

Predicting Long-Term Wind Speed in Wind Farms of Northeast Brazil: A Comparative Analysis Through Machine Learning Models

M. de Paula, M. Colnago, J. Fidalgo, and W. Casaca

Abstract—The rapid growth of wind generation in northeast Brazil has led to multiple benefits to many different stakeholders of energy industry, especially because the wind is a renewable resource – an abundant and ubiquitous power source present in almost every state in the northeast region of Brazil. Despite the several benefits of wind power, forecasting the wind speed becomes a challenging task in practice, as it is highly volatile over time, especially when one has to deal with long-term predictions. Therefore, this paper focuses on applying different Machine Learning strategies such as Random Forest, Neural Networks and Gradient Boosting to perform regression on wind data for long periods of time. Three wind farms in the northeast Brazil have been investigated, whose data sets were constructed from the wind farms data collections and the National Institute of Meteorology (INMET). Statistical analyses of the wind data and the optimization of the trained predictors were conducted, as well as several quantitative assessments of the obtained forecast results.

Index Terms—Wind Power, Wind Speed Forecasting, Machine Learning, Long-Term, Regression, Northeastern Brazil.

I. INTRODUÇÃO

Nos últimos anos tem se observado na indústria de energia mundial um grande crescimento do uso de fontes renováveis de energia, o que também tem sido constatado no contexto nacional. Dentre as principais formas de geração sustentável da matriz nacional, a energia eólica desempenha um grande protagonismo, visto que o vento é um recurso abundante e onipresente em diversas regiões do país. Com a crescente preocupação não só acerca da geração de energia sustentável, mas também devido à delicada questão da preservação ambiental, o Brasil passou a investir na construção de parques de geração eólica, em especial, na região nordeste do país.

This work was supported by the FAPESP – São Paulo Research Foundation under grant 2018/05341-4, by the ERDF – European Regional Development Fund through the Operational Programme for Competitiveness and Internationalisation - COMPETE 2020 Programme within project POCI-01-0145-FEDER-006961, and by National Funds through the FCT – Fundação para a Ciência e Tecnologia (Portuguese Foundation for Science and Technology) as part of project UID/EEA/50014/2013.

M. de Paula, M. Colnago and W. Casaca are with Department of Energy Engineering, São Paulo State University (UNESP), Av. dos Barrageiros, 1881, Primavera, Rosana-SP, Brazil (e-mail: matheus.paula@unesp.br; marilainecolnago@gmail.com; wallace.casaca@unesp.br). J. Nuno Fidalgo is with the Power Systems Unit of INESC TEC and with the Faculty of Engineering of University of Porto, Rua Dr Roberto Frias, s/n, 4200-465 Porto, Portugal (email: jfidalgo@inesctec.pt).

De fato, devido ao expressivo crescimento da energia eólica, tal como observado nos últimos anos, o Nordeste atingiu a inconcebível cifra de 13 GW de capacidade instalada no ano de 2018, o que equivale a cerca de aproximadamente 85% da energia eólica gerada em solo nacional [1]. Segundo a Agência Nacional de Energia Elétrica (ANEEL), é esperado que até 2024, a produção de energia eólica alcance a marca de 21 GW de potência instalada [1], indicando a ascensão dessa fonte renovável junto à fonte hidroelétrica, que hoje representa a maior parcela de geração de eletricidade no Brasil.

Em paralelo à perspectiva de crescimento das fontes renováveis, é importante também ressaltar a necessidade de se implementar planos mais otimizados de eficiência energética, isto é, que promovam a conciliação de geração e distribuição de eletricidade com a demanda de consumo, evitando assim possíveis racionamentos de energia no futuro. Assim, a fim de promover estratégias que sejam mais efetivas quanto à geração e uso da energia eólica nos próximos anos, é imprescindível o emprego de ferramentas de análise de dados, o que inclui modelos de previsão de geração eólica [2], da direção predominante do vento [3], e do fluxo aéreo em parques eólicos, sendo este último tópico o objetivo do presente trabalho.

Conforme destacam [4], [5], [6], os trabalhos de predição do vento podem ser agrupados consoante os seguintes horizontes de planejamento: (I) de extremo curto prazo (30 minutos), (II) curto prazo (de 30 minutos à 6 horas), (III) médio prazo (de 6 horas à 1 dia) e, finalmente, (IV) longo prazo (de 1 dia à 1 semana adiante). A literatura tem apresentado algumas soluções que endereçam a questão, em especial para os casos (I) e (II), como em [7], em que foi empregado métodos do tipo *ensemble* para prever a velocidade do vento em horizontes de curto prazo, e em [8], em que um modelo de predição de geração eólica via Redes Neurais Artificiais (RNA) foi desenvolvido e validado com dados de um parque eólico nordestino. Já no trabalho [9], os autores construíram um novo *pipeline* de RNA, tendo como objetivo avaliar o comportamento das funções *tansig* e *logsig* na previsão do vento, no caso, em intervalos de longo prazo. Na obra apresentada em [10], os autores utilizaram uma metodologia híbrida para prever a velocidade do vento dado um horizonte de planejamento do tipo (II). Além disso, no referido estudo, foram também utilizadas RNA, SVM (*Support Vector Machines*) e o método RF (*Random Forest*).

Ademais, em [11], os autores empregaram uma metodologia de seleção de *features* para predições em horizontes do tipo (II)

a partir da combinação dos modelos RNA, SVM, RF e *Gradient Boosting*. Por fim, no trabalho [12], foram utilizadas técnicas de clássicas de aprendizado de máquina para prever, em diferentes horizontes de tempo, a velocidade do vento no parque eólico de Osório, situado no Rio Grande do Sul.

