

Exploração de Técnicas para Previsão da Velocidade do Vento no Longo prazo

Felipe Whitaker, Rafaela Ribeiro, Raphael Bittencourt, Fernando Cyrino

Departamento de Engenharia Industrial

Pontifícia Universidade Católica do Rio de Janeiro

R. Marquês de São Vicente, 225, Gávea, Rio de Janeiro, RJ, 22452-900

`felipewhitaker@aluno.puc-rio.br`,

`rafaelamaribeiro@aluno.puc-rio.br`, `raphadombit@gmail.com`,

`cyrino@puc-rio.br`

RESUMO

A geração eólica no Brasil e no mundo vem crescendo e, com ela, estudos sobre sua fonte também. Assim, com o objetivo de apoiar o planejamento da operação, previsões da velocidade do vento para o longo prazo são feitas utilizando duas classes de modelos: SARIMA e Redes Neurais, com o objetivo de testar a robustez dos modelos. Foram analisados dois locais, Parque Eólico de Praia Formosa no Ceará (CE) e Usina Eólica Elebrás Cidreira 1 no Rio Grande do Sul (RS). Para ambos os parques, os melhores modelos foram modelos estatísticos, alcançando RMSE de 1,142 m/s e 1,730 m/s, respectivamente; enquanto o melhor modelo de Redes Neurais, uma rede com uma camada escondida de 100 neurônios, teve como RMSE 3,913 m/s e 4,186 m/s, respectivamente.

PALAVRAS CHAVE. Velocidade do Vento. Previsão. Incerteza. Séries Temporais. Redes Neurais.

ABSTRACT

The wind generation in Brazil and worldwide is increasing, and with it, studies on its source. To support the planning of operations, long-term wind speed predictions are made using two classes of models: SARIMA and Neural Networks, with the aim of testing the robustness of the models. Two locations were analyzed, Praia Formosa Wind Park in Ceará (CE) and Elebrás Cidreira 1 Wind Farm in Rio Grande do Sul (RS). For both parks, the best models were statistical models, achieving RMSE values of 1,142 m/s and 1,730 m/s, respectively; while the best neural network model, a network with a hidden layer of 100 neurons, had RMSE values of 3,913 m/s and 4,186 m/s, respectively.

KEYWORDS. Wind Speed. Prediction. Uncertainty. Time Series. Neural Networks.

1. Introdução

No final de 2022, de acordo com a *Energy Research Office* [EPE, 2023], 47,4% da Oferta Interna de Energia foi proveniente de fontes renováveis, enquanto que a participação de fontes renováveis na Matriz Elétrica Brasileira foi de 92,1% de renovabilidade (geração nacional mais importação líquida), considerando apenas o Sistema Interligado Nacional (SIN). Destes, 61,9% foi originada da fonte hídrica, 11,8% da eólica, 8,0% da biomassa e 4,4% da solar. Ainda, durante 2022, as fontes hídrica e solar tiveram aumentos consideráveis nas suas contribuições individuais para a Matriz Elétrica Brasileira; enquanto as contribuições de todas as fontes não renováveis, carvão e seus produtos como o gás natural e nuclear, decresceram.

A geração eólica teve 12,9% de crescimento de 2021 a 2022 [EPE, 2023], tornando-se a segunda maior fonte de energia no Brasil. Um total de 109 parques eólicos foram construídos durante esse tempo, todos na Região Nordeste [ABEEOLICA, 2023]. De acordo com [ABEEOLICA, 2024], os cinco estados com maiores produções em Março de 2024 foram Rio Grande do Norte (RN), Bahia (BA), Piauí (PI), Ceará (CE) e Rio Grande do Sul (RS), onde os quatro primeiros estão localizados na Região Nordeste e o último na Região Sul. Ainda, no total, o subsistema Nordeste gerou mais energia do que consumiu, mostrando sua relevância na geração eólica, enquanto que no subsistema Sul 16,96% da energia consumida foi proveniente da geração eólica [ABEEOLICA, 2023].

Dessa forma, estudos sobre o comportamento do vento tornam-se de extrema importância para o setor elétrico, tornando a análise e a previsão das séries temporais de velocidade do vento no curto prazo necessárias para o planejamento da operação.

Esse estudo visa comparar as previsões realizadas a partir de dois tipos de modelos para previsão de séries temporais. Os modelos SARIMA, da família Box & Jenkins, são os modelos clássicos, amplamente utilizados por possuírem uma metodologia bem definida e apresentar boas soluções [Hyndman, 2021]; e redes neurais, que são modelos flexíveis de caixa opaca capazes de aproximar funções arbitrárias [Hornik et al., 1989]. Aqui se faz importante ressaltar que todos os modelos são treinados com valores autorregressivos, mas apenas o primeiro modelo inclui informação de sazonalidade. Para isso, são realizadas previsões diárias até 14 dias a frente, considerado de longo prazo, para analisar se os modelos conseguem ajudar no planejamento da operação, considerando ambos os tipos de modelos para previsão de séries temporais.

