

## Research paper

# Short-term PV-Wind forecasting of large-scale regional site clusters based on FCM clustering and hybrid Inception-ResNet embedded with Informer

Daogang Peng<sup>a,\*</sup>, Yu Liu<sup>a</sup>, Danhao Wang<sup>a,\*</sup>, Ling Luo<sup>b</sup>, Huirong Zhao<sup>a,\*</sup>, Bogang Qu<sup>a</sup>

<sup>a</sup> College of Automation Engineering, Shanghai University of Electric Power, Shanghai 200090, China

<sup>b</sup> Electric Power Science Research Institute, State Grid Shanghai Electric Power Company, Shanghai 200437, China



## ARTICLE INFO

## Keywords:

Wind and photovoltaic cluster forecasting  
Fuzzy c-means clustering  
The improved gray wolf algorithm  
Informer

## ABSTRACT

In order to cope with the challenge that the high proportion of new energy generation for the stable operation of the power grid, this paper proposes an innovative short-term power forecasting model for regional site clusters based on fuzzy c-means (FCM) clustering and hybrid Inception-ResNet deep neural network embedded with Informer. Firstly, multiple wind farms and photovoltaic sites are clustered into different groups for popular clustering prediction owing to FCM clustering algorithm. Secondly, numerous strong factors are selected based on the combination of the linear and nonlinear correlation analysis between the variables and power generation. Furthermore, the improved gray wolf algorithm (GWO) can determine the optimal parameters of deep network model and the Informer and Inception are integrated which is fairly advanced to capture temporal relationship and potent feature extraction. Finally, the wind and photovoltaic dataset in western China is employed to verify our model and the results demonstrate that ours outperforms other algorithms with 5.400% and 4.200% higher R2 and 2.525% and 2.090% lower MAPE in the wind and solar forecasting, which simultaneously improves the accuracy and efficiency of prediction.

## 1. Introduction

As energy and power systems lower carbon and cleaner in response to climate change and energy crisis, wind and photovoltaic power are increasingly popular due to their clean and sustainable nature [1]. Since wind power relies on external factors such as wind direction, temperature, air pressure [2] and photovoltaic generation is dependent on variables as solar radiation, relative humidity, and temperature [3], the randomness and fluctuations of power generation pose significant challenges to the safe and economical operation of the power grid system. Precise forecasting can enhance wind and photovoltaic site operation and maintenance standard, facilitating efficient resource allocation and grid management and promotes energy integration [4].

Power forecasting can be characterized into the two main categories: mathematical statistical methods and artificial intelligence methods [5]. Statistical methods examine the relationship between historical time series of power generation and utilize mathematical models to anticipate performance, for instance, Autoregressive Integrated Moving Average (ARIMA) and Exponentially Weighted Moving Average (EWMA). Abdulla et al. [6] employed different seasonal Holt-Winters models to anticipate power generation in Kuwait from 2020 to 2030. Nevertheless, power generation is a nonlinear stochastic process and the above struggles to capture the complicated nonlinear

interactions. Artificial intelligence methods with outstanding nonlinear function approximation and computational capability, which can manage nonlinear interactions and are introduced into power generation forecasting. Recurrent Neural Networks (RNN) is experted in handling long-term dependencies in series data and can be effective for power forecasting, for example, Mohamad et al. [7] developed a framework for accurately predicting offshore wind power combined Long Short-Term Memory Recurrent Neural Networks (LSTM) with Swarm Intelligence (SI). Wu et al. [8] suggested a short-term PV power prediction model based on Extreme Gradient Boosting (IXGBoost), with similar day clustering and signal mode decomposition. Convolution Neural Networks (CNN) excels at capturing spatial information and is applied to picture processing but can also handle time series by treating the data as image-like inputs. Hu et al. [8] introduced a novel forecasting method termed Temporal Collaborative Attention. Liu et al. [9] intergrated Bidirectional Long Short-Term Memory (BiLSTM) with CNN to forecast photovoltaic (PV). Adeel et al. [10] proposed a hybrid embedded deep neural network including ResNet, CNN and Inception module, which improved feature extraction and forecating accuracy over time.

The above models are deterministic, which only can forecast the expected value of the output and describe the uncertainty of power

\* Corresponding authors.

