Hybrid Time-Series Forecasting with VMD and LogTrans for Wind Energy Applications

Bala Saibabu Bommidi
Department of Electrical and Electronics Engineering, Prasad V Potluri Siddhartha Institute of Technology, Vijayawada, Andhra Pradesh, India
saibabu.dec31@gmail.com

Hemanth Sai Madupu
Department of Electrical and Electronics Engineering, Prasad V Potluri Siddhartha Institute of Technology, Vijayawada, Andhra Pradesh, India
hemanthsai.madupu@pvpsiddhartha.ac.in (corresponding author)

Vijaya Anand Nidumolu
Department of Electrical and Electronics Engineering, Prasad V Potluri Siddhartha Institute of Technology, Vijayawada, Andhra Pradesh, India
nidumoluvijay@gmail.com

S. K. B. Pradeepkumar Chilakala
Department of Electrical and Electronics Engineering, Ramachandra College of Engineering, Eluru, Andhra Pradesh, India
pradeepkumar.chilakala@gmail.com

Received: 5 June 2025 | Revised: 22 June 2025 and 30 June 2025 | Accepted: 7 July 2025

ABSTRACT

Accurate Wind Speed Forecasting (WSF) is important for managing the wind energy, improving the turbine performance, and maintaining the power grid stability. This study proposes a new hybrid model that combines Variational Mode Decomposition (VMD) with Log-Sparse Transformer (LogTrans), a Transformer-based Deep Learning (DL) model that uses log-sparse attention to better capture the patterns in time-series data. VMD is used to break down the original Wind Speed (WS) signal into several simpler components, making it easier for the forecasting model to learn meaningful patterns. These components are then processed by the LogTrans model to make accurate predictions. The model is tested using WS data collected from Garden City, and its performance is evaluated at 5-min, 10-min, and 30-min forecasting intervals. The results show that the proposed VMD-LogTrans model gives more accurate and reliable predictions than benchmark DL models and other hybrid approaches. The proposed model offers an efficient solution for short-term WSF and can help improve the integration of wind energy into the power systems.

Keywords-time-series forecasting; decomposition; wind energy; transformer; deep learning

I. INTRODUCTION

A. *Background*

Accurate WSF is crucial for the efficient integration of the wind energy into modern power systems. As a sustainable energy source, the wind power poses challenges due to its variability and intermittency, affecting the grid reliability and turbine performance. Precise forecasting enables better turbine control, power dispatch scheduling, and grid stability.

However, WS is a highly complex, nonstationary, and stochastic time-series signal influenced by a variety of atmospheric and geographical factors [1]. This complexity necessitates the use of advanced modeling techniques capable of capturing the intricate temporal dependencies, fluctuating patterns, and non-linear behaviors that occur across multiple time scales. Moreover, effective WSF models must be adaptable to exogenous inputs, such as temperature, pressure, and humidity, which significantly impact the wind dynamics [2].

Advances in Machine Learning (ML) and DL have proven effective in electrical engineering tasks, such as load forecasting, microgrid control, and machinery diagnostics.

Studies have demonstrated over 97% accuracy in electricity load prediction [3], improved PID control using LSTM in microgrids [4], and reliable fault detection in rotating machinery using SVM and ensemble methods [5]. The findings support the application of ML/DL in WSF.

B. Related Work

To address the aforementioned challenges, research has turned to hybrid forecasting frameworks that combine powerful DL architectures with robust signal decomposition methods. DL models, such as Convolutional Neural Networks (CNNs) [6], Recurrent Neural Networks (RNNs) [7], Long Short-Term Memory (LSTM) neural networks [8], Bidirectional LSTMs (BiLSTMs) [9], and Transformer-based models [10] have been proved promising due to their capacity to learn complex features and long-range dependencies in time-series data. CNNs excel at extracting local features and spatial patterns from input sequences, while RNNs and LSTM variants are well-suited for modeling the sequential dependencies and temporal correlations. Authors in [11] reviewed the use of ML and DL methods for Wind Energy Forecasting (WEF), focusing mainly on standalone neural models with limited discussion on the signal decomposition or hybrid approaches. In [12], a dedicated survey on RNNs and their variants, including LSTM, GRU, and Bidirectional RNNs was performed, in the context of WEF.

To improve the robustness and forecasting accuracy, DL models are often combined with signal decomposition techniques, such as Empirical Mode Decomposition (EMD) [13], VMD [14], and wavelet transforms [15]. These methods break down the WS signals into components with distinct frequency and temporal traits, enhancing the input representation. When integrated with DL models, such hybrid approaches, they yield more accurate and reliable forecasts. In [16], the effectiveness of combining EMD with AI algorithms for WS prediction is highlighted.

