Wind speed prediction model based on DWT and Randon Forest

Modelo para predição da velocidade do vento baseado em DWT e Randon Forest

Received: 2023-10-21 | Accepted: 2023-11-25 | Published: 2023-11-29

Leandro Brito dos Santos
ORCID: https://orcid.org/0000-0000-0000-0000
Universidade Federal do Recôcoava da Bahia, CETENS, Brazil
e-mail: lbsantos@ufrb.edu.br

Roberto Luiz Souza Monteiro
ORCID: https://orcid.org/0000-0001-7576-9267
State University of Bahia, Department of Earth and Agricultural Sciences, Brazil

Maria Cristina Cunha de Oliveira Carvalho
ORCID: https://orcid.org/0000-0000-0000-0000
Bahia State University, Brazil
e-mail: biocris2005@hotmail.com

Marcos Antônio Felipe Santos
ORCID: https://orcid.org/0000-0000-0000-0000
State University of Bahia, Department of Earth and Agricultural Sciences, Brazil
e-mail: reciclabahia@yahoo.com.br

Marcos Batista Figueredo
ORCID: https://orcid.org/ 0000-0002-8193-5419
State University of Bahia, Department of Earth and Agricultural Sciences, Brazil
e-mail: mbfigueredo@uneb.br

ABSTRACT

Wind energy is one of the fastest power generation technologies in the power generation industry and one of the most cost-effective methods of generating electrical power. For system reliability, improving highly appropriate wind speed forecasting methods is desirable. The wavelet transform is a powerful mathematical technique that converts an analyzed signal into a time-frequency representation. This technique helps forecast non-stationary time series. The objective is to evaluate the performance of the "DWT-Random Forest" model in predicting wind speed through a comparative analysis of performance metrics (MSE, RMSE, and MAPE) with similar studies. Our motivation is rooted in the pressing need to improve wind forecasting methods to optimize renewable energy generation. The method involved implementing the model, which presented performance metrics: MSE of 0.0099, RMSE of 0.0996, and MAPE of 0.0779. However, comparative analysis with previous studies reveals that our model demonstrates competitive performance. The main result of this study is the finding that the "DWT-Random Forest" model exhibits a respectable performance in predicting wind speed, although there is room for improvement.

Keywords: wind forecast; random forest; wavelets;
RESUMO

A energia eólica é uma das tecnologias de geração de energia mais rápidas na indústria de geração de energia e um dos métodos mais econômicos de geração de energia elétrica. Para a confiabilidade do sistema, é desejável aprimorar métodos altamente apropriados de previsão de velocidade do vento. A transformada wavelet é uma poderosa técnica matemática que converte um sinal analisado em uma representação tempo-frequência. Essa técnica tem se mostrado útil na previsão de séries temporais não estacionárias. O objetivo é avaliar o desempenho do modelo “DWT-Random Forest” na previsão de velocidade do vento por meio de uma análise comparativa de métricas de desempenho (MSE, RMSE e MAPE) com estudos semelhantes. Nossa motivação está enraizada na necessidade premente de aprimorar os métodos de previsão de vento para otimizar a geração de energia renovável. O método envolveu a implementação do modelo, que apresentou métricas de desempenho: MSE de 0.0099, RMSE de 0.0996 e MAPE de 0.0779. No entanto, a análise comparativa com estudos anteriores revela que nosso modelo demonstra um desempenho competitivo. O principal resultado deste estudo é a constatação de que o modelo “DWT-Random Forest” exibe um desempenho respeitável na previsão de velocidade do vento, embora haja espaço para melhorias.

Keywords: Previsão vento; Random Forest; Wavelets.

INTRODUCTION

The transition to renewable and sustainable energy sources is a crucial response to today’s environmental challenges and energy crises (Khelil et al., 2021; Zhang et al., 2022). As an abundant and pollution-free resource, wind emerges as one of the leading candidates in the field of renewable energy (Silva et al., 2020). The effectiveness of wind energy, characterized by its affordability, cost-effectiveness, and eco-friendly nature, is strongly influenced by the predictability of wind speed (Shi, 2019; Fanel et al., 2021; Akhlagi et al., 2022). This forecasting is vital for optimizing wind power generation and has significant applications in various industries, including aviation, marine, environmental sciences, and electricity production (Charakopoulos, Karakasidis et al., 2019). While wind resources gain prominence for their renewable and sustainable characteristics, the challenge lies in wind speed variability, which is highly influenced by geographical and climatic conditions (Lipu et al., 2021; Tiwari, 2022). This variability can lead to irregular power generation, negatively affecting the quality and stability of electrical systems. Therefore, accurate wind speed forecasting methods are essential for efficiently controlling wind turbines in wind farms (Hanifi et al., 2020).

