Wind Speed Forecast based on Combined Theory, Multi-objective Optimisation, and Sub-model Selection

Tonglin Fu
Longdong University
Shenghui Zhang (zhangshh17@lzu.edu.cn)
University of Macau https://orcid.org/0000-0002-3047-5395

Research Article

Keywords: wind energy, wind speed forecasting, sub-model selection, combined model, multi-objective optimisation

Posted Date: April 5th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1492142/v1

Abstract

Wind energy is the primary energy source for a sustainable and pollution-free global power supply. However, because of its characteristic irregularity, nonlinearity, non-stationarity, randomness, and intermittency, previous studies have only focused on stability or accuracy, and the forecast performances of their models were poor. Moreover, in previous research, the selection of sub-models used for the combined model was not considered, which weakened the generalisability. Therefore, to further improve the forecast accuracy and stability of the wind speed forecasting model, and to solve the problem of sub-model selection in the combined model, this study developed a wind speed forecasting model using data preprocessing, a multi-objective optimisation algorithm, and sub-model selection for the combined model. Simulation experiments showed that our combined model not only improved the forecasting accuracy and stability but also chose different sub-models and different weights of the combined model for different data; this improved the model generalisability. Specifically, the MAPEs of our model are less than 4.96%, 4.60% and 5.25% in one, two and three step forecast.Thus, the proposed combined model is demonstrated as an effective tool for grid dispatching.

1. Introduction

For decades, the demand for electricity has represented a major global constraint, and the use of renewable energy has become increasingly important. From the perspective of economic growth, energy plays a vital role in procuring power from nature. Fossil fuels have been used to generate electricity, however, owing to the related fossil fuel crisis and current global environmental issues, the energy mixture is changing. Renewable energy sources (e.g., wind, tidal, and solar energy) are gaining increasing amounts of attention [1]. Wind energy is clean, abundant, environmentally friendly, inexhaustible, and inexpensive, it represents a feasible alternative to fossil fuels.

In recent decades, significant progress has been made in wind turbines, from the small to large scales. The reason for this rapid growth is that many parts of the world are rich in the required raw materials, the turbines are ecologically friendly and achieve low-carbon power generation, and new policies have been implemented towards the introduction of new renewable generators [2,3]. For example, in France, wind power installation increased by 14.04% [4] in 2017, mainly owing to tariff subsidies. According to statistical research, the global cumulative installed wind capacity reached nearly 591 Gigawatts (GW) by the end of 2018, with an annual growth rate of 9.6% [5]. However, owing to changes in wind speed and direction (especially the former), the integration of a large portion of wind power in wind power systems faces major challenges [6]. These challenges can be classified into two categories: operational issues [7] and planning and economic issues [8]. Improving wind forecasting is one of the most effective ways to overcome these challenges.

In recent research, improving the forecast accuracy of wind speed has become a hot topic, and new development directions and numerous wind speed forecasting methods have been proposed. Some of these technologies provide more accurate wind speed forecasts for specific wind speed data, whereas others are effective for multiple wind speed dataset [9]. These different wind-speed forecasting models can be classified into the following six categories:

(1) Persistence models. Persistence models, which assume that the wind speed at a certain time in the future matches the forecasting wind speed [10], are used for short-term forecasts. However, when the forecast timescale increases, the forecast accuracy of the model decreases. These models can be used as reference models to test new models for short-term wind-speed forecasts [11].

(2) Physical models. Physical models are used in atmospheric weather forecast modelling, they require copious amounts of numerical weather forecast data, including humidity, temperature, pressure, speed, and topological parameters, this leads to data accumulation. To forecast the wind speed, a long calculation time is required for data correlation. Therefore, the model is best suited for long-term wind speed forecasting [12].

(3) Statistical models. Historical data are used in statistical models to forecast the wind speed. Statistical models include linear and nonlinear models, and these have also been used in time series forecasting [13,14]. However, for linear methods such as the auto-regressive moving average [15, 16], auto-regressive integrated moving average (ARIMA) [17], Box–Jenkins method, and Markov chain models [18], when the main component of the wind speed data set is a non-linear feature, most statistical models assume that wind speed to be normally distributed, hence, the accuracy of these forecasting methods decreases rapidly. These nonlinear statistical methods require a large amount of historical data to train and develop forecasts, they primarily include fuzzy logic methods [19], support vector machines (SVMs) [20,21] and probabilistic methods[22].

(4) Hybrid models. Hybrid methods combine complementary characteristics (e.g., linear and nonlinear data) and the advantages of various methods to obtain the best forecasting performance.

(5) Artificial intelligence methods. Artificial intelligence methods, including artificial neural networks (ANNs) such as the back propagation neural network (BPNN) [23], generalised regression neural network (GRNN) [24], radial basis function neural network (RBFNN) [24], multiple

layer perceptron [25], and long short-term memory (LSTM) [26], have also been proposed to forecast wind speeds [28-30]. To improve the forecasting accuracy, researchers have also used optimisation algorithms such as differential evolution [31] and cuckoo search [32], amongst others.

(6) Combined models. Depending on the above methods, the stability and forecasting accuracy of forecasting beyond these methods do not meet the levels desired by wind farm operators. Therefore, to achieve a higher level of accuracy and more stable forecasting results, researchers have developed a combination model that incorporates the advantages of a single forecast model and can be widely used for wind-speed forecasting [33-35].

In general, the combined models not only overcome certain difficulties of the single models but also combine the advantages of these models, making them better than the models mentioned above. However, for combined models, the selection of sub-models has always been difficult. Some researchers choose linear and non-linear models so that the combined models can satisfy both linear and nonlinear data. Some researchers have selected models that perform better on a given dataset.

However, certain problems remain: (a) The lack of complementary theoretical knowledge for selecting linear and non-linear models, in other words, why are these methods selected as sub-models? (b) The listed sub-models are limited. For the resulting combined model, the sub-models represent the best of these models. Because the sub-models do not cover most models, the single model selected is not sufficiently convincing. (c) For different datasets, because of the different data characteristics, it is impossible to find a single combined model suitable for multiple datasets. (d) Most of the models feature single objectives. For wind speed data, it is usually difficult to guarantee both forecasting accuracy and stability using a single objective function in the combination model, owing to the nonlinear characteristics of the wind-speed time series.

To the best of our knowledge, no model has been proposed that can resolve the problems stated above. Thus, we propose a forecasting model based on model selection, multi-objective functions, and combined model theory. In the following section, we use CM to denote our proposed combined model.

Our contributions and innovations are as follows:

(1) A combined model based on outlier detection and processing is constructed, and the negative effects of outliers are eliminated using data analysis methods, whilst retaining the main trend of the wind-speed time series. Outliers in the original wind-speed data can result in poor forecasting results, and the data analysis module of the CM eliminates such outliers.

(2) Based on singular spectrum analysis (SSA) technology, data preprocessing technology is used to extract the main features of the wind speed data. The wind speed data are denoised to make them smoother and more reflective of the trend of the original data.

(3) An optimal sub-model selection criterion based on multiple forecast criteria is proposed. A new predictive evolution criterion is proposed to select the optimal predictive sub-model. This criterion is called the weighted information criterion (WIC), and it combines six criteria [mean absolute percentage error (MAPE), root mean square error (RMSE), Akaike information criterion (AIC), Bayesian information criterion (BIC), and direction accuracy (DA)]. These criteria are used not only to select the best sub-models from various forecasting models but also to improve the forecasting accuracy.

(4) The weights of the sub-models used to build the combined model are obtained using a multi-objective optimisation algorithm.

(5) The single model and weight of the ideal combination model vary with respect to the data, indicating the lack of a consistent model suitable for all datasets.

The remainder of this paper is organised as follows. The theories and methods are introduced in Section 2. In Section 3, we introduce the multiple objective functions, non-dominated sorting genetic algorithm-III, and model-building process. In Section 4, the performance metrics, three numerical experiments, and summary are presented. Finally, Section 5 presents the conclusions.

2 Theories And Methods

This section introduces the theories and methods used in CM. In the attached table in the appendix, we fit the original wind-speed time series using linear and nonlinear functions; we found that the wind speed had both linear and nonlinear characteristics. Therefore, linear and nonlinear models were selected to forecast and study the wind-speed time series. The basic methods and theories were shown in the appendix. The flowchart of the proposed combined model is shown in Fig.1.

3 The Non-dominated Sorting Genetic Algorithm-iii And Model Proposal

For a single-objective model, only one objective function must be determined for forecasting. However, theoretical and practical examples have shown that it is difficult to achieve high stability and accuracy by relying on only one objective. For multi-objective problems, multiple goals must be proposed for the optimisation algorithm to optimise, therefore, an objective function and optimisation algorithm [the non-dominated sorting genetic algorithm-III (NSGA-III)] are introduced in this section.

3.1 Non-dominated Sorting Genetic Algorithm-III (NSGA-III)

The NSGA-III [36] includes five main steps: (1) Population initialisation: Decision variables are randomly generated based on the given upper and lower boundaries. (2) Offspring selection (this is the selection algorithm for the next generation of individual evolution, it is based on a genetic algorithm): According to natural section theory, offspring selection is introduced in the algorithm to determine which is more likely to produce a solution. (3) Generation of new offspring: The parent generation generates new offspring via crossovers and mutations. (4) Non-dominated sorting: The solutions are sorted according to non-dominated relations. (5) Reference-point-based selection mechanism: To select a new group of size N in the next generation, a reference-point-based selection mechanism is introduced in the NSGA-III, this guarantees the uniformity of the distribution and is an enhanced optimisation drive for multi-objective optimisation problems. We expected the entire process to identify each group member corresponding to the reference point close to the Pareto optimal frontier. These group members constitute a set of Pareto-optimal solutions.

3.2 Model Proposal

The structure and construction process of the proposed two-objective combined model CM are introduced in this section.

