

Modelos de Séries Temporais em Cascata para Previsão da Velocidade do Vento^{*}

Wallace P. P. Porto^{*} Ângelo M. O. Sant'Anna^{**} Daniel Barbosa^{*}
Nyegirton Barreiros dos Santos Costa^{***}

^{*} Departamento de Engenharia Elétrica e da Computação,
Universidade Federal da Bahia, BA, (e-mail:
wallaceppporto@gmail.com, dbarbosa@ufba.br).

^{**} Departamento de Engenharia Mecânica, Universidade Federal da
Bahia, BA (e-mail: angelo.santanna@ufba.br)

^{***} Instituto Federal do Sertão Pernambucano, PE (e-mail:
nyegirton.costa@if-sertao-pe.edu.br).

Abstract: This paper proposes developing cascade time series models for forecasting of wind speed. It was developed the Seasonal Autoregressive Integrated Moving Average (SARIMA), Prophet, Cascaded Autoregressive Integrated Moving Average (ARIMAc) and Cascaded Prophet. The called cascaded times series models are an approach in which, unlike the others, the forecast was made using data from each time to distinct time series. From predicted data by time series, its was reassembled in the original order. The analysis and comparison of the performances of each model was carried out based on different evaluation criteria to determine the most accurated forecasting. Finally, the Bergmann-Hommel post-hoc test was applied to assess the level of difference between developed times series models.

Resumo: Este artigo propõe desenvolver modelos de séries temporais em cascata para a previsão da velocidade do vento. Foram desenvolvidos os modelos Auto-Regressivo Integrado de Médias Móveis Sazonal (SARIMA), Prophet, Auto-Regressivo Integrado de Médias Móveis em cascata (ARIMAc) e Prophet em cascata. Os modelos ditos em cascata são uma abordagem na qual, diferentemente dos demais, a previsão foi realizada utilizando os dados de cada horário como séries temporais distintas e, em posse dos dados previstos, a série temporal foi remontada na ordem original. Foi feita a análise e comparação dos desempenhos de cada modelo com base em diferentes critérios de avaliação de modo a determinar a previsão mais acurada. Por fim, foi aplicado o teste *post-hoc* de Bergmann-Hommel para avaliar o nível de diferença entre os modelos desenvolvidos.

Keywords: Forecasting; SARIMA; ARIMA; Prophet; Time Series; Wind Speed; Performance.

Palavras-chaves: Previsão; SARIMA; ARIMA; Prophet; Series Temporais; Velocidade do Vento; Desempenho.

1. INTRODUÇÃO

A geração de energia elétrica no Brasil é historicamente baseada em dois pilares principais: as hidrelétricas e as termelétricas. Entretanto, com o crescimento da energia demandada no país, a necessidade do uso de fontes mais limpas de geração e a alta disponibilidade de recursos naturais da região — devido à sua extensão e posição geográfica — abrem espaço para um aumento expressivo na potência instalada de geração renovável de energia nos últimos anos (EPE, 2021a).

Nesse contexto, a geração eólica vem se desenvolvendo e crescendo de modo que, em 2019, 9% da energia elétrica gerada no Brasil era eólica, colocando o país como o 7^o maior produtor deste tipo de energia no mundo e 8^o em capacidade instalada (EPE, 2021a; ABEEólica,

^{*} Agradecimentos ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo suporte financeiro por meio do programa PIBITI / UFBA - CNPq.

2019). Além disso, é importante salientar que essa fonte de energia ainda tem muito potencial a ser explorado no Brasil, visto que a capacidade instalada em 2019 era de 17,13 GW (EPE, 2021a), enquanto é estimado um potencial elétrico de 500 GW em terra (Cepel, 2017) e 700 GW no mar (EPE, 2020).

Entretanto, é preciso considerar que a geração, a transmissão e a distribuição de energia elétrica são realizadas de forma organizada, e este planejamento é naturalmente complexo, pois envolve diversos agentes e horizontes de operação.

O planejamento da expansão do setor elétrico estima a necessidade e a possibilidade de instalação de novas usinas e da expansão do sistema elétrico em geral, buscando suprir ao aumento da demanda a longo prazo. O estudo considera projeções do aumento da carga, do aumento da geração em cada fonte e dos contratos já estabelecidos de geração, conforme o Plano Decenal de Expansão de Energia (PDE)

(EPE, 2022), e as previsões de necessidade de ampliação do sistema de transmissão, conforme o Programa de Expansão da Transmissão / Plano de Expansão de Longo Prazo (PET/PELP) (EPE, 2021b).

O planejamento de operação determina como o sistema e as usinas irão operar, de modo a atender à variação, a curto prazo, da carga com o menor custo possível. O estudo considera a demanda prevista, o custo presente e futuro de das gerações hidráulicas e térmicas, a energia acumulada em cada reservatório, a previsão de chuvas e a média prevista para a geração das outras fontes, segundo o Programa Mensal de Operação (PMO) feito pelo Operador Nacional do Sistema Elétrico (ONS) (ONS, 2022). O ONS controla, ainda, a operação em tempo real. Por ter uma característica de controle mais imediato, o planejamento e controle de operação trabalha com horizontes temporais curtos, então requer informações mais detalhadas e precisas a respeito da disponibilidade de energia em cada fonte.