O restante do artigo é dividido em: na Seção 2 apresenta-se a revisão da literatura, na Seção 3 tem-se uma explicação sobre os dados e o pré-processamento, na Seção 4 tem-se uma breve explicação sobre os métodos, na Seção 5 pode-se ver os resultados, e, por fim, a Seção 6 traz a conclusão.

2. Revisão da literatura

A geração de energia eólica tem recebido crescente investimento ao longo dos anos, mas conta com um importante desafio: enquanto métodos tradicionais, como geração térmica, possuem uma geração previsível, a geração eólica depende do vento, um recurso incerto e fora do controle humano [Giebel, 2003]. Assim, se torna importante fazer boas previsões da velocidade do vento por dois motivos: melhorar a eficiência da operação, e reduzir a necessidade de equilibrar a energia dos mercados de reserva, melhorando também a lucratividade do parque eólico [Algarvio et al., 2019].

Segundo [Yang et al., 2020] as técnicas de previsão de vento podem ser divididas em três categorias distintas. A primeira, conhecida como abordagens físicas, utilizam modelos numéricos (*Numerical Weather Prediction*, abreviado NWP). A segunda categoria, conhecida como abordagem estática, assume uma estrutura e estima parâmetros utilizando o histórico da série de vento. Por fim, a abordagem híbrida combina dois ou mais modelos para que desta maneira se obtenha um resultado mais satisfatório do que se alcançaria ao utilizar apenas um modelo.

[Soman et al., 2010] destacam que a previsão da velocidade podem ser divididas em quatro categorias baseadas no período da previsão:

- Previsão de curtíssimo prazo: De alguns segundos a 30 minutos à frente.
- Previsão de curto prazo: De 30 minutos a 6 horas à frente.
- Previsão de médio prazo: De 6 horas a 1 dia à frente.
- Previsão de longo prazo: De 1 dia a 1 semana à frente

[Torres et al., 2005] utilizaram um modelo ARMA para prever a velocidade do vento das próximas 10 horas em cinco locais diferentes da Espanha obtendo um RMSE menor que 1,5 m/s e, desta maneira, mostraram que o ARMA pode ser um modelo eficiente a ser considerado para previsões de curto prazo. [Yatiyana et al., 2017] utilizaram um modelo ARIMA para prever a velocidade do vento na costa oeste da Austrália. Usando dados históricos para verificar o desempenho do modelo, foi possível perceber que o mesmo alcançou um erro menor que 5% quando se observava a velocidade do vento e menor que 16% na direção do vento, mostrando-se um resultado satisfatório considerando a simplicidade do modelo.

No campo de *machine learning*, [Filik and Filik, 2017] implementaram uma Rede Neural Artificial (*artificial neural network*, abreviado ANN) considerando diferentes aspectos para realizar a previsão da velocidade do vento em Eskisehir, tais como: velocidade do vento, temperatura e pressão atmosférica. Foi obtido um resultado satisfatório na análise realizada. Um aspecto importante da previsão do vento é que na maioria dos casos é muito importante ter a série histórica para realizar a previsão. Entretanto, em alguns casos, não é possível ter esse dado de entrada, aumentando a possibilidade de se obter resultados ruins. Com base nessa realidade, [Hu et al., 2016] introduziram um modelo de redes neurais profundas, treinadas por dados de locais conhecidos, para extrair padrões de velocidade do vento de locais conhecidos e assim ter uma melhor previsão em locais que não possuem dados, obtendo resultados bastante satisfatórios nos casos analisados.

3. Dados

A seguir, apresentamos a obtenção e o tratamento dos dados de série temporal de velocidade do vento na localidade de ambas as usinas.

3.1. Sobre os dados

Duas usinas eólicas no Brasil foram selecionadas. A primeira está localizada no litoral do Ceará, nomeada Parque Eólico de Praia Formosa que tem 105,00 kW de potência outorgada; e a segunda no litoral do Rio Grande do Sul, Usina Eólica Elebrás Cidreira 1 com 70,000 kW de potência outorgada. Dada a distância entre as usinas, é esperado que as séries não tenham relação. Após a coleta dos dados, os perfis de ambas as localizações foram analisados. Na Figura 1, é possível ver a localização das duas usinas no Brasil.