Within this context, we investigate the application of advanced modeling techniques to predict wind speed, with a particular focus on wavelet decomposition methodology. This technique divides the time series into approximation and detail coefficients, capturing general trends and specific wind fluctuations. Our study combined
this approach with the Random Forest machine learning model, known for its ability to model complex, nonlinear relationships between variables. By integrating wavelet decomposition coefficients as input variables for Random Forest, we seek to explore not only the effectiveness of this combination in capturing the dynamic nature of wind speed but also in improving the accuracy of short-term forecasting, a critical aspect of efficient wind energy management.

This work presents a wind power prediction method (WPP) that combines the robustness of the Discrete Wavelet Transform (DWT) with the predictive efficacy of the Random Forest model. The main contribution of this approach lies in the use of DWT to decompose the time series of wind energy data into approximation and detail components, each representing different frequencies and characteristics of the signal. Subsequently, these components are used as independent inputs to separately trained Random Forest models to capture the intrinsic dynamics of each. The final result of the forecast is obtained by aggregating the forecasts of each model, providing a comprehensive estimate of future wind power. Through accurate data from three wind farms, we demonstrated the superiority of the proposed method over conventional forecasting techniques, translating into greater predictive accuracy.

The rest of this article is structured as follows: the Problem Description Section outlines the description of the problem; the Materials and Methods Section explains the proposed methodology, emphasizing the application of DWT and modeling via Random Forest; and details the experimental design; the results are presented and discussed in the Results and Discussions Section. Finally, the Conclusion section closes the article, highlighting the conclusions reached and suggesting directions for future investigations in this field.

PROBLEM DESCRIPTION

Accurate forecasting of the power generated by wind turbines is challenged by the intrinsically volatile, non-stationary nature of wind. LSTM-based models have been prevalent in time series forecasting due to their ability to capture long-term dependencies (Srivastava et al., 2020). However, when applied directly to raw wind energy data, these models often fail to represent the acceptable variability expressed by the coefficient of detail in wavelet decomposition (Bali et al., 2019). This limitation is due to the architecture of LSTM, which, although efficient in detecting patterns over time, may not
capture the rapid and transient nuances of the detail signals, especially in time series with a high degree of randomness and abrupt fluctuations.

In this study, we address this methodological gap by incorporating wavelet decomposition as a preliminary step to isolate the components of detail and wind speed time series approximation. By treating these components separately, we seek to overcome the limitations of LSTM in capturing rapid and transient fluctuations, using the Random Forest model to predict the behavior of the coefficients of detail. The model is then challenged to capture the complex dynamics of the wind, which is critical for accurate wind power prediction and the subsequent stability and efficiency of power generation.

Wind power generation is represented as a time series, where it is the energy generated over time. This time series is influenced by previous data and is characterized by its non-stationary nature. \( \{P_t\} \)

The time series is modeled as a stochastic process, where the value of is determined by a function that maps previous values of the series of Equation 1. \( P_t \in \mathbb{R} \)

\[
P_t = f(P_{t-1}, P_{t-2}, \ldots, P_{t-h}, \epsilon_t, q)
\]

And a white noise term. White noise follows a normal distribution with zero average. \( \#t \)

The motivation is to predict the value accurately. Due to the complexity and nonlinear nature of the function, traditional analytical methods are inadequate to obtain an accurate prediction. \( P_t \in \mathbb{R} \)

The first step involves decomposing the wind energy time series into approximation and detail components using the Discrete Wavelet Transform (DWT). This decomposition helps to deal with the non-stationarity of the time series.

An independent Random Forest model then processes each component of the decomposed time series. Random Forest is chosen because of its ability to handle non-linear and complex relationships in the data.

After processing by the Random Forest models, the predicted values of each component are summed to obtain the final wind generation forecast. Then, the predicted value of energy generation over time is obtained by Equation 2:

\[
\hat{P}_t = \hat{f}(P_{t-1}, P_{t-2}, \ldots, P_{t-h}, \epsilon_t, q)
\]
Here, it is the approximate prediction function obtained through the DWT model combined with Random Forest, which considers both the previous values of the time series and the white noise in predicting the value of $\hat{f}_P_t$

This method proposes an innovative and more accurate approach to wind energy forecasting, which is vital for efficiently managing modern power grids.