E-mail addresses: [pengdaogang@shiep.edu.cn](mailto:pengdaogang@shiep.edu.cn) (D. Peng), [danhao.wang@qq.com](mailto:danhao.wang@qq.com) (D. Wang), [hrzhao@shiep.edu.cn](mailto:hrzhao@shiep.edu.cn) (H. Zhao).

<https://doi.org/10.1016/j.enconman.2024.118992>

Received 19 June 2024; Received in revised form 24 July 2024; Accepted 26 August 2024

Available online 4 September 2024

0196-8904/© 2024 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

generation. To obtain information about future output and reduce the decision-making risk of the power system, it is particularly significant to design an accurate probability model of power generation, which can be classified as probability density forecasts, interval forecasts and quantile forecasts. Specifically, probability density provides the probability density function of future power output, interval forecasts provides the approximate range of the future output and quantile prediction exports the value under a certain quantile [11]. Chen et al. [12] validated a quantile regression model on the dataset from an offshore wind farm in Penglai District, Shandong Province. Yang et al. [13] proposed wind-power farm cluster prediction model based Graph Convolution Neural Networks (GCN) and fluctuation correlation. In reference [14], an innovative Capsule Network (ACCNet) stood out in interval prediction tasks. Machine learning can evaluate the uncertainty of forecasts in probability distributions, which can help adapting the future uncertainty [15].

Previous research has concentrated on the prediction of centralized power farms, whereas the power system has focused on the total generation of distributed generation sites and there are still relatively few studies on the power prediction of distributed sites. In general, there are three common frameworks for power sites generation: cumulative prediction, predicting cumulatively, and clustering prediction, among which the cumulative prediction accumulates the value of each distributed site to obtain the total in a certain region, the predicting cumulative simply superimposes the results of each site and the clustering prediction categorizes the sites into different clusters, and then accumulates each cluster. However, it is essential for power generation forecasting to strike a balance between accuracy and efficiency. For one thing, the cumulative prediction can improve the prediction efficiency but may increase the forecasting error. For another, predicting cumulatively improves the prediction accuracy but occupies much calculation time and storage. Based on the above, the clustering prediction not only realizes famous prediction accuracy but increases forecasting efficiency, which has arisen strong interest in recent years. However, fewer studies have been conducted in this area. Hou et al. [16] combined the DBSCAN clustering and LSTM to cluster wind turbines and select representative turbine to predict power output.

The forecasting performance of power generation is affected by various characteristic variables. To strengthen the prediction performance of the forecasting model, multiple variables from dataset must be carefully selected. Literature [10] used the Pearson correlation coefficient (PCC) to only analyze the linear relationship between meteorological environmental factors and historical generation data. Literature [14] calculated the Pearson correlations (PCs) among the meteorological parameters and the PV power. Literature [17] evaluated of the correlation between meteorological data and PV power using Pearson correlation coefficient. In summary, while it is important to examine the crucial variables based on the correlation between the different loads and their associated characteristics, most studies focus solely on linear or nonlinear correlations which may lead to the omission of key information.

Although the above models demonstrate exceptional prediction performance, the clustering forecasting model is worth studying [18] and most research have applied the traditional correlation method to analyze single relationship [19]. Furthermore, it is difficult for an individual model to handle complex and variable power generation fluctuation and it may be necessary to optimize the model parameters, which improves the prediction performance [20]. Therefore, this paper proposes an innovative short-term power forecasting model for regional site clusters based on variable selection, FCM clustering algorithm, improved GWO algorithm and hybrid Inception-ResNet deep neural network embedded with Informer which achieves a tradeoff between accuracy and efficiency. The main contributions of this paper are as follows:

- (1) In order to select variables with strong correlation to characterize power sites, Pearson correlation coefficient and Spearman rank correlation coefficient derived from Copula function are used to analyze linear and nonlinear correlation simultaneously between influencing factors and power generation, which is a creative combination and utilization.
- (2) Large-scale power clusters are classified for the popular clustering prediction algorithm by FCM clustering algorithm and the center of the clusters is selected as the representative site where Silhouette Coefficient (SC) is established to evaluate the clustering results. Besides, the strong correlation variables of the representative site are combined with the power generation which constructs the data input set.
- (3) A hybrid Inception-ResNet embedded Informer deep neural network model is proposed based on ResNet which can effectively alleviate the gradient vanishing problem. The Inception module can capture multi-scale features in the time-series and the Informer layer benefiting from self-attention is proficient in capturing the long-term dependency of time-series.
- (4) The GWO algorithm with outstanding global search ability is improved by introducing a new adaptive position update strategy and a new nonlinearly adjusted convergence factor to hunt for optimal parameters, which lays a solid foundation for satisfying prediction accuracy and robustness.