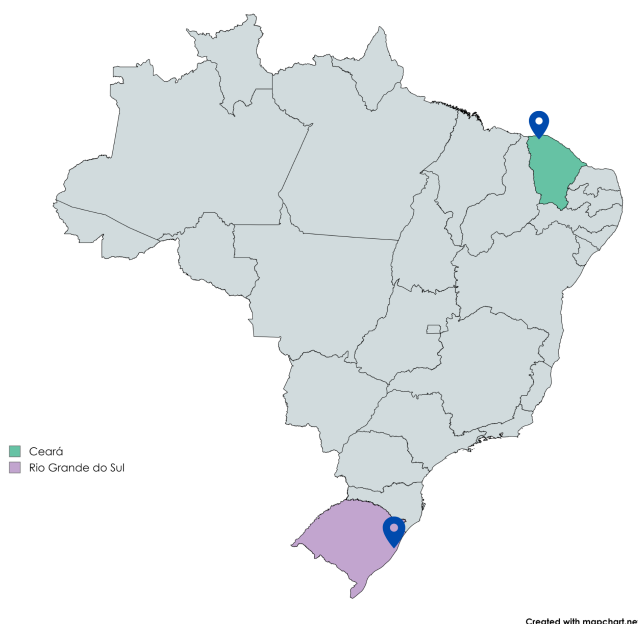


Figura 1: Localização das usinas escolhidas para realizar o estudo

Os dados usados foram coletados através do *Renewables Ninja* [Staffell and Pfenninger, 2016]. O *Renewables Ninja* disponibiliza diferentes conjuntos de dados, dentre eles velocidade do vento e geração horários. Esses dados têm como fonte o conjunto de reanálise *MERRA-2 (global)*, da NASA, e o próprio site trata algumas limitações previamente estudadas [Staffell and Pfenninger, 2016]. Os dados foram medidos considerando a turbina Vestas V15 4000 a uma altura de 100 metros, de 2018 a 2023. Ademais, [de Aquino Ferreira et al., 2022] estuda a qualidade dos dados de reanálise referentes ao território Brasileiro.

Embora os dados disponíveis sejam horários, toda análise desse trabalho se faz sobre dados diários. Para o agrupamento, de horário para diário, é feita a média das velocidades do vento. Esse agrupamento mantém a sazonalidade, mesmo que os padrões intradiários tenham sido perdidos. Finalmente, são analisados seis anos, 2018 à 2023, ou 1825 dias. As Figuras 2 e 3 representam as séries temporais de Praia Formosa e Elerbás Cidreira dos seis anos em análise, de 2018 à 2023.

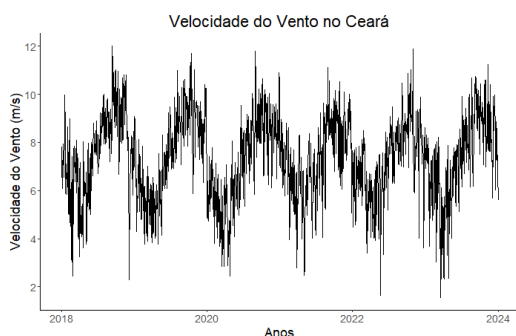


Figura 2: Série Temporal da Velocidade do Vento no CE

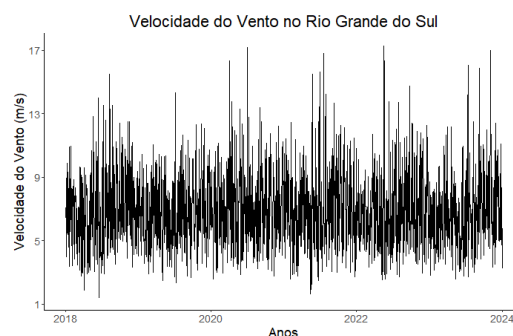


Figura 3: Série Temporal da Velocidade do Vento no RS

3.2. Estatísticas das Séries

Nesta Seção, analisamos os histogramas das séries e suas funções de autocorrelação (ACF) e funções de autocorrelação parcial (PACF).

3.2.1. Parque Eólico de Praia Formosa - CE

A série do Ceará possui: uma ACF que decai lentamente e uma PACF que decai rapidamente com um leve pico na defasagem sazonal, indicando um processo autorregressivo. Além disso, seu histograma indica que a variável não é normal, com curtose inferior a 3 (2,70) e simetria diferente de zero (-0,25). Por fim, foram realizados os testes Kwiatkowski-Phillips-Schmidt-Shin (KPSS) e o teste de Phillips-Perron (PP), para calcular as métricas da série, concluindo que a série é estacionária.

Além disso, foram feitas a decomposição da série pelo método aditivo (Figura 4) indicando uma pequena tendência e clara sazonalidade; e a decomposição por ano (Figura 5) indicando que há um perfil semelhante entre os anos.