**MATERIALS AND METHODS**

The proposed forecasting method, DWT Random Forest, has two essential components: Random Forest and DWT. The two components are used to implement the divide-and-conquer strategy. Specifically, the proposed method uses DWT to decompose the original wind energy data into sub-signals and independent Random Forest models to learn the temporal relationship of each sub-signal, respectively.

The quality and quantity of datasets are crucial to achieving better forecasting results. Getting real-time data is always a limitation in wind speed forecasting problems. For this study, wind speed data with daily sampling intervals were collected from the wind station in Salvador, Bahia State, Brazil. The data were provided by the National Institute of Meteorology (INMET), covering the period from May 12, 2000 to July 14, 2023.

The Salvador station, code A401, is located at the geographical coordinates of latitude -13.005515 and longitude -38.50576, with an altitude of 47.56 meters. The station is in continuous operation, ensuring the reliability and consistency of the data collected. In total, 34,477 data were collected, divided into two main categories: Average Speed and Maximum Gust of the wind.

The data collected is essential for analyzing and forecasting wind power generation in the region. The daily periodicity of the measurement provides a comprehensive view of the variations and trends in wind speed over time. The analysis of this data provides valuable insights for wind energy modeling and forecasting, contributing to better management and planning in the renewable energy sector.

Detailed statistics of the data collected from the wind season are presented in Table 1. In addition, the distribution of wind speed data is illustrated in Figures 1 and 2 using a distribution plot and a compass rose diagram. The compass rose diagram
highlights the frequency and direction of the wind, providing a visual understanding of
the prevailing wind characteristics in the region.

**Figure 1** – Average variation of wind speed in the season over the years and by season

**Figure 2** – Average variation of the maximum gust in the season over the years and by season
Table 1 – Statistics of Station 401 data

<table>
<thead>
<tr>
<th></th>
<th>RMD</th>
<th>VMD</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>1.388286</td>
<td>8.216410</td>
</tr>
<tr>
<td>Std</td>
<td>0.198763</td>
<td>0.925158</td>
</tr>
<tr>
<td>min</td>
<td>0.973173</td>
<td>6.175991</td>
</tr>
<tr>
<td>25%</td>
<td>1.220129</td>
<td>7.477127</td>
</tr>
<tr>
<td>50%</td>
<td>1.345403</td>
<td>8.298977</td>
</tr>
<tr>
<td>75%</td>
<td>1.554575</td>
<td>8.879313</td>
</tr>
<tr>
<td>max</td>
<td>1.885666</td>
<td>10.251588</td>
</tr>
</tbody>
</table>

Source: prepared by the authors (2023)

The Wavelet Transform is a data decomposition method that finds applications in several areas, such as image processing for feature extraction in image segmentation and data denoise (Berrezek et al., 2019). Noise is defined as a rapid variation in measurement, and data is often corrupted by noise during its acquisition or transmission (Ahmed & Mohammed, 2023). Noise removal in signals is separating the noise part from the actual signal. One of the goals of the Wavelet Transform is to denoise the actual signal, thus improving the predictive ability of the prediction model. The wavelet-based technique requires a sufficient amount of data to produce better results.

Wavelets are mathematical functions that decompose signals into different frequency components, with an approximation signal (low frequency) and a detail signal (high frequency). The Wavelet Transform can be of two types: continuous and discrete.

The Continuous Wavelet Transform (CWT) of a signal can be defined by Equation 3, while the Discrete Wavelet Transform (DWT) is described by Equation 4. CWT is useful for signal analysis, where the exact location of frequency variations is crucial. At the same time, DWT is better suited for analyses where computational efficiency and compact signal representation are essential.

\[
\text{CWT}(x)(a, b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} x(t) \cdot \psi^* \left( \frac{t-b}{a} \right) \, dt \tag{3}
\]

Here, \( \psi(t) \) is the wavelet function, \( a \) is the scaling factor, \( b \) is the translation factor, and \( \psi^*(t) \) is the complex conjugate of \( \psi(t) \). CWT provides a continuous representation of the signal regarding scale and position.

