Construction of an Ensemble Model of LSTM-IDW-GBRT for Wind Speed Interpolation: Case Study Over a Multi-Terrain Region

Bo Wang, Chun Liu, Shu Wang, Shuanglong Jin, Chang Liu, Zongpeng Song, Hu Yang, Haoyuan Ma and Hongqing Wang

Abstract—**The assessment of wind resources is crucial for wind farm siting and wind power forecasting. Currently, wind measurement data is often impacted by natural and human factors, leading to a high incidence of anomalies and missing values. This study introduces an LSTM-IDW-GBRT spatiotemporal ensemble model that leverages both temporal and spatial correlations of wind resources to interpolate regional wind speed fields effectively. The model integrates the Long Short-Term Memory (LSTM) neural network for capturing temporal dependencies in wind speed, the Inverse Distance Weighting (IDW) method for spatial correlation, and the Gradient Boosting Regression Tree (GBRT) model to integrate these single models and optimize the final interpolated wind speed predictions. By comparing the proposed ensemble model with its constituent sub-models, traditional wind speed interpolation approaches, and state-of-the-art neural network architectures, the results indicate that the LSTM-IDW-GBRT ensemble model outperforms both individual models and the traditional MCP linear regression model in wind speed interpolation, with particularly improved performance in inland and high-altitude areas and during periods of wind speed fluctuations. The ensemble model achieved mean error metrics of 0.8362 (MSE), 0.6778 (MAE), and 0.8746 (RMSE), representing improvements of 56.9%, 32.5%, and 37.9%, respectively. The proposed model shows strong applicability for wind speed interpolation and prediction, significantly enhancing the accuracy of wind resource assessments.**

Index Terms— **Long Short-Term Memory neural network, Inverse Distance Weighting, the Gradient Boosting Regression Tree, wind speed interpolation**

Manuscript received November 3, 2024; revised July 7, 2025.

This work was supported by the National Natural Science Foundation of China(Grant No.52177121) and Science and Technology Project of State Grid Corporation of China (5108-202218280A-2-315-XG).

Bo Wang is a manager of China Electric Power Research Institute, Beijing, China (e-mail: wangbo@epri.sgcc.com.cn).

Chun liu is a manager of China Electric Power Research Institute, Beijing, China (e-mail: liuchun@epri.sgcc.com.cn).

Shu Wang is a Researcher of China Electric Power Research Institute, Beijing, China (corresponding author, Tel: (+86)010-82814130; e-mail: wangshu2011@pku.edu.cn).

Shuanglong Jin is a researcher of China Electric Power Research Institute, Beijing, China (e-mail: jinshuanglong@epri.sgcc.com.cn).

Chang Liu is a graduate of Mathematics Department, Beijing Forestry University, Beijing, 100083 China(e-mail: 18813153618@163.com).

Zongpeng Song is a researcher of China Electric Power Research Institute, Beijing, China (e-mail: songzongpeng@epri.sgcc.com.cn).

Hu Yang is a graduate of Mathematics Department, Beijing Forestry University, Beijing, 100083 China(e-mail: yhu340321@bjfu.edu.cn).

Haoyuan Ma is a graduate of Mathematics Department, Beijing Forestry University, Beijing, 100083 China(e-mail: mhyjwg@126.com).

Hongqing Wang is a professor of Mathematics Department, Beijing Forestry University, Beijing, 100083 China(e-mail: wanghq@bjfu.edu.cn).

I. Introduction

THE global development of renewable energy has witnessed remarkable progress. According to the International Renewable Energy Agency (IRENA), the global renewable energy generation capacity reached 4,448 GW in 2024, reflecting a year-on-year increase of 15.1% [1]. Among these sources, wind and solar energy have emerged as the fastest-growing forms, accounting for over 30% of global renewable energy production in 2023 [2]. China plays a pivotal role in this global transition toward renewable energy. In 2023, China's renewable energy installed capacity exceeded 1,000 GW, constituting 43.5% of the global total [3]. Notably, in the wind energy sector, China achieved a wind power capacity of 440 GW, which is the world leader [4]. The Chinese government has set ambitious targets, aiming to expand wind and solar capacity to 1,200 GW by 2030 and increase the share of non-fossil fuels in total energy consumption to 80% by 2060 [5].

Wind power constitutes a cornerstone of China's renewable energy strategy, contributing significantly to carbon emission reductions and reducing reliance on fossil fuels. In 2023, wind power generation accounted for a substantial portion of the national electricity output, playing a vital role in ensuring the stability of the power system and advancing the green transition [6]. However, the planning and operation of wind power projects hinge on precise wind resource assessments, with the acquisition of reliable wind speed data being the core requirement [7]. Manwell et al. highlights that the accuracy of wind speed data directly influences the economic viability of wind farms and the efficiency of grid scheduling [8]. Accurate wind speed data are essential for estimating the potential power output of wind farms, yet this process is fraught with challenges. Environmental and human-induced factors—such as equipment malfunctions, extreme weather, communication failures, and inadequate maintenance—can lead to anomalies or gaps in the data [9][10].

Currently, wind resource assessment is categorized into regional and site-specific evaluations. Regional wind resource assessment predominantly employs numerical simulation techniques. For instance, Wang et al. [11] developed a mesoscale high-resolution numerical simulation system tailored to China's climatic characteristics using the Climate Four-Dimensional Data Assimilation (CFDDA) climate four-dimensional data assimilation technique and the

Weather Research and Forecasting (WRF) numerical weather prediction model. This system generated a 30-year gridded dataset of wind energy resources, elucidating the spatiotemporal distribution characteristics of wind resources across China. Similarly, Yang et al [12] utilized the WRF model with the Mellor-Yamada-Janjic (MYJ) boundary layer scheme and the Monin-Obukhov surface layer scheme to simulate wind speed data, albeit at the cost of substantial high-performance computing resources. Meng et al [13] explored the relationship between the third moment of wind speed and Weibull distribution parameters, proposing a method to calculate the third moment that enhances parameter estimation accuracy and improves fitting precision.

Site-specific wind resource assessment typically involves installing wind measurement towers or meteorological stations in the vicinity of prospective wind farm sites, conducting measurements and observations over an annual cycle. The Measure-Correlate-Predict (MCP) algorithm is commonly employed, leveraging the principle of spatial correlation to establish a model between observed wind speeds and the target wind farm site, thereby estimating local wind resources based on historical data [10]. Several mature wind resource assessment software packages integrate a variety of MCP algorithms. For example, Windographer offers eight widely used MCP methods, including linear regression, Weibull fitting, quicksort, orthogonal regression, wind speed ratio, variance ratio, matrix time series, and vertical stratification algorithms [14][15]. Nevertheless, wind speed interpolation and prediction methods, such as statistical models (e.g., MCP, AutoRegressive ensemble Moving Average (ARIMA)) and physical models (e.g., Numerical Weather Prediction (NWP)), are limited to establishing mathematical relationships between observations. These approaches often overlook the statistical characteristics of spatiotemporal correlations within the data, struggling to capture the complex spatiotemporal variability of wind speeds. Moreover, single models exhibit inherent limitations when addressing both temporal and spatial information [16]. Consequently, numerous researchers have turned to artificial intelligence (AI)-based hybrid models for wind speed interpolation and prediction, achieving promising results. In recent years, deep learning models, particularly Long Short-Term Memory (LSTM) networks, have garnered attention due to their capability to handle temporal dependencies in sequential data. Hybrid models integrating LSTM with other techniques have demonstrated remarkable performance. For instance, Li et al. [17]proposed a hybrid model combining Improved Complete ensemble Empirical Mode Decomposition with Adaptive Noise (ICEEMDAN), Multi-scale Fuzzy Entropy (MFE), LSTM, and Informer for short-term wind speed forecasting, outperforming traditional methods. Similarly, Zhang et al. [18]introduced a novel hybrid model for short-term wind speed prediction, leveraging advanced machine learning algorithms enhanced by data preprocessing and feature selection techniques to improve accuracy. Liu et al. [19]developed a deep learning-based wind speed forecasting approach that integrates Empirical Wavelet Transform (EWT), LSTM, and Elman neural networks(ENN), effectively capturing both linear and nonlinear patterns in wind speed data and thereby enhancing prediction precision. Hybrid models often incorporate optimization algorithms to fine-tune parameters. For example, Wang et al. [20] employed a hybrid model based on grey relational analysis and support vector machines optimized by the cuckoo search algorithm, achieving superior short-term wind speed forecasting results compared to standalone models. Gradient Boosting Regression Trees (GBRT), as an ensemble learning method, enhance model generalization by combining multiple weak learners through gradient boosting, with its mathematical foundation established by Friedman [21]. Mohandes et al. demonstrated GBRT's efficacy in wind speed prediction, particularly when integrating diverse features [22].

Furthermore, compared to wind speed interpolation based solely on a single station's time series, incorporating data from neighboring stations yields more accurate results. Lima et al. [23] combined NWP models with statistical techniques, employing a regional atmospheric modeling system for wind speed forecasting in two distinct Brazilian regions and using Kalman filtering to reduce systematic errors. Their findings offer valuable insights for achieving more precise wind power studies. Wang et al. [24] established and trained multiple neural network models using wind speed data from wind farms and nearby meteorological stations, optimizing model parameters with a particle swarm optimization algorithm to enhance wind resource assessment accuracy. Lin et al. [25] proposed a hybrid approach based on Graph Attention Networks (GAT) and an optimizable graph matrix, utilizing geographic information, Dynamic Time Warping (DTW), and Maximal Information Coefficient (MIC) to construct novel parameters for optimizing the matrix, effectively integrating spatial and temporal features to improve wind speed prediction accuracy. Baïle et al. investigated how utilizing wind speed data from nearby meteorological stations can enhance short-term wind speed forecasts at a specific location. Their study demonstrated that incorporating spatiotemporal information from neighboring stations can significantly improve forecast accuracy, particularly within the 1 to 6-hour prediction window, with root mean square error (RMSE) reductions of up to 20% [26]. Additionally, Reinhardt evaluated 12 wind speed interpolation methods (e.g., Inverse Distance Weighting(IDW), Kriging, and linear regression) to address data gaps at wind farm sites using neighboring station data, finding Kriging to be the most effective in most cases, while IDW excelled in areas with dense station coverage [27]. Palutikof et al. validated the applicability of IDW in complex terrains, further reinforcing its utility [28].

Building on these advancements, this study proposes an ensemble LSTM-IDW-GBRT model for wind speed interpolation. The model first employs LSTM to capture temporal dependencies within wind speed time series, then utilizes IDW to leverage wind speed data from neighboring stations for spatial interpolation, and finally integrates these components using GBRT to enhance overall prediction accuracy. This approach fully exploits the temporal and spatial characteristics of the data, offering a novel solution for wind speed interpolation in multi-terrain regions.