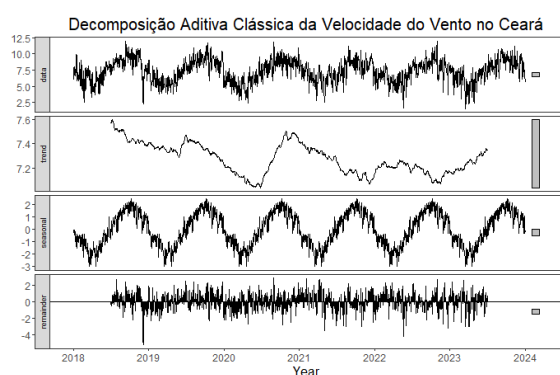


Figura 4: Decomposição Aditiva

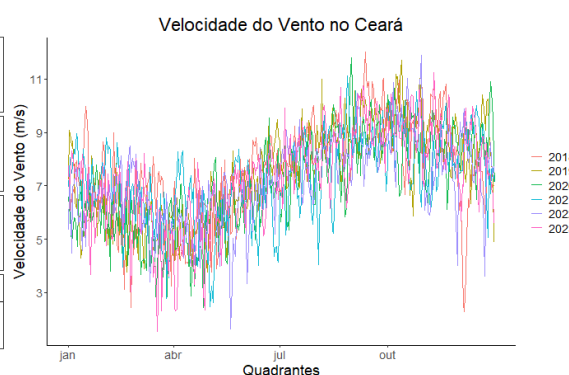


Figura 5: Decomposição Anual

3.2.2. Usina Eólica Elebrás Cidreira - RS

Já a série do Rio Grande do Sul não é tão bem comportada. Analisando o ACF e o pACF, temos que o ACF decai rapidamente, chegando a valores insignificantes a partir do *lag* 3 e o PACF tem comportamento semelhante ao ACF, decaindo rapidamente, chegando a valores insignificantes a partir do *lag* 3. Além disso, seu histograma, como é o caso para Formosa, indica que a série não possui uma distribuição Normal: sua curtose é superior a 3 (3,66) e sua simetria não é zero (0,64). Da mesma maneira que anteriormente, os testes Kwiatkowski-Phillips-Schmidt-Shin (KPSS) e o teste de Phillips-Perron (PP) foram feitos, para calcular as métricas da série, e se conclui que a série é estacionária.

Por fim, as decomposições da série pelo método aditivo e por ano podem ser observadas nas Figuras 6 e 7, respectivamente. Infelizmente as visualizações indicam apenas que a série não possui padrões claros que podem ser explorados.

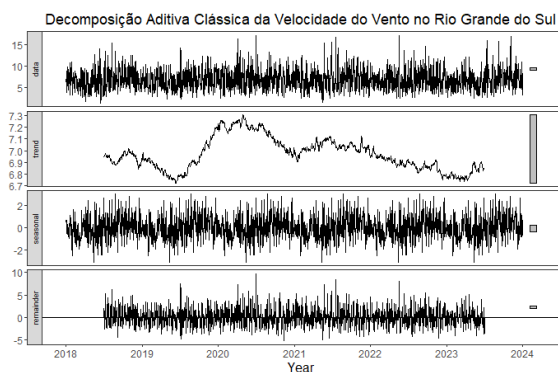


Figura 6: Decomposição Aditiva

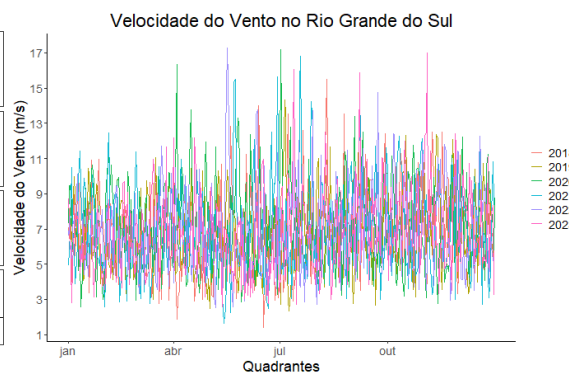


Figura 7: Decomposição Anual

4. Métodos

A seguir, é apresentada uma breve explicação sobre os métodos utilizados neste trabalho. O ARIMA é um método univariado e, portanto, é utilizado um modelo para cada série; a seção seguinte explora as redes neurais, que são modelos multivariados, recebendo ambas séries e fazendo a previsão de ambas simultaneamente.

4.1. Modelos ARIMA

Enquanto que os modelos autorregressivos AR(p) explicam a série em análise a partir de uma combinação linear entre p valores antigos da série, os modelos de média móvel MA(q), explicam a série em análise a partir de uma combinação linear entre q erros anteriores da série.

Ao juntar os dois modelos temos modelos autorregressivos de média móvel ARMA(p, q). Os mesmos explicam a série em análise a partir de uma combinação linear entre p valores antigos da série e os q erros anteriores da série, dada pelas fórmulas a seguir, onde ε_t é o erro da previsão:

$$Z_t = \sum_{k=1}^p \phi_k Z_{t-k} + \varepsilon_t - \sum_{j=1}^q \theta_j \varepsilon_{t-j} \quad (1)$$

Os modelos ARIMA(p, d, q) combinam os modelos ARMA com a diferenciação da série. d representa a ordem de diferenciação da série, onde $d = 1$ significa que a série foi diferenciada uma vez.

Por fim, modelos $SARIMA[(p, d, q)(P, D, Q)]_s$ são uma versão dos modelos ARIMA que levam em consideração a sazonalidade dos dados sendo que a parte (P, D, Q) , em que (P) representa a ordem da parte autorregressiva, (Q) a média móvel e (D) a diferenciação das partes sazonais da série [Hyndman, 2021].