1. We select the data to be tested and use SSA [37] to denoise it.

(1) Embedding:

We construct a trajectory matrix $X_{L\times K}$as

$$\boldsymbol{X} = [x_1, \cdots, x_N] = (x_{ij})_{i,j=1}^{L,K} = \begin{bmatrix} x_1 & x_2 & x_3 & \cdots & x_K \\ x_2 & x_3 & x_4 & \cdots & x_{K+1} \\ x_3 & x_4 & x_5 & \cdots & x_{K+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & x_{L+2} & \cdots & x_N \end{bmatrix}, \quad (1)$$

for the time series data, where $X = \left(x_1, \cdots, x_{i+L-1}\right)^T$, $(1 < i < K,\ L \leq K)$ is a lag vector of length L, and X is a Hankel matrix for which the elements on each of the sub diagonals are equal.

(2) Singular Value Decomposition

Singular value decomposition (SVD) is applied to the trace matrix X. We let $S = XX^T$ and calculate the eigenvalues $\lambda_1, \cdots, \lambda_L$ of S; because S is a symmetric matrix, we have that $\lambda_1 \geq \cdots \geq \lambda_L \geq 0$, and the standard orthogonal basis of the matrix S corresponding to these eigenvalues is obtained as $U_1, \ldots, U_L$.

Let $d = \mathrm{rank}(X)$ and $V_i = X^T U/\sqrt{\lambda_i}$, $(i = 1, \ldots, d)$; then, the SVD of the trace matrix X can be expressed as $X = X_1 + \cdots + X_N$, $X_i = \sqrt{\lambda_i} U_i V_i^T$, $(i = 1, \ldots, d)$. The rank of every matrix X_i is one, and these matrices are referred to as elementary. Vectors U_i are left as singular vectors of matrices X_i, and set $\sqrt{\lambda_i}$, $(i = 1, \ldots, d)$ is called the spectrum of the trace matrix; hence, this is a singular spectrum decomposition.

(3) Grouping eigenvalues:

By grouping the elementary matrices $X_i (i = 1, \ldots, d)$, the index set $\{1, \ldots, d\}$ is divided into m disjoint subsets, $I_1, \cdots, I_m$. Let I = $\left\{i_1, \cdots, i_p\right\}$. Then, the matrices X_I corresponding to Group I are defined as $X_I = X_{i_1} + \cdots + X_{i_p}$. Because the subsets are divided into m groups, X can be expressed as $X = X_{I_1} + \ldots + X_{I_m}$.

(4) Diagonal average:

Each matrix X_I of the group decomposition is Hankelised; then, by forming a one-to-one correspondence between the obtained Hankel matrix and time-series data, the previously obtained Hankel matrix will be transformed into a new sequence of length N. For any matrix $Z_{L\times K}$, its elements are z_{ij}, and Hankel $Z_{L\times K}$; then, the k-th value denotes the average over all elements of XZ satisfying $i + j = k + 2$.

The diagonal average method is applied to generate a reconstruction sequence $\tilde{X}^{(k)} = (\tilde{X}_1^{(k)}, \cdots, \tilde{X}_N^{(k)})$ from the generated matrix $x_1, \cdots, x_N$. In this way, the initial sequence $x_1, \cdots, x_N$ is decomposed into the sum of *m* reconstruction subsequences:

$$x_n = \sum_{k=1}^{m} \tilde{x}_n^{(k)}, (n = 1, 2, \dots, N).$$

(2)

Meanwhile, we take $\frac{1}{2}m$ and $\dot{x}_n = \sum_{k=1}^{\frac{1}{2}m} \tilde{x}_n^{(k)}$, $(n = 1, 2, \dots, N)$; then, $\dot{x}_1, \cdots, \dot{x}_N$ is the time series data after $x_1, \cdots, x_N$ is denoised.

2. According to the model selection algorithm WIC, we select N single models M and record each single model. Then, we reconstruct the data. The reconstructed data are expressed as follows:

$$\begin{bmatrix} \dot{x}_1 & \dot{x}_2 & \dot{x}_3 & \cdots & \dot{x}_K \\ \dot{x}_2 & \dot{x}_3 & \dot{x}_4 & \cdots & \dot{x}_{K+1} \\ \dot{x}_3 & \dot{x}_4 & \dot{x}_5 & \cdots & \dot{x}_{K+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \dot{x}_L & \dot{x}_{L+1} & \dot{x}_{L+2} & \cdots & \dot{x}_N \end{bmatrix}.$$

(3)

3. For the N selected single models $M_i(i = 1, 2, \dots, N)$, the forecasting value is set to $\hat{y}_i$, the initial weight $\hat{\omega}_i$ is given (to construct a combined model M_0), and the forecasting value is $\hat{y}_t = \sum_{i=1}^{M} \hat{\omega}_i \hat{y}_{it}$, $t = 1, 2, \dots, L$, where t represents each time point in the time series. In CM, to achieve a high accuracy and stability, the objective functions can be defined as

$$\text{Minimise} = \begin{cases} f_1 = \text{MSE}\left(\hat{Y}, Y\right) \\ f_2 = \text{VarSE}\left(\hat{Y}, Y\right) \end{cases},$$

(4)

where

$$MSE\left(\hat{Y}, Y\right) = \frac{1}{L}\sum_{t=1}^{L}\left(\sum_{i=1}^{M}\hat{\omega}_i\hat{y}_{it} - y_t\right)^2,$$

(5)

$$\text{VarSE}\left(\hat{Y}, Y\right) = \frac{1}{L}\sum_{t=1}^{L}\left(\left(\sum_{i=1}^{M}\hat{\omega}_i\hat{y}_{it} - y_t\right)^2 - \frac{1}{L}\sum_{t=1}^{L}\left(\sum_{i=1}^{M}\hat{\omega}_i\hat{y}_{it} - y_t\right)^2\right)^2.$$

(6)

Here, $\{y_t, t = 1, 2, \dots, L\}$ is the real value of the time series, and t is the time point.

4. A multi-objective optimisation algorithm NSGA-III is used to optimise the weight of the forecasting value for each forecasting model $M_i(i = 1, 2, \dots, N)$ in the forecasting value set $\hat{y}_t = \sum_{i=1}^{M} \hat{\omega}_i \hat{y}_{it}$, $t = 1, 2, \dots, L$. This includes five steps: population initialisation, offspring

selection (the offspring selection algorithm is a selection algorithm that produces the next generation of individuals in the genetic algorithm), new offspring generation, non-dominated sorting, and reference point-based selection.

5. A solution set S exists. The solutions in S are Pareto optimal solutions, and each solution corresponds to a set of weight values $\hat{\omega}_{ij}$, where i represents the weight of the i-th model $M_i(i = 1, 2, \ldots, N)$ and j represents the j-th solution in set S. Because the number of objective functions is three, the elements in S are three-dimensional arrays, which are expressed as follows: if one of the objective function values is reduced, the other objective function values will increase. Finally, the solution with the closest Euclidean distance to the origin in solution set S is selected as the weight:

$$d_{min} = \sqrt{\left(MSE_{best}\left(\hat{Y}, Y\right) \right)^2 + \left(VarSE_{best}\left(\hat{Y}, Y\right) \right)^2}.$$

(7)

Because each solution s corresponds to a set of weight values, the weight $\hat{\omega}_{ibest}$ of a single model in CM (with three objective functions) is found; hence, the forecasting value of the optimised CM is $\hat{y}_t = \sum_{i=1}^{M} \hat{\omega}_{ibest} \hat{y}_{it}$.

In our study, the parameters of NSGA-III are as follows: generated reference points: 10; maximum number of iterations: 50; population size: 80; crossover percentage: 0.5; mutation percentage: 0.5; and mutation rate: 0.02.

4 Numerical Experiment

To evaluate the forecasting accuracy and stability of our CM, 10-minute wind speed datasets from four stations were selected for multi-step forecasting. In this study, three datasets were selected as the research objects, and the data differed slightly depending on the forecasting steps. The ratio between the training and test sets was 125:18. When forecasting the second value, the training set and test set were each moved one dataset forward, and the numbers of the training and test sets were kept unchanged. For example, when forecasting the first value, the first 1000 sets of data were used for training, Datasets 1001 to 1144 were used for testing, and the first value was forecasted. When the second value was needed for forecasting, the second set (up to Dataset 1001) was used for training, and Datasets 1002 to 1145 were used for testing, thus, we obtained the second forecasting value, the process was repeated for a total of 1008 forecasted values. This method stopped learning when the training error reached MSE = 10^{-6} (after normalisation). Three datasets were used for the one-step, two-step, and three-step forecasting.

4.1 Performance metrics

To evaluate the characteristics of the model more comprehensively, certain performance indices were considered. Eight metrics, mean absolute error (MAE), RMSE, standard deviation of absolute percentage error (STDAPE), direction accuracy (DA), Theil U statistic 1 of forecasting results (U1), Theil U statistic 2 of forecasting results (U2), MAPE, and coefficient of determination (R^2) were used, as shown in Table 1, furthermore, the Diebold–Mariano (DM) test and forecasting availability test were used. These metrics were taken from a study by Wang et al. [38].

To evaluate the characteristics of the model more comprehensively, certain performance indices were considered. Eight metrics, mean absolute error (MAE), RMSE, standard deviation of absolute percentage error (STDAPE), direction accuracy (DA), Theil U statistic 1 of forecasting results (U1), Theil U statistic 2 of forecasting results (U2), MAPE, and coefficient of determination (R^2) were used, as shown in Table 1; furthermore, the Diebold–Mariano (DM) test and forecasting availability test were used. These metrics were taken from a study by Wang et al. [38].

Table 1
Performance metric.