Embora a literatura esteja majoritariamente inclinada a atacar os casos (I), (II) e (III), o grupo (IV) tem despertado grande interesse da comunidade científica, especialmente em razão do vento ser um fenômeno de alta volatilidade ao longo do dia, tornando sua previsão desafiadora na prática. Além disso, este tipo de estimativa pode beneficiar o setor eólico nacional nos seguintes aspectos: (i) possibilidade de melhorias na escolha de localidades para a construção de parques eólicos, auxiliando assim nas tomadas de decisão do setor, (ii) direcionar os planejamentos de manutenção dos parques, bem como melhorar a eficiência operacional dos sítios eólicos, (iii) mitigar situações climáticas desfavoráveis aos sistemas eólicos, detectando variações climáticas específicas e padrões de comportamento ao longo do ano, (iv) permitir uma maior otimização dos custos operacionais, reduzindo assim perdas econômicas por parte dos conjuntos eólicos como um todo.

Desta maneira, este trabalho visa explorar a previsibilidade do vento em horizontes de longo prazo (até 1 semana à frente) em três parques eólicos da região nordeste brasileira, a saber: Parque Eólico Morro do Chapéu Sul (Bahia), Parque Eólico Praia Mansa (Ceará), e Parque Eólico Mangue Seco (Rio Grande do Norte). Para realizar o presente estudo, a pesquisa empregou dados coletados de 2010 a 2013 dos registros diários do fluxo aéreo fornecidos por esses parques, bem como outras grandezas físicas associadas ao problema, extraídas do repositório do Instituto Nacional de Meteorologia (INMET).

Para a construção dos preditores, foram utilizados três modelos de aprendizado de máquina: Floresta Aleatória (*Random Forest*), Redes Neurais Multicamadas (*Multi-Layer Neural Networks*), e Aumento do Gradiente (*Gradient Boosting*).

Este trabalho está organizado da seguinte forma: a Seção II apresenta uma análise inicial dos dados, enquanto a Seção III traz a metodologia. Já as Seções IV e V apresentam as etapas de pré-processamento, métricas, resultados e as conclusões.

II. DATA SETS E ANÁLISE EXPLORATÓRIA DOS DADOS

A fim de realizar a previsão do vento nos parques eólicos, foi criado um banco de dados a partir do repositório público do INMET, que reúne os seguintes atributos/dados: temperatura média diária (°C), umidade relativa diária (%) e velocidade média do vento (m/s). Para cada um dos três parques eólicos, as amostras foram obtidas no período compreendido entre 01/01/2010 e 31/12/2013, resultando assim em três *data sets* distintos, abrangendo 4 anos completos cada um.

Com o objetivo de, inicialmente, analisar a influência dos atributos primários nos *data sets*, foi realizada uma análise estatística dos mesmos, o que inclui sumarizações básicas como: média, desvio padrão e valores máximos e mínimos de cada uma das variáveis do problema (vide Tabelas I-III). A partir das estatísticas computadas, foi possível constatar que há um comportamento padrão presente nas três coleções de dados: a umidade relativa do ar apresenta um desvio padrão maior em

relação as outras duas grandezas, o que demonstra que a umidade sofre alterações mais acentuadas ao longo do tempo.

Outro ponto a ser destacado é que o Parque Eólico Mangue Seco (RN) apresenta um maior fluxo de vento ao longo dos anos, seguido do Parque Morro do Chapéu Sul (BA) e do Parque Praia Mansa (CE). Esses resultados estão em sinergia com o verificado em [13], que conclui que as regiões do Brasil com maiores potenciais eólicos e capacidades instaladas são, nesta ordem: Rio Grande do Norte, Bahia e Ceará.

TABELA I
MÉTRICAS ESTATÍSTICAS: PARQUE EÓLICO MORRO DO CHAPÉU SUL.

Métrica	Temperatura	Umidade	Velocidade
Média	20,81 °C	72,12 %	3,41 m/s
Desvio Padrão	2,02 °C	11,45 %	0,94 m/s
Valor Máximo	26 °C	98,75 %	6,66 m/s
Valor Mínimo	14 °C	39,50 %	0,20 m/s

TABELA II
MÉTRICAS ESTATÍSTICAS: PARQUE EÓLICO PRAIA MANSÁ.

Métrica	Temperatura	Umidade	Velocidade
Média	27,26 °C	76,52 %	2,91 m/s
Desvio Padrão	0,87 °C	6,07 %	0,91 m/s
Valor Máximo	29,66 °C	96,75 %	5,43 m/s
Valor Mínimo	23,78 °C	61,75 %	0,33 m/s

TABELA III
MÉTRICAS ESTATÍSTICAS: PARQUE EÓLICO MANGUE SECO.

Métrica	Temperatura	Umidade	Velocidade
Média	26,68 °C	80,66 %	4,12 m/s
Desvio Padrão	1,27 °C	4,83 %	1,12 m/s
Valor Máximo	29,78 °C	96,50 %	7,46 m/s
Valor Mínimo	22,50 °C	69,50 %	0,26 m/s

Visando analisar o comportamento da temperatura e da umidade do ar (variáveis previsoras de entrada) em relação à velocidade do vento (variável *target* – de saída, a qual pretende-se prever), foram originados gráficos de dispersão (*scatter's plots*: Figs. 1-3).

Na Fig. 1, é possível constatar que tanto a temperatura como a umidade do ar (em relação à velocidade do vento) não apresentam um comportamento padrão, isto é, estas não são bem-comportadas ao longo do tempo. Há períodos em que a temperatura é mais alta, e a incidência de vento, mais baixa, porém, o contrário também ocorre. Já em relação à umidade relativa do ar, é possível observar que, em intervalos intermediários, esta acaba sendo associada à valores intermediários da velocidade do vento para o parque estudado.