DWT, on the other hand, is defined as:

\[
\text{DWT}(x)(j, k) = \sum_{n} x[n] \cdot \psi_{(j,k)}[n] x(t) \cdot \psi^* \left( \frac{t-b}{a} \right) \, dt \tag{4}
\]
Where \( \psi_{j,k}(t) \) is the discretized wavelet function, with \( j \) and \( k \) representing the scale and translation indices, respectively. DWT efficiently represents the signal in terms of approximation coefficients and detail at different resolution levels.

The choice between CWT and DWT depends on the signal's nature and the analysis's objectives. In many practical applications, especially in time series analysis and signal processing, DWT is preferred due to its efficiency and ability to capture essential signal characteristics at different levels of detail.

Random Forest is a machine learning method for classification, regression, and other tasks, which operates by constructing a multiplicity of decision trees at the time of training and producing the class that is the mode of the classes (classification) or the mean/median of the predictions (regression) of the individual trees.

Each decision tree is constructed using a random subset of the training dataset. This process is known as bootstrap aggregating or bagging. A vector of random variables is selected for an individual decision tree, and the best-split point between the variables is used to divide the training set into subsets. This process is repeated recursively for each derived subset.

After building multiple decision trees, Random Forest combines the predictions from each tree. For classification problems, the final prediction is made by majority vote, where the most frequent class among all trees is chosen. In regression problems, the mean or median of the predictions for all trees is calculated.

Let \( Y \) be the output variable and the input variables. An individual decision tree is represented as \( h(X, \Theta_k) \), where \( \Theta_k \) are the random parameters of the \( k \)th tree. The Random Forest predictor for regression and classification is given, respectively, in Equation 5.\( X_1, X_2, \ldots, X_p \)

\[
\hat{Y} = \frac{1}{B} \sum_{k=1}^{B} h(X, \Theta_k)
\]

Where \( B \) is the number of trees in the forest.

For ranking, the final prediction is the class with the most votes among all trees.
Data quality and quantity are crucial to achieving better forecasting results. This study used wind speed data collected from the Salvador weather station (Station Code: A401), located in Salvador, Bahia, Brazil. The data covers the period from 12 May 2000 to 14 July 2023, with a daily measurement periodicity. Thirty-four thousand four hundred seventy-seven observations were collected, including average wind speed (VMD) and maximum gusts (RMD) measurements. This data was provided by the National Institute of Meteorology (INMET), an organization dedicated to monitoring and analyzing meteorological data in Brazil. The predictive accuracy of the proposed hybrid model for wind speed prediction is analyzed using these datasets.

Data transformation is a pre-processing step that converts numerical data to the required scale without degrading the results. Among the various data transformation techniques, Min-Max normalization was adopted in this study. The resized data will be between 0 and 1. The equation for Min-Max normalization is shown in Equation 6, where $x$ is the original data, $x_0$ is the normalized data, min is the minimum value, and max is the maximum value of the samples.

$$x_0 = \frac{x - \text{min}}{\text{max} - \text{min}}$$  \hspace{1cm} (6)
Wavelet Transform is one of the most popular signal processing algorithms for performing signal denoising and decomposition for wind speed prediction. This study employs the Wavelet Transform to decompose the wind speed data into low and high-frequency components. The Discrete Wavelet Transform (DWT) is selected due to its computational efficiency and data compression capability compared to the Continuous Wavelet Transform (CWT). The Wavelet Transform is applied to enhance the predictive ability of the models by removing the noise present in the data. Decomposition results in a single low-frequency or approximation component and many high-frequency or detail components. The signal’s identity is stored in the approximation component, while the detail components reveal the essence of the signal. Because DWT breaks down signals in frequency and timing, it will provide a more accurate representation of the sequential behavior of the data. This study applies the Discrete Wavelet Transform (DWT) with Daubechies (db7) to the original wind series up to two levels.

The multi-step time series forecasting task aims to predict the following K values (xt+1, xt+2, ..., xt+K) of a historical time series with t observations (x1, x2, ..., xt), where K is the forecast horizon. This study’s forecast horizons selected for analysis are the next 5 hours. A recursive multi-step forecasting strategy is employed in this work to predict wind speed. This recursive multi-step method initially develops a single-step prediction model, which is repeatedly employed to predict the entire horizon. The previously predicted values are input to predict the next step in time.

The proposed WT-Random Forest model is developed by combining the characteristics of the wavelet transform and the Random Forest. Random Forest is known for handling long-term dependency issues and offers reliable performance with time-series data. Therefore, a Random Forest is selected to predict low-frequency components. A Random Forest-based predictive model with 100 trees in the hidden layer is developed and trained to predict low-frequency signals. An optimal delay value of 27 is chosen for the experiments. The calculation of the optimal delay value is reported in section the model. The entire dataset is separated into training and testing datasets in a 70:30 ratio, of which 70% of the data is used to train the model, and the remaining 30% is used to test the model’s effectiveness. In 70% of training datasets, 10% is used to validate model performance.

