta forma garantir para a empresa o aumento da receita e ao consumidor o fornecimento de uma energia de fonte renovável durante um determinado período de tempo. Muitas técnicas foram desenvolvidas com este objetivo e permitem estimar o recurso eólico e a energia gerada desde algumas horas, dias ou mesmo semanas. Neste trabalho é apresentada uma técnica baseada em teoria da informação para a densificação de uma base de dados de medições de vento, gerando novos dados com a mesma função densidade de probabilidade do conjunto original e utilizando estes dados virtuais como cenários de previsão.

Palavras-chave: *Mean Shift*; Teoria da informação; Previsão; Geração de energia; Energia eólica.

Abstract

Brazil in the last 10 years has a high potential for wind energy production, showing significant growth each year with the installation of new parks and an increase in installed power. In view of the power generation capacity, which makes it possible to generate wind important, how to carry out power which can generate the wind capacity important, as the factors that generate power may generate the wind important. In this sense, the projected wind speed forecast models, numerical weather forecast (NWP) models, many of the weather forecast models occur during generation. Others, such as statistical models and hybrids, perform analyzes based on data from a historical series. Wind power generation is used by many energy companies for the purpose of safety estimation and in this way guaranteeing the resource for the revenue company and the consumer to increase an energy source from source during a certain period of time. Many techniques were calculated with resources and working hours and with the same efficiency, working days and hours. This is an information theory based technique for a wind measurement database, generating new data with the same probability density function as the original set and using these virtual data as forecast scenarios.

Keywords: Mean Shift; Information theoretic; Forecast; Energy Generation; Wind energy.

Agradecimentos

Agradeço antes de tudo a Deus, que me deu força, saúde e inteligência, para superar as dificuldades.

Agradeço em seguida aos meus pais, pela educação, amor, carinho e apoio incondicional, sem os quais seria impossível trilhar o caminho que sigo.

Aos familiares, se fazem presentes nos momentos de dificuldade e também nos de alegria.

A todos os meus amigos. Aos que estão longe e aos que estão perto. Sem eles, não daria para sorrir depois das dificuldades.

Ao meu orientador e co-orientador, Shigeaki Leite de Lima e Denisson Queiroz Oliveira, que dedicaram tempo e acreditaram que seria possível este trabalho, pela orientação, paciência e pelo apoio ao longo do tempo em que trabalhamos juntos.

Ao professor Osvaldo Ronald Saavedra Mendez pelos ensinamentos passados ao longo desse período de mestrado.

Aos amigos do Instituto de Engenharia Elétrica (IEE/UFMA), por compartilhar momentos no almoço, no café, nas viagens, fazendo multirão para resolver os problemas e pela companhia de estudo ao longo do período do mestrado.

Em especial agradeço a minha namorada, Elinete, pela compreensão, incentivo, amor e pela sua paciência em ouvir minhas reclamações e em revisar este trabalho.

À Coordenação do curso de pós-graduação em Engenharia Elétrica, ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) - Brasil, Gera maranhão e a Equatorial Maranhão pelo Apoio Financeiro e cessão dos dados para o estudo.

*“Na natureza nada se cria, nada se perde;
tudo se transforma”.*

Antoine Lavoisier

SUMÁRIO

LISTA DE FIGURAS	8
1 INTRODUÇÃO	12
1.1 Objetivos	14
1.2 Estrutura do Trabalho	14
2 REVISÃO BIBLIOGRÁFICA	16
2.1 Estado da arte	16
2.2 Bibliometria	19
3 GERAÇÃO DE ENERGIA EÓLICA	24
3.1 Energia dos ventos	24
3.2 Modelos de Previsão	28
3.2.1 Modelos Físicos	30
3.2.2 Modelos Estatísticos	31
3.2.3 Modelo baseado em Inteligência Artificial	34
3.2.4 Modelos Híbridos	37
4 TEORIA DA INFORMAÇÃO <i>MEAN SHIFT</i>	39
4.1 Entropia de Rényi	39
4.2 Informação Mútua	40
4.3 Janela de Parzen	42
4.4 Algoritmo Mean Shift	43
5 <i>MEAN SHIFT</i> APLICADO A PREVISÃO EÓLICA	46

5.1	Modelo de Previsão	48
5.2	Análise dos Resultados	50
6	CONCLUSÃO	57
6.1	Trabalhos Futuros	58
6.2	Publicações	58
	REFERÊNCIAS	59

LISTA DE FIGURAS

1.1	<i>Potencial eólico brasileiro</i>	13
2.1	<i>Análise por autor</i>	19
2.2	<i>Análise por quantidade de citações</i>	20
2.3	<i>Análise por país</i>	21
2.4	<i>Análise por ano de publicação</i>	21
2.5	<i>Análise das palavras chaves</i>	22
2.6	<i>Análise das palavras chaves com mais ocorrência</i>	23
3.1	<i>Curva de distribuição de Weibull para κ</i>	26
3.2	<i>Perfil de vento</i>	27
3.3	<i>Curva de potência de uma aerogerador</i>	27
3.4	<i>Previsão de ventos via NWP</i>	30
3.5	<i>Esquema de um neurônio artificial</i>	35
3.6	<i>Resultados obtidos após um período de treinamento</i>	37
4.1	<i>Influência do σ na estimativa da FDP pela Janela de Parzen</i>	43
5.1	<i>Sodar(a) e Lidar(b)</i>	47
5.2	<i>Metodologia</i>	48
5.3	<i>Dimensão da Janela de Parzen</i>	49
5.4	<i>Medição de outubro a novembro</i>	50
5.5	<i>Direção x Velocidade</i>	51
5.6	<i>Mean Shift após 30 interações</i>	52
5.7	<i>Mean Shift após 30 interações em 3D</i>	52

5.8	<i>Mean Shift para 30 dias a frente</i>	53
5.9	<i>Mean Shift para 7 dias a frente</i>	54
5.10	<i>Mean Shift para 1 dias a frente</i>	55
5.11	<i>Energia prevista e esperada para 1 , 7, 30 dias a frente</i>	56

LISTA DE ABREVIATURAS

PROINFA	<i>Programa de Incentivo às Fontes Alternativas de Energia Elétrica</i>
NWP	<i>Numerical Weather Prediction</i>
WPF	<i>Wind Power Forecasting</i>
NN	<i>Neural Network</i>
BP	<i>Back-Propagation Algorithm</i>
SVM	<i>Support Vector Machine</i>
WPD	<i>Wavelet Packet Decomposition</i>
WSF	<i>Wind Speed Forecasting</i>
CNN	<i>Convolutional Neural Network</i>
CNNLSTM	<i>Convolutional Long Short Term Memory Network</i>
ANN	<i>Artificial neural networks</i>
GBM	<i>Gradient Boosting Machine</i>
NMAE	<i>Mean Absolute Error</i>
NRMSE	<i>Root Mean Square Error</i>
WNN	<i>Wavelet Neural Network</i>
MCC	<i>Maximum Correntropy Criterion</i>
KF	<i>Kalman Filter</i>
ARIMA	<i>Autorregressive Integrate Moving Average</i>
GPR	<i>Gaussian Process Regression</i>
ELM	<i>Extreme Learning Machine</i>
LSSVM	<i>Least Square Support Vector Machine</i>
WRF	<i>Weather Research and Forecasting</i>
GSA	<i>Gravitational Search Algorithm</i>
FDP	<i>Função de Distribuição de Probabilidade</i>
ONS	<i>Operador nacional do Sistema Elétrico</i>

LISTA DE ABREVIATURAS

RAMS	<i>Atmosferic Modeling System</i>
BRAMS	<i>Brazilian Developments on the Regional Atmospheric Modeling System</i>
ARMA	<i>Autorregressive Moving Average</i>
IA	<i>Inteligência Artificial</i>
AR	<i>Auto-regressivo</i>
MA	<i>Media Móvel</i>
RNA	<i>Rede neural Artificial</i>
MLP	<i>Multilayer Perceptron</i>
ANEEL	<i>Agencia Nacional de Energia Elétrica</i>
SVR	<i>Support Vector Regression</i>
PSR	<i>Phase Space Reconstruction</i>
GA	<i>Genetic Algorithm</i>
CS	<i>Cuckoo Search</i>
FS	<i>Fuzzy System</i>
E	<i>Ensemble</i>
ICSA	<i>Improved Clonal Selection Algorithm</i>

1 INTRODUÇÃO

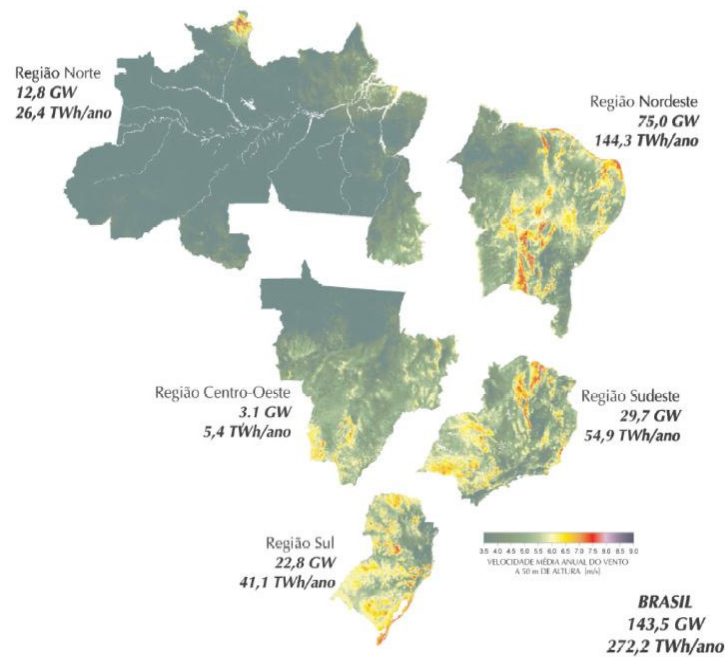
A crescente demanda de energia elétrica tem provocado o mercado elétrico a aumentar a oferta de energia, contudo para sustentar esse mercado consumidor outros meios de geração de energia não convencionais estão ganhando espaço, com isso a geração de energia de origem renovável vem sendo mais explorada [1].

Em conjunto com a alta demanda os preços dos combustíveis fósseis tem aumentado e conseqüentemente encarecido esta energia, outro fator que afeta diretamente a produção de energia baseada em combustíveis fósseis, são os resíduos produzidos por essa fonte, que apesar de toda a tecnologia já empregada para reduzir a emissão de gases, ainda provoca questionamentos e ações ambientais [5].

O Brasil se destaca em comparação com outros países em termos de produção de energia limpa, tendo sua matriz energética cerca de 87 % de fontes renováveis composta principalmente pela geração de energia das hidrelétricas com 82,8 %. Contudo, a instalação de usinas hidrelétricas demandam de grandes áreas, principalmente áreas alagadas e conseqüentemente provocam a alteração do ambiente devido a essa necessidade. Estas características restringem os locais em que essas usinas podem ser instaladas, oportunizando desta forma, a geração de energia por meio de outras fontes como eólica e solar [6].

Identificando a necessidade de aumentar o fornecimento de energia e aproveitar os recursos naturais, o governo brasileiro buscou incentivar a geração de energia elétricas oriundas de fontes renováveis e entendendo a importância econômica e energética instituiu a Lei nº10.438, de 26 de abril de 2002, criando o Programa de Incentivos às Fontes Alternativas de Energia Elétrica (PROINFA) [1].

Com a criação do PROINFA, foram vendidos 1430 MW de capacidade eólica, obtidos e redistribuídos ao setor consumidor pela Eletrobras. Os primeiros estudos com medições de vento no Brasil ocorreram nos anos 70, desde então outros estudos foram realizados ao longo das décadas de 1980, 1990 e 2000. Em 2001, foi elaborado o Atlas do Potencial Eólico Brasileiro, utilizando os dados produzidos das medições anteriores, que pode ser ilustrado pela Figura 1.1 [1].

Figura 1.1: *Potencial eólico brasileiro*

Fonte: Adaptado de [7]

Assim como estudos mais atualizados, o Atlas de 2001 constatou que em decorrência da diferença de propriedades atmosféricas, como relevo, vegetação e obstáculos, o potencial de geração de energia para cada região seria diferente, implicando na necessidade de avaliação individual de cada região a ser explorada [5, 6].

Como a fonte de energia eólica é intermitente e estocástica, o conhecimento dos fatores influenciadores da formação dos ventos é importante para estimar a potencialidade de um região. Em vista do alto potencial das região nordeste, que possui influência de ventos alísios em toda extensão do litoral e ventos constantes, tem ocorrido grandes investimentos no setor eólico nessa região [8].

Estudos que estimem a capacidade de geração considerando a velocidade dos ventos, a rugosidade, demais fatores climáticos e de relevo, obstáculos entre outros, devem ser considerados para a realização de projetos eólicos [5]. Os modelos como os *Numeric Weather Prediction* (NWP), realizam o monitoramento das condições atmosféricas, gerando informações que permitem realizar previsões para o setor, outros como os modelos estatísticos e os híbridos fazem as análises baseados em dados de medição de uma série histórica [8].

Neste trabalho é apresentado uma proposta para previsão de ventos e consequentemente geração de energia eólica, por meio de uma abordagem híbrida, em que foram utilizados um conjunto de séries temporais, adquiridas em uma base de dados, com medições entre 50 m a 280 metros (a cada 10 metros, por meio de perfiladores) e a aplicação da teoria da informação *Mean Shift*, para densificação dos dados.

1.1 Objetivos

O objetivo geral do trabalho é desenvolver uma metodologia e o algoritmo para a previsão de geração de energia eólica e explorar estratégias que auxiliem na aplicação da teoria da informação *Mean Shift* para densificação dos dados obtidos em base de dados existentes.

Já os objetivos específicos são:

- Realizar a revisão bibliográfica e estudos bibliométricos, afim de identificar os métodos para previsão de geração de energia eólica existentes;
- Obter base de dados de vento para estudos de geração de energia eólica no Brasil;
- A partir dos dados obtidos, identificar os grupos que precisam ser densificados por meio do algoritmo *Mean Shift*;
- Desenvolver os algoritmos para a densificação e para gerar o modelo de previsão de energia eólica;
- Testar e Avaliar o modelo de previsão eólica desenvolvido.

1.2 Estrutura do Trabalho

Neste trabalho é realizado a aplicação de um algoritmo baseado em teoria da informação para previsão da velocidade do vento de um parque eólico. Assim, este trabalho esta organizado em quatro capítulos incluindo a introdução. No capítulo dois é realizado a revisão bibliográfica e estudos de bibliometria. No capítulo três, a conceituação da geração de energia eólica das diferentes abordagem de previsão de vento e a teoria da informação *Mean Shift*. No capítulo quatro é descrito a área de estudo e a base de dados

utilizados, bem como a análise e os resultados obtidos no trabalho. No capítulo cinco são apresentadas as conclusões oriundas dos resultados deste trabalho e trabalhos futuros.

2 REVISÃO BIBLIOGRÁFICA

Em busca de compreender a utilização de técnicas inteligentes para previsão de ventos, foi realizada uma revisão bibliográfica em bases de dados científicas sendo selecionada uma amostra entre os anos de 2015 até 2021. Esta amostra com principal palavra-chave *Weather Forecasting* representou 2043 artigos, sendo 1427 em periódicos, 495 em congressos, 14 livros e 8 outros. A partir dos artigos mais relevantes, que discorrem sobre as temáticas de previsão de vento, modelos estatísticos, utilização de redes neurais, algoritmos, geração de energia, entre outros, foi realizada uma descrição sucinta para melhor caracterizar o problema da previsão de vento.

2.1 Estado da arte

Em [10] uma nova abordagem híbrida foi proposta para *Wind Power Forecasting* (WPF) probabilístico, que emprega as técnicas de transformada *Wavelet*, rede neural convolucional e técnica de *ensemble*. Tal método foi amplamente analisado comparando com métodos de persistência de *benchmark* e modelos *Neural Network* (NN) rasos, como *Back-Propagation Algorithm* (BP) e *Support Vector Machine* (SVM). Os resultados obtidos demonstraram que essa nova abordagem foi superior as outras alternativas testadas, nos quesitos de confiabilidade, nitidez e habilidade geral. Assim, o método proposto apresenta alta estabilidade e forte robustez, sendo superior a todas as alternativas com as quais foi comparado.

No trabalho desenvolvido por [11], foi aplicado uma transformada *ridgelet* para energia eólica, originando um novo mecanismo híbrido de previsão de malha fechada baseado em método inteligente. O método proposto consiste em duas partes, na primeira seção uma previsão preliminar é implementada e na segunda seção um comparador retorna a saída da previsão para atingir o valor de previsão desejado. Essa metodologia em comparação com previsões tradicionais se mostrou um modelo com maior precisão.

