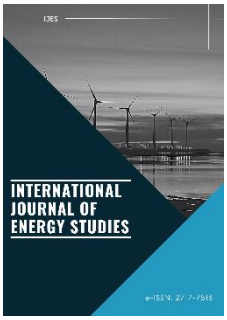
INTERNATIONAL JOURNAL OF ENERGY STUDIES

e-ISSN: 2717-7513 (ONLINE); homepage: https://dergipark.org.tr/en/pub/ijes



Research Article
Int J Energy Studies 2026; 11(1): 703-731
DOI: 10.58559/ijes.1847604

Received : 30 Dec 2025
Revised : 12 Mar 2026
Accepted : 12 Mar 2026

A comparative analysis of iterative and variational decomposition strategies for wind speed forecasting: LMD and VMD

Tuğçe İnağ[a*], Yasin İnağ[b]

[a]*Department of Industrial Engineering, Gazi University, Ankara, Türkiye, ORCID: 0000-0002-8800-6727*

[b]*Information Technologies Department, Council of Higher Education (YÖK), Ankara, Türkiye, ORCID: 0000-0002-7590-9345*

*(*Corresponding Author:tugceduzce@gazi.edu.tr)*

Highlights

- Variational and iterative decomposition methods are comparatively analyzed under identical modeling conditions.
- Decomposition-based preprocessing significantly improves LSTM forecasting accuracy compared to the baseline LSTM model.
- The VMD–LSTM framework demonstrates superior robustness and statistical performance over the LMD–LSTM approach.
- At least 3, max 5 highlights should be listed here

You can cite this article as: Inag T, Inag Y. A comparative analysis of iterative and variational decomposition strategies for wind speed forecasting: LMD and VMD. Int J Energy Studies 2026; 11(1): 703-731.

ABSTRACT

This paper investigates the role of data preprocessing techniques in improving the forecasting performance of highly volatile time series. Although prior studies predominantly emphasize complex hybrid model architectures or extensive hyperparameter optimization, the isolated effect of the mathematical foundations of decomposition methods on prediction accuracy remains relatively underexplored. To address this issue, two fundamentally different decomposition techniques—Variational Mode Decomposition (VMD), which follows a variational optimization framework, and Local Mean Decomposition (LMD), which is based on an iterative scheme—are comparatively applied to nonlinear and highly fluctuating wind speed time series. Following decomposition, each extracted component is forecast individually and subsequently reconstructed using a Long Short-Term Memory (LSTM) network. In addition, an LSTM model trained directly on raw data is employed as a benchmark to assess the effectiveness of the proposed hybrid approaches. The experimental results demonstrate that incorporating a preprocessing stage substantially enhances forecasting accuracy. Both hybrid models outperform the baseline LSTM model in terms of prediction error metrics. Among them, the VMD-based approach yields the lowest error values and exhibits superior robustness and stability when compared with the LMD-based model. The statistical significance of the observed performance differences is further validated through the Diebold–Mariano test, confirming the dominance of the proposed VMD–LSTM framework at the 1% significance level ($p < 0.01$). Overall, the findings underline the critical importance of data preprocessing in forecasting tasks involving highly noisy and volatile time series and indicate that variational-based decomposition offers notable advantages in terms of stability and reproducibility over iterative methods.

Keywords: Wind speed forecasting, Data preprocessing, Variational mode decomposition (VMD), Local mean decomposition (LMD), Long short-term memory (LSTM)

1. INTRODUCTION

The increasing penetration of renewable energy sources into modern power systems has intensified the need for accurate and reliable wind speed forecasting. [1]. Due to the inherently stochastic and highly variable nature of wind, precise prediction of wind speed plays a crucial role in ensuring efficient planning, stable operation, and effective integration of wind energy systems into electrical grids. Inaccurate forecasts may lead to operational inefficiencies, increased balancing costs, and reduced system reliability.[2]. Consequently, developing robust forecasting frameworks capable of handling the nonlinear and non-stationary characteristics of wind speed time series has become a central research focus in the field of wind energy systems.

In recent years, decomposition-based approaches employed in the data preprocessing stage have demonstrated remarkable results [3]. These methods decompose complex and stochastic time series into their intrinsic subcomponents, thereby revealing the underlying structural characteristics and hidden patterns within the signal. Techniques such as Variational Mode Decomposition (VMD) and Local Mean Decomposition (LMD) have emerged as effective preprocessing methods, particularly for handling high-frequency and noise-contaminated signals [4].

This study aims to quantitatively investigate the influence of data preprocessing techniques on the predictive efficacy of high-volatility time series, including wind speed. In this context, three different scenarios were established using the same dataset. First, a direct LSTM model (i) without any preprocessing was applied to the raw data. This model served as the reference model. In the second scenario, the LMD-based preprocessing step was applied, followed by training the LMD-LSTM model (ii). In the third scenario, a VMD-LSTM model (iii) was built using the data preprocessed with the VMD method.

These three models were trained under identical hyperparameter settings to isolate the effect of preprocessing differences. Model performance was assessed using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2). Furthermore, the Diebold–Mariano (DM) test was applied to evaluate the statistical significance of differences in forecasting accuracy. Accordingly, the effects of various preprocessing techniques on predictive performance were systematically analyzed.

The original contribution of this study to the literature lies in the quantitative demonstration of the impact of two distinct signal processing approaches—the locally iterative mean-based LMD and the globally variational optimization-based VMD—on forecasting performance. Previous studies in the literature have typically compared methods belonging to the same family [5]. In contrast, this study aims to identify the most robust approach for stochastic time series by comparing two fundamentally different mathematical frameworks in an environment isolated from model complexity. The findings indicate that the variational approach based on VMD significantly improves forecasting accuracy for high-variance data in a statistically meaningful way, thereby providing a methodological foundation for the development of future hybrid models.

2. LITERATURE REVIEW

Research on wind speed forecasting has gained significant momentum in recent years (Figure 1). Accurate forecasting of wind speed directly influences the reliability of energy management in the planning, production optimization, and grid integration processes of renewable energy systems [2]. In this context, a wide range of approaches—both statistical and machine learning-based—have been developed in the literature for wind speed forecasting. However, due to the inherently high volatility of wind speed, forecasting models trained directly on raw data often exhibit reduced performance. This challenge has made the investigation of the potential impact of data preprocessing and signal decomposition methods on forecasting accuracy an important research direction in recent literature. Recent studies have demonstrated that decomposition-based preprocessing steps can separate the structural components of complex signals, thereby enabling learning models to produce more stable and accurate forecasting results.

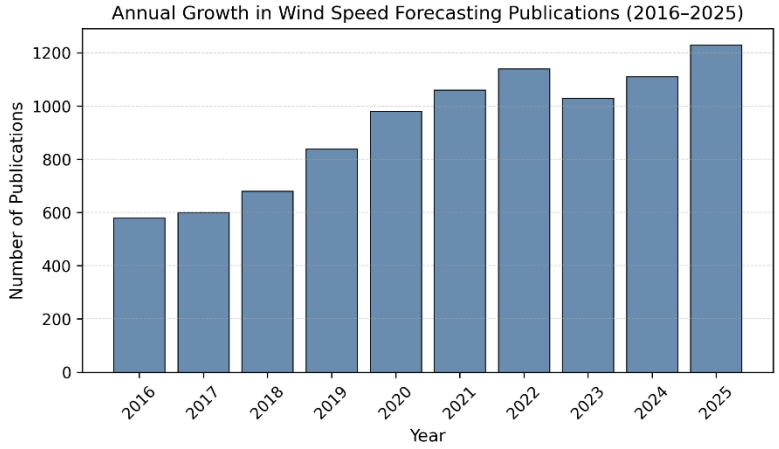


Figure 1. The annual increase in the number of publications on wind speed forecasting between 2015 and 2025 (data obtained from the Web of Science database).

Within this framework, the methodological evolution of wind speed forecasting approaches can be better understood by examining both traditional and data-driven models.

Wind speed forecasting plays a critical role in various fields such as energy production, aviation, maritime operations, and disaster management. Traditionally, numerical weather prediction (NWP) models have been employed for this purpose. However, these methods are constrained by high computational costs and low temporal resolution due to the complexity of the differential equations involved. Consequently, in recent years, machine learning (ML) and deep learning (DL)-based approaches have gained increasing prominence in wind speed forecasting [3].

In the early stages of wind speed forecasting, methods such as AutoRegressive Integrated Moving Average (ARIMA), AutoRegressive Moving Average (ARMA), Weibull distribution, and Fuzzy Logic were frequently used for short-term forecasting [6].