Com a expansão da geração de energia por fontes renováveis, surge o problema da estocasticidade, pois muitas dessas fontes dependem diretamente de fenômenos naturais que estão fora do controle do ser humano, como volume de chuvas, irradiação solar e velocidade do vento. Em especial, a energia eólica possui uma incerteza associada bastante elevada, visto que variações abruptas de velocidade do vento são comuns ao longo do dia e, dessa forma, a complexidade do planejamento aumenta, pois não há garantia de que a energia estará disponível quando for necessária. Devido a isso, a geração eólica é incluída no planejamento apenas como um abatimento no valor da carga baseado na média da geração, deixando de considerar as variações e prejudicando a otimização.

Buscando uma solução para este problema, Silva (2014) propõe um modelo de otimização do planejamento a médio prazo considerando a previsões mensais de velocidade do vento — baseadas na distribuição normal multivariada, na matriz de correlação e na decomposição de Cholesky — e perspectivas de instalações de novos parques. Muthemba (2019) utiliza previsões mensais baseadas na função *Predict2* do *Toolbox TSTOOL* no *software* Mathworks Matlab e inclui a modelagem das incertezas para a proposta de otimização do planejamento de operação. Gouveia (2011) aplica modelos de Redes Neurais Artificiais (RNA), Lógica Fuzzy e Análise Wavelet para tentar encontrar uma previsão horária da velocidade do vento e da geração de energia eólica com um bom desempenho. Stefsos (2001) propõe modelos ARIMA e de RNA diretas para a previsão da velocidade do vento a cada 10 minutos. Camelo et al. (2017) aplica os modelos ARIMA, Holt-Winters e RNA para a previsão da velocidade do vento no Nordeste brasileiro. Carneiro et al. (2014) utiliza duas metodologias de RNA para a previsão da velocidade do vento em Maracanaú, Ceará.

Nesse contexto, este artigo propõe desenvolver modelos de séries temporais em cascata para a previsão da velocidade do vento a curto prazo, de modo a tornar possível um planejamento de operação do sistema elétrico mais seguro e otimizado com base em dados futuros de alta confiabilidade. Para tanto, foram utilizados dados de velocidade do vento, medidos durante um ano, numa estação meteorológica localizada em uma cidade do interior do Nordeste

brasileiro. Quatro modelos de séries temporais foram desenvolvidos: Auto-Regressivo Integrado de Médias Móveis Sazonal (SARIMA), Prophet, Auto-Regressivo Integrado de Médias Móveis (ARIMA) em cascata e Prophet em cascata (A Tabela 1 mostra as contribuições de cada trabalho citado como referência e deste projeto). Cada modelo de previsão foi submetido a seis diferentes critérios de avaliação de desempenho para comparar objetivamente os seus resultados. Por fim, os testes de hipótese de Friedman e de Bergmann e Hommel foram aplicados aos modelos para avaliar a diferença entre eles. Todos os cálculos e simulações deste projeto foram feitos com o uso do *software* R (Venables et al., 2021).

Tabela 1. Contribuições dos trabalhos relacionados

Projetos	Métodos de Previsão
Silva (2014)	Distribuição normal multivariada, matriz de correlação e decomposição de Cholesky.
Muthemba (2019)	Função <i>Predict2</i> do <i>Matlab</i> e modelagem de incertezas.
Gouveia (2011)	RNA, Lógica Fuzzy e Análise de Wavelet.
Stefsos (2001)	Modelos ARIMA e RNA.
Camelo et al. (2017)	Modelos ARIMA, Holt-Winters e RNA.
Carneiro et al. (2014)	RNA.
Este Projeto	SARIMA, Prophet, ARIMA em cascata e Prophet em cascata.

Este trabalho está organizado como segue: na seção Materiais e Métodos é feita a descrição do banco de dados de velocidade do vento utilizado e das condições de medição, os procedimentos e modelos de previsão utilizados são expostos e explicados, e os métodos de análise do desempenho dos modelos são também mostrados; a seção de Resultados contém as previsões calculadas através de cada modelo, além da comparação gráfica e numérica entre estes; por fim, na Conclusão, é feita a discussão dos resultados obtidos.

2. MATERIAIS E MÉTODOS

2.1 Base de Dados

Para as análises feitas nesse artigo, foi utilizado um banco de dados produzido por uma estação solarimétrica situada em uma cidade no interior do Nordeste brasileiro e contém medições realizadas durante o período de um ano (365 dias) entre os dias 20/03/2013 e 19/03/2014.

Cada variável do banco de dados é organizada no formato de séries temporais, com intervalos de 10 minutos entre si, e é formada por médias das medições feitas a cada segundo nos 10 minutos passados. Diversas variáveis, como a temperatura, a umidade e a radiação solar são coletadas pelo sistema de medição, mas neste trabalho apenas a velocidade do vento foi utilizada. A série conta com 144 medições a cada dia, num total de 52560 dados em m/s. Uma análise de pré-processamento foi realizada para evitar possíveis falhas nas medições e nenhuma foi encontrada, de modo que não foi preciso fazer nenhuma interpolação corretiva para imputação de dados.