Authors in [17, 18] combined VMD with neural networks, like CNNs, TCNs, and Transformers to enhance WSF. These hybrid models decompose the signals into modes for improved temporal feature extraction, particularly in multi-step forecasting. The proper tuning of the VMD parameters, like the number of modes and penalty factors, is crucial for effective decomposition and accuracy.

Research has explored advanced EMD variants, like EEMD and CEEMDAN, to overcome limitations, such as the mode mixing and end effects. These methods are often combined with DL for a more robust forecasting. For example, authors in [19] deployed a CEEMD-BiLSTM-Transformer hybrid to capture both the short- and long-term dependencies, while authors in [20] proposed a CEEMDAN-CNN-Transformer model, where CNNs extract features from decomposed modes before the Transformer-based temporal modeling.

Similarly, authors in [21] proposed a VMD-Transformer model for multi-horizon WSF, leveraging the Transformer's contextual processing of decomposed features. Authors in [22] introduced a framework combining Improved EMD (IEMD) with an enhanced Transformer, reflecting the ongoing advances in decomposition and neural architectures. In [23], a hybrid CNN-GRU model outperformed traditional and DL methods for short-term electricity consumption forecasting, improving the accuracy and smart grid resource management.

The LogTrans [24] model enhances time series forecasting by addressing two key Transformer limitations: poor local pattern recognition and high memory usage. These are critical in WSF, where data are of high-frequency, noisy, and involve both short-term fluctuations and long-term cycles. While RNNs and LSTMs struggle with long-term dependencies and standard Transformers face quadratic memory issues, LogTrans introduces two innovations. Initially, the convolutional self-attention uses causal convolutions to improve the local trend detection. Second, the LogSparse attention reduces the memory complexity from $O(L^2)$ to $O(L(\log L)^2)$, enabling efficient long-sequence modeling while preserving long-range dependencies.

This study proposes a novel hybrid WSF model that integrates VMD with LogTrans, a Transformer variant employing log-sparse attention to efficiently capture the long-range temporal dependencies. In this study, a total of 8,640 WS samples were used for each time interval. The data were collected from the National Renewable Energy Laboratory (NREL) website for the Garden City site [25]. VMD decomposes the WS signal into IMFs, isolating the frequency components and mitigating the signal nonstationarity. These components are then modeled using LogTrans for accurate temporal feature extraction with a reduced computational cost. The model is evaluated over 5-min, 10-min, and 30-min forecasting horizons, consistently outperforming the conventional DL and hybrid models, and offering a robust, scalable solution for improved WSF.

II. PROPOSED HYBRID MODEL FOR WSF

A. Denoising Wind Speed Data Using VMD

In WSF, the raw data often contain non-stationary, noisy, and multi-scale components that hinder the model accuracy. To mitigate this issue, VMD is used as a preprocessing step to denoise and extract intrinsic signal components. VMD adaptively decomposes the input signal into a predefined number of sub-signals, each with limited bandwidth and centered on specific frequencies.

The decomposition task in VMD is framed as a constrained variational optimization problem, aiming to find a set of mode functions $\{u_j\ (t)\}$ and their corresponding center frequencies $\{w_j\}$ that collectively reconstruct the original signal. The optimization problem is expressed as:

$$\min_{\{u_j\}\{w_j\}} \left\{ \sum_{j=1}^{J} \left\| \partial_t \left[\left(\delta(t) + \frac{i}{\pi t} \right) * u_j(t) \right] e^{-j w_j t} \right\|_2^2 \right\} \qquad (1)$$

subject to $\sum_{j=1}^{J} u_j(t) = x(t)$

where $\delta(t)$ is the Dirac delta function, $*$ denotes convolution, and ∂t represents the time derivative. To solve this, an augmented Lagrangian is introduced:

$$L(\{u_j\},\{w_j\},\lambda) = \alpha \sum_{j=1}^{J} \left\| \partial_t \left[\left(\delta(t) + \frac{i}{\pi t} \right) * u_j(t) \right] e^{-jw_jt} \right\|_2^2 + \left\| x(t) - \sum_{j=1}^{J} u_j(t) \right\|_2^2 + \langle \lambda(t), x(t) - \sum_{j=1}^{J} u_j(t) \rangle \quad (2)$$

where α is a quadratic penalty factor, and $\lambda(t)$ is the Lagrangian multiplier. The solution involves iteratively updating each mode function u_j and its center frequency w_j in the frequency domain using the method of multipliers:

$$\hat{u}_j^{n+1}(w) = \frac{\hat{x}(w) - \sum_{k\neq j} \hat{u}_k(w) + \hat{\lambda}(w)/2}{1+2\alpha(w-w_j)^2} \quad (3)$$

$$w_j^{n+1} = \frac{\int_0^\infty w \left|\hat{u}_j(w)\right|^2 dw}{\int_0^\infty \left|\hat{u}_j(w)\right|^2 dw} \quad (4)$$

These iterations continue until convergence is achieved by decomposing the input WS signal into multiple IMFs, with VMD effectively separating low-frequency trends and high-frequency noise. The low-frequency components (e.g., IMF1) typically capture the dominant signal behavior, while the high-frequency components (e.g., IMF4 and above) often represent noise or transient fluctuations. Selecting and reconstructing the relevant IMFs enables the denoising of theWS data.