No caso da Fig. 2, pode-se notar que há um certo “padrão de comportamento” de ambas as variáveis previsoras em relação à velocidade do vento. É possível constatar ainda que, se verifica certa correlação entre a temperatura e a velocidade do vento (na maior parte das amostras do *data set*). Já para a umidade, é possível observar o contrário, isto é, maiores taxas de umidade do ar (%) proporcionam menores fluxos de vento.

Por fim, na Fig. 3, pode-se observar que, no caso da temperatura, os dados acabam não apresentando um padrão bem definido, já que existem certos períodos em que se tem altas temperaturas e elevados fluxos aéreos, assim como acontece a situação contrária. Em relação à umidade relativa, é visível que maiores taxas de umidade relativa do ar levam à

menores fluxos de velocidade do vento, embora tal comportamento não seja absolutamente evidente como no caso do parque anterior.

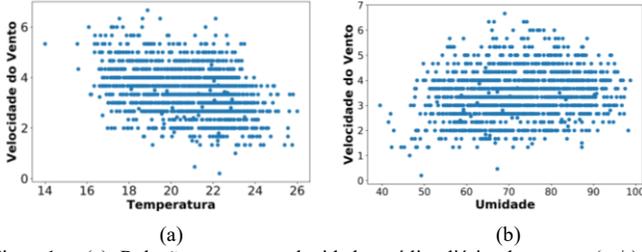


Fig. 1. (a) Relação entre a velocidade média diária do vento (m/s) e temperatura média diária (°C). (b) Relação entre a velocidade média diária do vento (m/s) e umidade relativa do ar diária (%) no Parque Eólico Morro do Chapéu Sul.

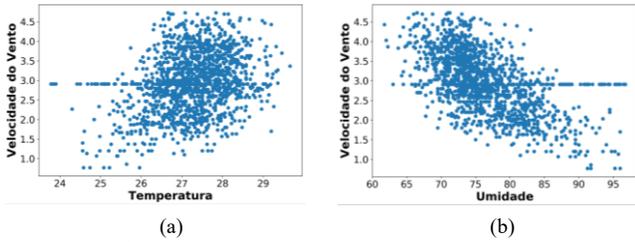


Fig. 2. (a) Relação entre a velocidade média diária do vento (m/s) e temperatura média diária (°C). (b) Relação entre a velocidade média diária do vento (m/s) e umidade relativa do ar diária (%) no Parque Eólico Praia Mansa.

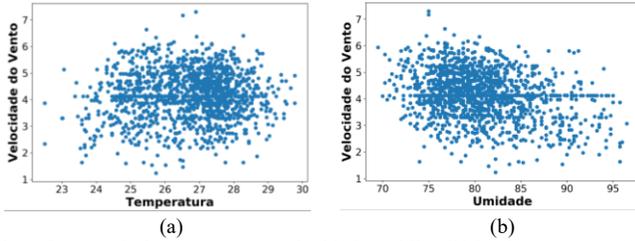


Fig. 3. (a) Relação entre a velocidade média diária do vento (m/s) e temperatura média diária (°C). (b) Relação entre a velocidade média diária do vento (m/s) e umidade relativa do ar diária (%) no Parque Eólico Mangue Seco.

III. PREVISÃO DA VELOCIDADE DO VENTO: PROCESSO METODOLÓGICO

A fim de prever o fluxo aéreo nos três parques eólicos objetos-de-estudo deste trabalho, três diferentes técnicas de Aprendizado de Máquina (AM) foram implementadas e devidamente ajustadas para o problema em questão: Árvores de Decisão via Floresta Aleatória (*Random Forest*), Redes Neurais Multicamadas (*Neural Networks Multi-Layer Perceptron*), e Aumento de Gradiente (*Gradient Boosting*). Tais métodos foram selecionados devido ao sucesso obtido no tratamento anterior de problemas que envolvem séries temporais, tendo resultado em preditores de alta assertividade com um bom nível de customização de seus parâmetros de controle [14], [15].

As subseções a seguir descrevem as principais características dos modelos de predição usados, bem como as etapas metodológicas de ajuste dos parâmetros dos modelos.

A. *Random Search*

Os algoritmos do tipo *ensemble* implementados neste trabalho possuem diversos hiperparâmetros de ajuste utilizados

a fim de maximizar o desempenho de cada modelo e a aderência aos dados. Para esse propósito, utilizou-se a estratégia *Random Search*, que testa um número limitado de possibilidades para a construção dos modelos de predição a partir de árvores de regressão, com um número pré-fixado de iterações. Desta forma, é possível reduzir consideravelmente o número de verificações e obter, de modo mais acelerado, uma configuração adequada para os dados do problema de regressão estudado [15].

B. *Árvores de Decisão – Random Forest*

O *Random Forest* é um método do tipo *ensemble*, que integra um conjunto de estimadores que induzem à aprendizagem definitiva do modelo [16]. Essa metodologia utiliza um subconjunto aleatório dos dados no decorrer da escolha dos nós, de forma que a randomização faz com que o algoritmo atribua pesos maiores às variáveis que mais influenciam na predição, independente da correlação entre elas [17]. Utilizando *Random Search*, foram testadas todas as configurações possíveis considerando o universo de parâmetros previamente definido em [15], tendo sido utilizadas 20 iterações com esse modelo.

C. *Redes Neurais Multicamadas – Multi-Layer Perceptron*

Com o propósito de estimar a velocidade do vento, um algoritmo de Redes Neurais do tipo *Multi-Layer Perceptron* (MLP) foi implementado, buscando reproduzir um paradigma neuro-matemático através de uma estrutura neural inteligente.