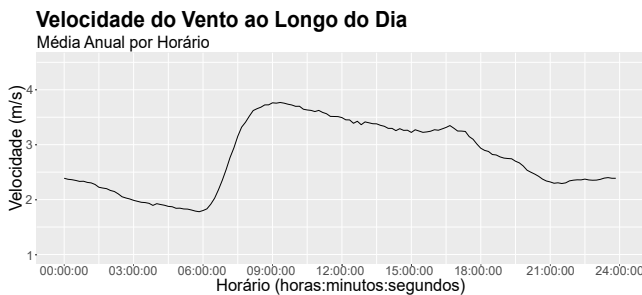


Figura 1. Média anual dos dados de velocidade do vento de cada horário.

Para a previsão, o bando de dados foi dividido em dois: os primeiros 360 dias (51840 medições) formaram os dados de treinamento para os modelos, os últimos 5 dias (720 medições) formaram os dados de teste e serviram de base para a análise da acurácia das previsões feitas por cada modelo.

Vale ressaltar que valores de velocidade do vento para a geração de energia eólica depende da altura na qual foram realizadas as medições e a altura do cubo de uma turbina eólica típica, pois a velocidade do vento varia diretamente com a altura, mas mantém o perfil temporal da localização do parque eólico. Caso deseje-se conhecer os valores correspondentes à altura da turbina eólica, deve ser feita uma correção com a altura (Castro, 2007).

2.2 Modelos em Cascata

Através da observação do perfil diário da velocidade do vento verifica-se uma relação entre o período do dia e a velocidade registrada. A Figura 1 mostra a média dos dados de cada horário ao longo de todo o ano e expõe este padrão de comportamento. É possível notar um aumento abrupto de velocidade das 06 às 09 horas e um decréscimo gradual ao longo do dia, que se estende até a manhã seguinte quando ocorre um novo aumento abrupto. Este padrão se repete com frequência ao longo do ano em todas as estações.

Levando isso em conta, planejou-se desenvolver uma modelagem de previsão que dependesse diretamente das medições do mesmo horário nos dias anteriores. Para isso, a série original, formada pelos dados de treinamento, com intervalos de 10 minutos e 51840 registros, foi decomposta em 144 séries — uma para cada horário — com intervalos de 24 horas, cada uma com 360 registros.

Feito isso, cada série passou por uma modelagem e uma previsão distinta (com a mesma técnica) visando prever as velocidades do vento nos próximos 5 dias em cada horário, formando juntas 720 dados previstos. Em seguida, as estimativas foram reagrupadas na sequência horária para corresponder ao formato da série original e então foram feitas as análises de desempenho das previsões em comparação com os dados de teste. A Figura 2 mostra um fluxograma que ilustra o mecanismo utilizado neste método.

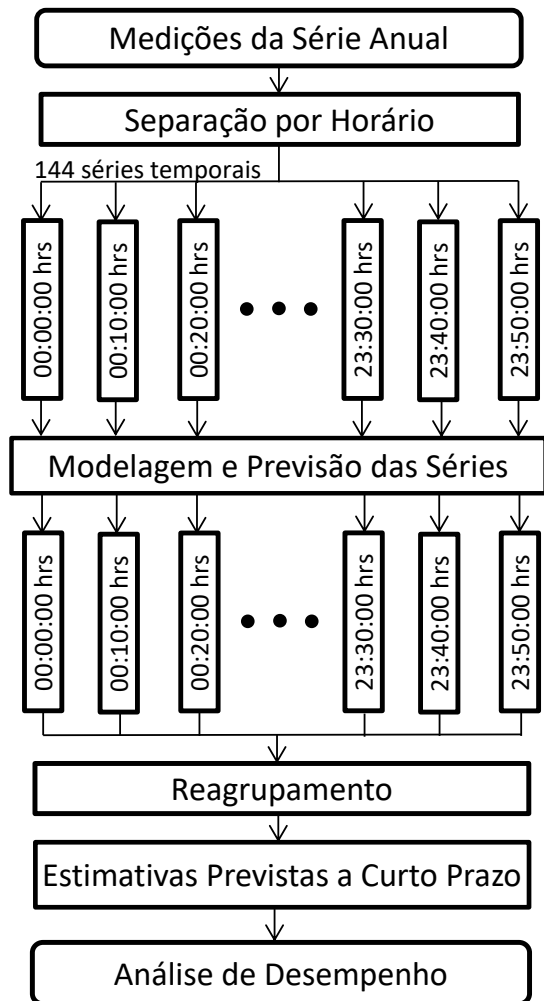


Figura 2. Método de previsão em cascata proposto.

2.3 Modelo Auto-Regressivo Integrado de Médias Móveis (ARIMA)

Os modelos ARIMA são modelos de previsão que têm como base uma combinação aditiva entre os modelos auto-regressivos e de médias móveis. A parcela auto-regressiva é composta por valores passados da variável estudada, sendo os valores mais recentes os mais significativos e, portanto, têm coeficientes maiores. A parcela de médias móveis é composta pelos erros entre as previsões anteriores e os valores reais e, assim como a anterior, os valores mais recentes são mais significativos que os passados. Além disso, existe a parte correspondente à integração, que determina a ordem da diferenciação aplicada (Hyndman and Athanasopoulos, 2021).

O número de valores passados utilizados em cada parcela depende da série à qual se pretende aplicar o modelo, de modo que um modelo ARIMA é definido como $ARIMA(p, d, q)$, sendo p e q , respectivamente, as ordens das parcelas auto-regressiva e de médias móveis, e d , o grau da diferenciação aplicada.

