



A novel model for wind speed point prediction and quantifying uncertainty in wind farms

Chunsheng Yu¹ · Yutao Ma²

Received: 28 May 2024 / Accepted: 17 November 2024

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract

Wind speed prediction plays a critical role in the operation and maintenance of wind farms. This paper introduces a wind speed point and interval prediction model, named CEEMDAN-SE-BiLSTM-MK, which is based on decomposition, reconstruction, prediction, and quantification, specifically designed for wind speed prediction in wind farms. Initially, the wind speed series is decomposed into multiple modal components, and components with similar complexity are reconstructed using the sample entropy (SE) and the K-Means method. Subsequently, the reconstructed components are forecasted using the Bi-directional Long Short Term Memory (BiLSTM) model, and the predictions are aggregated to generate the final wind speed point prediction. Addressing the issue of lag in point prediction affecting interval prediction, the model integrates the Markov Chain Monte Carlo (MCMC) and Kernel Density Estimation (KDE) methods. Furthermore, the uncertainty of component prediction is quantified and assessed for its influence on the final prediction outcomes. The model's performance is validated using data from two wind farms in China, demonstrating its superior robustness compared to other models. The proposed approach significantly enhances interval prediction accuracy, revealing distinct sources of uncertainty in the wind speed components for the two wind farms.

Keywords Wind speed prediction · CEEMDAN · MK method · Uncertainty quantification

1 Introduction

1.1 Background

Wind energy is a promising and sustainable renewable energy source that has gained increasing attention in recent years due to its clean, stable, and cost-effective characteristics. According to the Global Wind Energy Report 2023 published by the Global Wind Energy Council [1], wind energy has become one of the fastest-growing energy sources worldwide, with mature technology, high cost-effectiveness, and resource efficiency. The report also predicts that the global installed capacity of onshore wind power will increase significantly to 68.8 GW in 2022, indicating a bright future for the

wind energy industry. Accurate wind speed prediction is crucial for various aspects of wind energy utilization, including wind farm planning, grid connection, and power scheduling. It serves as the foundation for ensuring the power system's safe, reliable, and cost-effective operation [2, 3].

1.2 Literature review

In recent years, the landscape of wind speed prediction has been enriched by the adoption of diverse methodologies, notably encompassing physical [4–6], statistical [7–10], and artificial intelligence (AI) models [11–14]. Among these, AI models have garnered significant attention due to their exceptional predictive prowess, adeptness in capturing intricate nonlinear relationships, and advanced feature extraction capabilities. Notably, the fusion of decomposition techniques with AI has emerged as a vibrant research frontier, exemplified by Liu et al. [15] innovative approach that leverages Empirical Mode Decomposition (EMD) coupled with Long Short-Term Memory (LSTM) networks for high-frequency sub-sequences and ARIMA for low-frequency components

✉ Chunsheng Yu
csy05050513@163.com

¹ College of Mathematics and Statistics, Northwest Normal University, Lanzhou 730070, China

² College of Resources and Environment, Lanzhou University, Lanzhou 730000, China

and residuals, yielding enhanced prediction accuracy and stability. Similarly, Lin et al. [16] introduced a hybrid framework integrating EMD and phase space reconstruction, showcasing high precision in capturing wind speed dynamics and emphasizing the value of decomposition in boosting prediction accuracy. However, a notable limitation of these studies lies in their exclusive focus on point predictions, which neglect crucial uncertainty information. As wind farm operations confront heightened uncertainty, point predictions alone become insufficient, posing challenges for effective decision-making [17]. Thus, integrating uncertainty quantification becomes imperative to augment the practical utility of predictions.

Prediction intervals, characterized by upper and lower bounds, serve as potent vehicles for conveying uncertainty, offering insights into the likely range of future values and prediction accuracy. For wind farm stakeholders, range predictions are not only more reliable but also richer in information compared to point predictions [17]. Existing strategies for generating prediction intervals can be broadly classified into two categories: direct interval prediction and prediction error-based approaches. The former category encompasses methods such as Han et al. [18] multivariate linear regression utilizing WRF variables and Li et al. [19] and Wang et al. [20] multi-objective optimization frameworks, while Naik et al. [21] innovatively fused Variational Mode Decomposition (VMD) with low-rank Multi-Kernel Ridge Regression (MKRR) for direct interval construction. The latter category leverages prediction errors, with Sreekumar et al. [22] optimizing joint probability distributions using MCMC and C-vine copulas, while Shi et al. [23] proposed an MCMC-based approach within a CLM framework to generate narrow, symmetric confidence intervals. Despite their merits, many of these methods rely on assumptions about the shape of underlying distributions, which may not always align with reality. Nonparametric probability density estimation, particularly Kernel Density Estimation (KDE), emerges as a promising alternative, offering flexibility and robustness without imposing distributional constraints [24, 25]. Wahbah et al. [26] demonstrated the superiority of KDE based on unbiased cross-validation, while Fan et al. [27] integrated LSTM with KDE, enhancing the model's adaptability to diverse error distribution patterns.

In summary, while substantial progress has been made in wind speed prediction, there remains a need to bridge the gap between point predictions and comprehensive uncertainty quantification. The integration of advanced decomposition techniques, AI models, and nonparametric probability density estimation offers a promising avenue to address this challenge, motivating the development of our proposed model, which aims to provide more reliable and informative range predictions for wind farm operations.

This paper presents an integrated framework for point and interval prediction of wind speeds in wind farms. The main contributions are as follows:

1. Proposed a novel wind speed point and interval prediction model, CEEMDAN-SE-BiLSTM-MK, which addresses the challenges of wind speed prediction in wind farms.
2. Integrated various methods, including CEEMDAN decomposition and reconstruction, BiLSTM model, and MK methods, to improve the prediction performance and robustness of the model.
3. Quantified and evaluated the uncertainty of component prediction in the model, providing more accurate and reliable prediction results for wind speed prediction. This contributes to optimizing the operation and maintenance of wind farms.

The paper is organized as follows. Section 2 outlines the methodology and model structure proposed in this study. Section 3 presents the experimental results. Finally, Sect. 4 concludes the paper and provides suggestions for future research.

2 Methodology

2.1 CEEMDAN

By iteratively adding noise and extracting the IMF, the CEEMDAN [28] algorithm effectively separates the mixing components and provides a more accurate decomposition of the original signal. The CEEMDAN decomposition process is shown in Fig. 1.

1. Adding Gaussian white noise $\omega^i(t)$ ($i = 1, 2, 3, \dots, n$) and signal-to-noise ratio to the signal to be decomposed produces a new signal $x^i(t)$. The new signal is then subjected to EMD decomposition to obtain the first-order intrinsic mode function (IMF):

$$x^i(t) = x(t) + \varepsilon_0 \omega^i(t), \quad i = 1, 2, 3, \dots, n \quad (1)$$

$$E(x^i(t)) = \text{IMF}_1^i(t) + r^i \quad (2)$$

, where t represents different time points and the number of times white noise is added.

2. The overall average of the n -mode components generated is used to obtain the first IMF of CEEMDAN decomposition, as well as the calculation of the residual after

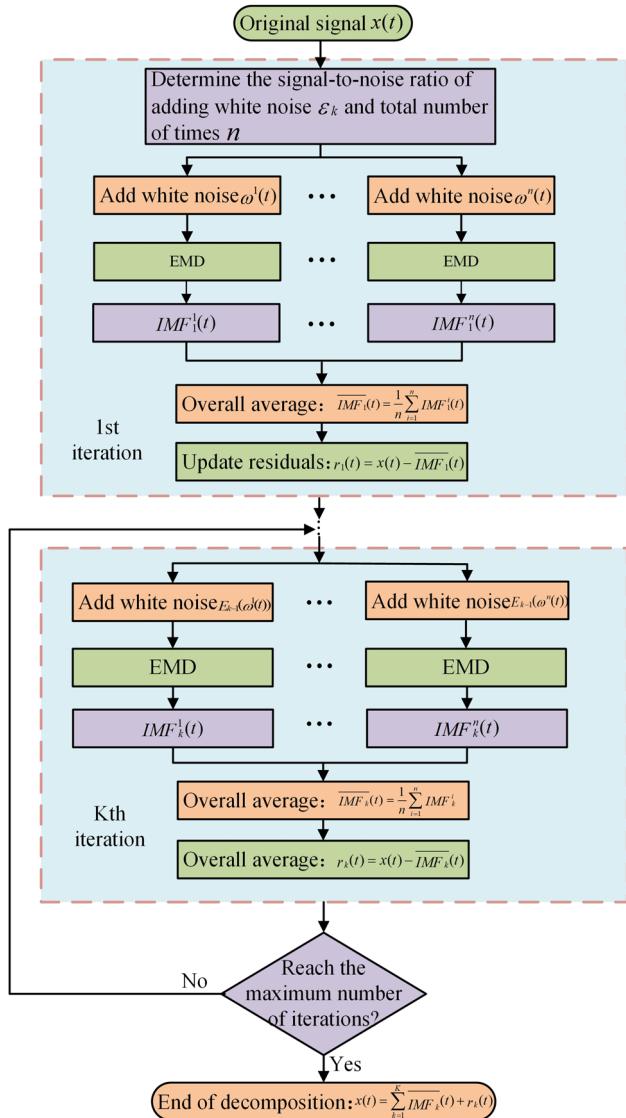


Fig. 1 CEEMDAN decomposition process

removing the first mode component:

$$\overline{\text{IMF}_1}(t) = \frac{1}{n} \sum_{i=1}^n \text{IMF}_1^i(t) \quad (3)$$

$$r_1(t) = x(t) - \overline{\text{IMF}_1}(t) \quad (4)$$

3. The IMF of the first white noise mode decomposition with signal-to-noise ratio ε_1 is added to $r_1(t)$ obtain the second intrinsic mode function of CEEMDAN decomposition, and the residual after removing the first mode component is calculated:

$$\overline{\text{IMF}_2}(t) = \frac{1}{n} \sum_{i=1}^n E_1(r_1(t) + \varepsilon_1 E_1(\omega^i(t))) \quad (5)$$

$$r_2(t) = r_1(t) - \overline{\text{IMF}_2}(t) \quad (6)$$

4. Repeat the iteration of the above steps until the residual signal obtained is a monotonic function and cannot be decomposed further, terminate the CEEMDAN algorithm, at which time K eigenmode components are obtained, and then the original signal is decomposed as:

$$x(t) = \sum_{k=1}^K \overline{\text{IMF}_k}(t) + r_K(t) \quad (7)$$

2.2 Sample entropy

Sample entropy is a metric used to quantify the complexity of a time series. It measures the probability of generating new patterns within the signal. A smaller sample entropy suggests higher self-similarity in the sequence and a lower likelihood of generating new patterns. Sample entropy is also used in this study to evaluate the complexity between component sequences and guide the process of sequence reconstruction. By analyzing the sample entropy, we can determine the level of complexity and generate more accurate reconstructed sequences.

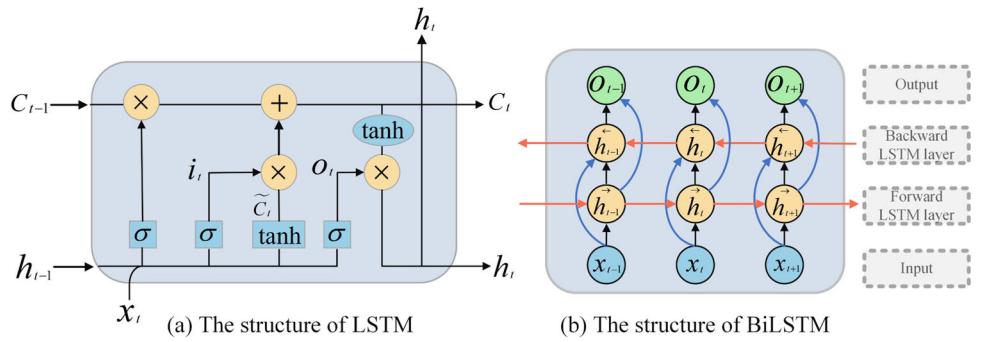
2.3 BiLSTM

BiLSTM is a powerful deep neural network that extends LSTM. A complete BiLSTM network includes an input layer, a forward LSTM layer, a reverse LSTM layer, and an output layer. By using both the forward and reverse LSTM layers, a BiLSTM network can capture information from both past and future contexts, making it useful for tasks such as series prediction [29], sentiment analysis [30], machine translation, and many others. Figure 2 shows the LSTM and BiLSTM structures.

2.4 MCMC

Based on the Bayesian principle, we can infer and construct the posterior distribution function of the model using the observed data, combining the likelihood distribution and the prior distribution. However, this posterior distribution often has a complex, nonstandard form, and it becomes very difficult to sample directly from it or to compute its properties (e.g., mean, variance, etc.). MCMC (Markov Chain Monte Carlo) is a suitable method for random sampling from complex distributions. The MCMC method iteratively generates a series of samples that follow the target distribution, and when the generated Markov chain achieves convergence, the resulting Markov chain is used to generate a distribution that follows the target distribution. Obtained smooth

Fig. 2 Bidirectional structure composed of LSTMs



distribution is used as an estimate of the desired joint posterior distribution. The basic steps of MCMC based on the Metropolis–Hastings algorithm are as follows:

1. *Initialization*: Select an initial sample $x^{(0)}$ as the starting point of the Markov chain.
2. *Generate a candidate sample*: Generate a candidate sample $x^{(t)}$ based on the current sample.
3. *Calculate the acceptance rate*: Sample from a uniform distribution, where the random number u follows a uniform distribution in the range $[0,1]$. Calculate the probability of accepting the candidate sample, which is the acceptance rate σ_{ij} .

$$\begin{aligned}\sigma_{ij} &= \min \left\{ \frac{p(x^*|y)}{p(x^{(t-1)}|y)}, 1 \right\} \\ &= \min \left\{ \frac{p(d|x^*) p(x^*)}{p(d|x^{(t-1)}) p(x^{(t-1)})}, 1 \right\}\end{aligned}\quad (8)$$

4. *Decide whether to accept the candidate sample*: If the candidate sample is accepted, update the current sample to the candidate sample; otherwise, keep the current sample unchanged.

$$x^{(t)} = \begin{cases} x^*, & \sigma_{ij} > u; \\ x^{(t-1)}, & \sigma_{ij} \leq u. \end{cases}\quad (9)$$

5. Repeat steps 2–4 until the predetermined number of iterations is reached, generating a series of samples that follow the predictive error distribution. Through these randomly sampled samples from the stationary distribution, the mean m and standard deviation v can be estimated, obtaining parameter estimates for the predictive error distribution.
6. Calculate the lower and upper bounds of the prediction interval using the confidence level $1 - \alpha$ and the inverse function z of the cumulative distribution function of the

normal distribution:

$$\begin{cases} \text{LB_MCMC} = S_p - z * v \\ \text{UB_MCMC} = S_p + z * v \end{cases}\quad (10)$$

, where S_p represents the predicted wind speed value.

In this paper, the MCMC method is chosen over the direct method for obtaining distribution parameters and prediction intervals because of its many advantages. For example, MCMC is flexible in incorporating complex priors, providing complete posterior distributions, and rigorously propagating uncertainty. It enables a more thorough exploration of the parameter space and ensures robustness to local optima. In addition, MCMC captures the full range of predictive uncertainties that direct methods may miss. Its generalization to complex models and rigorous convergence diagnostics make it an excellent choice for studies seeking novelty and reliability.

2.5 KDE

Kernel Density Estimation (KDE) represents a nonparametric, data-driven approach that treats individual data points as minute probability masses, subsequently smoothing these masses via kernel functions. KDE provides a continuous and adaptable estimation of probability density, outperforming traditional histogram methods in capturing intricate data details and shapes. The kernel density function is given by the following formula:

$$f(e) = \frac{1}{Nh} \sum_{i=1}^N K\left(\frac{e - e_i}{h}\right)\quad (11)$$

, where N is the total number of samples, $K(\cdot)$ is the kernel function, e_i is the i th sample value of the prediction error. The selection of the kernel function $K(\cdot)$ and the bandwidth h is crucial for kernel density estimation.

The selection of kernel functions, which are inherently smooth and symmetric, underpins local density estimates

surrounding data points. Gaussian kernels are often preferred due to their advantageous mathematical properties and computational efficiency. While the choice of kernel influences each point's contribution, its impact on accuracy is generally overshadowed by the bandwidth parameter. Narrow bandwidth settings risk overfitting, while overly broad bandwidths can lead to underfitting. Thus, the careful selection of bandwidth is paramount for striking a delicate balance between estimation accuracy and smoothness. A larger bandwidth promotes smoother density curves but may mask subtle data features, whereas a smaller bandwidth enhances local nuances but may introduce instability. Visualizing KDE curves across various bandwidths offers a powerful tool for identifying the optimal bandwidth, thereby achieving an optimal equilibrium between fidelity to the data and smoothness. The most commonly used Gaussian kernel function is expressed as:

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad (12)$$

The bandwidth selection method is based on the standard deviation of the data and the sample size to achieve a balance between smoothing and estimation accuracy. The KDE method PI is calculated using the following formula:

$$\begin{cases} LB_KDE = S_p + F(\alpha/2) \\ UB_KDE = S_p + F(1 - \alpha/2) \end{cases} \quad (13)$$

where $F(\cdot)$ is the cumulative probability distribution function, and $1 - \alpha$ is the confidence level.