The rest of the paper are shown as follows. Section two provides a detailed description of different algorithms and models. Section three presents and analyzes the experiment results of various forecasting models. Finally, the section concludes the research and draws the directions of the future.

## 2. Related methods

The detailed architecture is divided into four components, as illustrated in Fig. 1. Firstly, the dataset is preprocessed including outliers filtering and missing values filling. Secondly, Pearson analysis and Spearman analysis based on copula are applied to select influencing factors and determine the input and output. Then distributed power sites are clustered using the FCM clustering algorithm and an hybrid Inception-ResNet embedded Informer forecasting model is constructed whose parameters are optimized through the improved GWO algorithm. Finally, A wind and photovoltaic dataset in western China is used to evaluate and confirm our model compared with different forecasting models.

### 2.1. The analysis of influencing factor correlation

Since wind power is strongly randomly affected by meteorological factors, it is necessary to analyze the correlation between power generation and meteorological factors and appropriately screen the input variables. In order to more accurately and specifically describe above correlation, Pearson correlation coefficient and Spearman rank correlation coefficient based on copula theory are employed to analyze linear and nonlinear correlation and select relevant variables accordingly.

Pearson correlation analysis measured the linear correlation between random variables  $X$  and  $Y$  where  $\bar{x}$  and  $\bar{y}$  denote the averages of  $X$  and  $Y$ . Pearson correlation coefficient  $r_{xy}$  between the two variables is defined as the quotient of the covariance and the standard deviation of two elements, which can be expressed as follows:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

The correlation classification of Pearson correlation coefficient is shown in Table 1 where the absolute value of  $r_{xy}$  represents the strength of the correlation between the two elements. When  $r_{xy} = 1$ , the two variables are completely positively correlated, unlike  $r_{xy} = -1$ , the

