

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

SAMARA DEON

**PREVISÃO DA GERAÇÃO DE ENERGIA EÓLICA: UMA ABORDAGEM
UTILIZANDO TÉCNICAS DE COMBINAÇÃO DE MODELOS DE PREVISÃO**

PATO BRANCO

2022

SAMARA DEON

**PREVISÃO DA GERAÇÃO DE ENERGIA EÓLICA: UMA ABORDAGEM
UTILIZANDO TÉCNICAS DE COMBINAÇÃO DE MODELOS DE PREVISÃO**

**WIND POWER FORECASTING: AN APPROACH USING ENSEMBLE LEARNING
MODELS**

Trabalho de conclusão de curso de graduação
apresentado como requisito para obtenção do título de
Licenciado em Matemática pela Universidade
Tecnológica Federal do Paraná (UTFPR).

Orientador(a): Prof. Dr. Matheus Henrique Dal Molin
Ribeiro.

Coorientador(a): Marlova Estela Caldato.

PATO BRANCO

2022



[4.0 Internacional](https://creativecommons.org/licenses/by/4.0/)

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

SAMARA DEON

**PREVISÃO DA GERAÇÃO DE ENERGIA EÓLICA: UMA ABORDAGEM
UTILIZANDO TÉCNICAS DE COMBINAÇÃO DE MODELOS DE PREVISÃO**

Trabalho de Conclusão de Curso de Graduação
apresentado como requisito para obtenção do título de
Licenciado em Matemática pela Universidade
Tecnológica Federal do Paraná (UTFPR).

Data de aprovação: 17/dezembro/2022

Matheus Henrique Dal Molin Ribeiro
Doutorado
Universidade Tecnológica Federal do Paraná

José Donizetti de Lima
Doutorado
Universidade Tecnológica Federal do Paraná

Géremi Gilson Dranka
Doutorado
Universidade Tecnológica Federal do Paraná

PATO BRANCO

2022

Dedico este trabalho à minha família, que me apoiou em tudo que precisei para conseguir concluí-lo.

AGRADECIMENTOS

Primeiramente, agradeço à Deus por ter me dado o dom da vida e a oportunidade de estudar, além de uma família que em tudo me apoiou para o sucesso desse trabalho.

Agradeço ao meu orientador Prof. Dr. Matheus Henrique Dal Molin Ribeiro, que foi um excelente orientador do início ao fim dessa jornada, sempre soube esclarecer o caminho a ser percorrido, esteve disponível para auxiliar em tudo que necessitei e por tornar esse trabalho muito agradável de ser realizado.

Agradeço também à Prof^a Dra. Marlova Estela Caldato que coorientou esse trabalho e explica sobre projetos de pesquisa como ninguém.

Agradeço à Secretaria do Curso, pela cooperação e aos meus colegas de sala pelo suporte mútuo em momentos de aflição.

Gostaria de deixar registrado também, o meu reconhecimento à minha família, pois o apoio deles tornou tudo isso possível.

Enfim, a todos os que por algum motivo contribuíram para a realização desta pesquisa.

RESUMO

A energia eólica possui um expressivo papel na sociedade, pois ela representa 11% da matriz energética no Brasil. Além disso, detém um importante caráter ambiental e econômico. Nesse contexto, obter a previsão de energia de fonte eólica é importante para realizar tomadas de decisões no que se refere à exploração dessa energia e para o planejamento de operações dos sistemas elétricos de potência, visando o aproveitamento de recursos energéticos. Porém, devido às características de séries temporais relacionadas à produção de energia eólica como intermitência e alteração no padrão dos dados devido a fatores externos, como clima e características geográficas dos parques eólicos, o desenvolvimento de modelos estatísticos no contexto de previsões de séries temporais torna-se uma tarefa desafiadora. Assim, nesse trabalho quarenta e dois comitês de máquina também conhecidos como *ensembles learning models*, foram propostos para a previsão de séries temporais de energia eólica. Os *ensembles* foram gerados a partir de diferentes modelos base combinados por meio de: média aritmética, média harmônica, mediana e média ponderada (com pesos definidos por intermédio das métricas raiz quadrada do erro médio, erro médio absoluto, e erro absoluto médio percentual). Os *ensembles* foram aplicados a fim de gerar a previsão de série temporal de energia eólica para horizontes de previsão de 10, 30, 60 e 120 minutos para dois parques eólicos, localizados na Bahia, Brasil. Os resultados mostram que a combinação de previsões de modelos individuais realizadas através de média ponderada com pesos definidos a partir dos erros absolutos médios percentuais foi considerada a com melhor desempenho em termos de menor erro percentual em 87,5% dentre as combinações realizadas. Enquanto que a combinação realizada por meio da média harmônica apresentou maior erro em relação às outras em 75% das situações. O modelo que pondera (com pesos definidos usando os erros absolutos percentuais médios) as previsões das abordagens máquina de vetores suporte, máquinas de aprendizado extremo, processos Gaussianos e regressão com regularização foi considerado com o melhor desempenho em termos de menor erro em 62,5% das combinações.

Palavras-chave: Geração de energia eólica; Previsão de séries temporais; *ensembles*.

ABSTRACT

Wind power has a large role in society, as it represents 11% of the energy matrix in Brazil. Furthermore, has an important environmental and economic character. In this sense, obtaining wind power forecast it is important for making decisions regarding to exploration of this kind of energy and for operations planning, aiming the use of energy resources. However, due to the characteristics of the series related to wind power generation, such as intermittence and non-standard external data, such as the climate and the geographical characteristics of wind farms, the development of mathematical models in the context of a series of temporal factors becomes a difficult task. Thus, in this work forty-two learning model ensembles were proposed for wind power time series forecast. The ensembles were generated from different base models combined by means of arithmetic mean, harmonic mean, median and weighted mean (with weights defined by Root Mean Squared Error, Mean Absolute Error and Mean Absolute Percentage Error). The ensembles were applied in order to generate a wind power time series forecast for the forecast horizons of 10, 30, 60 and 120 minutes for two wind farms, located in Bahia, Brazil. The results show that the combination of forecasts of individual models performed through weighted average with weights defined from the Mean Absolute Percentage Error was considered the one with the best performance in terms of the lowest percentage error in 87.5% among the combinations performed. While the combination performed using the harmonic mean presented a larger error in relation to the others in 75% of the situations. The model that weights (with weights defined using the Mean Absolute Percentage Error) the predictions of the support vector machine, extreme learning machines, Gaussian processes and regression with regularization approaches was considered to have the best performance in terms of the smallest error at 62.5 % of ratings.

Keywords: Wind power generation; Time series forecasting; ensembles.

LISTA DE ILUSTRAÇÕES

Figura 1 - Evolução da capacidade instalada da energia eólica em MW no Brasil	18
Figura 2 - Etapas de análise e previsão de séries temporais	20
Figura 3 - Diagrama combinação de métodos de previsão	25
Figura 4 - Produção de energia eólica de acordo com cada WDF	28
Figura 5 - Valores ACF e PACF para as amostras	29
Figura 6 - Diagrama combinação de métodos de previsão	31
Figura 7 - Parte da série temporal dos valores observados de energia eólica produzida em WDF1	33
Figura 8 - Parte da série temporal dos valores observados de energia eólica produzida em WDF2	33
Figura 9 - Nomenclatura ensembles	35
Figura 10 - Valores observados e previstos em WDF1: horizonte de previsão 10 min, ensemble 2Comb1d	41
Figura 11 - Valores observados e previstos em WDF1: horizonte de previsão de 30 min, ensemble 3Comb4f	41
Figura 12 - Valores observados e previstos em WDF1: horizonte de previsão de 60 min, ensemble 4Comb6f.....	42
Figura 13 - Valores observados e previstos em WDF1: horizonte de previsão de 120 min, ensemble 4Comb6f	42
Figura 14 - Valores observados e previstos em WDF2: horizonte de previsão de 10 min, ensemble 4Comb6f	44
Figura 15 - Valores observados e previstos em WDF2: horizonte de previsão de 30 min, ensemble 3Comb4f	44
Figura 16 - Valores observados e previstos em WDF2: horizonte de previsão de 60 min, ensemble 4Comb6f	45
Figura 17 - Valores observados e previstos em WDF2: horizonte de previsão de 120 min, ensemble 4Comb6f	45

LISTA DE TABELAS

Tabela 1 - Revisão dos modelos de previsão utilizados para a previsão de energia eólica	24
Tabela 2 - Medidas estatísticas para cada série temporal	29
Tabela 3 - Métricas RMSE, MAE e MAPE das previsões geradas a partir de dois e três modelos base	36
Tabela 4 - Métricas RMSE, MAE e MAPE das previsões geradas a partir de quatro e cinco modelos base	37
Tabela 5 – Modelos que apresentaram melhor desempenho em termos de menor erro em cada categoria	40

LISTA DE ABREVIATURAS E SIGLAS

ABEEólica	Associação Brasileira de Energia Eólica
ANN	Redes Neurais Artificiais (do inglês, <i>Artificial Neural Networks</i>)
ARIMA	Média Móvel Integrado Autorregressivo (do inglês, <i>Autoregressive Integrated Moving Average</i>)
ARIMAX	AutoRegressivo Integrado de Média Móvel com Entradas Exógenas (do inglês, <i>Auto-Regressive Integrated Moving Average with eXogenous inputs</i>)
CEEMD	Decomposição em Modo Empírico de Conjunto Completo (do inglês, <i>Complete Ensemble Empirical Mode Decomposition</i>)
CPSO–CM	<i>Chaos Particles Warm Optimization Algorithm For Combination Method</i>
ELM	Máquinas de Aprendizado Extremo (do inglês, <i>Extreme Learning Machines</i>)
GFS	Sistema de Previsão Global (do inglês, <i>Global Forecast System</i>)
GP	Processo Gaussiano (do inglês, <i>Gaussian Process</i>)
GPR	Regressão do Processo Gaussianico (do inglês, <i>Gaussian process regression</i> , GPR).
GWEC	Conselho global de energia eólica (do inglês, <i>Global Wind Energy Council</i>)
HW	<i>Holt-Winters</i>
IA	Inteligência Artificial
IDHM	Índice de Desenvolvimento Humano Municipal
I-REC	Certificado Internacional de Energia Renovável (do inglês, <i>International Renewable Energy Certificate</i>)
MAE	Erro Absoluto Médio (do inglês, <i>Mean Absolute Error</i>)
MAPE	Erro Percentual Médio Absoluto (do inglês, <i>Mean Absolute Percentage Error</i>)
ML	Aprendizado de máquina (do inglês, <i>Machine learning</i>)
NRMSE	Raiz do Erro Quadrático Médio Normalizado (do inglês, <i>Normalized Root Mean Square Error</i>)
PIB	Produto Interno Bruto
RF	Floresta Aleatória (do inglês, <i>Random Forest</i>)
RMSE	Raiz do Erro Quadrático Médio (do inglês, <i>Root Mean Squared Error</i>)
RR	Regressão de Ridge (do inglês, <i>Ridge Regression</i>)
SVM	Máquina de Vetor de Suporte (do inglês, <i>Support Vector Machine</i>)
SVR	Regressão de Vetor Suporte (do inglês, <i>Support Vector Regression</i>)
U_2	Coefficiente Estatístico de Theil U2

SUMÁRIO

1	INTRODUÇÃO	13
1.1	PROBLEMATIZAÇÃO	16
1.2	JUSTIFICATIVA.....	17
1.3	Questão Investigativa	19
1.4	Objetivos	19
1.4.1	Objetivo Geral.....	19
1.4.2	Objetivos Específicos	19
2	FUNDAMENTAÇÃO TEÓRICA	20
2.1	Série temporal.....	20
2.2	<i>Machine learning</i>	21
2.3	Previsão de Geração Energia Eólica.	22
2.4	Combinação de técnicas de previsão.....	24
2.5	Estratégias para combinação de métodos de previsão.....	26
2.6	Medidas de avaliação dos modelos.....	27
3	METODOLOGIA DA PESQUISA.....	28
3.1	Fonte dos dados.....	28
3.2	Procedimentos de coleta, tratamento e análise dos dados.....	30
4	APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS	32
4.1	Análise exploratória dos dados	32
4.2	Nomenclatura dos <i>ensembles</i> gerados.....	34
4.3	Comparação entre os modelos propostos para o WDF1 e para o WDF2	35
4.4	Análise gráfica	40
4.5	Comparação com as técnicas individuais	46
4.6	Comparação entre as previsões obtidas com o trabalho de referência	47
5	CONCLUSÃO	50
	REFERÊNCIAS.....	52

1 INTRODUÇÃO

De acordo com Vieira (2021), o crescimento econômico exige dos países uma produção massiva de produtos, e tal produção está amparada pelo uso intensivo de energia. Dentre essas energias estão as fósseis e as renováveis. A primeira categoria engloba as energias consideradas "suja", as quais são provenientes da lenha, do carvão, do petróleo e de outros, e que provocam emissão de poluentes e colaboram para o aquecimento global. A segunda, compõem as energias renováveis, as quais são consideradas energias "limpas" devido a serem obtidas de fontes que não geram grandes impactos ambientais negativos e não podem ser esgotadas.

Levando isso em consideração, de acordo com Suárez-Cetrulo (2022), e Jin *et al.* (2021), a energia renovável está em ascensão devido às vantagens de seu caráter sustentável e menos polutivo. Ademais, de acordo com Ribeiro *et al.* (2022) as empresas também estão voltando seu olhar para o uso de energias renováveis por consequência do interesse pelo Certificado Internacional de Energias Renováveis (do inglês, *International Renewable Energy Certificate Standard, I-REC Standard*), os quais são úteis para operações globais visto sua capacidade de compra e venda.

Dentre as energias renováveis, destacam-se a energia solar fotovoltaica (proveniente do calor e luz do sol), energia eólica (vinda da força dos ventos), energia hídrica (oriunda da força dos rios), energia de biomassa (procedente da matéria orgânica), energia geotérmica (obtida por meio do calor do interior da Terra) e energia do mar (vinda das marés e ondas) (Nascimento, 2016). A eólica é uma das energias renováveis que ganha destaque no que se refere a produção de energia elétrica (Martins, 2008). Ela converte a energia cinética dos ventos em eletricidade, e "vem sendo apontada como a fonte de energia renovável mais promissora para a produção de eletricidade, em curto prazo, considerando aspectos de segurança energética, custo socioambiental e viabilidade econômica" (MARTINS; v. 30, p. 1304-2, 2008), e será objeto de investigação nesse estudo.

Diversos estudos tem sido realizados no âmbito das energias renováveis, como, impacto e ganhos ambientais (De Castro, 2019), impacto social, custo benefício, efeito no crescimento da renda nacional (Presotto, 2021), análises de oferta e consumo de energias renováveis (Vieira, 2021), e previsão de energia eólica e solar (Andrade, 2021; Gebin, 2020). Algumas das ferramentas utilizadas nesses estudos

são de origem matemática e estatística, como é o caso dos estudos sobre análise e previsão de séries.