4.2. Redes Neurais

Redes Neurais Artificiais são pilhas de regressões intercaladas por funções de ativação não lineares, tornando-as capazes de aproximar qualquer função contínua universal, dado que possuam graus de liberdade suficiente [Hornik et al., 1989]. No entanto, sua capacidade de generalização também traz dois desafios: (i) não há uma forma explícita para determinar a arquitetura de uma rede para o problema em questão, então diferentes redes precisam ser treinadas e avaliadas - elevando o custo computacional; (ii) dificuldade de interpretabilidade - potencialmente impedindo sua utilização em contextos que exigem a compreensão da razão por trás do resultado.

4.2.1. Transformação dos Dados

Uma etapa de extrema importância para o treinamento de redes neurais é o tratamento dos dados antes de iniciar o treinamento. Com o objetivo de restringir os valores dos gradientes propagados durante a otimização dos parâmetros, é recomendado que os valores sejam escalados para possuírem módulo inferior a um. Para esse fim, escolheu-se transformar os dados para o intervalo entre 0 e 1, seguindo a equação 2:

$$MinMax(y_i) = \frac{y_i - \min(Y)}{\max(Y) - \min(Y)} \quad (2)$$

Considerando Y a totalidade da série, e y_i os valores individuais da mesma. Ademais, $\min$ e $\max$ são funções que computam o mínimo e máximo de um conjunto.

Outras formas de adequar os dados ao intervalo sugerido poderiam ter sido usadas, inclusive transformações que alteram a distribuição dos dados, mas $MinMax$ foi escolhido pela simplicidade de implementação.

4.2.2. Arquiteturas Escolhidas

A atual Seção se dedica a determinação da estrutura da rede e outros hiperparâmetros importantes. Duas arquiteturas foram escolhidas, descritas nos itens abaixo. Ambas de redes neurais artificiais totalmente conectadas, ou seja, todos os neurônios de uma camada se conectam com a camada seguinte.

- DenseModel, uma rede sem camadas escondidas; e
- MLPModel, possui uma camada escondida com diferentes números de neurônios, [10, 50, 100].

Todas as configurações das redes foram treinadas utilizando cinco (5) passos autorregressivos, com o objetivo de permitir que a rede explore padrões não capturados por modelos menores. Além disso, todas as redes foram testadas com cinco (5) *seeds* distintas, [36, 64, 81, 100, 121].

Finalmente, com o objetivo de simular possíveis resultados para a série, foi utilizada a técnica de *bootstrapping* [Hyndman, 2021], na qual os resíduos do treino são inseridos a cada passo da predição, gerando um intervalo para as predições.

4.2.3. Treinamento

As diferentes redes foram treinadas para otimizar o Erro Quadrático Médio (MSE) utilizando o otimizador Adam Kingma and Ba [2015] com taxa de aprendizado (*learning rate*) igual a $1e-3$. Além disso, foi utilizado um tamanho de lote (*batch size*) de 64, máximo de 100 épocas (*epochs*), e Interrupção Precoce (*early stopping*) monitorando a perda da validação por até 5 épocas para uma alteração inferior a 0,001.

O treinamento foi feito em um computador com processador AMD Ryzen 5 5600G 3,90 GHz, 64GB de memória RAM e placa de vídeo RTX 3060 12GB. Para fins de informação, o principal gargalo para o treinamento foi a movimentação dos dados para a placa de vídeo, indicando que o tamanho do lote poderia ter sido maior.

4.3. Avaliação das Metodologias

Há várias métricas que avaliam a qualidade de uma simulação. O MAPE (*Mean Absolute Percentage Error*) e o MPE (*Mean percentage error*) indicam a qualidade da previsão por meio de porcentagem. O MSE (*Mean Squared Error*) calcula a média das diferenças entre as previsões e os valores observados ao quadrado, podendo dar mais peso a valores maiores. E o MAE (*Mean Absolute Error*) calcula a média das diferenças absolutas entre os dois casos. Todavia, essas métricas apresentam pontos negativos, quando comparadas com a métrica RMSE (*Root Mean Square Error*). O RMSE é um mais interpretável, pois essa a métrica tem a mesma unidade de medida que os

dados, e, dessa forma, foi escolhida para comparar a qualidade das simulações realizadas com cada um dos métodos descritos acima. A seguir, tem-se a equação do RMSE.

$$\text{RMSE}(y, \hat{y}) = \sqrt{\frac{\sum_{i=0}^{N-1} (y_i - \hat{y}_i)^2}{N}} \quad (3)$$

5. Resultados

O principal objetivo do presente estudo é realizar previsões. Para isso, a base de dados foi dividida em dois conjuntos. O conjunto de treino consiste nos dados desde Janeiro de 2018 até o final de Dezembro de 2022. O conjunto de teste consiste nos dados de Janeiro de 2023 até o final de Dezembro de 2023. Utilizando o conjunto de testes são feitas previsões no horizonte de 14 dias, para explorar como os modelos respondem para uma previsão de longo prazo.