O método híbrido desenvolvido por [12] com base em Regressão Vetorial de Suporte para *Wind Speed Forecasting* (WSF), é denominado *Phase Space Reconstruction*

Support Vector Regression Genetic Algorithm ($PSR - SVR_{GA}$) que é uma metodologia para previsão de séries temporais univariadas. Os parâmetros do *Support Vector Regression* (SRV) são ajustados por meio de algoritmo genético, obtidas as previsões desse modelo apresentam melhor desempenho para horizontes de curto prazo, sendo mais precisos para WSF e *Wind Power Forecasting* (WPF).

Em [13] foi desenvolvido um modelo de previsão da velocidade do vento combinando as abordagens *Wavelet Packet Decomposition* (WPD), *Convolutional Neural Network* (CNN) e *Convolutional Long Short Term Memory Network* (CNNLSTM). Nesse estudo o WPD é empregado para decompor a série temporal de velocidade do vento original em várias subcamadas, diminuindo assim, a não estacionariedade dos dados originais. As redes CNN e CNNLSTM foram utilizadas para o calcular a previsão e usufruir de suas capacidades de processamento não linear, a fim de obter resultados satisfatórios de previsão de velocidade do vento, se mostrando um modelo robusto e eficaz para previsão de séries temporais de velocidade de vento 1D.

Na pesquisa de [14] foi trabalhado um modelo de previsão de vento multimodelo orientada por dados, sendo desenvolvida com uma técnica de aprendizado de máquina de conjunto de duas camadas. Algoritmos como *Artificial neural networks* (ANN), *Support Vector Machine* (SVM), *Gradient Boosting Machine* (GBM) e *Random forest* (RF) com diferentes kernels foram testados e ajustados em ambas as camadas. A metodologia multimodelo desenvolvida superou os modelos de *benchmark*, *Mean Absolute Error* (NMAE) e *Root Mean Square Error* (NRMSE), bem como, a previsão probabilística gerada pelo modelo desenvolvido quantificou a incerteza das previsões juntamente com a previsão determinística.

Um novo mecanismo de previsão de energia eólica é proposto por [15] uma rede *Wavelet Neural Network* (WNN) com *Wavelet* multidimensional de Morlet como função de ativação dos neurônios ocultos e *Maximum Correntropy Criterion* (MCC) como critério de treinamento. Neste artigo ainda é proposto uma nova técnica de busca estocástica, o *Improved Clonal Selection Algorithm* (ICSA) , consistindo em um melhoramento do algoritmo de seleção clonal, adequada para o procedimento de treinamento para otimizar os parâmetros livres do mecanismo de previsão. A técnica baseada em função *Wavelet* de Morlet proposta, resulta em uma previsão de energia eólica mais precisa do que a função *Wavelet Mexican-hat* e o critério de treinamento MCC leva a erros de previsão mais baixos do que a medida de erro de treinamento tradicional de *Mean Squared Error* (MSE) .

A abordagem adotada por [16] propõe o aprimoramento do modelo de previsão de vento *Autorregressive Integrate Moving Average* (ARIMA), desse modo buscando uma abordagem mais adequada usando uma rede neural artificial (RNA) e um *Kalman Filter* (KF) para tratar os problemas de não linearidade e incerteza. Os ganhos do modelo híbrido foram o resultado de hibridizar as partes lineares e não lineares do modelo ARIMA em associação com um KF para tratar com a incerteza estocástica e ANN trabalhando com a não linearidade dentro do modelo híbrido KF-ANN.

A partir da combinação de vários modelos [17] propõe um modelo de combinação para previsão probabilística de velocidade do vento de curto prazo. Assim, o modelo *Gaussian Process Regression*(GPR) foi usado para combinar previsões independentes geradas por outros métodos, como ARIMA, *Extreme Learning Machine* (ELM), SVM e *Least Square Support Vector Machine* (LSSVM) em uma forma não linear em vez da forma linear comumente usada. A abordagem desenvolvida fornece informações mais probabilísticas para previsões de velocidade do vento, bem como melhora a precisão da previsão para previsões de valor único.

De acordo com [18] técnicas utilizadas recentemente não atendem as necessidades reais de despacho de uma rede. Nesse sentido, foi desenvolvido um novo modelo de previsão multi-passos para previsão operacional do vento. O modelo *96-steps of the next day*, nomeado de modelo *Cuckoo Search Fuzzy System Weather Research and Forecasting Ensemble* (CS-FS-WRF-E), que foi elaborado com base em uma previsão *Weather Research and Forecasting* (WRF) ensemble, um algoritmo *Cuckoo Search* (CS) e um sistema *fuzzy* otimizado, para reduzir as incertezas nas simulações determinísticas de *Numeral Weather Prediction* (NWP) de valor único.

Após obtenção dos resultados foi constatado que esse método é adequado para a aplicação prática da operação em um parque eólico, não só pela eficácia, como também pelo custo computacional aceitável. Podendo ainda, ser importante na regulamentação real de relatórios e verificações de energia para parques eólicos, fornecendo uma previsão operacional aprimorada de vento [18].

Entendendo a necessidade de realizar uma previsão precisa da potência de saída de um parque eólico foi elaborado por [19], um modelo híbrido (LSSVM-GSA) baseado na máquina de vetores de suporte de mínimos quadrados (LSSVM) e algoritmo *Aravitational Search Algorithm* (GSA) para previsão de curto prazo. Estabelecendo um

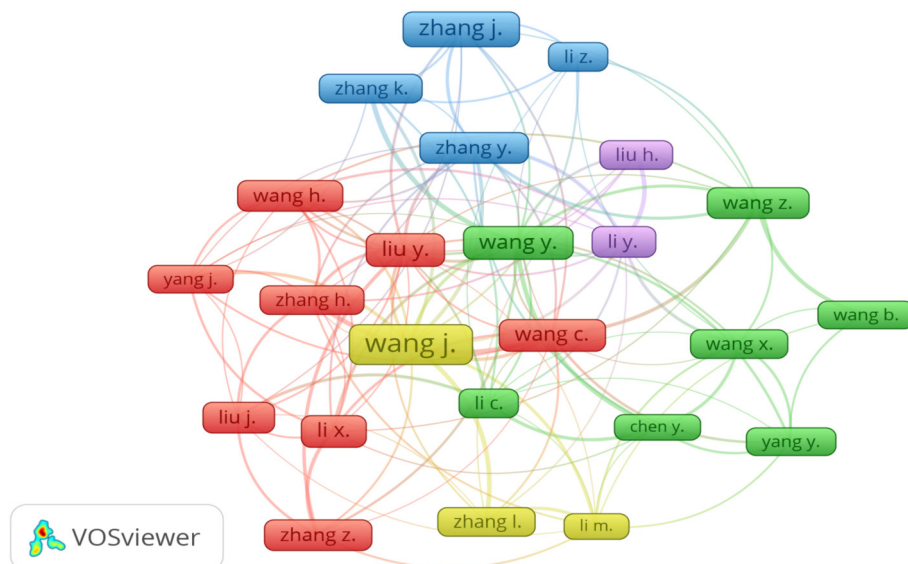
modelo a partir de diferentes aplicações de funções do *kernel*. Estabelecida uma função ideal, foram determinados os parâmetros do modelo LSSVM para otimização utilizando GSA.

2.2 Bibliometria

Em complemento a revisão bibliográfica foi realizado um breve estudo bibliométrico considerando aspectos quantitativos de produção, disseminação, socialização e evidenciação da informação registrada nas bases de dados científicas abordadas neste trabalho.

Com base em uma análise bibliométrica foi possível interpretar as conexões entres os artigos revisados, sendo as principais conexões realizadas considerando o autor, o número de citações dos periódicos, os países de origem destes trabalhos, os anos em que foram publicadas e as palavras-chave com maior aderência ao problema da previsão. A partir da análise por nome de autores, foi possível identificar a formação de 5 *clusters* predominantes, que se destacam pelo número de citações dos trabalhos publicados, conforme ilustrado por meio da Figura 2.1.

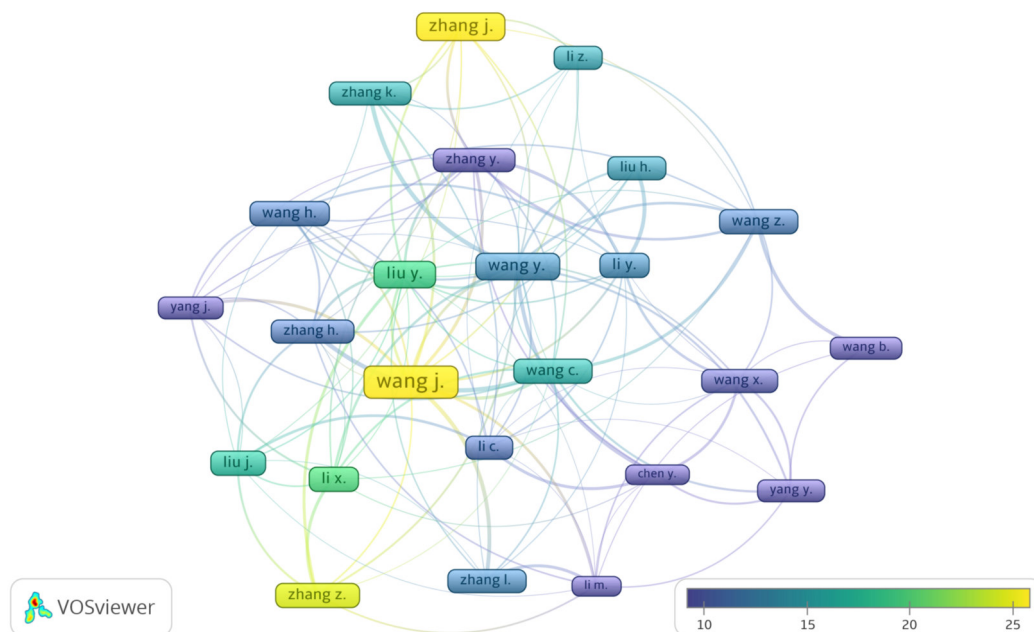
Figura 2.1: *Análise por autor*



Fonte: Elaborada pelo Autor.

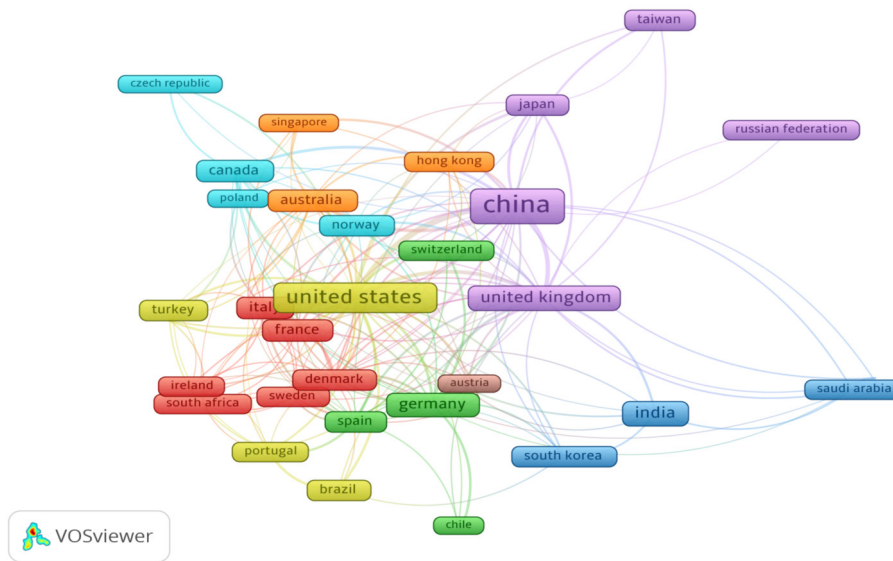
É possível observar que grande parte das publicações em destaque são produzidas por pessoas de nacionalidade asiática, contudo não foram em sua totalidade produzidas na Ásia e sim em países da Europa e Estados Unidos, onde há uma grande concentração de parques eólicos. Um outro ponto de destaque é o quadro em amarelo ao centro do gráfico, o que indica a origem da maior parte dos trabalhos, ou seja o artigo de referência nesta área. Na Figura 2.2 é ilustrado com base em cores os autores mais citados.

Figura 2.2: *Análise por quantidade de citações*



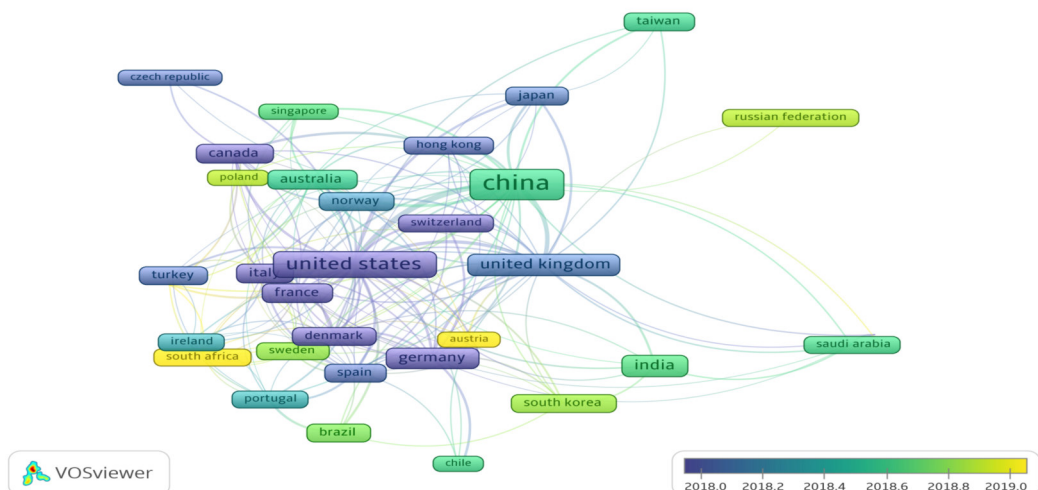
Fonte: Elaborada pelo Autor.

Da mesma forma que na Figura 2.1 a cor amarela ao centro e o tamanho do quadro com o nome do autor reforça a indicação do número de trabalhos citados (barra de cores no gráfico) e a existência de um artigo de referência. Assim, para entender qual a origem destes trabalhos, foi realizado a análise dos países em que foram originados os estudos (Figura 2.3).

Figura 2.3: *Análise por país*

Fonte: Elaborada pelo Autor.

Ainda na Figura 2.3 é possível observar uma maior heterogeneidade que na Figura 2.1, destacando 8 *clusters* bem definidos. Os dois *clusters* principais (amarelo e lilás) indicam como os maiores produtores de artigos voltados a previsão, Estados Unidos e China, ou seja embora os autores tenham correlação com o Ásia, um dos locais de estudo é a América. Na Figura 2.4 é relacionados os países e o ano de publicação.

Figura 2.4: *Análise por ano de publicação*

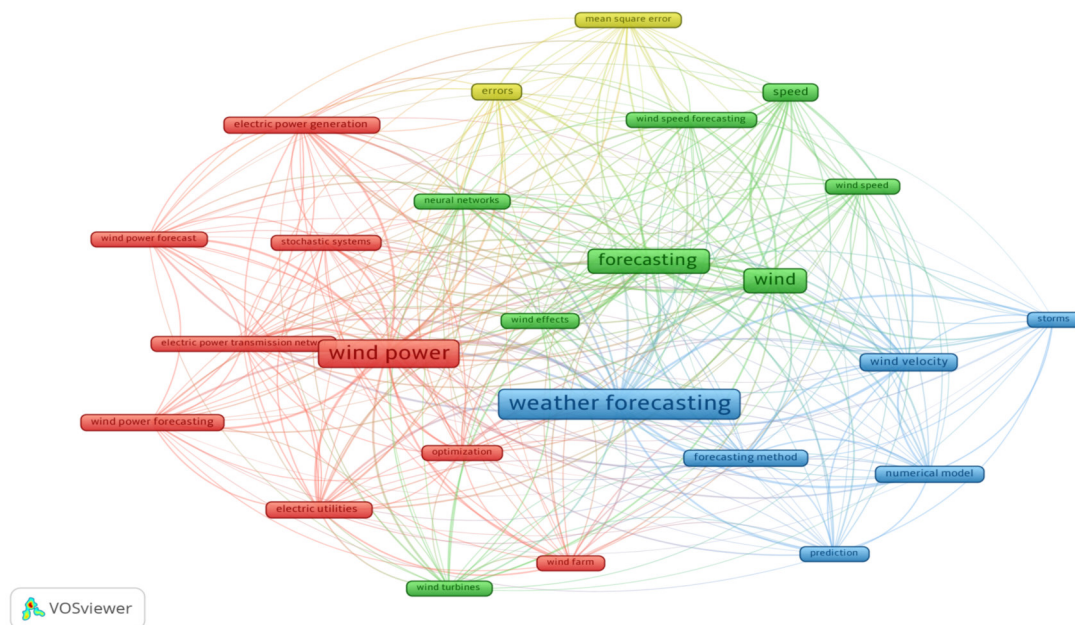
Fonte: Elaborada pelo Autor.