Although these models are capable of capturing linear dependencies in time series, they are inadequate for representing nonlinear and complex structures [7]. As alternatives to numerical models, methods such as Support Vector Regression (SVR) and Decision Trees have been developed, demonstrating better performance than ARIMA in modeling nonlinear relationships. Among machine learning techniques, one of the most widely used approaches is Artificial Neural Networks (ANNs) [8, 9].

In recent years, deep learning-based models have gained significant attention in wind speed forecasting. The primary models employed in this context include LSTM, Bidirectional LSTM (Bi-LSTM), Convolutional Neural Networks (CNNs), and Graph Neural Networks (GNNs) [3, 10-16]. Deep learning models have substantially improved forecasting accuracy for complex time series data. In particular, LSTM models have been reported to outperform other models with considerably lower error values [3]. Furthermore, several studies have proposed hybrid models that combine multiple algorithms, which enhance learning capability while minimizing prediction errors [17]. Despite these methodological advances, the inherent characteristics of wind speed data continue to pose challenges for forecasting accuracy.

Due to the nonlinear and non-stationary nature of wind speed time series, it has been reported that forecasting models trained on raw data often produce high error values [1]. Therefore, in the

literature, Empirical Mode Decomposition (EMD) and its variants, EEMD and CEEMDAN, have been widely used as data preprocessing techniques. These methods decompose the signal into Intrinsic Mode Function (IMF) components representing different frequency bands, allowing learning models to capture each component in a more stable manner. Previous studies have shown that EMD–LSTM, EEMD–LSTM, and CEEMDAN–LSTM hybrid architectures provide significant improvements in forecasting accuracy compared to LSTM models trained with raw data. Typically, these hybrid approaches have been reported to reduce RMSE values by approximately 20–45% [18-25]. Especially in short-term forecasting, it has been emphasized that noise reduction and component-based learning significantly enhance forecasting accuracy [3].

However, most existing studies have focused on a specific decomposition method and a specific forecasting model. The effects of different decomposition techniques on the same model and dataset have been examined only in a limited number of studies. Even these studies are generally restricted to methods derived from the same algorithmic family [26]. In recent years, due to the structural limitations of EMD-based methods—such as mode mixing and end effects—VMD method has been proposed as an alternative preprocessing approach in wind speed forecasting [27-29]. VMD decomposes the signal within a variational optimization framework, simultaneously optimizing the center frequency and bandwidth of each mode. This mathematical structure makes VMD more robust to noise and spectrally stable.

Several recent investigations have demonstrated that VMD-based hybrid models, including VMD–LSTM, VMD–CNN, and VMD–GRU, yield improved RMSE and MAE performance relative to EMD-derived approaches [30-32]. Although the effectiveness of decomposition-based preprocessing techniques in wind speed forecasting has been widely recognized, the majority of existing research has focused on a single decomposition method and a single forecasting model. In particular, studies that perform a direct and controlled comparison of decomposition methods with fundamentally different mathematical principles under the same dataset, model architecture, and hyperparameter configuration are limited [33, 34]. In addition to methodological limitations, the evaluation strategies adopted in existing studies also warrant further scrutiny.

Furthermore, in most existing studies, the statistical significance of the performance differences between models has not been evaluated. The obtained results have generally been interpreted only based on the absolute values of error metrics [25–27]. This makes it difficult to objectively assess

whether different preprocessing techniques truly provide a significant improvement in forecasting performance.

In response to these gaps, the original contribution of this study is positioned as follows. The original contribution of this study lies in the quantitative analysis of the impact of the data preprocessing step on wind speed forecasting performance, isolating the effect of methodological differences. In this context, three approaches were compared using the same dataset, identical LSTM architecture, and fixed hyperparameters: the LSTM model without preprocessing, the iterative-based LMD-LSTM, and the variational-based VMD-LSTM. In addition, the performance differences between the models were evaluated not only through error metrics but also statistically using the DM test. In this respect, this study presents a more comprehensive and reliable framework for revealing the role of decomposition-based preprocessing techniques in wind speed forecasting.

3. DATASET

The dataset used in this study was obtained from the publicly available "Wind Turbine SCADA Dataset" on the Kaggle platform [35]. This dataset, which is widely accepted as a benchmark standard in the literature, encompasses a wide variety of characteristics, including periods of stationary states, sudden ramps, and storm regimes with high volatility. The choice of this standard dataset, rather than using data from multiple locations, aims to minimize external geographical uncertainties. Thus, the performance differences between the proposed hybrid models are analyzed in an isolated environment, where the influence of data diversity is eliminated, allowing for a direct comparison of the structural properties of the mathematical decomposition frameworks under examination.

The dataset consists of Supervisory Control and Data Acquisition (SCADA) measurements collected from a real wind turbine. As a result, the measurements were collected in a high-frequency and continuous time series format. The dataset includes measurements taken every 10 minutes and contains key parameters such as wind speed. The general behavior of the wind speed time series is illustrated in Figure 2. A visual inspection reveals that the series exhibits a clear trend with high amplitude oscillations, demonstrating a nonlinear and volatile structure. This behavior aligns with the atmospheric turbulence and chaotic nature of wind speed, presenting a significant challenge for modeling the series.

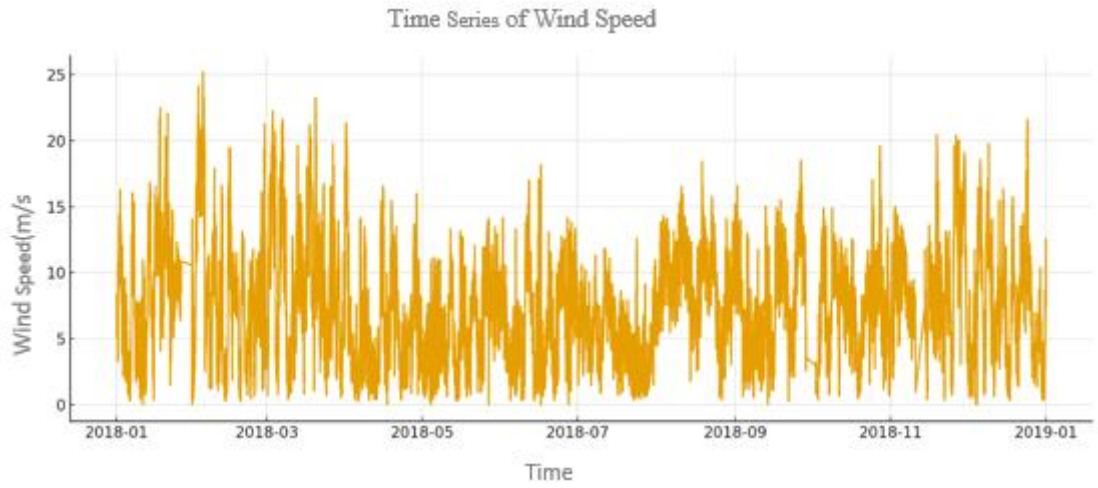


Figure 2. Time series of wind speed measurements.

The summary statistics of the wind speed variable are presented in Table 1.

Table 1. Descriptive statistics of the wind speed data.

Statistic	Value (m/s)
Mean	7.56
Standard deviation	4.23
Minimum	0
Maximum	25.20
Number of observations	50530

The mean wind speed of the series is 7.56 m/s, and the relatively high standard deviation (4.23 m/s) indicates the high volatility of the wind speed data.

The graphical analysis of the time series reveals the presence of seasonal fluctuations and short-term irregular components. This observation necessitates the decomposition of trend and seasonality effects during the modeling stage. Figure 3 presents the results of the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) analyses of the series. The ACF plot starts with a high value close to 1 at Lag 1 and then gradually and smoothly decays toward zero. This indicates that the series exhibits strong persistence and a distinct trend component, while the influence of high-frequency noise remains limited. In the PACF plot, a sharp positive spike is observed at Lag 1, followed by a rapid decline of correlations toward zero. This pattern suggests

the presence of strong short-term dependence, whereas long-term linear relationships are relatively weak.