Devido a importância da previsão de séries temporais de energias renováveis, muita atenção tem sido dada a esse campo de pesquisa. Junior *et al.* (2018) projetaram um modelo combinado com base nas técnicas Holt-Winters (HW) e de redes neurais artificiais (do inglês, *Artificial Neural Networks*, ANN) para previsão de velocidade do vento com horizonte de previsão mensal, doze passos à frente ou doze meses à frente (longo prazo) e de um dia (curto prazo). Para a previsão mensal, o modelo projetado obteve erros de aproximadamente 7,7%, 0,45 m/s e 0,32 m/s, em termos das métricas de erro percentual médio absoluto (do inglês, *Mean Absolute Percentage Error*, MAPE), raiz do erro quadrático médio (do inglês, *Root Mean Squared Error*, RMSE) e erro absoluto médio (do inglês, *mean absolute error*, MAE), respectivamente. Isto significa que a previsão obtida pode apresentar erro de 7,7% para mais ou para menos em termos da métrica MAPE e de 0,45 m/s e 0,32 m/s para mais ou para menos em termos de RMSE e MAE, respectivamente. O modelo proposto foi comparado com a técnica individual HW, e concluindo-se que o modelo projetado é eficiente, proporcionando maior precisão da série temporal prevista e sendo capaz de representar características sazonais da velocidade do vento na região de estudo.

Ademais, Yong *et al.* (2019) desenvolveram um modelo *ensemble* combinando as técnicas de rede neural de retropropagação (do inglês, *backpropagation*, BP) e de máquina de vetor de suporte (do inglês, *support vector machine*, SVM), com horizonte de previsão diário, 365 passos à frente. O modelo adquirido pelo método *chaos particles warm optimization algorithm for combination method* (CPSO–CM) obteve erros de aproximadamente 22,24%, 0,4008 m/s, 0,2514 m/s em termos das métricas MAPE, MAE e MSE, respectivamente. Além disso, o *ensemble* desenvolvido foi comparado com os modelos individuais, e chegou-se à conclusão que o mesmo tem alto grau de precisão, e um desempenho melhor que os modelos individuais.

Ainda, Gebin *et al.* (2020) propuseram um modelo com base nas técnicas: modelo auto-regressivo integrado de médias móveis (do inglês, *Autoregressive Integrated Moving Average* (ARIMA), *eXtreme Gradient Boosting* (XGBoost) e Floresta aleatória (do inglês, *Random Forest*, RF), com horizonte de previsão de uma hora. O modelo de previsão proposto obteve erros de aproximadamente 8% e 1.028,69

megawatts (MW), em termos das métricas MAPE e RMSE, respectivamente. O modelo proposto foi comparado com técnicas individuais de previsão, resultando em um melhor desempenho, mais acurácia e sendo capaz de suprir a deficiência dos outros. Por fim, Ribeiro *et al.* (2022) propuseram um modelo de comitês de máquinas (do inglês, *ensembles*) que combina as técnicas *bagging* e *stacking* para previsão de séries temporais, com horizonte de previsão de 10 e 30 minutos (muito curto prazo) e 60 e 120 minutos (curto prazo). O modelo proposto obteve erros de aproximadamente 21,20%, 2.123,97 kWh; 1.538,24 kWh para a previsão de 120 min, em relação as métricas MAPE, RMSE, MAE, respectivamente. O modelo proposto foi comparado com vários modelos, entre eles *stacking*, aprendizado de máquina, ANN, modelos estatísticos, sendo que os resultados apontam que o modelo *ensemble* adotado apresenta resultados melhores que os modelos individuais citados.

Devido às características de séries temporais relacionadas às energias renováveis como intermitência e alteração no padrão dos dados devido a fatores externos, como clima e características geográficas dos parques eólicos, o desenvolvimento de modelos estatístico matemáticos no contexto de previsões de séries temporais torna-se uma tarefa desafiadora (XING, 2020; YONG, 2019). Nesse contexto, modelos relacionados à inteligência artificial (IA) ganham um importante papel nas previsões de energias renováveis, uma vez que eles agregam um "conjunto de ferramentas computacionais e matemáticas, de forma que os modelos computacionais possam se ajustar aos dados" (DA SILVA, 2022, p. 3), ou seja, os modelos possuem a capacidade de aprender padrões complexos. Consequentemente, tal capacidade de se adaptar a novos cenários é adequado no processo de previsões de séries temporais de energias renováveis.

Outrossim, os modelos híbridos ou comitês de máquina - também conhecidos como *ensembles* – ganham visibilidade no que se refere à previsão de séries temporais, pois de acordo com Gebin *et al.* (2020), eles consistem na combinação de modelos individuais de IA ou estatísticos com o intuito da geração de um modelo com melhor desempenho em termos da redução dos erros de previsão, visto que têm como objetivo aumentar a capacidade de generalização dos modelos individuais em um novo modelo conjunto.

Além do acima mencionado, De Lima *et al.* (2016) discorrem sobre a combinação de diferentes técnicas de previsões, salientando seu potencial em melhorar a acurácia das previsões e justificando seu uso para a geração de

ensembles. Dentre as justificativas apontadas pelos autores, estão o fato de não existir uma técnica de previsão perfeita, a variabilidade de precisão de uma técnica aplicada em períodos, horizontes de previsão e conjunto de dados diferentes, a sensibilidade à diversos fatores das técnicas individuais, e que a combinação das técnicas pode apresentar um resultado melhor que o obtido por técnicas individuais.

Ademais, esses mesmos autores reforçam que apesar das combinações de previsões terem o potencial de apresentar resultados significativamente melhores em termos de acurácia, nem sempre a combinação dos métodos gerará um resultado com menor erro em relação às técnicas individuais. Desse modo, para escolher a previsão mais adequada é necessário analisar a acurácia de cada uma das previsões geradas, através da comparação de medidas de erro relativas ou absolutas como MAPE, MAE e MSE.

Desta maneira, diante do contexto exposto para o campo das energias renováveis e a importância do estudo e desenvolvimento de modelos de previsão o objetivo deste estudo é: comparar a performance das técnicas individuais estatísticas e de IA, bem como suas combinações para a previsão de séries temporais de energia eólica, a fim de avaliar qual a técnica mais adequada para realizar tal previsão. As contribuições de estudo são destacadas a seguir: (i) diferentes *ensembles* de previsões de séries temporais são desenvolvidos, aplicados e avaliados para previsão da geração de energia de dois parques eólicos; (ii) um estudo comparativo entre as técnicas individuais e os modelos *ensembles* gerados no que se refere à acurácia na previsão de séries temporais de energia eólica; (iii) comparação e análise crítica das previsões de séries temporais de energia eólica obtidas por meio dos modelos de combinação desenvolvidos nesse estudo usando as métricas de desempenho.

1.1 PROBLEMATIZAÇÃO

Tendo em consideração o olhar do mundo para as energias renováveis, surge a necessidade de se criar modelos matemáticos capazes de prever a produção dessas energias, seja para a melhor negociação de energia e redução do risco de exposição financeira (Suárez-Cetrulo, *et al.*, 2022), seja para "facilitar o despacho de energia razoável e organizar o desligamento e a manutenção para garantir a operação estável do sistema de energia" (Jin, Huaiping *et al.* 2021).

No entanto, como a produção de energia eólica está fortemente associada a velocidade do vento, e devido à característica de instabilidade e de difícil previsão dos

mesmos, a previsão de energia eólica ainda é difícil de se prever (Melo, 2020; De Farias, RIBEIRO *et al.*, 2022). Assim, na literatura diversas técnicas são utilizadas para a previsão de séries temporais de energia eólica. Contudo, ainda não existe um consenso de qual é a abordagem mais adequada em termos de acurácia.

No estudo realizado por Ribeiro *et al.* (2022) é avaliado um modelo de *ensemble learning* que combina métodos *bagging* e *stacking* aplicados à previsão de séries temporais de geração de energia eólica para horizontes de previsão de 10, 30, 60 e 120 minutos. Além disso, os resultados obtidos mostraram que o modelo de *ensemble learning* proposto alcançou um desempenho de previsão melhor do que modelos de previsão únicos, como processo Gaussiano (do inglês, Gaussian Process, GP), regressão de vetor suporte (do inglês, *support vector regression*, SVR), RF, regressão de Ridge (do inglês, *Ridge Regression*, RR) e *stacking*, ANNs, como máquinas de aprendizado extremo (do inglês, *extreme learning machines*, ELM), e modelos estatísticos, como ARIMA, TBATS, THETA, *Naive*. Outrossim, os resultados com média ponderada foram 87,5% superior aos com média aritmética para previsão de energia eólica, e mostraram que a associação de estratégias para *ensembles* pode fornecer resultados precisos de previsão no campo das energias renováveis.

Nesse contexto, a questão investigativa que norteará essa pesquisa é: dentre as técnicas individuais de previsão adotadas por Ribeiro *et al.* (2022), bem como suas combinações por intermédio de média aritmética simples, média harmônica, mediana e média aritmética ponderada, qual é a abordagem mais adequada em termos de redução dos erros de previsão no contexto de previsão de séries temporais de energia eólica?

1.2 JUSTIFICATIVA

De acordo com os dados extraídos do INFOVENTO de 2022 (Associação Brasileira de Energia Eólica - ABEEólica, 2022), em 2020, no Brasil, 11,8% da capacidade de energia elétrica instalada é de fonte eólica, cuja geração representou 28,8 milhões de residências supridos por mês e 86,4 milhões de habitantes beneficiados com energia eólica. Além disso, essa produção evitou a emissão de 21,2 milhões de toneladas de CO₂, quantidade que é equivalente a emissão de CO₂ correspondente à aproximadamente 20,9 milhões de carros. Outrossim, a instalação de parques eólicos está associada ao aumento do Produto Interno Bruto (PIB) e do Índice de Desenvolvimento Humano do Município (IDHM). Pelos dados

obtidos no INFOVENTO de 2022, a comparação entre os municípios que possuem parques eólicos e os que não possuem, indicaram o aumento no PIB de 21,15% de 1999 para 2017, e de 20% no IDHM.

Levando em consideração todos os benefícios provenientes do uso de energias renováveis, procura-se expandir esse meio de produção de energia. Assim, a capacidade de energia eólica instalada no Brasil em MW e a previsão dessa capacidade nos próximos quatro anos é apresentada na Figura 1.

Figura 1 - Evolução da capacidade instalada da energia eólica em MW no Brasil



Fonte: Associação Brasileira de Energia Eólica - ABEEólica, 2022

No panorama mundial, pelos dados expostos no Conselho Global de Energia Eólica (do inglês, *Global Wind Energy Council*, GWEC), a capacidade de produção de energia eólica é de 837 gigawatts (GW), ajudando o mundo evitar 1,2 bilhões de toneladas de CO₂ anualmente, o que equivale à emissão de gás carbônico anual de toda América do Sul. Além disso, a China está em primeiro lugar no que se refere a produção de energia elétrica de fonte eólica, atingindo 50,91% de toda a capacidade de energia eólica instalada no mundo. Apesar de todos esses dados promissores, informações contidas no GWEC expõem que as previsões indicam que a energia eólica não está crescendo o suficiente para que ocorra uma transição de energia global segura.

Atentando para todos os benefícios citados pelo uso da energia eólica e em virtude de que a quantidade de energia de fonte eólica utilizada no Brasil e no mundo é considerável, surge a necessidade da obtenção de previsões precisas da energia eólica, seja para realizar tomadas de decisões no que se refere à exploração da

energia eólica de uma determinada região (Nascimento Camelo, 2018) ou para o planejamento de operações, visando a otimização no aproveitamento de recursos energéticos (Melo, 2020). Como resposta a esse panorama, surgiu o uso dos modelos de previsão, os quais são utilizados para realizar previsões acuradas de energia eólica.

Portanto, visto que ainda não existe um consenso de qual é a abordagem mais adequada em termos de acurácia para a previsão de energia eólica, o presente trabalho visa comparar a performance de algumas das técnicas individuais de inteligência artificial, bem como suas combinações para a previsão de séries temporais de energia eólica, a fim de avaliar o melhor desempenho em termos de menor erro.

1.3 Questão Investigativa

Dentre as combinações das técnicas individuais propostas por Ribeiro *et al.* (2022), por intermédio de média aritmética simples, média harmônica, mediana e média aritmética ponderada, qual é a abordagem mais adequada em termos de redução dos erros de previsão no contexto de previsão de séries temporais de energia eólica?

1.4 Objetivos

1.4.1 Objetivo Geral

Avaliar a performance de técnicas individuais e de suas combinações para a previsão de séries temporais de energia eólica.

1.4.2 Objetivos Específicos

- Realizar a análise exploratória dos dados, através de uso de gráficos e testes para séries temporais;
- Construir combinações de modelos individuais a partir dos modelos publicados em Ribeiro *et al.* (2022);
- Comparar as previsões obtidas no objetivo anterior com o trabalho referência;
- Computar as métricas (MAPE, RMSE, MSE) para as previsões desenvolvidas, e validar o modelo através da análise dos erros.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Série temporal

De acordo com Morettin e Toloí (2006), “uma série temporal é qualquer conjunto de dados ordenados no tempo”. Ainda, a série temporal permite ao usuário “gerar previsões baseadas em informações do seu passado e fornece uma visão da dinâmica do relacionamento entre as variáveis” (WOLFART *et al.*, 2020, p. 22). Ademais, de acordo com Fraccanabbia (2021), tais previsões podem ser obtidas por meio de modelos relacionados ao aprendizado de máquina, o que vem atraindo pesquisadores desse campo, os quais abordam em seus trabalhos as limitações dos métodos individuais e propõem novas abordagens, ou combinam as já existentes, em busca de melhor desempenho nas previsões.

Outrossim, para Morettin e Toloí (2006) a análise de séries temporais pode ter como objetivos: (a) investigar o mecanismo gerador da série temporal, ou seja, qual o acontecimento que determina a série; (b) fazer previsões de valores futuros da série (as quais podem ser de curto ou longo prazo); (c) descrever o comportamento da série (tais como sazonalidade ou tendência); e (d) procurar periodicidades relevantes nos dados.

De modo geral, de acordo com Hyndman e Athanasopoulos (2021) a análise e previsão de séries temporais envolvem diversas etapas, as quais são destacadas na Figura 2.

Figura 2 - Etapas de análise e previsão de séries temporais



Fonte: Elaborado pela autora

No contexto desse estudo, a definição do problema é a previsão de energia eólica. A coleta de dados, geralmente se dá entrando em contato com o parque eólico que se deseja realizar a previsão, mas nesse caso, foi obtida por meio do professor orientador, que já possuía tais dados e que foram utilizados na pesquisa que gerou Ribeiro *et al.* (2022). A análise exploratória dos dados engloba a criação de gráficos dos dados observados e de testes estatísticos como média, mediana, desvio-padrão, máximo e mínimo. Já a etapa de escolha de modelos de previsões deve ao fato de existirem muitos métodos de previsão de energia eólica, e deve-se escolher o que melhor se adequa para o cenário pretendido. Por fim, a avaliação de performance dos modelos se dá através das métricas de erro, como MAPE, MAE, RMSE e outras.

2.2 Machine learning

Atualmente o aprendizado de máquina, ou *machine learning* (ML) é muito utilizado em diversos campos, como por exemplo na medicina para diagnósticos e prognósticos, em mecanismos de busca na internet, em sistemas bancários para a detecção de fraudes e em sistemas de segurança de dados (De Mattos Paixão, 2022). Dentre as utilidades do *machine learning* podemos citar também a previsão de séries temporais (Carlotto, 2020).