Em (1) é apresentada a definição do modelo genérico, sendo y'_t , o termo t da série diferenciada (a ordem da diferenciação pode ser superior a um); c , a média das diferenças entre observações consecutivas; ϕ e θ , as repre-

sentações dos coeficientes dos termos auto-regressivos e de médias móveis, respectivamente; e ε , os erros das previsões passadas.

$$y'_t = c + \phi_1 y'_{t-1} + \dots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (1)$$

A mesma equação pode ser reescrita de forma simplificada utilizando a notação conhecida como *backshift*, na qual o operador B representa um retrocesso de uma unidade no índice na amostra, ou seja, $By_t = y_{t-1}$, e $(1-B)$ representa a aplicação de uma diferenciação. Dessa forma o modelo ARIMA é descrito por (2).

$$(1 - \phi_1 B - \dots - \phi_p B^p)(1 - B)^d y_t = c + (1 + \theta_1 B + \dots + \theta_q B^q) \varepsilon_t \quad (2)$$

Os coeficientes das parcelas são obtidos pelo método de máxima verossimilhança, cujos valores calculados têm a maior probabilidade de gerar uma série correspondente à série em estudo.

Neste artigo, os modelos ARIMA foram aplicados na abordagem em cascata mostrada na seção 2.2, ou seja, para cada horário do dia foi proposto um modelo de ARIMA que se adequasse melhor à sua respectiva série temporal. Com o uso do *software* R, os modelos mais adequados foram calculados automaticamente utilizando o Critério de Informação de Akaike (Hyndman and Athanasopoulos, 2021). A Tabela 2 mostra os modelos encontrados para cada horário.

2.4 Modelo Auto-Regressivo Integrado de Médias Móveis Sazonal (SARIMA)

Os modelos ARIMA são muito úteis, porém são indicados para séries nas quais a sazonalidade é inexistente ou pouco importante, já que não inclui esse aspecto na sua modelagem. Para o caso de séries com uma periodicidade acentuada, existem os modelos SARIMA. Em geral, o período considerado na sazonalidade se refere à um ano, mas o modelo pode ser aplicado para qualquer periodicidade, seja esta horária, diária, semanal, etc.

Nestes modelos, as parcelas referentes ao modelo ARIMA supracitado são modificadas por fatores que incluem o efeito periódico nas estimativas. Estes fatores são calculados de maneira análoga ao ARIMA, porém consideram os valores passados com um intervalo de um período ao invés de uma amostra (Hyndman and Athanasopoulos, 2021).

O modelo é descrito como $ARIMA(p, d, q)(P, D, Q)_m$, onde P , D e Q são, respectivamente, análogos à p , d e q para os termos referentes à sazonalidade e m é a quantidade de amostras em um período.

A equação geral do modelo, usando a notação *backshift*, pode ser vista em (3). Nesta, além das variáveis descritas anteriormente, temos Φ e Θ que correspondem, respectivamente, à ϕ e θ para os termos sazonais e também são obtidos pelo método de máxima verossimilhança.

Tabela 2. Modelos ARIMA para cada horário.