2.6 MK

In order to make the interval results more generalizable and avoid instability and large biases in single interval calculation methods, this paper integrates two easily implementable basic methods, MCMC and KDE, to achieve high coverage and narrow width interval performance. The main idea is to weigh and add the upper and lower boundary distributions calculated by both methods:

$$\begin{cases} LB_MK = \omega * LB_MCMC + (1 - \omega) * LB_KDE \\ UB_MK = \omega * UB_MCMC + (1 - \omega) * UB_KDE \end{cases} \quad (14)$$

where ω is the weighting coefficient.

2.7 Evaluation metrics

To evaluate the performance of prediction intervals (PI) and select appropriate PI, this study uses the following evaluation metrics: prediction interval coverage probability (PICP),

prediction interval average normalized width (PIANW), and a comprehensive index F -value [18].

- PICP measures the probability that the true wind speed values are covered by the prediction intervals.

$$PICP = \frac{1}{N} \sum_{i=1}^N c_i, \quad c_i = \begin{cases} 1, & S_i \in [LB_i, UB_i] \\ 0, & S_i \notin [LB_i, UB_i] \end{cases} \quad (15)$$

where N is the number of samples, S_i represents the actual wind speed values, and c_i records whether the i th actual wind speed value is covered by the interval. A higher PICP indicates better performance of the PI.

- Narrow PI provide more valuable information compared to wider PI. Therefore, PIANW is used to measure the narrowness of the intervals:

$$PIANW = \frac{1}{NR} \sum_{i=1}^N (UB_i - LB_i) \quad (16)$$

where R represents the range of the target values. A smaller PIANW indicates better performance of the PI.

- Generally, a higher PICP is often accompanied by a larger PIANW, and a smaller PIANW may result in a smaller PICP. To balance the trade-off between PICP and PIANW, a comprehensive index F -value, is used to consider both metrics. The calculation formula is as follows:

$$F-value = \frac{2 \times PICP \times \frac{1}{PIANW}}{PICP + \frac{1}{PIANW}} \quad (17)$$

A higher F -value indicates better overall performance of the PI.

2.8 CEEMDAN-SE-BiLSTM-MK model

Figure 3 shows the main steps of the proposed model as follows:

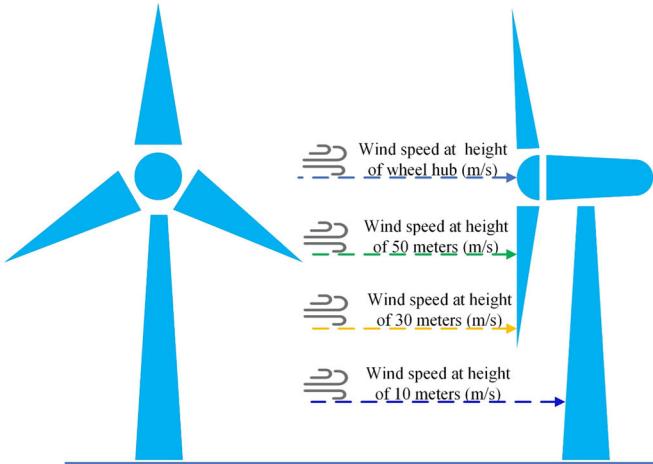
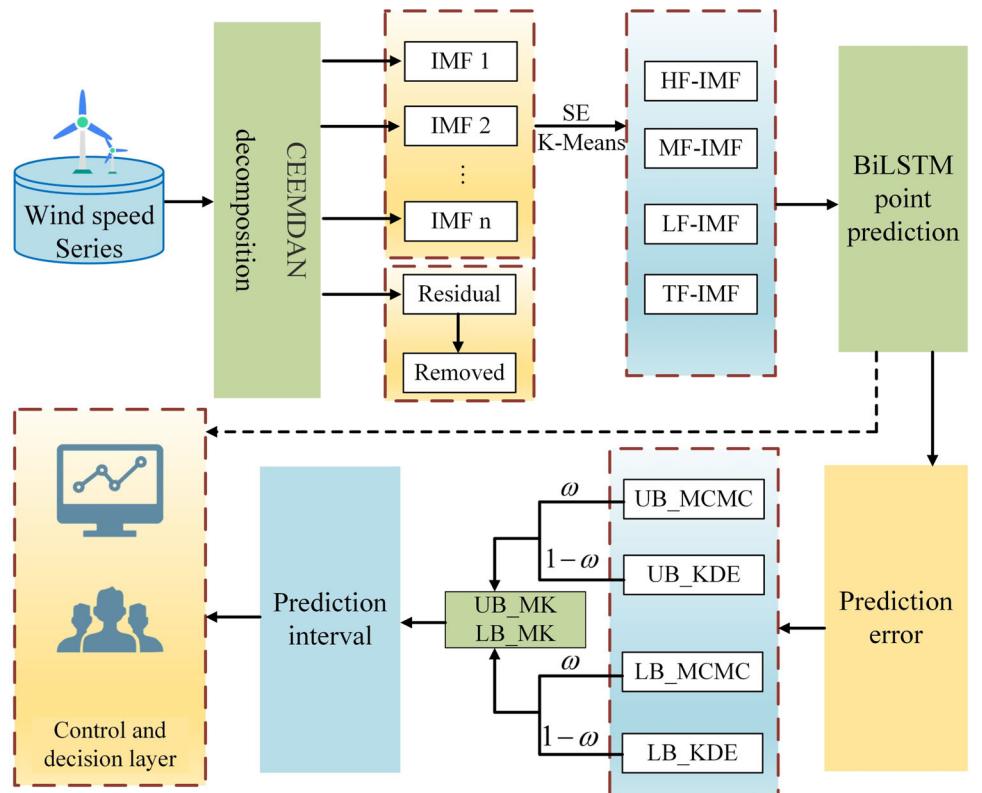
Step 1: The wind speed sequence is decomposed into different frequency modal components using CEEMDAN. The SE and K-Means methods are then used to reconstruct the components into four new components.

Step 2: The reconstructed components are predicted using the BiLSTM model, and the prediction results are summed to obtain the wind speed point prediction.

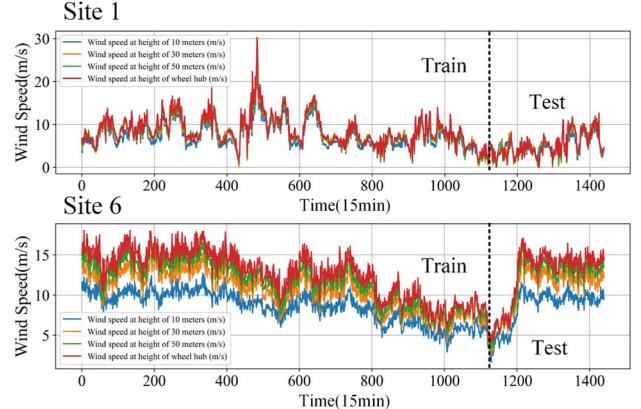
Step 3: Prediction errors are analyzed, and the MCMC and KDE methods are used to construct upper and lower bounds for the interval prediction. The MK method is then used to combine the interval results through weighted summation.

Step 4: Control and decision-making are performed based on the point prediction and interval prediction results.