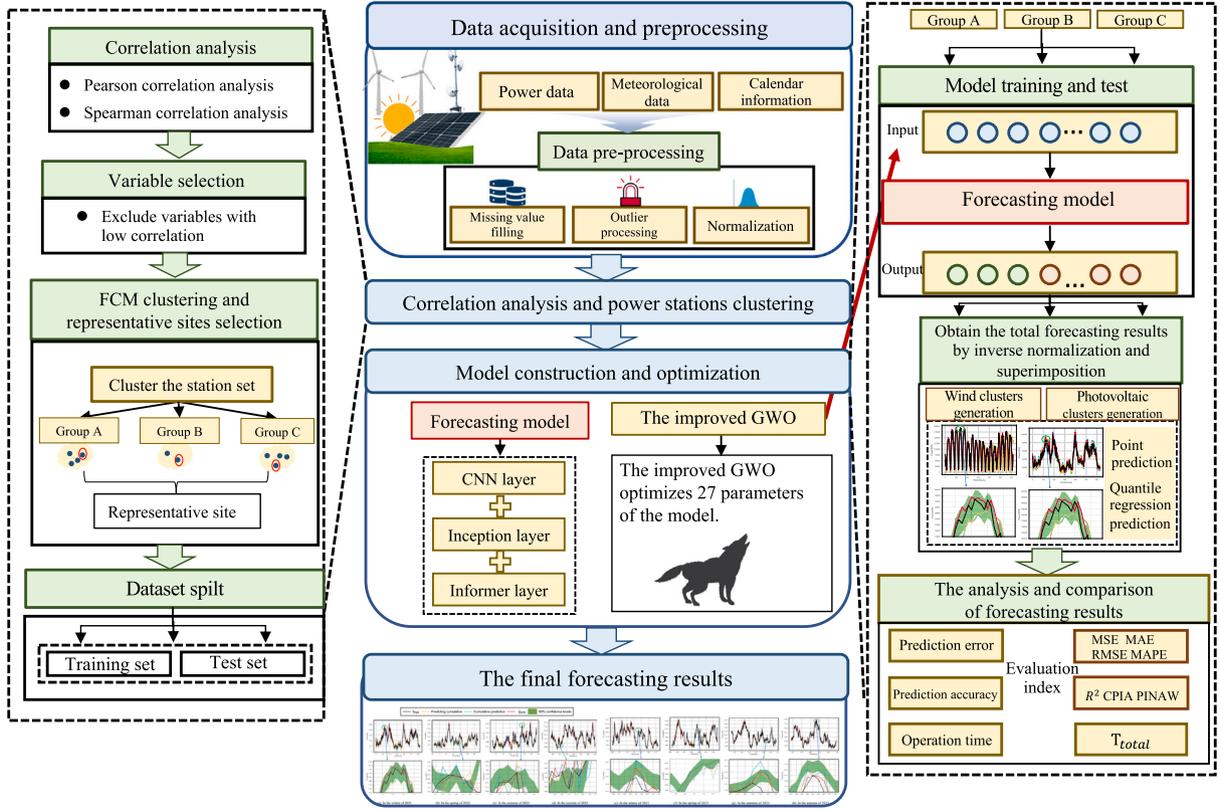


Fig. 1. The flowchart of the proposed model based on FCM clustering and hybrid Inception-ResNet embedded with Informer.

two variables are completely negatively correlated. Additionally, the correlation coefficient is 0 when there is no correlation.

Spearman correlation analysis is less sensitive to outliers and it converts raw data into rank data where  $x_i$  and  $y_i$  are sorted in descending order and the corresponding ranks  $x'_i$  and  $y'_i$  are assigned, the Spearman rank difference  $d_i$  and the Spearman rank correlation coefficient  $\rho_s$  are:

$$d_i = x'_i - y'_i \quad (2)$$

$$\rho_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (3)$$

Copula functions construct joint distribution probability functions that can characterize nonlinear correlations between variables and have been introduced into renewable energy correlation and volatility analysis in recent years [21]. According to Sklar's theorem, the joint distribution function of an N-dimensional component can be described by the marginal distributions of the N variables and a copula function [22]. And a series of random variables  $x_1, x_2, \dots, x_n$ , whose respective marginal distribution functions are  $F_1(x_1), F_2(x_2), \dots, F_n(x_n)$ , then the common distribution function between the variables based on the copula function is:

$$F(x_1, x_2, \dots, x^n) = C(F_1(x_1), F_2(x_2), \dots, F_n(x_n)) \quad (4)$$

Considering that it is complicated for joint distribution function to characterize the correlation between variables, the Spearman rank correlation coefficient derived from the copula function is used to quantitatively represent the nonlinear correlation between variables for two random variables  $X$  and  $Y$ , whose corresponding distribution functions are  $F(X)$  and  $G(Y)$ . Furthermore, the Spearman rank correlation coefficient  $\rho_s$  obtained from the copula function  $C(F(X), G(Y))$  is:

$$\rho_s = 12 \int_0^1 \int_0^1 F(X)G(Y)dC(F(X), G(Y)) - 3 \quad (5)$$

The Spearman rank correlation coefficient  $\rho_s$  indicates the direction of correlation between  $X$  independent variables and  $Y$  dependent

Table 1

The correlation classification on Pearson correlation analysis.

Pearson correlation coefficient	Classification
$0.6 \leq  r_{xy}  \leq 1$	Strong correlation
$0.4 \leq  r_{xy}  \leq 0.6$	Moderate correlation
$0 \leq  r_{xy}  \leq 0.4$	Weak correlation

Table 2

The correlation classification on Spearman rank correlation coefficient based on copula function.

Spearman rank correlation coefficient	Classification
$0.7 \leq  \rho_s  \leq 1$	Strong correlation
$0.5 \leq  \rho_s  \leq 0.7$	Moderate correlation
$0 \leq  \rho_s  \leq 0.5$	Weak correlation

variables and its magnitude is between  $-1$  and  $1$ . While the direction of change is the same and  $X$  and  $Y$  are perfectly positively correlated,  $\rho_s = 1$ . Otherwise,  $\rho_s = -1$ . The correlation classification of Spearman rank correlation is demonstrated in Table 2. As shown in Fig. 2, variables whose absolute values of the Pearson correlation coefficient and the Spearman rank correlation coefficient are both below the specified threshold of 0.6 are excluded to simplify the model and improve prediction accuracy. On the one hand, the variable is retained as long as it reaches the above-mentioned threshold values. On the other hand, if the correlation coefficients of a variable are both below the threshold value, the variable is excluded, as the selected variables with a high correlation can convey significant information. After this, the variables with a high correlation are employed to characterize the power stations, which can effectively improve the clustering performance of the clustering algorithm and enhance the forecasting performance of each forecasting model.