Horário	Modelo	Horário	Modelo	Horário	Modelo
00:00:00	(3,1,0)	08:00:00	(2,1,0)	16:00:00	(1,2,0)
00:10:00	(0,3,0)	08:10:00	(2,2,0)	16:10:00	(4,3,0)
00:20:00	(0,3,0)	08:20:00	(2,1,0)	16:20:00	(1,1,0)
00:30:00	(0,3,0)	08:30:00	(1,1,0)	16:30:00	(1,1,0)
00:40:00	(1,1,0)	08:40:00	(1,1,0)	16:40:00	(2,1,0)
00:50:00	(3,1,0)	08:50:00	(2,1,0)	16:50:00	(1,1,0)
01:00:00	(0,3,0)	09:00:00	(2,1,0)	17:00:00	(2,1,0)
01:10:00	(0,4,0)	09:10:00	(1,2,0)	17:10:00	(2,1,0)
01:20:00	(0,3,0)	09:20:00	(1,1,0)	17:20:00	(2,1,0)
01:30:00	(1,1,0)	09:30:00	(2,1,0)	17:30:00	(0,2,0)
01:40:00	(1,1,0)	09:40:00	(1,1,0)	17:40:00	(2,1,0)
01:50:00	(0,3,0)	09:50:00	(1,2,0)	17:50:00	(0,2,0)
02:00:00	(1,1,0)	10:00:00	(3,1,0)	18:00:00	(1,1,0)
02:10:00	(0,3,0)	10:10:00	(2,1,0)	18:10:00	(1,1,0)
02:20:00	(1,1,0)	10:20:00	(2,1,0)	18:20:00	(0,2,0)
02:30:00	(1,1,0)	10:30:00	(2,2,0)	18:30:00	(0,2,0)
02:40:00	(2,2,0)	10:40:00	(1,1,0)	18:40:00	(2,1,0)
02:50:00	(2,1,0)	10:50:00	(1,2,0)	18:50:00	(2,1,0)
03:00:00	(2,1,0)	11:00:00	(1,3,0)	19:00:00	(2,1,0)
03:10:00	(2,1,0)	11:10:00	(1,1,0)	19:10:00	(2,1,0)
03:20:00	(2,1,0)	11:20:00	(1,1,0)	19:20:00	(2,1,0)
03:30:00	(2,1,0)	11:30:00	(4,1,0)	19:30:00	(2,1,0)
03:40:00	(2,1,0)	11:40:00	(2,1,0)	19:40:00	(0,3,0)
03:50:00	(1,1,0)	11:50:00	(2,1,0)	19:50:00	(2,1,0)
04:00:00	(1,1,0)	12:00:00	(2,1,0)	20:00:00	(2,1,0)
04:10:00	(2,1,0)	12:10:00	(1,1,0)	20:10:00	(2,1,0)
04:20:00	(2,1,0)	12:20:00	(1,1,0)	20:20:00	(2,1,0)
04:30:00	(2,1,0)	12:30:00	(1,1,0)	20:30:00	(2,1,0)
04:40:00	(1,1,0)	12:40:00	(1,1,0)	20:40:00	(2,1,0)
04:50:00	(3,1,0)	12:50:00	(4,3,0)	20:50:00	(2,1,0)
05:00:00	(0,3,0)	13:00:00	(2,1,0)	21:00:00	(0,2,0)
05:10:00	(1,1,0)	13:10:00	(0,4,0)	21:10:00	(1,1,0)
05:20:00	(1,1,0)	13:20:00	(1,1,0)	21:20:00	(1,1,0)
05:30:00	(0,3,0)	13:30:00	(2,2,0)	21:30:00	(0,3,0)
05:40:00	(0,2,0)	13:40:00	(0,5,0)	21:40:00	(1,1,0)
05:50:00	(0,2,0)	13:50:00	(1,1,0)	21:50:00	(1,1,0)
06:00:00	(0,2,0)	14:00:00	(2,2,0)	22:00:00	(2,1,0)
06:10:00	(0,2,0)	14:10:00	(2,1,0)	22:10:00	(0,3,0)
06:20:00	(0,2,0)	14:20:00	(1,0,0)	22:20:00	(2,1,0)
06:30:00	(0,2,0)	14:30:00	(1,1,0)	22:30:00	(0,3,0)
06:40:00	(0,2,0)	14:40:00	(1,1,0)	22:40:00	(1,1,0)
06:50:00	(0,2,0)	14:50:00	(1,1,0)	22:50:00	(2,2,0)
07:00:00	(0,2,0)	15:00:00	(0,2,0)	23:00:00	(2,2,0)
07:10:00	(0,2,0)	15:10:00	(4,2,0)	23:10:00	(2,1,0)
07:20:00	(2,2,0)	15:20:00	(0,4,0)	23:20:00	(2,2,0)
07:30:00	(0,2,0)	15:30:00	(1,4,0)	23:30:00	(3,1,0)
07:40:00	(0,2,0)	15:40:00	(1,1,0)	23:40:00	(3,1,0)
07:50:00	(2,2,0)	15:50:00	(1,2,0)	23:50:00	(3,1,0)

$$(1 - \phi_1 B - \dots - \phi_p B^p) \cdot (1 - \Phi_1 B^m - \dots - \Phi_P (B^m)^P) \cdot (1 - B)^d \cdot (1 - B^m)^D y_t = c + (1 + \theta_1 B + \dots + \theta_q B^q) \cdot (1 + \Theta_1 B^m + \dots + \Theta_Q (B^m)^Q) \varepsilon_t \quad (3)$$

Diversos modelos SARIMA foram testados durante o andamento do projeto, inclusive o modelo obtido automaticamente com o uso do *software* R, e, dentre estes, o que obteve melhor desempenho foi o modelo $ARIMA(2, 1, 5)(0, 1, 1)_{144}$. A análise gráfica e numérica dos seus resultados será feita na seção Resultados, juntamente com os outros métodos de previsão testados.

2.5 Prophet

O Prophet é um modelo de previsões de séries temporais desenvolvido pelo antigo *Facebook* (hoje *Meta*). Foi desen-

volvido de modo intuitivo para que um usuário que não necessariamente tenha tanto conhecimento a respeito de séries temporais e modelos de previsão pudesse utilizar.

Como mostra (4), sua modelagem se baseia na soma de diferentes parcelas, todas variando com o tempo: $g(t)$ se refere a componente de tendência, $s(t)$ é a componente sazonal (periódica), $h(t)$ adiciona o efeito de feriados nos cálculos, ε_t inclui uma variação aleatória e juntas elas formam a previsão $y(t)$ (Taylor and Letham, 2017).

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t \quad (4)$$

A tendência pode ser calculada de três maneiras diferentes escolhidas pelo usuário: linear, não linear e constante. A tendência constante, como o nome indica, corresponde a uma série que não aumenta ou diminui temporalmente e varia em torno de um valor fixo. O cálculo da tendência no modo linear usa pontos de mudança, que podem ser indicados pelo usuário ou calculados automaticamente, e, no intervalo entre esses pontos, o comportamento da tendência é aproximado por uma função afim. Por fim, a tendência não linear utiliza também de pontos de mudança, porém a função entre estes é calculada considerando fatores como saturação e taxa de crescimento e decrescimento (Taylor and Letham, 2017).

A componente periódica do modelo é calculada a partir de Séries de Fourier geradas pelo próprio Prophet. Três diferentes períodos podem ser incluídos ou excluídos do modelo conforme a série estudada: sazonalidade diária, semanal e anual. Além disso, ainda é possível determinar se a componente sazonal será aplicada de maneira aditiva ou multiplicativa.