De acordo com De Mattos Paixão (2022) o *machine learning* aprende através de seguidas observações e

[...] é um ramo da inteligência artificial (IA) que explora o estudo e a construção de algoritmos computacionais a partir do aprendizado por dados ao invés de instruções pré-programadas. O objetivo principal de um modelo de ML é construir um sistema de computador que aprenda com um banco de dados pré-definido e gere, ao final, um modelo de predição, classificação ou detecção. (DE MATTOS PAIXÃO, 2022, p. 95)

O desenvolvimento de um código de *machine learning* passa por algumas etapas. A primeira etapa é a de pré-processamento e engloba a organização do banco de dados, a definição da pergunta da pesquisa e a divisão dos dados em treinamento e teste. A segunda fase compreende o treinamento dos dados, e a última fase inclui a avaliação do modelo, na qual o modelo gerado é comparado com os dados de teste. (De Mattos Paixão, 2022).

2.3 Previsão de Geração Energia Eólica.

Nascimento Camelo *et al.* (2018) propuseram um modelo híbrido que combinava as técnicas Auto-Regressivo Integrado de Médias Móveis com entradas de Variáveis Exógenas (do inglês, *Auto-Regressive Integrated Moving Average with eXogenous inputs*, ARIMAX) e ANN para previsão de séries temporais, com horizonte de previsão horário e mensal. O modelo híbrido proposto, denominado ARXAN - Auto-Regressivo (AR), Exógeno (E), Artificial-Neural (AN) – obteve erros de aproximadamente 0,35 m/s; 0,46 m/s e 7,2% para a previsão mensal, em relação às métricas MAE, RMSE e MAPE, respectivamente. O modelo proposto foi comparado com as técnicas individuais ARIMA e ANN, e conclui-se que o modelo híbrido proposto apresenta um desempenho melhor que as técnicas individuais.

De Farias *et al.* (2019) avaliaram dois procedimentos de previsão de energia eólica envolvendo o método ANN com parâmetros de entrada GFS (*Global Forecast System*). O primeiro procedimento consistia em prever “diretamente a soma de geração do parque inteiro com o auxílio de um vetor de indisponibilidade”, enquanto o segundo constituiu-se em prever a geração de um aerogerador equivalente que multiplicada pelo número de aerogeradores funcionando resultou na previsão de geração do parque. O horizonte de previsão adotado foi de 120 horas, com janela temporal de 30 minutos. Dos procedimentos empregados, o segundo resultou em um melhor desempenho, com erros de aproximadamente 22,2% e 6669 MW, em relação às métricas erro quadrado médio da raiz normalizada (do inglês, *Normalized Root Mean Square Error*, NRMSE) e RMSE, respectivamente.

Melo *et al.* (2020) aplicaram ANN para a previsão de velocidade do vento, com horizonte de previsão de 1 hora, 160 passos à frente. A ANN foi treinada com os dados extraídos de uma estação meteorológica da cidade de Juiz de Fora. Vale dizer que foram realizadas cinco simulações, as quais se diferenciaram quanto aos parâmetros de entradas adotados. O modelo de previsão que apresentou menor erro, foi o que teve como variáveis de entrada a velocidade do vento, temperatura do ar e radiação solar global, apresentando erros de 25,791%; 0,8532 km/h e 0,6559; em termos das métricas MAPE, RMSE e U_2 , respectivamente.

Xing *et al.* (2020) propuseram um modelo híbrido combinando as técnicas ELM e decomposição em modo empírico de conjunto completo (do inglês, *complete ensemble empirical mode decomposition*, CEEMD), com horizonte de previsão de

quatro dias. O modelo proposto, denominado CEEMD-ELM, foi comparado com outros modelos, entre eles o ELM único, e concluiu-se que o modelo híbrido adotado apresentou resultados com maior acurácia que os outros, com erros aproximados de 9,0396%, 0,3071 m/s, 0,4305 m/s, em relação às métricas MAPE, MAE e RMSE, respectivamente.

Jin *et al.* (2021) propuseram um método *ensemble* (SEFMGPR) para previsão de séries temporais de energia eólica, com janela temporal de 15 min, e horizonte de previsão de 4, 8 e 16 passos à frente. O modelo proposto obteve erro de 0,7771 MW com relação às métricas RMSE, para o horizonte de previsão de 4 passos à frente. O modelo proposto foi comparado com vários outros, sendo que os resultados apontam que o adotado apresenta resultados melhores, em quesito de menor erro, que o método de Persistência (do inglês, *Persistence Method*) e que o modelo único de regressão do processo Gaussiano (do inglês, *Gaussian process regression*, GPR).

Karthikeyan e Rengaraj (2021) desenvolveram três modelos de previsões de velocidade do vento: *bagged trees*, *boosted trees* e SVR, com resolução temporal de 10 minutos, e horizonte de previsão de um ano. Dentre os modelos, o que apresentou melhor desempenho em termos de acurácia foi o modelo *bagged trees*, com erros de aproximadamente 1,12 m/s, 1,26 m/s e 0,89 m/s para o mês de janeiro, com relação às métricas RMSE, MSE e MAE, respectivamente. Suárez-Cetrulo *et al.* (2022) aplicaram uma ampla variedade de algoritmos de aprendizado de máquina em oito parques eólicos distintos, com a finalidade de avaliar qual destes algoritmos apresenta menores erros de previsão de todos os parques eólicos. Dentre os quinze algoritmos testados, haviam modelos básicos de regressão, ANNs e técnicas de *ensemble* competitivas. Os autores verificaram que os algoritmos de *xGBoost* e *Gradiente Boosting* são considerados melhores, no que se refere ao desempenho, em todos os parques eólicos analisados.

De forma ampla, a Tabela 1 resume os trabalhos detalhados acima, assim como os apresentados na seção 1.

Tabela 1 - Revisão dos modelos de previsão utilizados para a previsão de energia eólica

Referência	Ano de publicação	Técnicas utilizadas	Resolução temporal	Horizonte de previsão - passos à frente	Méridas de Desempenho
JUNIOR <i>et al.</i> (2018)	2018	HW, Híbrido	1 mês	12	MAE, MAPE, RMSE, NS
NASCIMENTO CAMELO <i>et al.</i> (2018)	2018	ARIMAX, ANN, Híbrido:ARNAX	1 dia, 1 mês	24 e 12	MAE, MAPE, RMSE, NS
DE FARIAS <i>et al.</i> (2019)	2019	ANN parâmetros de entrada GFS	30 min	60	RMSE, NRMSE, validação cruzada (cross-validation)
YONG <i>et al.</i> (2019)	2019	BP, SVM, Híbridos (EEMD-BP, EEMD-SVM, WT-BP, WT-SVM), Combinados (CM-NNCT, CPSO-CM)	1 dia	365	MAE, MSE, MAPE
MELO <i>et al.</i> (2020)	2020	ANN	1 hora	160	MAPE, RMSE, algoritmo U de Theil
XING <i>et al.</i> (2020)	2020	Híbrido (CEEMD-ELM), ELM único, WD-ELM, EMD-ELM, EEMD-ELM			MAE, MAPE, RMSE
GEBIN <i>et al.</i> (2020)	2020	XG Boost, Random Forest, Arima, Ensemble	1 hora	150	MAPE, RMSE
JIN <i>et al.</i> (2021)	2021	método de persistência, GPR, EGPR, EFMGPR, SEFMGPR_nãoadaptado, SEFMGPR	15 min	4, 8 e 16	RMSE, R ²
KARTHIKEYAN e RENGARAJ (2021)	2021	Bagged trees, boosted trees, SVR	10 min	4320	RMSE, R ² , MSE, MAE
RIBEIRO <i>et al.</i> (2022)	2022	<i>Ensemble: bagging com moving block bootstrap (MBB).</i> Aprendizado de máquina: <i>stacking ensemble learning</i> . Ajuste de pesos (MOO)	10 min	1, 3, 6 e 12	MAE, MAPE, RMSE, NRMSE, U ₂ , IP, DM
SUÁREZ-CETRULO <i>et al.</i> (2022)	2022	<i>gradiente boosting, linear regression</i> , SVR, LSTM, Dense LSTM, Cayesian Ridge, LSTM Ens. [G], <i>Neural Net</i> , ANN Ens. [M], ANN Ens. [G], <i>Boosting Ens.</i> , xGBoost, Reference Curve	10 min	144	MAE, RMSE

Fonte: Elaborado pela autora

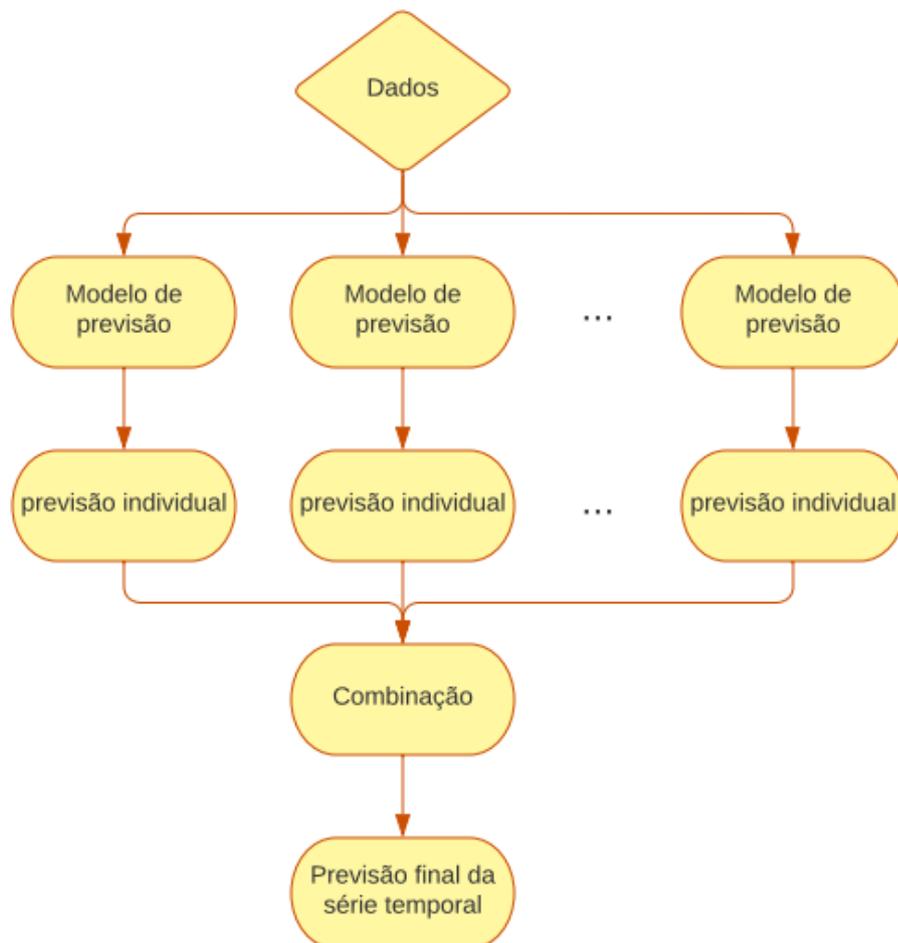
2.4 Combinação de técnicas de previsão

De Lima *et al.* (2016) discorrem sobre a combinação de diferentes técnicas de previsões, salientando seu potencial em melhorar a acurácia das previsões. Nesse método, diferentes técnicas são combinadas para se obter a previsão final, em que a intensão de combinar a diferentes técnicas é que se possa melhorar a acurácia ao se extrair informações úteis e independentes dos dados. Os autores também justificam a utilização de combinações de previsões, e, dentre as apontadas pelos autores, estão: o fato de não existir uma técnica de previsão perfeita, a variabilidade de precisão de

uma técnica aplicada em períodos, horizontes de previsão e conjunto de dados diferentes, a sensibilidade à diversos fatores das técnicas individuais, e que a combinação das técnicas pode apresentar um resultado melhor que o obtido por técnicas individuais.

A Figura 3 ilustra como se gera uma previsão a partir de combinação de modelos de previsão. A partir dos dados obtidos, se gera n previsões de séries temporais individuais, cada uma baseada em diferentes modelos de previsão. Então, todas essas previsões individuais são combinadas com a finalidade de se obter a previsão final da série temporal.

Figura 3 - Diagrama combinação de métodos de previsão



Fonte: Elaborado pela autora

Na seção 2.4 são detalhadas diferentes maneiras de combinar modelos de previsão.

2.5 Estratégias para combinação de métodos de previsão

Considerando o exposto na seção 2.2, são limitados os estudos que abordam a metodologia utilizada por De Lima *et al.* (2016) no contexto da previsão de energia eólica. Portanto, esta abordagem será adotada neste estudo, como maneira de expandir o estudo de Ribeiro *et al.* (2022) descrito na seção 1.2. Na sequência, serão apresentados os principais aspectos relacionados a combinação de métodos de previsão conforme descrito em De Lima *et al.* (2016).

A combinação de técnicas de previsão é dada por:

$$F_c(t) = \sum_{i=1}^k w_i F_i(t) \text{ com } \sum_{i=1}^k w_i \leq 1 \text{ e } 0 \leq w_i \leq 1 \quad (1)$$

Na qual:

$F_c(t)$: é o resultado da previsão no instante t ;

k : é o número de técnicas de previsões que são combinadas;

$F_i(t)$: é o resultado da i – ésima previsão;

w_i : é o peso atribuído à i – ésima previsão.

De acordo com De Lima *et al.* (2016), dentre as formas de combinações existentes temos a combinação pela média aritmética, na qual todos os pesos recebem o valor igual a 1, e é expressa por:

$$F_c(t) = \frac{1}{k} \sum_{i=1}^k F_i(t) \text{ com } w_i = \frac{1}{k} \quad (2)$$

Além disso, esses mesmos autores apresentam a combinação pela média harmônica que é dada por:

$$F_c(t) = \frac{k}{\sum_{i=1}^k \frac{1}{F_i(t)}} \quad (3)$$

Por outro lado, também é possível considerar como combinação a mediana das previsões dos métodos individuais escolhidos durante o estudo.

Outra abordagem apontada por De Lima *et al.* (2016), é a combinação de previsões em que os pesos são definidos de acordo com o desempenho de cada

técnica individual. Nesse caso, a técnica que apresenta o melhor desempenho, recebe maior peso. As fórmulas para calcular os pesos em relação as medidas MAPE, RMSE e MAE, são respectivamente:

$$w_i = \frac{1}{\frac{\overline{MAPE}_i}{\sum_{i=1}^k \frac{1}{\overline{MAPE}_i}}} \quad (4)$$

$$w_i = \frac{1}{\frac{\overline{RMSE}_i}{\sum_{i=1}^k \frac{1}{\overline{RMSE}_i}}} \quad (5)$$

$$w_i = \frac{1}{\frac{\overline{MAE}_i}{\sum_{i=1}^k \frac{1}{\overline{MAE}_i}}} \quad (6)$$

2.6 Medidas de avaliação dos modelos

Para avaliar a efetividade dos modelos desenvolvidos neste estudo, em termos de previsão fora da amostra (para conjunto de teste), serão adotadas as medidas de desempenho MAPE, RMSE, e MAE, as quais são computadas nas equações (7), (8) e (9), respectivamente (Ribeiro et al., 2022):

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (7)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (8)$$

$$MAE = \sum_{i=1}^n \frac{1}{n} |y_i - \hat{y}_i| \quad (9)$$

y_i representa o i – ésimo valor observado.