Na RNA, o processo de aprendizagem ocorre a partir de um *threshold* (limiar), onde o neurônio passa a emitir uma série de “impulsos nervosos” que, quando acumulados pelo sistema, acabam por representar uma saída que pode ser um impulso ou não [18]. Em termos matemáticos, uma RNA pode ser descrita pelas equações 1 e 2, ou seja, por uma função de ativação φ que processa o sinal gerado pela combinação linear das entradas x_j 's, do neurônio net_i , com seus respectivos pesos w_{ij} , somada ao valor de limiarização θ (*thresholding*) [19]:

$$net_i = \sum_{j=1}^n x_j w_{ij} + \theta_i \quad (1)$$

$$y_k = \varphi(net_i) \quad (2)$$

Os parâmetros clássicos do modelo, como número de camadas ocultas, função de ativação e outros, estão descritos na Seção IV.

D. *Aumento do Gradiente – Gradient Boosting*

O último modelo implementado foi o algoritmo *Gradient Boosting* (GB). Tal método emprega a chamada estratégia de *Boosting* nas árvores de regressão, com o objetivo de obter um menor impacto de sobreajuste nos resultados. A partir dessa estratégia, as árvores acabam não possuindo os mesmos pesos em suas configurações, o que resulta em pesos mais adequados para tratar o problema de predição de vento [20], [21].

O GB utiliza uma variação do método do gradiente descendente para otimizar as árvores durante as iterações a partir de uma função de custo. Assim, a decisão, que é orientada

pelo gradiente, é definida a partir da soma das estimativas das árvores [22]. É válido salientar que o modelo GB depende de três pontos-chaves: uma função de custo a ser otimizada (*loss function*), submodelos preditivos baseados em árvores de regressão, e submodelos extras (“*weak learners*”) para a minimização da função de custo. Os parâmetros utilizados pelo modelo em cada parque eólico investigado, bem como as configurações testadas, encontram-se na Seção IV. Finalmente, cabe também salientar que o universo de parâmetros considerado para a aplicação do *Random Search* foi o mesmo do estabelecido em [15].

E. Pré-processamento e Engenharia de Recursos

A fim de se utilizar os modelos descritos, foi necessária a realização do pré-processamento dos dados, isto é, a padronização (*standardization*) dos *data sets*. Essa etapa é necessária, visto que os atributos devem ser uniformizados em uma escala comum, evitando assim que valores discrepantes influenciem nos cálculos dos pesos e na etapa de treinamento dos modelos [23]. Assim, a partir do uso da equação (3), tem-se a *standardization* dos dados (fórmula do *z-score*):

$$z = \frac{x - u}{s}, \quad (3)$$

em que x representa uma dada variável de entrada, u a média dessa variável, e s seu desvio padrão.

Além da etapa de pré-processamento, foram criadas variáveis auxiliares (*features*) a partir do uso da engenharia de recursos.

Essas *features* são geradas a partir da utilização de estatísticas descritivas que permitem melhorar significativamente o desempenho preditivo por parte dos modelos de aprendizado de máquina. Com isso, no presente trabalho, foram criadas as seguintes variáveis complementares de suporte:

- Média Móvel da Velocidade do Vento (m/s) – (janelas de 2, 3 e 5 amostras).
- Desvio Padrão da Velocidade do Vento (m/s) – (janelas de 3 e 5 amostras).
- Mediana Móvel da Velocidade do Vento (m/s) – (janelas de 3 e 5 amostras).
- Subtração Móvel da Velocidade do Vento (m/s) – (janela de 2 amostras).
- Multiplicação da Temperatura Média do Ar (°C) e Umidade Relativa Média do Ar (%).
- Soma da Temperatura Média do Ar (°C) e Umidade Relativa do Ar (%).

F. Métricas de Validação

Para a análise do desempenho dos preditores, as seguintes métricas foram consideradas: *Mean Squared Error* (MSE), *Mean Absolute Error* (MAE), e o *Mean Absolute Percentage Error* (MAPE). As métricas acima foram adotadas por serem medidas estatísticas já consagradas na literatura de aprendizado de máquina, tendo sido largamente utilizadas em outras obras tais como [15], [24], [25], [26], [27]. Além disso, no caso do

MAPE, erros percentuais têm a vantagem de serem independentes de escala, sendo utilizados para comparar o desempenho da previsão entre diferentes conjuntos de dados. Isso resulta em um melhor suporte de análise estatística, além de facilitar a avaliação de desempenho dos métodos.

As métricas foram computadas a partir do valor real Y_i (do fluxo aéreo), e da estimativa \hat{Y}_i , obtida pelos modelos já treinados, conforme as equações (4), (5) e (6) abaixo:

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (4)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i| \quad (5)$$

$$MAPE = \frac{1}{n} \sum_{i=0}^n \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| * 100 \quad (6)$$

IV. RESULTADOS E DISCUSSÃO

Para construir os modelos preditivos, foram utilizados 75% (anos de 2010 até 2012) dos dados disponíveis na etapa de treinamento, e 25% dos dados na etapa de validação (ano de 2013), em cada um dos parques eólicos analisados. Além disso, todas as implementações computacionais envolvidas no presente estudo foram realizadas em linguagem *Python*, com o auxílio da biblioteca de aprendizado de máquina *Sklearn*.

A. Parque Eólico Morro do Chapéu Sul

No primeiro caso analisado – Parque Eólico Morro do Chapéu Sul (BA) – o modelo *Random Forest* (RF) foi otimizado pelo uso do *Random Search*, conforme mencionado anteriormente. A Tabela IV apresenta os parâmetros ótimos utilizados pelo método.

Já com relação ao preditor originado pelas Redes Neurais, foram utilizados os parâmetros estabelecidos na Tabela V. Os demais parâmetros do modelo (que não foram listados na tabela) não foram alterados, tendo sido adotados os valores *default* definidos na biblioteca *Scikit-learn*.

Por fim, em relação ao algoritmo *Gradient Boosting*, foram utilizados os parâmetros listados na Tabela VI.