B. WSF using LogTrans Model

In this study, the LogTrans architecture is used for WSF to effectively capture both the short-term fluctuations and long-term seasonal dependencies inherent in wind patterns. Given a collection of univariate WS time series $\{y_{i,1:t_0}\}_{j=1}^{n}$, where $y_{j,t} \in R$ represents the WS at time t for site j, and corresponding time-based covariates $\{y_{j,1:t_0+\tau}\}_{j=1}^{n}$, the objective is to forecast the WS over the next τ time steps.

$$p\left(y_{j,t_0+1:t_0+\tau} \middle| y_{j,1:t_0}, x_{j,1:t_0+\tau}; \Phi\right) = \prod_{t=t_0+1}^{t_0+1} p\left(y_{j,t} \middle| y_{j,1:t-1}, x_{j,1:t}; \Phi\right) \quad (5)$$

where Φ denotes the model parameters, which defines the input at each time step as a concatenation of the previous WS and covariates:

$$u_t = [y_{t-1} \odot u_t] \in R^{d+1} \quad (6)$$

and the sequence input to the model is:

$$U_t = [u_1, u_2, \dots\dots\dots u_t]^T \in R^{t\times(d+1)}$$

The LogTrans model enhances the canonical Transformer by introducing two key innovations: convolutional self-attention and LogSparse attention. The self-attention mechanism, rather than computing queries Q and keys K using simple linear projections, is generated using causal convolution with kernel size k, and is denoted as:

$$Q = Conv_k(Y)W_Q \quad (7)$$

$$K = Conv_k(Y)W_K \quad (8)$$

$$V = YW_V \quad (9)$$

where W_Q, W_K, and W_V are learnable weights. The use of causal convolution ensures that the attention mechanism captures the local context, improving the sensitivity to sudden changes or anomalies in WS patterns. To overcome the quadratic memory bottleneck in standard Transformers, LogTrans replaces full attention with a logarithmically sparse attention pattern, where each position l attends only to a subset of previous positions defined as:

$$J_l = \{l - 2^{\lceil log_2 l \rceil}, l - 2^{\lceil log_2 l \rceil - 1}, \dots\dots, l-1, l\} \quad (10)$$

This leads to a total attention cost of $O(L(\log L)^2)$ for a sequence of length L, allowing the efficient modeling of long sequences, which iscritical in high-resolution wind datasets.

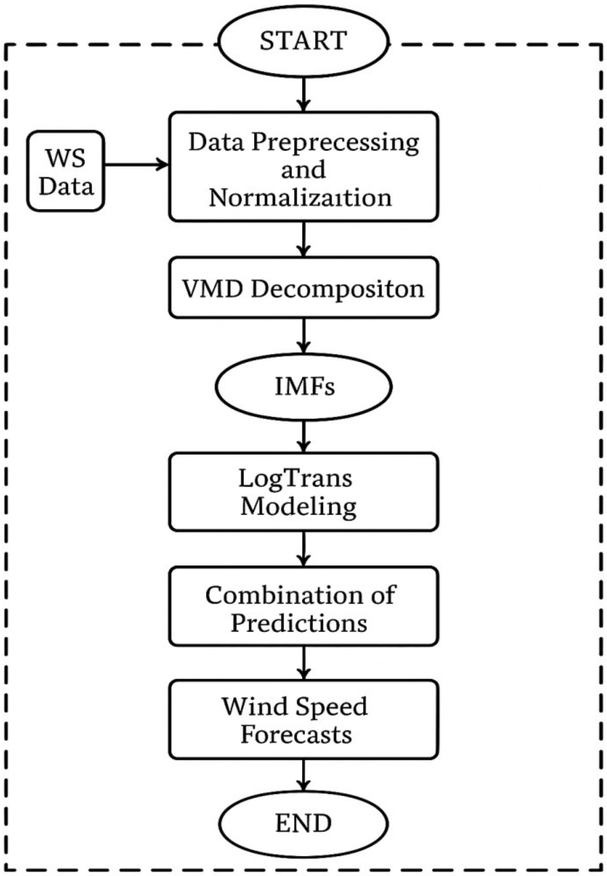


Fig. 1. The flowchart of the proposed hybrid model.