Fig. 3 Framework diagram of the proposed CEEMDAN-SE-BiLSTM-MK model



(a) Schematic diagram of wind speed at various heights



(b) Wind speed fluctuations at various heights

Fig. 4 Wind speed at various heights of the wind farm

3 Experimental results

3.1 Dataset

This study uses a dataset from a wind farm in China [31], which is collected using the SCADA system. The SCADA system ensures the accuracy and reliability of data collection, with a sampling interval of 15 min. The data include wind speed (m/s), wind direction (°), barometric pressure (hPa),

temperature (°C), humidity (%), power generation (MW), etc. We selected wind speed data samples from Site 1 during June 1–15, 2020 (summer) and Site 6 during December 1–15, 2020 (winter). Each site has 1440 samples, and the train data accounts for 80% of the total data, while the test data accounts for 20%. As shown in Fig. 4b, the wind speed at Site 1 in summer has a large fluctuation range, and the wind speed at different heights is similar. The wind speed at Site 6 in winter has a smaller fluctuation range, and the

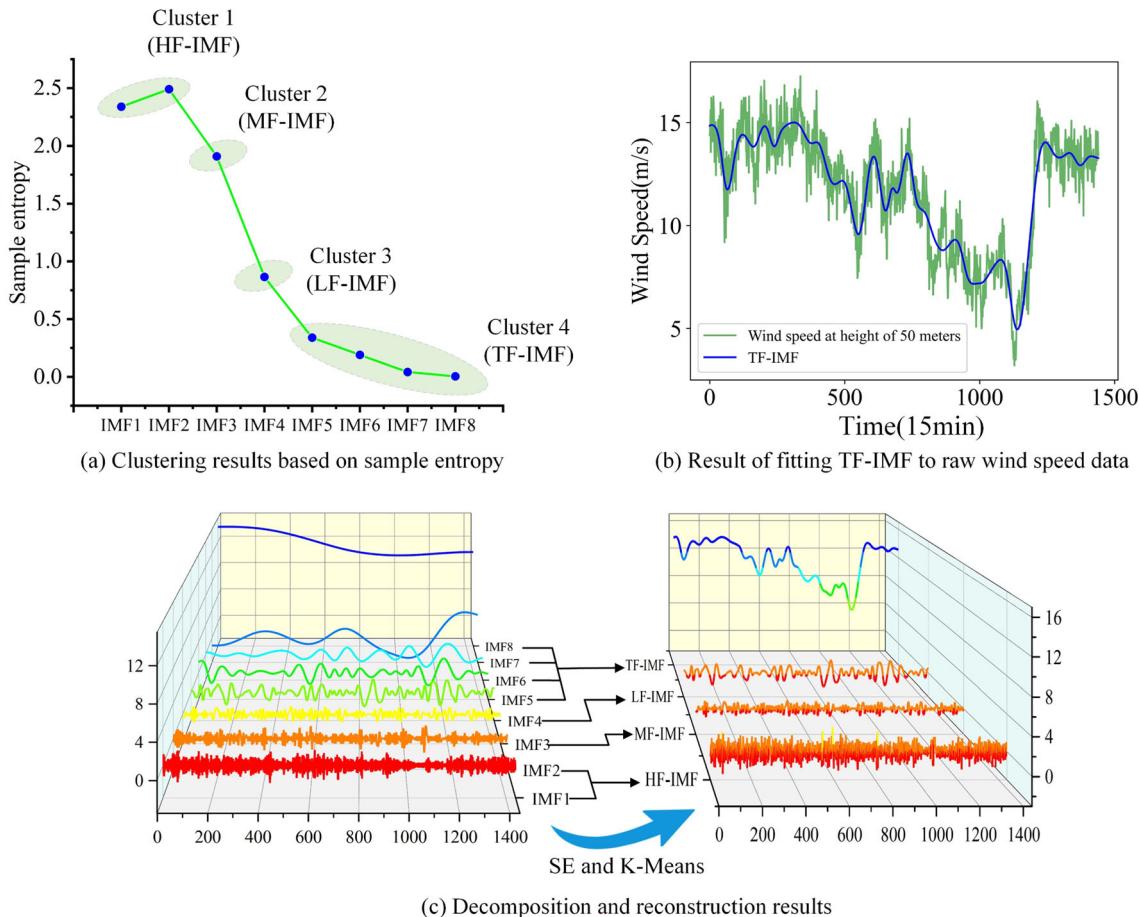


Fig. 5 Decomposition–reconstruction process diagram

wind speed at various heights shows a layered phenomenon. However, overall, both site's wind speed data exhibit strong randomness and nonlinear characteristics. The wind speeds at 30 m and 50 m heights have the greatest impact on the wind turbine's rotational speed, making them the focus of wind speed prediction in the wind farm.

3.2 Experimental parameter setup

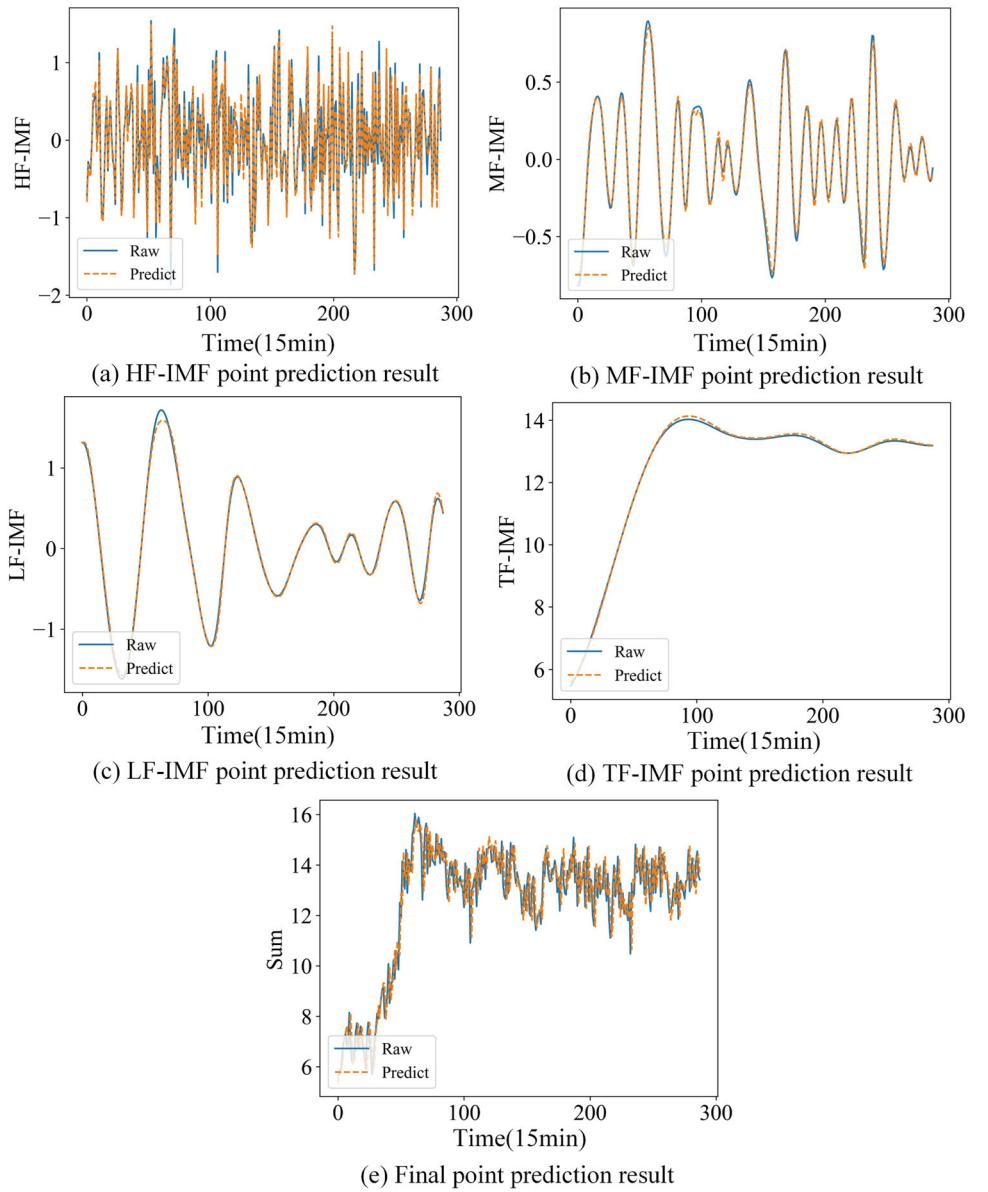
A total of four models were set up for the experiment, namely CEEMDAN-SE-LSTM (Model 1), CEEMDAN-SE-GRU (Model 2), CEEMDAN-SE-BiGRU (Model 3), and CEEMDAN-SE-BiLSTM (Proposed). All four models have three hidden layers L_1 , L_2 , and L_3 with values of 128, 64, and 32, respectively. Dropout is 0.2, epoch is 50, and batch_size is 50.

3.3 Decomposition and reconstruction

In this section, we use the CEEMDAN to decompose the original wind speed data and extract modal components with different frequency characteristics. Taking the 50 m height

wind speed at Site 6 as an example, the decomposition and reconstruction process is shown in Fig. 5. CEEMDAN decomposes the original wind speed sequence into 8 components with different frequencies, and the volatility and complexity of IMF1-IMF8 gradually decrease. To measure the time complexity of each modal component and provide a theoretical basis for reconstruction, we also calculate the sample entropy, where the sample entropy of each modal component represents the difference in time complexity. Next, we apply the K-Means clustering algorithm to aggregate and reconstruct modal components with similar sample entropy. The clustering result is shown in Fig. 5a. Through this approach, we aim to reduce the random fluctuations in the wind speed data and improve the predictive performance of the model. The reconstruction result is shown in Fig. 5c. IMF1 and IMF2 are reconstructed as the new high-frequency component (HF-IMF), representing the extremely unstable nature of the original wind speed sequence. IMF3 is reconstructed as the new mid-frequency component (MF-IMF), and IMF4 is reconstructed as the new low-frequency component (LF-IMF), indicating certain periodic characteristics in the original wind speed sequence. IMF5, IMF6, IMF7, and

Fig. 6 Point prediction results for each component after reconstruction



IMF8 are reconstructed as the trend component (TF-IMF), which represents the overall trend of wind speed fluctuations. As shown in Fig. 5b, TF-IMF can effectively capture the trend characteristics of the original wind speed data. Through the reconstruction process, we can observe that the number of components for prediction is reduced by half, effectively reducing the cost of wind speed prediction.