TABELA IV
PARÂMETROS UTILIZADOS PELO ALGORITMO FLORESTA ALEATÓRIA.
CASO DE ESTUDO I: PARQUE EÓLICO MORRO DO CHAPÉU SUL

NÚMERO DE ÁRVORES DA FLORESTA	25
NÚMERO MÍNIMO DE AMOSTRAS NECESSÁRIAS PARA DIVIDIR UM NÓ INTERNO	1
PROFUNDIDADE MÁXIMA DA ÁRVORE	NONE
NÚMERO DE RECURSOS A SEREM UTILIZADOS	AUTO
AMOSTRAS DE AUTO-INICIALIZAÇÃO	TRUE

TABELA V

PARÂMETROS UTILIZADOS PELO ALGORITMO REDES NEURAIS MLP.	
CASO DE ESTUDO 1: PARQUE EÓLICO MORRO DO CHAPÉU SUL	
NÚMERO DE NEURÔNIOS NA CAMADA INTERMEDIÁRIA	100
FUNÇÃO DE ATIVAÇÃO	RELU
MÁXIMO DE INTERAÇÕES	200

TABELA VI

PARÂMETROS UTILIZADOS PELO ALGORITMO AUMENTO DO GRADIENTE.	
CASO DE ESTUDO 1: PARQUE EÓLICO MORRO CHAPÉU DO SUL	
FUNÇÃO DE CUSTO	REGRESSÃO DE MÍNIMOS QUADRADOS ('Ls')
NÚMERO DE ÁRVORES DA FLORESTA	100
TAXA DE APRENDIZADO	0,1

Para fins de validação numérica dos resultados, a Tabela VII relaciona os *scores* obtidos por cada algoritmo para o ano de 2013, referente ao Parque Eólico Morro do Chapéu Sul, enquanto a Fig. 4 apresenta a previsão realizada pelos três modelos no período dos doze primeiros dias de 2013.

TABELA VII

SCORES OBTIDOS PELOS PREDITORES (PARQUE EÓLICO MORRO CHAPÉU SUL).

MÉTRICA DE AVALIAÇÃO	RANDOM FOREST	MULTI-LAYER PERCEPTRON	GRADIENT BOOSTING
MSE	0,064969	0,074638	0,085420
MAE	0,138770	0,184009	0,215118
MAPE	6,11%	7,39%	8,27%

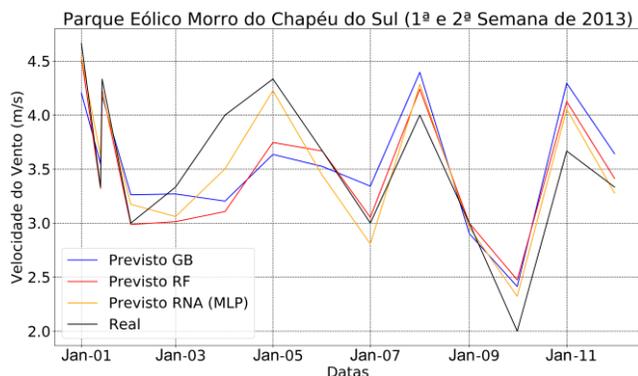


Fig. 4. Gráfico da velocidade prevista do vento em função da velocidade real (em m/s) no subintervalo de previsão dos doze primeiros dias de 2013, obtido pelos modelos preditivos para o Parque Eólico Morro do Chapéu Sul (BA).

A Fig. 4 mostra uma alta assertividade por parte dos modelos nas duas primeiras semanas de 2013. Nesse período em específico, fica visível observar um melhor comportamento do RF e das RNA (MLP), sendo os modelos com menores métricas em relação ao MSE, indicando que a distância entre os valores previstos e reais são menores nesses modelos em relação ao GB.

A partir do uso das métricas de validação, fica evidente que durante todo o período de verificação (ano de 2013), o modelo *Random Forest* foi aquele que apresentou um melhor desempenho (vide, por exemplo, o valor obtido para o MAPE). Entretanto, é válido ressaltar que os erros ficaram todos represados em faixas de valores semelhantes.

B. Parque Eólico Praia Mansa (CE)

Considerando agora nosso segundo caso de estudo – Parque Eólico Praia Mansa (CE) – inicialmente foram encontrados os melhores parâmetros para o preditor correspondente ao *Random Forest*, tendo sido estes apresentados na Tabela VIII. De maneira análoga ao estudo de caso anterior, também foram realizadas previsões utilizando RNA, além do *Gradient Boosting*. Para esses casos, as Tabelas IX e X listam os parâmetros ótimos empregados por ambos os modelos.

TABELA VIII

PARÂMETROS UTILIZADOS PELO ALGORITMO FLORESTA ALEATÓRIA.	
CASO DE ESTUDO 2: PARQUE EÓLICO PRAIA MANSÁ	
NÚMERO DE ÁRVORES DA FLORESTA	100
NÚMERO MÍNIMO DE AMOSTRAS NECESSÁRIAS PARA DIVIDIR UM NÓ	3
PROFUNDIDADE MÁXIMA DA ÁRVORE	NONE
NÚMERO DE RECURSOS A SEREM UTILIZADOS	AUTO
AMOSTRAS DE AUTO-INICIALIZAÇÃO	TRUE

TABELA IX

PARÂMETROS UTILIZADOS PELO ALGORITMO REDES NEURAIS MLP.	
CASO DE ESTUDO 2: PARQUE EÓLICO PRAIA MANSÁ	
NÚMERO DE NEURÔNIOS NA CAMADA INTERMEDIÁRIA	150
FUNÇÃO DE ATIVAÇÃO	RELU
MÁXIMO DE INTERAÇÕES	200