Metric	Definition	Equation
MAE	Mean absolute error of N forecasting results	$MAE=\frac{1}{N}\sum_{i=1}^{N}\lvert\hat{y}_i-y_i\rvert$
RMSE	Square root of average of the error squares	$RMSE=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i-y_i)^2}$
STDAPE	Standard Deviation of N absolute percentage errors	$STDMAPE=\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\lvert y_i-\hat{y}_i\rvert}{y_i}-\frac{1}{N}\sum_{i=1}^{N}\frac{\lvert y_i-\hat{y}_i\rvert}{y_i}\right)^2}$
DA	Direction accuracy of forecasting results	$DA=\frac{1}{N}\sum_{i=1}^{N}w_i,\ w_i=\begin{cases}1, & if(y_{i+1}-y_i)(\hat{y}_{i+1}-y_i)>0\\0, & otherwise\end{cases}$
U1	Theil U statistic 1 of forecasting results	$U1=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i-y_i)^2}/\left(\sqrt{\frac{1}{N}\sum_{i=1}^{N}y_i^2}+\sqrt{\frac{1}{N}\sum_{i=1}^{N}\hat{y}_i^2}\right)$
U2	Theil U statistic 2 of forecasting results	$U2=\sqrt{\frac{1}{N}\sum_{i=1}^{N}((y_{i+1}-\hat{y}_{i+1})/y_i)^2}/\sqrt{\frac{1}{N}\sum_{i=1}^{N}((y_{i+1}-\hat{y}_i)/y_i)^2}$
MAPE	Mean Absolute Percentage Error	$MAPE=\frac{1}{N}\sum_{i=1}^{N}(\lvert y_i-\hat{y}_i\rvert/y_i)\times 100\backslash\%$
R^2	Coefficient of determination	$R^2=\sum_{i=1}^{N}(\hat{y}_i-\bar{y})^2/\sum_{i=1}^{N}(y_i-\bar{y})^2$

4.2 Experiment I: One-step-ahead forecasting comparison between our proposed combined model (CM) and other models

This experiment aimed to compare the performance of our forecasting model against some widely used statistical forecasting models, ANNs, and other established models. The forecasting results for one-step-ahead forecasting are shown in Table 2 and Figure 2. We can see that extreme learning machine (ELM) consistently obtained the optimal results at Sites 1, 2, and 3. At Site 4, the adaptive network-based fuzzy inference system (ANFIS) performance was optimal among the branch models. CM consistently realised the optimal performance of all models.

Table 2 and Figure 2 show the numerical simulation results and model performance, respectively.

1. At Site 1, for the branch models, ELM achieved the best results in terms of MAE, RMSE, U1, U2, and R2. The STDAPE value (4.96%) of ELM was 0.01% higher than those of ARIMA and ANFIS, which achieved a value of 4.95%. The DA ensures that the trend of the forecasting results is consistent with the true data, for this metric, ANFIS produced an optimal value of 43.20%. ANFIS also achieved the lowest MAE value, of 5.79%.

Table 2 One-step-ahead forecasting performance of four sites (the optimal results of the single models are highlighted in bold).

	Metric	ELM	ARIMA	ANFIS	GRNN	SVM	BPNN	ELMAN	RBFNN	LSTM	CM
Site 1	**MAE**	**0.4038**	0.406	0.4073	0.4239	0.6144	0.4213	0.9499	0.5367	0.7929	0.3511
	RMSE	**0.5225**	0.526	0.5682	0.6201	0.8261	0.5691	1.3259	0.712	0.95	0.4879
	STDAPE	4.96%	**4.95%**	**4.95%**	5.59%	7.70%	5.25%	13.52%	6.62%	10.28%	4.31%
	DA	42.83%	35.95%	**43.20%**	42.20%	27.41%	41.81%	21.35%	26.32%	24.81%	49.45%
	U1	**0.0343**	0.0349	0.0376	0.041	0.0546	0.0378	0.0883	0.0472	0.0636	0.0324
	U2	**1.4215**	1.4304	1.5252	1.7016	2.504	1.5819	5.8825	2.0123	3.7514	1.2332
	MAPE	5.81%	6.00%	**5.79%**	5.99%	9.02%	6.17%	14.18%	7.78%	10.55%	4.96%
	R^2	**0.9754**	**0.9754**	0.974	0.9691	0.9416	0.9725	0.842	0.9562	0.955601	0.9788
	Weight	0.6575	0.2248	0.1051	0.0093	-	0.0037	-	-	-	1
Site 2	**MAE**	**0.3794**	0.3831	0.4001	0.4491	0.4757	0.663	0.9143	0.5472	0.7353	0.3237
	RMSE	**0.5115**	0.5188	0.5797	0.7014	0.6458	0.955	1.3934	0.7425	1.0962	0.4636
	STDAPE	**4.64%**	4.87%	4.79%	7.19%	6.01%	9.25%	13.59%	6.69%	9.21%	4.02%
	DA	37.00%	39.92%	**46.47%**	44.39%	36.64%	28.20%	25.12%	26.42%	26.41%	53.92%
	U1	0.033	**0.0327**	0.0365	0.0441	0.0407	0.0603	0.0881	0.0468	0.0664	0.0292
	U2	**1.2064**	1.21	1.2465	1.7946	1.6963	2.7925	5.2131	1.9879	3.8267	0.9497
	MAPE	**5.36%**	5.49%	5.42%	6.22%	6.81%	9.26%	12.81%	7.77%	10.36%	4.45%
	R^2	0.9802	**0.9809**	0.9792	0.9707	0.975	0.9398	0.8601	0.9622	0.930856	0.9847
	Weight	0.6896	0.2395	0.0108	0.0538	0.0062	-	-	-	-	1
Site 3	**MAE**	**0.4358**	0.4418	0.4478	0.4752	0.6034	0.4995	0.999	0.5897	0.7468	0.3612
	RMSE	**0.5735**	0.5918	0.606	0.6983	0.822	0.6741	1.4849	0.7769	1.1704	0.5149
	STDAPE	5.56%	**5.33%**	5.37%	5.73%	7.25%	6.16%	13.79%	6.65%	9.41%	4.52%
	DA	33.82%	36.44%	**41.61%**	39.23%	28.40%	37.04%	24.03%	24.93%	25.84%	51.64%
	U1	**0.0316**	0.0356	0.0364	0.0418	0.0495	0.0405	0.0897	0.0469	0.0703	0.0311
	U2	**1.3986**	1.4247	1.5013	1.6408	2.1929	1.7757	5.5955	2.0144	3.6164	1.1925
	MAPE	**5.68%**	5.96%	5.96%	6.15%	8.15%	6.64%	13.26%	7.92%	11.61%	4.76%
	R^2	**0.9799**	0.9747	0.9747	0.9677	0.9507	0.9681	0.836	0.9562	0.931703	0.9805
	Weight	0.6896	0.2395	0.0108	0.0538	-	0.0062	-	-	-	1
Site 4	**MAE**	0.3874	0.391	**0.3604**	0.4004	0.4879	0.4493	0.8928	0.5042	0.6891	0.3209
	RMSE	0.5085	0.5123	**0.5068**	0.6207	0.6832	0.6027	1.2611	0.6657	1.0111	0.4567
	STDAPE	4.85%	4.87%	**4.67%**	5.71%	7.36%	6.49%	13.62%	6.88%	10.64%	4.06%
	DA	44.12%	36.44%	**48.16%**	46.08%	34.46%	36.15%	22.14%	28.20%	25.92%	52.63%
	U1	0.0332	0.0338	**0.0331**	0.0405	0.0448	0.0394	0.0835	0.0436	0.0665	0.0301
	U2	1.3828	1.3891	**1.3478**	1.5768	2.1616	1.7852	5.6962	2.0656	3.5943	1.1837
	MAPE	5.41%	5.72%	**5.08%**	5.60%	7.20%	6.74%	13.49%	7.52%	10.54%	4.49%
	R^2	**0.9816**	0.9771	0.9789	0.9689	0.9594	0.9699	0.8589	0.9617	0.9265	0.9817
	Weight	0.846	0.0858	0.0307	0.0348	0.0036	-	-	-	-	1

Note: CM: Proposed combined model.

2. For Site 2 in Table 2 and Figure 2, the optimal results of MAE, RMSE, STDAPE, U1, and MAPE were obtained by the ELM for the five single models. ARIMA achieved the best U1 and R2 values, at 0.0327 and 0.9809, respectively. ANFIS optimally ensured that the trend of forecasting results was consistent with the original data (with a value of 46.47% DA).

3. The inequality coefficients (U1 and U2) were also effective for measuring the predictive powers of the models. The ELM obtained better values (0.0316 and 1.3986 for U1 and U2, respectively) for Site 3 than the other models. ELM also performed best in terms of MAE, RMSE, MAPE, and R^2. ARIMA and ANFIS achieved the best value of STDAPE and DA, with values of 5.33% and 41.61%, respectively.

4. To illustrate the relationship between the true and forecasting values, R^2 was employed in this study, furthermore, in the results at Site 4, ELM performed better than the other single models. For the other metrics, ANFIS obtained the best results for all single models.

Remarks:

1. No single model could obtain the best results for all metrics in Table 2.

2. Except for Site 4, the ELM performed better than every other single model. The weight of the ELM was highest among the four combined models for the four datasets.

3. For the four different combined models of the four sites, the sub-models and weights were different. For the different datasets, no single model performed best. To achieve the best forecasting results, the sub-models had to be modified by the dataset.

4. For the combined model, five models with the lowest MAPE were selected as the single models in CM (e.g., the BPNN at Site 3, with a value of 6.74%). This is because the MAPE values for the single models were not the only indexes used to select the sub-models and build CM.

5. For the single models selected to build CM, each achieved one or more optimal scores according to the metrics.

4.3 Experiment II: Two-step-ahead forecasting comparison between our proposed model (CM) and other models

This experiment aimed to compare the two-step-ahead forecasting performance of CM with some widely used statistical forecasting models, ANNs, and other established models. Table 3 and Figure 3 show the two-step-ahead forecasting results for the four sites. From these results, ELM performed better than the other single models at the four sites (four datasets), and CM obtained the optimal values of all models under the eight metrics.

Table 3 and Figure 3 show the results, which are as follows:

1. For Site 1, LSTM achieved the optimal results for the MAE, DA, and MAPE, with values of 0.3257, 58.59%, and 4.81%, respectively. The other optimal values of the metrics were obtained by ELM. The weight of ELM was the highest among the single models, at 0.5689. The other models selected for CM were ARIMA, ANFIS, GRNN, and LSTM, with weights of 0.0199, 0.1149, 0.0568, and 0.2404, respectively.