O efeito dos feriados é bastante útil para modelos que buscam prever o comportamento humano ou social e varia muito conforme o foco do estudo, mas, como este artigo trata de um fenômeno da natureza, este componente foi anulada do modelo. Finalmente a componente de erro aleatório é estimada como parte de uma distribuição normal.

O modelo Prophet foi usado neste estudo tanto para a previsão direta da série, como no modo em cascata. Testes foram feitos em busca do modelo mais adequado modificando as sazonalidades anual, semanal e diária, além do tipo de sazonalidade aplicado. Para o modelo direto, os resultados otimizados foram obtidos usando a sazonalidade aditiva com todos os períodos sendo considerados (dia, semana e ano). Já para o modelo em cascata, os melhores resultados foram obtidos para a sazonalidade apenas semanal aditiva e isso vale para todos os horários, diferente do ARIMA em cascata, que teve modelos distintos entre os horários.

2.6 Análises de Desempenho

Com o objetivo de analisar a qualidade das previsões realizadas e possibilitar uma comparação objetiva entre os modelos, seis métodos de análise de desempenho foram empregados, conforme (5) a (10) (Acosta et al., 2021b). Nestas, y_i , \hat{y}_i , \bar{y}_i e n correspondem aos valores observados, aos valores estimados, a média dos valores observados e o número de observações/estimativas, respectivamente.

No método raiz quadrada do erro-médio (RMSE, do inglês *Root Mean Squared Error*), o erro de cada estimativa é elevado ao quadrado, então é calculada a média desses valores e, por fim, é extraída a raiz quadrada, como mostra (5). O valor ótimo é obtido para RMSE nulo, portando o modelo com o menor valor é o mais acurado.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (5)$$

O erro médio absoluto (MAE, do inglês *Mean Absolute Error*) é determinado através do cálculo da média dos módulos dos erros cada estimativa, como em (6). O melhor modelo será aquele com o menor valor de MAE.

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (6)$$

No erro percentual absoluto médio (MAPE, do inglês *Mean Absolute Percentage Error*), é calculado o módulo do erro percentual em relação ao valor real para cada estimativa e, em seguida, é feito o cálculo da média entre tais valores, como pode ser visto em (7). Neste método, o melhor modelo também será aquele que tiver o menor valor.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \cdot 100 \quad (7)$$

O coeficiente de determinação (R^2) é calculado através da razão entre o somatório do quadrado dos resíduos e o somatório do quadrado das diferenças entre os valores observados e a sua média, como pode ser observado em (8). Este método corresponde à capacidade do modelo de explicar a variância dos dados e resulta em valores entre 0 e 1, sendo o valor ótimo obtido para um R^2 igual a 1, logo, o modelo com o maior valor é avaliado como o melhor.

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

O método *variance account for* (VAF) calcula a razão entre as variâncias dos erros de cada estimativa e a variância dos valores observados, como em (9). Neste caso o melhor modelo terá VAF mais próximo à 100%.

$$VAF = \left[1 - \frac{\text{var}(y_i - \hat{y}_i)}{\text{var}(y_i)} \right] \cdot 100 \quad (9)$$

Por fim, para o cálculo da taxa de acerto (HR, do inglês *Hit Rate*) é determinado um valor de referência que corresponde ao máximo erro considerado aceitável para cada estimativa e calcula-se o percentual de estimativas que tiveram sucesso, como mostra (10), portanto o modelo mais bem-sucedido será aquele com o maior percentual.

$$HR = \frac{\text{quantidade de } (|\hat{y}_i - y_i| < \text{referência})}{n} \quad (10)$$

2.7 Testes de Friedman e Bergmann-Hommel

Diferentemente dos medidores de desempenho mostrados na seção anterior, os testes de hipótese de Friedman e de Bergmann e Hommel foram aplicados de modo a comparar os modelos entre si, avaliando as diferenças. O objetivo é verificar se há modelos equivalentes entre si ou se de fato os resultados são significativamente distintos, de modo a validar o propósito da comparação entre eles.

O teste de Friedman é um teste não-paramétrico cuja hipótese nula é que todos os modelos têm medianas equivalentes, enquanto a hipótese alternativa é que pelo menos dois entre os modelos têm medianas diferentes entre si (Carrasco et al., 2020). Para a hipótese nula ser rejeitada, o p – valor resultante do teste deve ser inferior à 0,05.

O teste *post-hoc* de Bergmann-Hommel faz a comparação, dois a dois, entre os modelos para determinar exatamente quais modelos diferem entre si (Derrac et al., 2011). De modo análogo ao teste de Friedman, os modelos são considerados significativamente diferentes se o p – valor entre eles for inferior à 0,05 (Acosta et al., 2021a).

3. RESULTADOS

Figura 3 mostra as previsões obtidas dos melhores modelos para cada um dos quatro métodos de modelagem testados. É possível verificar que todos os modelos conseguem acompanhar adequadamente o padrão de velocidade do vento, com os picos e vales mais evidentes coincidindo. Além disso, é notável o perfil de velocidades citado na seção 2.2 se repetindo, tanto nos valores reais quanto nas previsões ao longo dos 5 dias, com variações, mas ainda assim tendo um comportamento similar. Adicionalmente, nota-se que o modelo Prophet tem um ajuste fino mais acurado, de modo que as grandes variações abruptas são seguidas pela previsão de maneira mais próxima que os outros modelos.