TABELA X

PARÂMETROS UTILIZADOS PELO ALGORITMO AUMENTO DO GRADIENTE.	
CASO DE ESTUDO 2: PARQUE EÓLICO PRAIA MANSÁ	
FUNÇÃO DE CUSTO	REGRESSÃO DE MÍNIMOS QUADRADOS ('Ls')
NÚMERO DE ÁRVORES DA FLORESTA	75
TAXA DE APRENDIZADO	0,1

Os *scores* obtidos para as métricas de avaliação encontram-se tabulados na Tabela XI. Pode-se observar que os algoritmos apresentaram desempenhos bastante similares em relação às três métricas de validação, o que demonstra que eles podem ser empregados sem distinção para explorar a previsão do vento para esse parque eólico em particular. Outro ponto a ser observado é que, embora esse parque tenha apresentado uma boa correlação entre as variáveis originais do problema (vide Figura 2), os erros de previsão foram mais acentuados. Isso se deve ao fato de que as variáveis velocidade, umidade e temperatura não foram aquelas de maior peso na etapa de treinamento, as quais foram superadas pelas variáveis artificialmente criadas para otimizar os modelos (vide Seção III.E). Em termos gerais, a média móvel da velocidade do vento (janela de 3 amostras) foi aquela que apresentou maior importância nos modelos de previsão construídos.

A Fig. 5 ilustra as previsões realizadas nos doze primeiros dias do período de teste, onde pode-se observar um maior erro por parte dos modelos em relação à Seção IV.A.

TABELA XI

SCORES OBTIDOS PELOS PREDITORES: PARQUE EÓLICO PRAIA MANSA.

MÉTRICA DE AVALIAÇÃO	RANDOM FOREST	MULTI-LAYER PERCEPTRON	GRADIENT BOOSTING
MSE	0,1425954	0,1298846	0,1392863
MAE	0,2421639	0,2405208	0,2564593
MAPE	9,95%	9,72%	10,28%

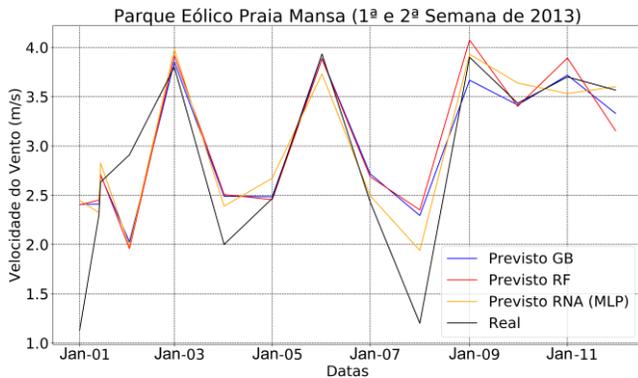


Fig. 5. Gráfico da velocidade prevista do vento em função da velocidade real (em m/s) no subintervalo de predição dos doze primeiros dias de 2013, obtido pelos modelos preditivos para o Parque Eólico Praia Mansa (CE).

C. Parque Eólico Mangue Seco (RN)

Por fim, foram realizados os testes no Parque Eólico Mangue Seco (RN). As Tabelas XII, XIII e XIV a seguir relacionam os parâmetros utilizados por cada um dos algoritmos de aprendizado de máquina estudados.

TABELA XII

PARÂMETROS UTILIZADOS PELO ALGORITMO FLORESTA ALEATÓRIA.

CASO DE ESTUDO 3: PARQUE EÓLICO MANGUE SECO	
NÚMERO DE ÁRVORES DA FLORESTA	75
NÚMERO MÍNIMO DE AMOSTRAS NECESSÁRIAS PARA DIVIDIR UM NÓ INTERNO	3
PROFUNDIDADE MÁXIMA DA ÁRVORE	NONE
NÚMERO DE RECURSOS A SEREM UTILIZADOS	AUTO
AMOSTRAS DE AUTO-INICIALIZAÇÃO	TRUE

TABELA XIII

PARÂMETROS UTILIZADOS PELO ALGORITMO REDES NEURAIS MLP.

CASO DE ESTUDO 3: PARQUE EÓLICO MANGUE SECO	
NÚMERO DE NEURÓNIOS NA CAMADA INTERMEDIÁRIA	100
FUNÇÃO DE ATIVAÇÃO	RELU
MÁXIMO DE INTERAÇÕES	200

TABELA XIV

PARÂMETROS UTILIZADOS PELO ALGORITMO AUMENTO DO GRADIENTE.

CASO DE ESTUDO 3: PARQUE EÓLICO MANGUE SECO	
FUNÇÃO DE CUSTO	REGRESSÃO DE MÍNIMOS QUADRADOS ('Ls')
NÚMERO DE ÁRVORES DA FLORESTA	100
TAXA DE APRENDIZADO	0,1

Tal como nos casos anteriores, a Fig. 6 a seguir apresenta as previsões obtidas pelos três modelos de aprendizado para os doze primeiros dias do período de validação/teste.

Além disso, a Tabela XV apresenta os *scores* obtidos mediante à aplicação das métricas de validação para os três modelos de predição, onde dessa vez o preditor correspondente ao *Random Forest* acabou obtendo um melhor desempenho ao longo de todo o período de teste (ano de 2013), cujo MSE, MAE e MAPE atingiram 0,21, 0,30, e 7,72%, respectivamente.