\hat{y}_i representa o i – ésimo valor previsto.

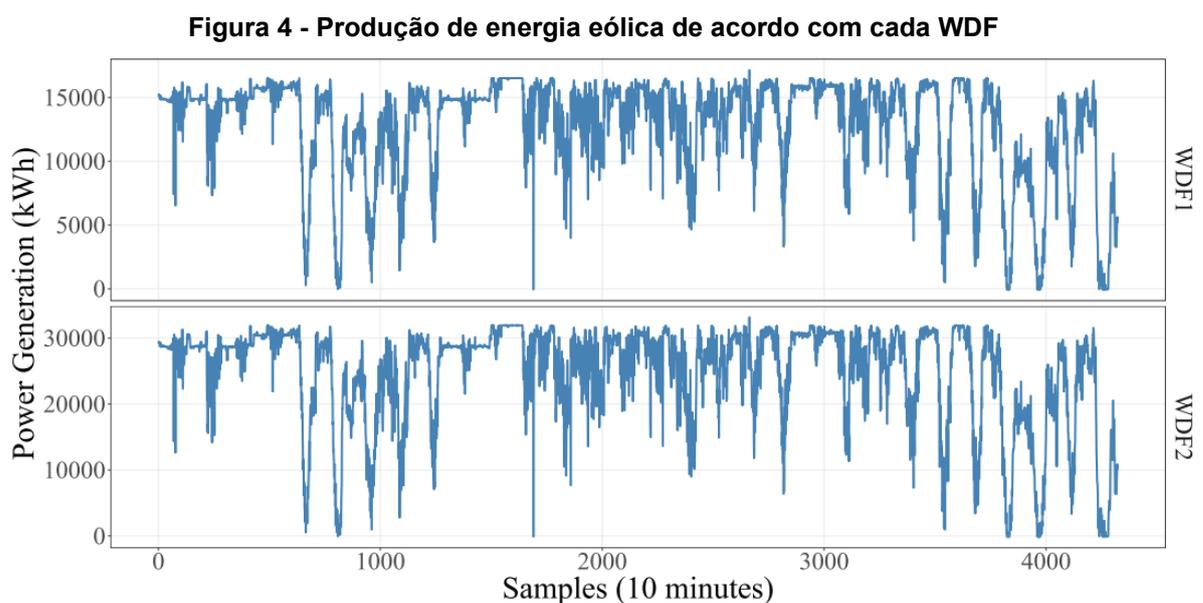
n corresponde ao número de observações utilizadas.

3 METODOLOGIA DA PESQUISA

3.1 Fonte dos dados

Os dados adotados para o presente trabalho são os mesmos utilizados por Ribeiro *et al.* (2022), e se tratam de duas séries temporais constituídas de valores de geração de energia eólica em kWh gravados a cada 10 min de dois parques eólicos (WDF1 e WDF2), ambos localizado na Bahia, Brasil, referente ao mês de janeiro de 2021. O parque WDF1 é localizado em São Sebastião do Alto, é composto de 12 turbinas e tem capacidade instalada de 19,20 *megawatts-hour* (MW). O parque WDF2 é localizado em Rio Verde e é composto de 19 turbinas, cada uma com capacidade de geração de 1,68 MWh.

Além disso, a duração da série temporal atingiu 4.322 amostras para janeiro de 2021, coletadas a cada 10 min, conforme mostrado na Figura 4 (Ribeiro *et al.*, 2022). Ainda, as medidas estatísticas mediana, média, valor máximo e mínimo, e desvio padrão para cada série temporal adotada no estudo de (Ribeiro *et al.*, 2022) é apresentada na Tabela 2. Os dados completos apresentados na Tabela 2 representam toda série de dados de geração de energia eólica, o conjunto treino exprime 70% dos dados completos e são utilizados para treinar os modelos de previsão, e o conjunto teste é composto pelos 30% restantes dos dados completos e é utilizado para testar a performance em dados desconhecidos pelos modelos.



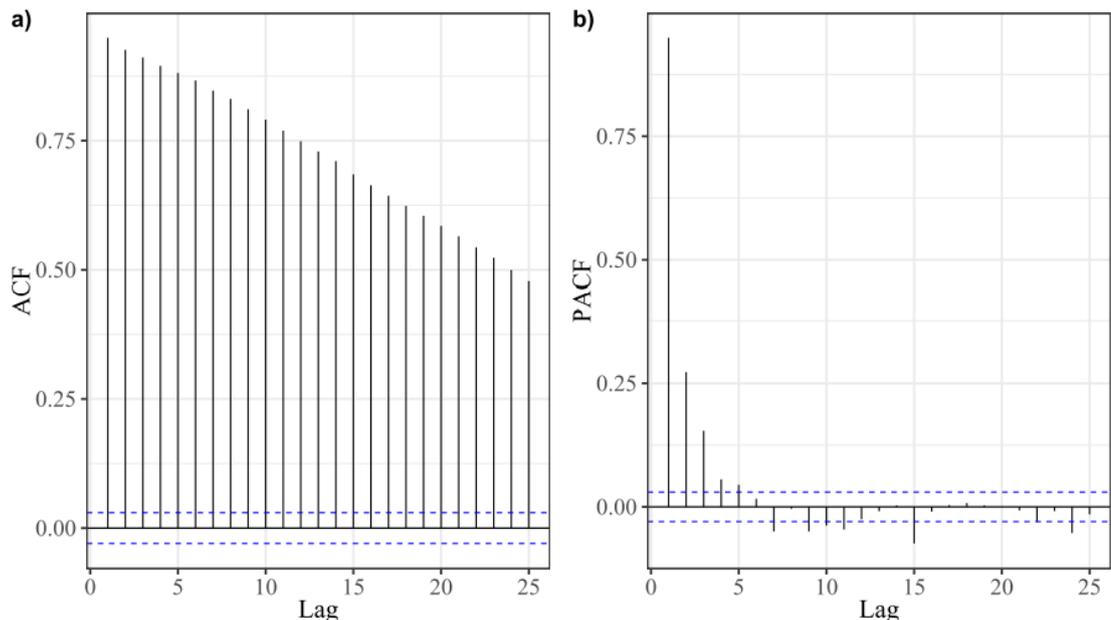
Fonte: Ribeiro *et al.* (2022, p. 6)

Tabela 2 - Medidas estatísticas para cada série temporal

WDF	Conjunto de dados	Medidas estatísticas					
		#Amostras	Mínimo (kW)	Mediana (kW)	Média (kW)	Máximo (kW)	Desvio-padrão (kW)
1	Completo	4322	0	14.470,81	12.752,47	17.146,42	3.973,87
	Treino	3025	0	14.796,41	13.502,21	17.146,42	3.169,87
	Teste	1297	0	12.750,91	11.003,86	16.523,81	4.983,32
2	Completo	4322	0	27.976,90	24.654,78	33.149,74	7.682,82
	Treino	3025	0	28.606,39	26.104,27	33.149,74	6.128,42
	Teste	1297	0	24.651,76	21.274,12	31.946,03	9.634,41

Fonte: Adaptado de Ribeiro *et al.* (2022, p. 5)

Por fim, a função de autocorrelação (do inglês, *autocorrelation function*, ACF) e a função de autocorrelação parcial (do inglês, *partial autocorrelation function*, PACF) são mostradas na Figura 5. Onde tem-se que a ACF “sugere um padrão não estacionário uma vez que há um decaimento linear” (Ribeiro e colaboradores (2022)) e a PACF “mostra defasagens de 1 a 5, 7, 9, 11,15 e 24 como candidatos de entrada.” (Ribeiro *et al.*, 2022). Além disso, as linhas azuis pontilhadas indicam intervalos de 95% de confiança para as correlações da produção de energia eólica no instante t (1,...,25) e seus atrasos.

Figura 5 - Valores ACF e PACF para as amostras

Fonte: Ribeiro e colaboradores (2022, p. 6)

3.2 Procedimentos de coleta, tratamento e análise dos dados

Para realizar o presente estudo, os dados adotados serão os já utilizados por Ribeiro e colaboradores (2022). Após a obtenção dos dados, será realizada uma análise exploratória dos dados, através do uso de gráficos e testes para séries temporais, conforme proposto em Morettin e Toloï (2006) e Paiva e Sáfiadi (2021). Em seguida, serão utilizadas seis abordagens de combinação das previsões dos modelos individuais publicados em Ribeiro e colaboradores (2022). A saber, médias aritmética, harmônica e ponderada, sendo que a última destas gerará três previsões distintas. Além disso, a medida de tendência central mediana será adotada, pois ela elimina o viés de valores extremos. No contexto da combinação por média ponderada os pesos serão obtidos de acordo com as equações (4), (5) e (6). Da mesma maneira, para as combinações através de média aritmética e harmônica as equações (2) e (3) serão utilizadas.

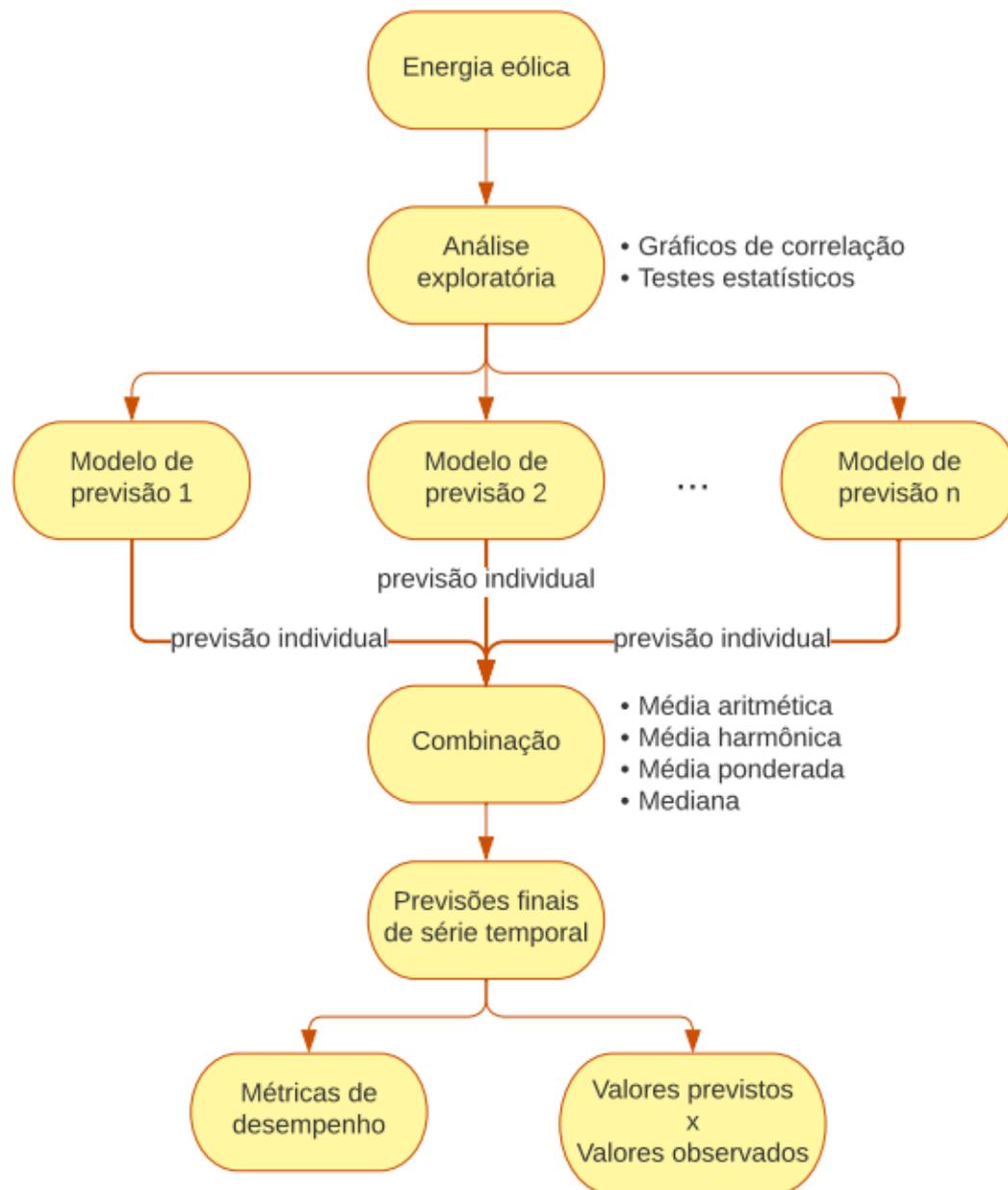
Realizada as previsões dos modelos combinados, serão computadas as métricas MAPE, RMSE e MSE para analisar a acurácia das previsões. Por fim, as previsões desenvolvidas serão comparadas entre si, e com as previsões obtidas no trabalho de referência, com a finalidade de avaliar qual modelo apresenta melhor desempenho, isto é, menor erro com relação às métricas indicadas no início deste parágrafo.

Este trabalho se trata de uma pesquisa aplicada, pois se trata de um estudo voltado à aquisição de conhecimentos com a finalidade de uma aplicação em uma situação específica (Gil, 2019). Também se trata de uma pesquisa descritiva, pois visa identificar padrões nas séries temporais, como tendências ou sazonalidade. Por fim, trata-se ainda de uma pesquisa quantitativa, dada a origem e características dos dados, por utilizar de “técnicas e recursos padronizados já existentes na literatura” (WOLFART, 2020, p. 24).

A realização da análise de dados, criação de gráficos, assim como o desenvolvimento das combinações e computação das métricas se dará por meio do *software* R 4.1.3 (R Core Team, 2022).

A Figura 6 ilustra os passos a serem seguidos para a realização do presente estudo, os quais estão detalhados nos três primeiros parágrafos desta seção.