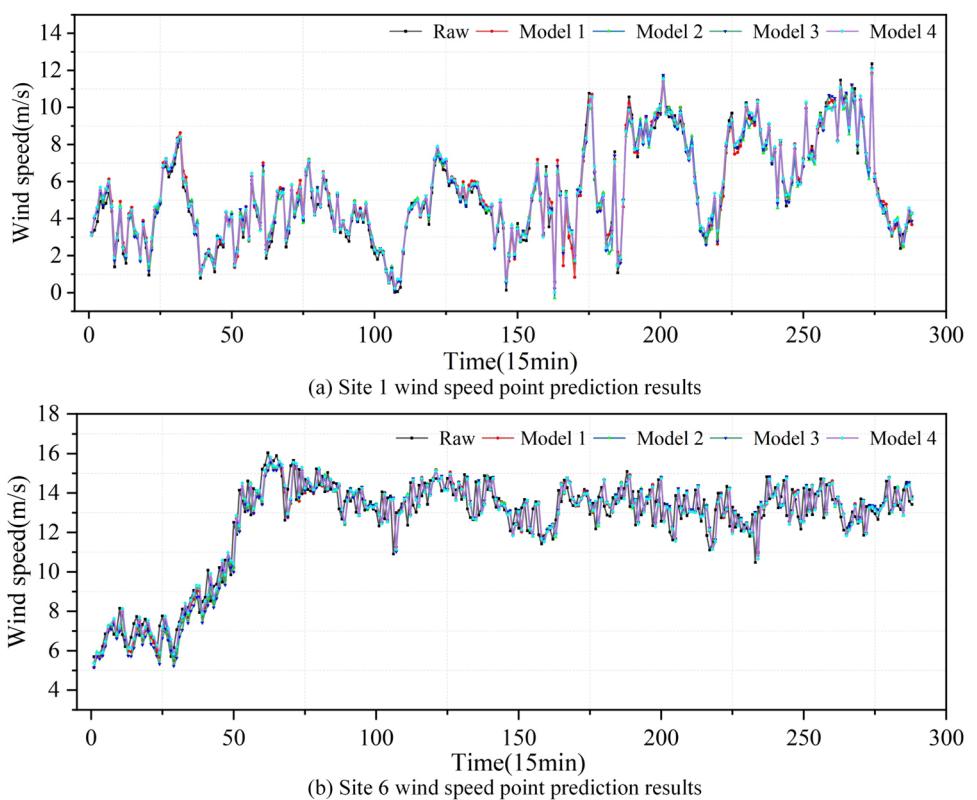
3.4 Wind speed point prediction

After predicting the reconstructed modal components using the forecasting model, Fig. 6 shows the fluctuation curves of the predicted values and actual values for each component. Among the four components, the high-frequency component (HF-IMF) has larger prediction errors at a few points,

while the other three components have good prediction performance. The specific values of the evaluation metrics are shown in Table 1. Next, the predicted results of each component are aggregated to obtain the final prediction result, as shown in Fig. 7. Different forecasting models are compared in terms of their prediction performance using the evaluation metrics. It is found that the four models have similar prediction performance, but under the same parameter conditions, the proposed CEEMDAN-SE-BiLSTM model has stronger generalization ability and more stable prediction performance. In the wind speed prediction results at Site 6, all four models exhibit a certain degree of lag, but within a controllable range.

Table 1 Prediction results for each model component

Models	IMFs	Site 1			Site 6		
		MAE	RMSE	R2	MAE	RMSE	R2
Model 1	HF-IMF	0.135	0.169	0.934	0.245	0.325	0.939
	MF-IMF	0.015	0.021	0.997	0.139	0.203	0.886
	LF-IMF	0.031	0.048	0.996	0.073	0.115	0.993
	TF-IMF	0.089	0.098	0.998	0.146	0.185	0.988
Model 2	HF-IMF	0.126	0.161	0.941	0.178	0.237	0.961
	MF-IMF	0.024	0.031	0.993	0.150	0.211	0.866
	LF-IMF	0.034	0.054	0.995	0.119	0.152	0.983
	TF-IMF	0.114	0.139	0.996	0.195	0.229	0.982
Model 3	HF-IMF	0.120	0.150	0.948	0.153	0.200	0.972
	MF-IMF	0.015	0.021	0.997	0.135	0.193	0.895
	LF-IMF	0.027	0.049	0.996	0.027	0.049	0.996
	TF-IMF	0.162	0.224	0.989	0.105	0.126	0.995
Model 4 (Proposed)	HF-IMF	0.127	0.158	0.942	0.161	0.214	0.974
	MF-IMF	0.025	0.033	0.992	0.135	0.189	0.904
	LF-IMF	0.026	0.037	0.997	0.069	0.093	0.995
	TF-IMF	0.042	0.051	0.999	0.156	0.202	0.987

Fig. 7 Point prediction results of each model

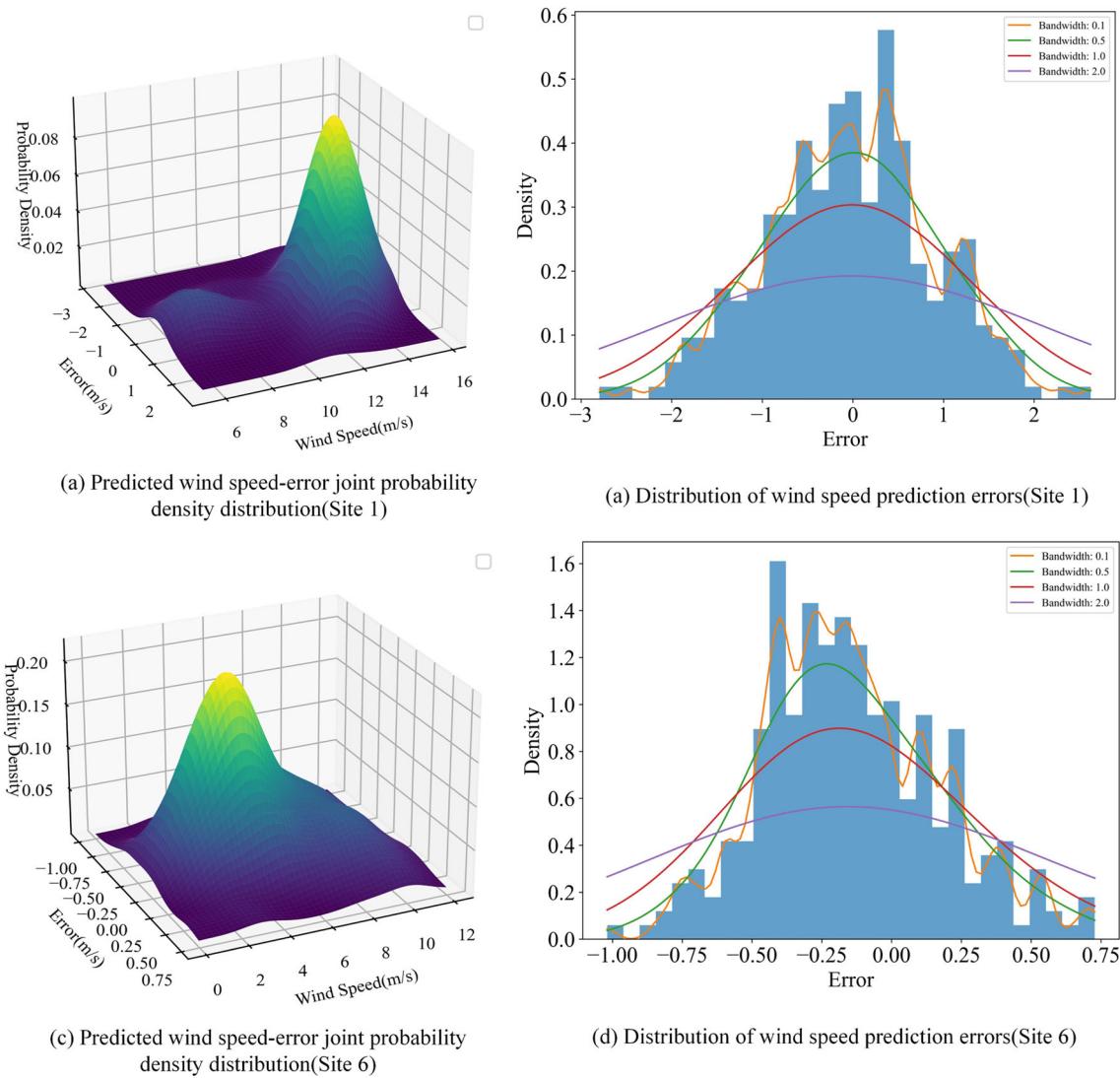


Fig. 8 Joint probability density graphs and error distributions of errors in predicted wind speeds