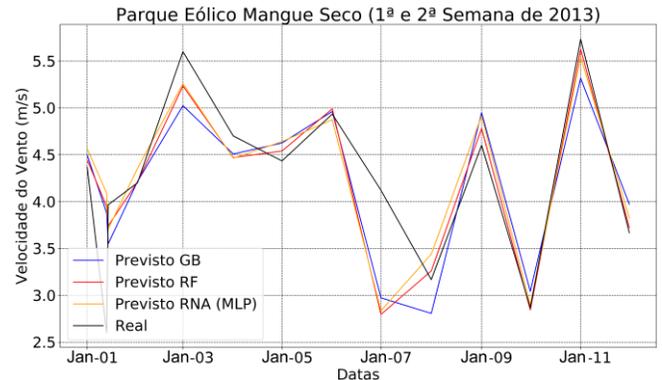


Fig. 6. Gráfico da velocidade prevista do vento em função da velocidade real (em m/s) no subintervalo de predição dos doze primeiros dias de 2013, obtido pelos modelos preditivos para o Parque Eólico Mangue Seco (RN).

TABELA XV

SCORES OBTIDOS PELOS PREDITORES: PARQUE EÓLICO MANGUE SECO.

MÉTRICA DE AVALIAÇÃO	RANDOM FOREST	MULTI-LAYER PERCEPTRON	GRADIENT BOOSTING
MSE	0,2107762	0,2404582	0,2260700
MAE	0,3000884	0,3456549	0,3332530
MAPE	7,72%	8,81%	8,54%

Segundo a Fig. 6, fica evidente um comportamento semelhante por parte dos modelos ao longo do período considerado. Nesse parque em particular, obteve-se uma melhor predição por parte do RF, seguido do GB, e por fim, das RNA.

D. Comparação dos Resultados dos Parques Eólicos

Como sumário dos resultados obtidos em cada caso de estudo conduzido, foi observado para o Parque Eólico Chapéu do Sul (BA) uma maior assertividade por parte do modelo *Random Forest*, com um MAPE de 6,11%. Já para o Parque Eólico Praia Mansa (CE), o melhor resultado foi alcançado pelas Redes Neurais, cujo MAPE foi de 9,72%. Por fim, no Parque Eólico Mangue Seco (RN), o melhor *score* obtido para o MAPE foi o do preditor *Random Forest*, cujo valor foi de 7,72%.

V. CONCLUSÃO

A crescente expansão da energia eólica nos últimos anos, sobretudo no nordeste brasileiro, tem permitido atingir uma grande capacidade de geração de potência elétrica. O vento é um fenômeno contínuo e abundante em diversas localidades do nordeste brasileiro, o que torna sua análise um ponto chave nos mais variados complexos eólicos instalados na região a fim de melhor dimensionar a geração futura de energia e elaborar planos estratégicos de eficiência energética.

A previsão do fluxo aéreo em janelas de longo prazo, tal como foi explorado neste trabalho, pode auxiliar o controle de despacho de potência não somente em território nordestino, mas também para todo o país – por exemplo, via SIN (Sistema Integrado Nacional) – visto que os preditores obtidos apresentaram uma boa taxa de assertividade ao longo de um ano completo. Dentre os casos investigados, o *score* para o MAPE é de 10% para o caso mais drástico, e 6% no melhor caso.

Embora não se tenha observado a imposição de um único modelo “ideal” para todos os três parques avaliados, é importante mencionar que em problemas práticos de regressão, diferentes abordagens de aprendizado de máquina podem levar à diferentes resultados mesmo quando aplicadas em um mesmo espaço de solução, conforme discutido em [28]. Desse modo, conclui-se que o emprego de mais de um modelo de predição em problemas relacionados à regressão não linear podem trazer benefícios, especialmente no contexto de predição em horizontes de longo prazo tal como o explorado neste trabalho.