Figura 6 - Diagrama combinação de métodos de previsão



Fonte: Elaborado pela autora

4 APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS

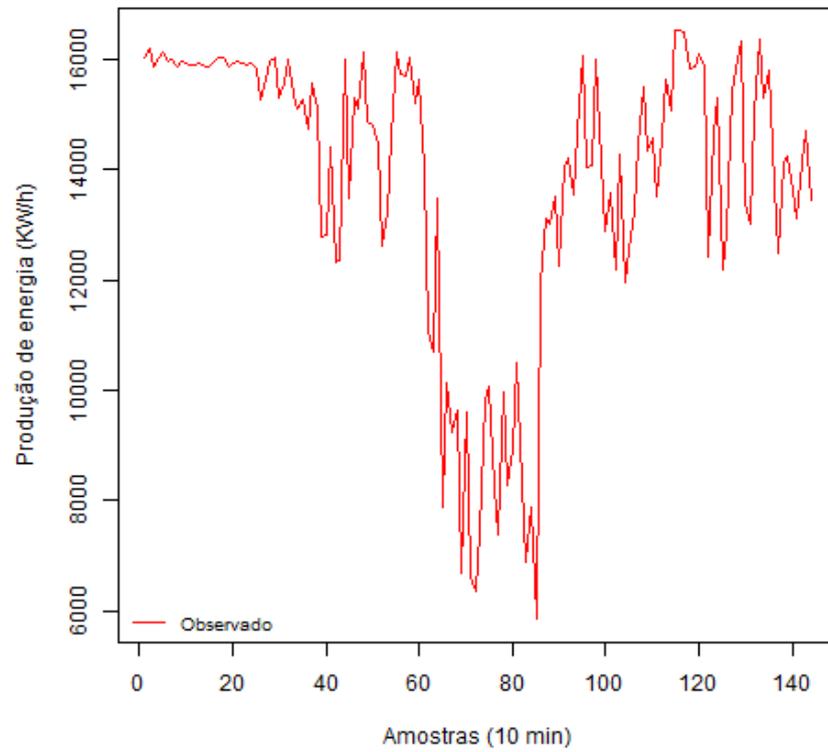
Nessa seção, os resultados de seis modelos de combinações aplicados em sete conjuntos de modelos base serão apresentados e discutidos. Na seção 4.1 é exibida a análise exploratória dos dados, já na seção 4.2 são apresentadas as nomenclaturas dos *ensembles* gerados. Na seção 4.3 são discutidos os resultados obtidos para os parques WDF1 e WDF2, respectivamente. Na seção 4.4 são apresentados gráficos das séries temporais observadas e previstas para o parque WDF1 e WDF2. Ademais, na seção 4.5 se encontra a comparação entre os desempenhos obtidos pelas combinações adotadas nesse estudo e pelas técnicas individuais. Por fim, na seção 4.6 é apresentada a comparação entre as previsões obtidas em Ribeiro e colaboradores (2022) com as obtidas nesse trabalho. Nas tabelas 5 e 6 os modelos com melhor desempenho, isto é, menores valores para as métricas de erro, estão destacados em negrito.

4.1 Análise exploratória dos dados

Através do *software R* foram gerados os gráficos de séries temporais de energia eólica dos parques WDF1 e WDF2, e computadas os valores de máximo e mínimo, média, mediana e desvio-padrão dos valores observados. As figuras 7 e 8, mostram parte da série temporal gerada para os parques WDF1 e WDF2, respectivamente. Além disso, o valor mínimo observado no WDF1 foi de -71,43 kWh (sendo que o valor negativo representa que o parque não gerou energia naquele momento, apenas consumiu), enquanto que o valor máximo foi de 16.523,81 kWh, e os valores de média, mediana e desvio padrão foram: 10.996,13 kWh, 12.744,48 kWh e 4.983,28 kWh, respectivamente.

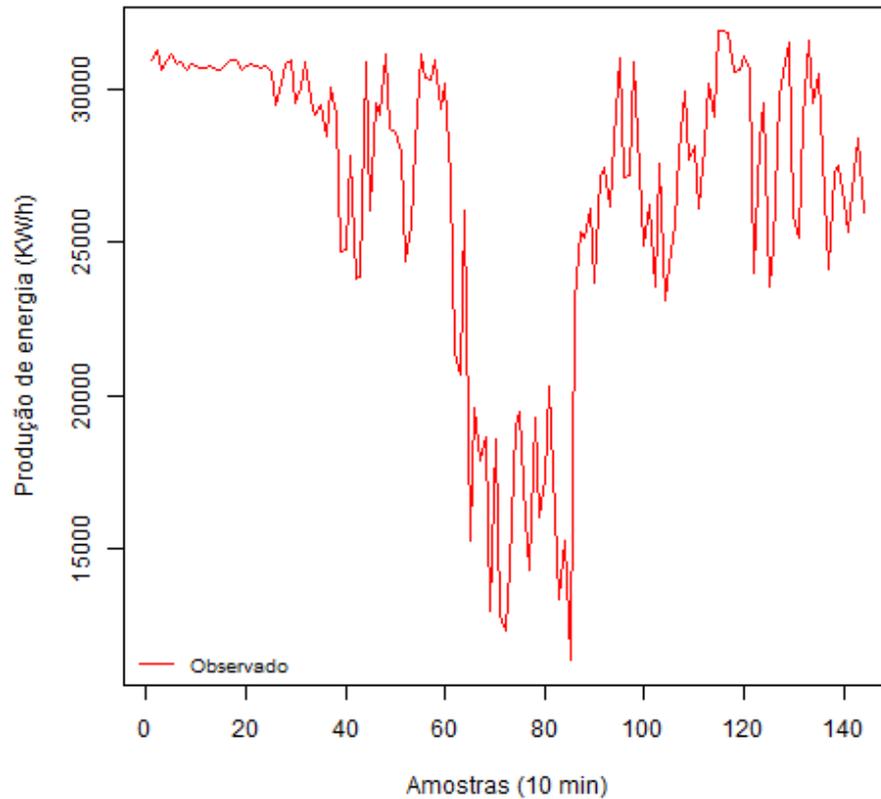
Já, o parque WDF2 apresentou valores de mínimo de -138,1 kWh e de máximo de 31.946,0 kWh. Além disso, os valores de média, mediana e desvio-padrão foram: 21.369,2 kWh, 24.773,6 kWh e 9.600,43 kWh, respectivamente.

Figura 7 - Parte da série temporal dos valores observados de energia eólica produzida em WDF1



Fonte: Elaborado pela autora

Figura 8 - Parte da série temporal dos valores observados de energia eólica produzida em WDF2



Fonte: Elaborado pela autora

4.2 Nomenclatura dos *ensembles* gerados

Inicialmente foram gerados *ensembles* combinando-se dois, três, quatro e cinco modelos base de previsão (SVR, ELM, GP, RR e RF) por meio da combinação por média aritmética, mediana, média harmônica e média ponderada, esta última com pesos definidos a partir das métricas RMSE, MAE e MAPE, conforme detalhado na seção 3.2.

A escolha dos modelos base de cada *ensemble* se baseou no modelo com menor erro RMSE apresentado por Ribeiro *et al.* (2022) (p. 13), levando em consideração o horizonte de previsão. Exemplificando, para gerar um *ensemble* a partir de dois modelos base para o horizonte de previsão de 10 minutos, foi escolhido os dois modelos base com menor erro RMSE no horizonte de previsão de 10 minutos. A partir desse critério, foram criados quarenta e dois *ensembles* no total.

Os *ensembles* serão nomeados da seguinte forma:

XCombYz,

em que:

X indica a quantidade de modelos base utilizados.

Y indica quais foram os modelos base utilizados a partir da Tabela 3.

z indica qual foi a combinação utilizada com base na Tabela 4.

Quadro 1 – Índices dos modelos base utilizados

Índice (Y)	Modelos base
1	SVR - ELM
2	SVR - RR
3	SVR - ELM - GP
4	SVR - ELM - RR
5	SVR - RR - GP
6	SVR - ELM - GP - RR
7	SVR - ELM - GP - RR - RF

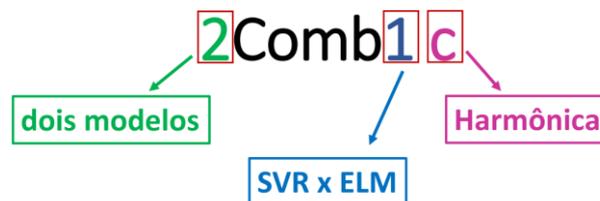
Fonte: Elaborado pela autora

Quadro 2 – Índices da combinação utilizada

Índice (z)	Combinação
a	Média
b	Mediana
c	Harmônica
d	Ponderada RMSE
e	Ponderada MAE
f	Ponderada MAPE

Fonte: Elaborado pela autora

Assim, o modelo 2Comb1c será o *ensemble* gerado a partir dos modelos SVR e ELM combinados por meio da média harmônica. Isso é ilustrado na Figura 9:

Figura 9 - Nomenclatura *ensembles*

Fonte: Elaborado pela autora

Após serem gerados os modelos, foram computadas as métricas MAE, MAPE e RMSE por meio do *software* R. As Tabelas 3 e 4, apresentam os erros obtidos para as métricas RMSE, MAE e MAPE de cada uma das previsões geradas.

4.3 Comparação entre os modelos propostos para o WDF1 e para o WDF2

Para os dados WDF1, a previsão com horizonte de 10 min que obteve resultado com menor erro de acordo com a métrica RMSE foi a obtida pelo modelo 2Comb1d (SVR-ELM combinados através de média ponderada com RMSE), apresentando um erro de 122,59 kWh e desempenho 0,13% melhor que o segundo modelo com menor erro (2Comb1a, SVR-ELM combinados por intermédio da média aritmética), 97,46% melhor que o modelo que apresentou maior erro (2Comb1c, SVR-ELM combinados por meio da média harmônica), e 89,67% melhor que o segundo modelo com maior erro (5Comb7f, SVR-ELM-GP-RR-RF combinados através da média ponderada com MAPE).

Tabela 3 - Métricas RMSE, MAE e MAPE das previsões geradas a partir de dois ou três modelos base

WDF	Horizonte de previsão	Critério	2Comb1					3Comb3						
			Média (a)	Mediana (b)	Harmônica (c)	Ponderada RMSE (d)	Ponderada MAE (e)	Ponderada MAPE (f)	Média (a)	Mediana (b)	Harmônica (c)	Ponderada RMSE (d)	Ponderada MAE (e)	Ponderada MAPE (f)
1	10 min	RMSE (kWh)	122,75	122,75	4827,61	122,59	124,36	126,13	411,30	1164,90	1126,32	413,54	403,84	399,21
		MAE (kWh)	78,42	78,42	755,19	78,36	79,91	81,71	290,61	811,45	469,75	292,21	285,31	282,02
		MAPE	10,45%	10,45%	1187,29%	10,35%	10,97%	11,36%	20,22%	64,45%	181,56%	20,36%	19,76%	19,47%
	30 min	RMSE (kWh)	1449,32	1449,32	1450,14	1448,62	1449,38	1448,97	500,59	1410,32	1225,45	454,84	491,49	452,43
		MAE (kWh)	1044,11	1044,11	1044,21	1043,01	1044,22	1043,56	355,76	994,13	600,80	322,38	349,18	320,73
		MAPE	101,42%	101,42%	100,76%	100,83%	101,48%	101,13%	28,40%	93,75%	254,01%	24,45%	27,72%	24,43%
	60 min	RMSE (kWh)	1795,19	1795,19	1795,72	1794,64	1795,25	1794,92	1798,29	1803,70	1797,84	1798,04	1798,35	1798,23
		MAE (kWh)	1307,51	1307,51	1307,24	1306,43	1307,62	1306,97	1312,71	1321,77	1311,64	1312,25	1312,81	1312,58
		MAPE	145,47%	145,47%	144,72%	144,72%	145,55%	145,10%	149,20%	154,74%	148,39%	148,89%	149,27%	149,12%
	120 min	RMSE (kWh)	2435,77	2435,77	2634,32	2435,76	2435,78	2435,76	2437,00	2438,84	2436,54	2436,97	2437,02	2437,02
		MAE (kWh)	1747,83	1747,83	1783,61	1746,83	1747,93	1747,33	1753,13	1762,25	1752,43	1752,70	1753,23	1753,02
		MAPE	230,11%	230,11%	231,07%	229,09%	230,21%	229,60%	235,15%	242,48%	234,29%	234,75%	235,24%	235,05%
2	10 min	RMSE (kWh)	2313,10	2313,10	2312,39	2313,09	2313,19	2312,82	2317,33	2326,23	2317,61	2317,32	2317,43	2317,34
		MAE (kWh)	1644,64	1644,64	1644,89	1644,62	1644,78	1644,23	1651,12	1664,12	1651,04	1651,10	1651,26	1651,13
		MAPE	70,39%	70,39%	69,97%	70,38%	70,44%	70,25%	72,50%	75,97%	71,92%	72,49%	72,54%	72,50%
	30 min	RMSE (kWh)	2802,02	2802,02	2803,64	2802,00	2802,16	2801,62	967,81	2726,68	2369,50	973,01	950,97	848,38
		MAE (kWh)	2018,64	2018,64	2018,85	2018,61	2018,85	2018,00	687,82	1922,08	1161,66	691,58	675,64	601,11
		MAPE	101,43%	101,43%	100,77%	101,42%	101,49%	101,25%	28,41%	93,77%	254,07%	28,62%	27,76%	23,45%
	60 min	RMSE (kWh)	3470,73	3470,73	3471,77	3470,72	3470,84	3470,41	3476,72	3487,16	3475,87	3476,70	3476,83	3476,77
		MAE (kWh)	2527,90	2527,90	2527,39	2527,87	2528,11	2527,27	2537,93	2555,42	2535,89	2537,90	2538,13	2537,98
		MAPE	145,49%	145,49%	144,74%	145,48%	145,56%	145,26%	149,20%	154,74%	148,40%	149,19%	149,27%	149,23%
	120 min	RMSE (kWh)	4709,19	4709,19	5201,74	4709,19	4709,19	4709,18	4711,56	4715,10	4710,73	4711,55	4711,59	4711,63
		MAE (kWh)	3379,19	3379,19	3455,77	3379,16	3379,38	3378,61	3389,42	3407,02	3388,13	3389,39	3389,61	3389,49
		MAPE	230,12%	230,12%	231,17%	230,11%	230,22%	229,82%	235,16%	242,47%	234,30%	235,15%	235,25%	235,20%

Fonte: Elaborado pela autora

Tabela 4 - Métricas RMSE, MAE e MAPE das previsões geradas a partir de quatro ou cinco modelos base