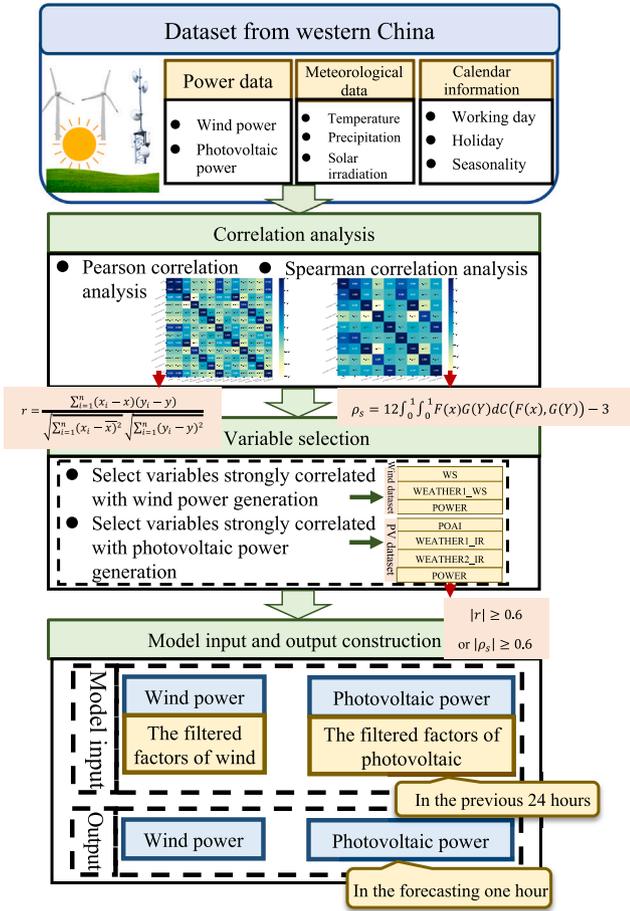


Fig. 2. The correlation analysis and variables selection.

2.2. Fuzzy c-means clustering for classifying power sites

Fuzzy C-means (FCM) is a widely used soft clustering method that optimizes the objective function to obtain the membership of each data point with respect to all clustering centers and assigns the data points to the class with the largest membership to complete the accurate classification of the sample set. Suppose N example points  $X = \{x_1, x_2, \dots, x_n\}$  divided into K classes, where the example centers of the K classes are  $\{q_1, q_2, \dots, q_k\}$ . The FCM attempts to minimize the sum of the weighted distances of the grouped sample points to the centers of all samples:

$$J_{min} = \sum_{j=1}^k \sum_{i=1}^n u_{ij}^m \|x_i - q_j\|^2 \quad (6)$$

where  $u_{ij}$  is the membership of the sample point  $x_i$  with respect to  $q_j$  in all clusters,  $m$  is the membership factor and  $u_{ij}$  must satisfy the following equation:

$$\sum_{j=1}^k u_{ij} = 1, \forall i = 1, 2, \dots, n \quad (7)$$

To obtain the minimum value of the objective function, it is necessary to iteratively calculate the degree of membership  $u_{ij}$  and the center of the sample  $q_j$  shown in Eq:

$$u_{ij} = \frac{1}{\sum_{z=1}^k \left( \frac{\|x_i - q_j\|}{\|x_i - q_z\|} \right)^{\frac{2}{m-1}}} \quad (8)$$