5.1. Modelos Univariados

Para realizar a previsão dos modelos ARIMA foi utilizada a linguagem *R* e os pacotes *forecast* [Hyndman and Khandakar, 2008] e *tseries* [Trapletti and Hornik, 2020]. Diferentes modelos foram comparados com as métricas AIC, AICc e BIC. Como já apresentado, a série do Parque Eólico de Praia Formosa (CE) possui sazonalidade e, portanto, é adicionada a componente sazonal. Ao contrário, a série na Usina Elebrás Cidreira (RS) não possui forte sazonalidade. As Tabelas 1 e 2 apresentam os resultados dessa análise. O asterisco (*) indica modelo escolhido pelo *auto.arima*.

Tabela 1: Modelos testados para a série do Parque Eólico de Praia Formosa (CE)

Modelo	AIC	AICc	BIC	Modelo	AIC	AICc	BIC
ARIMA(1,0,0) x (0,1,0)	4731,31	4731,32	4741,89	ARIMA(1,0,0)	8051,14	8051,15	8067,67
ARIMA(1,0,1) x (0,1,0)	4692,51	4692,52	4708,36	ARIMA(1,0,1)	8027,16	8027,18	8049,19
ARIMA(2,0,0) x (0,1,0)	4695,60	4695,61	4711,46	ARIMA(2,0,0)	8022,81	8022,83	8044,85
ARIMA(2,0,1) x (0,1,0)	4694,50	4694,53	4715,65	ARIMA(2,0,1)*	8022,05	8022,09	8049,60
ARIMA(3,0,0) x (0,1,0)*	4693,58	4693,61	4714,73	ARIMA(3,0,0)	8023,26	8023,30	8050,81
ARIMA(3,0,1) x (0,1,0)	4692,91	4692,95	4719,34	ARIMA(3,0,1)	8022,38	8022,42	8055,43
ARIMA(4,0,0) x (0,1,0)	4694,92	4694,96	4721,35	ARIMA(4,0,0)	8022,77	8022,82	8055,83
ARIMA(4,0,1) x (0,1,0)	4694,71	4694,76	4726,42	ARIMA(4,0,1)	8022,27	8022,33	8060,83

Tabela 2: Modelos testados para a série da Usina Elebrás Cidreira (RS)

Para a série referente ao Ceará, o modelo sugerido pelo *auto.arima* não foi o que possui o melhor critério de avaliação, mas para a série referente ao Rio Grande do Sul o modelo sugerido pelo *auto.arima* obteve os melhores valores de AIC e AICc. Todavia, pode-se perceber que o *ARIMA*(2, 0, 0) é mais simples que o *ARIMA*(2, 0, 1), escolhido pelo *auto.arima* e apresenta métricas AIC e AICc extremamente próximas, além da métrica BIC ser consideravelmente menor. Assim, o *ARIMA*(1, 0, 1)x(0, 1, 0) e o *ARIMA*(2, 0, 0) serão os modelos considerados para as séries do Ceará e do Rio Grande do Sul, respectivamente. Definido o melhor modelo para cada série, são realizadas previsões para o horizonte de 14 dias. As Figuras 8 e 9 apresentam o resultado.

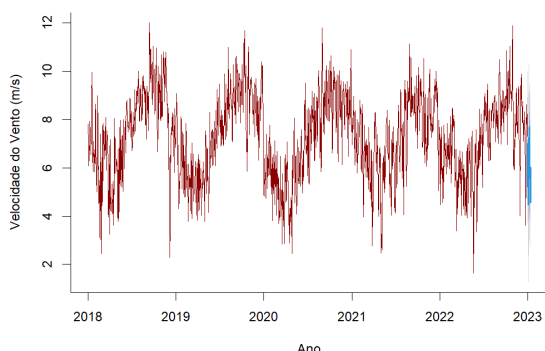


Figura 8: Previsão da série do Parque Eólico de Praia Formosa (CE) para 14 dias

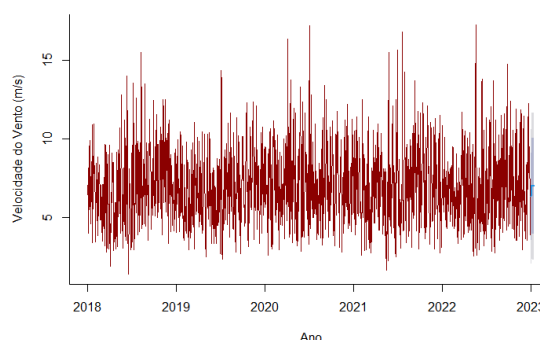


Figura 9: Previsão da série da Usina Elebrás Cidreira (RS) Cidreira para 14 dias

Além da análise visual, é importante analisar as métricas de avaliação das previsões. Podemos observar que os resultados são bastante similares, com exceção das métricas ME e MPE. O ME de Elebrás Cidreira é relativamente mais baixo que o de Praia Formosa enquanto que o MPE de Praia Formosa é bem mais baixo que o da Elebrás Cidreira. Esse fato é explicado pelo motivo de Elebrás Cidreira variar no entorno de uma média, fazendo os erros positivos e negativos se anularem. Já o MAPE mostra valores mais próximos.