Diante disto, com base no ano os países que estão em tons mais amarelos se

encontram publicados em períodos mais recentes. Logo é possível visualizar que os estudos sobre previsão mais recentes vem ocorrendo em países no qual o recurso eólico vem sendo explorado recentemente. Já países no qual há uma tradição em geração eólica a previsão é uma área bastante estudada e por isso apresenta maior número de conexões (em lilás).

A Figura 2.5 ilustra uma investigação realizada a partir das palavras-chave mais utilizadas na amostra de artigos selecionadas. Com base nesta investigação foi possível identificar 4 *clusters* bem definidos, que se destacam pelo maior numero de ocorrências.

Figura 2.5: *Análise das palavras chaves*



Fonte: Elaborada pelo Autor.

As palavras-chave em destaque foram previsão de tempo (*weather forecasting*), energia eólica (*wind power*), apenas vento/eólica (*wind*) e apenas previsão (*forecasting*), uma vez que em buscas separadas de forma que não haja sobreposição de previsão de vento com vento nem de energia eólica com vento/eólica.

Por fim, ao analisar as palavras-chave a partir do período (Figura 2.5), foi observado que as palavras-chave definidas na Figura 2.6 são as que possuem mais aderência ao tema ao longo dos anos, logo permitem encontrar as publicações necessárias para revisão bibliográfica, bem como é uma forma de pôr um marcador para que o trabalho seja citado

3 GERAÇÃO DE ENERGIA EÓLICA

A geração de energia oriunda da produção eólica tem crescido em todo o Mundo, desde países menos desenvolvidos até superpotências. Isto pode ser explicado pela pegada ambiental, no qual há um incentivo internacional em cuidado ao meio ambiente e consequentemente a eliminação progressiva da geração fóssil e nuclear. Há também a descoberta de novas áreas (jazidas de vento) em que é possível extrair energia verde, com eficiência e com bom rendimento financeiro por parte dos investidores.

Todavia, uma vez que a indústria eólica cresce a passos largos, muitos países tem se beneficiado destas tecnologias, pois há um mercado latente em expansão e isto pode ser mensurado pelo aumento nas dimensões das turbinas e na capacidade de geração. Logo conhecer e estudar o recurso eólico é um ponto importante para entendimento destas tecnologias, assim como aplicação em locais com abundância de ventos.

3.1 Energia dos ventos

Os ventos resultam principalmente da diferença de pressão na atmosfera devido ao aquecimento diferenciado das massas de ar, sendo também influenciado pelo movimento de rotação da Terra que ocasiona variações sazonais na intensidade e direção. Além do sistema de vento global, há sistemas locais de ventos que são sistemas menores e atuam em microrregiões, a exemplo das brisas marítimas e terrestres que são resultado da diferença de capacidade de absorção de calor da terra e do mar [1].

A energia cinética proveniente dos ventos possui um elevado potencial de geração de energia elétrica, cujo aproveitamento do potencial energético depende de variáveis como: localização, a constância das correntes de ventos, a sazonalidade dos ventos e a velocidade dentro dos parâmetros necessários para geração de energia entre outros [6].

Em decorrência de tais peculiaridades, projetos relacionados ao aproveitamento do potencial eólico para geração de energia elétrica, precisam dispor de informações confiáveis e de qualidade para estimar a produção de energia a ser gerada. Em vista do alto investimento para captação de energia elétrica a partir do vento, é indispensável um

período de análise técnica da região a ser utilizada, objetivando minimizar os erros de estimativa de produção e eventuais prejuízos financeiros [5].

Com base nos dados gerados das análises técnicas de uma região, a partir da modelagem, tratamento estatísticos e das informações associadas ao vento, é possível realizar o dimensionamento adequado de parques eólicos. Equipamentos como anemômetro, usualmente instalados a 10 m do solo, permitem o registro de medidas como direção e intensidade, que possibilitam estimar o comportamento dos ventos por meio do processamento destes dados [9]. Ao extrair a curva de frequência da velocidade durante o processamento destes dados, é obtido várias curvas importantes, dentre as quais está a curva de potência média bruta ou fluxo de potência eólica [4].

Em virtude da dificuldade de previsão e aleatoriedade do fluxo de ventos, o processamento dos dados é realizado por meio de função de distribuição de probabilidade (FDP). Para dados de velocidade de ventos é admitido a probabilidade de existir uma velocidade do vento em um determinado intervalo [2]. Portanto, quando menor for o intervalo, menor será a probabilidade, sendo assim no momento em que o intervalo iguala a zero, a probabilidade também zera. Contudo a razão referente ao comprimento do intervalo tende ao infinito, diante disto a FDP deve ser zero ou maior que zero, mas não deve ser maior que 1 [3], logo:

$$0 \leq FDP \leq 1 \quad (3.1)$$

Em que a probabilidade de existir uma velocidade de vento Pv maior que v_1 e menor que v_2 é representada pela integral [3]:

$$Pv = \int_{v_1}^{v_2} FDP(v)dv \quad (3.2)$$

Ainda que as principais funções de distribuição de probabilidades aptas para utilização na engenharia sejam a Normal e Gaussiana, destaca-se a distribuição de *Rayleigh* e a distribuição de *Weibull* como FDPs eficientes para modelagem dos dados de vento [4]. Na distribuição de *Rayleigh* é conhecido apenas a média de velocidade do vento \bar{v} , logo é suposto que o vento obedece a distribuição de *Rayleigh*, portanto a $FDP(v)$ é definida como [3]:

$$FDP(v) = \frac{v}{\sigma^2} \exp \left[-\frac{1}{2} \left(\frac{v}{\sigma} \right)^2 \right] \quad (3.3)$$

Em que, σ é o moda da distribuição, ou seja o valor cujo a FDP atinge o pico e v é a velocidade média do vento. Apesar de σ não ser o valor médio, existe uma relação entre a velocidade média e a moda da distribuição da FDP de *Rayleigh* [3].

$$\sigma^2 = \frac{2}{\pi} \bar{v}^2 \quad (3.4)$$

Portanto,

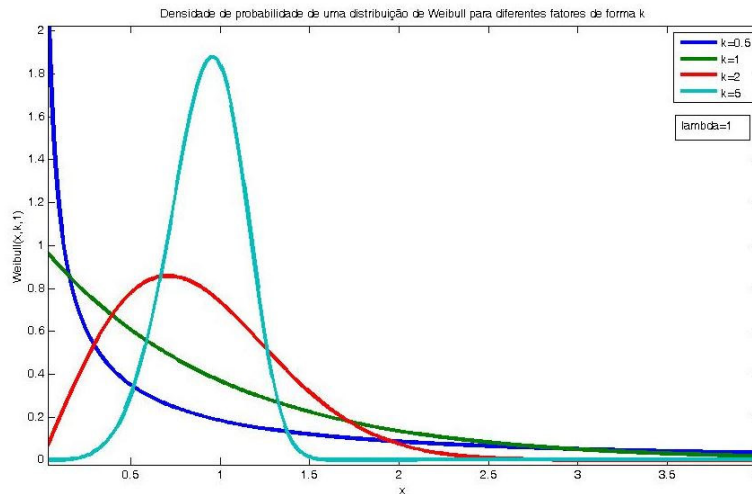
$$FDP(v) = \frac{\pi v}{2 \bar{v}^2} \exp \left[-\frac{\pi}{4} \left(\frac{v}{\bar{v}} \right)^2 \right] \quad (3.5)$$

Contudo, a distribuição de *Rayleigh* utiliza apenas um único parâmetro para representar o comportamento do vento, direcionando os dados observados para a função de distribuição de *Weibull* com dois parâmetros, em que a FDP de *Weibull* depende de κ [2]:

$$FDP(v) = \frac{\kappa}{c} \left(\frac{v}{c} \right)^{\kappa-1} \exp \left[-\left(\frac{v}{c} \right)^\kappa \right] \quad (3.6)$$

Em que c é o fator de escala e κ é um fator de formato ajustável. A distribuição de *Weibull* adere a diferentes formatos de acordo com o valor de κ , conforme ilustrado por meio da Figura 4.1.

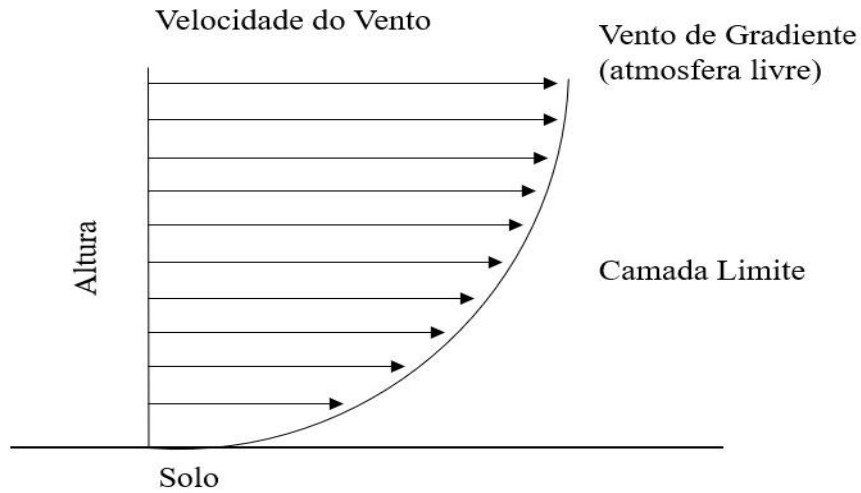
Figura 3.1: Curva de distribuição de Weibull para κ



Fonte: Adaptado de [4]

A partir da Figura 4.1 é possível verificar como o gradual aumento da variável κ tende a concentrar a ocorrência de registros em torno da média de velocidade média, assim há uma menor variabilidade da velocidade de vento, tendendo para o ideal.

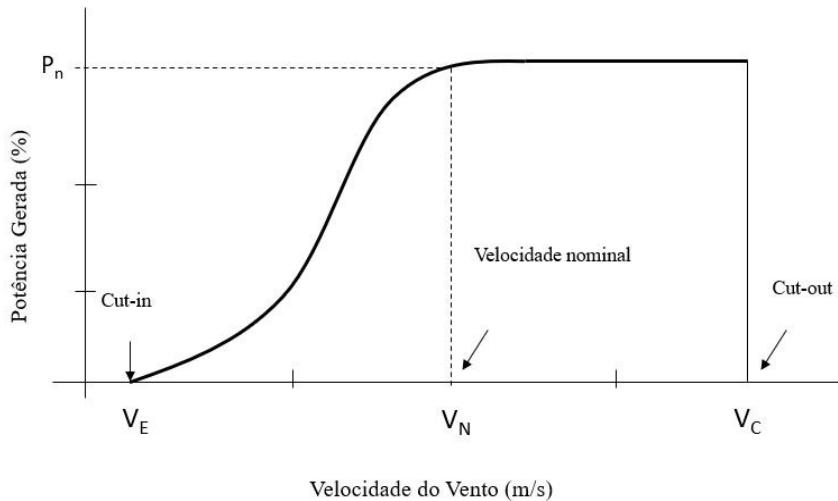
Figura 3.2: Perfil de vento



Fonte: Adaptado de [1]

Cada região tem seu próprio perfil de vento, contendo velocidades e especificidades que diferem de um local para outro, esse perfil é ilustrado na Figura 3.2, em que pode ser observado que a velocidade do vento pode variar conforme a altura, bem como tem um limite de atuação. Identificar o perfil do vento de uma localidade interfere diretamente na gestão do planejamento da instalação das turbinas, visto que com base no perfil de vento é possível determinar a vida útil das pás do rotor [2].

Figura 3.3: Curva de potência de uma aerogerador



Fonte: Adaptado de [2]

Como ilustrado na Figura 3.3, os aerogeradores começam a produzir energia a partir de uma específica velocidade de vento de partida (*cut in*), sendo parados quando

a velocidade ultrapassa o valor limite estipulado (*cut out*), instituído por medida de segurança [2, 3].

Portanto, é fundamental o conhecimento da velocidade média do vento para estimar a energia a ser gerada, e registrar a frequência da duração das variações dos ventos fortes (turbulências) e das calmarias. Dessa forma, a potencia disponível (P_d) para uma massa de ar em velocidade V , com uma densidade d , que flui perpendicularmente por uma área A , pode ser expressa por [3]:

$$P_d = \frac{1}{2}dAV^3 \quad (3.7)$$

Por meio de (3.7) a potencia é proporcional ao cubo da velocidade, o que sugere que qualquer variação de vento, mesmo que pequena, pode provocar grandes variações de potência. A alteração de velocidade do vento de acordo com a altura pode ser estimada com base em [1]:

$$V = V_0 \left(\frac{H}{H_0} \right)^n \quad (3.8)$$

Em que, V é a velocidade do vento; V_0 é a velocidade disponível na altura conhecida; H é a altura; H_0 é a altura conhecida e n é o fator de rugosidade do terreno. O potencial eólico convertido em eletricidade depende da área do rotor e do rendimento do aerogerador, podendo ser obtido por[4]:

$$\eta = \eta_B \eta_A \eta_M \eta_r \eta_G \quad (3.9)$$

No qual, η_B é o rendimento de Betz; η_A é o rendimento do aerogerador; η_M é o rendimento do multiplicador de velocidade; η_r é o rendimento do rotor e η_G é o rendimento do gerador [2].

O conhecimento sobre o recurso eólico e as equações que modelam e convertem a energia cinética dos ventos em energia elétrica é base importante para os modelos de previsão e são atualizados a partir das medições pontuais em campo.

3.2 Modelos de Previsão

Os modelos de previsão eólicas são métodos que estimam a produção esperada de turbinas ou de um parque eólico em um horizonte que pode variar de segundos a meses a

frente. Para realizar este procedimento de previsão são utilizados modelos computacionais que mediante ao processamento de determinadas variáveis produzem as estimativas da produção de geração de energia em um horizonte futuro [20].

Como o recurso eólico apresenta uma alta variabilidade, os métodos de previsão da velocidade do vento são importantes indicadores de capacidade produtiva, nível de manutenção e gerenciamento dos parques. A variabilidade do vento pode causar incertezas sobre a produção de energia eólica, o que atinge diretamente a estabilidade do mercado elétrico e os despachos do Operador Nacional do Sistema Elétrico (ONS) [21].

Diante da necessidade de prever e controlar a produção de energia, houve um aumento no número de pesquisas realizadas e no desenvolvimento de novos modelos, buscando uma previsão mais eficiente com menor margem de erro e com dados mais precisos. A análise dos dados de ventos deve considerar fatores que influenciam os resultados dos modelos de previsão, tais como: topologia do terreno, irregularidade da superfície, clima, turbulências, obstáculos, entre outros que podem interferir na produção e na vida útil dos equipamentos [23].

Os modelos de previsões podem ser diferenciados de acordo com a linha temporal que são realizadas as previsões, apesar disso, não existe unanimidade na definição que distingue uma linha temporal de atuação das previsões, entretanto é possível dividir em 4 categorias [21]:

- Curtíssimo prazo: de minutos a 1 hora à frente. É aplicado em ações regulatórias, previsão da geração e análise de mercado;
- Curto prazo: de 1 hora a várias horas à frente. É utilizado para planejamento de despacho de carga e segurança operacional;
- Médio prazo: de várias horas a uma semana à frente. Esta abordagem é utilizada para obter informações sobre tendências e padrões;
- Longo prazo: a partir de uma semana à frente. O principal objetivo desse modelo é o planejamento da manutenção, a gestão de operação e os estudos de viabilidade.