WDF	Horizonte de previsão	Critério	4Comb6					5Comb7						
			Média (a)	Mediana (b)	Harmônica (c)	Ponderada RMSE (d)	Ponderada MAE (e)	Ponderada MAPE (f)	Média (a)	Mediana (b)	Harmônica (c)	Ponderada RMSE (d)	Ponderada MAE (e)	Ponderada MAPE (f)
1	10 min	RMSE (kWh)	606,27	1183,21	859,34	585,12	600,97	581,85	722,98	1186,36	944,57	711,70	720,90	703,88
		MAE (kWh)	431,00	835,04	525,94	415,75	427,24	413,49	518,47	840,78	594,33	510,58	517,10	504,87
		MAPE	34,10%	70,19%	125,25%	32,59%	33,78%	32,47%	44,01%	72,83%	110,59%	43,27%	43,94%	42,75%
	30 min	RMSE (kWh)	736,66	1433,49	1132,20	711,07	730,27	707,17	877,78	1430,73	1129,06	864,16	875,27	854,72
		MAE (kWh)	528,82	1025,04	676,02	510,04	524,21	507,32	632,08	1023,22	733,10	622,10	630,32	615,26
		MAPE	48,73%	101,21%	183,26%	46,57%	48,26%	46,37%	57,14%	99,49%	145,70%	55,82%	56,92%	55,30%
	60 min	RMSE (kWh)	910,17	1775,76	1560,29	878,67	902,26	873,76	1085,76	1770,07	2580,78	1069,09	1082,66	1057,31
		MAE (kWh)	657,78	1284,06	899,24	634,32	651,98	630,83	790,10	1279,86	999,81	777,68	787,88	768,97
		MAPE	67,21%	145,22%	260,71%	64,06%	66,51%	63,75%	74,84%	138,40%	197,23%	72,70%	74,43%	72,11%
	120 min	RMSE (kWh)	1246,50	2403,46	2636,19	1204,74	1235,91	1198,06	1452,00	2390,45	6753,03	1454,77	1472,87	1438,93
		MAE (kWh)	879,62	1710,12	1345,06	848,72	871,89	843,97	1027,58	1694,65	1509,53	1033,58	1047,02	1022,15
		MAPE	119,61%	229,74%	491,99%	115,20%	118,64%	114,75%	115,85%	218,34%	246,78%	126,03%	128,47%	125,23%
2	10 min	RMSE (kWh)	1172,13	2287,55	1661,44	1175,55	1162,42	1106,24	1396,28	2293,46	12617,35	1402,42	1393,15	1354,32
		MAE (kWh)	833,28	1614,43	1016,86	835,71	826,38	786,13	1001,06	1624,63	1476,07	1005,54	999,04	971,56
		MAPE	34,10%	70,19%	125,27%	34,21%	33,80%	31,88%	43,59%	72,91%	121,10%	43,81%	43,54%	42,25%
	30 min	RMSE (kWh)	1424,21	2771,44	2189,02	1428,33	1412,51	1344,69	1692,82	2765,59	3181,17	1700,20	1689,00	1642,07
		MAE (kWh)	1022,39	1981,80	1307,02	1025,38	1013,96	964,59	1218,56	1976,60	1475,39	1223,94	1215,88	1181,80
		MAPE	48,73%	101,22%	183,28%	48,89%	48,29%	45,52%	56,84%	99,38%	146,99%	57,09%	56,66%	54,67%
	60 min	RMSE (kWh)	1759,68	3433,20	3016,74	1764,78	1745,17	1661,41	2094,42	3420,84	3819,08	2103,57	2089,72	2031,93
		MAE (kWh)	1271,73	2482,60	1738,65	1275,49	1261,10	1199,15	1523,45	2472,12	1912,32	1530,29	1520,16	1477,47
		MAPE	67,21%	145,24%	260,74%	67,45%	66,56%	62,47%	73,82%	138,34%	190,60%	74,11%	73,43%	70,37%
	120 min	RMSE (kWh)	2409,93	4646,78	5097,02	2416,75	2390,51	2279,03	2836,54	4631,77	5515,34	2848,55	2829,77	2751,81
		MAE (kWh)	1700,62	3306,36	2600,67	1705,61	1686,46	1604,47	2015,23	3293,38	2744,01	2024,03	2010,44	1953,51
		MAPE	119,62%	229,75%	492,05%	119,96%	118,70%	112,95%	128,62%	219,07%	345,39%	129,02%	128,07%	123,75%

Fonte: Elaborado pela autora

Ainda no WDF1, para o horizonte de 30 min, o modelo 3Comb4f (SVR-ELM-RR combinados através da média ponderada com MAPE) obteve menor erro que os demais (320,73 kWh) segundo a métrica MAE, apresentando desempenho 0,51% melhor que o segundo modelo com menor erro nessa categoria (3Comb4d, SVR-ELM-RR combinados por meio da média ponderada com RMSE) e 69,27% melhor que o modelo que apresentou maior erro (2Comb2f, SVR-RR combinados através da média ponderada com MAPE). Já para o horizonte de previsão de 60 min, o modelo 4Comb6f (SVR-ELM-GP-RR combinados por intermédio da média ponderada com MAPE) obteve menor erro segundo a métrica MAPE, apresentando um erro de 63,75% e desempenho 0,48% melhor que o segundo modelo com menor erro (4Comb6d, SVR-ELM-GP-RR, combinados através da média ponderada com RMSE) e 75,55% melhor que o modelo que apresentou maior erro (4Comb6c, SVR-ELM-GP-RR, combinados através da média harmônica).

Seguindo no WDF1, para o horizonte de previsão de 120 min, o modelo 4Comb6f (SVR-ELM-GP-RR combinados através da média ponderada com MAPE) obteve menor erro segundo a métrica RMSE, apresentando um erro de 1198,06 kWh e desempenho 0,55% melhor que o segundo modelo com menor erro (4Comb6d, SVR-ELM-GP-RR combinados através da média ponderada com RMSE), 82,26% melhor que o modelo que apresentou maior erro (5Comb7c, SVR-ELM-GP-RR combinados através da média harmônica), e 54,55% melhor que o segundo modelo de maior erro (4Comb6c, SVR-ELM-GP-RR-RF combinados por intermédio da média harmônica).

Para os dados WDF2, a previsão com horizonte de 10 min que obteve resultado com menor erro de acordo com a métrica RMSE foi a obtida pelo modelo 4Comb6f (SVR-ELM-GP-RR combinados através da média ponderada com MAPE), apresentando um erro de 1106,24 kWh e desempenho 4,83% melhor que o segundo modelo com menor erro (4Comb6e, SVR-ELM-GP-RR combinados através da média ponderada com MAE) e 52,44% melhor que o modelo que apresentou maior erro (3Comb5b, SVR-RR-GP combinados por intermédio da mediana).

Ainda em WDF2, para o horizonte de 30 min, o modelo 3Comb4f (SVR-ELM-RR combinados através da média ponderada com MAPE) obteve menor erro que os demais (601,11 kWh) segundo a métrica MAE, apresentando desempenho 11,03% melhor que o segundo modelo com menor erro nessa categoria (3Comb4e, SVR-ELM-RR combinados por meio da média ponderada com MAE) e 70,23% melhor que o

modelo que apresentou maior erro (2Comb2c, SVR-RR combinados por intermédio da média harmônica). Já para o horizonte de previsão de 60 min, o modelo 4Comb6f (SVR-ELM-GP-RR, combinados através da média ponderada com MAPE) obteve menor erro segundo a métrica RMSE, apresentando um erro de 1661,41 kWh e desempenho 4,80% melhor que o segundo modelo com menor erro (4Comb6e, SVR-ELM-GP-RR, combinados por meio da média ponderada com MAE) e 56,50% melhor que o modelo que apresentou maior erro (5Comb7c, SVR-ELM-GP-RR-RF, combinados através da média harmônica).

Continuando em WDF2, para o horizonte de previsão de 120 min, o modelo 4Comb6f (SVR-ELM-GP-RR, combinados através da média ponderada com MAPE) obteve menor erro segundo a métrica MAPE, apresentando um erro de 112,95% e desempenho 4,85% melhor que o segundo modelo com menor erro (4Comb6e, SVR-ELM-GP-RR combinados por intermédio da média ponderada com MAE), 77,05% melhor que o modelo que apresentou maior erro (4Comb6c, SVR-ELM-GP-RR combinados por meio da média harmônica), e 67,30% melhor que o segundo modelo de maior erro (5Comb7c, SVR-ELM-GP-RR-RF combinados através da média harmônica).

Portanto, para o parque WDF1, os modelos com melhor desempenho em termos de menor erro foram 2Comb1d (SVR-ELM, combinados através da média ponderada com RMSE), 3Comb4f (SVR-ELM-RR, combinados através da média ponderada com MAPE), 4Comb6f (SVR-ELM-GP-RR, combinados através da média ponderada com MAPE) e 4Comb6f (SVR-ELM-GP-RR, combinados através da média ponderada com MAPE) para os horizontes de previsão de 10, 30, 60 e 120 minutos, e segundo as métricas RMSE, MAE, MAPE e RMSE, respectivamente. Enquanto que para o parque WDF2 os *ensembles* com melhor desempenho foram 4Comb6f (SVR-ELM-GP-RR, combinados através da média ponderada com MAPE), 3Comb4f (SVR-ELM-RR, combinados através da média ponderada com MAPE), 4Comb6f (SVR-ELM-GP-RR, combinados através da média ponderada com MAPE) e 4Comb6f (SVR-ELM-GP-RR, combinados através da média ponderada com MAPE), para os horizontes de previsão de 10, 30, 60 e 120 minutos e em relação as métricas RMSE, MAE, RMSE e MAPE, respectivamente. A tabela 5 resume o descrito nesse parágrafo.

Tabela 5 – Modelos que apresentaram melhor desempenho em termos de menor erro em cada categoria

WDF	Horizonte de previsão (min)	Critério	Modelo	Combinação
1	10	RMSE	SVR - ELM	Ponderada com RMSE
	30	MAE	SVR - RR - ELM	Ponderada com MAPE
	60	MAPE	SVR - RR - GP - ELM	Ponderada com MAPE
	120	RMSE	SVR - RR - GP - ELM	Ponderada com MAPE
2	10	RMSE	SVR - RR - GP - ELM	Ponderada com MAPE
	30	MAE	SVR - RR - ELM	Ponderada com MAPE
	60	RMSE	SVR - RR - GP - ELM	Ponderada com MAPE
	120	MAPE	SVR - RR - GP - ELM	Ponderada com MAPE

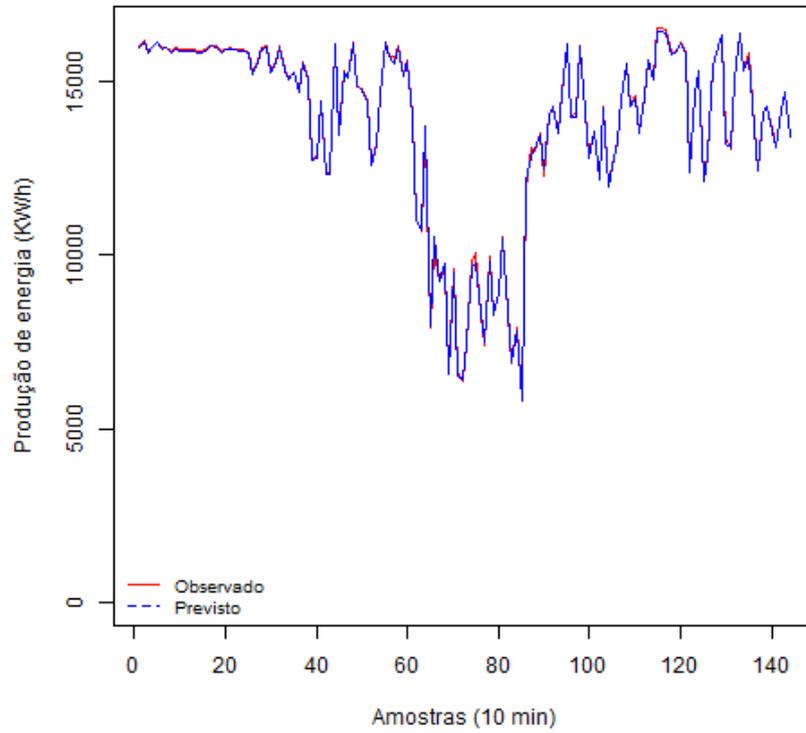
Fonte: Elaborado pela autora

Além disso, a combinação realizada através da média ponderada MAPE foi considerada a com melhor desempenho em 87,5% das situações. Enquanto que a combinação realizada por meio da média harmônica apresentou o maior erro em relação às outras em 75% das situações.

4.4 Análise gráfica

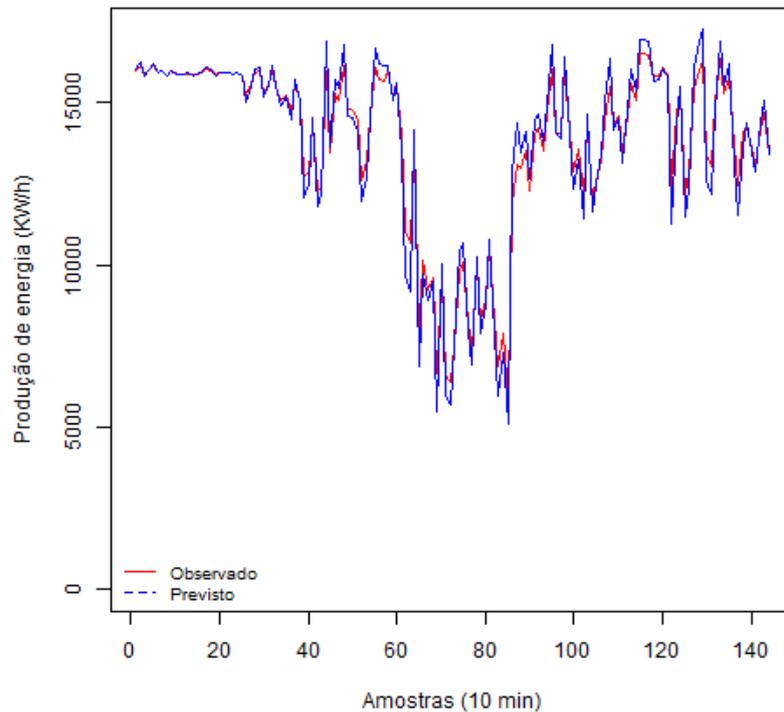
A Figura 10 mostra a comparação entre as séries temporais dos valores observados e dos valores previstos com horizonte de previsão de 10 min por meio do *ensemble 2Comb1d* (SVR-ELM combinados através da média ponderada com RMSE), para o WDF1. Sendo que a linha vermelha representa os dados observados e a linha pontilhada azul os valores previstos. Pode-se perceber que os valores são muito próximos e que ambas apresentam comportamento semelhante. Já a Figura 11 mostra a comparação entre as séries temporais dos valores observados (linha vermelha) e dos valores previstos (linha pontilhada azul) com horizonte de previsão de 30 min obtida por meio do *ensemble 3Comb4f* (SVR-ELM-RR combinados por intermédio da média ponderada com MAPE), para o WDF1. Pode-se perceber que os valores são muito próximos, mas que os mesmos se distanciam nos pontos máximos e mínimos do gráfico.