2. ELMAN optimally ensured that the trend of the forecasting result was consistent with the true data, with a DA value of 55.61% at Site 2. ELM achieved the best values under the other metrics. The combined model was built using the ELM, ARIMA, ANFIS, SVM, and BPNN. ELMAN was not selected to build CM, although the forecasting trend of ELMAN more resembled the original data than the other single models.

3. In the forecasting results for the two-step-ahead forecast (shown in Table 3 and Figure 3), ELM performed best among the nine single models in seven metrics, except for DA. However, the weight of ELM did not exceed 0.5 (it was 0.4914). The trend of the results forecasted by BPNN most resembled that of the original data.

4. The MAE value can better reflect the actual forecasting value error, here, the BPNN obtained the lowest MAE value (0.3197). The BPNN also obtained the best result for DA, with a value of 61.67%. For the data of Site 4, ELM achieved the best values of RMSE, STDAPE, U1, U2, MAPE, and R2, with values of 0.4633, 4.13%, 0.0306, 1.2056, 4.71%, and 0.9813, respectively.

Table 3 Two-step-ahead forecasting performance of four sites (the optimal results of single models are highlighted in bold).

	Metric	ELM	ARIMA	ANFIS	GRNN	SVM	BPNN	ELMAN	RBFNN	LSTM	CM
Site 1	**MAE**	0.3554	0.4193	0.4065	0.4261	0.5541	0.5862	0.5939	0.5333	**0.3257**	0.3184
	RMSE	**0.5046**	0.5519	0.5605	0.7745	0.7654	0.8129	0.7953	0.7069	0.7628	0.4610
	STDAPE	**4.50%**	5.37%	5.10%	7.28%	7.28%	7.64%	7.65%	6.60%	11.39%	4.60%
	DA	48.96%	35.05%	45.98%	45.18%	31.08%	29.31%	28.92%	26.32%	**58.59%**	55.51%
	U1	**0.0336**	0.0367	0.0371	0.0512	0.0508	0.0512	0.0539	0.0469	0.0505	0.0306
	U2	**1.2511**	1.3907	1.4880	1.9633	2.2347	2.3641	2.4449	1.9858	2.2126	1.0712
	MAPE	5.03%	6.20%	5.74%	6.04%	7.98%	8.32%	8.02%	7.75%	**4.81%**	4.53%
	R^2	**0.9775**	0.9729	0.9737	0.9511	0.9497	0.9116	0.9401	0.9565	0.9487	0.9812
	Weight	0.5689	0.0199	0.1149	0.0568	-	-	-	-	0.2404	1
Site 2	**MAE**	**0.3297**	0.3894	0.4035	0.4881	0.5975	0.4287	0.3518	0.5414	0.5459	0.3128
	RMSE	**0.4727**	0.5258	0.5900	0.8728	0.8368	0.5988	0.7368	0.7342	0.7386	0.4430
	STDAPE	**4.09%**	4.96%	5.11%	10.90%	7.56%	5.52%	10.82%	6.62%	6.81%	4.06%
	DA	54.42%	40.81%	46.67%	42.40%	28.30%	37.84%	**55.61%**	27.51%	26.31%	55.91%
	U1	**0.0299**	0.0332	0.0371	0.0548	0.0530	0.0377	0.0464	0.0464	0.0464	0.0280
	U2	**0.9023**	1.1469	1.2156	2.7947	2.3875	1.3770	1.9866	1.9483	2.1022	0.8065
	MAPE	**4.50%**	5.58%	5.54%	6.80%	8.34%	5.89%	5.09%	7.70%	7.96%	4.32%
	R^2	**0.9842**	0.9806	0.9785	0.9580	0.9541	0.9755	0.9623	0.9628	0.9377	0.9861
	Weight	0.5249	0.1473	0.1003	-	0.0297	0.1983	-	-	-	1
Site 3	**MAE**	**0.3628**	0.4406	0.4673	0.4964	0.5545	0.3755	0.5336	0.5830	0.5995	0.3457
	RMSE	**0.5209**	0.5845	0.6460	0.9366	0.7379	0.9203	0.7057	0.7673	0.7824	0.4865
	STDAPE	**4.53%**	5.38%	5.95%	8.16%	6.63%	13.33%	6.55%	6.62%	7.20%	4.54%
	DA	53.23%	37.34%	40.71%	41.51%	33.27%	**59.48%**	33.17%	24.33%	22.36%	52.63%
	U1	**0.0315**	0.0354	0.0387	0.0559	0.0443	0.0553	0.0424	0.0463	0.0479	0.0294
	U2	**1.1576**	1.3105	1.5529	2.1317	2.0285	2.9943	1.9707	1.9807	2.0679	1.0269
	MAPE	**4.74%**	5.97%	6.23%	6.43%	7.52%	5.19%	7.35%	7.83%	8.12%	4.60%
	R^2	**0.9802**	0.9752	0.9716	0.9447	0.9640	0.9384	0.9647	0.9575	0.9359	0.9828
	Weight	0.4914	0.2073	0.0994	0.0313	-	0.1713	-	-	-	1.0000
Site 4	**MAE**	0.3321	0.3917	0.3872	0.4129	0.5514	**0.3197**	0.4459	0.5053	0.5509	0.3044
	RMSE	**0.4633**	0.5158	0.5398	0.6247	0.7794	0.6585	0.5729	0.6663	0.6975	0.4258
	STDAPE	**4.13%**	5.07%	4.85%	5.94%	8.91%	12.51%	5.90%	6.92%	7.04%	4.61%
	DA	50.35%	37.44%	45.48%	44.59%	30.39%	**61.67%**	35.05%	28.70%	28.57%	53.13%
	U1	**0.0306**	0.0340	0.0352	0.0408	0.0512	0.0432	0.0376	0.0437	0.0467	0.0280
	U2	**1.2056**	1.3233	1.3914	1.5893	2.5146	2.5806	1.7637	2.0813	2.1797	1.0322
	MAPE	**4.71%**	5.77%	5.46%	5.72%	8.20%	5.04%	6.79%	7.57%	8.27%	4.40%
	R^2	**0.9813**	0.9767	0.9768	0.9685	0.9463	0.9623	0.9731	0.9613	0.9227	0.9843
	Weight	0.5994	0.0967	0.0113	0.0366	-	0.2560	-	-	-	1.0000

Note: CM: Proposed combined model.

Remarks:

1. For the two-step-ahead forecasting (see Table 3), the optimal results in all metrics could not be obtained from any single model.

2. The forecasting results of ELM almost achieved the optimal values at the four sites, as shown in Table 3 and Figure 3. The ELM weights selected to build CM for the four sites were 0.5698, 0.5249, 0.4914, and 0.5994, respectively.

3. At Site 2, ELMAN achieved the best DA value, however, the CM for Site 2 was not built by ELMAN. This indicates that a single standard or metric cannot determine whether a single model should be selected to construct CM.

4. For Sites 3 and 4, the models selected to build CM were the same. The single models were ELM, ARIMA, ANFIS, GRNN, and BPNN. However, the weights of these models for Sites 3 and 4 were different. This indicates that the CMs for these two sites differed.

5. For the four sites or four datasets, the CMs differed.

6. ELMAN and BPNN were not selected to construct the CMs for the four sites.

4.4 Experiment III: Three-step-ahead forecasting comparison between our proposed model (CM) and other models

This experiment aimed to compare the three-step-ahead forecasting performance of CM with several widely used statistical forecasting models, ANNs, and other established models. ELM and BPNN outperformed the other single models at the four sites (four datasets).

Table 4 and Figure 4 show the results, which can be summarised as follows:

1. From the results of Site 1, the proportion of variation in the dependent variable of indicator R^2 can be explained by the independent variable via a regression relationship. LSTM achieved the optimal value (0.9709). The optimal values of U1 and RMSE were also achieved by LSTM, at 0.6707 and 32.96%, respectively. The RMSE measures the deviation between the observed and real data. It is often used as a standard to measure the forecasting results of machine-learning models. ELM achieved the optimal values under the other five metrics: MAE, STDAPE, U1, U2, and MAPE.

2. For the experiment at Site 2, ELM and SVM were the two best of the nine single models, and they obtained the best values in five and three metrics, respectively. Although SVM only achieved the three best values of the metrics (lower than ELM), the weight of the SVM used to build CM was 0.5055, higher than that of ELM.

3. At Site 3, the BPNN performed best among the nine single models, and the weight of the BPNN used to build CM was 0.4829, which was the highest among the nine single models.

4. Similar to Site 3, an optimal single model was observed at Site 4, ELM had an overwhelming advantage among the nine models. The CM of Site 4 was built using ELM, ARIMA, ANFIS, GRNN, and LSTM.

Remarks:

1. For the three-step ahead forecasting from table 4, the best results of all metrics in four sites could not be obtained by any single model.