3.5 Prediction error analysis

The joint probability density distribution of predicted wind speed and error was established using KDE, as shown in Fig. 8. It can be observed that when the wind speed is high, the model exhibits higher prediction accuracy, with a low probability of large errors. This indicates that the model performs well in accurately predicting wind speeds when they are high. Conversely, when the wind speed is low, the model's accuracy decreases, and the probability of larger prediction errors increases. To enhance the reliability of wind power prediction intervals, it is crucial to accurately fit the distribution of prediction errors and capture the variation in historical prediction errors. The wind speed error distribution was fitted using kernel density estimation, as shown in Fig. 8. It is

worth noting that the choice of bandwidth during the estimation process affects the quality of the fit, so in this paper the best bandwidth is chosen for the fit.

3.6 Wind speed interval prediction

If the final prediction interval is directly constructed using the final results, it may amplify the errors of the components, resulting in an excessively wide average width of the wind speed interval. Additionally, the lag of the point prediction results will also propagate into the interval prediction, leading to a low PICP. Therefore, in this study, the prediction errors for each component were obtained based on point predictions. These prediction error sequences were then combined with MCMC and KDE to obtain MCMC prediction intervals and KDE prediction intervals for each component. Figure 9 illustrates the MCMC prediction intervals for the 50 m height

Fig. 9 Interval prediction results for each component

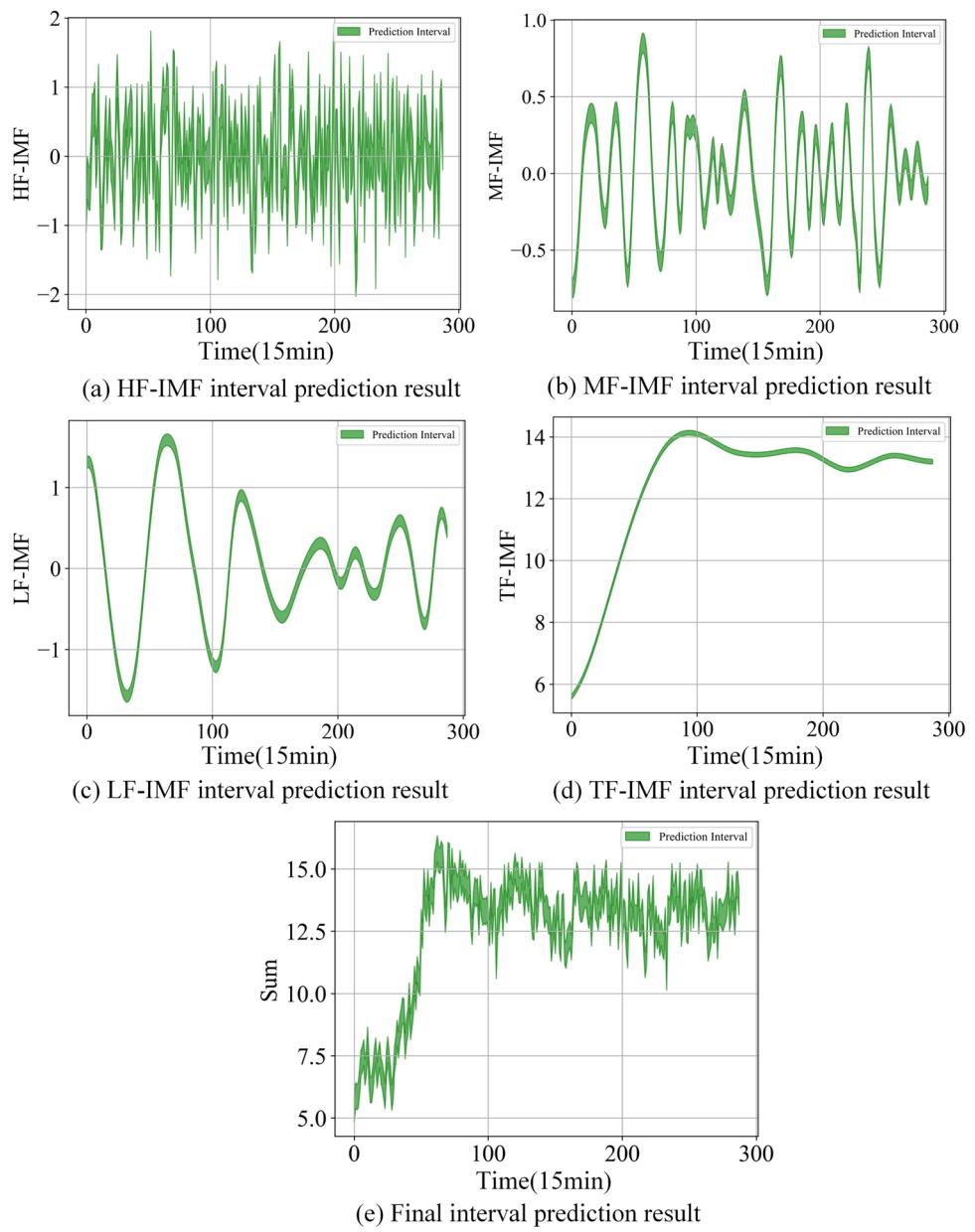


Table 2 Component interval prediction results

Data	IMFs	MCMC			KDE		
		PICP	PINAW	F-value	PICP	PINAW	F-value
Site 1	HF-IMF	0.9375	0.1010	1.7128	0.9861	0.3537	1.4622
	MF-IMF	0.9514	0.1491	1.6663	0.9930	0.4810	1.3441
	LF-IMF	0.9444	0.0917	1.8023	1.0	0.6551	1.2084
	TF-IMF	0.7813	0.5712	1.4580	1.0	0.8946	1.0556
Site 6	HF-IMF	0.9271	0.1781	1.5915	1.0	1.1091	0.9483
	MF-IMF	0.9410	0.0745	1.7586	1.0	3.9832	0.4013
	LF-IMF	0.9375	0.0436	1.8014	1.0	2.2442	0.6165
	TF-IMF	0.8715	0.0176	1.7168	1.0	1.0146	0.9928

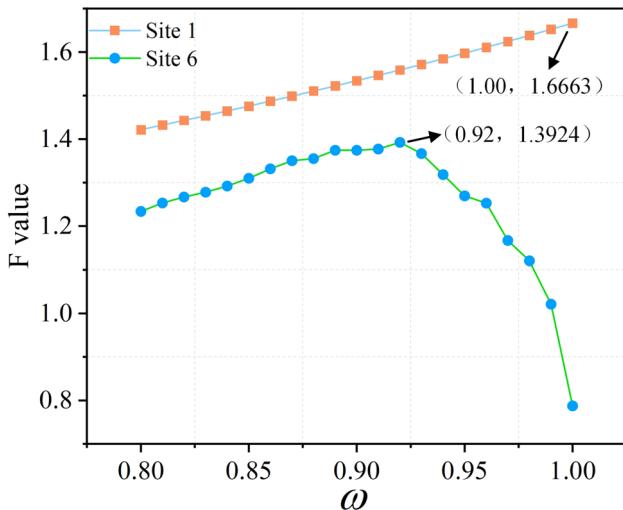


Fig. 10 Weighting parameters grid search optimization results

wind speed components at Site 6. Table 2 presents the PICP, PINW, and F -values for each component. It can be observed that using MCMC prediction intervals has the advantage of low width and high coverage in wind speed interval prediction at Site 1, but performs poorly at Site 6. On the other hand, KDE prediction intervals have high coverage but at the expense of widening the interval width.