Recurrent neural networks are efficient models for multi-step temporal series prediction. Random Forest networks, a variant of recurrent neural networks, take longer
to train when dealing with recursive multi-step problems on large datasets. In this study, there are three sets of high-frequency signals, and predicting these high-frequency signals using Random Forest can be computationally expensive in terms of training time. Therefore, the Random Forest model, with radial-based kernel function, was selected to predict high-frequency signals. Support vector regression (SVR) exhibits a remarkable generalization ability to achieve an optimal solution. In addition, SVR adopts structural risk minimization rather than empirical risk minimization, and therefore, SVR offers superior performance. These are the main reasons for selecting the Random Forest in this study to predict high-frequency subseries of wind speed. Finally, the prediction results of each sub-series are aggregated to obtain the final prediction results.

The performance of the proposed hybrid model is extensively evaluated using two criteria. Statistical error indicators such as MAE (Mean Absolute Error), RMSE (Root Mean Squared Error), and MAPE (Mean Absolute Percentage Error) are employed to evaluate the effectiveness of the model. The different statistical error indices used to evaluate the performance of the models are defined using Equations 7 to 9.

\[
MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i| \quad (7)
\]

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

\[
MAPE = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (9)
\]

The \(y_i\) represents the observed values, \(\hat{y}_i\) represents the values predicted by the model, and \(n\) is the total number of observations.

These metrics provide a quantitative assessment of the accuracy of the model's predictions. MAE measures the average magnitude of errors in a set of forecasts without regard to direction. The RMSE measures the dispersion of prediction errors and gives an idea of the magnitude of the errors. MAPE expresses accuracy as a percentage and helps compare the accuracy of different forecasting methods in the same data set.

RESULTS AND DISCUSSION

Figure 4 illustrates the Mean Square Error (MSE) with the number of lags for a wind speed data set modeled using Random Forest. The graph's Y-axis represents the
MSE, which measures how well the model fits the data, with smaller values indicating a better fit. The X-axis shows the lags used in the model, where each point indicates the use of data from a specific previous time interval to predict future values. The graph shows a decrease in MSE, with the number of lags increasing to approximately 10, suggesting that the model improves as more historical data is included. However, from then on, the MSE begins to fluctuate and increase, which may indicate an overfitting of the model to the training data. This suggests that a moderate number of lags, possibly around 10, may be optimal for wind speed prediction with the Random Forest algorithm, balancing model accuracy and complexity.

**Figure 4** – Lag Mean Square Error (MSE) for Wind Speed Data Trained by Random Forest

The original wind speed data is decomposed into various approximation and detail components using the wavelet transform, aiming to improve the prediction model's overall performance. In this study, the Discrete Wavelet Transform (DWT) with the Daubechies function (db7) is employed to decompose the original wind speed series into up to three levels. The results of this decomposition of the original wind speed data are illustrated in Figure 5. It can be seen that each series of wind velocities is disintegrated into approach and detail components, which are denoted by cA4, cD1, cD2, cD3, and
cD4, respectively, where cA4 represents the approach component, and cD1, cD2, cD3, cD4 represent the detail components.

The effectiveness of the proposed DWT-Random Forest model is evaluated by comparing it with the literature, as we see in Table 2, which demonstrates the accuracy of the multi-step prediction of all MAE, RMSE, and MAPE-based models for various prediction horizons. The proposed model achieves lower error values for all prediction horizons than other models. The lower the error values, the better the accuracy of the prediction. It can be understood from the table that error values increase as the prediction horizon increases for all prediction horizons.

This implies that the model's efficiency decreases as the prediction horizon's duration increases. The results illustrate that hybrid models perform better than single models.