Table 4 Three-step-ahead forecasting performance of the four sites (the optimal results of single models are highlighted in bold)

	Metric	ELM	ARIMA	ANFIS	GRNN	SVM	BPNN	ELMAN	RBFNN	LSTM	CM
Site 1	**MAE**	**0.4884**	0.5139	0.5214	0.5519	0.7146	0.6197	0.7113	0.6104	0.5008	0.3230
	RMSE	0.6734	0.6776	0.7135	0.7740	0.9383	0.8069	0.8058	0.8066	**0.6707**	0.5048
	STDAPE	**6.26%**	6.90%	6.56%	7.86%	9.79%	7.89%	9.57%	7.65%	6.73%	5.25%
	DA	30.09%	26.42%	30.19%	29.49%	23.63%	25.72%	22.38%	21.75%	**32.96%**	53.53%
	U1	**0.0448**	0.0451	0.0472	0.0513	0.0624	0.0540	0.0578	0.0535	0.0458	0.0335
	U2	**1.7701**	1.8702	1.9441	2.3163	3.6966	2.6244	2.9395	2.4018	1.7841	0.9911
	MAPE	**7.05%**	7.71%	7.43%	7.96%	10.94%	9.25%	10.05%	8.94%	7.38%	4.67%
	R^2	0.9593	0.9590	0.9558	0.9484	0.9318	0.9417	0.9282	0.9428	**0.9709**	0.9774
	Weight	0.1672	0.0877	0.1042	0.0231					0.6189	
Site 2	**MAE**	0.4342	0.4792	0.5060	0.5645	**0.3755**	0.4574	0.4616	0.6165	0.8692	0.3732
	RMSE	**0.6165**	0.6674	0.7186	0.8242	0.9316	0.6978	0.7519	0.8374	1.2517	0.5426
	STDAPE	**5.53%**	6.39%	6.81%	8.07%	7.31%	7.58%	8.41%	7.65%	14.25%	5.68%
	DA	40.02%	33.66%	34.26%	34.16%	**54.42%**	43.03%	40.13%	23.04%	24.73%	47.17%
	U1	**0.0390**	0.0421	0.0453	0.0519	0.0585	0.0594	0.0575	0.0530	0.0786	0.0342
	U2	**1.4324**	1.6120	2.0153	2.4036	3.6572	2.6758	2.5493	2.3707	5.2523	1.2270
	MAPE	5.98%	6.90%	7.02%	7.83%	**5.53%**	6.95%	7.32%	8.75%	12.49%	5.19%
	R^2	**0.9729**	0.9683	0.9669	0.9593	0.9397	0.9587	0.9251	0.9510	0.8864	0.9790
	Weight	0.2687	0.0237	0.1641		0.5055	0.0390				
Site 3	**MAE**	0.5030	0.5353	0.5614	0.5815	0.9965	**0.4711**	0.6284	0.6600	0.5888	0.3330
	RMSE	0.7082	0.7237	0.7727	0.7937	1.4372	**0.7045**	1.0911	0.8691	0.7919	0.7079
	STDAPE	6.37%	6.78%	6.85%	6.84%	10.09%	**5.90%**	9.46%	7.53%	6.85%	5.46%
	DA	34.66%	30.78%	32.87%	31.98%	19.66%	**48.46%**	42.50%	21.45%	27.31%	60.28%
	U1	0.0429	0.0437	0.0464	0.0476	0.0885	**0.0433**	0.0671	0.0525	0.0478	0.0426
	U2	1.7606	1.7663	2.0134	2.0101	3.9966	**1.6841**	2.8524	2.3686	2.0938	1.9161
	MAPE	6.69%	7.30%	7.42%	7.72%	12.12%	**6.50%**	8.68%	8.88%	7.82%	4.62%
	R^2	0.9629	0.9612	0.9583	0.9574	0.8428	**0.9633**	0.9481	0.9450	0.9543	0.9782
	Weight	0.1435	0.1969	0.1402			0.4829			0.0348	
Site 4	**MAE**	**0.4571**	0.4971	0.4985	0.5319	0.6071	0.5294	0.5132	0.5897	0.4768	0.3570
	RMSE	**0.6281**	0.6483	0.6814	0.7783	0.8205	0.6871	0.7820	0.7793	0.6878	0.5119
	STDAPE	**0.0611**	0.0679	0.0735	0.0775	0.0862	0.0698	0.0835	0.0843	0.0732	0.0647
	DA	**34.56%**	29.39%	32.97%	32.87%	27.41%	27.51%	29.82%	22.94%	30.08%	46.47%
	U1	**0.0415**	0.0427	0.0446	0.0509	0.0536	0.0452	0.0507	0.0512	0.0451	0.0337
	U2	**1.7722**	1.9064	2.2269	2.3708	2.8225	2.1802	2.3534	2.7017	2.4159	1.3093
	MAPE	**6.59%**	7.45%	7.34%	7.69%	9.12%	7.96%	7.78%	8.94%	7.23%	5.25%
	R^2	**0.9652**	0.9629	0.9606	0.9488	0.9432	0.9599	0.9367	0.9465	0.9591	0.9772
	Weight	0.4405	0.0884	0.0703	0.0467					0.3513	

Note: CM: Proposed combined model.

2. ELM, ARIMA, and ANFIS were selected to build the three-step-ahead forecasting CM for the four sites, which indicated that these four models were fit for three-step-ahead forecasting and the four datasets.

3. For the four sites, the CMs were built with different single models, to obtain the best CM for the forecasting results.

4. The weight of LSTM was 0.6189, greater than the weight of ELM, however, ELM obtained the best metric scores. Thus, the weights of the branch models were unrelated.

5. For Sites 1 and 4, the single models of the combined model were identical. The single models were ELM, ARIMA, ANFIS, GRNN, and LSTM. However, the weights of these models for Sites 1 and 4 differed. This indicates that the CMs for these two sites were different. Therefore, for the four sites or four datasets, the CMs chosen by our proposed method differed.

6. ELMAN and RBFNN were not selected to build CMs for the four sites.

7. For constructing CM, five models with the lowest MAPE were selected, including the ELMAN at Site 4 (with a value of 6.59%). This is because the MAPE of the single models did not propose a single index to help select the branch model of the CM.

4.5 Experiment IV: DM test and forecasting availability

To further evaluate our CM, we used two evaluation methods–the DM test and forecasting availability–to evaluate the model quality.

DM testing, proposed by Diebold and Mariano [39], focuses on forecasting accuracy and evaluates the forecasting performance of two or more time-series models, as well as forecasting availability [40]. The effectiveness of the forecasting was measured by the sum of the squares and the mean square deviation of the forecasting errors, to further evaluate and analyse the performance of CM. Among all models, the optimal performance was achieved by CM.

The results of the DM test and forecasting availability are shown in Tables 5 and 6, respectively.

(1) As shown in Table 5, CM differed considerably from the other models, regardless of the dataset or order.

(2) The results of forecasting availability are listed in Table 6. In wind speed forecasting, the first- and second-order forecasting availability of CM for the four datasets and one-step, two-step, and three-step forecasting outperformed those of the other models. For example, at Site 1 (one-step forecasting), the first-order forecasting availabilities of each model were 0.9419, 0.9400, 0.9421, 0.9401, 0.9098, 0.9383, 0.8582, 0.9222, 0.8945, and 0.9504, respectively.

Remark:

(1) The results of the DM test showed that the CM differed from other models. The higher the value, the greater this difference.

(2) The forecasting availability results show the differences between CM and other models. The higher the value, the greater the difference. The results show that the first- and second-order values of the CMs are close to 1, which indicates that CM is significantly better than the other models.

Table 5 Results for the Diebold and Mariano (DM) test.

One-Step				
Site	Site 1	Site 2	Site 3	Site 4
ELM vs CM	1.3065*	0.1395*	0.1661*	1.4595*
ARIMA vs CM	1.6757*	1.6586*	1.6739*	1.8051*
ANFIS vs CM	0.0274*	1.3118*	1.4458*	0.6710*
GRNN vs CM	1.5687*	1.7102*	1.7210**	1.6786*
SVM vs CM	2.4982**	1.9551*	2.6040**	2.2348**
BPNN vs CM	1.8466*	2.5931***	1.9504*	2.0062**
ELMAN vs CM	6.7752***	6.8146***	6.8772***	7.0040***
RBFNN vs CM	2.0196**	2.1254*	2.1854*	2.6704***
LSTM vs CM	3.3721***	4.5251***	4.8914***	6.0206***
Two-Step				
Site	Site 1	Site 2	Site 3	Site 4
ELM vs CM	1.2993*	0.1374*	0.1637*	0.6674*
ARIMA vs CM	1.8668*	1.6942*	1.6809*	1.9734**
ANFIS vs CM	1.5772*	1.6655*	1.7305*	1.6777*
GRNN vs CM	1.6942*	2.0944**	1.9513*	1.8093*
SVM vs CM	2.4858*	6.3136***	2.5871***	5.7858***
BPNN vs CM	6.2349***	1.9270*	1.4301*	1.4517*
ELMAN vs CM	3.2725***	1.3073*	2.1733**	2.2234**
RBFNN vs CM	2.0444**	2.5292*	4.6024***	2.6808***
LSTM vs CM	0.0271*	4.3469***	6.3290***	6.3992***
Three-Step				
Site	Site 1	Site 2	Site 3	Site 4
ELM vs CM	0.0282*	1.3041*	1.4711*	0.6785*
ARIMA vs CM	1.7416*	1.6725*	1.7229*	1.8285*
ANFIS vs CM	1.6226*	1.9251*	1.7691*	1.7009*
GRNN vs CM	1.9223*	2.4704**	1.9977**	2.0000**
SVM vs CM	6.5018***	0.1377*	6.7895**	6.3215***
BPNN vs CM	2.5322**	1.7014*	0.1693*	2.6339***
ELMAN vs CM	3.3788***	2.0647**	2.6412***	2.2112**
RBFNN vs CM	2.0801**	4.2901**	4.6880***	5.7065***
LSTM vs CM	1.3544*	6.56452***	2.1912**	1.4819*

Note: *: 1% significant difference, **: 5% significant difference, ***: 10% significant difference, CM: our proposed combined model.

Table 6 Forecasting availability results (the optimal results of single models are highlighted in bold).