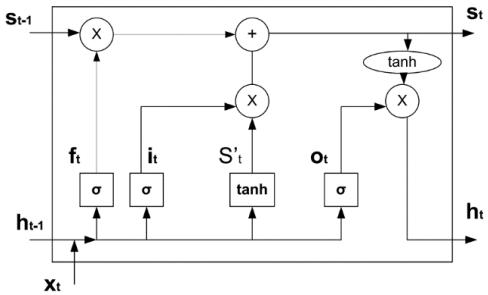

Fig.1. Illustration of the LSTM model

II. Technical PRINCIPLES

A. LSTM

Based on Recurrent Neural Network (RNN), the LSTM network adds a gating mechanism. It introduces a "forget gate" to discard unnecessary information, an "input gate" to store new useful information, and an "output gate" to produce output based on the current cell state, thereby effectively alleviating the gradient vanishing and exploding problems that are inherent to RNNs when modeling long-range dependencies[29]. Through these gate mechanisms, LSTM can more effectively capture long-term dependencies, which is essential for time-varying data like wind speed. The working principle of LSTM in wind speed prediction is detailed in [17] , and its workflow is shown in Figure 1.

B. IDW

The IDW interpolation method is a distance-based spatial interpolation technique [30]. Its core idea is to estimate the value of an unknown point by taking a weighted average of known point values, where each weight is determined by a distance function between the known and unknown points. Specifically, given a set of known points and their corresponding values on a two-dimensional plane, the estimation process for an unknown point's value is as follows:

Step 1 Calculate the distance between each known point and the unknown point.

$$d_i = \sqrt{(x_i - x_0)^2 + (y_i - y_0)^2} \quad (1)$$

where (x_i, y_j) represents the coordinates of the *i*-th known point (x_0, y_0) represents the coordinates of the unknown point, and d_i denotes the distance between the *i*-th known point and the unknown point.

Step 2 Calculate the weight for each known point value based on the distance obtained in Step 1:

$$w_i = \frac{1}{d_i^P} \quad (2)$$

where w_i represents the weight of the *i*-th known point value, and *P* is a distance exponent parameter that controls the influence of distance on the weight。

Step 3 Based on the distances and weights calculated in the previous steps, the value of the unknown point can be estimated using the following formula:

$$z_0 = \frac{\sum_{i=1}^{n} w_i z_i}{\sum_{i=1}^{n} w_i} \quad (3)$$

where z_0 and z_i represent the estimated value at the unknown point and the value at the *i*-th known point, respectively, and *n* is the total number of known points.

The Inverse Distance Weighting (IDW) interpolation method is recognized for its simplicity, ease of implementation, and its ability to account for spatial correlations in the data. This method is particularly effective for interpolating in cases where the data are irregularly distributed. In the context of meteorological data collected from sparsely and unevenly distributed wind farm stations, the IDW method facilitates interpolation by utilizing data from proximate stations, thereby generating gridded estimates for the wind farm locations and their surrounding areas.

C. GBRT

Gradient Boosting Regression Tree (GBRT) is a decision tree algorithm within the gradient boosting framework. It 和 adds new tree models in a Boosting iterative process to correct the residuals of previous models, thereby enhancing accuracy. In each iteration t , the GBRT algorithm calculates the residual $r_{i,t} = y_i - F_{t-1}(x_i)$, where y_i is the observed value, and $F_{t-1}(x_i)$ represents the model output at the current iteration. A new decision tree $h_t(x)$ is then trained on the residuals as target values, and the model is updated as $F_t(x) = F_{t-1}(x) + \nu \cdot h_t(x)$, where *v* is the learning rate. The final output is the weighted sum of all tree models. In this study, the GBRT model is used to integrate the interpolation results from the LSTM model and the IDW method, thereby producing the final ensemble interpolated wind speed values.

A. Pearson Correlation Coefficient Method

As a phenomenon of air flow, wind speed in neighboring regions within the same time period exhibits spatial correlation. To verify the spatial correlation of wind speed in adjacent regions at the same time, this study employs the Pearson correlation coefficient method. The Pearson correlation coefficient is a measure used to assess the strength of the linear relationship between two variables, and its calculation formula is as follows:

$$\rho_{X,Y} = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E\left((X-\mu_X)(Y-\mu_Y)\right)}{\sqrt{\sum_{i=1}^{n}(X-\mu_X)^2 \sum_{i=1}^{n}(Y-\mu_Y)^2}} \quad (4)$$

where X and Y are continuous variables, $\mathrm{cov}(X,Y)$ denotes the covariance matrix between X and Y, σ_X and σ_Y represent the standard deviations of X and Y, and μ_X and μ_Y are the means of X and Y, respectively.

The Pearson correlation coefficient ranges from -1 to 1, where -1 indicates a perfect negative correlation, 1 indicates a perfect positive correlation, and 0 means no linear correlation between the variables. The closer the Pearson correlation coefficient is to 1 or -1, the stronger the linear relationship between the variables. Generally, a coefficient greater than 0.4 indicates a strong positive correlation between the two variables.

III. BUILDING the ensemble Model

Given the temporal and spatial correlations in wind speed data, an ensemble model based on the spatiotemporal characteristics of LSTM-IDW-GBRT is constructed. This model integrates data from neighboring meteorological stations, which exhibit spatial correlation, along with historical wind speed data from the target station, which reflects temporal correlation. The specific workflow of the model is illustrated in Figure 2.

Based on the model workflow, for a meteorological station with wind speed data that requires interpolation, two nearby stations are selected. Initially, historical data from all three stations undergoes data prepossessing, where single anomalous or missing values are filled using the mean of adjacent time points. Subsequently, to verify the spatial correlation among the stations' wind speeds, the Pearson correlation coefficient method is applied to calculate the correlation coefficients between each pair of the three stations' wind speed data.

Using the processed data from each station, the LSTM model and IDW interpolation method are applied for fitting and training. The historical data of the target station needing interpolation is input into the LSTM model through a 12-step rolling training procedure. The wind speed data from nearby stations is then used to compute the IDW interpolated wind speed for the target station, with the optimal parameters for IDW selected based on the target station's historical wind speed data. Finally, the outputs from the LSTM and IDW models, alongside the actual historical data, are integrated into the GBRT model for ensemble learning."

This study evaluates the interpolation results using three error metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Square Error (RMSE). The formulas for these error metrics are as follows:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{5}$$

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{6}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{7}$$

where n is the number of interpolated output values, y_i represents the true wind speed value, and $\hat{y}_i$ denotes the wind speed value output by the interpolation model.

IV. Case Study Analysis

A. Data Description and Preprocessing

Shandong Province, located in the eastern coastal region of China, has a warm temperate monsoon climate, primarily featuring plains and hills, along with mountain ranges and low-lying coastal areas. The region's wind resources are favorable, influenced mainly by the East Asian monsoon and sea-land winds. To evaluate the interpolation effectiveness of the proposed ensemble model for wind speed data with partial missing values, four sets of meteorological stations are selected from national basic meteorological stations within Shandong Province, based on the regional topography. Each group consists of three meteorological stations and includes data on wind speed, station ID, and latitude and longitude. The wind speed data spans the entire year of 2022, with a time resolution of 1 hour and a sampling interval of 2 minutes, yielding 8,760 wind speed entries per station. Figure 3 shows the spatial distribution of each group of meteorological stations, labeled from left to right as Group A, Group B, Group C, and Group D. The wind speed data and basic information for each station are summarized in Table 1.

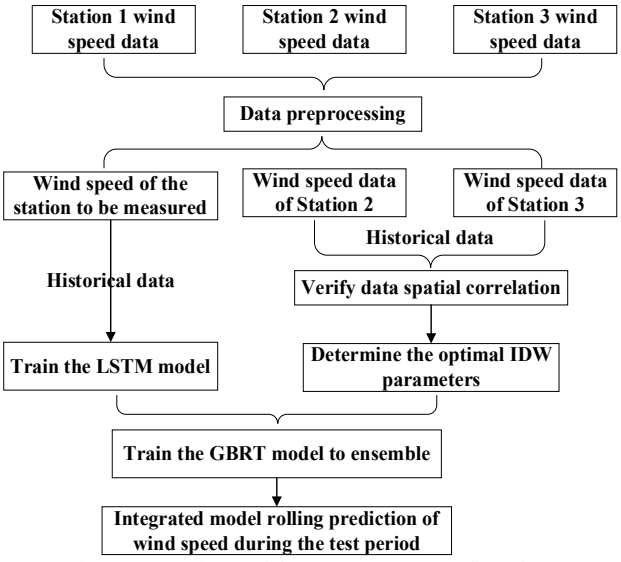

Fig. 2. Integration model LSTM-IDW-GBRT flow chart

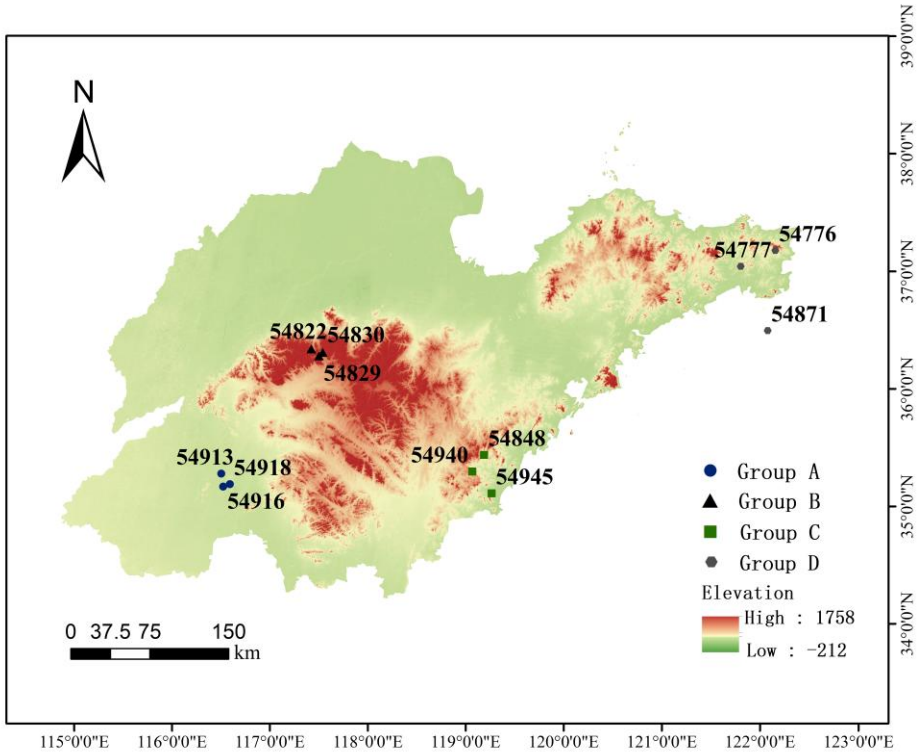

Fig. 3. Map of location of weather stations

TABLE I
TABLE OF BASIC WEATHER STATION INFORMATION

Group	Station ID	Longitude	Latitude	Elevation (m)	Data Volume	Characteristics
A	54913	116.49	35.45	62.3	8760	Uniform elevation, inland area
	54916	116.51	35.34	51.7	8760	
	54918	116.58	35.36	62.3	8760	
B	54822	117.44	36.52	47.4	8760	Low average elevation, mountainous region
	54829	117.52	36.46	85.7	8760	
	54830	117.56	36.49	34.4	8760	
C	54848	119.25	35.59	82.6	8760	High average elevation
	54940	119.12	35.45	148.2	8760	
	54945	119.32	35.26	36.9	8760	
D	54776	122.42	37.24	47.7	8760	Coastal area
	54777	122.04	37.12	118.0	8760	
	54871	122.29	36.56	9.8	8760	
Total					105120	

Among the groups, Group A represents an inland area with relatively uniform altitude, Group B is a mountainous region with a lower average altitude, Group C has a higher and more uneven altitude distribution, and Group D is a coastal area.