The final output layer of the model predicts the parameters (e.g., mean μ_t, standard deviation σ_t) of a Gaussian distribution. At this point forecasts or probabilistic intervals can be derived:

$$y_t \sim \mathbb{N}(\mu_t, {\sigma_t}^2) \quad (11)$$

The model is trained by maximizing the log-likelihood over all prediction steps using:

$$\mathcal{L} = \sum_{j=1}^{n} \sum_{t=t_0+1}^{t_0+\tau} \log p\left(y_{j,t} \middle| y_{j,1:t-1}, x_{j,1:t}; \Phi\right) \quad (12)$$

This methodology enables LogTrans to learn rich temporal dependencies and localized patterns in WS time series, offering a scalable and accurate solution for short-term and long-term WSF tasks.

Figure 1 illustrates the flow of the proposed hybrid model. WS data are first preprocessed and normalized, then decomposed into several IMFs using VMD. Each IMF, being more stationary, is forecasted individually using the LogTrans model, which leverages the log-sparse attention to capture the long-term dependencies. A trial-and-error approach is employed to determine the optimal hyperparameters of the proposed model, ensuring the best forecasting performance. The final WS forecast is obtained by summing the predicted IMFs. The LogTrans model uses 2 encoder layers, each with 3 attention heads and an embedding dimension of 64. The kernel size is set to 3. The model has been trained using the Adam optimizer with a learning rate of 0.001 and batch size of 32. This integration of VMD and LogTrans enables a more accurate and efficient forecasting process by combining signal simplification with advanced sequence modeling.

III. RESULTS AND DISCUSSION

This study compares several DL models for forecasting WS at 5-min, 10-min, and 30-min intervals. Tables I-III present the WSF results of the proposed, individual, and hybrid models at 5-min, 10-min, and 30-min intervals, respectively. A detailed comparison of the model performance across different forecasting horizons is provided, highlighting the effectiveness of the proposed approach in various time frames.

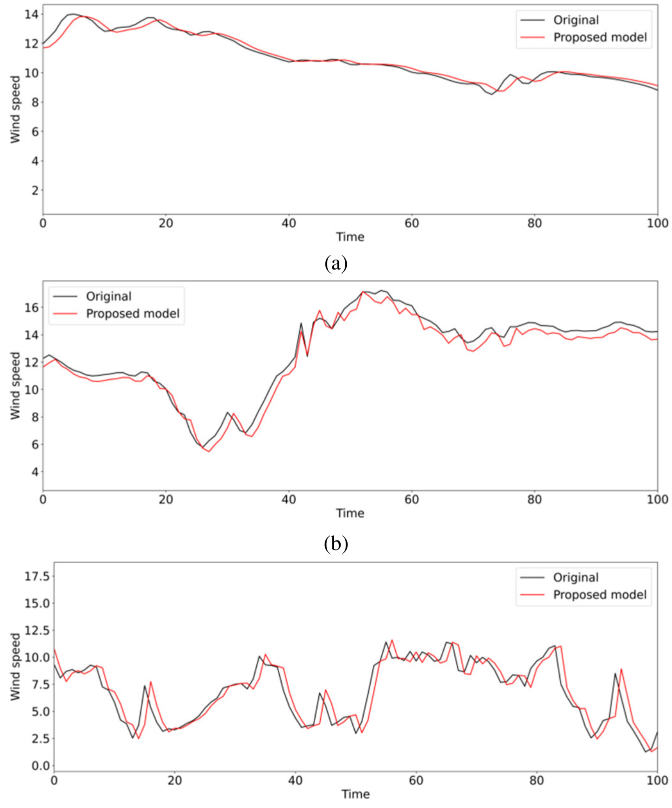
TABLE I. WSF RESULTS OF THE PROPOSED, INDIVIDUAL, AND HYBRID MODELS AT THE 5-MIN INTERVAL

Model	RMSE	MAE	MSE	MAPE (%)	R^2
ANN	0.4839	0.4051	0.2341	6.89	0.9675
CNN	0.4489	0.3578	0.2015	5.76	0.9731
LSTM	0.5044	0.3812	0.2544	6.46	0.9587
BiLSTM	0.5178	0.3979	0.2681	6.94	0.9609
TCN	0.4297	0.3515	0.1847	5.94	0.9720
Trans	0.3593	0.2833	0.1291	4.55	0.9678
LogTrans	0.3071	0.2322	0.0943	4.14	0.9720
VMD-ANN	0.4351	0.3636	0.1893	5.99	0.9744
VMD-CNN	0.2940	0.2422	0.0865	3.70	0.9783
VMD-LSTM	0.3003	0.2489	0.0902	3.92	0.9775
VMD-BiLSTM	0.2741	0.2114	0.0751	3.38	0.9772
VMD-TCN	0.2587	0.2028	0.0669	3.15	0.9729
VMD-Trans	0.2231	0.1715	0.0498	2.87	0.9719
Proposed	0.1922	0.1581	0.0369	2.37	0.9769