A previsão de produção ainda pode ser enquadrado em mais uma categoria denominada "*Nowcasting*", responsável pelas previsões para os próximos segundos, sendo utilizada para controle e gestão dos parques. Da mesma maneira que as previsões podem

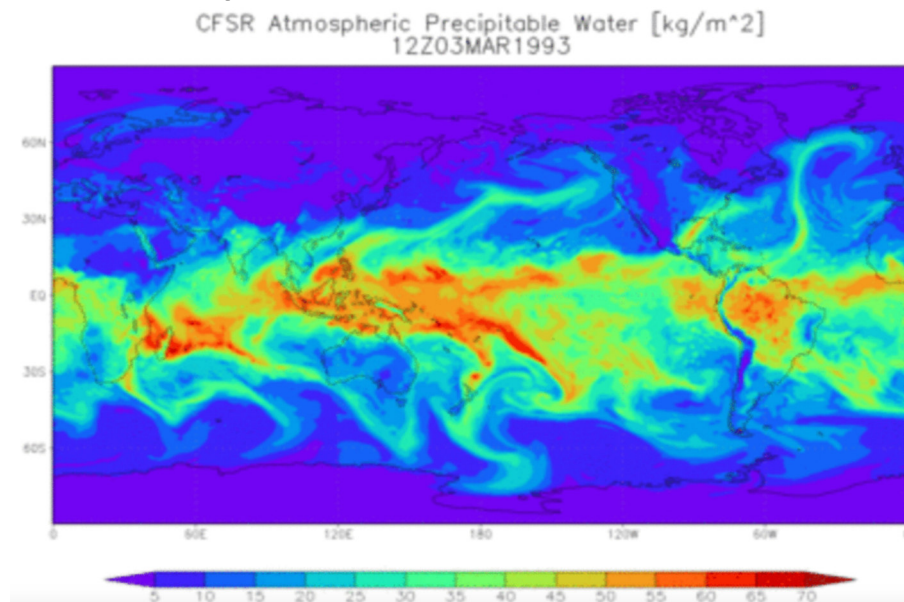
ser divididas pela linha temporal, podem também ser classificadas pelo método abordado: modelos físicos, estatísticos e inteligência artificial e híbridos [21].

3.2.1 Modelos Físicos

A abordagem física é baseada no processamento de informações das condições atmosféricas, topografia do terreno, entre outras medidas obtidas por meio de um sistema de previsão do tempo numérica (*Numerical Weather Prediction* - NWP) [26]. Os NWP utilizam as informações como temperatura, pressão, rugosidade e obstáculos, e convertem os dados fornecendo uma previsão da velocidade dos ventos [22].

O método físico é formado por um conjunto de modelos, que unificados permitem a conversão de dados capazes de estimar o vento em um quadrante da rede. Para realizar as previsões, os NWP necessitam de uma alta capacidade computacional, devido ao volume elevado de dados a serem tratados para desenvolver uma previsão [24][25]. Na Figura 3.4 é apresentado um modelo de previsão usando o NWP.

Figura 3.4: *Previsão de ventos via NWP*



Fonte: Adaptado de [28]

Os modelos físicos fazem uso de bases de dados meteorológicas globais e regionais, que exigem uma robustez computacional para obter previsões mais precisas. Nesse contexto, a abordagem física apresenta um bom desempenho para horizontes de previsão de médio a longo prazo, sendo inadequado para horizontes de previsões de curto e

curtíssimo prazo [26].

No Brasil há o BRAMS que é baseado no Sistema de Modelagem Atmosférica Regional (*Regional Atmospheric Modeling System - RAMS*) originalmente desenvolvido na CSU/EUA. O BRAMS/RAMS são modelos numéricos de previsão meteorológica multifuncionais que foram projetados para simular circulações atmosféricas que vão desde escalas hemisféricas até grandes simulações de vórtices da camada limite planetária [27].

3.2.2 Modelos Estatísticos

Os modelos estatísticos de modo geral, são mais fáceis de interpretar e implementar em comparação aos demais métodos. Afim de realizar as previsões, os métodos estatísticos processam as informações contidas em séries históricas, que armazenam dados de medições [29]. Apesar deste método oferecer uma maior praticidade, apresenta uma desvantagem, uma vez que a proporção que o tempo de previsão aumenta, o erro relacionado a este aumenta. Sendo assim é interessante a utilização para desenvolver previsões de curto prazo [32].

Há uma pluralidade de modelos estatísticos utilizados para previsões com séries temporais. Dos modelos comumente utilizados para realizar estas análises, estão os modelos auto regressivos de média móvel (*Autoregressive Moving Average - ARMA*), e modelos auto regressivos integrado de média móvel (*Autoregressive Integrated Moving Average - ARIMA*), bem como a utilização de inteligência artificial(IA) [31, 32].

O desenvolvimento de modelos baseados em séries temporais demanda o conhecimento prévio do comportamento da série de dados, em vista disto são compostos os primeiros gráficos que facultam o entendimento do comportamento desses dados. A partir das observações preliminares é possível destacar componentes sazonais, tendências, *outliers* e descontinuidades [29, 30].

Com base no processamento estatístico inicial é possível verificar com mais precisão o comportamento dos dados que trazem informações como média, desvio padrão, variância e valores máximos e mínimos, sendo a média um valor de tendência central, que resulta no valor médio do conjunto, podendo ser escrito como [33][34]:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad (3.10)$$

Em que, N é a dimensão da amostra formada pelos elementos x_1, x_2, \dots, x_n , no qual a variância e o desvio padrão são medidas de dispersão e podem ser do tipo [31]:

$$\sigma^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2 \quad (3.11)$$

No qual, a variância amostral é uma medida de dispersão que aponta o desvio padrão quadrático médio em relação a medida dos dados.

E conhecido que a variância não é capaz de preservar os elementos originais dos dados que tenha sido aplicada a variância. Logo, com a finalidade de manter os elementos originais é empregado o desvio padrão, definido como a raiz quadrada da variância [33]:

$$s = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2} \quad (3.12)$$

Os modelos ARMA e ARIMA são métodos estatísticos de previsão desenvolvidos por Box e Jenkins na década de 70, em que uma das características desta metodologia é a realização de previsão com base em series de dados de medições passadas [31]. O modelo ARMA é a resultante da soma dos processos auto-regressivos (AR) de ordem maior ou igual a zero e por médias moveis (MA) de ordem maior ou igual a zero, sendo representados por ARMA (p,q). Os processos auto-regressivos (AR) podem ser descritos em [35][33].

Considerando que p são valores passados, a_t são choques aleatórios, uma determinada função de regressão pode ser representada por $\bar{z} = z_t - \mu$ em que [31][35]:

$$\bar{z} = \phi_1 \bar{z}_{t-1} + \phi_2 \bar{z}_{t-2} + \dots + \phi_p \bar{z}_{t-p} + a_t \quad (3.13)$$

Logo, ϕ representa os valores dos parâmetros autorregressivos. Se for reutilizado (3.13) pode ser reescrita utilizando o operador B [35]:

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (3.14)$$

Em que, B é operador de deslocamento atrasado, podendo ser escrito na forma reduzida [35]:

$$\phi(B) \bar{z}_t = a_t \quad (3.15)$$

Para modelos de média móvel (MA), as definições são realizadas de maneira similar aos autorregressivos. Um processo aleatório com média zero, cuja série temporal seja \bar{z}_t e de variância σ_2^a , e uma metodologia de média móveis que seja linearmente dependente de valores q de a_t , podendo ser descrito como [31]:

$$\bar{z} = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_p a_{t-p} \quad (3.16)$$

No qual, θ_i são os valores referentes aos parâmetros de média móvel. Essa equação pode ser definida utilizando o operador B de média móvel de ordem q , assim como [35][33]:

$$\theta(B) = 1 - \theta_1(B) - \theta_2(B^2) - \dots - \theta_q B^q \quad (3.17)$$

Sendo ainda possível utilizar na forma simplificada [35]:

$$\bar{z}_t = \theta_1(B) a_t \quad (3.18)$$

A aglutinação ente os processos autorregressivos e os de média móvel, originam o modelo denominado ARMA ou modelo autorregressivo de média móvel, sendo do tipo [30][33]:

$$\bar{z} = \phi_1 \bar{z}_{t-1} + \phi_p \bar{z}_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (3.19)$$

Assim como as demais, (3.19) também pode ser escrita na fórmula simplificada como [35]:

$$\phi(B) \bar{z}_t = \theta(B) a_t \quad (3.20)$$

A metodologia ARMA foi elaborada para realização dos processos autorregressivos em séries temporais estacionárias. Diferente desta metodologia, o modelo ARIMA(p, d, q) foi desenvolvido para séries temporais não estacionárias. Neste caso, os valores não oscilam em torno de uma média fixa. Contudo, apesar de uma série temporal não ser estacionária, pode demonstrar em um dado intervalo de tempo, algum comportamento homogêneo. O modelo ARIMA pode ser representado na forma [31][35]:

$$\varphi(B) \bar{z}_t = \theta(B) a_t \quad (3.21)$$

Em que $\varphi(B)$ é um operador autorregressivo não estacionário, assumindo que o polinômio

$\varphi(B) = 0$ contem d raízes unitárias e as restantes estiverem distante do centro unitário, (3.21) pode ser reescrita como [35]:

$$\varphi(B)\bar{z}_t = \phi(B)(1 - B)^d \bar{z}_t = \theta(B)a_t \quad (3.22)$$

Se for considerado o operador diferencial ∇ igual a $(1 - B)$, a (3.22) pode ser reescrita substituindo os termos, assim [35]:

$$\phi(B)\nabla^d z_t = \theta(B)a_t \quad (3.23)$$

Logo,

$$\nabla^d \bar{z}_t = \nabla^d z_t \quad (3.24)$$

Sendo ainda possível, a partir da interação $\nabla^d z_t = w_t$ reformular a (3.23):

$$\phi(B)w_t = \theta(B)a_t \quad (3.25)$$

Portanto, o modelo ARIMA é uma derivação do modelo ARMA com a finalidade de processar séries temporais diferenciadas. A peculiaridade apresentada por este modelo ao inserir o termo “integrado”, ocorre em virtude da iteração de $\nabla^d z_t = w_t$, que também pode ser escrita de forma inversa, sendo[35]:

$$z_t = S^d w_t \quad (3.26)$$

Em que,

$$S = \nabla^{-1} = (1 - B)^{-1} = 1 + B + B^2 + \dots + B^n \quad (3.27)$$

No qual, S o operador somatório (3.27), ainda é possível transcrever como:

$$S w_t = \sum_{j=0}^{\infty} w_{t-j} = w_t + w_{t-1} + w_{t-2} + \dots \quad (3.28)$$

Não obstante a metodologia ARIMA é resultado da soma do método ARMA w_t diferenciado d vezes [35].

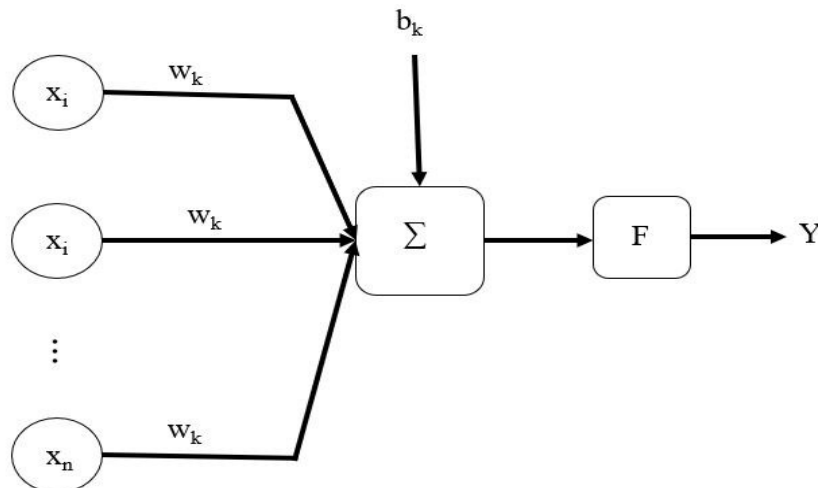
3.2.3 Modelo baseado em Inteligência Artificial

Quando se aborda o uso de Inteligência Artificial (IA) para resolução de problemas de engenharia, está se falando de uma diversidade de metaheurísticas (redes neurais artificiais,

algoritmos evolutivos, estratégias evolutivas, enxame de partículas, máquinas de suporte vetorial, lógica nebulosa, etc) que podem ser usadas em otimização, classificação, clusterização, *tracking* entre outras aplicações [33] [36].

No caso em particular das previsão, há muitas aplicações baseada no uso de redes neurais artificiais (RNA). Com o surgimentos das RNAs, várias áreas da pesquisa tem utilizados esta ferramenta, viabilizando estudos com melhores desempenhos. Não diferente de outras áreas, o setor das energias tem aplicado as RNAs em vários ambientes. Nas análises de previsão tem sido usadas em conjunto com as séries temporais, possibilitando o processamento de dados mais complexos [36]. Na Figura 3.5 é ilustrada a composição de uma RNA clássica, que contém uma primeira camada denominada neurônios de entrada ($x_1 - x_n$), uma camada intermediária ou oculta (representada pelo Σ) e uma última denominada neurônios de saída (Y).

Figura 3.5: *Esquema de um neurônio artificial*



Fonte: Adaptado de [36]

Uma RNA pode ter n camadas, no qual a quantidade de camadas depende da complexidade da aplicação, e de forma geral possui [37]:

- Sinapses ou elos de conexão: são caracterizadas individualmente por peso sináptico, em que as entradas x_i são multiplicadas por pesos sinápticos w_{ki} ;
- *Bias*: que representa uma entrada externa que tem a finalidade de aumentar ou diminuir regulando a informação de saída de um neurônio;

- Somador: que faz a somatória dos sinais de entrada multiplicados pelos respectivos pesos sinápticos dando origem a v_k e;
- Função ativação: em que $\varphi(\cdot)$ delimita a amplitude de um neurônio.

Conforme a complexidade e o objetivo da aplicação, uma RNA pode apresentar diferentes funções de ativação, sendo necessário escolher a função que melhor se habilite ao estudo. Por exemplo uma função sigmoide, limiar, linear e tangente hiperbólica são funções de ativação [36].

- Função limiar

$$\varphi(v) = \begin{cases} 1 & \text{se } v \geq t \\ 0 & \text{se } v < t \end{cases} \quad (3.29)$$

- Função linear

$$\varphi(v) = \begin{cases} 1 & \text{se } v \geq t \\ v & \text{se } \frac{1}{2} > v > -\frac{1}{2} \\ 0 & \text{se } v \leq -\frac{1}{2} \end{cases} \quad (3.30)$$

- Função sigmoide

$$\varphi(v) = \frac{1}{1 + e^{-v}} \quad (3.31)$$

- função tangente hiperbólica

$$\varphi(v) = \begin{cases} 1 & \text{se } v > 0 \\ 0 & \text{se } v = 0 \\ -1 & \text{se } v < 0 \end{cases} \quad (3.32)$$

Redes do tipo *Multilayer Perceptron* - MLP, originalmente foram desenvolvidas para processamento de dados e classificação. Conforme foram sendo desenvolvidas outras pesquisas, essas redes foram evoluindo, sendo capazes de realizar atividades de regressão. Para problemáticas que envolvem regressão comumente são utilizadas funções não lineares [36][37].

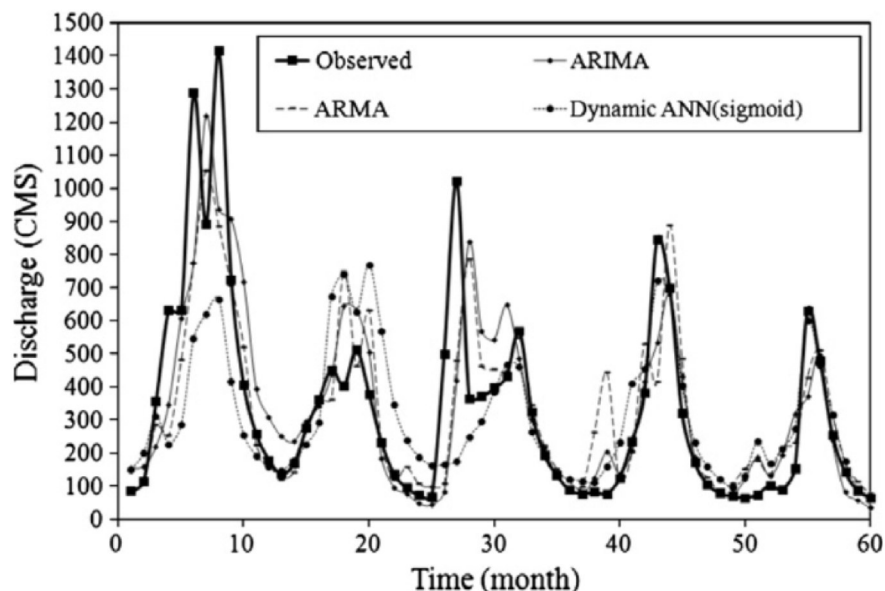
3.2.4 Modelos Híbridos

Nos modelos híbridos são combinados diferentes abordagens para realizar as previsões, como modelos de curto e médio prazo ou físicos e estatísticos. Este tipo de método tem como finalidade se beneficiar das vantagens de ambos os métodos combinados, buscando assim ter previsões mais precisas [38].