Site	Forecasting Availabilities	Forecasting Step	ELM	ARIMA	ANFIS	GRNN	SVM	BPNN	ELMAN	RBFNN	LSTM	CM
Site 1	1-Order	One-Step	0.9419	0.9400	0.9421	0.9401	0.9098	0.9383	0.8582	0.9222	0.8945	**0.9504**
		Two-Step	0.9497	0.9380	0.9426	0.9396	0.9202	0.9168	0.9198	0.9225	0.9519	**0.9547**
		Three-Step	0.9295	0.9229	0.9257	0.9204	0.8906	0.9075	0.8995	0.9106	0.9262	**0.9533**
	2-Order	One-Step	0.9147	0.9121	0.9159	0.9126	0.8694	0.9098	0.7993	0.8872	0.8474	**0.9270**
		Two-Step	0.9259	0.9093	0.9158	0.9112	0.8847	0.8797	0.8843	0.8880	0.9293	**0.9336**
		Three-Step	0.8970	0.8881	0.8917	0.8841	0.8437	0.8665	0.8559	0.8711	0.8921	**0.9314**
Site 2	1-Order	One-Step	0.9464	0.9451	0.9458	0.9378	0.9319	0.9074	0.8719	0.9223	0.8964	**0.9555**
		Two-Step	0.9550	0.9442	0.9446	0.9320	0.9166	0.9411	0.9491	0.9230	0.9204	**0.9568**
		Three-Step	0.9402	0.9310	0.9298	0.9217	0.9447	0.9305	0.9268	0.9125	0.8751	**0.9481**
	2-Order	One-Step	0.9219	0.9201	0.9202	0.9093	0.9003	0.8665	0.8165	0.8877	0.8507	**0.9349**
		Two-Step	0.9338	0.9187	0.9194	0.9011	0.8799	0.9136	0.9254	0.8887	0.8853	**0.9364**
		Three-Step	0.9124	0.8994	0.8974	0.8857	0.9190	0.8987	0.8929	0.8739	0.8213	**0.9241**
Site 3	1-Order	One-Step	0.9432	0.9404	0.9404	0.9385	0.9185	0.9336	0.8674	0.9208	0.8839	**0.9524**
		Two-Step	0.9526	0.9403	0.9377	0.9357	0.9248	0.9481	0.9265	0.9217	0.9188	**0.9540**
		Three-Step	0.9331	0.9270	0.9258	0.9228	0.8788	0.9350	0.9132	0.9112	0.9218	**0.9538**
	2-Order	One-Step	0.9174	0.9132	0.9125	0.9100	0.8823	0.9031	0.8121	0.8848	0.8328	**0.9299**
		Two-Step	0.9304	0.9131	0.9089	0.9067	0.8901	0.9241	0.8937	0.8868	0.8825	**0.9324**
		Three-Step	0.9021	0.8933	0.8925	0.8874	0.8270	0.9051	0.8749	0.8708	0.8861	**0.9322**
Site 4	1-Order	One-Step	0.9459	0.9428	0.9492	0.9440	0.9280	0.9326	0.8651	0.9248	0.8946	**0.9551**
		Two-Step	0.9529	0.9423	0.9454	0.9428	0.9180	0.9496	0.9321	0.9243	0.9173	**0.9560**
		Three-Step	0.9341	0.9255	0.9266	0.9231	0.9088	0.9204	0.9222	0.9106	0.9277	**0.9475**
	2-Order	One-Step	0.9204	0.9168	0.9254	0.9175	0.8955	0.9021	0.8090	0.8902	0.8488	**0.9337**
		Two-Step	0.9313	0.9161	0.9198	0.9162	0.8807	0.9260	0.9012	0.8901	0.8806	**0.9353**
		Three-Step	0.9034	0.8912	0.8927	0.8890	0.8676	0.8852	0.8870	0.8706	0.8949	**0.9229**

Note: The higher value, the better the forecasting effectiveness of the model.

CM: Our proposed combined model.

4.6 Summary

From the four experiments, we obtained the following findings:

(1) The sub-models of CM were not static. Across the three different datasets, four sites, and different step-ahead-forecasting scenarios, the preferred model was always varied to achieve the best results.

(2) No single model could consistently achieve the best results among the sub-models, owing to the complexity of the data.

(3) The multi-objective optimisation algorithm was used to optimise the weights of the combined model's sub-models from the former three experiments. The weights of the sub-models in CM differed. In addition, the multi-objective optimisation algorithm could balance the objective functions of the combined model, thereby ensuring the accuracy and effectiveness of the forecasting.

(4) The model selection chose the best sub-models to construct CM in different situations (datasets, forecasting steps, and sites), this made the selection of the sub-models more reasonable, and the combined model helped achieve the best results.

Our experiments demonstrate that CM has a stronger forecasting power and higher forecasting accuracy than the benchmark model.

5. Conclusion

In this study, our proposed combined model obtained an optimal result compared to the other models. There was no single ANN that could perfectly solve the problem for different levels of step-ahead forecasting and different datasets. Furthermore, based on the experiments, certain innovations of the forecast system developed in this study [i.e., our proposed model (CM)] considered not only different time and site data but also the disadvantages of the combined model. To overcome these disadvantages, a model was proposed using multiple objective functions, and the model selection theory and innovations can be summarised as follows:

(1) The multi-objective functions guaranteed the stability and accuracy of the CM's results, because they considered both aspects. (2) The model selection theory made the single models of the CM more reasonable, instead of manually selecting single models. (3) Our proposed model confirmed that for different data and different step-ahead forecasting processes, the optimal combined model was not fixed. For these results, it was necessary to adjust the sub-models used to construct CM for the different data. These developments in our model are rarely seen in other studies; therefore, this study fills that research gap.

To summarise, by overcoming the disadvantages and making innovations, our proposed model based on multiple objective functions and model selection was found to be stable and accurate (the MAPE is less than 4.60% and the STDAPE is less than 5.68%); it overcame the difficulty of selecting sub-models for the combined model. With the forecasting results and theories, CM can also be applied to futures, forwards, securities, house prices, and other forecasting fields. More benchmark models should be added to the model selection, to help CM achieve better results. However, more models will increase the model runtime; thus, the number of models and runtime should be kept balanced in future studies.

Declarations

Acknowledgements

This work was supported by Western Project of the National Social Science Foundation of China (Grant No.18XTJ003).

Compliance with ethical standards

Conflict of interest: The authors declare that they have no conflict of interest.

References

1. Wang X, Sideratos G, Hatziargyriou N et al (2004) Wind speed forecasting for power system operational planning, International Conference on Probabilistic Methods Applied to Power Systems. IEEE, 470–474
2. Enrique R, David B, Jorge-Juan B, Ana P (2019) Review of wind energy technology and associated market and economic conditions in Spain, Renew. Sustain. Energy Rev. 101; 415–427
3. Aguilar Vargas S, Telles Esteves GR, Medina Maçaira P, Quaresma Bastos B, Cyrino Oliveira FL (2019) Castro Souza, Wind power generation: a review and a research agenda. J Clean Prod 218:850e870
4. Global Wind Energy Council. Global wind statistics (2019) p. 2019
5. Fried L, Qiao L, Sawyer S Global wind report, global wind energy council. https://gwec.net/members-area-market-intelligence/reports/
6. Meng K, Yang H, Dong ZY, Guo W, Wen F, Xu Z (2016) Flexible operational planning framework considering multiple wind energy forecasting service providers. IEEE Trans Sustain Energy 7(2):708–717
7. Yan J, Li F, Liu Y, Gu C (2017) Novel cost model for balancing wind power fore- casting uncertainty. IEEE Trans Energy Convers 32(1):318–329
8. Bruninx K, Bergh KVD, Delarue E et al (2016) Optimization and allocation of spinning reserves in a low-carbon framework, IEEE Power and Energy Society General Meeting (PESGM). IEEE Trans Power Syst 31(2):872–882
9. Chang W-Y (2014) A literature review of wind forecasting methods.Power Energy Eng. 2,161–168
10. Soman SS, Zareipour H, Malik O et al (2010) A review of wind power and wind speed forecasting methods with different time horizons[C]// North American Power Symposium (NAPS), IEEE, 2010. 1–8
11. Sfetsos A (2000) A comparison of various forecasting techniques applied to mean hourly wind speed time series. Renew Energy 21(1):23–35