Among all individual models, LogTrans consistently delivered the most accurate results across all forecasting intervals, outperforming the next-best Transformer model. At the 5-min interval, LogTrans achieved lower errors (RMSE: 0.3071, MAPE: 4.14%) compared to the Transformer (RMSE: 0.3593, MAPE: 4.55%), with similar R^2 values. As the interval increased to 10 and 30 min, LogTrans maintained its advantage with lower errors and a stronger correlation with actual WS values. This indicates its effectiveness in capturing the short-term patterns and seasonal trends. Among the hybrid models, VMD-LogTrans emerged as the most effective across all horizons. At the 5-min interval, it recorded the lowest error values with an RMSE of 0.1922, MAE of 0.1581, MSE of 0.0369, MAPE of 2.37%, and a high R^2 of 0.9769, indicating excellent short-term prediction accuracy. As the forecast horizon extended to 10 min, the model maintained a robust performance, with RMSE increasing slightly to 0.2460, MAE to 0.1995, MSE to 0.0605, and MAPE to 2.46%, while still achieving a high R^2 of 0.9734.

TABLE II. WSF RESULTS OF THE PROPOSED, INDIVIDUAL, AND HYBRID MODELS AT THE 10-MIN INTERVAL

Model	RMSE	MAE	MSE	MAPE (%)	R^2
ANN	0.8049	0.6294	0.6479	7.58	0.9129
CNN	0.8805	0.6436	0.7753	8.18	0.9175
LSTM	0.6908	0.5268	0.4772	6.71	0.9219
BiLSTM	0.6266	0.4419	0.3927	5.70	0.9325
TCN	0.5726	0.4068	0.3279	5.13	0.9471
Trans	0.5192	0.3590	0.2696	4.47	0.9486
LogTrans	0.4878	0.3358	0.2379	4.33	0.9552
VMD-ANN	0.7001	0.6328	0.4901	8.52	0.9344
VMD-CNN	0.5960	0.4945	0.3552	6.31	0.9443
VMD-LSTM	0.5018	0.3747	0.2518	4.69	0.9570
VMD-BiLSTM	0.3931	0.3188	0.1545	4.14	0.9670
VMD-TCN	0.3675	0.2754	0.1350	3.59	0.9638
VMD-Trans	0.2931	0.2191	0.0859	2.79	0.9665
Proposed	0.2460	0.1995	0.0605	2.46	0.9734



(a)

(b)

(c)

Fig. 2. Forecasting results of proposed model: (a) 5 min WSF, (b) 10 min WSF, (c) 30-min WSF.

At the 30-min interval, where forecasting becomes more challenging due to increased uncertainty and variability in the wind patterns, VMD-LogTrans continues to lead, with an RMSE of 0.4255, MAE of 0.3582, MSE of 0.1810, and MAPE of 5.95%, along with a solid R^2 of 0.9520. These results are significantly better compared to traditional models, like CNN, which suffers a drastic drop in performance at 30 min, showing an RMSE of 2.2451, MAPE of 50.95%, and a poor R^2 of 0.6356. The consistent superiority of VMD-based models—especially VMD-LogTrans—demonstrates that decomposing WS time series into IMFs allows the model to better capture the nonlinearities and fluctuations, making the data more interpretable and learnable for deep architectures. This is particularly important in wind energy applications, where accurate forecasting is critical for operational planning, turbine control, and integration into power systems. The results confirm that VMD-LogTrans is a powerful approach for enhancing the accuracy and reliability of WSF across different time horizons.

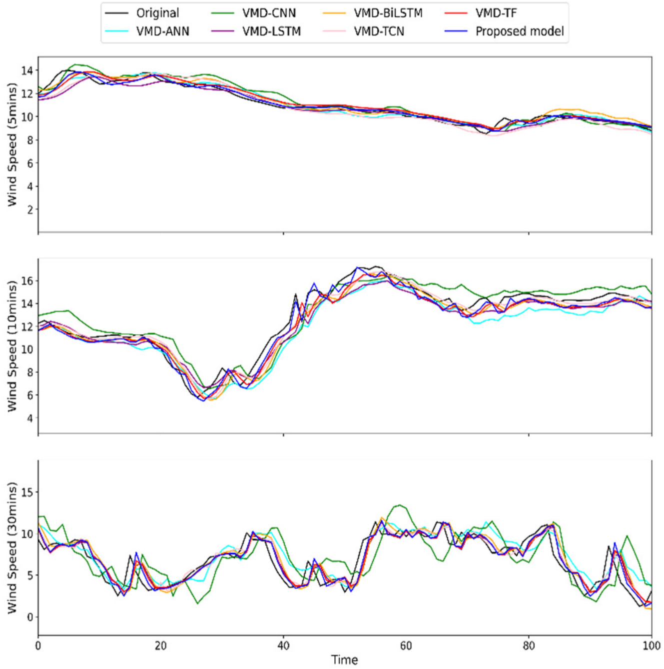


Fig. 3. Forecasting results of all hybrid models.