Portanto, é possível afirmar que os modelos anteriores tem uma limitação, já nos modelos híbridos é possível maximizar o alcance da previsão uma vez que usufrui das vantagens dos métodos em que está associado, impulsionando a capacidade de previsão [38].

Os modelos híbridos podem ser oriundos das seguintes combinações: abordagens físicas e estatísticas; modelos de curto e médio prazo; modelos físicos e inteligência artificial entre outros. Os modelos de abordagem híbrida no geral tem sido utilizados para previsões de médio e curto prazo, sendo muito utilizados para previsões de geração de energia eólica [38]. Na Figura 3.6 é ilustrado o gráfico de resultados obtidos ao longo de um período de simulação, no qual representa um modelo híbrido e a comparação entre ARMA, ARIMA e Redes Neurais.

Figura 3.6: *Resultados obtidos após um período de treinamento*



Fonte: Adaptado de [30]

Independente do método utilizado seja físico, estatístico ou baseado em IA, a área de previsão tem crescido nos últimos 10 anos, devido ao aumento do número de

parques instalados em todo o mundo. Logo há uma necessidade de portanto, prever a potência contratada que deve ser entregue a rede de concessionária. Dessa maneira, é uma área em que trabalhos que envolvam o tema previsão, estes devem obter um maior destaque.

4 TEORIA DA INFORMAÇÃO *MEAN SHIFT*

A teoria da informação foi inicialmente abordada por Shannon em 1948, sendo originalmente voltada para fins de viabilizar transmissão, processamento e compressão de informação. As bases da teoria proposta por Shannon, teria posterior impacto no desenvolvimento de sistemas de comunicação bem como em outras áreas, visto que a informação transita em todos os lugares, sendo objeto de estudo em varias áreas do conhecimento [39][37].

Inicialmente a teoria da informação teve os conceitos utilizados para transmissão de sinais e com a evolução foi adaptada em outras áreas para análise de dados. Logo, Shannon, definiu o conceito de entropia como uma métrica para a incerteza contida na informação [40].

Nesse sentido, a entropia tem como finalidade a qualificação da informação compreendida na mensagem, identificando qualquer deformidade que provoque redundância na informação [41]. Desta forma, para uma variável aleatória X em que $X = x_1, x_2, \dots, x_n$ com probabilidade $P = p_1, p_2, \dots, p_n$ e $k = 1, 2, \dots, n$, a entropia pode ser expressa por:

$$H_s(P) = \sum_{k=1}^n p_k \cdot \log \left(\frac{1}{p_k} \right) \quad (4.1)$$

Sendo que,

$$\sum_{k=1}^n p_k = 1; p_k \geq 0 \text{ e } 0 \cdot \log_2(0) = 0. \quad (4.2)$$

De modo que a entropia de X não seja dependente dos valores de X , mas do valor da propriedade de cada X , e portanto, seja uma medida de incerteza de uma variável aleatória comum e composta [40].

4.1 Entropia de Rényi

Em 1950, Rényi introduziu uma nova medida de entropia baseada em uma generalização da entropia de Shannon. A entropia desenvolvida por Rényi buscava descobrir uma ca-

tegoria mais geral de medidas de informação que conservasse a propriedade aditiva dos sistemas independentes estatísticos [40]. Logo, para definir a entropia de Rényi é necessário um dado parâmetro α de uma variável aleatória discreta P e com uma FDP p_k , resultando em:

$$H_{R\alpha}(P) = \frac{1}{1-\alpha} \log \left(\sum_{k=1}^N p_k^\alpha \right) \quad (4.3)$$

Para,

$$\alpha > 0, \alpha \neq 1 \quad (4.4)$$

Desse modo, a entropia de Rényi é mais versátil em função do parâmetro α , que permitiu a medição de vários níveis de incerteza. Considerando que α tende para 1 a entropia de Rényi se apresenta da mesma maneira que a entropia de Shannon, em outros termos é possível informar que a entropia de Shannon é um caso especial da entropia de Rényi [39].

A entropia de Rényi tem sido bastante utilizada como estimador não paramétrico devido a facilidade de implementação computacional, em destaque a forma quadrática da entropia [41] [39]. A entropia quadrática acontece quando $\alpha = 2$:

$$H_{R2}(P) = -\log \sum_{k=1}^N p_k \log p_k \quad (4.5)$$

Nesse contexto, em uma situação de agrupamento de dados é habitual não se conhecer qualquer informação sobre os dados [40]. Em vista disto, uma das maneiras mais eficientes de estimar uma FDP é utilizando a janela de Parzen, permitindo que a entropia de Rényi execute a estimação diretamente dos dados.

4.2 Informação Mútua

Diante da necessidade de analisar um sistema de informações que continham tanto entradas quanto saídas, em que cada uma possui as incertezas, metodologias foram desenvolvidas para o processamento dessas informações. Devido a isso o estudo sobre informação mútua, dentro da teoria da informação se fez de suma importância [41].

Shannon desenvolveu um conceito de entropia prevendo uma medida para apenas uma variável aleatória. Dentre os diferentes conceitos de informação mútua, o modelo de entropia de Rényi e a desigualdade proposta por Cauchy-Schwartz são uns dos mais

utilizados para variáveis aleatórias [40]. Intitulada como entropia relativa, a informação mútua de Renyi pode ser descrita como:

$$D_{R\alpha}(P_1 || P_2) = \frac{1}{\alpha - 1} \log \int \frac{P_1^\alpha}{P_2^{\alpha-1}} dx \quad (4.6)$$

Considerando que P_1 e P_2 como duas FDPs e α a divergência de Rényi. Assumindo uma entropia relativa $I_R(P_1; P_2)$ entre a FDP conjunta $P_{P_1 P_2}(p_{k1}; p_{k2})$ e as funções marginais $P_{P_1}(p_{k1})$ e $P_{P_2}(p_{k2})$ e presumindo que a igualdade existe somente se $P_{P_1}(p_{k1})$ e $P_{P_2}(p_{k2})$ forem independentes [41].

$$I_R(P_1; P_2) = \log \int_{P_1} \int_{P_2} \frac{P_{P_1 P_2}^2(p_{k1}; p_{k2})}{P_{P_1}(p_{k1}) P_{P_2}(p_{k2})} dx dy = D_{R_2}(P_{P_1 P_2}^2(p_{k1}; p_{k2}) || P_{P_1}(p_{k1}) P_{P_2}(p_{k2})) \geq 0 \quad (4.7)$$

No caso da divergência de Cauchy-Schwartz sendo fundamentada na entropia de Rényi, as aplicações são realizadas por meio de FDPs e em densidades de informação mútua de variáveis aleatórias [39].

$$D_{CS} = -\log \frac{P_1^T P_2}{\sqrt{\|P_1\|^2 \|P_2\|^2}} \quad (4.8)$$

Desenvolvendo a expressão de acordo com a entropia de Rényi, admitindo que $h_{R_2}(P_1)$ e $h_{R_2}(P_2)$ são entropias quadráticas de Rényi em que P_1 e P_2 e $h_{R_2}(P_1 \times P_2)$ e sendo a entropia cruzada de ambas as funções densidade [41].

$$D_{CS}(P_1 || P_2) = -\log \sum_{p_{k1} \in P_1} P_1(p_{k1}) P_2(p_{k1}) - \frac{1}{2} (-\log \sum_{p_{k1} \in P_1} P_1^2) - \frac{1}{2} (-\log \sum_{p_{k1} \in P_1} P_2^2)$$

$$\Rightarrow D_{CS}(P_1 || P_2) = h_{R_2}(P_1 \times P_2) - \frac{1}{2} h_{R_2}(P_1) - \frac{1}{2} h_{R_2}(P_2) \quad (4.9)$$

A partir de (4.9) é possível aplicar a desigualdade de Cauchy-Schwartz na entropia de Rényi, para as densidades de P_1 e P_2 :

$$D_{CS} = -\log \frac{\int_{P_1} P_1(p_{k1}) P_2(p_{k1})}{\sqrt{\left(\int_{P_1} P_1^2(p_{k1})\right) \left(\int_{P_1} P_2^2(p_{k1})\right)}} \quad (4.10)$$

Assim como na (4.8) a (4.10) pode ser desenvolvida com base na entropia de Rényi, possibilitando a simplificação da relação dos termos [41]. A desigualdade de Cauchy-Schwartz pode ainda ser utilizada para obter a informação mútua de duas variáveis aleatórias [40].

4.3 Janela de Parzen

A janela de Parzen, desenvolvida por Parzen em 1962, permite estimar uma FDP com base nos próprios dados, visto que é possível a extração determinada amostras N de uma dada variável aleatória. A finalidade do método é estimar a FDP de N com base no N original, logo tal particularidade tipifica esse método como não paramétrico [42].

Nas metodologias paramétricas a função densidade de probabilidade é conhecida, sendo desconhecido apenas alguns parâmetros. A utilização dos métodos paramétricos está condicionada a informações que atendam aos requisitos para estimar a FDP. Caso isto ocorra, faz-se necessário a utilização do método não paramétrico que irá estimar a FDP a partir do conjunto de dados [41].

Utilizando a janela de Parzen para estimar a densidade de um conjunto de dados, sabendo que a FDP $p(x)$ é desconhecida, dado a escolha aleatória de um valor x extraído de $p(x)$ e pertencente a R , de modo que $p(x)$ não varie, o método pode ser definido como [42]:

$$\hat{P} = \int_R p(x)' dx' = p(x) \cdot V \quad (4.11)$$

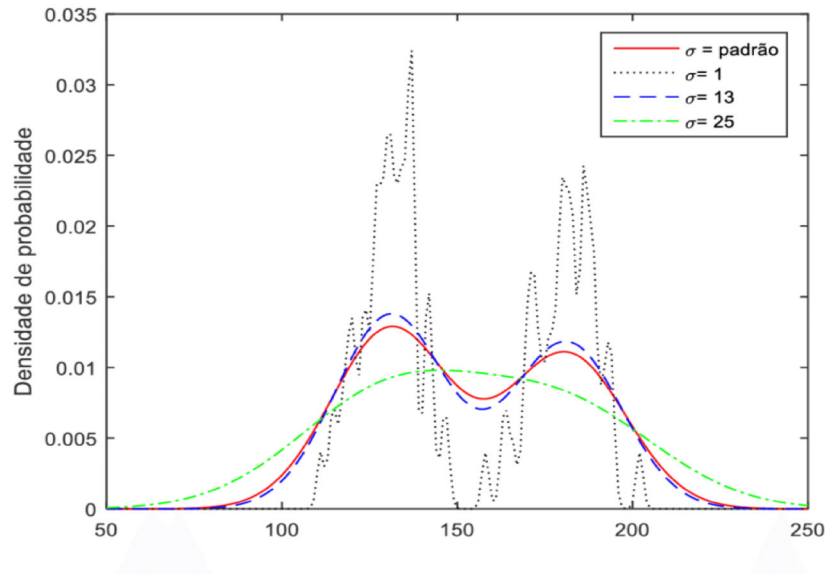
Em que, V é volume que pertence a R .

Várias funções podem ser utilizadas através da janela de Parzen para determinar a densidade de uma FDP, a mais utilizada é a função de kernel gaussiano [37]. Nesse processo a função kernel mapeia os pontos criando um conjunto contínuo derivado da FDP [41]. A função densidade de probabilidade é resultado das somas individuais do kernel, logo este processo pode ser expresso:

$$P_X(x) = \frac{1}{N} \sum_{i=1}^N K_\sigma \left(\left\| \frac{x - x_i}{\sigma} \right\|^2 \right) \quad (4.12)$$

Para, $K_\sigma = e^{-\frac{1}{2}}$ e $\sigma > 0$, no qual σ é o desvio padrão da função gaussiana e essencial para estabelecer a precisão da densidade prevista.

Assim quando o valor de σ é muito pequeno a tendência da densidade é apresentar muitos picos, enquanto que para valores maiores a tendência é de suavizar a FDP (Figura 4.1), podendo provocar a falta de interação entre os pontos estabelecidos pelo kernel, gerando perda na informação [41].

Figura 4.1: *Influência do σ na estimativa da FDP pela Janela de Parzen*

Fonte: Adaptado de [43]

4.4 Algoritmo Mean Shift

O Algoritmo *Mean Shift* foi desenvolvido baseado na teoria da informação e entropia de Rényi, sendo comumente utilizado para determinar a moda de conjuntos, a clusterização, a classificação, a segmentação, o *tracking* e recentemente aplicado a densificação [45][46].

Este algoritmo possibilita a captura das estruturas preponderantes de n conglomerados de dados por meio de uma função densidade de probabilidade com Kernel Gaussiano, sendo um método muito utilizado em análises de processamento de imagens [47][44].

A utilização do algoritmo *Mean Shift* no contexto da análise de previsão de sistemas eólicos implica em uma abordagem diferenciada para provisão em séries temporais [46]. Logo, admitindo uma amostra de dados $x_0 = (x_1)_{i=1}^n ER^D$, com base no modelo não-paramétrico de janela de Parzen, a função de densidade de probabilidade é definida por:

$$p(x, \sigma) = 1/n \sum_{i=1}^n G_{\sigma}(x - x_1) \quad (4.13)$$

Em que G_0 um kernel gaussiano dado por $G(t) = e^{-1/2}$ com largura de banda $\sigma > 0$.

Por meio do algoritmo *Mean Shift* é possível obter a moda de um conjunto em uma base de dados, em que $\rho(x) = 0$ [44]. Para isso é necessário que o ponto estacionário seja dado por:

$$m(x) = \frac{\sum_{i=1}^n G_{\sigma}(x - x_i) * x_i}{\sum_{i=1}^n G_{\sigma}(x - x_i)} \quad (4.14)$$

Em que a diferença entre $m(x)-x$, é denominado de *Mean Shift* [46][48]. Os parâmetros de utilização deste algoritmo reduzem a entropia de x , $H(x)$, a medida em que se conserva a distância de Cauchy-Schwartz $D_{cs}(x, x_0)$, entre x e x_0 anexado em algum valor de K [45]. Desta forma, a função FDP pode ser calculada segundo a entropia de Renyi's, como:

$$H(x) = -\log \int_{-\infty}^{+\infty} p_2(x) dx \quad (4.15)$$

Substituindo (4.13) em (4.15) tem-se:

$$H(x) = -\log V(x) \quad (4.16)$$

Em que o $\sigma' = \sqrt{2\sigma}$ e $V(x)$ é denominado como a informação potencial da FDP de $p(x)$ [46]. Na sequencia a entropia cruzada entre duas FDPs pode ser calculada por:

$$H(x, x_0) = -\log V(x, x_0) \Rightarrow V(x, x_0) = \frac{1}{n^2} \sum_{j=1}^n G_{\sigma'}(x, x_0) \quad (4.17)$$

Para duas FDPs p e q , a distância de Cauchy-Schwartz pode ser expressa por [45][46]:

$$D(x, x_0) = \log \frac{(\int p^2(x) dx)(\int q^2(x) dx)}{(\int p(x)q(x) dx)^2} \quad (4.18)$$

Considerando o conceito de que o cosseno de um ângulo entre dois vetores e a medida da distância, tem-se:

$$D_{cs}(x, x_0) = -[H(x) + H(x_0) - 2H(x, x_0)] \quad (4.19)$$

O deslocamento entre os dois objetivos pode ser otimizado é representado pela fórmula de otimização irrestrita sob o parâmetro λ [46].

$$J(x) = \min H(x) + \lambda [D_{cs}(x, x_0) - K] \quad (4.20)$$

Com base na diferença de $J(x)$ em relação a cada $x_i \in x$, faculta a transmutação de x_0 em outro conjunto x^{t+1} na interação $t + 1$, fazendo uso da informação inclusa em x na interação [45].

O agrupamento das verificações em distintas classes são obtidas pelo deslocamento de cada verificação. Cada deslocamento de um ponto $x_i^t \in x^t$, descreve uma FDP de x^t que contem informações da FDP de x_0 [47].

O conjunto $x_v = [x^1 \cup x^2 \dots \cup x^t]$ equivalente aos dados virtuais concebidos pelo algoritmo *Mean Shift*, que formam um *cluster* denso possuindo propriedades compatíveis com a serie original x_0 [46]. Este novo arquivo com novos dados virtuais que conservam as mesmas FDPs originais são utilizados como cenários para a previsão.