For occasional single missing or anomalous values in the raw data, a nearest-neighbor interpolation method is applied due to the continuity of time series data. Specifically, if x_i is a wind speed value that needs to be filled, the average of the wind speeds from the previous time point x_{i-1} and the next time point x_{i+1} is used to fill x_i, as follows:

$$x_i = \frac{x_{i-1} + x_{i+1}}{2} \tag{8}$$

Following data cleaning, the complete wind speed data for each meteorological station in 2022 was divided, with the first 80% designated as the training set for training the LSTM, IDW, and GBRT ensemble models, and the remaining 20% allocated as the test set for interpolation validation.

Figure 4 presents the box plots of actual wind speeds for each station after data preprocessing. In each box plot, the central line represents the median wind speed at the respective station, while the height of the box corresponds to the interquartile range (IQR), spanning from the first quartile (Q1) to the third quartile (Q3). A taller box indicates a larger IQR, reflecting greater variability in the data. Outliers are marked by diamond symbols, signifying values that fall

substantially outside the interquartile range.

From Figure 4, it is evident that although the distribution of actual wind speed data exhibits minor variations across stations, it remains relatively consistent overall. Notably, each station shows a significant number of high wind speed peaks. The median wind speed for Group D, comprising three coastal meteorological stations, is comparatively higher, with a greater frequency of high wind speeds. This suggests that coastal stations experience more pronounced wind speed fluctuations and a higher occurrence of strong winds.

B. Verifying Spatial Correlation

To investigate the inter-station correlations of wind speed sequences within each group and to further inform the integration strategy of spatial information in the wind speed interpolation model, this study calculates the Pearson correlation coefficients between all pairs of 12 meteorological stations. The results are illustrated in Figure 5. In this figure, both the numerical values and the shape of ellipses in the heatmap intuitively convey the spatial correlation of wind speeds between stations. As observed in Figure 5, the wind speed sequences of stations within the same group exhibit generally strong correlations, with most inter-station coefficients exceeding 0.5. For instance, in Group A, the coefficient between stations A-54913 and A-54916 reaches as high as 0.71; in Group B, the correlation between B-54829 and B-54822 is 0.65, and between B-54829 and B-54830 is 0.75. These values reflect a high degree of temporal synchronicity, indicating consistent wind speed variation patterns under similar climatic and topographical conditions, thus laying a solid foundation for using data from neighboring stations in spatial interpolation.

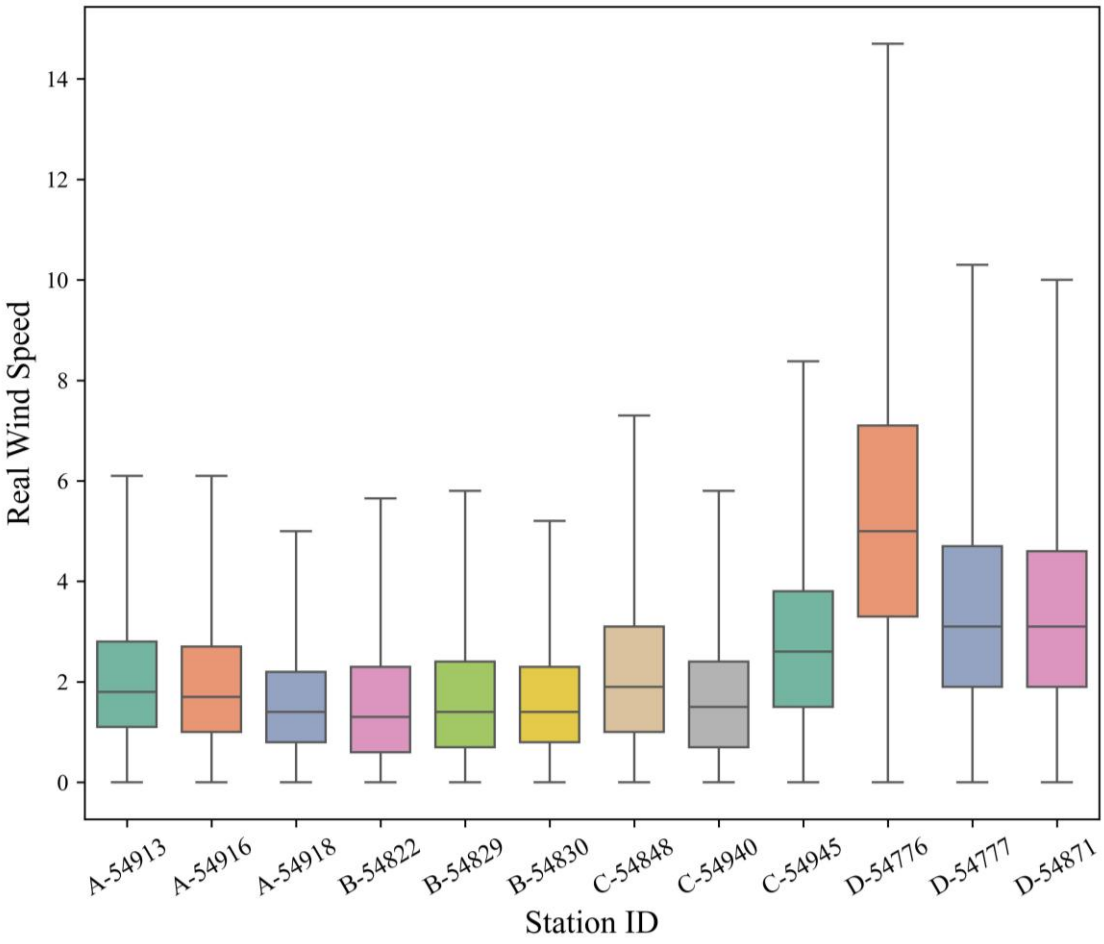

Fig.4. Box plot of wind speed at each meteorological station

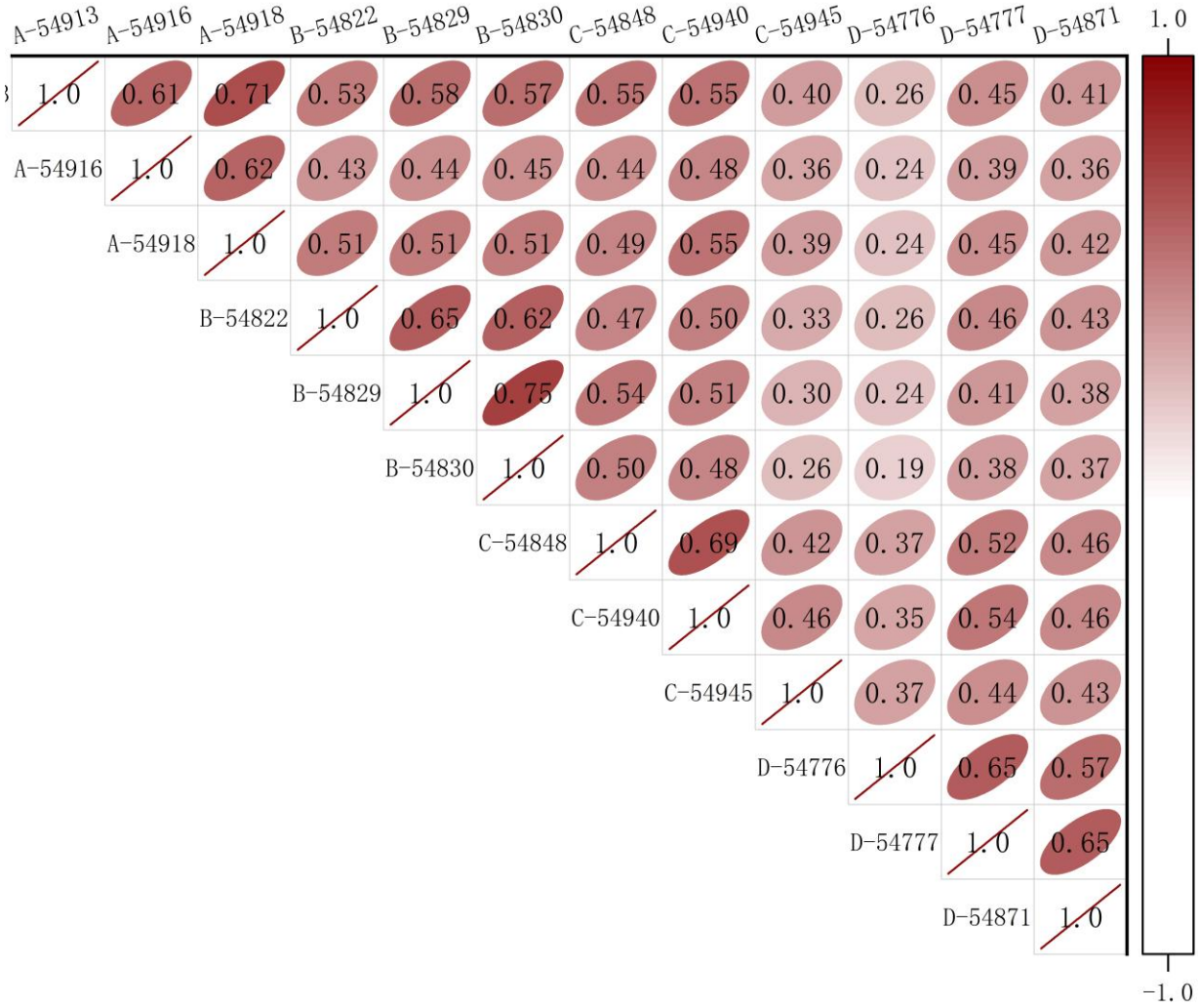

Fig.5. Box plot of wind speed at each meteorological station

In contrast, wind speed correlations between stations from different groups are significantly lower, with most coefficients below 0.4 and some even under 0.3. For example, the correlation between A-54913 and D-54776 is only 0.26, and between B-54830 and D-54777 merely 0.19. These results confirm the existence of pronounced spatial attenuation in wind speed: as geographical distance increases, the spatial correlation of wind speed diminishes accordingly. This spatial heterogeneity poses a challenge to interpolation model construction, particularly in regions with large spatial spans or complex terrain, where traditional distance-based methods such as IDW may suffer from performance degradation.