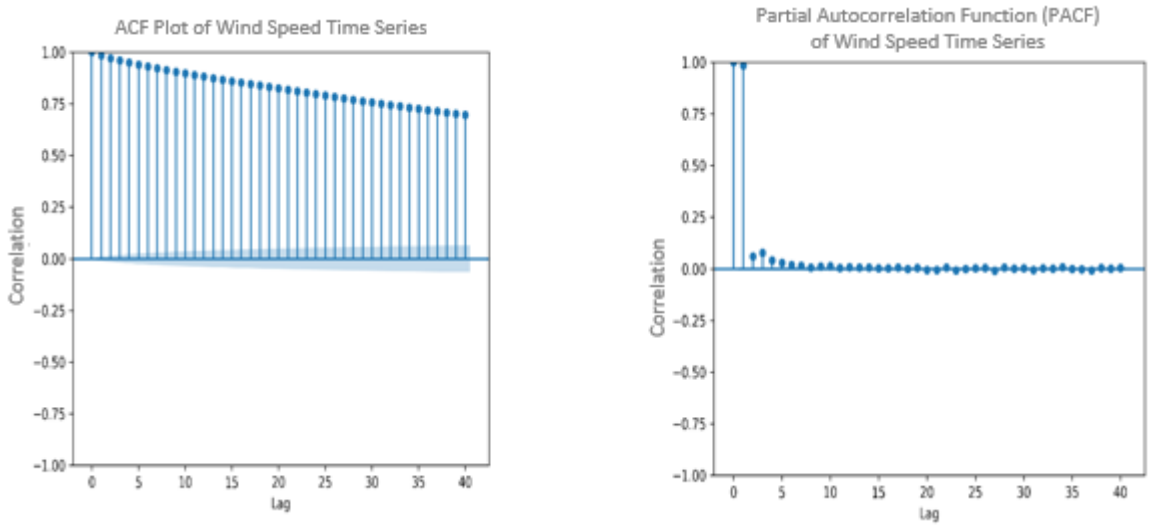


Figure 3. ACF and PACF analysis of the wind speed time series.

These findings indicate that the raw series is non-stationary and trend-dominant, implying that direct modeling of the signal may lead to high forecasting errors. Therefore, decomposing the signal into its low-frequency components through preprocessing constitutes an essential step in improving forecasting accuracy.

3.1. Construction of Datasets

The data used in this study consist of the raw wind speed time series obtained from a single wind turbine. In the modeling process, the purpose was not to transform the dataset into different forms but to examine the effect of different data preprocessing approaches applied to the same data on forecasting performance. Therefore, the dataset was created only once, and each preprocessing method was subsequently applied independently. In this way, the input structure of the model remained unchanged, while only the statistical representation of the data was altered.

The temporal order of the data was preserved, and the dataset was divided into 80% training and 20% testing subsets. The same split was used consistently across all experiments. To ensure scale consistency among model inputs, all data were normalized to the range [0, 1] using Min–Max normalization. The Min–Max scaler was fitted exclusively on the training data and then applied to the test data using the same scaling parameters, thereby preventing data leakage.

3.2. Data Decomposition

In highly volatile time series such as wind speed, factors such as noise, trend, and seasonality directly affect forecasting performance. Therefore, the data preprocessing step is critically important for noise reduction, signal balancing, and the decomposition of underlying structural components.

Although EMD and its variants have been frequently used in the literature for this purpose, these methods often suffer from stability issues due to structural problems such as mode mixing and end effects [18, 20-23].

In contrast, VMD formulates the signal within a variational framework, optimizing modes under a predefined bandwidth constraint. This mathematical structure makes VMD more robust to noise and theoretically more stable. On the other hand,LMD , although similar in its iterative structure to EMD, partially mitigates end effects and oscillation errors analytically through its local mean envelope estimation technique.

While the requirement of prior parameter specification in VMD—such as the number of modes (K)—may appear to be a limitation, it also provides flexibility in decomposition. Conversely, LMD balances this complexity through its self-adaptive, parameter-free nature. In this study, VMD, representing mathematical rigor, and LMD, representing a stable adaptive decomposition approach, were selected to comparatively analyze the performance of two fundamentally different methodological frameworks.

3.2.1. Variational mode decomposition

VMD is a variational signal processing technique proposed by Dragomiretskiy and Zosso (2014) that decomposes a signal into a set of mode components with distinct bandwidths [36]. The main objective of the method is to represent a given signal as the sum of a finite number of sub-modes and to decompose it in the frequency domain by minimizing the bandwidth of each mode.

The VMD algorithm aims to determine a set of mode functions $u_k(t)$and their corresponding center frequencies $\omega_k$by solving the following variational optimization problem given in Eq. (1).

$$\min_{u_k,w_k}\left\{\sum_{k=1}^{K}\left\|\partial_t\left[\left(\delta(t)+\frac{j}{\pi t}\right)*u_k(t)\right]e^{-jw_kt}\right\|_2^2\right\} \quad (1)$$

subject to the reconstruction constraint given in Eq. (2).

$$\sum_{k=1}^{K}u_k(t)=f(t) \quad (2)$$

where K denotes the total number of modes to be extracted and is defined as a positive integer. u_k, denotes the k-th mode component and ω_k indicates the center frequency corresponding to each mode obtained after decomposition. $\delta(t)$, is the Dirac delta function.

An unconstrained variational representation is obtained from Eqs. (1)–(2) by introducing a Lagrange multiplier (λ) and a penalty parameter (α).

As a result of this transformation, the augmented Lagrangian expression given below is obtained, which preserves the reconstruction accuracy while ensuring the stability of the solution. The resulting augmented Lagrangian functional is defined in Eq. (3).

$$\begin{aligned}L(\{u_k\},\{w_k\},\lambda)&=\alpha\sum_{k=1}^{K}\left\|\partial_t\left[\left(\delta(t)+\frac{j}{\pi t}\right)*u_k(t)\right]e^{-jw_kt}\right\|_2^2\\&+\|f(t)-\sum_{k=1}^{K}u_k(t)\|_2^2+\langle\lambda(t),f(t)-\sum_{k=1}^{K}u_k(t)\rangle\end{aligned} \quad (3)$$

L denotes the augmented Lagrangian functional. α is the penalty parameter that controls the bandwidth constraint, while $\lambda(t)$ represents the Lagrange multiplier enforcing the reconstruction constraint.

$$\sum_{k=1}^{K}u_k(t)=f(t) \quad (4)$$

The reconstruction constraint expressed in Eq. (4) ensures that each mode is concentrated within a narrow bandwidth around its corresponding center frequency, while preserving the overall data integrity. Thus, the VMD algorithm enables the effective separation of the signal into its distinct frequency components.

3.2.1. Local mean decomposition

LMD method, proposed by Jonathan S. Smith in 2005, is a time-domain analysis technique [37].

This method is a self-adaptive analytical approach that can be applied to nonlinear and non-stationary signals, and it was first used in computer-based electrical signal processing. LMD is an adaptive signal decomposition method capable of decomposing a complex, multi-component signal into a sequence of Product Function (PF) components and a residual function, ordered from high to low frequency. The fundamental assumption of LMD is that any signal $x(t)$can be modeled as the product of an instantaneous amplitude function $a_i(t)$and a pure frequency modulation signal $s_i(t)$ as expressed in Eq. (5):

$$PFi(t) = ai(t) * si(t) \quad (5)$$

The decomposition process is performed iteratively. First, the local mean of the signal is subtracted to obtain a pure frequency modulation (pure FM) component. Then, using the Hilbert transform, the instantaneous envelope $a_i(t)$ s calculated and normalized to extract the pure frequency modulation $s_i(t)$. The obtained PF component is subtracted from the signal, and this procedure is repeated on the remaining residual signal. The decomposition form of the signal in the LMD method is given in Eq. (6):

$$x(t) = \sum_{i=1}^{n} PF_i(t) + u_n(t) \quad (6)$$

Here, the components $PF_i(t)$represent the high-frequency and rapidly varying parts of the signal, while $u_n(t)$denotes the low-frequency residual component. During the decomposition process, the signal energy is preserved, and the sum of all components can reconstruct the original signal completely and accurately.