In this study, two interval prediction methods were integrated to propose the MK interval prediction method. This method combines the upper and lower boundary prediction results of the two methods by weighted overlay to construct a new prediction interval. The objective is to find the maximum F -value, and the weight parameter ω is continuously adjusted through grid search. Based on some prior knowledge, the optimal parameter ω is likely to be in the range of [0.80, 1.00], with a search step size of 0.01. The grid search results are shown in Fig. 10. From Fig. 10, it can be observed that for wind speed interval prediction at Site 1, the F -value is maximized when the weight parameter ω is equal to 1 (i.e., original MCMC), indicating the optimal interval. For Site 6, the maximum F -value is achieved when ω is 0.92. The final interval prediction evaluation results are shown in Table 3. For wind speed prediction intervals at Site 1, both MCMC and KDE achieve a PICP of 100% at a 95% confidence level. The PINAW for MCMC is only 0.2003 m/s, while for KDE

it is 1.2334 m/s. For wind speed prediction intervals at Site 6, the performance of MCMC prediction intervals is not stable. Compared to the original MCMC results, the MK method improves the PICP and F -value by 114.67% and 76.84%, respectively.

The final PI are shown in Fig. 11, Fig. 11a shows the results of the interval prediction of MCMC at different confidence levels, which basically covers all the original data. Figure 11b shows the results of PI for MCMC and MK.

3.7 Quantification of uncertainty in components

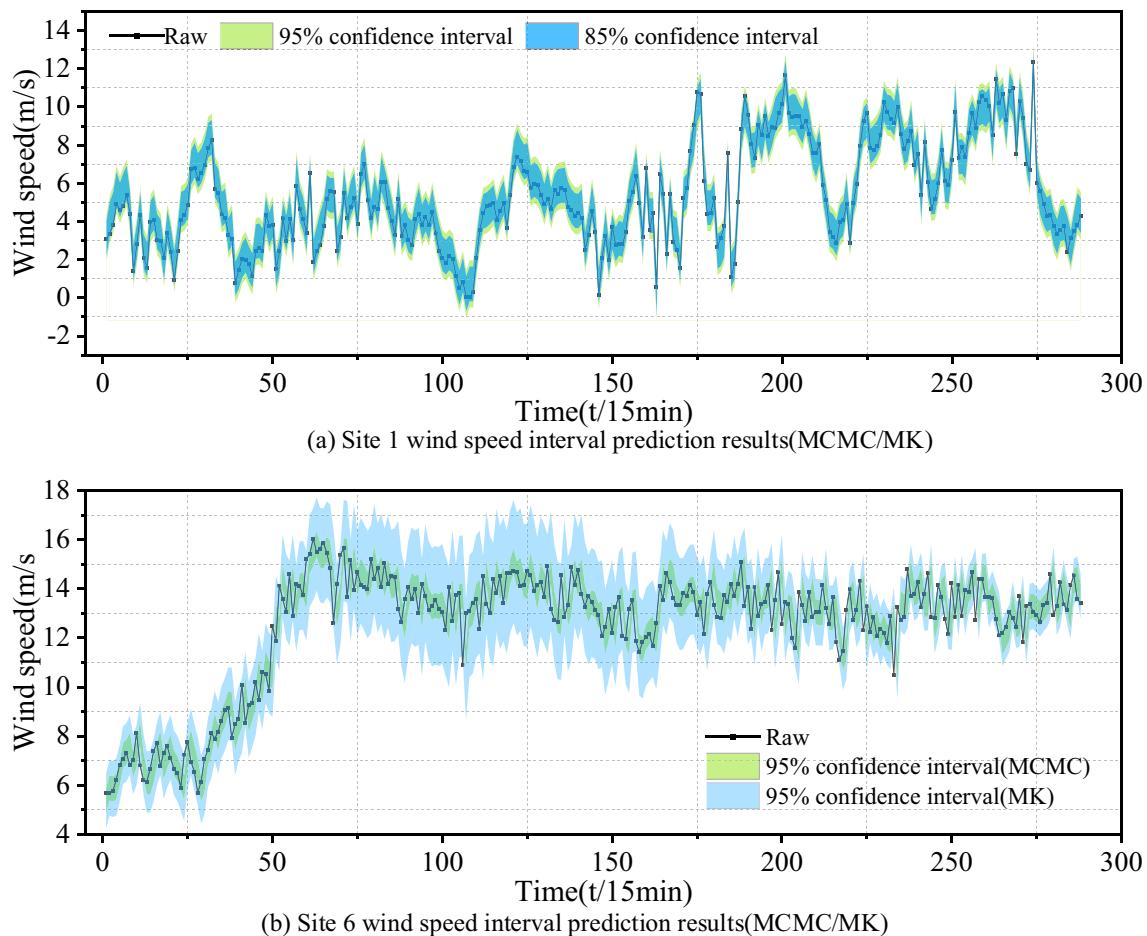
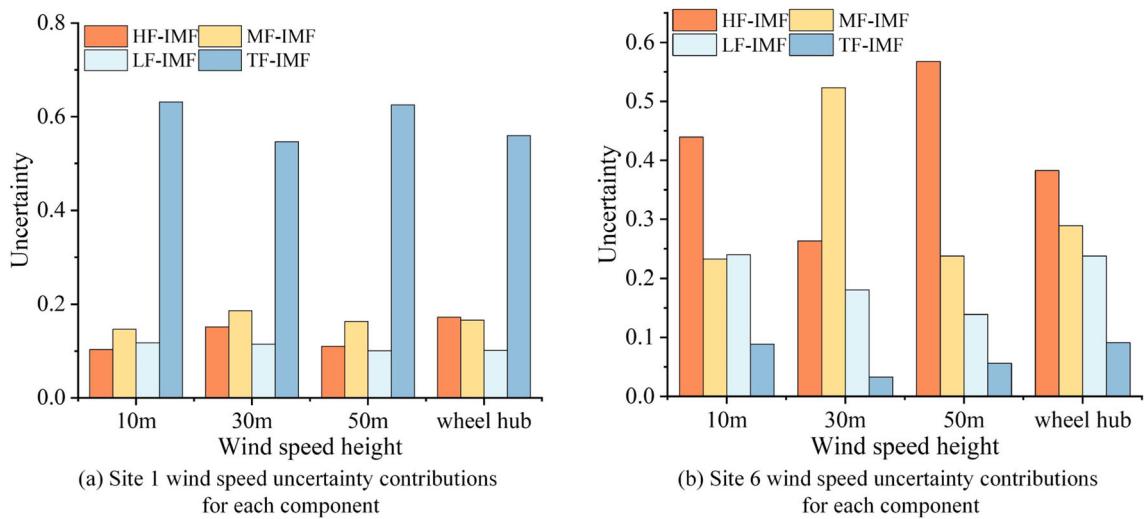
In the process of constructing wind speed intervals, the individual components are separately evaluated to explore and quantify the impact of component prediction uncertainty on the overall wind speed prediction uncertainty. The uncertainty is quantified using the interval area, where a larger area indicates higher uncertainty. The results, as shown in Fig. 12, indicate that the uncertainty in wind speed prediction at Site 1 is mainly concentrated in the trend component (TF-IMF) prediction process, which may be attributed to the less significant trend in the wind speed data at Site 1. On the other hand, the uncertainty in wind speed prediction at Site 6 is concentrated in the high-frequency (HF-IMF) and mid-frequency (MF-IMF) prediction processes, indicating a more pronounced trend in the wind speed data at Site 6. Therefore, it is possible to dynamically adjust the model parameters based on the fluctuation characteristics of wind speed to reduce the prediction uncertainty of individual components and improve the overall accuracy of wind speed prediction.

4 Conclusion

The proposed CEEMDAN-SE-BiLSTM-MK model aims to achieve accurate wind speed prediction and reliable prediction intervals (PI), providing wind farm managers with richer decision-making information. In terms of point prediction, the proposed model exhibits greater stability and accuracy compared to other models. For interval prediction, considering the lag between point prediction and interval prediction, the MK interval prediction method is proposed. This method integrates the MCMC and KDE interval prediction methods and determines the optimal weight parameter

Table 3 Comparison of results of interval prediction methods

Methods	Site 1			Site 6		
	PICP	PINAW	F -value	PICP	PINAW	F -value
MCMC	1.0	0.2003	1.6663	0.4097	0.0993	0.7874
KDE	1.0	1.2334	0.8955	1.0	2.5836	0.5581
MK	1.0	0.2003	1.6663	0.8795	0.2980	1.3924

**Fig. 11** PI results**Fig. 12** Quantitative results of component uncertainty

through grid search. Experimental results demonstrate that the MK method can improve prediction accuracy and outperform other evaluation metrics on certain sites. Furthermore, the study evaluates and quantifies the impact of component prediction uncertainty on the final prediction results, providing a better understanding of the sources of uncertainty in the decomposition-reconstruction wind speed prediction process. Measures are taken to enhance the reliability of predictions. Therefore, the proposed CEEMDAN-SE-BiLSTM-MK model can provide wind farm managers with more accurate and reliable decision-making information, maximizing the potential of wind energy resources and offering practical value.