A closer examination of intra-group spatial structure reveals micro-level differences as well. For example, in Group C, while the correlation between C-54848 and C-54940 is 0.69, that between C-54848 and C-54945 drops to 0.42. This disparity indicates that even within the same subregion, wind speed variation can still be affected by factors such as terrain features, microclimate conditions, or the layout of measurement stations. Such local spatial heterogeneity further underscores the necessity of incorporating a correlation-weighted mechanism into the interpolation process.

Additionally, although Group D exhibits relatively lower overall correlation compared to inland groups, the inter-station correlation among D-54776, D-54777, and D-54871 remains above 0.65. This suggests that in complex terrains such as coastal areas, spatial correlation—while generally weaker—can still be maintained at moderate to high levels within local subregions. This provides practical feasibility for interpolation at edge stations or in areas with complex geographical features.

In summary, the spatial correlation structure revealed by Figure 5 offers important insights for model development. On one hand, the strong intra-group correlation supports the effectiveness of IDW-based methods and spatially weighted regressions such as MCP. On the other hand, the observed spatial heterogeneity and correlation decay with distance offer a theoretical basis for incorporating spatial feature extraction mechanisms in the ensemble model—specifically, leveraging IDW for spatial trend estimation and LSTM for capturing temporal dependencies. Future studies may further incorporate a spatial correlation weighting matrix to dynamically adjust the influence of neighboring stations during interpolation, thereby improving both the accuracy and robustness of wind speed interpolation models.

C. Results Analysis

This section evaluates the performance of the LSTM-IDW-GBRT ensemble model in wind speed interpolation tasks. By integrating observational data from target meteorological stations and comparing the fitting results of various models, the study assesses the proposed

model's superiority in terms of interpolation accuracy and stability across multiple dimensions, including error metrics, fitting curves, terrain differences, and model structural advantages.

Figure 6 presents the radar charts of interpolation error metrics for four representative target stations from different groups (A-54918, B-54830, C-54940, and D-54777), comparing the proposed LSTM-IDW-GBRT model against baseline models such as LSTM, IDW, linear regression, segmented spline interpolation, and polynomial interpolation. The radar plots clearly indicate that the proposed ensemble model achieves the smallest coverage area in MSE, MAE, and RMSE, while the R^2 values are significantly higher than those of the other methods, highlighting its superiority in both fitting accuracy and stability.

Specifically, for station B-54830, the RMSE of the ensemble model is substantially lower than that of individual models such as LSTM and IDW, indicating that it not only fits the overall trend but also maintains high accuracy during periods of substantial wind speed fluctuation. Similarly, for stations A-54918 and C-54940, the ensemble model demonstrates lower error volatility compared to spline and polynomial interpolation, reflecting its enhanced ability to capture spatiotemporal patterns. Quantitatively, the average MSE, MAE, and RMSE across the four target stations are 0.639, 0.5045, and 0.897, respectively—more than 30% lower than those of traditional interpolation models—demonstrating a significant improvement in overall performance.

To further validate the model's ability to track temporal wind speed fluctuations, Figure 7-10 illustrates the interpolation curves of actual wind speed values versus predicted values across multiple randomly selected time intervals at the four target stations. The results show that the LSTM-IDW-GBRT interpolation curves closely follow the actual wind speed trends in most cases, particularly during periods of high-frequency oscillations and abrupt changes. For example, in Figure 8, station 54940 experiences a sharp rise and fall in wind speed between October 10 and 14, 2022. During this period, LSTM and IDW models exhibit delayed responses or underfitting, whereas the ensemble model accurately captures the trend and maintains a low prediction error. Additionally, in Figure 9, station 54777 shows a steady wind speed increase over several consecutive periods. Traditional methods such as linear regression or polynomial interpolation often suffer from edge-effect errors under such conditions, while the ensemble model yields consistent outputs aligned with the actual trend, demonstrating robustness to both short-term stability and mid-term variability.

This study also explores the model's generalization ability by examining performance across two major terrain categories: inland and coastal stations. Stations A-54918, B-54830, and C-54940 represent inland areas with high inter-station correlation, while station D-54777 represents a coastal area with complex topography. Results show that the ensemble model achieves higher accuracy and consistency in inland areas, with station A-54918 yielding an MSE of 0.3813 and an MAE of 0.473, outperforming all baseline methods. Even in coastal terrain, such as at station D-54777, the ensemble model maintains an RMSE of 1.319, about 40% lower than that of the IDW method. This demonstrates the model's robustness in adapting to varying terrain conditions and superior error suppression performance, especially in regions where traditional interpolation methods tend to struggle. Further analysis reveals that the average RMSE for inland stations is 0.7265, compared to 1.319 for coastal stations—a difference of 0.5925—which aligns with the well-known physical relationship between terrain complexity and wind speed variability. It also suggests that the ensemble model can adaptively adjust its weighting mechanism based on input features, thereby exhibiting strong terrain adaptability.

In summary, the ensemble model's outstanding interpolation accuracy and stability are attributable to its ensemble modeling of temporal and spatial features. The LSTM module effectively captures long-term dependencies and seasonal trends in wind speed sequences via its memory gate mechanisms. The IDW module incorporates spatial correlation by leveraging data from nearby stations, enhancing the model's sensitivity to local spatial variations. The GBRT module, acting as a nonlinear regression integrator, combines the outputs of LSTM and IDW, delivering a more adaptive interpolation outcome. This multicomponent architecture—characterized by temporal sensing, spatial reasoning, and model fusion—not only compensates for the limitations of single models under complex conditions but also enhances the model's robustness to data anomalies and abrupt changes, ensuring interpolation results remain closely aligned with actual wind speed dynamics.

To quantify the model's advantages more precisely, Table 2 lists the percentage reductions in error metrics across the four target stations. The ensemble model achieves average reductions of 56.4% in MSE, 37.15% in RMSE, and 32.55% in MAE, indicating not only improved accuracy but also greater stability with reduced error variability.

Further analysis of the standard deviation of prediction errors across the four stations confirms that the ensemble model yields significantly lower error fluctuations than traditional models. For instance, at station B-54830, the RMSE from the IDW model varies widely across different time points, while the ensemble model consistently maintains prediction error within ±0.3. This result highlights the superior error control capacity of the proposed method, effectively minimizing interpolation bias caused by data non-stationarity or irregular station placement.

To further demonstrate the superiority of the proposed model, we conduct comparative experiments against several state-of-the-art neural network architectures. Specifically, we include Informer, a cutting-edge temporal transformer-based forecasting model; CNN-BiLSTM-AM, which utilizes a convolutional neural network for feature extraction followed by a BiLSTM enhanced with a self-attention mechanism for prediction; and VMD-GRU-MC, which integrates Variational Mode Decomposition (VMD) for signal decomposition, a Gated Recurrent Unit (GRU) network for forecasting, and a Markov Chain-based (MC) component for error correction. The comparative results are presented in Table 3 and Figure 11-18.

Among all evaluated models, the proposed model achieved the best overall performance at Station 54918, with the

lowest Mean Squared Error (MSE = 0.3813), Mean Absolute Error (MAE = 0.473), and Root Mean Squared Error (RMSE = 0.6175), alongside the highest coefficient of determination (R^2 = 0.6451). In contrast, the Informer, CNN-BiLSTM-AM, and VMD-GRU-MC models yielded significantly higher error metrics and lower R^2 values, with VMD-GRU-MC recording the poorest performance. As illustrated in Figure 8, the predicted wind speed series from the proposed model closely follows the actual fluctuations, accurately capturing both the sharp rises and sudden drops. This alignment suggests a superior ability to track local variations and temporal trends, which baseline models fail to achieve consistently.

At Station 54830, the proposed model again exhibited strong predictive capabilities, outperforming all baselines with the lowest MSE (0.5461), MAE (0.5717), and RMSE (0.739), and the highest R^2 value (0.882). While the Informer model showed relatively better performance than CNN-BiLSTM-AM and VMD-GRU-MC, it still lagged behind the proposed model in all metrics. From the time series comparison in Figure 8, the proposed model demonstrates a remarkable ability to trace the wind speed trends, particularly in capturing turning points and trend reversals. In contrast, other models tend to either oversmooth the predictions or introduce significant lag, leading to suboptimal forecasting accuracy.

For Station 54940, the proposed model delivered the most accurate results, achieving the lowest MSE (0.6775), MAE (0.6527), and RMSE (0.8231), as well as the highest R^2 value (0.6775). The gap between the proposed model and the baselines, especially VMD-GRU-MC (MSE = 1.6903, R^2 = 0.432), is notably large. As shown in Figure 8, the proposed model's predictions maintain a strong correspondence with the observed wind speed curve, effectively reflecting both high-frequency oscillations and long-term trends. This indicates the model's strong capability in adapting to both short-term fluctuations and overall temporal structure—an aspect where baseline methods frequently falter.

Although Station 54777 presented greater forecasting challenges for all models, the proposed model still outperformed the baselines in most metrics, recording the lowest MAE (1.0139) and RMSE (1.3190), and a competitive R^2 of 0.6401. While the Informer model slightly outperformed in terms of R^2 (0.7376), it incurred higher errors in other metrics. From the temporal trend depicted in Figure 8, the proposed model achieves a better fit with the observed wind speed curve, particularly in capturing broader trend patterns and abrupt wind changes. Meanwhile, the baseline models exhibit noticeable lag or fail to adapt to rapid transitions, undermining their practical applicability in dynamic scenarios.

Across all four stations, the proposed model consistently outperforms the baseline methods in terms of MSE, MAE, RMSE, and R^2, demonstrating its superior accuracy and generalization ability. The improvements are particularly evident in stations with more complex or volatile wind patterns, where conventional models struggle to adapt to sudden changes. Visual comparisons in Figure 15-18 further reinforce these findings, showing that the proposed model is better aligned with the ground truth time series, especially in capturing both sharp transitions and smooth trend variations. These results collectively validate the effectiveness of the proposed approach in delivering robust and precise wind speed forecasting under diverse conditions.