3.3. LSTM (Long Short-Term Memory)

LSTM networks were developed to overcome the vanishing gradient problem commonly encountered in traditional RNNs. They contain special memory blocks called *cells*, which consist of three main gates: the input gate, the forget gate, and the output gate. These gates control the flow of information within the cell. The forget gate is defined in Eq. (7):

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (7)$$

The forget gate determines how much of the information from the previous cell state should be retained or discarded. Here, $f_t$represents the output of the forget gate and takes a value between 0 and 1, defining the proportion of information to be retained. $\sigma$is the sigmoid activation function, which compresses the output to the range [0, 1]. $W_f$denotes the weight matrix of the forget gate, and $[h_{t-1}, x_t]$represents the concatenation of the

previous hidden state $h_{t-1}$and the current input x_t. $b_f$is the bias vector of the forget gate. This equation dynamically regulates the information flow by determining which information should be forgotten (values close to 0) and which should be retained (values close to 1) [38] . The input gate and the candidate cell state are defined in Eqs. (8) and (9), respectively:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (8)$$

$$\tilde{C}_t = \tanh\,(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (9)$$

The input gate is responsible for deciding which new information will be added to the cell. Here, $i_t$represents the output of the input gate, controlling which information is to be stored in the cell. $\tilde{C}_t$denotes the candidate cell state, while $W_i$and $W_C$correspond to the weight matrices of the input and candidate cell, respectively. Similarly, $b_i$and $b_C$are the bias vectors of these layers. The input information is scaled through the sigmoid (σ) and hyperbolic tangent (tanh) functions, which regulate the flow of information into the memory cell by constraining values within the [0, 1] interval, thereby contributing to a more stable learning process. The cell state update mechanism is given in Eq. (10):

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (10)$$

The cell state is refreshed by integrating the information preserved from the previous time step with the newly incorporated content. In this equation, $C_t$represents the updated cell state. The term $f_t * C_{t-1}$indicates the portion of past information preserved by the forget gate, while $i_t * \tilde{C}_t$represents the new information added by the input gate. Thus, the LSTM cell can maintain long-term dependencies by selectively remembering useful past information while discarding irrelevant details. The output gate and the hidden state are defined in Eqs. (11) and (12):

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (11)$$

$$h_t = o_t * \tanh\,(C_t) \quad (12)$$

The output gate determines how much of the information stored in the cell will be passed to the next layer or the next time step. Here, $o_t$represents the output of the output gate, controlling which information is propagated forward. $W_o$denotes the weight matrix of the output gate, and $b_o$is its bias vector. $h_t$represents the current hidden state, which is the network's output at that time step, while $C_t$denotes the updated cell state that carries the accumulated information within the cell. This mechanism enables LSTM networks to maintain a balanced flow of information by considering both past and present data during the forecasting process.

These equations collectively define how the LSTM cell determines which information should be exposed to the network output. They demonstrate that LSTM networks simultaneously consider the current input and the previous hidden state, allowing them to effectively capture long-term dependencies in time series data. The general architecture of the LSTM cell is illustrated in Figure 4, which is composed of three fundamental mechanisms; the input gate, forget gate, and output gate.

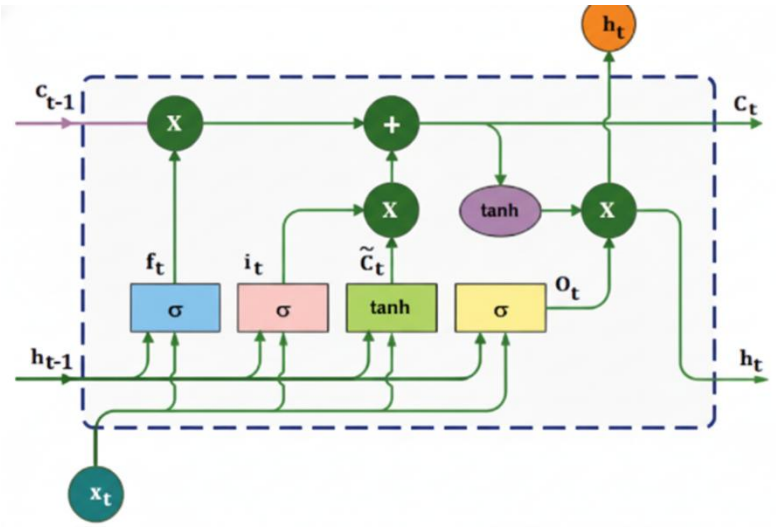


Figure 4. Schematic diagram of the LSTM cell [39]

This structure enables the LSTM to preserve long-term dependencies learned from past observations while remaining sensitive to short-term fluctuations. Therefore, in dynamic time series such as wind speed, it provides higher forecasting accuracy compared to traditional RNN models.

4. EXPERIMENTAL SETUP

In this study, the "Wind Turbine SCADA Dataset" publicly available on the Kaggle platform was used. The dataset contains 10-minute resolution SCADA measurements collected from a single wind turbine. The wind speed time series was used as the target variable in the analysis. The dataset was divided chronologically to preserve the temporal structure of the time series. Approximately 80% of the observations were used for model development (training and validation), while the remaining 20% were reserved for testing. A portion of the training data was used as a validation set during the training process for early stopping. The training, validation, and testing procedures were performed without shuffling the data. To avoid data leakage, the parameters of the Min–Max normalization were estimated using only the training data and then applied to the validation and

test sets. Early stopping was monitored using the validation set while ensuring that no information from the test data was used during the training process.

All experiments were conducted on three different model architectures. First, the raw wind speed series was directly forecasted using an LSTM model without any preprocessing.

Second, a hybrid VMD–LSTM model was constructed, in which the wind speed series was decomposed into a set of variational modes using the VMD method, and the final forecast was obtained by aggregating the individual mode-wise predictions.

Finally, a hybrid LMD–LSTM model was developed, where the wind speed series was decomposed into components using the LMD method, and each component was forecasted independently by a separate LSTM model.

All models were trained using identical hyperparameters to ensure a fair comparison. The LSTM layer consisted of 64 neurons with a tanh activation function, followed by a Dense(1) output layer. The Adam optimizer with its default learning rate (0.001) was employed, using a batch size of 32. A look-back window of 24 time steps was used to construct the input sequences. To prevent overfitting and ensure optimal convergence, an early stopping criterion was applied: training was monitored on a validation set and terminated if no improvement in validation loss was observed for five consecutive epochs, with a maximum of 50 epochs. This approach ensured that all models achieved a similar level of convergence while mitigating overfitting, allowing a direct and fair comparison of the impact of different preprocessing methods on forecasting performance.

Table 2. Hyperparameter settings used for all LSTM-based models.

Hyperparameter	**Value**
Model structure	LSTM(64, tanh) → Dense(1, linear)
Number of LSTM layers	1
Look-back window (time steps)	24
Activation function	tanh
Optimizer	Adam
Learning rate	Default (0.001)
Batch size	32
Maximum epochs	50 (early stopping applied)

The forecasting performance of the models was evaluated using three commonly adopted error metrics in the literature; RMSE, MAE, and the Coefficient of Determination (R^2). In addition to these metrics, the DM test was employed to assess the statistical significance of the differences in forecasting performance among the models. The DM test was used to determine whether the error distributions of two competing forecasting models differ significantly in a statistical sense.

All experiments were implemented using the Python programming language. The deep learning models were developed using the TensorFlow framework with the Keras API. The experiments were conducted in the Google Colab computational environment. Standard scientific computing libraries including NumPy, Pandas, and Matplotlib were used for data processing and visualization. To enhance reproducibility, all experiments were conducted under identical training conditions.

4.1. LSTM Model

The LSTM model was trained directly on the unprocessed wind speed data, without applying any signal decomposition or preprocessing method. The model's performance is illustrated in Figure 5. A visual inspection of the figure shows that the model successfully follows both the overall trend and the local variations of the wind speed series. This behavior indicates that the LSTM layer effectively captures short-term dependencies, although some high-frequency noise components are only partially represented.

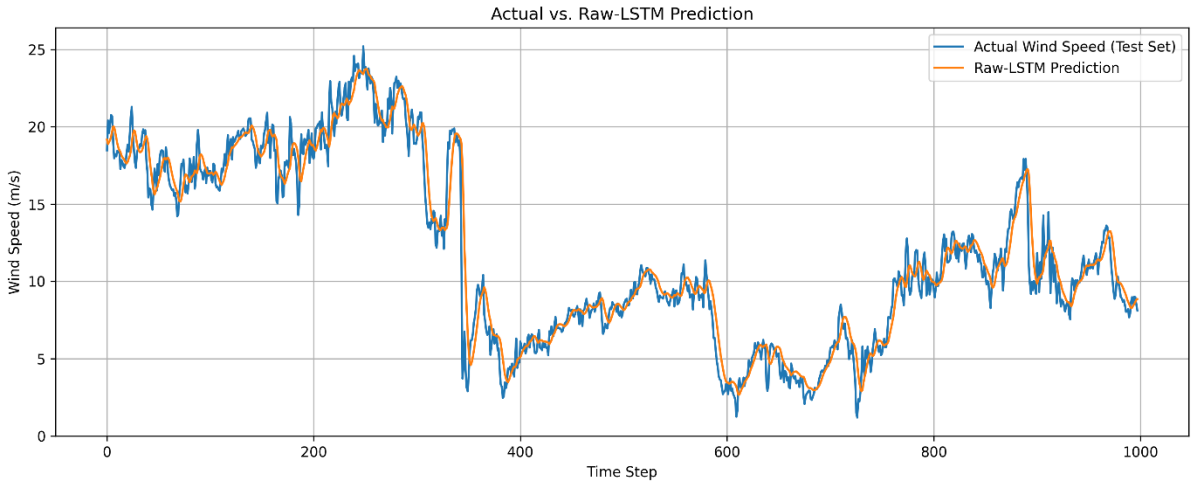


Figure 5. Comparison of actual and predicted wind speed values for the LSTM model.