5 *MEAN SHIFT* APLICADO A PREVISÃO EÓLICA

O Maranhão apresenta um alto potencial para instalação de parques eólicos, a região costeira do estado possui grande influência dos ventos alísios, o recorte do litoral e a estrutura do relevo favorece a baixa variação dos picos de velocidades dos ventos. A região denominada Parque Nacional dos Lençóis Maranhenses, composta pelos Municípios de Barreirinhas, Paulinho Neves, Tutóia, Santo Amaro do Maranhão e Humberto de Campos, tem apresentado um potencial latente para geração de energia eólica, tendo despertado interesse de empresas do setor elétrico/eólico.

Em busca de identificar os locais que apresentem o melhores condições para geração de energia, projetos como o EOSOLAR vem sendo desenvolvidos . Este visa avaliar os efeitos micrometeorológicos em diferentes escalas temporais e espaciais, e são voltados ao planejamento e operação de parques eólicos e fotovoltaicos. O EOSOLAR é liderado pelo Instituto de Energia Elétrica da Universidade Federal do Maranhão e tem financiamento de empresas do setor elétrico do Maranhão.

Os estudos realizados pelo EOSOLAR visam mapear e avaliar o potencial eólico (recurso eólico) e solar do estado do Maranhão, fazendo medições na região de Barreirinhas e Paulino Neves em vários pontos de interesse, e para isto utiliza a tecnologia mais avançada existente no mercado internacional com foco na prospecção do vento.

A captação dos dados vem sendo realizada por meio de microcentrais anemométricas e perfiladores como o Sodar e Lidar, equipamentos que fazem o monitoramento de variáveis como: velocidade e direção de vento, pressão atmosférica, temperatura, pluviosidade, entre outras que permitem entender o comportamento do vento e dos efeitos associados como rugosidade, definição de camada limite atmosférica, turbulência, etc., além de permitir analisar estes efeitos sobre a estrutura física (impacto sobre a manutenção) e elétrica (possíveis desvios de geração).

Um ponto em destaque é que os perfiladores fazem medições de direção e velocidade de vento em frequências que permitem decompor o vento em escalas temporais

de segundos, sendo este registro realizado a partir dos 50 metros a cada 10 metros até a altura máxima de 280 metros.

O monitoramento realizado pelo Sodar ocorre a partir de disparos de pulsos sonoros em direção a atmosfera, que ao entrarem em contato com as partículas de aerossóis em movimento, retorna esta informação para o equipamento. O Lidar aplica a mesma idéia do Sodar, porém executa essa função emitindo feixes de laser para a atmosfera. Na Figura 5.1 é ilustrado os perfiladores Sodar (ilustração a) e Lidar (ilustração b) com aplicação em campo.

Figura 5.1: *Sodar(a) e Lidar(b)*



Fonte: Elaborada pelo Autor

A utilização desses equipamentos com alto grau de precisão, devido a medição a cada 10 metros, viabiliza a criação de um banco de dados para a realização de pesquisas de instituições públicas e privadas que tenham interesse em explorar o potencial dos locais estudados, comparando os que apresentam melhor desempenho.

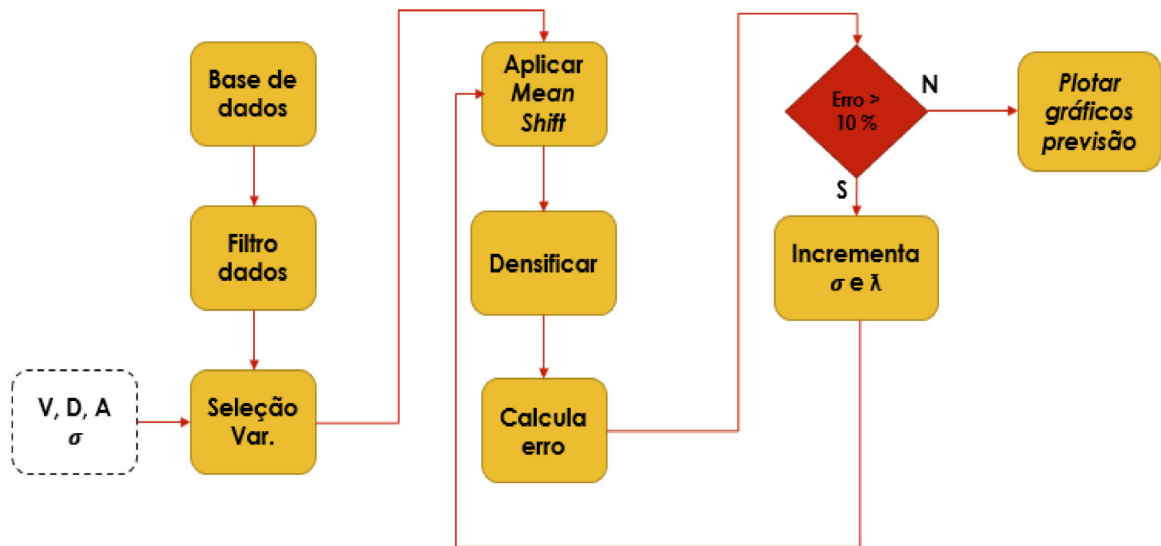
Para realização desta proposta de dissertação de mestrado, foram utilizados dados de medição do município de Paulinho Neves, concedidos dentro do projeto EOSOLAR. Essas informações viabilizam a proposta dessa pesquisa em aplicar do algoritmo *Mean Shift* com foco em uma metodologia inteligente para a previsão de geração de energia elétrica.

5.1 Modelo de Previsão

O modelo de previsão abordado neste estudo tem características híbridas, com o emprego de dados numéricos medidos por equipamentos e a aplicação de uma técnica inteligente para realizar as previsões acerca do potencial de geração de energia do local estabelecido.

O banco de dados meteorológicos contém medições dos meses de setembro a dezembro de 2021, em uma escala diária de dados armazenados a cada 10 minutos. Para este modelo proposto foram verificadas três métricas que possuem mais aderência para realizar as previsões de geração: velocidade, direção e altura.

Figura 5.2: Metodologia



Fonte: Elaborada pelo Autor

Na Figura 5.2 é ilustrado o fluxograma com as etapas de interação do algoritmo de previsão. Estas etapas incluem a aquisição e tratamento dos dados de entrada, a aplicação do *Mean Shift* para a geração dos cenários de previsão e a plotagem dos gráficos para análise.

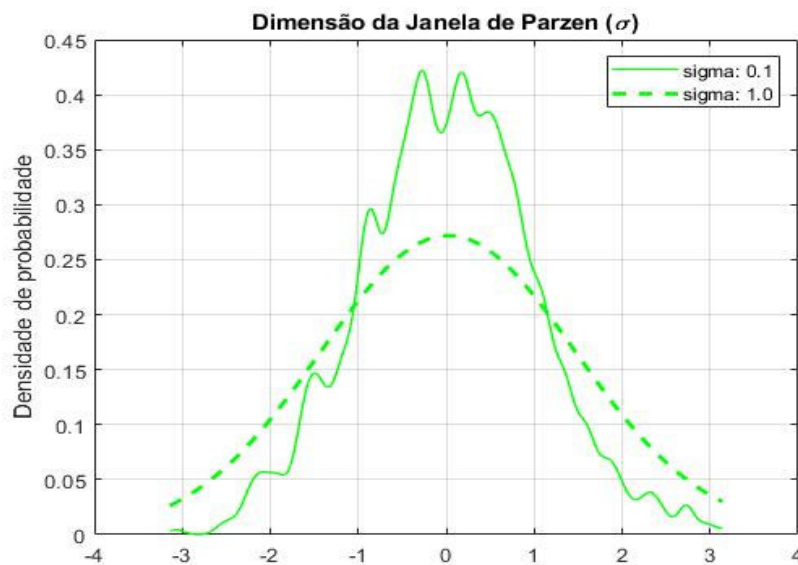
O processo tem início com a obtenção do banco de dados que contém dois meses de medições, seguido da filtragem dessas medições, buscando identificar algum erro dentro da série contida no banco de dados. Após esses primeiros passos, é realizada a escolha das variáveis, que são mais determinantes para a pesquisa e a delimitação do σ .

Conforme mostrado no fluxograma na Figura 5.2, o processo segue para aplicação do *Mean Shift*, densificação do conjunto e em seguida realizado o cálculo de erro, os dados com erros $> 10\%$ é realizado a plotagem dos gráficos, em casos de erros superiores ao estipulado ocorre a incrementação do σ e do λ , fazendo com que esses dados retornem para um novo processamento pelo algoritmo.

Uma vez que a previsão foi adquirida derivada do conjunto de cenários gerados pelas densificação do conjunto original e obtendo um novo conjunto densificado, são plotados os gráficos para a análise. As previsões obtidas são projetadas em escalas de curto e médio prazo, podendo variar em um horizonte de 1 minuto a 1 mês.

Definir o valor inicial do σ é necessário para estimar a densidade baseada na janelas de Parzen com base na aplicação da função kernel, no qual fará a identificação de um ponto local com o conjunto contínuo derivado da FDP discriminada pelo kernel (Figura 5.3).

Figura 5.3: *Dimensão da Janela de Parzen*



Fonte: Adaptado de [41]

Ainda na Figura 5.3 o σ foi estipulado entre os parâmetros 0.1 e 1.0, no qual quanto mais picos tem a curva o σ se aproxima de 0.1 e quanto mais suave a curva se torna o σ mais se aproxima de 1.0. Logo, com o uso do kernel gaussiano os pesos são os que estão mais próximos de x_i , havendo uma suavização da estimativa da densidade.

Entendendo que ao estimar a janela de Parzen o σ não pode assumir um valor

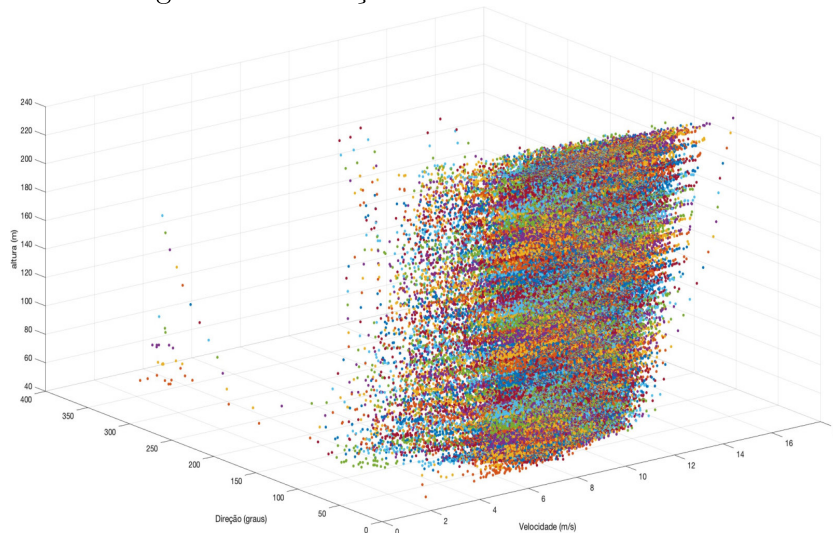
muito baixo, visto a falta de interação entre os pontos estabelecidos pelo kernel, o que irá gerar uma função cheia de modas é portanto diferente da estimação da FDP. O σ caso assuma um valor muito alto compromete a estimação dos cenários em decorrência da perda de informação, em razão da extinção dos detalhes.

Portanto, para atender a proposta desse trabalho, cuja função do *Mean Shift* permite densificar a amostra de dados iniciais, os valores de σ utilizados permitiram a criação dos cenários com a estimação da FDP até atingir a moda que é o ponto de parada do algoritmo.

5.2 Análise dos Resultados

Durante a construção do modelo de previsão e os dados de saída para análise, foi definido os parâmetro de velocidade, direção e altura, com o objetivo de plotar os gráficos em três dimensões, permitindo uma visão destas variáveis com um único ponto no espaço R^3 . Na Figura 5.4 são apresentados a base de dados iniciais, no qual é possível observar a associação das alturas (10 - 10 metros), as velocidades (4 m/s a 14 m/s) e direção (predominância nordeste).

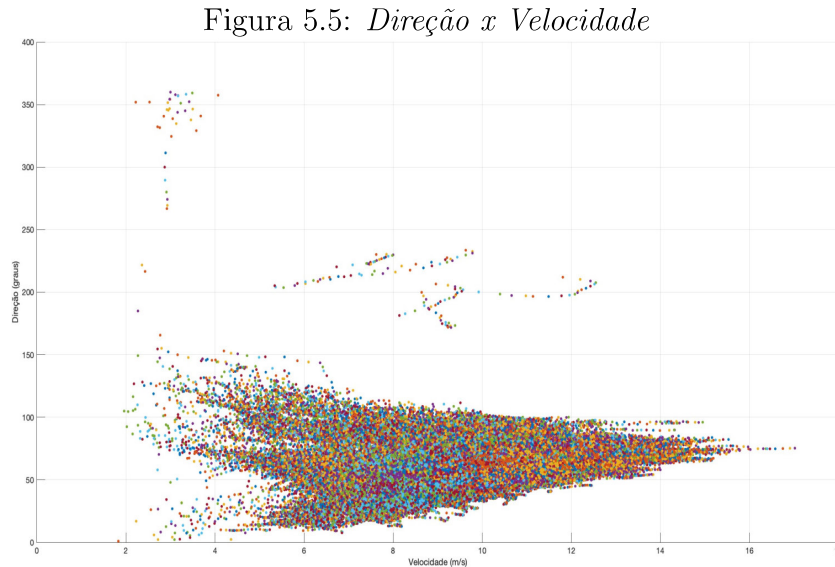
Figura 5.4: *Medição de outubro a novembro*



Fonte: Elaborada pelo Autor

Já na Figura 5.5 foi rotacionado para duas dimensões (direção e velocidade) de forma a poder ilustrar a predominância dos ventos de 7 m/s a 14 m/s e constância no quadrante de 0 - 100 graus referente a direção nordeste. É possível observar que

considerando uma velocidade de referência de 10 m/s - 12 m/s (ideal para aerogeradores, pois reduz a fadiga mecânica e mantém um fluxo de geração sem turbulências e parada da máquina), mas de 95 % do tempo os ventos se mantêm unidirecionais, embora seja possível encontrar alguns *outliers* no quadrante sul - sudeste, momento em que, num caso real, os aerogeradores teriam que parar e mudar de direção, o qual leva a um novo estudo de desvios de geração.



Fonte: Elaborada pelo Autor

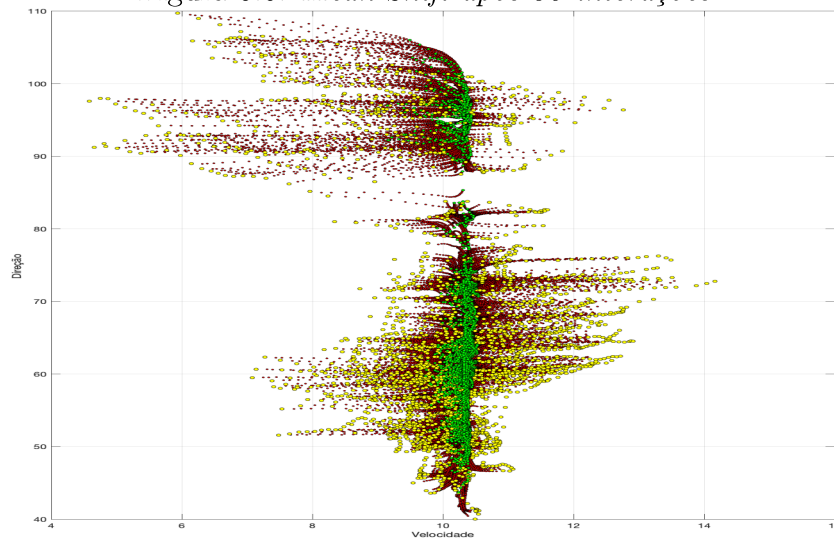
Estes valores de velocidade e direção de vento de referência supracitados, ocorrem basicamente em todas as alturas medidas, ou seja de 50 metros a 280 metros em mais de 90 % dos casos, logo é possível pensar em projetos com turbinas em menor escala, uma vez que o padrão atual (encontrado em parques com 5 a 10 anos de operação) são turbinas acima de 2 MW de potência instalada e torres de até 100 metros.

Na Figura 5.6 é ilustrado um gráfico de velocidade e altura, o que permite comprovar a predominância de ventos com velocidades (10-12 m/s), condições ótimas para geração em todas as alturas. O gráfico ilustrado na Figura 5.6 se refere a aplicação do algoritmo *Mean Shift* na base de dados de direção, velocidade e altura, no qual em amarelo são os dados originais que em cada interação do σ geram pontos virtuais (em vermelho) com a mesma FDP do arquivo original.