TABLE III. WSF RESULTS OF THE PROPOSED, INDIVIDUAL, AND HYBRID MODELS AT THE 30-MIN INTERVAL

Model	RMSE	MAE	MSE	MAPE (%)	R^2
ANN	1.6920	1.4362	2.8629	23.14	0.7731
CNN	2.2451	1.7872	5.0406	50.95	0.6356
LSTM	1.1689	0.8794	1.3664	15.94	0.8470
BiLSTM	1.2468	0.9162	1.5544	18.86	0.8253
TCN	1.1324	0.8310	1.2824	15.31	0.8682
Trans	1.0385	0.6853	1.0786	13.63	0.8903
LogTrans	1.0040	0.6428	1.0080	12.27	0.9038
VMD-ANN	0.8215	0.6920	0.6748	12.68	0.8674
VMD-CNN	0.8758	0.7021	0.7670	14.49	0.8190
VMD-LSTM	0.5495	0.4470	0.3019	8.20	0.9142
VMD-BiLSTM	0.6910	0.5833	0.4775	10.46	0.9173
VMD-TCN	0.5290	0.4385	0.2799	7.76	0.9293
VMD-Trans	0.4650	0.3736	0.2162	6.59	0.9317
Proposed	0.4255	0.3582	0.1810	5.95	0.9520

When comparing the VMD-LogTrans model with the second-best hybrid model, VMD-Transformer, it becomes clear that both models benefit from VMD, but LogTrans is still more accurate. At all three intervals—5 min, 10 min, and 30 min—VMD-LogTrans consistently gave lower error values and higher R^2 scores than the VMD-Transformer. For example, at the 10-min interval, VMD-LogTrans had an RMSE of 0.2460 and an MAPE of 2.46%, while VMD-Transformer had an RMSE of 0.2931 and an MAPE of 2.79%. At the 30-min interval, the difference remained in favor of VMD-LogTrans. This shows that while both models are improved by using VMD, LogTrans is better at learning from the decomposed WS data and is more reliable for forecasting across different time scales. Figure 2 displays the forecasting results of the proposed model in comparison with the original WS data, while Figure 3 illustrates the forecasting results of all hybrid models, including the proposed model, alongside the original data.

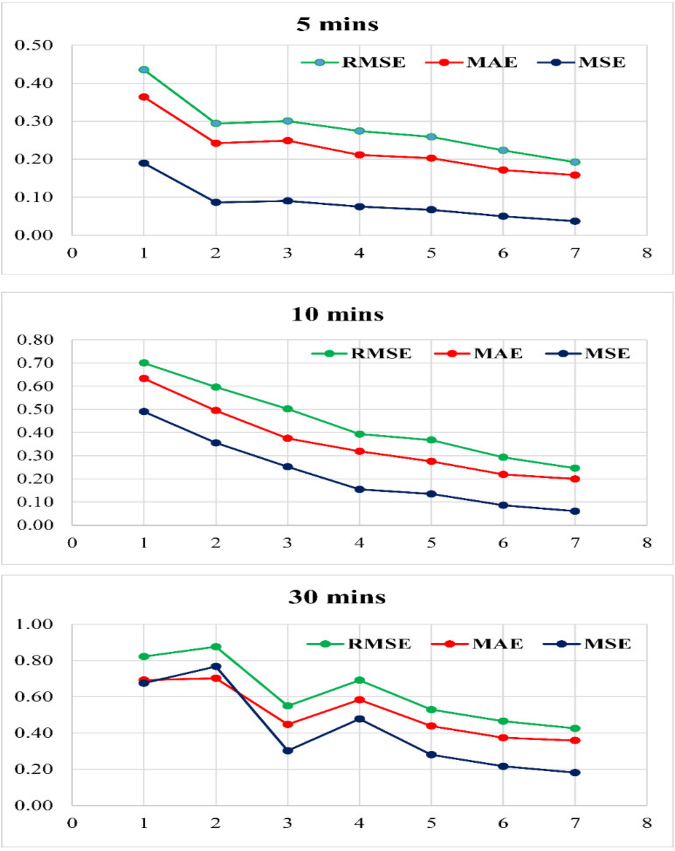


Fig. 4. Comparative analysis of hybrid models.

Figure 4 presents a comparative analysis of different VMD-based hybrid models for 5-min, 10-min, and 30-min WSF, using RMSE, MAE, and MSE as performance metrics. Each point on the x-axis corresponds to a specific model, ranging from VMD-ANN (1) to VMD-LogTrans (7). The overall trend across all three metrics—RMSE (green), MAE (red), and MSE (blue)—shows a consistent decline while moving from VMD-ANN to VMD-LogTrans, indicating a progressive improvement in the forecasting accuracy.

In summary, among the individual models tested, LogTrans delivers the most accurate WS forecasts. When integrated with VMD, the resulting VMD-LogTrans model consistently outperforms the other approaches in terms of accuracy and reliability. This is especially important in wind energy applications, where accurate forecasts help in planning, controlling turbines, and integrating wind energy into the power system.