An examination of the model performance results reveals that the LSTM model demonstrates strong performance in terms of error metrics. The obtained RMSE, MAE, and R^2 values of 0,9301,

0.5694, and 0.9440, respectively, indicate that the model successfully captures both the overall trend and the short-term variations of the wind speed series.

However, since no preprocessing was applied to the raw data, the model exhibited limited deviations during high-frequency and abrupt fluctuations. Therefore, the LSTM model was used as a baseline reference for comparison with all other methods evaluated in this study.

4.2. LMD-LSTM

LMD is an adaptive signal decomposition technique that separates complex time series into a set of PF components and a residual signal. This method captures instantaneous amplitude and frequency variations within the signal, enabling the extraction of structural components at different temporal scales. The number and structure of the PF components are determined iteratively based on the local characteristics of the signal, without requiring a predefined fixed number of modes specified by the user.

As shown in Figure 6, the extracted PF components exhibit oscillatory behaviors at different temporal scales. The first PF components are dominated by rapid, irregular fluctuations, capturing high-frequency variations and short-term disturbances in the wind speed signal. As the decomposition proceeds, the PF components tend to represent relatively smoother and more structured oscillations. The residual component reflects the slowly varying behavior of the series and corresponds to the long-term trend structure. It should be noted that, unlike variational decomposition methods, the PF indices in LMD do not impose a strict frequency ordering, as the decomposition is fully data-driven and adaptive.

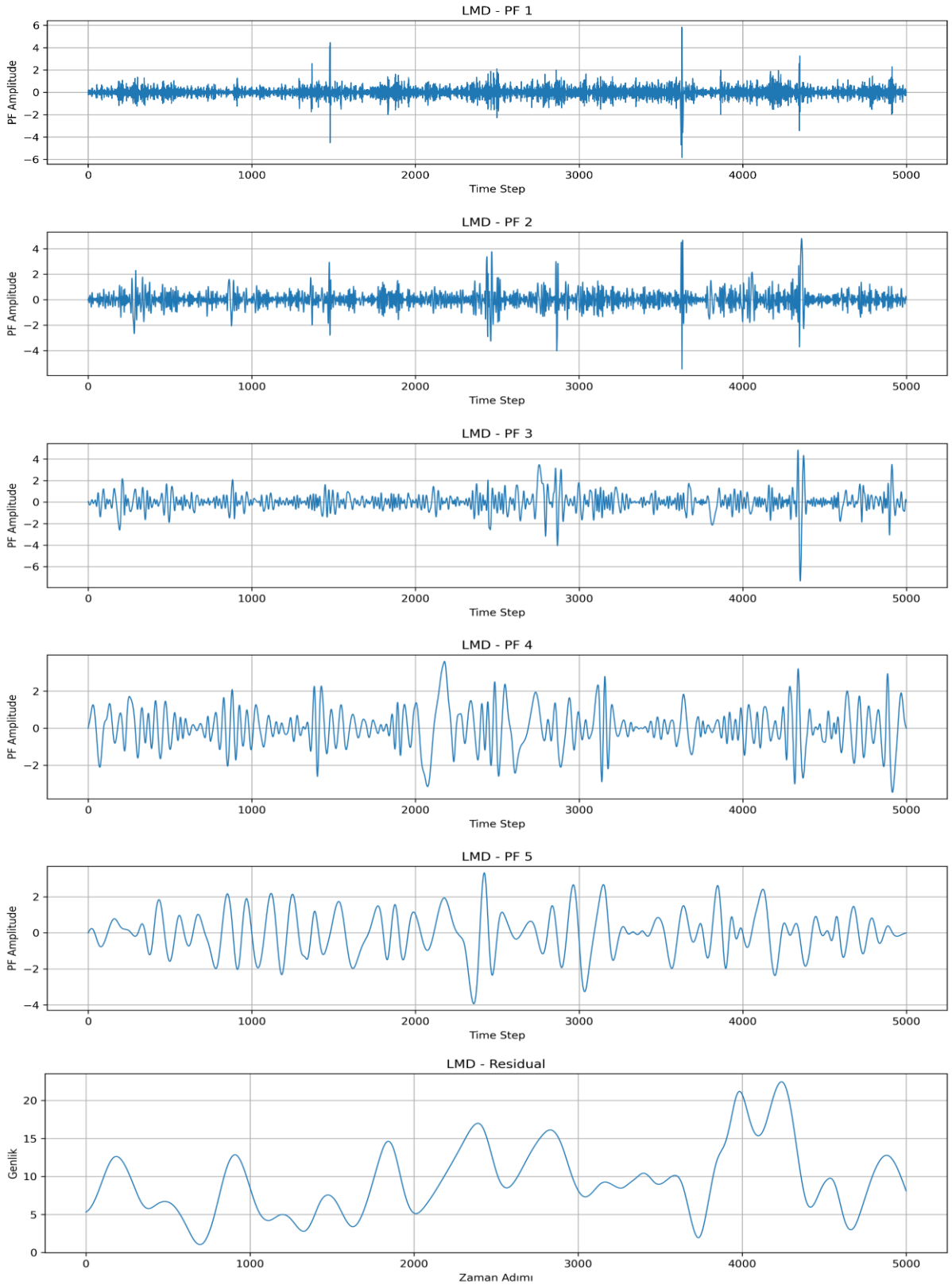


Figure 6. PF components and residual signal obtained using the LMD method.

To evaluate the accuracy of the LMD decomposition process, all PF components were summed together with the residual term to reconstruct the original wind speed series.

As shown in Figure 7, the reconstructed signal perfectly overlaps with the original series.

In addition, the calculated reconstruction error was found to be extremely low, with RMSE = 1.24×10^{-15}. This result confirms that the LMD decomposition process does not cause information loss and that the components generated by the method accurately represent the underlying structure of the signal. However, it should be noted that this reconstruction error reflects only the numerical accuracy of the decomposition process and should not be interpreted as a measure of forecasting performance.

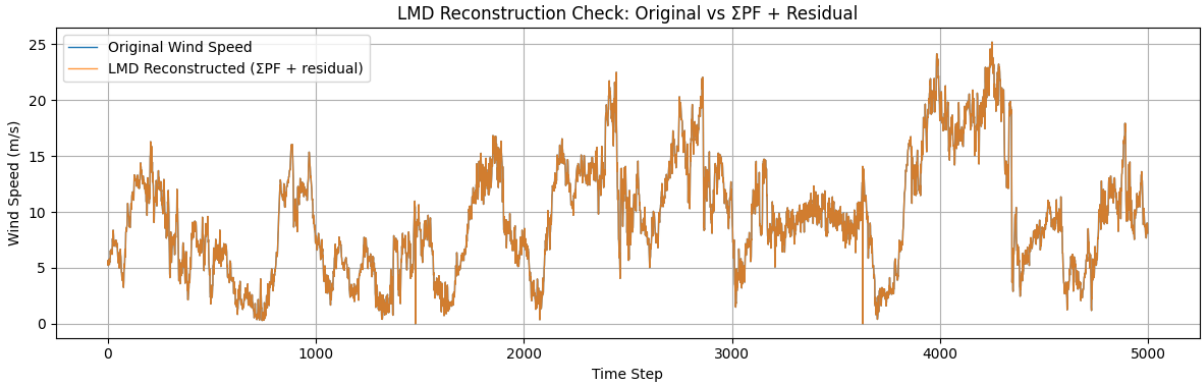


Figure 7. LMD Reconstruction Validation

Each PF component obtained from the LMD decomposition was forecasted using a separate LSTM model. The final forecast series was then obtained by summing all the predicted PF components. This structure constitutes the core architecture of the LMD–LSTM hybrid model.