V.I REFERÊNCIAS

- [1] Letras Ambientais, “Nordeste gera 85% da energia eólica do Brasil,” January 2018. Available: <https://www.letrasambientais.com.br/posts/nordeste-gera-85-da-energia-eolica-do-brasil>.
- [2] G. C. Godinho, P. M. Maçaira and R. C. Souza, “Combined forecast model for wind generation in Brazilian monthly dispatch scheduling,” *IEEE Latin America Transactions*, vol. 17, no. 09, pp. 1432-1438, 2019.
- [3] M. I. Garcia-Planas and T. Gondadze, “Wind profile prediction using linear Markov Chains: a linear algebra approach,” *IEEE Latin America Transactions*, vol. 16, no. 2, pp. 536-541, 2018.
- [4] S. S. Soman, H. Zareipour, O. Malik and P. Mandala, “A review of wind power and wind speed forecasting methods with different time horizons,” *IEEE North American Power Symposium*, pp. 1-8, 2010.
- [5] X. Wang, P. Guo and X. Huang, “A review of wind power forecasting models,” *Energy Proceedings*, vol. 12, pp. 770-778, 2011.
- [6] M. Santhosh, C. Venkaiah and D. M. V. Kumar, “Short-term wind speed forecasting approach using Ensemble Empirical Mode Decomposition and Deep Boltzmann Machine,” *IEEE Trans. Antennas Propagations*, vol. 19, pp. 100242, 2019.
- [7] A. Alonso, A. Torres and J. R. Dorronsoro, “Random forests and gradient boosting for wind energy prediction,” *International Conference on Hybrid Artificial Intelligence Systems*, Springer, pp. 26-37, 2015.
- [8] J. Fujise. *Desenvolvimento de Modelo de Previsão de Geração Eólica para Planejamento Energético Utilizando Redes Neurais Artificiais*. Tese de Doutorado, Universidade Federal do Rio de Janeiro, 2018.
- [9] B. Chainok, S. Tunyasirut, S. Wangnipparnto and W. Permpoonsinsup, “Artificial neural network model for wind energy on urban building in Bangkok,” *International Electrical Engineering Congress*, pp. 1-4, 2017.
- [10] Q. Han. Non-parametric hybrid models for wind speed forecasting. *Energy Conversion and Management*, vol. 148, pp. 554-568, 2017.
- [11] C. Feng. A data-driven multi-model methodology with deep feature selection for short-term wind forecasting. *Applied Energy*, vol. 190, pp. 1245-1257, 2017.
- [12] A. Khosravi, L. Machado, R. O. Nunes. Time-series prediction of wind speed using machine learning algorithms: A case study Osorio wind farm, Brazil. *Applied Energy*, vol. 224, pp. 550-566, 2018.
- [13] Produção de energia eólica atinge a marca de 14 gigawatts, Agência Brasil, Aug. 2018. Available: <http://agenciabrasil.ebc.com.br/economia/noticia/2018-11/producao-de-energia-eolica-no-pais-atinge-marca-de-14-gigawatts>.
- [14] L. Junho, W. Wang, F. Harrou and Y. Sun, “Wind power prediction using ensemble learning-based models,” *IEEE Access*, vol. 8, pp. 61517-61527, 2020.
- [15] J. V. Leme, W. Casaca, M. Colnago and M. A. Dias, “Towards assessing the electricity demand in Brazil: data-driven analysis and ensemble learning models,” *Energies*, vol. 13, no. 6, pp. 1407, 2020.
- [16] L. Breiman, “Random Forests,” *Machine Learning*, vol. 45, no. 1, pp. 5-32, 2001.
- [17] V. Svetnik, A. Liaw, C. Tong, J. C. Culberson, R. P. Sheridan and B. P. Feuston, “Random Forest: a classification and regression tool for compound classification and QSAR modeling,” *Journal of Chemical Information and Computer Sciences*, vol. 43, no. 6, pp. 1947-1958, 2003.
- [18] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, Pearson, 3rd edition, 2009.
- [19] F. J. V. Zuben, *Redes Neurais Artificiais*. Tech. Report, Available: ftp.dca.fee.unicamp.br/pub/docs/vonzuben/ia006_03/topico5_03.pdf, 2010.
- [20] J. H. Friedman, “Stochastic gradient boosting,” *Computational Statistics & Data Analysis*, vol. 38, no. 4, pp. 367-378, 2002.
- [21] A. Natekin and A. Knoll. “Gradient boosting machines, a tutorial,” *Frontiers in Neuroinformatics*, vol. 7, pp. 21, 2013.
- [22] C. Li. “A Gentle introduction to gradient boosting”. Dec. 2019, Available: http://www.ccs.neu.edu/home/vip/teach/MLcourse/4_boosting/slides/g_radien_boosting.pdf, 2016.
- [23] S. Haykin. *Redes neurais: princípios e prática*, Bookman Editora, 2007.
- [24] M. Paliwal and U. A. Kumar, “Neural networks and statistical techniques: A review of applications”, *Expert Systems with App.*, vol. 36, pp. 2-17, 2009.
- [25] A. De Myttenaere. Mean absolute percentage error for regression models. *Neurocomputing*, vol. 192, p. 38-48, 2016.
- [26] D. M. Allen. Mean square error of prediction as a criterion for selecting variables. *Technometrics*, vol. 13, no. 3, pp. 469-475, 1971.
- [27] D. G. Mayer and D. G. Butler. Statistical validation. *Ecological modelling*, vol. 68, no. 1-2, pp. 21-32, 1993.
- [28] Z. H. Zhou. *Ensemble Methods: Foundations and Algorithms*, Ed. 1, Chapman and Hall, 2012.



Matheus Pussainolli de Paula é graduando do curso de Engenharia de Energia da Universidade Estadual Paulista (UNESP). Atualmente é bolsista de Iniciação Científica da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP). Foi Jovem Aprendiz na empresa ABB (Asea Brown Boveri) por dois anos, atuando no seguimento de manutenção elétrica. Fez curso técnico em Mecatrônica pelo Centro Universitário ENIAC e curso de Aprendizagem Industrial de Eletricista de Manutenção pela Escola SENAI “Hermenegildo Campos de Almeida”. Sua área de interesse é Aprendizado de Máquina no Setor de Energia, Ciência de Dados, Visualização de Dados e Previsões em Séries Temporais.



Marilaine Colnago é graduada e mestre em Matemática Aplicada e Computacional pela Universidade Estadual Paulista (UNESP), e doutora em Ciências da Computação e Matemática Computacional pela Universidade de São Paulo (USP). Seus interesses de pesquisa incluem Ciência de Dados, Visualização, Inteligência Computacional Aplicada no Setor de Energia, Processamento Digital de Imagens, Matemática Computacional e Análise Numérica.



José Nuno Fidalgo nasceu no Porto, Portugal, em 30 de agosto de 1961. Atualmente é Professor Associado do Departamento de Engenharia Electrotécnica e de Computadores da Faculdade de Engenharia da Universidade da Universidade do Porto. Também é Pesquisador Sênior do Centro de Sistemas de Energia do Instituto de Engenharia de Sistemas e Computadores - Tecnologia e Ciência (INESC-TEC). Obteve seu doutorado em Engenharia Elétrica e Computação na mesma Universidade, em 1995. Seus interesses de pesquisa incluem Reconhecimento de Padrões, Inteligência Computacional e Redes Neurais e suas Aplicações em Sistemas de Energia.



Wallace Casaca obteve o bacharelado e o mestrado em Matemática pela Universidade Estadual Paulista (UNESP) em 2008 e 2010. Durante 2010-2014, realizou seu doutorado em Ciências da Computação e Matemática Computacional na Universidade de São Paulo (USP). Como parte de seus estudos de doutorado, também foi pesquisador visitante na Brown University, School of Engineering, EUA. Atualmente é docente da UNESP, atuando no curso de Engenharia de Energia. Seus interesses de pesquisa incluem Ciência de Dados, Problemas Computacionais em Energia, Visão Computacional, Sensoriamento Remoto, Visualização de Informações e Técnicas de Otimização.