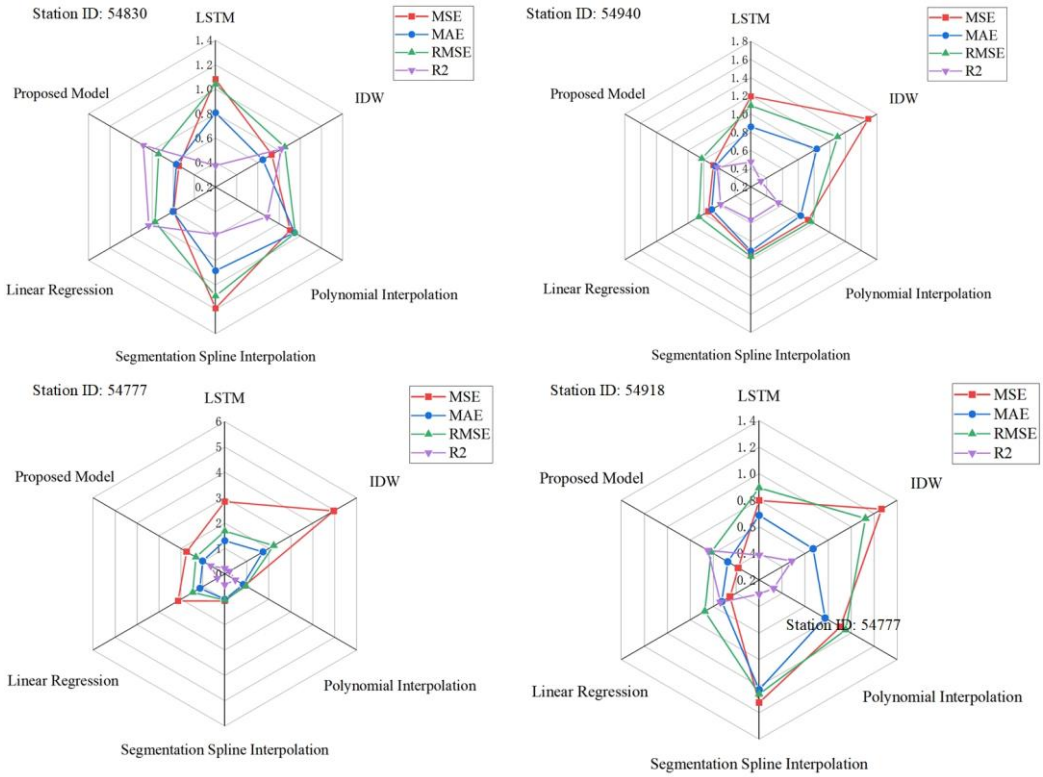

Fig. 6. Comparison of Error Metrics for Wind Speed Interpolation Results of Various Models

V. Conclusion

To address the issues of spatial unevenness, sparse observation points, and missing values in wind speed data over complex terrains, this study proposes a spatiotemporal wind speed interpolation method—the LSTM-IDW-GBRT ensemble model. This model innovatively integrates Long Short-Term Memory (LSTM) networks, the Inverse Distance Weighting (IDW) method, and Gradient Boosting Regression Trees (GBRT), which are respectively responsible for capturing temporal dependencies, spatial correlations, and nonlinear fitting capabilities of wind speed data. This integration enables accurate interpolation of missing wind speed values across multi-terrain regions. The model was systematically validated using real-world wind speed data from 12 meteorological stations in Shandong Province, and the experimental results demonstrate that the proposed method outperforms conventional and single-model approaches in terms of interpolation accuracy, stability, and adaptability. The main conclusions are summarized as follows.

Firstly, in terms of temporal modeling, the LSTM network effectively captures long-term trends and periodic patterns in wind speed series through its memory gate mechanism, and maintains strong predictive ability even during periods of high fluctuation. Secondly, the IDW method utilizes spatial information from neighboring stations to perform interpolation, demonstrating strong adaptability in regions with dense station distribution. Finally, the GBRT model nonlinearly integrates the outputs from the LSTM and IDW modules, effectively fusing temporal and spatial features, and thereby improving overall interpolation accuracy and compensating for the limitations of individual models in handling complex spatiotemporal dependencies.

In the experimental analysis, four representative target stations (A-54918, B-54830, C-54940, and D-54777) were selected, located across different terrain types including inland, mountainous, and coastal areas. A comprehensive comparison was conducted between the proposed model and several baseline methods, including LSTM, IDW, linear regression, polynomial interpolation, and segmented spline interpolation, using common error metrics such as MSE, MAE, RMSE, and R^2. The results indicate that the proposed ensemble model achieved the lowest error at all target stations, with average reductions in MSE, MAE, and RMSE of 56.4%, 32.55%, and 37.15%, respectively, significantly outperforming the comparison methods. The model exhibited the highest fitting accuracy in inland regions, and also maintained strong error suppression capabilities in complex coastal terrain, demonstrating excellent generalization and robustness. In addition, by evaluating the proposed model across four geographically distinct regions, we further validate its robustness and adaptability. The results show that the proposed approach consistently achieves the best performance under varying terrain conditions, outperforming both state-of-the-art neural architectures and ensemble-based models. Among the baseline methods, Informer demonstrates relatively stable and competitive performance across all sites, though it still falls short of the proposed model. In contrast, CNN-BiLSTM-AM and VMD-GRU-MC exhibit noticeably inferior accuracy and generalization capabilities, particularly in regions with more complex or rapidly changing wind dynamics. These findings suggest that the proposed model is better equipped to handle diverse environmental settings and spatial heterogeneity.

In conclusion, the proposed LSTM-IDW-GBRT ensemble model exhibits remarkable advantages in interpolation accuracy, structural rationality, and broad applicability. It provides a practical solution for various real-world applications such as wind resource assessment, wind farm site selection, and short-term wind power forecasting. Future work may extend this model to other geographical regions for wind speed imputation and forecasting, further enhancing its utility and decision-support capabilities within smart energy systems.

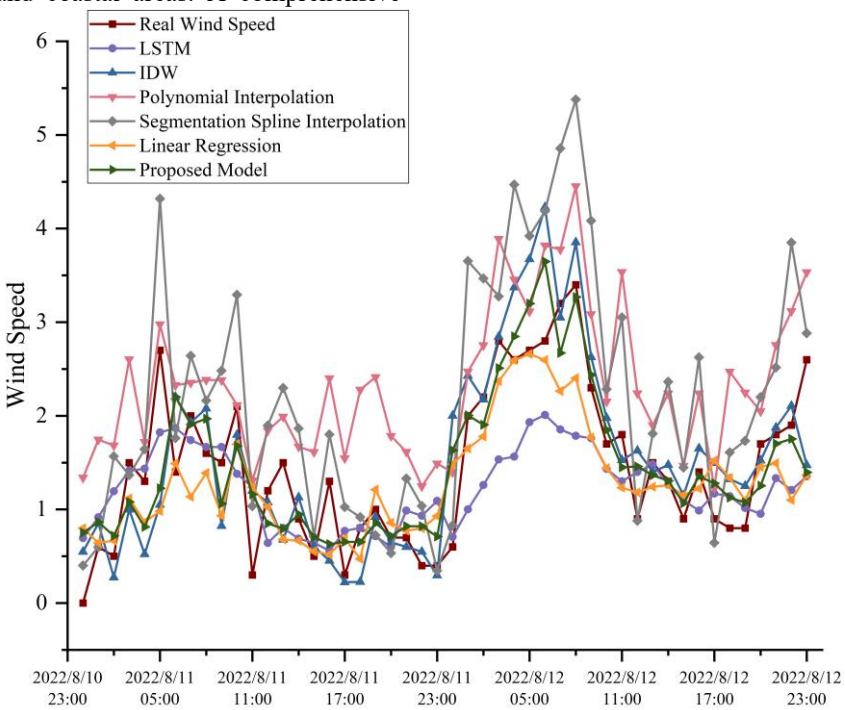

Fig. 7. Forecast results of Group A meteorological station 54918

TABLE II
TABLE OF PREDICTION ERROR RESULTS

Group	Station ID	Prediction Method	MSE	MAE	RMSE	R^2
A	54918	LSTM	0.7983	0.6872	0.8935	0.3863
		IDW	1.2662	0.6699	1.1253	0.4844
		Polynomial Interpolation	0.9072	0.7733	0.9525	0.3260
		Segmentation Spline Interpolation	1.1242	1.0271	1.0603	0.3070
		Linear Regression	0.4540	0.5260	0.6730	0.5373
		Proposed Model	0.3813	0.4730	0.6175	0.6451
B	54830	LSTM	1.0815	0.8075	1.0399	0.3755
		IDW	0.7306	0.6463	0.8548	0.8240
		Polynomial Interpolation	0.9059	0.9458	0.9518	0.6872
		Segmentation Spline Interpolation	1.1919	0.8845	1.0918	0.5884
		Linear Regression	0.5960	0.6020	0.7720	0.8312
		Proposed Model	0.5461	0.5717	0.7390	0.8820
C	54940	LSTM	1.1940	0.8607	1.0927	0.4740
		IDW	1.6920	1.0364	1.3008	0.3270
		Polynomial Interpolation	0.9265	0.8323	0.9625	0.5470
		Segmentation Spline Interpolation	0.9335	0.9050	0.9662	0.5610
		Linear Regression	0.7399	0.6978	0.8601	0.5840
		Proposed Model	0.6775	0.6527	0.8231	0.6270
D	54777	LSTM	2.8458	1.3068	1.6870	0.2161
		IDW	4.9619	1.7380	2.2275	0.2048
		Polynomial Interpolation	0.9138	0.8374	0.9559	0.4770
		Segmentation Spline Interpolation	1.0685	1.0013	1.0337	0.4290
		Linear Regression	2.1290	1.1320	1.4590	0.3463
		Proposed Model	1.7399	1.0139	1.3190	0.6401