In future work, different models or parameters can be applied in the component prediction process to reduce prediction uncertainty. Additionally, a smaller step size can be used to search for the optimal weight parameters.

Author contributions C.S.Y. and Y.T.M. conceptualized the study, conducted the methodology, performed formal analysis and investigation, developed the software, curated the data, verified the results, provided supervision, and contributed to writing the original draft. C.S.Y. was primarily responsible for writing the original draft, while both C.S.Y. and Y.T.M. were involved in reviewing and editing the manuscript.

Funding The authors declare no funding.

Data availability No datasets were generated or analyzed during the current study.

Declarations

Conflict of interest The authors declare no competing interests.

References

- Hutchinson M, Zhao F et al (2023) Global wind energy report 2023. Global Wind Energy Council (GWEC), 2023. <https://gwec.net/globalwindreport2023/>
- Demolli H, Dokuz AS, Ecemis A et al (2019) Wind power forecasting based on daily wind speed data using machine learning algorithms. *Energy Convers Manage* 19:111823. <https://doi.org/10.1016/j.enconman.2019.111823>
- Porté-Agel F, Bastankhah M, Shamsoddin S (2020) Wind-turbine and wind-farm flows: a review. *Bound-Layer Meteorol* 17:1–59. <https://doi.org/10.1007/s10546-019-00473-0>
- Di ZH, Ao J, Duan Q et al (2019) Improving WRF model turbine-height wind-speed forecasting using a surrogate—based automatic optimization method. *Atmos Res* 22:1–16. <https://doi.org/10.1016/j.atmosres.2019.04.011>
- Cheng WYY, Liu YB, Bourgeois AJ et al (2017) Short-term wind forecast of a data assimilation/weather forecasting system with wind turbine anemometer measurement assimilation. *Renew Energy* 107:340–351. <https://doi.org/10.1016/j.renene.2017.02.014>
- Wang S, Wang JZ, Lu HY et al (2021) A novel combined model for wind speed prediction—combination of linear model, shallow neural networks, and deep learning approaches. *Energy* 234:121275. <https://doi.org/10.1016/j.energy.2021.121275>
- Aasim SSN, Mohapatra A (2019) Repeated wavelet transform based ARIMA model for very short-term wind speed forecasting. *Renew Energy* 136:758–768. <https://doi.org/10.1016/j.renene.2019.01.031>
- Louka P, Galanis G, Siebert GN et al (2008) Improvements in wind speed forecasts for wind power prediction purposes using Kalman filtering. *J Wind Eng Ind Aerodyn* 96:2348–2362. <https://doi.org/10.1016/j.jweia.2008.03.013>
- Wang Y, Wang D, Tang Y (2020) Clustered hybrid wind power prediction model based on ARMA, PSO-SVM, and clustering methods. *IEEE Access* 8:17071–17079. <https://doi.org/10.1109/ACCESS.2020.2968390>
- Yang HF, Zhu ZJ, Li C et al (2020) A novel combined forecasting system for air pollutants concentration based on fuzzy theory and optimization of aggregation weight. *Appl Soft Comput* 87:105972. <https://doi.org/10.1016/j.asoc.2019.105972>
- Chen G, Li L, Zhang Z et al (2020) Short-term wind speed forecasting with principle-subordinate predictor based on Conv-LSTM and improved BPNN. *IEEE Access* 8:67955–67973. <https://doi.org/10.1109/ACCESS.2020.2982839>
- Liu ZJ, Liu HZ (2023) A novel hybrid model based on GA-VMD, sample entropy reconstruction and BiLSTM for wind speed prediction. *Measurement* 222:113643. <https://doi.org/10.1016/j.measurement.2023.113643>
- Yang QS, Huang GQ, Li T et al (2023) A novel short-term wind speed prediction method based on hybrid statistical-artificial intelligence model with empirical wavelet transform and hyper-parameter optimization. *J Wind Eng Ind Aerodyn* 240:105499. <https://doi.org/10.1016/j.jweia.2023.105499>
- Suo LM, Peng T, Song SH et al (2023) Wind speed prediction by a swarm intelligence based deep learning model via signal decomposition and parameter optimization using improved climp optimization algorithm. *Energy* 276:127526. <https://doi.org/10.1016/j.energy.2023.127526>
- Liu MD, Ding L, Bai YL (2021) Application of hybrid model based on empirical mode decomposition, novel recurrent neural networks and the ARIMA to wind speed prediction. *Energy Convers Manage* 233:113917. <https://doi.org/10.1016/j.enconman.2021.113917>
- Lin BQ, Zhang CC (2021) A novel hybrid machine learning model for short-term wind speed prediction in inner Mongolia, China. *Renewable Energy* 179:1565–1577. <https://doi.org/10.1016/j.renene.2021.07.126>
- Du BG, Huang S, Guo J et al (2022) Interval forecasting for urban water demand using PSO optimized KDE distribution and LSTM neural networks. *Appl Soft Comput* 122:108875. <https://doi.org/10.1016/j.asoc.2022.108875>
- Han Y, Mi LH, Shen L et al (2022) A short-term wind speed interval prediction method based on WRF simulation and multivariate line regression for deep learning algorithms. *Energy Conversat Manage* 258:115540. <https://doi.org/10.1016/j.enconman.2022.115540>
- Li JR, Wang JY, Li ZW (2023) A novel combined forecasting system based on advanced optimization algorithm—a study on optimal interval prediction of wind speed. *Energy* 264:126179. <https://doi.org/10.1016/j.energy.2022.126179>
- Wang JZ, Cheng ZS (2021) Wind speed interval prediction model based on variational mode decomposition and multi-objective optimization. *Appl Soft Comput* 113:107848. <https://doi.org/10.1016/j.asoc.2021.107848>
- Naik J, Bisoi R, Dash PK (2018) Prediction interval forecasting of wind speed and wind power using modes decomposition based low rank multi-kernel ridge regression. *Renew Energy* 129:357–383. <https://doi.org/10.1016/j.renene.2018.05.031>
- Sreekumar S, Khan NU, Rana AS et al (2022) Aggregated net-load forecasting using Markov-chain Monte-Carlo regression and C-vine copula. *Appl Energy* 328:120171. <https://doi.org/10.1016/j.apenergy.2022.120171>

23. Shi PF, Yang T, Yong B, Xu CY et al (2023) Some statistical inferences of parameter in MCMC approach and the application in uncertainty analysis of hydrological simulation. *J Hydrol* 617:128767. <https://doi.org/10.1016/j.jhydrol.2022.128767>
24. Saeed A, Li CS, Gan ZH, Xie Y et al (2022) A simple approach for short-term wind speed interval prediction based on independently recurrent neural networks and error probability distribution. *Energy* 238:122012. <https://doi.org/10.1016/j.energy.2021.122012>
25. Zhang YG, Pan GF, Zhao YP et al (2020) Short-term wind speed interval prediction based on artificial intelligence methods and error probability distribution. *Energy Convers Manage* 224:113346. <https://doi.org/10.1016/j.enconman.2020.113346>
26. Wahbah M, Mohandes B, EL-Fouly TH, El Moursi MS (2022) Unbiased cross-validation kernel density estimation for wind and PV probabilistic modelling. *Energy Convers Manage* 266:115811. <https://doi.org/10.1016/j.enconman.2022.115811>
27. Fan HJ, Zhen Z, Liu N, Sun YQ et al (2023) Fluctuation pattern recognition based ultra-short-term wind power probabilistic forecasting method. *Energy* 266:126420. <https://doi.org/10.1016/j.energy.2022.126420>
28. Torres ME, Colominas MA, Schlotthauer G, et al (2011) A complete ensemble empirical mode decomposition with adaptive noise. In: 2011 IEEE international conference on acoustics, speech and signal processing, Prague, Czech Republic pp 4144–4147. <https://doi.org/10.1109/ICASSP.2011.5947265>
29. Song DL, Yu M, Wang ZF et al (2023) Wind and wave energy prediction using an AT-BiLSTM model. *Ocean Eng* 281:115008. <https://doi.org/10.1016/j.oceaneng.2023.115008>
30. Aslan S (2023) A deep learning-based sentiment analysis approach (MF-CNN-BiLSTM) and topic modeling of tweets related to the Ukraine–Russia conflict. *Appl Soft Comput* 143:110404. <https://doi.org/10.1016/j.asoc.2023.110404>
31. Chen YB, Xu JJ (2022) Solar and wind power data from the Chinese state grid renewable energy generation forecasting competition. *Sci Data* 9:577. <https://doi.org/10.1038/s41597-022-01696-6>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.