Figura 10 - Valores observados e previstos em WDF1: horizonte de previsão 10 min, ensemble 2Comb1d



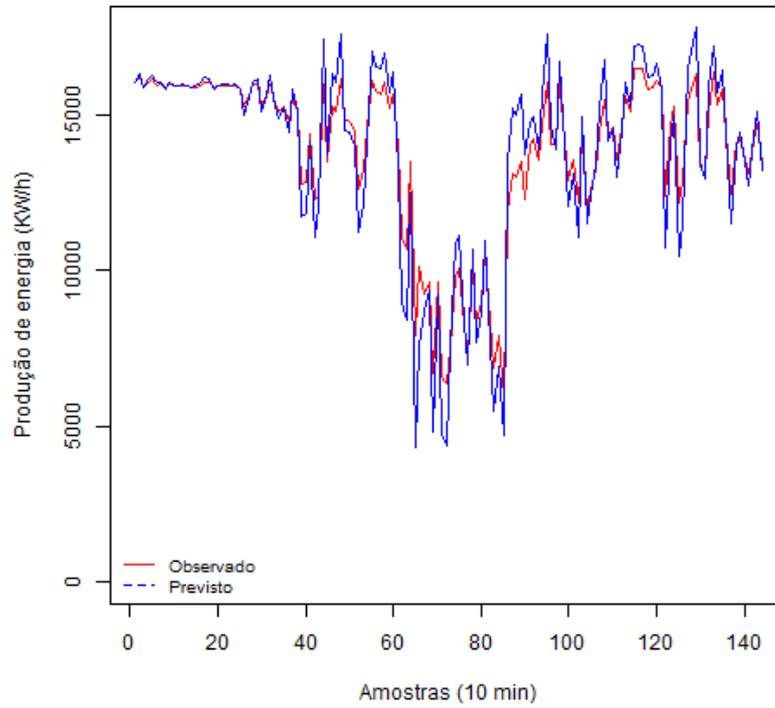
Fonte: Elaborado pela autora

Figura 11 - Valores observados e previstos em WDF1: horizonte de previsão de 30 min, ensemble 3Comb4f



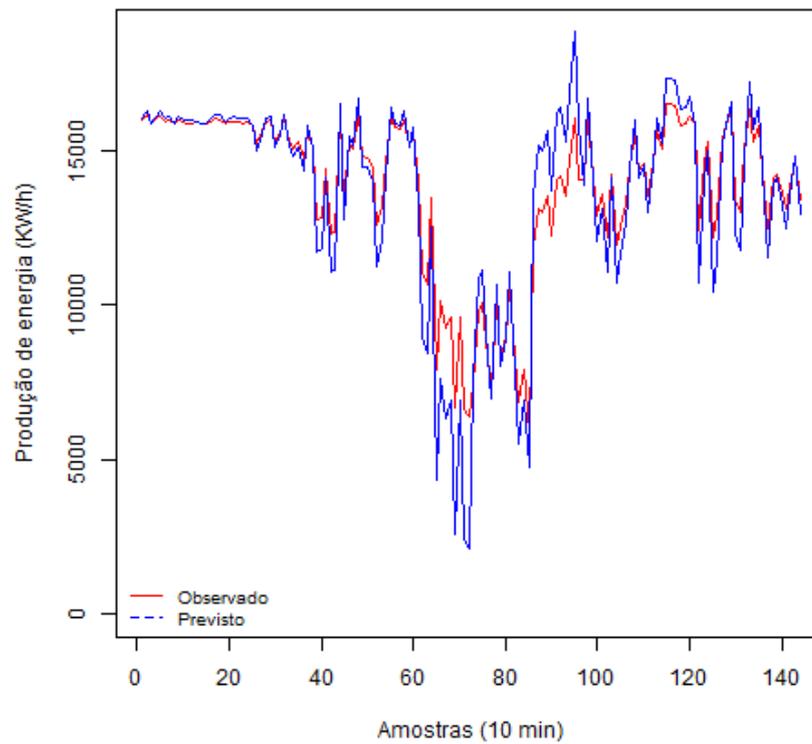
Fonte: Elaborado pela autora

Figura 12 - Valores observados e previstos em WDF1: horizonte de previsão de 60 min, ensemble 4Comb6f



Fonte: Elaborado pela autora

Figura 13 - Valores observados e previstos em WDF1: horizonte de previsão de 120 min, ensemble 4Comb6f



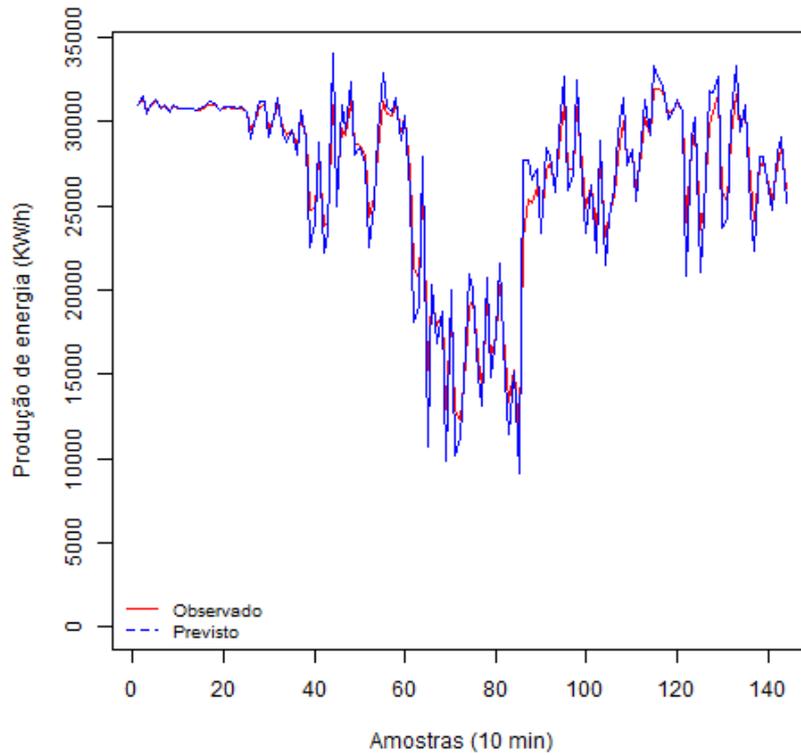
Fonte: Elaborado pela autora

As Figuras 12 e 13 mostram a comparação entre as séries temporais dos valores observados (linha vermelha) e dos valores previstos (linha pontilhada azul) com horizonte de previsão de 60 e 120 minutos, respectivamente, para o WDF1. Ambas previsões obtidas por meio do *ensemble 4Comb6f* (SVR-ELM-GP-RR combinados por intermédio da média ponderada com MAPE). Pode-se perceber que apesar de apresentarem o mesmo comportamento, existe distanciamento entre os valores nos pontos máximos e mínimos do gráfico.

A Figura 14 mostra a comparação entre as séries temporais dos valores observados e dos valores previstos com horizonte de previsão de 10 min por meio do *ensemble 4Comb6f* (SVR-ELM-GP-RR combinados através da média ponderada com MAPE), para o WDF2. Pode-se perceber que os valores são muito próximos e que ambas apresentam comportamento semelhante. Já a Figura 15 mostra a comparação entre as séries temporais dos valores observados (linha vermelha) e dos valores previstos (linha pontilhada azul) com horizonte de previsão de 30 min obtida por meio do *ensemble 3Comb4f* (SVR-ELM-RR combinados por intermédio da média ponderada com MAPE), para o WDF2. Pode-se perceber que os valores são muito próximos, mas que os mesmos se distanciam nos pontos máximos e mínimos do gráfico.

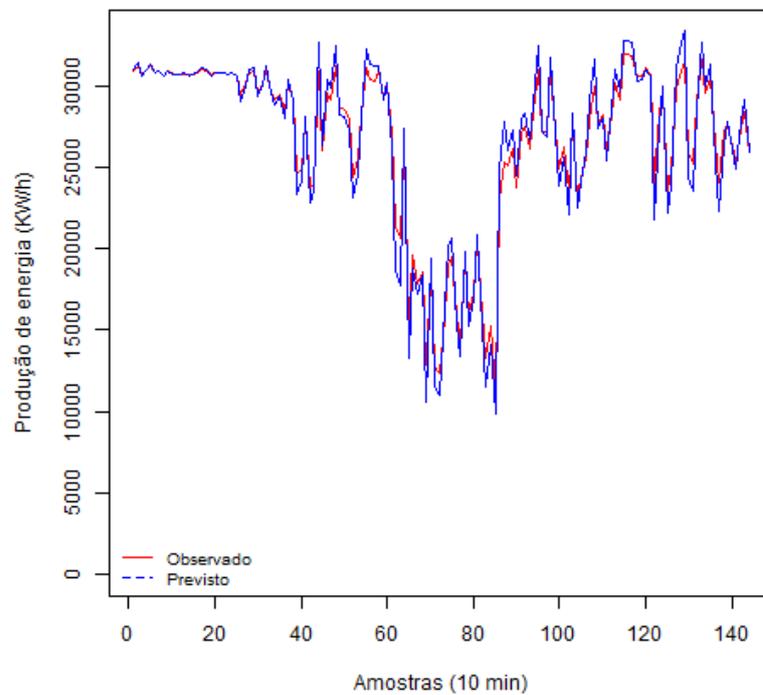
As Figuras 16 e 17 mostram a comparação entre as séries temporais dos valores observados (linha vermelha) e dos valores previstos (linha pontilhada azul) com horizonte de previsão de 60 e 120 minutos, respectivamente, para o WDF2. Ambas previsões obtidas por meio do *ensemble 4Comb6f* (SVR-ELM-GP-RR combinados por intermédio da média ponderada com MAPE). Pode-se perceber que apesar de apresentarem o mesmo comportamento, existe distanciamento entre os valores nos pontos máximos e mínimos do gráfico.

Figura 14 - Valores observados e previstos em WDF2: horizonte de previsão de 10 min, *ensemble 4Comb6f*



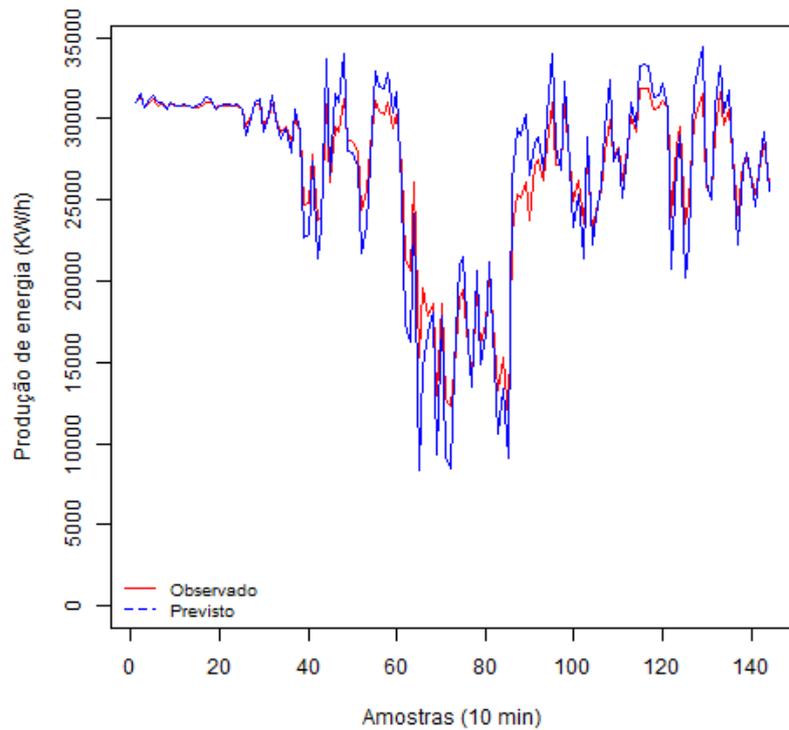
Fonte: Elaborado pela autora

Figura 15 - Valores observados e previstos em WDF2: horizonte de previsão de 30 min, *ensemble 3Comb4f*



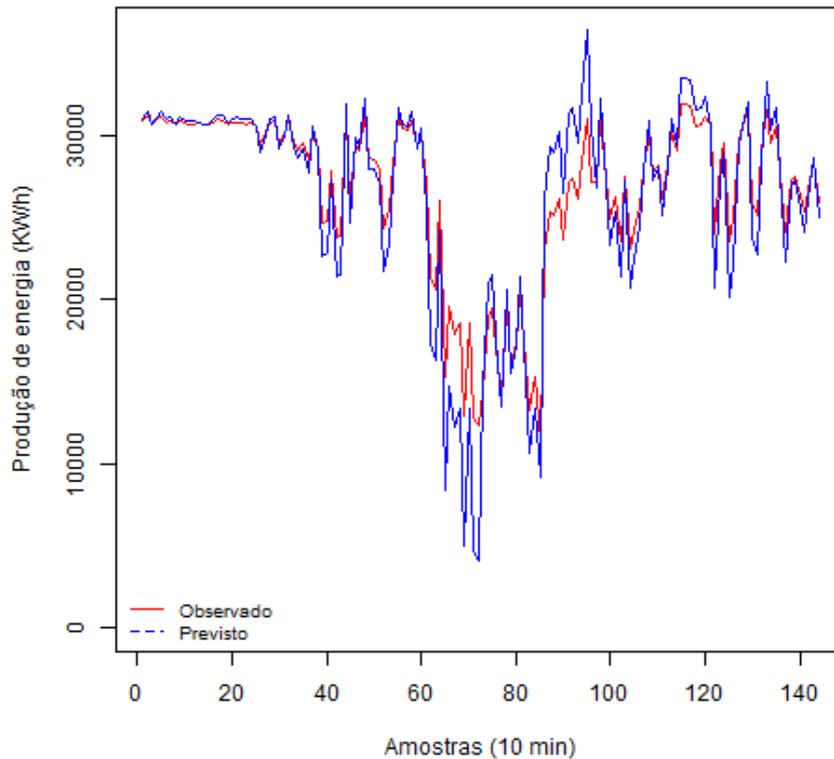
Fonte: Elaborado pela autora

Figura 16 - Valores observados e previstos em WDF2: horizonte de previsão de 60 min, ensemble 4Comb6f



Fonte: Elaborado pela autora

Figura 17 - Valores observados e previstos em WDF2: horizonte de previsão de 120 min, ensemble 4Comb6f



Fonte: Elaborado pela autora

4.5 Comparação com as técnicas individuais

Dentre os modelos individuais utilizados nesse trabalho (SVR, ELM, RR, GP e RF) foram escolhidos os que apresentavam melhor desempenho em termo de menor erro em cada uma das categorias, isto é, para os horizontes de 10 min, 30 min, 60 min e 120 min, dos parques WDF1 e WDF2. Os modelos designados de cada categoria foram comparados com as combinações de melhor desempenho em termos de menor erro obtidas nesse estudo.

Para o parque WDF1, o modelo individual que obteve menor erro em termos de RMSE para o horizonte de previsão de 10 min foi o SVR, apresentando erro de 1.191,70 kWh. Assim, o modelo 2Comb1d (SVR-ELM combinados através da média ponderada com RMSE) apresentou desempenho 89,71% melhor, com erro de 122,59 kWh em termos de RMSE. Já para o horizonte de previsão de 30 min, o modelo individual que obteve menor erro em termos de MAE foi o SVR, apresentando erro de 1.030,43 kWh. Logo, o modelo 3Comb4f (SVR-ELM-RR combinados por intermédio da média ponderada com MAPE) apresentou desempenho 68,87% melhor, com erro de 320,73 kWh em termos da métrica MAE.

Ainda no parque WDF1, o modelo individual que obteve menor erro em termos de MAPE para o horizonte de previsão de 60 min foi o GP, apresentando erro de 18,76%. O mesmo obteve desempenho 70,57% melhor que o modelo 4Comb6f (SVR-ELM-GP-RR combinados através da média ponderada com MAPE), cujo qual apresentou erro de 63,75% em termos de MAPE. Por fim, para o horizonte de previsão de 120 min, o modelo individual SVR obteve menor erro em termos de RMSE dentre os modelos individuais citados no início dessa seção, exibindo erro de 2.430,56 kWh. Desse modo, o modelo 4Comb6f (SVR-ELM-GP-RR combinados através da média ponderada com MAPE) apresentou desempenho 50,71% melhor, com erro de 1.198,06 kWh em termos da métrica RMSE.

Para o parque WDF2, o modelo individual que obteve menor erro em termos de RMSE para o horizonte de previsão de 10 min dentre os modelos individuais citados no início dessa seção foi SVR, apresentando erro de 2.303,97 kWh. Deste modo, o modelo 4Comb6f (SVR-ELM-GP-RR combinados através da média ponderada com MAPE) apresentou desempenho 51,99% melhor, com erro de 1.106,24 kWh em termos de RMSE. Já para o horizonte de previsão de 30 min, o modelo SVR obteve menor erro em termos de MAE dentre os modelos individuais

citados no início dessa seção, apresentando erro de 1.992,21 kWh. Diante disso, o modelo 3Comb4f (SVR-ELM-RR combinados por intermédio da média ponderada com MAPE) apresentou desempenho 69,83 % melhor, com erro de 601,11 kWh em termos da métrica MAE.