12. Lei M, Shiyan L, Chuanwen J, Hongling L, Yan Z (2009) A review on the forecasting of wind speed and generated power. Renew Sustain Energy Rev 13(4):915–920
13. Smith DA, Mehta KC (1993) Investigation of stationary and nonstationary wind data using classical Box-Jenkins models. J Wind Eng Indus Aerodyn 49:319–328
14. Riahy G, Abedi M (2008) Short term wind speed forecasting for wind turbine applications using linear prediction method. Renew Energy 33:35–41
15. Torres JL, Garca A, Blas MD, DeFrancisco A (2005) Forecast of hourly average wind speed with arma models in navarre (Spain). Sol Energy 79(1):65–77
16. Brown BG, Katz RW, Murphy AH (1984) Time series models to simulate and forecast wind speed and wind power. J Appl Meteorol 23:1184–1195
17. Grigonyte E, Butkeviciute E (2016) Short-term wind speed forecasting using ARIMA model. Energetika 62(1–2):45–55
18. Shamshad A, Bawadi M, Hussin WW, Majid T, Sanusi S (2005) First and second order Markov chain models for synthetic generation of wind speed time se- ries. Energy 30(5):693–708
19. Damousis IG, Alexiadis MC, Theocharis JB, Dokopoulos PS (2004) A fuzzy model for wind speed prediction and power generation in wind parks using spatial correlation. IEEE Trans Energy Convers 19(2):352–361
20. Zhang W, Qu Z, Zhang K, Mao W, Ma Y, Fan X (2017) A combined model based on CEEMDAN and modified flower pollination algorithm for wind speed forecasting. Energy Convers Manag 136:439–451
21. Niu T, Wang J, Zhang K, Du P (2018) Multi-step-ahead wind speed forecasting based on optimal feature selection and a modified bat algorithm with the cognition strategy. Renew Energy 118:213–229
22. Heng J, Hong Y, Hu J, Wang S (2022) Probabilistic and deterministic wind speed forecasting based on non-parametric approaches and wind characteristics information. Appl Energy 306:118029
23. Guo ZH, Wu J, Lu HY, Wang JZ (2011) A case study on a hybrid wind speed forecasting method using BP neural network. Knowl -Based Syst 24:1048–1056
24. Specht DF (1991) A general regression neural network. IEEE Trans Neural Netw 2(6):568–576
25. Wang J, Zhang W, Wang J et al (2014) A novel hybrid approach for wind speed prediction. Inf Sci 273:304–318
26. Dorvlo AS, Jervase JA, Al-Lawati A (2002) Solar radiation estimation using artificial neural networks. Appl Energy 71(4):307–319
27. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9:1735–1780
28. Li G, Shi J (2010) On comparing three artificial neural networks for wind speed forecasting. Appl Energy 87(7):2313–2320
29. Yu C, Li Y, Zhang M (2017) An improved wavelet transform using singular spectrum analysis for wind speed forecasting based on elman neural network. Energy Convers Manag 148:895–904
30. Wang S, Zhang N, Wu L et al (2016) Wind speed forecasting based on the hybrid ensemble empirical mode decomposition and GA-BP neural network method. Renew Energy 94:629–636
31. Yang Y, Chen Y, Wang Y, Li C, Li L (2016) Modelling a combined method based on ANFIS and neural network improved by DE algorithm: a case study for short- term electricity demand forecasting. Appl Soft Comput. https:// doi.org/10.1016/j.asoc.2016.07.053
32. Xiao L, Wang J, Dong Y, Wu J (2015) Combined forecasting models for wind energy forecasting: a case study in China, Renew. Sustain Energy Rev 44:271–288. https://doi.org/10.1016/j.rser.2014.12.012
33. Zhang S, Wang C, Liao P, Xiao L, Fu T (2022) Wind Speed Forecasting Based on Model Selection, Fuzzy Cluster, and Multi-Objective Algorithm and Wind Energy Simulation by Betz's Theory. Expert Systems with Applications, p 116509
34. Wang C, Zhang S, Xiao L, Fu T (2021) Wind speed forecasting based on multi-objective grey wolf optimisation algorithm, weighted information criterion, and wind energy conversion system: A case study in Eastern China. Energy Conv Manag 243:114402
35. Xiao L, Dong Y, Dong Y (2018) An improved combination approach based on Adaboost algorithm for wind speed time series forecasting. Energy Conv Manag 160:273–288
36. Deb K, Jain H (2014) An evolutionary many-objective optimization algorithm using reference-point based nondominated sorting approach, part I: solving problems with box constraints. IEEE Trans Evol Comput 18(4):577–601
37. Abdollahzade M, Miranian A, Hassani H, Iranmanesh H (2015) A new hybrid enhanced local linear neuro-fuzzy model based on the optimized singular spectrum analysis and its application for nonlinear and chaotic time series forecasting. Inf Sci 295:107–125
38. Wang JZ, Yang WD, Du P, Niu T (2018) A novel hybrid forecasting system of wind speed based on a newly developed multi-objective sine cosine algorithm. Energy Conv Manag 163:134–150
39. Diebold FX, Mariano R (1995) Comparing predictive accuracy. J Bus Econ Stat 20(1):134–144

40. Zhang S, Wang J, Guo Z, Research (2018) on combined model based on multi-objective optimization and application in time series forecast Soft Computing https://doi.org/10.1007/s00500-018-03690-w
41. Vapnik V The nature of statistic learning theory.Springer, Berlin Heidelberg New York
42. Greff K, Srivastava RK, Koutnik J, Steunebrink BR, Schmidhuber J (2017) LSTM: a search space odyssey. IEEE Trans Neural Netw Learn Syst 28:2222–2232
43. Gers FA, Schmidhuber J (2000) Recurrent nets that time and count. Ieee-Inns-Enns Int Jt Conf Neural Netw 3:189–194
44. Contreras J, Espinola R, Nogales F, Conejo A (2003) ARIMA models to predict next-day electricity prices. IEEE Trans Power Syst 18(3):1014–1020
45. Schwenker F, Kestler HA, Palm G (2001) Three learning phases for radial-basis- function networks. Neural Netw 14(4–5):439–458
46. Jang J-SR (1993) ANFIS: adaptive-network-based fuzzy inference system. IEEE Trans Syst Man Cybernetics 23(3):665–685
47. Neshat M, Adeli A, Sepidnam G (2012) Predication of concrete mix design using adaptive neural fuzzy inference systems and fuzzy inference systems. Int J Adv Manuf Technol 63(1–4):373–390
48. İnan G, GöKtepe AB, Ramyar K et al (2007) Prediction of sulfate expansion of PC mortar using adaptive neuro-fuzzy methodology. Building & Environment 42(3):1264–1269
49. Liu M, Ling YY (2003) Using fuzzy neural network approach to estimate contractors' markup. Build Environ 38(11):1303–1308
50. Wu J, Hsu C, Chen H (May 2009) An expert system of price forecasting for used cars using adaptive neuro-fuzzy inference. Expert Syst Appl Volume 36(4):7809–7817
51. Huang GB, Zhu QY, Siew CK (2006) Extreme learning machine: theory and applications. Neurocomputing 70(1–3):489–501
52. Elman JL (1990) Finding Structure in Time. Cogn Sci 14:179–211
53. Egrioglu E, Aladag CH, Günay S (2008) A new model selection strategy in artificial neural networks. Appl Math Comput 195:591–597
54. Bates JM, Granger CWJ (2001) The combination of forecasts. Essays in Econometrics. Cambridge University Press, Cambridge, UK, pp 451–468
55. Laumanns M, Thiele L, Zitzler E (2006) An efficient, adaptive parameter variation scheme for metaheuristics based on the epsilon-constraint method. Eur J Oper Res 169(3):932–942
56. Liu Z, Jiang P, Wang J, Zhang L (2022) Ensemble system for short term carbon dioxide emissions forecasting based on multi-objective tangent search algorithm. J Environ Manage 302:113951
57. Fu T, Zhang S, Wang C (2020) Application and research for electricity price forecasting system based on multi-objective optimization and sub-models selection strategy[J]. Soft Comput 24(20):15611–15637