As Tabelas 3 e 4 mostram os resultados dos testes de desempenho, apresentados na seção 2.6, aplicados nos diferentes modelos utilizados nesse trabalho, destacando-se os melhores resultados obtidos para cada teste em negrito. Analisando os dados de desempenho obtidos, verifica-se que o modelo Prophet apresentou os melhores valores, apenas ficando em segundo em segundo lugar no teste MAPE, atrás do Prophet em cascata.

Tabela 3. Métricas de erros de previsão dos modelos

Modelo	RMSE (m/s)	MAE (m/s)	MAPE (%)
SARIMA	0,7798075	0,6217323	39,43911
Prophet	0,6324670	0,5112149	29,26831
ARIMA*	0,6511635	0,5248339	29,60131
Prophet*	0,6592816	0,5262104	28,46251

*Modelos em cascata

Tais resultados mostram que o modelo Prophet possui uma melhor adequação aos dados apresentados, de modo que este é o modelo que permitiria um planejamento de operação do sistema elétrico mais acurado, viabilizando uma melhora na qualidade da otimização econômica e logística.

Tabela 4. Métricas de desempenho dos modelos

Modelo	R^2	VAF (%)	HR (%)
SARIMA	0,2853125	52,74352	46,66667
Prophet	0,5298709	53,18056	56,25000
ARIMA*	0,5016649	50,17105	55,97222
Prophet*	0,4891619	51,88660	55,55556

*Modelos em cascata

Aplicando o teste de Friedman aos modelos, obteve-se um p – valor igual a 0, bem abaixo de 0,05, o que significa que a hipótese nula foi rejeitada e há ao menos dois modelos diferentes entre si. Os p – valores para a comparação dois a dois entre os modelos através do teste *post-hoc* de Bergmann-Hommel foram todos aproximadamente iguais a zero, sendo o maior deles 0,0003 para os modelos ARIMA em cascata e Prophet, o que significa que todos os modelos são significativamente diferentes entre si.

4. CONCLUSÃO

Este artigo propôs o uso de modelos em cascata para a previsão de séries temporais de velocidades do vento, cujas medições foram feitas em uma estação solarimétrica localizada no interior do Nordeste brasileiro, buscando auxiliar no planejamento de operação do sistema elétrico.

Após a execução dos testes, verificou-se que todos os modelos obtiveram um resultado satisfatório, quando se considera a dificuldade em prever fenômenos da natureza. Cada um deles conseguiu captar o perfil de velocidades do vento ao longo do dia, de modo que os resultados dos testes de desempenho não foram tão diferentes um do outro. Apesar disso, através do teste de Friedman e do teste *post-hoc* de Bergmann-Hommel, constatou-se que todos os modelos diferem entre si.

Adicionalmente, observando os resultados obtidos, constatou-se que o modelo Prophet apresentou os melhores indicadores de desempenho para o conjunto de dados utilizado, indicando que este seria o selecionado para uma aplicação prática das previsões de velocidade do vento nesta região. Entretanto, tal observação não pode ser generalizada, pois, o comportamento e o padrão de velocidades do vento varia em função da região geográfica do país na qual foi observado, o que pode fazer com que outro modelo tenha um desempenho melhor.

Apesar disso, o procedimento utilizado neste artigo pode ser aplicado em bancos de dados de velocidade do vento de outros locais, os mesmos métodos de previsão e testes de desempenho podem ser feitos em busca do melhor modelo para cada caso.

Sendo assim, este é um dos possíveis caminhos para a expansão do projeto, buscar outros bancos de dados de diferentes locais e aplicar o método aqui exposto, visando comparar os resultados dos bancos de dados entre si e analisar se existe um modelo único que se adéqua bem ao perfil de velocidades do vento de qualquer lugar.

REFERÊNCIAS

- ABEEólica (2019). Energia eólica os bons ventos do brasil. Associação Brasileira de Energia Eólica.
Acosta, S.M., Amoroso, A.L., Sant’Anna, A.M.O., and Junior, O.C. (2021a). Predictive modeling in a steelma-