$$q_j = \frac{\sum_{i=1}^n u_{ij}^m x_i}{\sum_{i=1}^n u_{ij}^m} \quad (9)$$

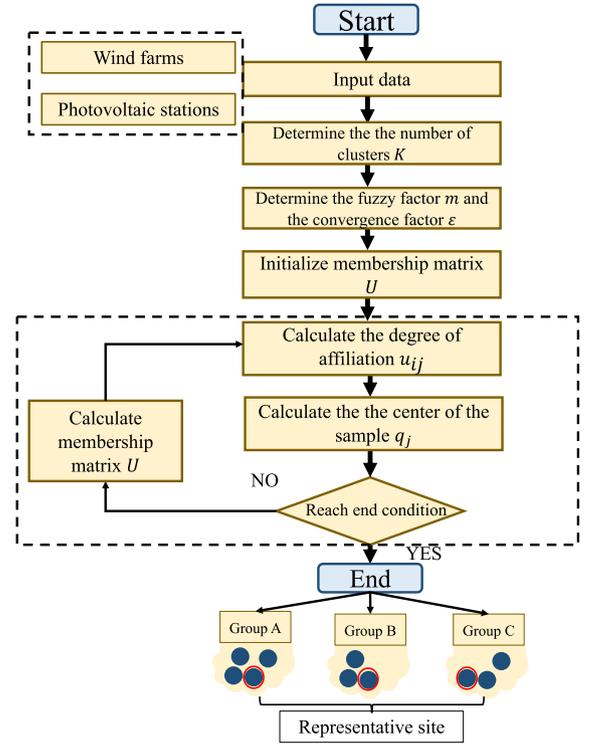


Fig. 3. Calculation procedure for the FCM clustering algorithm.

The iterative process begins with determining the number of clusters  $k$ , the fuzzy factor  $m$  and the convergence factor  $\epsilon$  as well as the cluster center  $u_{ij}^{(k+1)}$ . If  $\|u_{ij}^{(k+1)} - u_{ij}^{(k)}\| < \epsilon$  is satisfied in  $k + 1$  iterations, the iteration is stopped.

FCM clustering is an unsupervised algorithm and the silhouette coefficient (SC) is applied to evaluate the clustering results and its formula is calculated as follows:

$$s_i = \frac{n(i) - m(i)}{\max\{m(i), n(i)\}} \quad (10)$$

where  $s_i$  is the profile coefficient of power site  $i$ ;  $m_i$  is the average distance between power site  $i$  and other site during the class;  $n_i$  is the average distance between power site  $i$  and all site belonging to other classes. With the average profile coefficient calculated to get the average profile coefficient, the effect of clustering can be measured where the value of profile coefficient ranges from  $-1$  to  $1$ .

The power sites on FCM clustering algorithm are shown in Fig. 3:

1. Input the sites including generation data and selected variables and determine the number of clusters  $K$ ;
2. Determine  $k$ , the fuzzy factor  $m$  and the convergence factor  $\epsilon$ ;
3. Initialize the membership matrix ( $U$ );
4. Calculate the degree of membership  $u_{ij}$  and the center of the sample  $q_j$  based on the generation characteristic curve;
5. Recalculate the initializing the membership matrix ( $U$ );
6. If the end condition is satisfied, end the calculation ; otherwise, proceed 5;
7. Evaluate the clustering results based on SC and output the clusters of sites and representative sites.

2.3. A hybrid inception-ResNet deep neural network embedded with informer deep neural network model

2.3.1. CNN module

CNN-one dimensional (1D) which is popular technique specialized in capturing relative features and have been introduced into NLP [23]. CNN-1D consisting of input layer, convolutional layer, pooling layer