Tabela 3: Métricas de avaliação para previsão das séries do Parque Eólico de Praia Formosa (CE) e da Usina Elebrás Cidreira (RS)

Previsão	Modelo	ME	RMSE	MAE	MPE (%)	MAPE (%)
Praia Formosa (14 dias)	SARIMA(1,0,1)(0,1,0)	-0,8669347	1,417645	1,134087	-17,06388	20,84723
Elebrás Cidreira (14 dias)	ARIMA(2,0,0)	-0,06813356	1,729731	1,466326	-1,069121	21,06972

5.2. Redes Neurais

Em adição à divisão dos dados descrito na Seção 5.1, o conjunto de treino foi dividido em dois: um novo conjunto de treino, e um conjunto de validação. O conjunto de validação é utilizado para escolher os hiperparâmetros - nesse caso, o tamanho da camada escondida - para então comparar com outros modelos.

A partir do processo de validação, foi escolhido o modelo MLPModel com 100 neurônios na camada escondida.

A figura 10 apresenta as perdas de treino e validação da melhor arquitetura. O intervalo apresentado existe porque está agregando o resultado das diferentes *seeds* utilizadas. É interessante notar que inicialmente a queda com a atualização dos pesos é bastante brusca, mas então se estabiliza conforme os parâmetros são atualizados.

Já a Figura 11 apresenta a distribuição dos erros do melhor modelo. É interessante notar que as distribuições são bastante viesadas. Além disso, como esperado dada a análise feita na Seção 3.2, Elebrás Cidreira possui uma variância superior a Praia Formosa.

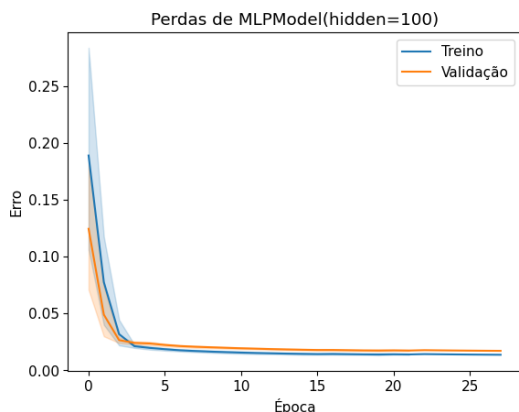


Figura 10: Erros de treino e validação por época

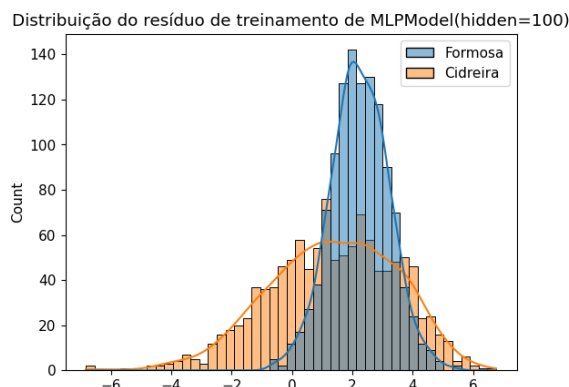


Figura 11: Distribuição do resíduo de treinamento do melhor modelo

Finalmente, a Figura 12 apresenta a previsão feita pela rede para as duas primeiras semanas do conjunto de teste. Os valores de RMSE para as séries de Formosa e Cidreira na melhor configuração foram 3,913 m/s e 4,186 m/s, respectivamente.

Embora os valores reais da séries estejam contidos dentro do intervalo simulado com *bootstrap*, as previsões praticamente persistem o último valor.

Previsão 14 dias à frente com MLPModel(hidden=100)