Table 2 presents the performance metrics of the Random Forest model in predicting wind speed, comparing the results with previous studies found in the literature.
Table 2 – Random Forest Model Performance Metrics

<table>
<thead>
<tr>
<th>Author</th>
<th>MAP</th>
<th>RMSE</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>DWT-Random Forest</td>
<td>0.0099</td>
<td>0.0996</td>
<td>0.0779</td>
</tr>
<tr>
<td>Liu, Wu, &amp; Li (2020)</td>
<td>0.9942</td>
<td>0.2084</td>
<td>1.2</td>
</tr>
<tr>
<td>KU &amp; Kovore (2021)</td>
<td>1.406</td>
<td>1.146</td>
<td>13.555</td>
</tr>
<tr>
<td>Mohammed, Ahmed, et al. (2023)</td>
<td>0.0072</td>
<td>0.02683</td>
<td>2.32400</td>
</tr>
</tbody>
</table>

Source: prepared by the authors (2023)

Regarding the "DWT-Random Forest" model, it is observed that it achieves an MSE of 0.0099, an RMSE of 0.0996, and a MAPE of 0.0779. These metrics indicate that the model performs generally well in predicting wind speed. The MSE and RMSE are relatively low, suggesting that the mean squared errors and their square roots are minor, which is positive. The MAPE of 0.0779 indicates that the forecasts have an average absolute error of 7.79%, which is also acceptable, especially if we consider the context of the wind speed forecast, which can be affected by many variables.

On the other hand, the study by Mohammed et al. (2023) obtained remarkable results, with an MSE of 0.0072, an RMSE of 0.02683, and a MAPE of 2.32400. These values represent an exceptional performance in predicting wind speed compared to the other models. The MSE and RMSE are very low, significantly minimizing mean squared errors and their square roots. The MAPE of 2.32400 is slightly higher, but it is still considered entirely accurate for predicting wind speed.

In the study by Liu et al. (2020), there is an MSE of 0.9942, an RMSE of 0.2084, and a MAPE of 1.2. Although the MSE is higher than the DWT-Random Forest model, it is still reasonably low. The RMSE is also acceptable, indicating a moderate dispersion of errors. The MAPE of 1.2 suggests relatively good accuracy in the forecasts.

Finally, the study by Ku et al. (2021) presents an MSE of 1,406, an RMSE of 1,146, and a MAPE of 13,555. These values indicate that the performance of this study is the least accurate among the four compared. The MSE and RMSE are higher, suggesting more significant mean squared errors and square roots of errors. The MAPE of 13,555 is relatively high, indicating a lower accuracy than the other studies.

The results show that the "DWT-Random Forest" model has a respectable performance in predicting wind speed, with MSE, RMSE, and MAPE metrics within acceptable limits. However, it is surpassed in terms of accuracy by the exceptional results of the study by Mohammed et al. (2023). The other studies also show reasonable performance, although with different levels of precision. Therefore, there is evidence that
the "DWT-Random Forest" model could benefit from adjustments and improvements to achieve more competitive results compared to previous studies.

In Figure 6, the blue line represents the actual data, and the orange line shows the predictions made by Random Forest. The figure shows that the forecasts follow the general trend of the actual data but with less variation. This is typical of Random Forest predictions, as the method tends to smooth out fluctuations by averaging predictions from multiple decision trees.

Figure 6 – Results obtained for prediction

![Previsões Random Forest](image)

Source: prepared by the authors (2023)

The model appears to capture the trend of the actual data reasonably. However, we cannot assess exact accuracy without statistical metrics such as Mean Square Error (MSE), Root Mean Square Error (RMSE), or R-squared. This data is commonly used to evaluate the performance of a prediction model and to visually check how well the predictions align with the observed data.
CONCLUSION

The integration between Wavelet Decomposition and the Random Forest model will prove to be a promising methodology in wind speed forecasting, as evidenced by the graph presented. The multi-pronged approach allows for a more granular and comprehensive analysis, capturing the specific trends and fluctuations inherent in the time series of wind energy data. Random Forest's ability to model nonlinear complexities, combined with wavelet decomposition accuracy in distinguishing different frequencies and signal characteristics, offers a significant advance over traditional forecasting techniques.

The described method improves the accuracy of short-term forecasts, a crucial element for the effective management of wind power generation. The ability to accurately anticipate wind power can lead to optimization in the operation of wind farms, facilitating more reliable and efficient energy planning. The technique also highlights the importance of hybrid approaches in predictive modeling, where merging different analytical methods can overcome the limitations of isolated methods.

Finally, validating this methodology through accurate data from three wind farms confirms its superiority in predictive accuracy. This reinforces the practical relevance of combining these advanced techniques and paves the way for future investigations and applications in other areas of time forecasting. The approach presented has the potential to significantly transform how the wind industry approaches power forecasting, positively impacting the integration of renewables into the energy system.

REFERENCES