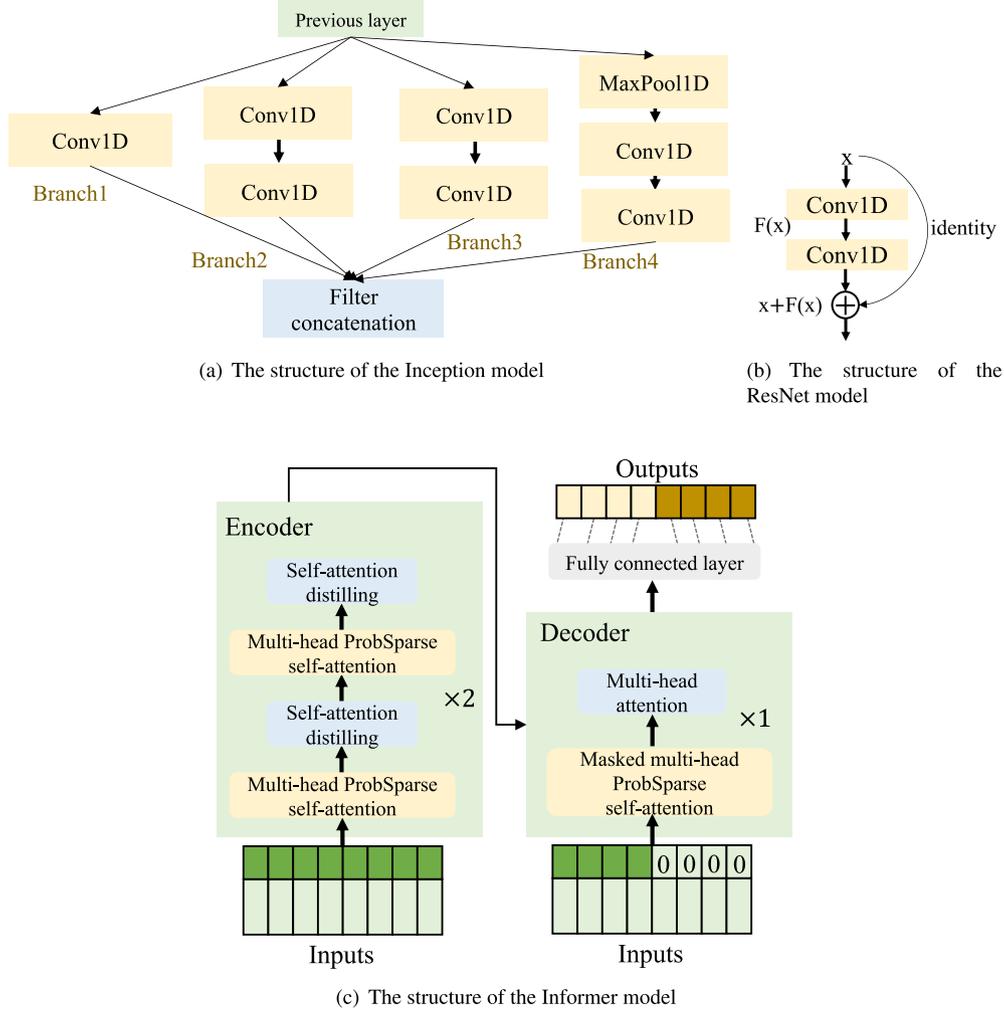


Fig. 4. Model structure.

and fully connected layer excels in extracting the sequence feature and fetch beneficial information of the data sequence. The primary distinction between CNN-two dimensional (2D) and CNN-1D lies in that 1D-CNN operates on one dimensional array instead of the matrix for 2D. Consequently, the convolution kernel and filter in 1D-CNN are also one-dimensional, which combines fewer layers and brings less calculation costs.

### 2.3.2. ResNet blocks

Residual Networks (ResNet) [24] was devised to address the issue of gradient vanishing or exploding in deep neural networks, which introduces skip connections (Shortcuts) between input and output and facilitates smooth gradient flow between layers. ResNet employs the BasicBlock structure as depicted in Fig. 4(a) where the parameter *filters* denotes the number of filters of the convolutional layer and *kernel\_size* signifies the size of the convolution kernel. ResNet is proficient in capturing both local and leads to achieving high accuracy. The BasicBlock initially employs 1D-CNN, followed by the addition of ReLU to introduce nonlinear characteristics. Furthermore, CNN-1D layers and activation functions are applied and the Shortcut operation adds the output tensor to obtain the final output through the activation function ReLU.

### 2.3.3. Informer model

The Transformer model abandons traditional CNN and RNN structures, instead comprising an attention mechanism and a feedforward neural network [25]. In the face of the limitations including high time complexity, large memory consumption, and slow single-step decoding prediction speed while conducting long sequence time-series forecasting (LSTF), Zhou et al. proposed the Informer model famous for accurately capturing long-term dependencies and the notable characteristics are as follows:

(1) Canonical self-attention is replaced by ProbSparse self-attention which reduces calculation complexity and effectively captures long-term dependencies in sequences. ProbSparse self-attention is defined as follows

$$A(Q, K, V) = \text{Softmax}\left(\frac{\bar{Q}K^T}{\sqrt{d}}\right)V \quad (11)$$

where the Kullback–Leibler (K–L) divergence between the distribution of all query vectors  $Q$  and the uniform distribution are calculated to obtain the query vector  $\bar{Q}$  that is far from the uniform distribution.

$$KL(q \parallel p) = \ln \sum_{l=1}^{L_K} \exp\left(\frac{q_l k_l^T}{\sqrt{d}}\right) - \frac{1}{L_K} \sum_{l=1}^{L_K} \frac{q_l k_l^T}{\sqrt{d}} - \ln L_K \quad (12)$$

where  $q_i, k_l$  denote the row  $i$  and  $l$  of the query vector  $Q$  and the value vector,  $\sqrt{d}$  is a normalization operation.  $l = 1^{L_K}$  is the attention score between the query vector  $q_i$  and each key vector  $k_l$ .