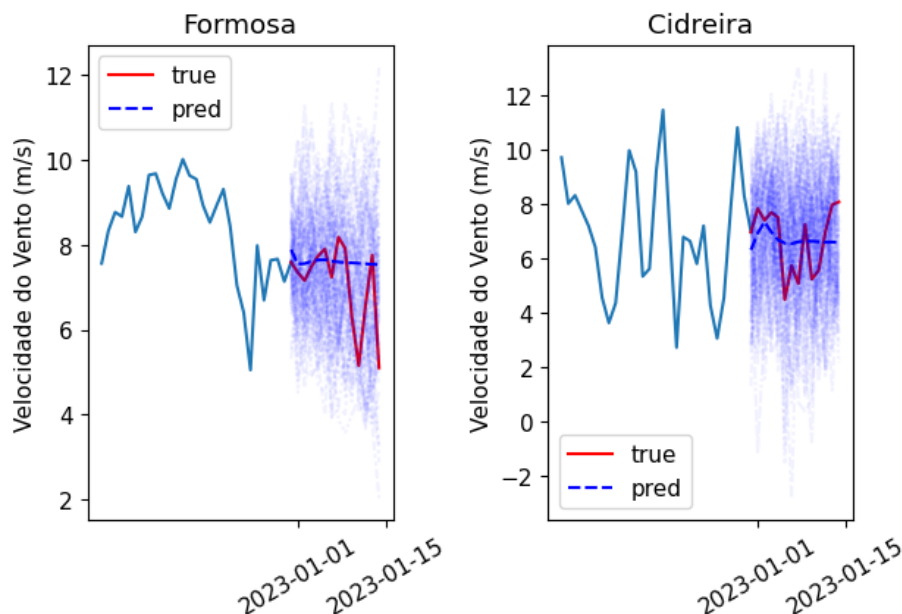


Figura 12: Previsão e Simulação 14 dias à frente

6. Conclusão

Neste trabalho, foi realizada a análise do comportamento de séries temporais da velocidade do vento em dois locais com altas gerações eólicas no Brasil, uma no Ceará e uma no Rio Grande do Sul. Foram executadas previsões de longo prazo, visando entender se os modelos conse-

guem auxiliar no planejamento da operação e se conseguem capturar o comportamento sazonal das séries.

O ARIMA se mostrou um método bastante eficaz, apresentando as melhores previsões. Ainda, vale destacar que a função *auto.arima* informou um modelo diferente do escolhido pelas métricas. Já o melhor modelo de Redes Neurais obteve resultados piores do que seria esperado dado sua flexibilidade, especialmente para a série da Usina Elebrás Cidreira (RS), já que os modelos clássicos teriam menos padrões claros para aprender. Uma razão possível é a distância física entre ambas as séries: como estão distantes, conhecer a velocidade de curto prazo de uma não favorece a previsão da outra.

Para trabalhos futuros com Redes Neurais para previsão de velocidade do vento, recomendamos usar séries diferenciadas ou modelar apenas a variação de um passo para o próximo. Outra possível análise seria estudar séries de velocidade do vento de locais mais próximos e com comportamentos semelhantes, como por exemplo Rio Grande do Norte e Ceará, para compararmos a semelhança dos modelos selecionados.

Referências

- ABEEOLICA (2023). Annual Wind Energy Report 2022. Technical report, ABEEólica.
- ABEEOLICA (2024). InfoWind March 2024. Technical report, ABEEólica.
- Algarvio, H., Lopes, F., do Couto, A. F., and Estanqueiro, A. (2019). Participation of wind power producers in day-ahead and balancing markets: An overview and a simulation-based study. *Wiley Interdisciplinary Reviews: Energy and Environment*, 8.
- de Aquino Ferreira, S. C., Oliveira, F. L. C., and Maçaira, P. M. (2022). Validation of the representativeness of wind speed time series obtained from reanalysis data for brazilian territory. *Energy*, 258:124746.
- EPE (2023). Brazilian Energy Balance: summary report 2023 - reference year 2022. Technical report, Energy Research Office – Brazil.
- Filik, Ü. B. and Filik, T. (2017). Wind speed prediction using artificial neural networks based on multiple local measurements in eskisehir. *Energy Procedia*, 107:264–269.
- Giebel, G. (2003). The state-of-the-art in short-term prediction of wind power. a literature overview.
- Hornik, K., Stinchcombe, M., and White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366. ISSN 0893-6080. URL <https://www.sciencedirect.com/science/article/pii/0893608089900208>.
- Hu, Q., Zhang, R., and Zhou, Y. (2016). Transfer learning for short-term wind speed prediction with deep neural networks. *Renewable Energy*, 85:83–95.
- Hyndman, R. *fpp3: Data for "Forecasting: Principles and Practice" (3rd Edition)*, 2021. URL <https://CRAN.R-project.org/package=fpp3>. R package version 0.4.0.
- Hyndman, R. J. and Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software*, 26(3):1–22. URL <https://www.jstatsoft.org/article/view/v027i03>.

- Kingma, D. P. and Ba, J. L. (2015). Adam: A method for stochastic optimization. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, p. 1–15.
- Soman, S. S., Zareipour, H., Malik, O., and Mandal, P. (2010). A review of wind power and wind speed forecasting methods with different time horizons. *North American Power Symposium 2010*, p. 1–8.
- Staffell, I. and Pfenninger, S. (2016). Using Bias-Corrected Reanalysis to Simulate Current and Future Wind Power Output. *Energy*, 114:1224–1239.
- Torres, J. L., García, A., de Blas, M., and de Francisco, A. (2005). Forecast of hourly average wind speed with arma models in navarre (spain). *Solar Energy*, 79:65–77.
- Trapletti, A. and Hornik, K. *tseries: Time Series Analysis and Computational Finance*, 2020. URL <https://CRAN.R-project.org/package=tseries>. R package version 0.10-48.
- Yang, B., Zhong, L., Wang, J., Shu, H., Zhang, X., Yu, T., and Sun, L. (2020). State-of-the-art one-stop handbook on wind forecasting technologies: An overview of classifications, methodologies, and analysis. *Journal of Cleaner Production*.
- Yatiana, E., Rajakaruna, S., and Ghosh, A. (2017). Wind speed and direction forecasting for wind power generation using arima model. *2017 Australasian Universities Power Engineering Conference (AUPEC)*, p. 1–6.