As shown in Figure 8, the LMD–LSTM model follows the trend of the actual data more closely compared to the raw model and effectively reduces high-frequency noise.

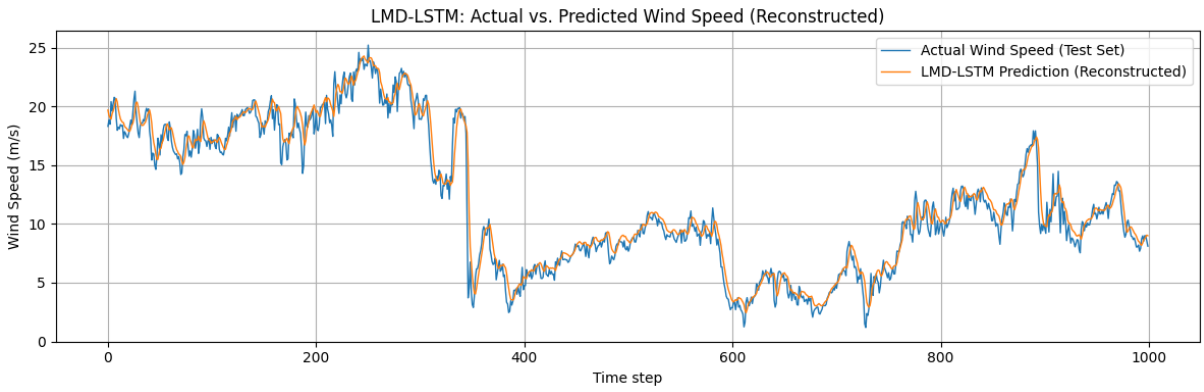


Figure 8. Comparison of actual and predicted wind speed values for the LMD–LSTM model.

The model achieved RMSE, MAE, and R^2 values of 0.4525, 0.3390 and 0.9615, respectively. These results demonstrate that the LMD-based approach improves the overall forecasting performance of the LSTM model.

Compared to the LSTM model trained on raw data, the proposed model produced lower error values, highlighting the positive impact of the data preprocessing step on forecasting accuracy.

4.2. VMD-LSTM

VMD method decomposes the wind speed time series into a set of mode components, each representing a distinct frequency band. In this study, it should be noted that the selection of the mode number K was not performed using a grid-search or optimization-based procedure. Instead, K was determined based on the convergence behavior and frequency separation characteristics of the normalized center frequencies produced by the VMD algorithm, as reported in Table 3. By examining the distribution and separation of the $|\omega_k|$ values across different K settings, the value K = 5 was identified as providing a clear and physically meaningful decomposition without introducing redundant components or mode splitting.

Although the same LSTM architecture was adopted for all modes, each mode was associated with an independently trained LSTM instance. This design ensures fair comparison across modes while allowing each frequency component to be learned individually.

Within the VMD algorithm, the bandwidth control parameter α (alpha) was fixed at 2000 throughout the experimental analysis. This value has been commonly adopted in previous studies addressing highly volatile time series, such as wind speed, and has been reported to offer an effective balance between mode smoothness and spectral separation [36].In this study, the α parameter was kept constant in order to ensure a controlled comparison between preprocessing methods. A comprehensive sensitivity analysis of the VMD parameters is considered an important direction for future research. As reflected in the normalized center frequency values presented in Table 3, the selected α value enabled clear separation between modes and mitigated excessive spectral overlap.

Table 3. Final normalized center frequencies ($|\omega_k|$) of VMD modes for different values of K.

Mode	K=3	K=4	K=5	K=6
1	1.81×10^{-4}	7.79×10^{-5}	5.50×10^{-5}	5.08×10^{-5}
2	2.50×10^{-2}	5.30×10^{-3}	3.92×10^{-3}	3.70×10^{-3}
3	3.46×10^{-1}	3.06×10^{-2}	2.58×10^{-2}	2.37×10^{-2}
4	—	3.46×10^{-1}	6.30×10^{-2}	5.13×10^{-2}
5	—	—	3.47×10^{-1}	1.02×10^{-1}
6	—	—	—	3.48×10^{-1}

A closer examination of Table 3 reveals that, for K = 5, five distinctly separated frequency bands were obtained. In particular, an intermediate frequency component emerging at approximately 0.063 was observed for K = 5, whereas this component was absent in the K = 4 case. This finding indicates the presence of an additional oscillatory mode that carries relevant information about the wind speed dynamics. When K was increased to 6, this intermediate band was observed to split into two sub-bands at approximately 0.051 and 0.102. This behavior corresponds to mode splitting, a form of over-decomposition that has been reported in the literature to reduce physical interpretability. Consequently, the choice of K = 5 provides an appropriate balance by capturing all meaningful frequency components of the signal while avoiding unnecessary model complexity.

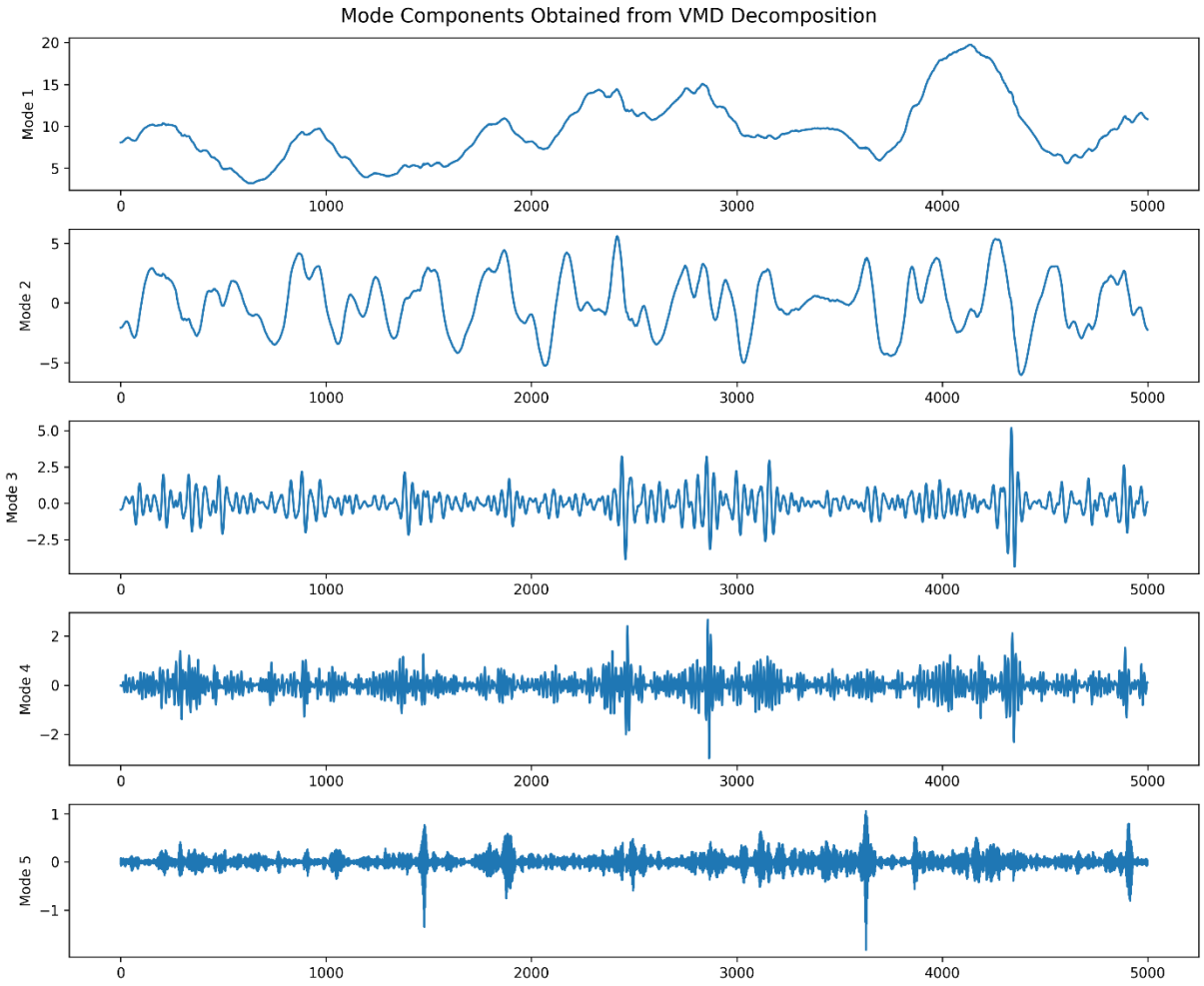


Figure 9. Mode components obtained from the VMD decomposition.