The model training was conducted on Google Colab, where the VMD-LogTrans model required approximately 24 s per epoch. For the comparison, the standard Transformer incurs a time and memory complexity of $O(L^2)$ due to its full self-attention across all token pairs. In contrast, LogTrans significantly improves efficiency by using a LogSparse attention pattern, reducing the complexity to $O(L(\log L)^2)$. This reduction led to noticeably lower memory usage and faster execution, particularly when handling longer input sequences.

This reduced computational and memory demand enabled LogTrans to train more efficiently, especially on longer sequences, and contributed to the observed speed and resource benefits during the experiments.

IV. CONCLUSIONS

This study presented a comprehensive evaluation of various Deep Learning (DL) models for short-term Wind Speed Forecasting (WSF), focusing on individual models and their hybrid forms integrated with Variational Mode Decomposition (VMD). Among the individual models, Log-Sparse Transformer (LogTrans) consistently outperformed others across all forecasting horizons, demonstrating a superior ability to capture the temporal dependencies and nonlinear trends in WS data. When combined with VMD, the forecasting accuracy of these models improved significantly, with the hybrid VMD-LogTrans model achieving the best overall performance. This model achieved the lowest error rates and highest R^2 score values at 5-min, 10-min, and 30-min intervals, indicating a strong predictive power and robustness across multiple time scales.

These findings are consistent with established research trends that show how combining the signal decomposition methods with DL enhances the forecasting accuracy. By isolating the frequency components and simplifying the input signal, the decomposition techniques allow deep models to learn more effectively. The proposed VMD-LogTrans model builds upon this approach and demonstrates improved accuracy, especially at longer forecasting horizons where traditional models often struggle.

Finally, the proposed model offers a reliable and scalable solution for multi-horizon WSF, supporting more accurate operational planning, improved turbine control, and effective integration of the wind energy into power systems.

REFERENCES

[1] S. S. Soman, H. Zareipour, O. Malik, and P. Mandal, "A review of wind power and wind speed forecasting methods with different time horizons," in *North American Power Symposium 2010*, Arlington, TX, USA, Sept. 2010, pp. 1–8, https://doi.org/10.1109/NAPS.2010.5619586.

[2] M. U. Yousuf, I. Al-Bahadly, and E. Avci, "Statistical wind speed forecasting models for small sample datasets: Problems, Improvements, and prospects," *Energy Conversion and Management*, vol. 261, June 2022, Art. no. 115658, https://doi.org/10.1016/j.enconman.2022.115658.

[3] T. N. Le, N. A. Nguyen, T. N. T. Huynh, Q. T. Le, T. T. H. Huynh, and T. T. H. Le, "Applying Intelligent Algorithms In Short-Term Electrical Load Forecasting," *Engineering, Technology & Applied Science Research*, vol. 14, no. 5, pp. 16365–16370, Oct. 2024, https://doi.org/10.48084/etasr.8304.

[4] S. Sahoo, S. C. Swain, R. Dash, and P. Samal, "Forecasting Tariff Rates and Enhancing Power Quality in Microgrids: The Synergistic Role of LSTM and UPQC," *Engineering, Technology & Applied Science Research*, vol. 14, no. 1, pp. 12506–12511, Feb. 2024, https://doi.org/10.48084/etasr.6481.

[5] A. F. Khalil and S. Rostam, "Machine Learning-based Predictive Maintenance for Fault Detection in Rotating Machinery: A Case Study," *Engineering, Technology & Applied Science Research*, vol. 14, no. 2, pp. 13181–13189, Apr. 2024, https://doi.org/10.48084/etasr.6813.

[6] Z. Wang, J. Zhang, Y. Zhang, C. Huang, and L. Wang, "Short-Term Wind Speed Forecasting Based on Information of Neighboring Wind Farms," *IEEE Access*, vol. 8, pp. 16760–16770, 2020, https://doi.org/10.1109/ACCESS.2020.2966268.

[7] Shivani, K. S. Sandhu, and A. Ramachandran Nair, "A Comparative Study of ARIMA and RNN for Short Term Wind Speed Forecasting," in *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Kanpur, India, July 2019, pp. 1–7, https://doi.org/10.1109/ICCCNT45670.2019.8944466.

[8] G. Memarzadeh and F. Keynia, "A new short-term wind speed forecasting method based on fine-tuned LSTM neural network and optimal input sets," *Energy Conversion and Management*, vol. 213, June 2020, Art. no. 112824, https://doi.org/10.1016/j.enconman.2020.112824.

[9] Y. M. Zhang and H. Wang, "Multi-head attention-based probabilistic CNN-BiLSTM for day-ahead wind speed forecasting," *Energy*, vol. 278, Sept. 2023, Art. no. 127865, https://doi.org/10.1016/j.energy.2023.127865.