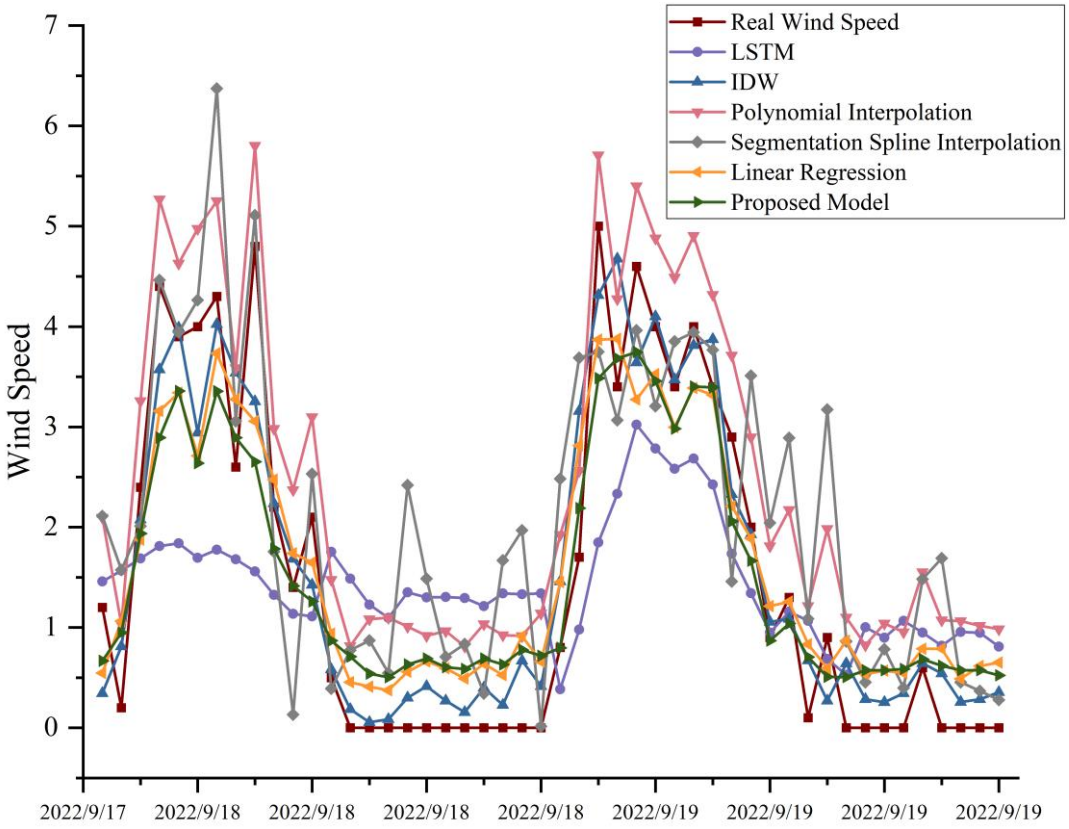

Fig. 8. Forecast results of Group B meteorological station 54830

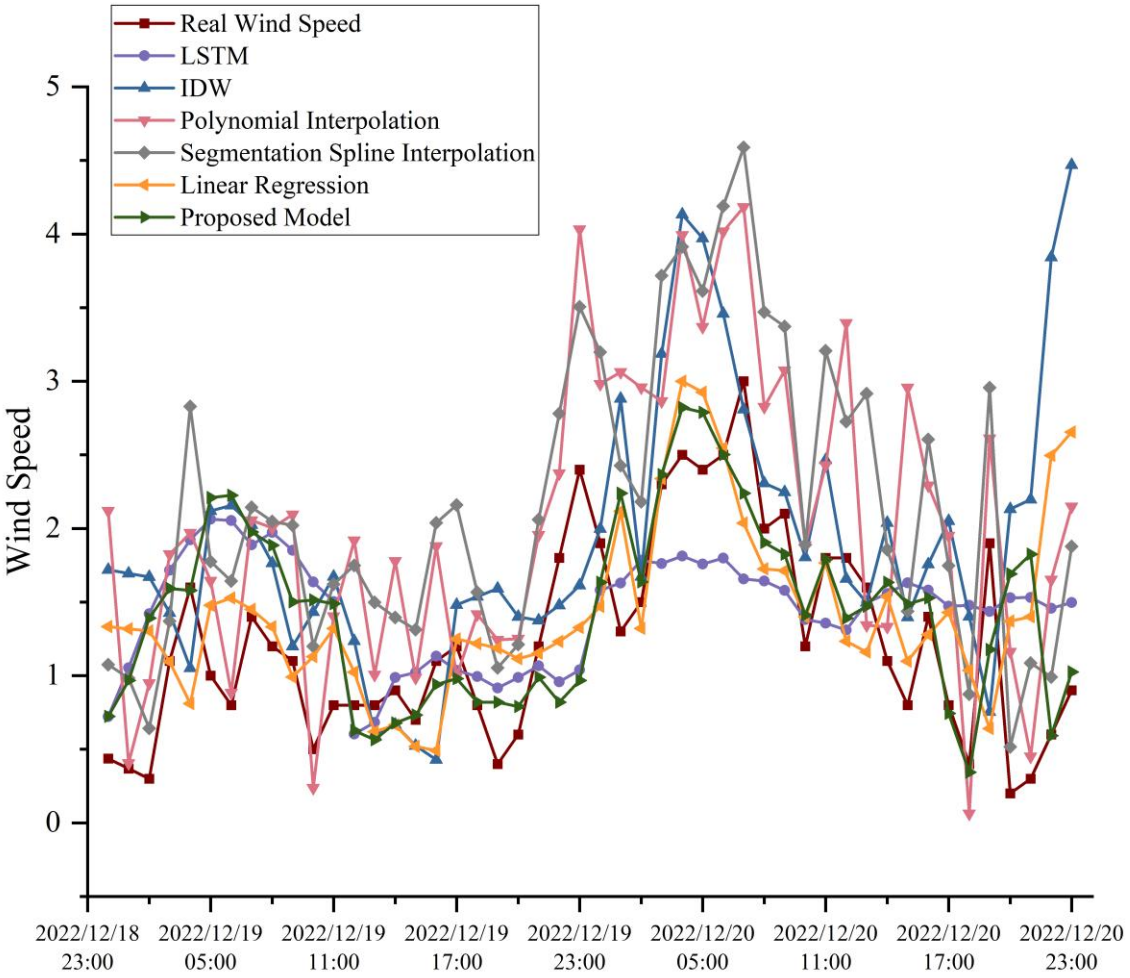

Fig. 9. Forecast results of Group C meteorological station 54940

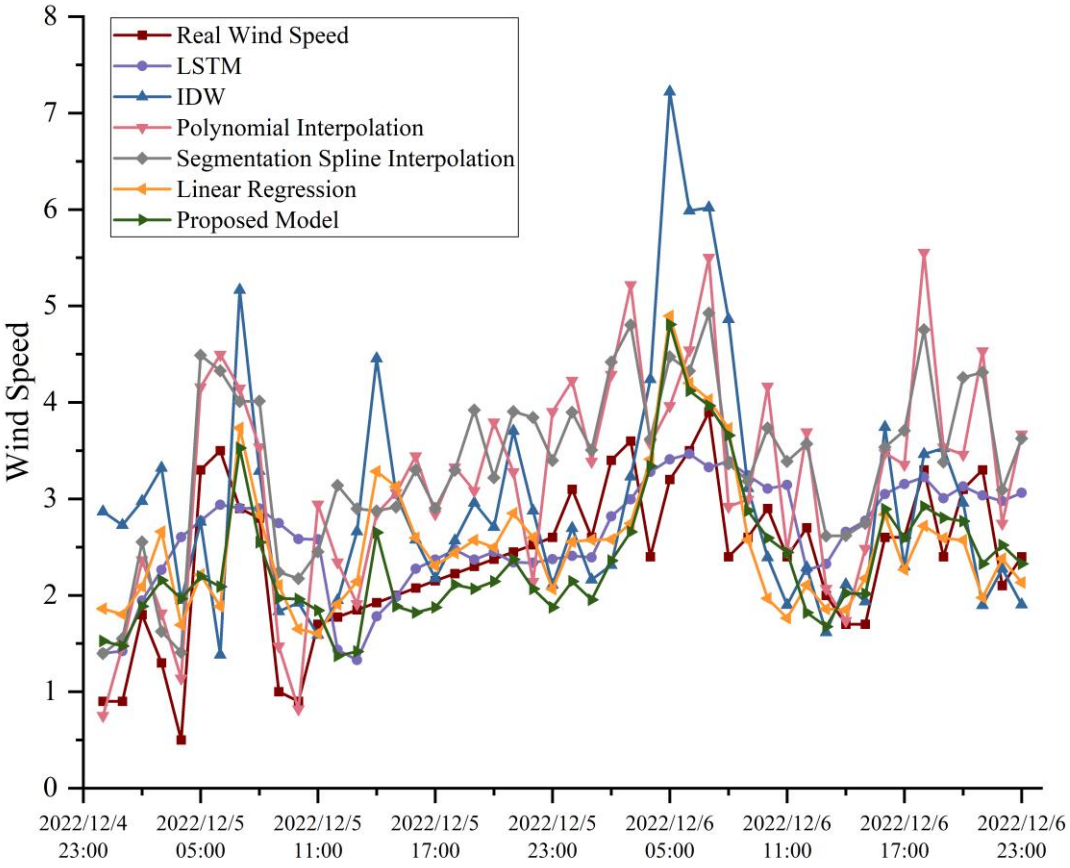

Fig. 10. Forecast results of Group D meteorological station 54777

TABLE III
RESULTS WITH MODERN MODELS

Group	Station ID	Prediction Method	MSE	MAE	RMSE	R^2
A	54918	Informer	0.7983	0.675	0.8889	0.3863
		CNN-BiLSTM-AM	0.9112	0.77	0.9546	0.4844
		VMD-GRU-MC	0.9552	0.756	0.9774	0.3260
		The Proposed Model	0.3813	0.473	0.6175	0.6451
B	54830	Informer	0.76	0.706	0.8718	0.7376
		CNN-BiLSTM-AM	0.87	0.773	0.9327	0.6996
		VMD-GRU-MC	0.95	0.744	0.9747	0.6719
		The Proposed Model	0.5461	0.5717	0.739	0.882
C	54940	Informer	1.2816	1.2816	1.2816	0.603
		CNN-BiLSTM-AM	1.4396	0.98	1.1998	0.578
		VMD-GRU-MC	1.6903	0.977	1.3001	0.432
		The Proposed Model	0.6775	0.6527	0.8231	0.6775
D	54777	Informer	2.2	1.244	1.48	0.7376
		CNN-BiLSTM-AM	2.39	1.174	1.5	0.6996
		VMD-GRU-MC	2.6	1.233	1.6	0.6719
		The Proposed Model	1.7399	1.0139	1.3190	0.6401

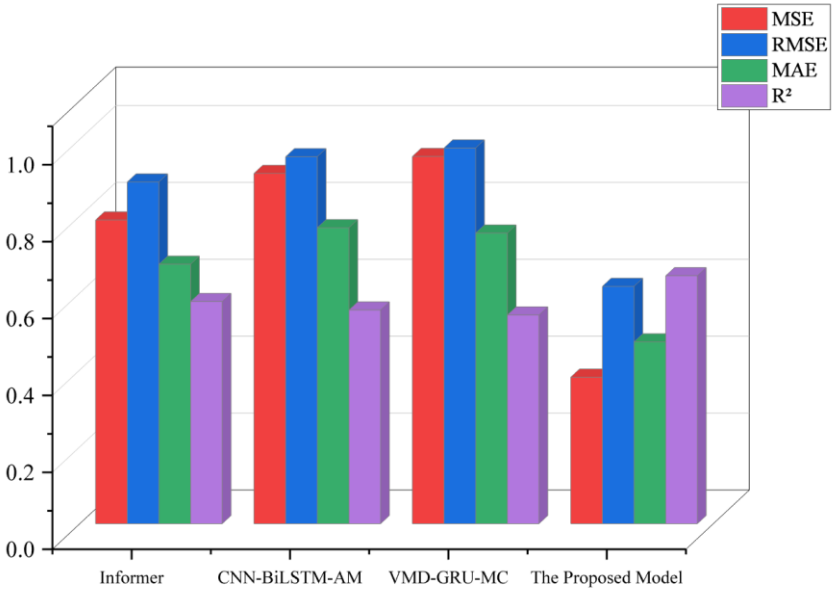

Fig. 11. Evaluation Metrics of Results compared with Modern Models of station 54918