Com a variação do σ em cada iteração, estes dados virtuais (em vermelho) vão descrevendo uma trajetória de pontos até atingir a moda ($\sigma = 1.0$) que são representados pelos pontos em verde. Esta trajetória de pontos como são FDPs da FDP do conjunto

principal (pontos em amarelo) são cenários de previsão de velocidade, direção e altura que serão convertidos em previsão de geração (potência e/ou energia).

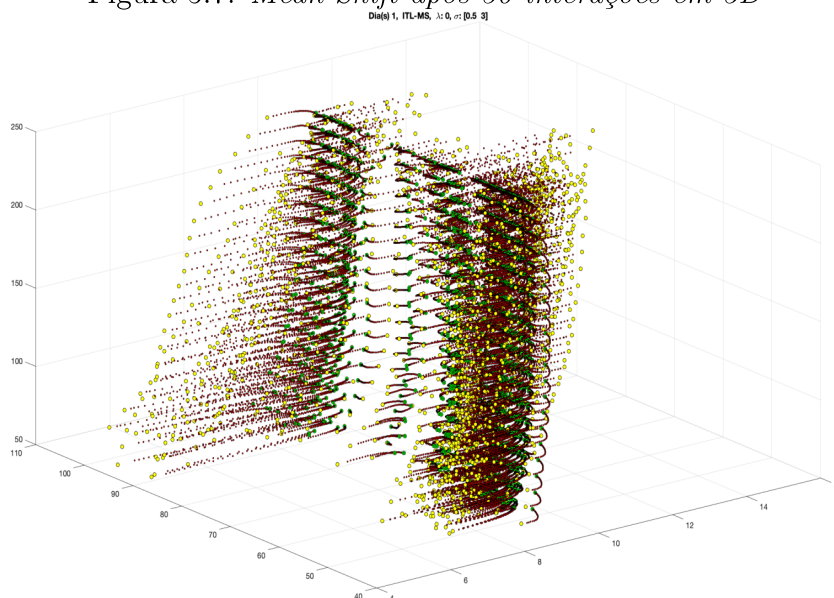
Figura 5.6: *Mean Shift* após 30 interações



Fonte: Elaborada pelo Autor

Uma vez que o algoritmo *Mean Shift* foi baseado em [41] além da variação do σ existe a componente λ que faz o *threshold* que permite ir em direção a moda (atraídos para a região de maior densidade do conjunto) ou em direção oposta, expandindo o conjunto em novas FDPs. Na Figura 5.7 a projeção dos pontos virtuais em cada iteração do σ e a trajetória em 3D.

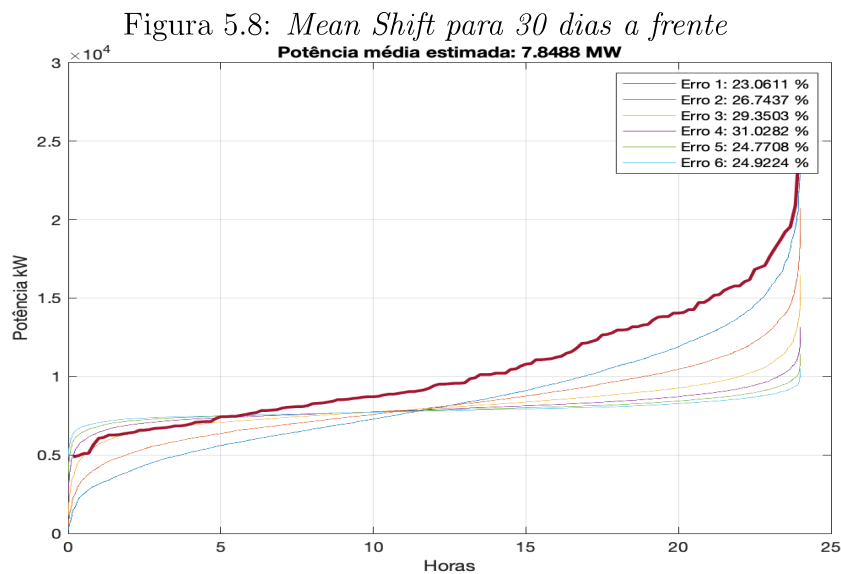
Figura 5.7: *Mean Shift* após 30 interações em 3D



Fonte: Elaborada pelo Autor

A partir dos resultados obtidos, ou seja os novos cenários de velocidade e direção do vento nas alturas pré-definidas (50-280 metros) foi realizado o cálculo da potência prevista para os horizontes de curta e média duração. Esta revisão de geração foi baseada em dados das turbinas existentes nos municípios de Barreirinhas/MA e Paulino Neves/MA, considerando diâmetro do rotor de 70 metros para cálculo da área de varredura das pás de uma turbina e a altura padrão de 100 metros.

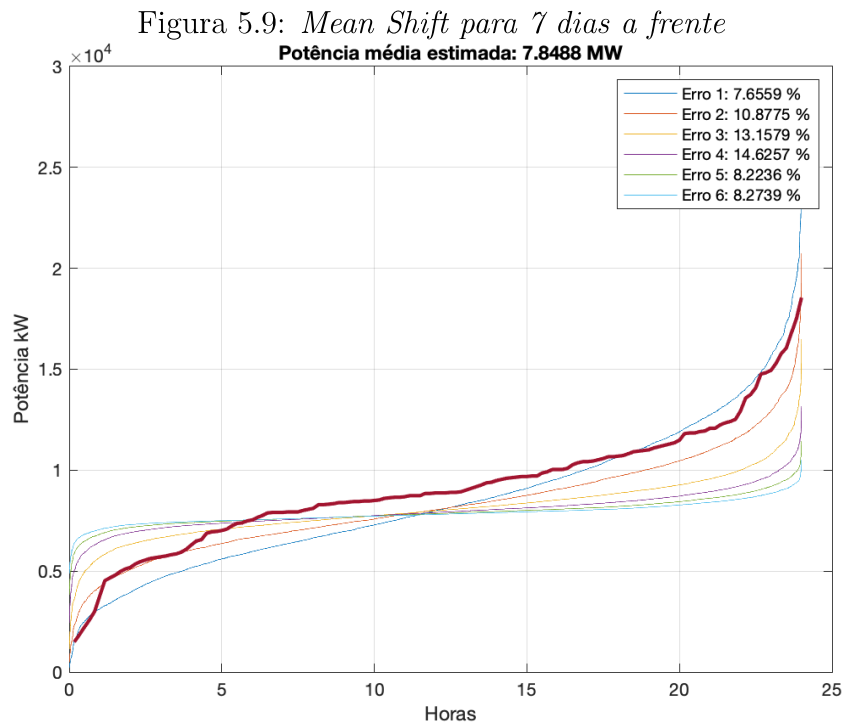
Na Figura 5.8 é ilustrado um dos gráficos de saída de análise da previsão de geração, no qual foram ordenados em ordem crescente a potência que poderia ser gerada ao longo de um dia (24 horas). O gráfico foi elaborado para a previsão de uma geração 30 dias a frente, sendo plotado os principais cenários e o erro de previsão.



Fonte: Elaborada pelo Autor

Ainda na Figura 5.8 a curva em vermelho representa o valor real da potência extraível do vento que poderia ser convertida em energia e as demais curvas os possíveis cenários de geração. Para o caso em destaque, da previsão 30 dias a frente (1 mês), o erro médio quadrático de 23 % em relação ao cenário mais próximo.

Já o segundo gráfico ilustrado por meio da Figura 5.9, representa os cenários projetados para sete dias a frente (1 semana). Neste segundo gráfico a curva vermelha também representa a informação original para aquele sétimo dia, e as demais curvas os cenários possíveis. Neste caso o erro calculado foi de 7.6 % em relação a potência real extraível para aquele dia.

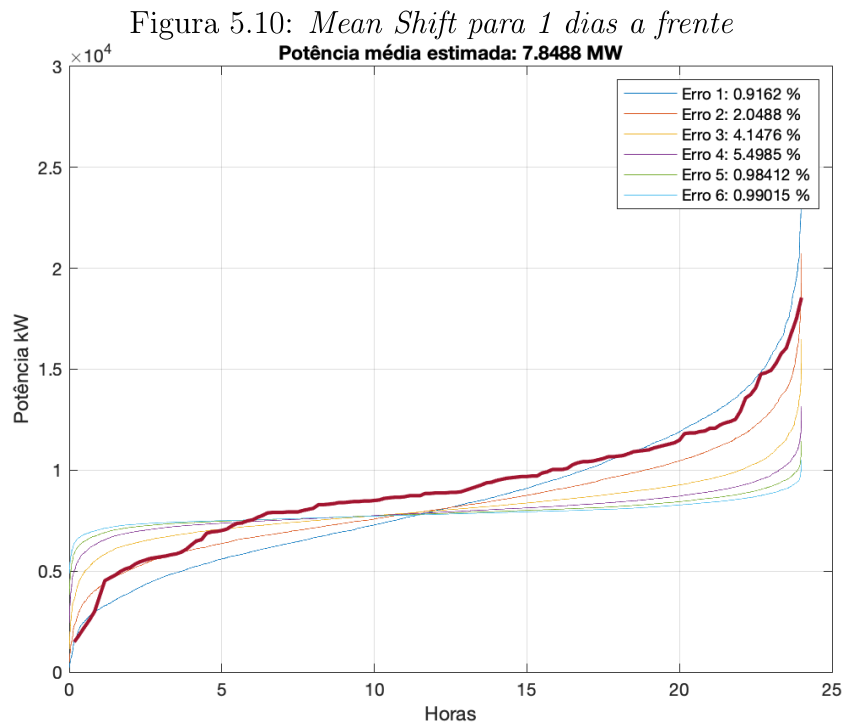


Fonte: Elaborada pelo Autor

Por fim, na Figura 5.10 é ilustrado a simulação de precisão de curta duração para 24 horas a frente (1 dia). Neste cenário o erro previsto entre medição esperada e estimada foi de 0.91 %, ou seja para previsões de curto prazo, hora a hora os cenários gerados de previsão baseados no algoritmo *Mean Shift* apresenta um melhor desempenho.

Entre os 40 testes realizados para cada cenários de previsão de 1 dia, 1 semana e 1 mês, os cenários diários foram os que apresentaram um melhor desempenho, isto se deve ao fato do refinamento realizado pelo algoritmo maximizando a entropia, ou seja a informação predominante, enquanto que isto não ocorre para os horizontes de previsão de médio e longo prazo, havendo perda de informação no processo de geração dos cenários de previsão.

Com uma base de dados completa, o que significa pelo menos uns 5 anos de medição e que possa contemplar os fenômenos meteorológicos como *El Niño* e *La Niña*, o modelo de previsão baseado na teoria da informação *Mean Shift* poderá apresentar um grau de refinamento muito maior para previsões de médio prazo. Contudo, baseado nas informações obtidas, os resultados também se mostraram satisfatórios.

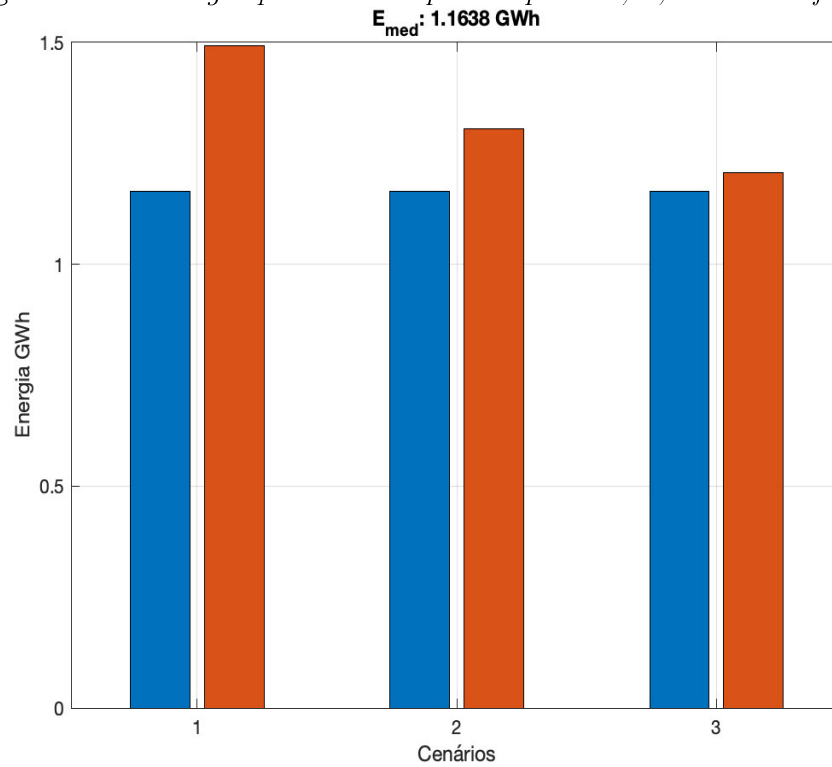


Fonte: Elaborada pelo Autor

Para testes de previsão do algoritmo foi definido os cenários de geração, considerando a curva de demanda para o dia, logo o que se fez foi gerar cenários que tentavam emular a curva de demanda real, o qual é positivo como ferramenta de gestão da operação para tomada de decisão. Todavia para questões relacionadas a contratos de fornecimento os cenários interessantes para uma empresa são os de energia, uma vez que o não fornecimento da demanda de energia contratada incide multas aplicadas pela Agência Nacional de Energia Elétrica - ANEEL, demandadas pelo Operador Nacional de Sistemas - ONS.

Em função deste entendimento foi gerado um gráfico, ilustrado por meio da Figura 5.11 que a partir dos cenários gerados para 1 dia, 1 semana e 1 mês a frente (Figura 5.8, Figura 5.9 e Figura 5.10) foi calculado a energia esperada (cenário real) e a energia estimada (cenário virtual). Também neste caso foram considerados os erros médios quadráticos entre o esperado e o estimado.

Os resultados plotados na forma de gráfico de barras ilustram da esquerda para a direita a previsão de energia para 1 dia, 1 semana e 1 mês a frente respectivamente. A energia média esperada seria de 1.16 GWh nesta simulação, sendo que o resultado da integral da potência foi bem superior para 1 dia em relação aos demais cenários, isto se deve ao refinamento do cenário para 1 dia a frente, o qual possui muito mais informação densificada que os demais e isto aparece quando se calcula a potência no tempo.

Figura 5.11: *Energia prevista e esperada para 1 , 7, 30 dias a frente*

Fonte: Elaborada pelo Autor

Logo, para os cenários de energia a previsão é mais eficaz em todos os horizontes permitindo uma gestão assertiva, uma vez que pelo menos a demanda de energia contratada será gerada pela empresa.

6 CONCLUSÃO

Com um elevado potencial na produção de energia eólica, com perspectiva de em 2023 atingir 23 GW de potência instalada em parques eólicos, o Brasil tem alcançado um crescimento geométrico a cada ano, com instalações de novos parques e aumento da potência instalada. É importante observar que grande parte destes parques estão sendo projetados e instalados na região nordeste, em uma área que vai desde Sergipe ao Maranhão e cuja predominância de ventos são alísios vindos do oceano.

No Maranhão também tem ocorrido um aumento nos parques instalados, inicialmente 228 MW em 2017 e um incremento para 426 MW em 2019, com perspectivas de chegar a 1 GW nos próximos anos. O financiamento de projetos como o EOSOLAR liderado pelo IEE/UFMA é prova que há uma grande necessidade em conhecer o recurso eólico e poder explorar este potencial de geração de energia renovável.

Em vista disso, é importante a realização de análises técnicas que permitam viabilizar a geração de energia eólica, com as medições dos fatores que influenciam na velocidade do vento e que podem comprometer a geração de energia. Isto é importante para se gerar modelos de previsão eólica confiáveis, com alta taxa de assertividade e que auxiliem na tomada de decisão de operação e manutenção.

Logo, foi proposto por meio desta dissertação de mestrado a aplicação de uma técnica baseada em teoria da informação *Mean Shift* para a densificação de uma base de dados de medições de vento com o objetivo de realizar a previsão de ventos e consequentemente geração de energia eólica, por meio de uma abordagem híbrida, em que foram utilizados um conjunto de séries temporais, adquiridas em uma base de dados, com medições entre 50 m a 280 metros (a cada 10 metros, por meio de perfiladores).