[10] H. Wu, K. Meng, D. Fan, Z. Zhang, and Q. Liu, "Multistep short-term wind speed forecasting using transformer," *Energy*, vol. 261, Dec. 2022, Art. no. 125231, https://doi.org/10.1016/j.energy.2022.125231.

[11] J. Manero, J. Béjar, and U. Cortés, "Wind Energy Forecasting with Neural Networks: A Literature Review," *Computación y Sistemas*, vol. 22, no. 4, Dec. 2018, https://doi.org/10.13053/cys-22-4-3081.

[12] S. K. Paramasivan, "Deep Learning Based Recurrent Neural Networks to Enhance the Performance of Wind Energy Forecasting: A Review," *Revue d'Intelligence Artificielle*, vol. 35, no. 1, pp. 1–10, Feb. 2021, https://doi.org/10.18280/ria.350101.

[13] Z. Guo, W. Zhao, H. Lu, and J. Wang, "Multi-step forecasting for wind speed using a modified EMD-based artificial neural network model," *Renewable Energy*, vol. 37, no. 1, pp. 241–249, Jan. 2012, https://doi.org/10.1016/j.renene.2011.06.023.

[14] T. Liang, Q. Zhao, and H. Shi, "A novel combined model based on VMD and IMODA for wind speed forecasting," *Journal of Intelligent & Fuzzy Systems*, vol. 42, no. 4, pp. 2845–2861, Mar. 2022, https://doi.org/10.3233/JIFS-201191.

[15] Y. Zhang, B. Chen, G. Pan, and Y. Zhao, "A novel hybrid model based on VMD-WT and PCA-BP-RBF neural network for short-term wind speed forecasting," *Energy Conversion and Management*, vol. 195, pp. 180–197, Sept. 2019, https://doi.org/10.1016/j.enconman.2019.05.005.

[16] N. Bokde, A. Feijóo, D. Villanueva, and K. Kulat, "A Review on Hybrid Empirical Mode Decomposition Models for Wind Speed and Wind Power Prediction," *Energies*, vol. 12, no. 2, Jan. 2019, Art. no. 254, https://doi.org/10.3390/en12020254.

[17] J. Zhou, H. Liu, Y. Xu, and W. Jiang, "A Hybrid Framework for Short Term Multi-Step Wind Speed Forecasting Based on Variational Model Decomposition and Convolutional Neural Network," *Energies*, vol. 11, no. 9, Aug. 2018, Art. no. 2292, https://doi.org/10.3390/en11092292.

[18] S. Zhang, C. Zhu, and X. Guo, "Wind-Speed Multi-Step Forecasting Based on Variational Mode Decomposition, Temporal Convolutional Network, and Transformer Model," *Energies*, vol. 17, no. 9, Apr. 2024, Art. no. 1996, https://doi.org/10.3390/en17091996.

[19] B. Jiang, Y. Liu, and H. Xie, "Super short-term wind speed prediction based on CEEMD decomposition and BILSTM-Transformer model," in *2023 IEEE 3rd International Conference on Power, Electronics and Computer Applications (ICPECA)*, Shenyang, China, Jan. 2023, pp. 876–882, https://doi.org/10.1109/ICPECA56706.2023.10076025.

[20] Q. Bing, P. Zhao, C. Ren, X. Wang, and Y. Zhao, "Short-Term Traffic Flow Forecasting Method Based on Secondary Decomposition and Conventional Neural Network–Transformer," *Sustainability*, vol. 16, no. 11, May 2024, Art. no. 4567, https://doi.org/10.3390/su16114567.

[21] K. Zhang, X. Li, and J. Su, "Variable Support Segment-Based Short-Term Wind Speed Forecasting," *Energies*, vol. 15, no. 11, June 2022, Art. no. 4067, https://doi.org/10.3390/en15114067.

[22] J. Huan, L. Deng, Y. Zhu, S. Jiang, and F. Qi, "Short-to-Medium-Term Wind Power Forecasting through Enhanced Transformer and Improved EMD Integration," *Energies*, vol. 17, no. 10, May 2024, Art. no. 2395, https://doi.org/10.3390/en17102395.

[23] M. H. Mohamed, E.-T. Nemat, M. Ramadan, and H. B. E. Amany, "Powering Up Efficiency: A Deep Learning Model For Accurate Electricity Consumption Forecasting," *Journal of Theoretical and Applied Information Technology*, vol. 102, no. 18, pp. 6623–6637, Sept. 2024.

[24] S. Li *et al.*, "Enhancing the Locality and Breaking the Memory Bottleneck of Transformer on Time Series Forecasting," in *33rd Conference on Neural Information Processing Systems*, Vancouver, Canada, Jan. 03, 2020, 1–11.

[25] "National Renewable Energy Laboratory Data CatLog, U.S. Department of Energy," Jun. 2025. https://data.nrel.gov/.