Figures

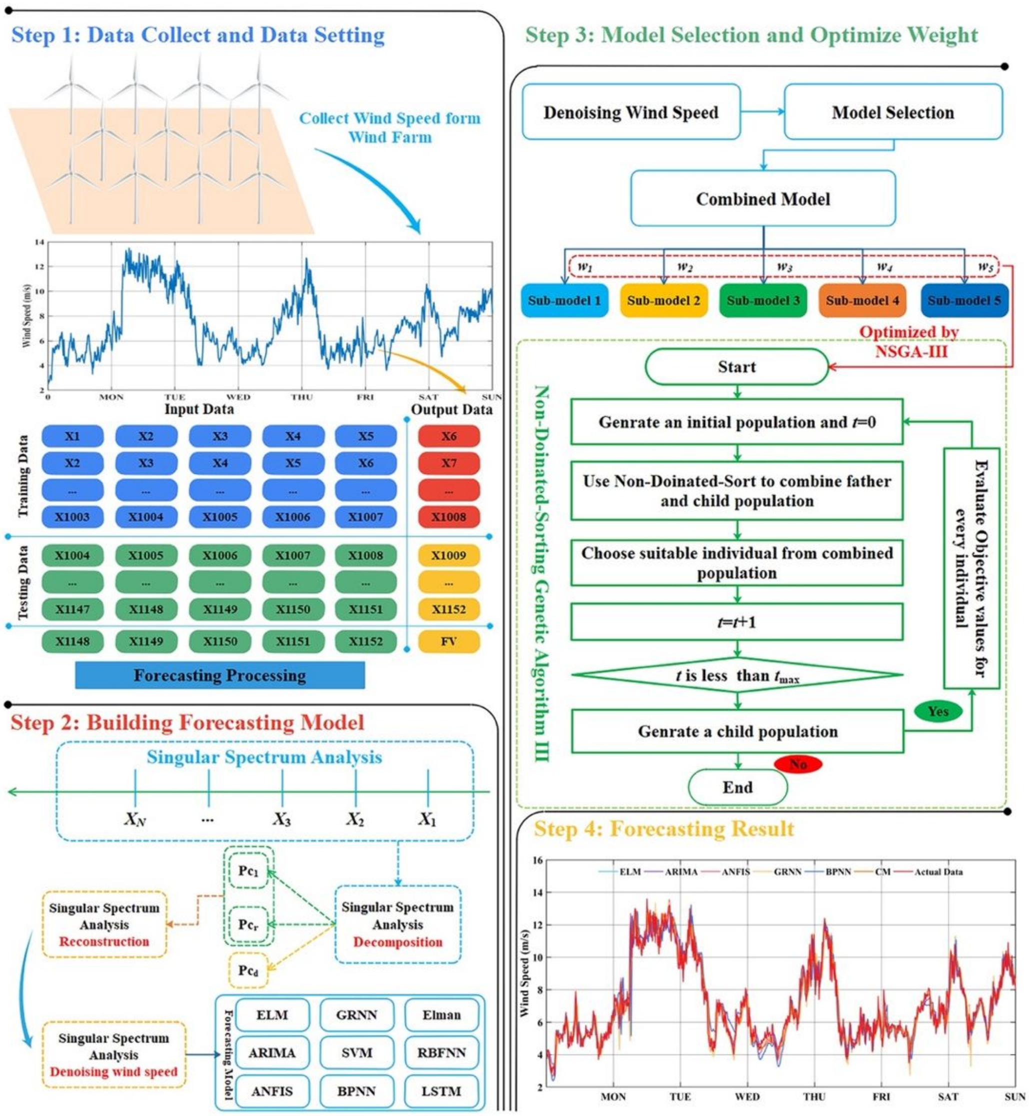

Figure 1

Flow chart of our proposed model

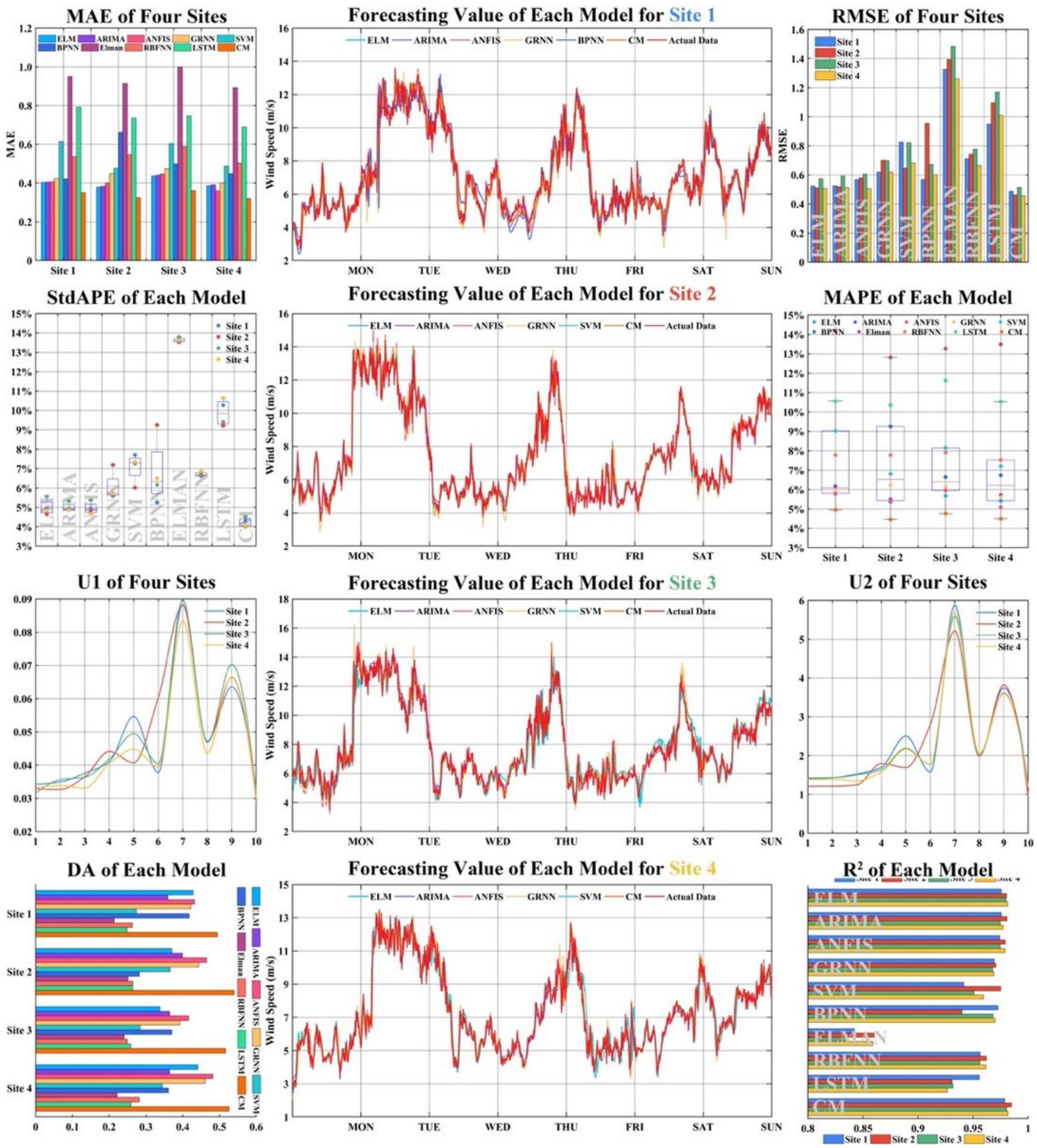

Figure 2

Forecasting results for the one-step ahead forecast

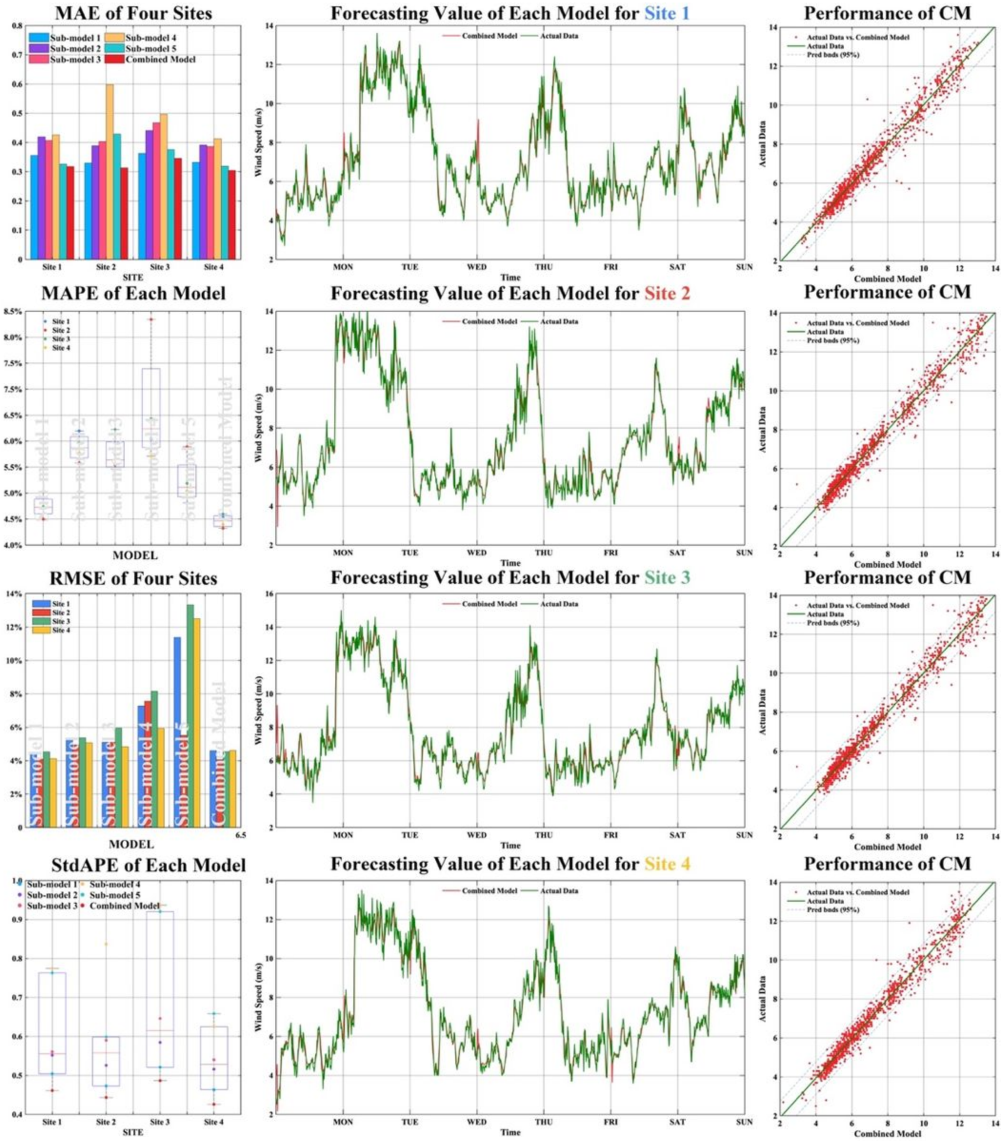

Figure 3

Forecasting results of two-step-ahead forecast

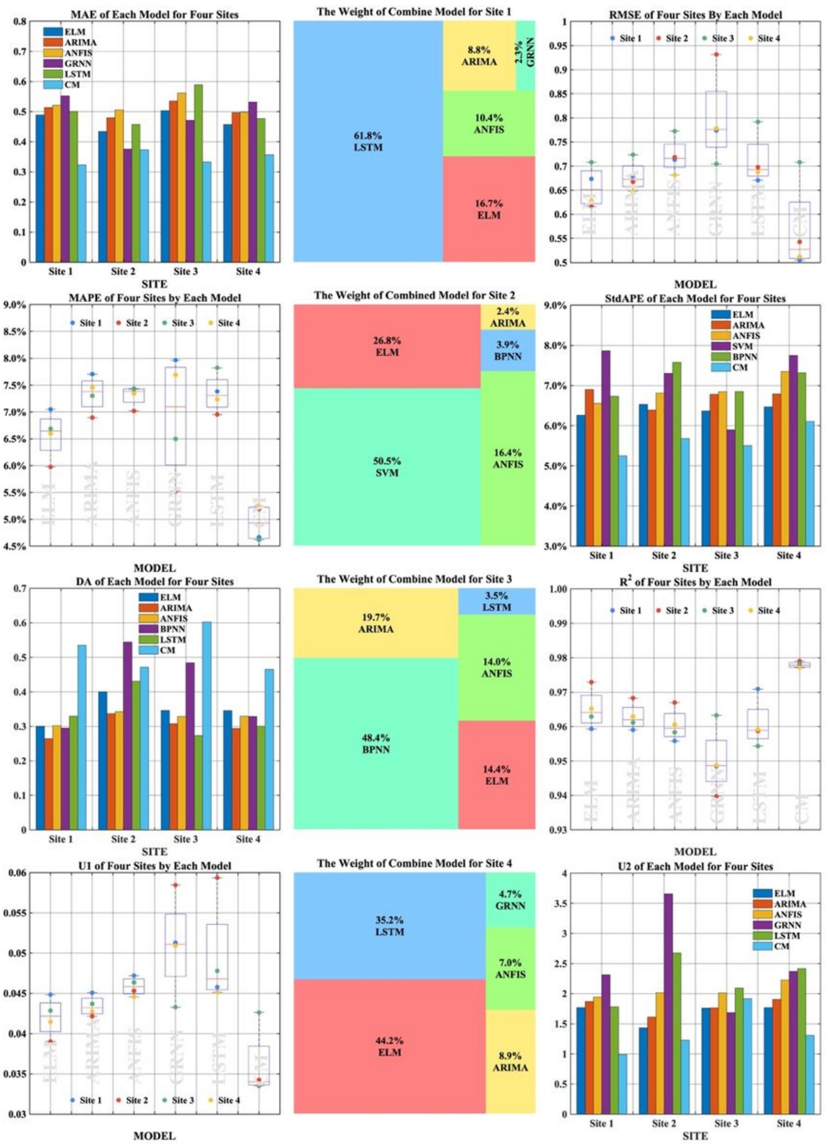

Figure 4

Forecasting results of three-step-ahead forecast.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Appendix.docx