Figure 9 illustrates the mode components obtained from the VMD decomposition. As shown in Figure 9, the first mode component represents the low-frequency trend of the series, while the subsequent modes capture progressively higher-frequency fluctuations. In particular, the final mode components reveal regions of the signal where abrupt variations and noise effects are more pronounced.

Each mode component obtained from the decomposition was forecasted using a separate LSTM model. The overall wind speed prediction was produced by aggregating the forecasts generated for each individual mode, which constitutes the fundamental structure of the VMD–LSTM hybrid framework.

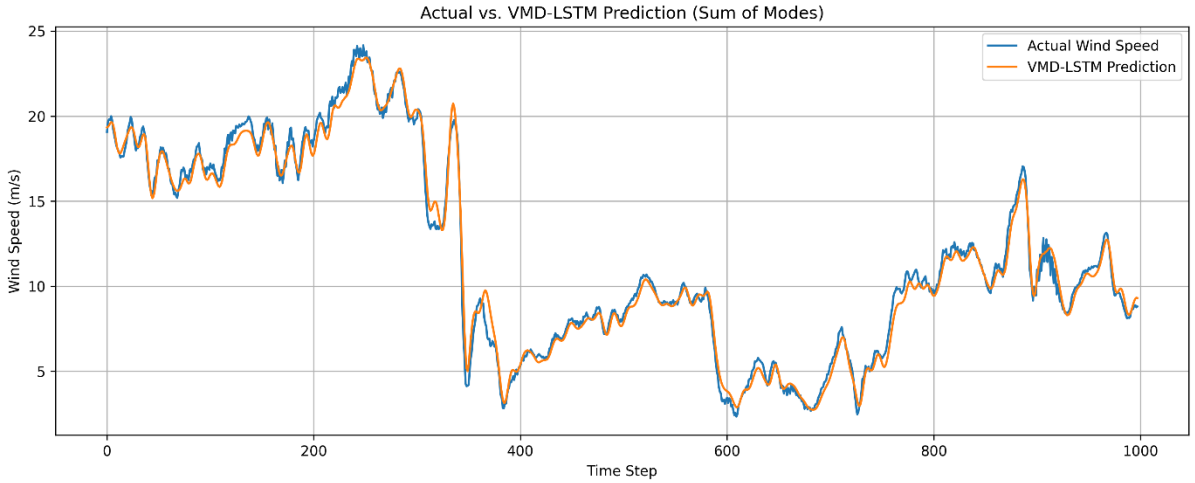


Figure 10. Comparison of actual and predicted wind speed values for the VMD–LSTM model.

As shown in Figure 10, the VMD–LSTM model exhibits a high degree of overlap with the actual wind speed values. Notably, a significant improvement in forecasting accuracy is observed in regions with abrupt wind speed fluctuations.

The model achieved RMSE, MAE, and R^2 values of 0.3825, 0.2798, and 0.9723, respectively. These results indicate that the proposed VMD–LSTM hybrid model demonstrates the highest forecasting performance compared to the other models. The model also shows superior generalization capability, effectively capturing short-term variations while maintaining the overall trend of the wind speed series.

4.5. Comparative Performance and Statistical Analysis

The error metrics of the three models are summarized in Table 4.

Table 4. Error metrics of the three models.

Model	RMSE	MAE	R^2
LSTM	0.9301	0.5694	0.9440
LMD-LSTM	0.4525	0.3390	0.9615
VMD-LSTM	**0.3825**	**0.2798**	**0.9723**

The DM test was applied to the one-step-ahead forecast errors obtained from the test dataset. The error series used in the test correspond to the residuals calculated as the difference between the observed wind speed values and the model forecasts.

The DM test results are presented in Table 5. The findings indicate that all the proposed hybrid models achieved a statistically significant improvement in forecasting performance, with all p-values < 0.001.

First, both the VMD–LSTM and LMD–LSTM models exhibited significantly better performance than the benchmark LSTM model in terms of both the Squared Error (SE) and Absolute Error (AE) metrics. This finding demonstrates that both signal decomposition methods substantially enhance the forecasting capability of the LSTM model when applied individually.

Even more noteworthy are the results obtained from the direct comparison between the two hybrid models. The VMD–LSTM model demonstrated statistically significant improvements over the LMD–LSTM model, yielding substantially lower forecasting errors. According to the Diebold–Mariano test, the test statistic based on squared errors was –11.80 with a significance level below 0.001, while the statistic based on absolute errors was –20.96 with a significance level below 0.001. These results indicate that, for wind speed forecasting applications, the VMD method constitutes a more effective preprocessing strategy than LMD, leading to superior predictive performance within the proposed hybrid framework.

Table 5. Results of the DM test.

Model Comparison	**DM(SE)**	**p(SE)**	**DM(AE)**	**p(AE)**
VMD-LSTM vs LSTM	–8.30	<0.001	–12.50	<0.001
LMD-LSTM vs LSTM	–19.51	<0.001	–40.39	<0.001
VMD-LSTM vs LMD-LSTM	–11.80	<0.001	–20.96	<0.001

5. CONCLUSION

In this study, the impact of data preprocessing methods on model performance and forecasting accuracy in wind speed prediction problems was comprehensively analyzed. For this purpose, three different scenarios were compared using the same dataset and a fixed LSTM architecture:
(i) LSTM model trained on unprocessed data, (ii) the LMD-LSTM model based on Local Mean Decomposition, and (iii) the VMD-LSTM model based on Variational Mode Decomposition. The results clearly demonstrate that data preprocessing significantly improves model performance.

The VMD-LSTM model achieved the highest accuracy. The VMD-LSTM model produced the lowest error values (RMSE 0.3825, MAE 0.2798) and achieved the highest coefficient of determination (R^2 0.9723), outperforming all other methods. This finding suggests that, due to its variational optimization framework, VMD effectively suppresses high-level stochastic noise in the wind speed signal and provides more stable and deterministic subcomponents for LSTM learning. Therefore, for highly volatile signals, VMD should be preferred over LMD as a preprocessing strategy.

The LMD-LSTM model achieved better results than the LSTM model trained on unprocessed data. This finding confirms that signal decomposition, regardless of the specific method used, facilitates the learning of complex wind speed data. However, due to the iterative structure of LMD, limitations such as end effects and mode mixing prevented it from reaching the same level of precision as the VMD-based approach.

The differences between the models were statistically significant. The results of the Diebold–Mariano (DM) test confirmed that the performance differences among the models were significant under both the Squared Error (SE) and Absolute Error (AE) criteria. In particular, the superiority of the VMD–LSTM model was statistically significant at the 1% level ($p < 0.01$).

These findings indicate that reducing signal roughness and decomposing stochastic components into more stable sub-modes significantly enhance the learning capability of deep learning models such as LSTM. The superior performance of the VMD–LSTM model can be attributed to the ability of the VMD algorithm to separate the signal into spectrally well-defined components. By isolating different frequency bands prior to the learning process, the LSTM network can capture more stable temporal patterns. This decomposition reduces the impact of noise and improves the interpretability of the forecasting model. Overall, the findings of this study demonstrate that the data preprocessing step is not merely a supportive operation but a critical component that determines the success of forecasting models. Based on the obtained results, several directions for future research can be suggested. First, hybrid models combining robust decomposition techniques such as VMD with advanced deep learning architectures, including BiLSTM, Transformer, and CNN-Attention networks, can be explored. Second, optimization of VMD hyperparameters (K, α) using automated search algorithms such as Optuna or Grid Search may further improve model

performance. Finally, incorporating spatial data from multiple meteorological stations may help assess and enhance the generalization capability of the forecasting framework.

NOMENCLATURE

VMD Variational Mode Decomposition
LMD Local Mean Decomposition
LSTM Long Short-Term Memory
RMSE Root Mean Square Error
MAE Mean Absolute Error
DM Diebold–Mariano test
u_k (t) k-th decomposed mode
ω_k Center frequency of k-th mode
α Penalty parameter in VMD
λ Lagrange multiplier
K Number of decomposition modes
PFi(t) i-th product function in LMD

DECLARATION OF ETHICAL STANDARDS

The authors of the paper submitted declare that nothing which is necessary for achieving the paper requires ethical committee and/or legal-special permissions.