Os resultados obtidos demonstraram que o melhor desempenho do algoritmo ocorreu nos cenários diários, devido ao fato do refinamento realizado pelo algoritmo ao maximizar a entropia, não ocorrendo este fato quando o horizonte de previsão é mais longo, havendo perda de informação no processo de geração dos cenários de previsão. Assim, foi possível implementar o algoritmo *Mean Shift* como ferramenta previsão de ventos de curto prazo, com a possibilidade de aprimoramento do algoritmo a partir de um banco de dados

que contemple medições com um período mais longo, bem como apresenta facilidade de ser integrado em plataformas de empresas como ferramenta de gestão da geração.

6.1 Trabalhos Futuros

Com o desenvolvimento do trabalho de dissertação de mestrado e da elaboração do texto, foi observado dois pontos que podem ser explorados em trabalhos futuros:

- Se os ventos são quase constantes em velocidades de 10-12 m/s adequadas para geração eólica na maioria das alturas estudadas, o uso de turbinas de menor estatura e conseqüentemente custos menores de operação e manutenção poderiam ser o mais adequado; e
- Em que ponto de operação há perdas de desvios de geração quando as turbinas mudam de direção (nordeste para sudeste) e quanto isto impacta financeiramente no ativo. O MW gerado quando muda de direção por poucas horas cobre esta logística ou seria melhor manter a turbina fixa na direção predominante.

6.2 Publicações

Foram enviados algumas propostas de artigos para eventos nacionais e periódicos de forma a validar pelos pares a proposta de uso do *Mean Shift* para previsão, uma vez que já foi utilizado para densificação de bases de dados e utiliza a FDP como elemento de suporte.

- Foi enviado um artigo para o IX Simpósio Brasileiro de Sistemas Elétricos (SBSE 2022), que deverá ocorrer entre 10 a 13 de julho de 2022, o qual aguarda avaliação; e
- Foi iniciado um artigo para o periódico do IEEE *Power Delivery* que deverá ser submetido até meados de março de 2022.

REFERÊNCIAS

- [1] MOREIRA, José Roberto Simões. Energias renováveis, geração distribuída e eficiência energética. Rio de Janeiro: LTC, 2017.
- [2] ROSA, Aldo. Processos de Energia Renováveis. Elsevier Brasil, 2016.
- [3] DOS REIS, Lineu Belico. Geração de energia elétrica. Editora Manole, 2017.
- [4] SILVA, EP da. Fontes renováveis de energia: produção de energia para um desenvolvimento sustentável. São Paulo: Livraria da Física, p. 90, 2014.
- [5] Boletim Anual de Geração Eólica 2020[Online]. Availabe: http://abeeolica.org.br/formato_docs/boletim_-_anual/. [Acesso em 10 de dezembro 2021].
- [6] BRASIL. Ministério de Minas e Energia, Plano Nacional de Energia 2030 / Ministério de Minas e Energia ; colaboração Empresa de Pesquisa Energética. Brasília : MME : EPE, 2007 [Online] Availabe: <https://www.epe.gov.br/pt/publicacoes-dados-abertos/publicacoes/Plano-Nacional-de-Energia-PNE-2030> . [Acesso em 10 de dezembro de 2021]
- [7] AMARANTE, O. A.; BROWER, M.; ZACK, J.; EOLICA, C. S. E.; SOLUTIONS, T. Atlas do potencial eólico brasileiro. 2001 Availabe: http://www.cresesb.cepel.br/publicacoes/download/atlas_eolico/Atlas%20do%20Potencial%20Eolico%20Brasileiro.pdf [Acesso em : 20 de outubro de 2021]
- [8] ALENCAR, David Barbosa de. Modelo híbrido baseado em series temporais e redes neurais para previsão da geração eolica. Tese (Doutorado). Universidade Federal do Pará, Belém:2018. [Online] Availabe: <http://repositorio.ufpa.br:8080/jspui/handle/2011/10416>. [Acesso em: 20 de outubro de 2021]
- [9] EPE - Empresa de Pesquisa Energética. 2017. Availabe: <http://www.epe.gov.br> [Acesso em: 10 de outubro de 2021]

- [10] WANG, Huai-zhi; Li, Gang-qiang; Wang, Gui-bin; Peng, Jian-chun; Jiang, Hui Ji-ang; Liu, Yi-tao. Deep learning based ensemble approach for probabilistic wind power forecasting. *Applied Energy*, Volume 188, 2017, Pages 56-70, 0306-2619. Available: <https://doi.org/10.1016/j.apenergy.2016.11.111>. [Acesso em: 10 de setembro de 2021]
- [11] LENG, H., Li, X., Zhu, J., Tang, H., Zhang, Z., Ghadimi, N. (2018). A new wind power prediction method based on ridgelet transforms, hybrid feature selection and closed-loop forecasting. *Advanced Engineering Informatics*, 36, 20-30. Available: <https://doi.org/10.1016/j.aei.2018.02.006>. [Acesso em: 10 de setembro de 2021]
- [12] SANTAMARÍA-BONFIL, G., Reyes-Ballesteros, A., Gershenson, C. J. R. E. (2016). Wind speed forecasting for wind farms: A method based on support vector regression. *Renewable Energy*, 85, 790-809. Available: <https://doi.org/10.1016/j.renene.2015.07.004>. [Acesso em: 10 de setembro de 2021]
- [13] LIU, Hui; MI, Xiwei; LI, Yanfei. Smart deep learning based wind speed prediction model using wavelet packet decomposition, convolutional neural network and convolutional long short term memory network. *Energy Conversion and Management*, v. 166, p. 120-131, 2018. Available: <https://doi.org/10.1016/j.enconman.2018.04.021>. [Acesso em: 10 de setembro de 2021]
- [14] FENG, Cong et al. A data-driven multi-model methodology with deep feature selection for short-term wind forecasting. *Applied Energy*, v. 190, p. 1245-1257, 2017. Available: <https://doi.org/10.1016/j.apenergy.2017.01.043> [Acesso em: 10 de setembro de 2021]
- [15] CHITSAZ, Hamed; AMJADY, Nima; ZAREIPOUR, Hamidreza. Wind power forecast using wavelet neural network trained by improved Clonal selection algorithm. *Energy conversion and Management*, v. 89, p. 588-598, 2015. Available: <https://doi.org/10.1016/j.enconman.2014.10.001> [Acesso em: 10 de setembro de 2021]
- [16] SHUKUR, Osamah Basheer; LEE, Muhammad Hisyam. Daily wind speed forecasting through hybrid KF-ANN model based on ARIMA. *Renewable Energy*, v. 76, p. 637-647, 2015. Available: <https://doi.org/10.1016/j.renene.2014.11.084>. [Acesso em: 10 de setembro de 2021]

- [17] WANG, Jianzhou; HU, Jianming. A robust combination approach for short-term wind speed forecasting and analysis—Combination of the ARIMA (Autoregressive Integrated Moving Average), ELM (Extreme Learning Machine), SVM (Support Vector Machine) and LSSVM (Least Square SVM) forecasts using a GPR (Gaussian Process Regression) model. *Energy*, v. 93, p. 41-56, 2015. Available: <https://doi.org/10.1016/j.energy.2015.08.045> [Acesso em: 10 de setembro de 2021]
- [18] ZHAO, Jing et al. An improved multi-step forecasting model based on WRF ensembles and creative fuzzy systems for wind speed. *Applied Energy*, v. 162, p. 808-826, 2016. Available: <https://doi.org/10.1016/j.apenergy.2015.10.145> [Acesso em: 10 de setembro de 2021]
- [19] YUAN, Xiaohui et al. Short-term wind power prediction based on LSSVM–GSA model. *Energy Conversion and Management*, v. 101, p. 393-401, 2015. Available: <https://doi.org/10.1016/j.enconman.2015.05.065>. [Acesso em: 10 de setembro de 2021]
- [20] SENATORE, Alfonso et al. A combined modelling system for short-term wind power forecasting based on mesoscale Numerical Weather Prediction. In: 2020 IEEE International Conference on Environment and Electrical Engineering and 2020 IEEE Industrial and Commercial Power Systems Europe (EEEIC/ICPS Europe). IEEE, 2020. p. 1-5. Available: [Acesso em: 10 de setembro de 2021]
- [21] OLIVEIRA, Felipe Maia Barbosa. Previsão da velocidade do vento utilizando redes neurais artificiais e modelos autorregressivos. 2020. Dissertação (Mestrado em Sistemas de Energia) - Universidade Tecnológica Federal do Paraná, Curitiba, 2020. Available: <http://repositorio.utfpr.edu.br/jspui/handle/1/23577> [Acesso em: 28 de novembro de 2021]
- [22] NOAA. Numerical Weather Prediction. 2020. Available: <https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/numerical-weather-prediction>. [Acesso em: 10 de outubro de 2021]
- [23] ABESAMIS, KreigAudrey et al. Short-Term Wind Power Forecasting Using Structured Neural Network. In: 2019 IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment,

- and Management (HNICEM). IEEE. p. 1-4. Available: [Acesso em:10 de setembro de 2021]
- [24] YAN, Jie et al. Forecasting the high penetration of wind power on multiple scales using multi-to-multi mapping. *IEEE Transactions on Power Systems*, v. 33, n. 3, p.3276-3284, 2017. Available: [Acesso em:12 de setembro de 2021]
- [25] LI, Menglin et al. Wind Power Probabilistic Forecasting Based on Wind Correction Using Weather Research and Forecasting Model. In: 2020 IEEE 3rd Student Conference on Electrical Machines and Systems (SCEMS). IEEE, 2020. p. 619-624. Available: [Acesso em:12 de setembro de 2021]
- [26] WU, Sheng-Hong; WU, Yuan-Kang. Probabilistic Wind Power Forecasts Considering Different NWP Models. In: 2020 International Symposium on Computer, Consumer and Control (IS3C). IEEE, 2020. p. 428-431. Available: 10.1109/IS3C50286.2020.00116 [Acesso em:10 de setembro de 2021]
- [27] BRAMS. Brazilian Regional Atmospheric Modeling System. Available: <http://brams.cptec.inpe.br/about/>. [Acesso em:5 de setembro de 2021]
- [28] NCEP. National Centers for Environmental Prediction. The Climate Forecast System (CFS) models the interactions between Earth's oceans, land, and atmosphere on a global scale. Available: <https://www.ncei.noaa.gov/products/weather-climate-models/climate-forecast-system>. [Acesso em: 5 de dezembro de 2021]
- [29] CHANG, G. W. et al. A hybrid model for forecasting wind speed and wind power generation. In: 2016 IEEE Power and Energy Society General Meeting (PESGM). IEEE, 2016. p. 1-5. Available: [Acesso em:15 de setembro de 2021]
- [30] MOHAMMAD Valipour, Mohammad Ebrahim Banihabib, Seyyed Mahmood Reza Behbahani. Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir. *Journal of Hydrology*. Volume 476. 2013. Pages 433-441. ISSN 0022-1694. <https://doi.org/10.1016/j.jhydrol.2012.11.017>. [Acesso em:10 de dezembro de 2021]
- [31] LIMA, Fábio Henrique de Andrade. Previsão de geração eólica de curto prazo em tempo real baseada em dados observacionais das centrais eólicas e da previsão numérica do tempo. Dissertação (Mestrado - Engenharia Mecânica), Universidade Federal

- de Pernambuco:2019. Disponível em: <https://repositorio.ufpe.br/handle/123456789/33809> [Acesso em: 30 de novembro de 2021]
- [32] FRANÇA, Ana Carolyne G and Faceroli, Silvana T and Nascimento, Paulo Sergio C and Kligerman, Alberto Sergio and others. ANALISE COMPARATIVA DE MODELOS DE PREVISÃO DE GERAÇÃO DE ENERGIA EÓLICA BASEADOS EM MACHINE LEARNING. Revista de Sistemas e Computação. vol.9, n. 2, 2020. [Acesso em: 30 de setembro de 2021]
- [33] SOARES, Paulo Henrique. Previsão da geração de energia eólica de curtíssimo prazo aplicada ao Parque Eólico de Palmas (PR). Dissertação (mestrado) - Universidade Federal do Paraná, Setor de Tecnologia, Programa de Pós-Graduação em Engenharia de Recursos Hídricos e Ambiental. Curitiba: 2015. Disponível em: <http://hdl.handle.net/1884/41342> [Acesso em: 5 de setembro de 2021]
- [34] ZHANG, Jiawei et al. Constructing probabilistic load forecast from multiple point forecasts: A bootstrap based approach. In: 2018 IEEE Innovative Smart Grid Technologies-Asia (ISGT Asia). IEEE, 2018. p. 184-189. Disponível em: [Acesso em: 12 de setembro de 2021]
- [35] GEORGE E.P. Box, Gwilym M. Jenkins, Gregory C. Reinsel, Greta M. Ljung, Time series analysis : forecasting and control. Fifth edition, John Wiley Sons: 2016
- [36] HAYKIN, S. Redes neurais princípios prática. Vol. 2. Editora Bookman, Porto Alegre, 2001.
- [37] ARAÚJO, Daniel Sabino Amorim de. Análise de agrupamentos com base na teoria da informação: uma abordagem representativa. Tese (Doutorado) , Programa de Pós- graduação em Engenharia Elétrica e computação da Universidade Federal do Rio Grande do Norte. Natal: 2013. Disponível em: <https://repositorio.ufrn.br/handle/123456789/15208> [Acesso em: 5 de dezembro de 2021]
- [38] MARTINS, Catarina Alexandra Couteiro. Aplicação de Modelos Estatísticos à Previsão de Produção de Energia Elétrica. Dissertação (Mestrado) Programa de Pós-graduação em Engenharia Eletrotécnica e de Computadores. Faculdade de Ciências e Tecnologia, Universidade NOVA de Lisboa. Lisboa: 2019. Disponível em: <http://hdl.handle.net/10362/93421> [Acesso em: 5 de dezembro de 2021]

- [39] PRINCIPE, Jose C. Information theoretic learning: Renyi's entropy and kernel perspectives. Springer Science Business Media, 2010.[Acesso em: 28 de novembro de 2021]
- [40] RAO, Sudhir; DE MEDEIROS MARTINS, Allan; PRÍNCIPE, José C. Mean shift: An information theoretic perspective. Pattern Recognition Letters, v. 30, n. 3, p. 222-230, 2009.Availabe:<https://doi.org/10.1016/j.patrec.2008.09.011> [Acesso em: 28 de novembro de 2021]
- [41] LIMA, Shigeaki Leite de et al. Diagnostico de falhas incipientes em transformadores de potência.Tese (Doutorado), Programa de Pós-graduação em Engenharia de eletricidade da Universidade Federal do Maranhão. São Luís: 2013. [Acesso em: 05 de setembro de 2021]
- [42] PARZEN, Emanuel. On estimation of a probability density function and mode. The annals of mathematical statistics, v. 33, n. 3, p. 1065-1076, 1962.[Acesso em: 05 de setembro de 2021]
- [43] SOUSA. J. C. S. , SAAVEDRA. O. R. and LIMA. S. L. , "Decision Making in Emergency Operation for Power Transformers With Regard to Risks and Interruptible Load Contracts,"in IEEE Transactions on Power Delivery, vol. 33, no. 4, pp. 1556-1564, Aug. 2018, doi: 10.1109/TPWRD.2017.2764466.[Acesso em:5 de dezembro de 2021]
- [44] TUZEL, Oncel; PORIKLI, Fatih; MEER, Peter. Kernel methods for weakly supervised mean shift clustering. In: 2009 IEEE 12th International Conference on Computer Vision. IEEE, 2009. p. 48-55. [Acesso em:5 de dezembro de 2021]
- [45] WU, Kuo-Lung; YANG, Miin-Shen. Mean shift-based clustering. Pattern Recognition, v. 40, n. 11, p. 3035-3052, 2007.Availabe: <https://doi.org/10.1016/j.patcog.2007.02.006> [Acesso em:5 de dezembro de 2021]
- [46] MIRANDA, Vladimiro; CASTRO, Adriana R. Garcez; LIMA, Shigeaki. Diagnosing faults in power transformers with autoassociative neural networks and mean shift. IEEE Transactions on Power Delivery, v. 27, n. 3, p. 1350-1357, 2012.Availabe: [Acesso em:5 de dezembro de 2021]

-
- [47] WU, Tianjun et al. Computationally efficient mean-shift parallel segmentation algorithm for high-resolution remote sensing images. *Journal of the Indian Society of Remote Sensing*, v. 46, n. 11, p. 1805-1814, 2018. Available: <https://doi.org/10.1007/s12524-018-0841-8> [Acesso em:5 de dezembro de 2021]
- [48] COHEN, Yishai; LAPIDOT, Itshak. Speaker clustering quality estimation with logistic regression. *Computer Speech Language*, v. 65, p. 101139, 2021. Available: <https://doi.org/10.1016/j.csl.2020.101139> [Acesso em:10 de dezembro de 2021]