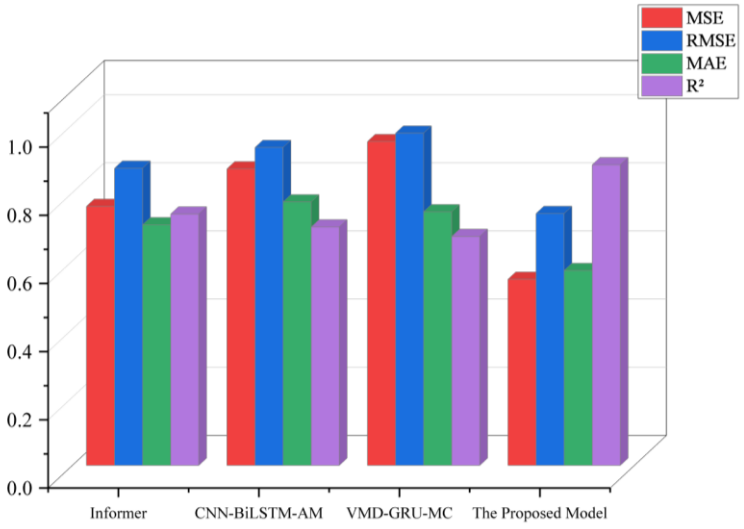

Fig. 12. Evaluation Metrics of Results compared with Modern Models of station 54830

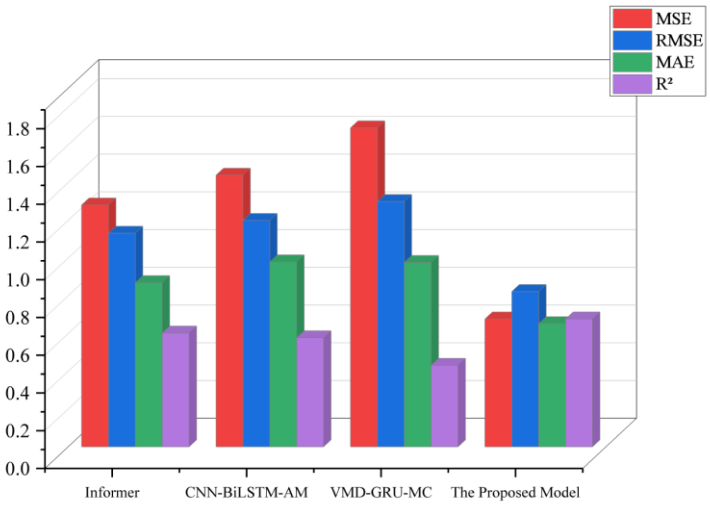

Fig. 13. Evaluation Metrics of Results compared with Modern Models of station 54940

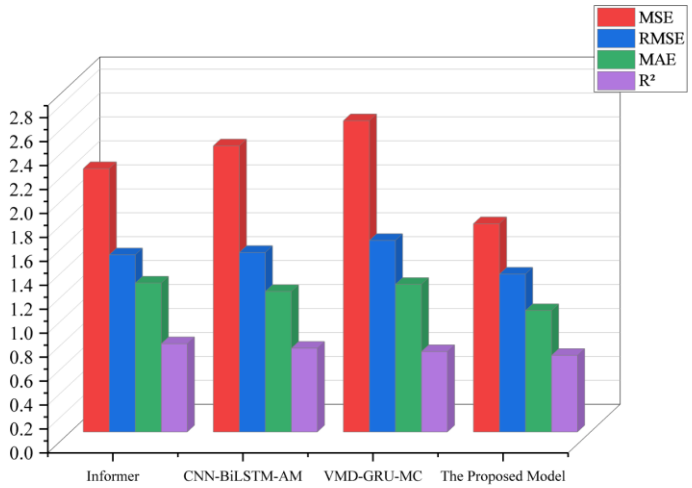

Fig. 14. Evaluation Metrics of Results compared with Modern Models of station 54777

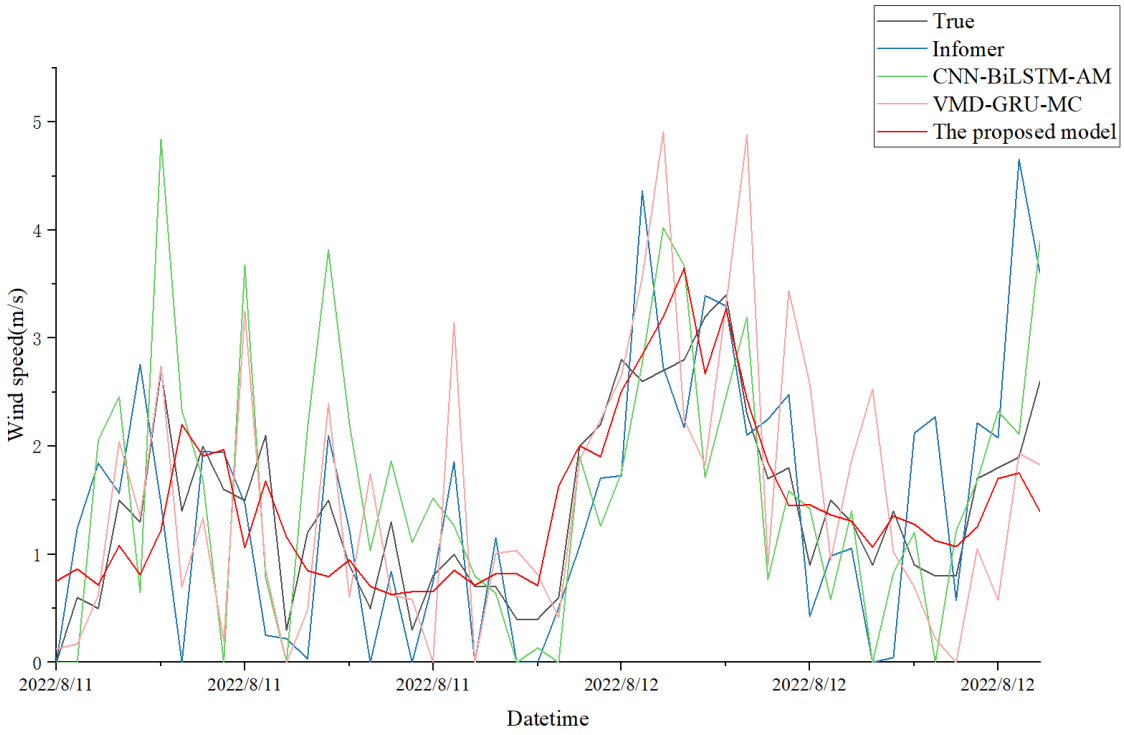

Fig. 15. Results compared with Modern Models of Group A meteorological station 54918

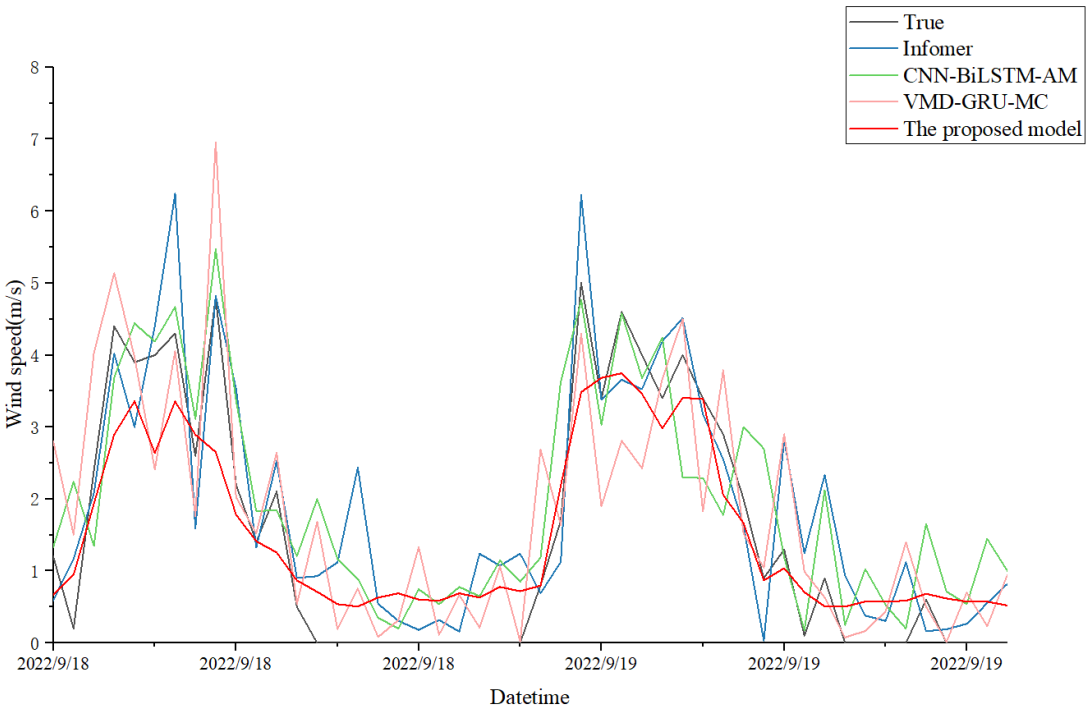

Fig. 16. Results compared with Modern Models of Group B meteorological station 54830

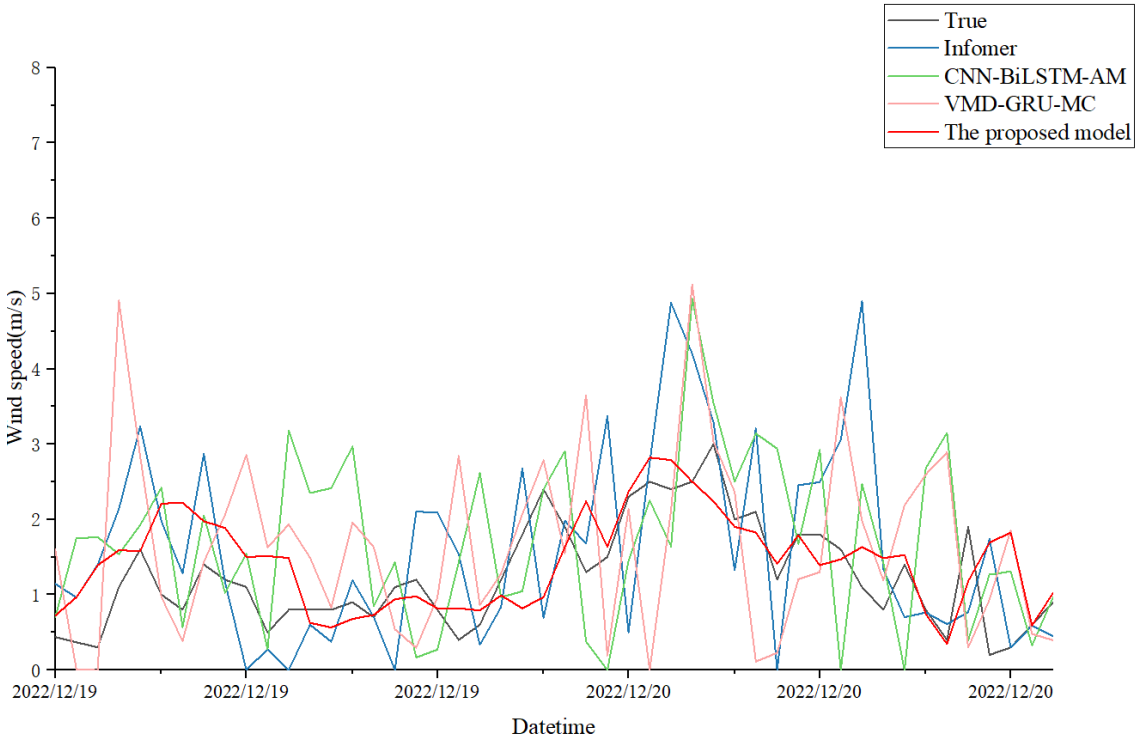

Fig. 17. Results compared with Modern Models of Group C meteorological station 54940