CONTRIBUTION OF THE AUTHORS

Tuğçe Inag: Methodology, Data Curation, Formal Analysis, Writing – Original Draft, Review & Editing.
Yasin Inag: Conceptualization, Supervision, Review & Editing

CONFLICT OF INTEREST

There is no conflict of interest in this study.

REFERENCES

[1] Sambane M, Mendu B, Monchusi BB. Advanced neural network and hybrid models for wind power forecasting: a comprehensive global review. Future Energy 2024; 3(4): 67-79.

[2] Amirteimoury F, Keynia F, Amirteimoury E, Memarzadeh G, Shabanian H. A novel wind speed prediction model based on neural networks, wavelet transformation, mutual information, and coot optimization algorithm. Scientific Reports 2025; 15(1): 10860.
[3] Alves D, Mendonça F, Mostafa SS, Morgado-Dias F. The potential of machine learning for wind speed and direction short-term forecasting: a systematic review. Computers 2023; 12: 206.
[4] Han Y, Zhang C, Li K. A new two-stage decomposition and integrated hybrid model for short-term wind speed prediction. Wind Engineering 2024; 48(5): 835-860.
[5] Şenkal S, Emeksiz C. The effect of data decomposition on prediction performance in wind speed prediction with artificial neural network. International Scientific and Vocational Studies Journal 2023; 7(2): 213-223.
[6] Alsamamra HR, Salah S, Shoqeir JH. Performance analysis of ARIMA model for wind speed forecasting in Jerusalem, Palestine. Energy Explor Exploit 2024; 42(5): 1727-1746.
[7] Liu H, Tian HQ, Li YF. An EMD-recursive ARIMA method to predict wind speed for railway strong wind warning system. Journal of Wind Engineering and Industrial Aerodynamics 2015; 141: 27-38.
[8] Wang Y, Zou R, Liu F, Zhang L, Liu Q. A review of wind speed and wind power forecasting with deep neural networks. Applied Energy 2021; 304: 117766.
[9] Wu Z, Luo G, Yang Z, Guo Y, Li K, Xue Y. A comprehensive review on deep learning approaches in wind forecasting applications. CAAI Transactions on Intelligence Technology 2022; 7(2): 129-143.
[10] Lim JY, Kim S, Kim HK, Kim YK. Long short-term memory (LSTM)-based wind speed prediction during a typhoon for bridge traffic control. Journal of Wind Engineering and Industrial Aerodynamics 2022; 220: 104788.
[11] Muslim R, Delfianti R, Harsito C, Palaloi S, Aryono NA, Indrarahmana AY, Joshua SR, Podder R. Optimizing the integration of wind and solar power for hybrid electrical energy in high-rise building energy systems. Engineered Science 2025; 35: 1-10.
[12] Gomes V, Carvalho D, Gouveia S. On the correction of GFS wind speed forecasts in Portugal using LSTM networks. Lecture Notes in Computer Science 2026; 15000: 1-12.
[13] Lu H, Mokarram M. Evaluation of environmental and climatic impacts of sand dune movement using geographic object-based image analysis and machine learning. Journal of Arid Environments 2026; 232: 105400.

[14] Tian G, Le Coz C, Charantonis AA, Tantet A, Goutham N, Plougonven R. Improving subseasonal wind speed forecasts in Europe with a nonlinear model. Monthly Weather Review 2025; 153(9): 1761-1779.

[15] Shahriar SA, Choi Y, Islam R. Advanced deep learning approaches for forecasting high-resolution fire weather index (FWI) over CONUS: Integration of GNN-LSTM, GNN-TCNN, and GNN-DeepAR. Remote Sensing 2025; 17(3): 1-20.

[16] Van der Bank D, Dalton A, Bekker B. Implementation and evaluation of a limited-area artificial intelligence wind speed forecasting model. International Conference on Clean Electrical Power (ICCEP), Naples, Italy, 2025.

[17] Liu H, Yang R, Wang T, Zhang L. A hybrid neural network model for short-term wind speed forecasting based on decomposition, multi-learner ensemble, and adaptive multiple error corrections. Renewable Energy 2021; 165: 573-594.

[18] Ayyavu S, Sayeed MS, Abdul Razak SF. A multi-step day-ahead wind power forecasting based on VMD-LSTM-EFG-ABC technique. Energy Informatics 2025; 8(1):111.

[19] Du M, Zhang Z, Ji C. Prediction for coastal wind speed based on improved variational mode decomposition and recurrent neural network. Energies 2025; 18(3):542.

[20] Li S, Guo L, Zhu J, Chen J, Liu M, Cui X, Li L. Medium-term offshore wind speed multi-step forecasting based on VMD and GRU-MATNet model. Ocean Engineering 2025; 325: 120737.

[21] Patil R, Suttraway PS, Shikhare T, Hire V. Wind power forecasts through hybrid model based on VMD and wavelet decomposition. International Conference on Energy, Power and Environment (ICEPE), NIT Meghalaya, India, 2025.

[22] Zhou J, Lu L, Wang Y. Port wind speed prediction model based on VMD and GRU-MLP. International Symposium on Computer Applications and Information Technology (ISCAIT), Nanjing, China, 2025.

[23] Qi B, Sun Y, Wang H, Song C, Wang X, Jin J. Wind speed prediction method based on EMD-LSTM with integrated spatial features. International Conference of Information and Communication Technology (ICTech), Nanjing, China, 2024.

[24] Sun Q, Che J, Hu K, Qin W. Deterministic and probabilistic wind speed forecasting using decomposition methods: accuracy and uncertainty. Renewable Energy 2025; 243: 122515.

[25] Wang Z, Wang L, Revanesh M, Huang C, Luo X. Short-term wind speed and power forecasting for smart city power grid with a hybrid machine learning framework. IEEE Internet of Things Journal 2023; 10(21): 18754-18765.

[26] Feng L, Wang Y, Yan Y, Wang X, Liu N, Ding W. Wind speed prediction model based on deep learning. E3S Web of Conferences, Prague, Czech Republic, 2023.
[27] Taha A, Nazih N, Makeen P. Wind speed prediction based on variational mode decomposition and advanced machine learning models in Zaafarana, Egypt. Scientific Reports 2025; 15(1): 15599.
[28] Shu C, Qin B, Wang X. Wind speed prediction based on improved VMD-BP-CNN-LSTM model. Journal of Power and Energy Engineering 2024; 12(1): 29-43.
[29] Li Y, Sun K, Yao Q, Wang L. A dual-optimization wind speed forecasting model based on deep learning and improved dung beetle optimization algorithm. Energy 2024; 286: 129604.
[30] Hu W, Yang Q, Zhang P, Yuan Z, Chen HP, Shen H, Zhou T, Guo K, Li T. A novel two-stage data-driven model for ultra-short-term wind speed prediction. Energy Reports 2022; 8: 9467-9480.
[31] Xia W, Che J, Hu K, Xu Y. A synchronized multi-step wind speed prediction with adaptive features and parameters selection: Insights from an interaction model. Expert Systems with Applications 2024; 255: 124764.
[32] Chen H, Lu T, Huang J, He X, Sun X. An improved VMD–EEMD–LSTM time series hybrid prediction model for sea surface height derived from satellite altimetry data. Journal of Marine Science and Engineering 2023; 11(12): 2386.
[33] Liang Y, Zhang D, Zhang J, Hu G. A state-of-the-art analysis on decomposition method for short-term wind speed forecasting using LSTM and a novel hybrid deep learning model. Energy 2024; 313: 133826.
[34] Chen Y, Yu S, Islam S, Lim CP, Muyeen SM. Decomposition-based wind power forecasting models and their boundary issue: An in-depth review and comprehensive discussion on potential solutions. Energy Reports 2022; 8: 8805-8820.
[35] Wind Turbine SCADA Dataset. UCI Machine Learning Repository, Irvine, CA, US, 2018.
[36] Dragomiretskiy K, Zosso D. Variational mode decomposition. IEEE Transactions on Signal Processing 2014; 62(3): 531-544.
[37] Chen G, Guan Z, Cheng Q. Improved LMD and its application in short-term wind power forecast. Communications in Computer and Information Science 2014; 463: 66-72.
[38] Singla P, Duhan M, Saroha S. Solar irradiation forecasting by long-short term memory using different training algorithms. Renewable Energy Optimization, Planning and Control: Proceedings of ICRTE; Springer; 2021: 81-89.

[39] Sharadga H, Hajimirza S, Balog RS. Time series forecasting of solar power generation for large-scale photovoltaic plants. Renewable Energy 2020; 150: 797-807.