Ainda no parque WDF2, o modelo que obteve menor erro para o horizonte de previsão de 60 min em termos de RMSE foi o SVR, com o valor de 3.455,34 kWh de erro. Por conseguinte, o modelo 4Comb6f (SVR-ELM-GP-RR combinados através da média ponderada com MAPE) apresentou desempenho 51,92% melhor, com erro de 1.661,41 kWh em termos da métrica RMSE. Por outro lado, o modelo que obteve menor erro em termos de MAPE para o horizonte de previsão de 120 min dentre os modelos individuais citados no início dessa seção foi o GP, exibindo 22,12% de erro. O mesmo obteve desempenho 80,42% melhor que o modelo 4Comb6f (SVR-ELM-GP-RR combinados através da média ponderada com MAPE), cujo qual apresentou erro de 112,95% em termos de MAPE.

Contudo, percebe-se que as combinações desenvolvidas nesse estudo citadas acima têm menor erro com relação a métrica RMSE e MAE em comparação com os modelos individuais. Por outro lado, os modelos individuais comparados acima apresentam menor erro em termos da métrica MAPE em comparação com as combinações adotadas nesse estudo.

4.6 Comparação entre as previsões obtidas com o trabalho de referência

Os modelos com melhor desempenho em termos de menor erro obtidos em Ribeiro e colaboradores (2022) foram comparados com os modelos com desempenho superior em termos de menor erro obtidos neste trabalho em cada uma das categorias, isto é, para os horizontes de 10 min, 30 min, 60 min e 120 min, dos parques WDF1 e WDF2.

Para o parque WDF1, o modelo que obteve menor erro em termos de RMSE para o horizonte de previsão de 10 min no artigo de referência foi o *Stacking+Bagging* (100 *Bootstraps*) combinado por média ponderada (1.094,12 kWh). Assim, o modelo 2Comb1d (SVR-ELM combinados através da média ponderada com RMSE) apresentou desempenho 88,8% melhor, com erro de 122,59 kWh em termos de RMSE. Já para o horizonte de previsão de 30 min, o modelo que obteve menor erro em termos de MAE no artigo de referência foi *Stacking+Bagging* (100 *Bootstraps*) combinados por média aritmética (932,19 kWh). Logo, o modelo 3Comb4f (SVR-ELM-

RR combinados por intermédio da média ponderada com MAPE) apresentou desempenho 65,59% melhor, com erro de 320,73 kWh em termos da métrica MAE.

Ainda no parque WDF1, o modelo que obteve menor erro em termos de MAPE para o horizonte de previsão de 60 min em Ribeiro e colaboradores (2022) foi o *Stacking+Bagging* (30 *Bootstraps*) combinado por média ponderada (17,13%). O mesmo obteve desempenho 73,13% melhor que o modelo 4Comb6f (SVR-ELM-GP-RR combinados através da média ponderada com MAPE), cujo qual apresentou erro de 63,75% em termos de MAPE. Por fim, para o horizonte de previsão de 120 min, o modelo que obteve menor erro em termos de RMSE no artigo de referência foi *Stacking+Bagging* (100 *Bootstraps*) combinados por média ponderada (2.123,97 kWh). Logo, o modelo 4Comb6f (SVR-ELM-GP-RR combinados através da média ponderada com MAPE) apresentou desempenho 43,59% melhor, com erro de 1.198,06 kWh em termos da métrica RMSE.

Para o parque WDF2, o modelo que obteve menor erro em termos de RMSE para o horizonte de previsão de 10 min no artigo de referência foi o *Stacking+Bagging* (100 *Bootstraps*) combinado por média ponderada (2.116,91 kWh). Deste modo, o modelo 4Comb6f (SVR-ELM-GP-RR combinados através da média ponderada com MAPE) apresentou desempenho 47,74% melhor, com erro de 1.106,24 kWh em termos de RMSE. Já para o horizonte de previsão de 30 min, o modelo que obteve menor erro em termos de MAE no artigo de referência foi *Stacking+Bagging* (100 *Bootstraps*) combinados por média aritmética (1.801,82 kWh). Diante disso, o modelo 3Comb4f (SVR-ELM-RR combinados por intermédio da média ponderada com MAPE) apresentou desempenho 66,64 % melhor, com erro de 601,11 kWh em termos da métrica MAE.

Ainda no parque WDF2, o modelo que obteve menor erro para o horizonte de previsão de 60 min em termos de RMSE no artigo de referência foi *Stacking+Bagging* (100 *Bootstraps*) combinados por média ponderada (2962,13 kWh). Por conseguinte, o modelo 4Comb6f (SVR-ELM-GP-RR combinados através da média ponderada com MAPE) apresentou desempenho 43,91% melhor, com erro de 1.661,41 kWh em termos da métrica RMSE. Por outro lado, o modelo que obteve menor erro em termos de MAPE para o horizonte de previsão de 120 min em Ribeiro e colaboradores (2022) foi o *Stacking+Bagging* (30 *Bootstraps*) combinado por média ponderada (20,84%). O mesmo obteve desempenho 81,55% melhor que o modelo 4Comb6f (SVR-ELM-GP-

RR combinados através da média ponderada com MAPE), cujo qual apresentou erro de 112,95% em termos de MAPE.

Contudo, percebe-se que combinações adotadas nesse têm menor erro com relação a métrica RMSE e MAE em comparação com os modelos admitidos pelo artigo de referência e citados acima. Por outro lado, os modelos adotados por Ribeiro e colaboradores (2022) apresentam menor erro em termos da métrica MAPE quando comparados com as combinações desenvolvidas nesse trabalho.

5 CONCLUSÃO

No presente trabalho, quarenta e dois *ensembles* foram propostos para a previsão de séries temporais de energia eólica. Os *ensembles* foram gerados a partir de diferentes modelos base combinados por meio de: média aritmética, média harmônica, mediana e média ponderada (com pesos definidos por intermédio das métricas RMSE, MAE e MAPE).

Os *ensembles* foram aplicados a fim de gerar a previsão de série temporal de energia eólica para horizontes de previsão de 10, 30, 60 e 120 minutos para dois parques eólicos WDF1 e WDF2, localizados na Bahia, Brasil. Para o parque WDF1, os modelos com melhor desempenho foram 2Comb1d (SVR-ELM, combinados através da média ponderada com RMSE), 3Comb4f (SVR-ELM-RR, combinados através da média ponderada com MAPE), 4Comb6f (SVR-ELM-GP-RR, combinados através da média ponderada com MAPE) e 4Comb6f (SVR-ELM-GP-RR, combinados através da média ponderada com MAPE) para os horizontes de previsão de 10, 30, 60 e 120 minutos, respectivamente.

E para o parque WDF2 os *ensembles* com melhor desempenho foram 4Comb6f (SVR-ELM-GP-RR, combinados através da média ponderada com MAPE), 3Comb4f (SVR-ELM-RR, combinados através da média ponderada com MAPE), 4Comb6f (SVR-ELM-GP-RR, combinados através da média ponderada com MAPE) e 4Comb6f (SVR-ELM-GP-RR, combinados através da média ponderada com MAPE), para os horizontes de previsão de 10, 30, 60 e 120 minutos, respectivamente.

Além disso, a combinação realizada através da média ponderada MAPE foi considerada a com melhor desempenho em termos de menor erro em 87,5% das situações. Enquanto que a combinação realizada por meio da média harmônica apresentou maior erro em relação às outras em 75% das situações. Ainda temos que o *ensemble* 4Comb6f (SVR-ELM-GP-RR, combinados através da média ponderada com MAPE) foi considerado com o melhor desempenho em termos de menor erro em 62,5% das situações, seguido de 3Comb4f (SVR-ELM-RR, combinados através da média ponderada com MAPE) que foi considerado com melhor desempenho em 25% das situações e, por fim 2Comb1d (SVR-ELM, combinados através da média ponderada com RMSE) considerado com melhor desempenho em 12,5% das situações.

Apesar dos resultados apresentados neste estudo serem promissores em termos de acurácia de previsão é possível identificar as seguintes oportunidades de estudos futuros: (i) desenvolver de uma revisão sistemática de literatura no que se refere a aplicação de métodos de combinação na previsão de séries temporais; (ii) estudar a aplicabilidade dos modelos adotados nesse estudo em bases de dados desses mesmos parques eólicos, nos meses de janeiro de anos anteriores, seriam interessantes para verificar se a eficácia desse método permanece para esse mês; (iii) realizar um estudo aplicando esses modelos na *dataset* utilizado por De Lima *et al.* (2016).

REFERÊNCIAS

- ANDRADE, Fillipe de A. *et al.* **Previsão da Geração de Energia Fotovoltaica Utilizando Inteligência Artificial em Séries Temporais.** In: Simpósio Brasileiro de Automação Inteligente-SBAI. 2021.
- Associação Brasileira de Energia Eólica - ABEEólica. **INFOWIND Brazil no. 25. 2021.** Acesso em 13 jun 2022. Disponível em: <https://abeeolica.org.br/wp-content/uploads/2022/04/2022_03_InfoVento25.pdf>.
- CARLOTTO, Giulia Bagatini. **Métodos de Machine Learning para previsão de Séries Temporais.** 2020.
- DE ASSIS PAIVA, Denise; SÁFADI, Thelma. **Study of tests for trend in time series.** *Revista brasileira de biometria*, v. 39, n. 2, p. 311-333, 2021.
- DE CASTRO, Marcelo Teixeira; DE SOUZA, Maciel Gama; DE OLIVEIRA CASTRO, Anderson. **Energias renováveis: energia eólica, seus efeitos e ganhos ambientais.** *ITEGAM-JETIA*, v. 5, n. 19, p. 103-108, 2019.
- DA SILVA, Francisco Eduardo Mendes *et al.* **Previsão de geração de energia elétrica renovável em curto prazo no estado do Ceará utilizando modelo de regressão prophet.** *Research, Society and Development*, v. 11, n. 7, p. e12711729579-e12711729579, 2022.
- DE FARIAS, João Gabriel *et al.* **Previsão de Geração no Médio Prazo em um Parque Eólico no Rio Grande do Sul utilizando o GFS e Redes Neurais Artificiais.**
- DE LIMA, José DONIZETTI *et al.* **A study of the performance of individual techniques and their combinations to forecast urban water demand.** *Revista ESPACIOS| Vol. 37 (Nº 22) Año 2016*, 2016.
- DE MATTOS PAIXÃO, Gabriela Miana *et al.* **Machine Learning na Medicina: Revisão e Aplicabilidade.** *Arquivos Brasileiros de Cardiologia*, v. 118, n. 1, p. 95, 2022.
- FRACCANABBIA, Naylene. **Previsão de séries temporais de fontes de energias renováveis aplicando abordagens de regressão.** 2021. Tese (Mestrado) - Curso de Engenharia Mecânica, Pontifícia Universidade Católica do Paraná, Curitiba, 2021
- GIL, Antonio Carlos, 1946. **Como elaborar projetos de pesquisa.** 6. Ed. São Paulo: Atlas, 2019.

- GEBIN, Luís Gustavo Gutierrez; SALGADO, Ricardo Menezes; NOGUEIRA, Denismar Alves. **Previsão de Energia Eólica: Modelo de Ensemble Baseado em Modelos Estatísticos e de Aprendizado de Máquina**. Research, Society and Development, v. 9, n. 12, p. e38291211251-e38291211251, 2020.
- GWEC – **Global Wind Energy Council**. *Global Wind Report 2022*. Disponível em: <<https://gwec.net/global-wind-report-2022/>>. Acessado em junho de 2022.
- HYNDMAN, R.J., & Athanasopoulos, G. (2021) **Forecasting: principles and practice**, 3rd edition, OTexts: Melbourne, Australia. OTexts.com/fpp3. Acesso em: 21 de junho de 2022.
- JIN, Huaiping *et al.* **Probabilistic wind power forecasting using selective ensemble of finite mixture Gaussian process regression models**. Renewable Energy, v. 174, p. 1-18, 2021.
- JUNIOR, João Bosco Verçosa Leal *et al.* **Modelo híbrido de previsão de séries temporais para possíveis aplicações no setor de geração eólica**.
- KARTHIKEYAN, M.; RENGARAJ, R. **Short-Term Wind Speed Forecasting Using Ensemble Learning**. In: 2021 7th International Conference on Electrical Energy Systems (ICEES). IEEE, p. 502-506, 2021.
- MARTINS, Fernando Ramos; GUARNIERI, Ricardo André; PEREIRA, Enio Bueno. **O aproveitamento da energia eólica**. Revista Brasileira de Ensino de Física, v. 30, p. 1304.1-1304.13, 2008.
- MELO, Denise S.; OLIVEIRA, Leonardo W.; OLIVEIRA, Janaína G. **Aplicação de Redes Neurais Artificiais para Previsão de Velocidade de Vento Considerando Variáveis Correlacionadas**. Simpósio Brasileiro de Sistemas Elétricos-SBSE, v. 1, n. 1, 2020.
- MORETTIN, P. A.; TOLOI, C. M. C. **Análise de séries temporais**. São Paulo: Edgar Blucher, 2006
- NASCIMENTO CAMELO, Henrique do *et al.* **Proposta para previsão de velocidade do vento através de modelagem híbrida elaborada a partir dos modelos ARIMAX e RNA**. Revista Brasileira de Meteorologia, v. 33, p. 115-129, 2018.
- NASCIMENTO, Raphael Santos; ALVES, Geziele Mucio. **Fontes alternativas e renováveis de energia no Brasil: Métodos e benefícios ambientais**. XX Encontro Latino Americano de Iniciação Científica, XVI Encontro Latino Americano de Pós-

Graduação e VI Encontro de Iniciação à Docência–Universidade do Vale do Paraíba, 2016.

PRESOTTO, Elen; TALAMINI, Edson. **O uso de recursos energéticos renováveis e não renováveis e sua influência na variação da renda nacional**. *Economia & Região*, v. 9, n. 2, p. 195-210.

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Disponível em:

<<https://www.R-project.org/>>. Acesso em: 28 out. 2022

RIBEIRO, Matheus Henrique Dal Molin *et al.* **Efficient bootstrap stacking ensemble learning model applied to wind power generation**

forecasting. *International Journal of Electrical Power & Energy Systems*, v. 136, p. 107-712, 2022.

SUÁREZ-CETRULO, Andrés L. *et al.* **Wind power forecasting using ensemble learning for day-ahead energy trading**. *Renewable Energy*, 2022.

VIEIRA, Ana Cândida Ferreira. **Energias renováveis e sua eficiência na nova economia energética no Brasil**. *Revista Brasileira de Gestão Ambiental e Sustentabilidade*, v. 8, n. 18, p. 211-223, 2021.

WOLFART, Simone Beatriz *et al.* **Estudo de métodos para previsão da produção de grãos no brasil via séries temporais**. *Cadernos do IME-Série Estatística*, v. 49, p. 17, 2020.

XING, Zhou *et al.* **Wind Speed Forecasting Model Based on Extreme Learning Machines and Complete Ensemble Empirical Mode Decomposition**. 5th Asia Conference on Power and Electrical Engineering (ACPEE). IEEE, p. 159-163, 2020.

YONG, Binbin *et al.* **Ensemble neural network method for wind speed forecasting**. IEEE International Workshop on Signal Processing Systems (SiPS). IEEE, p. 31-36, 2019.