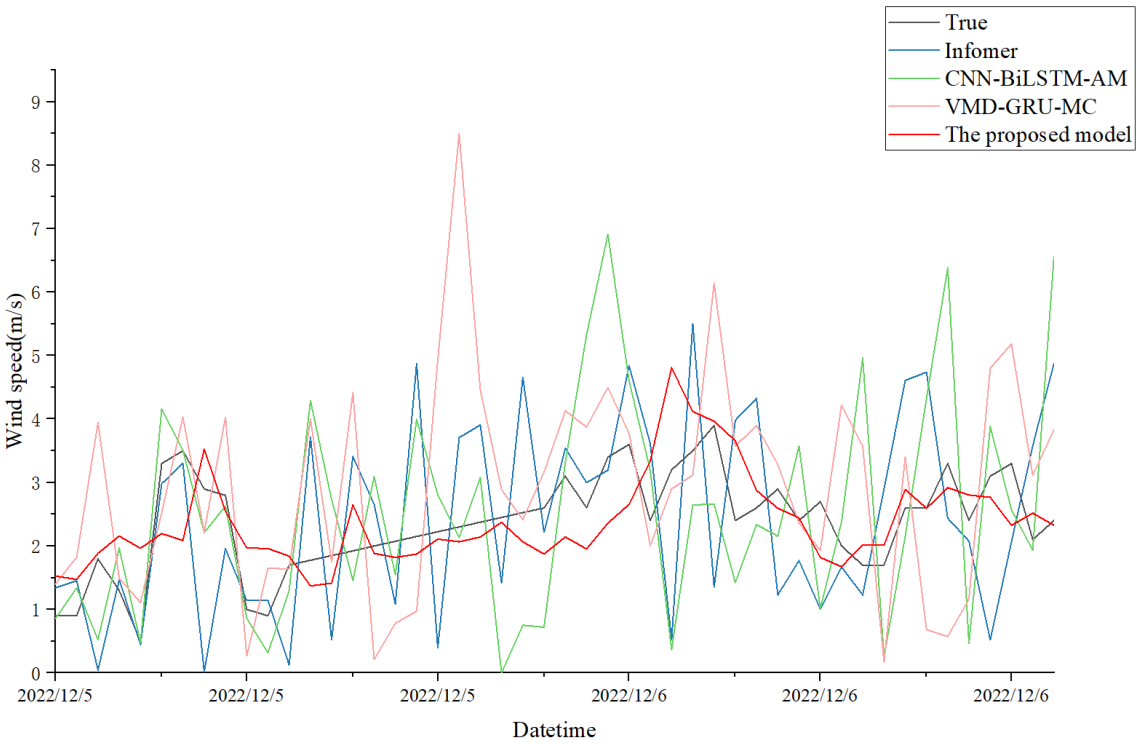

Fig. 18. Results compared with Modern Models of Group D meteorological station 54777

REFERENCES

[1] IRENA (2024), Renewable energy statistics 2024, International Renewable Energy Agency, Abu Dhabi.

[2] IEA (2024), Renewables 2024, IEA, Paris https://www.iea.org/reports/renewables-2024, Licence: CC BY 4.0

[3] China drives world renewables capacity addition in 2023, (n.d.). https://english.www.gov.cn/news/topnews/202401/13/content_WS65a22a99c6d0868f4e8e30aa.html (accessed April 10, 2025).

[4] GWEC Global Wind Report - WindEnergy Hamburg, (n.d.). https://www.windenergyhamburg.com/news-themen/reports-market-updates/gwec-global-wind-report/ (accessed April 10, 2025).

[5] Notice of the State Council on Issuing the Action Plan for Carbon Peaking Before 2030, Environmental Monitoring, Protection, and Governance, China Government Network, (n.d.). Available at: https://www.gov.cn/zhengce/content/2021-10/26/content_5644984.htm (accessed April 10, 2025).

[6] Summary of China's energy and power sector statistics in 2023 - China Energy Transformation Program, (n.d.). https://www.cet.energy/2024/03/18/summary-of-chinas-energy-and-power-sector-statistics-in-2023/ (accessed April 10, 2025).

[7] Wind Energy and the Electric Power System, in: Wind Energy Handbook, John Wiley & Sons, Ltd, 2011: pp. 565–612. https://doi.org/10.1002/9781119992714.ch10.

[8] W. Mackenzie, China leads global wind turbine manufacturers' market share in 2023, (2024). https://www.woodmac.com/press-releases/2024-press-releases/global-wind-oem-marketshare/ (accessed April 10, 2025).

[9] G. Li, J. Shi, On comparing three artificial neural networks for wind speed forecasting, Applied Energy 87 (2010) 2313–2320. https://doi.org/10.1016/j.apenergy.2009.12.013.

[10] Wind Energy Applications, in: Wind Energy Explained, John Wiley & Sons, Ltd, 2009: pp. 449–504. https://doi.org/10.1002/9781119994367.ch10.

[11] Shu, HU Ju, LV Chen, et al. A Hindcast Dataset on Domestic Wind Energy Resource in Fine Resolution via Observation-nudging Climate Four-dimension Data Assimilation Method[J]. Power System Technology,2019,43(11):3859-3866.

[12] YANG Guangyan, WU Xi, ZHOU Hai. Effect analysis of WRF on wind speed prediction at the coast wind power Station of Fujian province[J]. Journal of the Meteorological Sciences,2014,34(05):530-535.

[13] MENG Changbo, MA Jiming, YANG Jianshe. Study on the improvement of wind resources assessment method[J]. Journal of Hydroelectric Engineering,2010,29(6):237-242.

[14] J.A. Carta, S. Velázquez, P. Cabrera, A review of measure-correlate-predict (MCP) methods used to estimate long-term wind characteristics at a target site, Renewable and Sustainable Energy Reviews 27 (2013) 362–400. https://doi.org/10.1016/j.rser.2013.07.004.

[15] Windographer | Wind Data Analytics and Visualization Solution, UL Solutions (n.d.). https://www.ul.com/software/windographer-wind-data-analytics-and-visualization-solution (accessed April 10, 2025).

[16] H. Liu, H. Tian, Y. Li, Comparison of two new ARIMA-ANN and ARIMA-Kalman hybrid methods for wind speed prediction, Applied Energy 98 (2012) 415–424. https://doi.org/10.1016/j.apenergy.2012.04.001.

[17] W. Xinxin, S. Xiaopan, A. Xueyi, L. Shijia, Short-term wind speed forecasting based on a hybrid model of ICEEMDAN, MFE, LSTM and informer, PLOS ONE 18 (2023) e0289161. https://doi.org/10.1371/journal.pone.0289161.

[18] H. Hu, Y. Li, X. Zhang, M. Fang, A novel hybrid model for short-term prediction of wind speed, Pattern Recognition 127 (2022) 108623. https://doi.org/10.1016/j.patcog.2022.108623.

[19] F. Li, G. Ren, J. Lee, Multi-step wind speed prediction based on turbulence intensity and hybrid deep neural networks, Energy Conversion and Management 186 (2019) 306–322. https://doi.org/10.1016/j.enconman.2019.02.045.

[20] P. Jiang, Y. Wang, J. Wang, Short-term wind speed forecasting using a hybrid model, Energy 119 (2017) 561–577. https://doi.org/10.1016/j.energy.2016.10.040.

[21] Jerome H. Friedman "Greedy function approximation: A gradient boosting machine.," The Annals of Statistics, Ann. Statist. 29(5), 1189-1232, (October 2001)

[22] R.A. Sobolewski, M. Tchakorom, R. Couturier, Gradient boosting-based approach for short- and medium-term wind turbine output power prediction, Renewable Energy 203 (2023) 142–160. https://doi.org/10.1016/j.renene.2022.12.040.

[23] J.M. Lima, A.K. Guetter, S.R. Freitas, J. Panetta, J.G.Z. de Mattos, A Meteorological–Statistic Model for Short-Term Wind Power Forecasting, J Control Autom Electr Syst 28 (2017) 679–691. https://doi.org/10.1007/s40313-017-0329-8.

[24] WANG Na, ZHOU Youqing, SHAO Xia. Wind Energy Resource Assessment of Wind Farm Based on Hybrid Neural Network[J]. Transactions of china electrotechnical society, 2015,30(14):370-376.

[25] S. Lin, S. Wang, X. Xu, R. Li, P. Shi, GAOformer: An adaptive spatiotemporal feature fusion transformer utilizing GAT and optimizable graph matrixes for offshore wind speed prediction, Energy 292 (2024) 130404. https://doi.org/10.1016/j.energy.2024.130404.

[26] R. Baïle, J.-F. Muzy, Leveraging data from nearby stations to improve short-term wind speed forecasts, Energy 263 (2023) 125644. https://doi.org/10.1016/j.energy.2022.125644.

[27] K. Reinhardt, C. Samimi, Comparison of different wind data interpolation methods for a region with complex terrain in Central Asia, Clim Dyn 51 (2018) 3635–3652. https://doi.org/10.1007/s00382-018-4101-y.

[28] Palomino, I., and F. Martín, 1995: A Simple Method for Spatial Interpolation of the Wind in Complex Terrain. *J. Appl. Meteor. Climatol.*, 34, 1678–1693, https://doi.org/10.1175/1520-0450-34.7.1678.

[29] Alymov, Elena,Kudryavtsev, Oleg, "Artificial Neural Networks and Wiener-Hopf Factorization," IAENG International Journal of Computer Science, vol. 51, no.8, pp1182-1194, 2024

[30] Fihri, Abdelkader Fassi;Hajji, Tarik;El Hassani, Ibtissam;Masrour, Tawfik,"Particle Swarm Algorithm Setting using Deep Reinforcement Learning in the Artificial Neural Network Optimization Learning Process"IAENG International Journal of Computer Science, Vol.51, No.8, pp1195-1208,2024