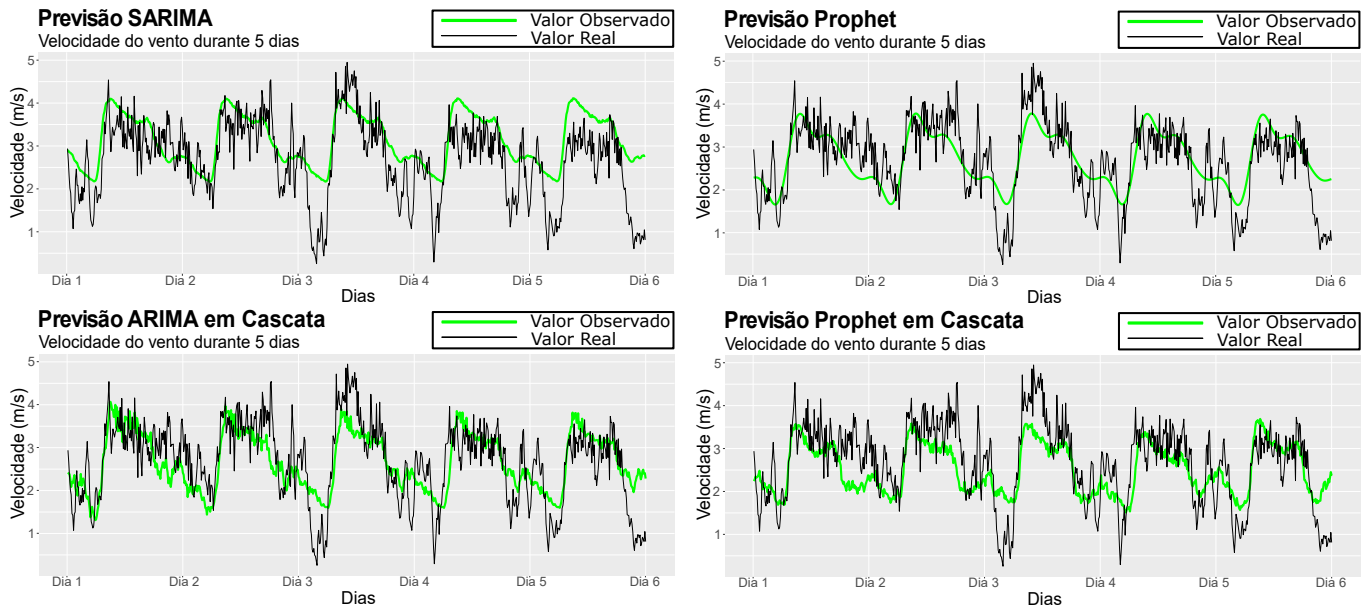


Figura 3. Previsões de 5 dias obtidas pela aplicação dos modelos.

- king process using optimized relevance vector regression and support vector regression. *Annals of Operations Research*, 1–22. doi:10.1007/s10479-021-04053-9.
- Acosta, S.M., Amoroso, A.L., Sant'Anna, A.M.O., and Junior, O.C. (2021b). Relevance vector machine with tuning based on self-adaptive differential evolution approach for predictive modelling of a chemical process. *Applied Mathematical Modelling*, 95, 125–142. doi:10.1016/j.apm.2021.01.057.
- Camelo, A.H., Lucio, P.S., and Júnior, J.V.L. (2017). Modelagem da velocidade do vento usando metodologias arima, holt-winters e rna na previsão de geração eólica no nordeste brasileiro. *Revista Brasileira de Climatologia*.
- Carneiro, T.C., Santos, H.A., Braga, A.P.S., and Carvalho, P.C.M. (2014). Redes neurais artificiais pa a previsão da velocidade do vento: Estudo de caso para maracanaú - ce. *Congresso Brasileiro de Automática*.
- Carrasco, J., García, S., Rueda, M.M., Das, S., and Herrera, F. (2020). Recent trends in the use of statistical tests for comparing swarm and evolutionary computing algorithms: Practical guidelines and a critical review. *Swarm and Evolutionary Computation*, 11. doi:10.1016/j.swevo.2020.100665.
- Castro, R.M.G. (2007). *Energias Renováveis e Produção Descentralizada: Introdução à Energia Eólica*. Universidade Técnica de Lisboa, Lisboa, 3rd edition.
- Cepel, E. (2017). *Atlas do Potencial Eólico Brasileiro - Simulações 2013 - 1ª Edição*. Centro de Pesquisas de Energia Elétrica, Rio de Janeiro.
- Derrac, J., García, S., Molina, D., and Herrera, F. (2011). A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms. *Swarm and Evolutionary Computation*, 13–14. doi:10.1016/j.swevo.2011.02.002.
- EPE (2020). *Roadmap Eólica Offshore Brasil*. Empresa de Pesquisa Energética, Brasília.
- EPE (2021a). *Balanço Energético Nacional*. Empresa de Pesquisa Energética, Brasília.
- EPE (2021b). *Programa de Expansão da Transmissão (PET) / Plano de Expansão de Longo Prazo (PELP) Ciclo 2021 – 2º Semestre*. Empresa de Pesquisa Energética, Brasília.
- EPE (2022). *Plano Decenal de Expansão de Energia 2031*. Ministério de Minas e Energia/Empresa de Pesquisa Energética, Brasília.
- Gouveia, H.T.V. (2011). *Previsão de Ventos e Geração Eólica do Sistema NE: Analisando Diversos Sítios e Buscando a Melhor Modelagem através da Inteligência Artificial*. Universidade Federal de Pernambuco, Recife.
- Hyndman, R.J. and Athanasopoulos, G. (2021). *Forecasting: Principles and Practice, 3rd edition*. OTexts, Melbourne.
- Muthemba, L.J. (2019). *Modelo Robusto Para o Planejamento da Operação Eólica-Hidrotérmica Considerando Incertezas*. Universidade Federal do Paraná, Curitiba.
- ONS (2022). *Sumário do Programa Mensal de Operação*. Operador Nacional do Sistema Elétrico, Brasília.
- Silva, S.R. (2014). *Planejamento Hidrotérmico Considerando a Penetração de Geração Eólica em Sistemas Elétricos de Potência*. Universidade Federal de Itajubá, Itajubá.
- Stefsos, A. (2001). A novel approach for the forecasting of mean hourly wind speed time series. *Renewable Energy*.
- Taylor, S.J. and Letham, B. (2017). Forecasting at scale. *PeerJ Preprints*, 5–18.
- Venables, W.N., Smith, D.M., and the R Core Team (2021). *An Introduction to R*. R Foundation.