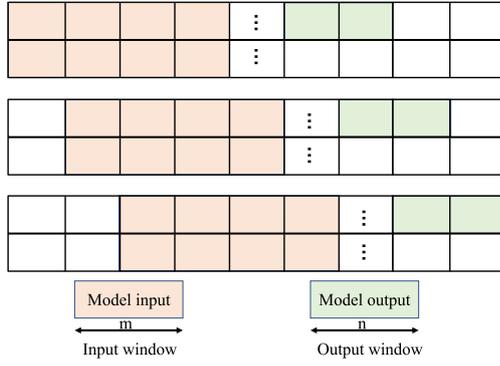


Fig. 5. The process of rolling forecasting.

(2) The Decoder adopts the self-attention distillation mechanism to diminish Encoder stacking and sequence dimensionality while assigning greater weights to dominant features. The distillation operation from the layer  $j$  to the layer  $j+1$  is outlined below:

$$X_{j+1}^t = \text{MaxPool}(\text{RELU}(\text{Conv1d}([\cdot]_{AB}^t))) \quad (13)$$

where  $[\cdot]_{AB}$  represents multi-head ProbSparse attention,  $\text{Conv1d}$  is convolution module with kernel size of 3,  $\text{RELU}$  is the activation function and  $\text{MaxPool}$  is the pooling layer with a stride of 2.

(3) Transformer employs step-by-step dynamic decoding, which is suited for short-term prediction but is tough for LSTF. Informer can address the efficiency challenge associated with LSTF and mitigate the accumulation and diffusion of prediction errors.

$$X_{de}^t = \text{Concat}(X_{token}^t, X_0^t) \in \mathbb{R}^{L_{token}+L_y} \times d_{model} \quad (14)$$

where  $X_{token}^t$  represents the token of the initial input for the Decoder layer,  $X_0^t$  is the encoded value of the target sequence and  $X_{de}^t$  denotes the output of the Decoder. Consequently, it is accessible to generate the target sequence in a single step and eliminate the necessity for dynamic decoding.

#### 2.3.4. Advantages of ours

Inception [26] is a multi-scale convolutional neural network (CNN) proposed by Google combining convolution kernels and pooling layers of various sizes to learn spatial and temporal features across different scales to facilitate richer and diverse information. It plays a vital role in capturing both short-term and long-term dependencies within output sequences in generation tasks. The Inception module in this paper consisting of four sections, as shown in Fig. 4(a) where Branch 1 employs CNN-1D and the ReLU activation function, Branches 2 and 3 are composed of two CNN-1D with different kernel sizes and Branch 4 utilizes the MaxPool1D and CNN-1D. Finally, the feature maps from different channels are combined to obtain a multi-channel spatial feature matrix followed the Concatenate.

Informer based on ProbSpace self-attention mechanism mainly reduces computational complexity and effectively capture long-term dependency in time series, which contributes to accurately understanding the historical trends and future performance of power generation. As shown in Fig. 5, the input of Inception-ResNet embedded with Informer is filtered factor variables and generation data for the previous 24 h where  $m$  is 24 and  $n$  is 1. The output of our model is the value of the generation value of next hour. Inception-ResNet module extracts high-level features from input data and Informer model can capture long-term dependencies in time series data. Finally, decoder gets the forecasting result for the next hour and sliding window method is used for 24 h prediction shown in Fig. 6.

#### 2.4. The improved GWO algorithm optimizes parameters

Gray Wolf Optimization (GWO) is an optimization algorithm inspired by nature, which is similar to simulate the social hierarchy and hunting strategy of wolves. Wolves are categorized into four levels represented by  $\alpha$ ,  $\beta$ ,  $\delta$ , and  $\omega$ . The algorithm flow is depicted in Fig. 7, which is divided into algorithm initialization, encircling prey, pursuing prey, and attacking prey.

(1) Initialization the wolves involves calculating the fitness of each individual in the population and marking the top three gray wolves with the best fitness as  $\alpha$ ,  $\beta$ ,  $\delta$ , and designate the rest as  $\omega$ .

(2) Gray wolves initial hunting strategy contains in encircling prey, which is represented as:

$$\begin{aligned} \vec{D} &= |\vec{C} * \vec{X}_p(t) - \vec{X}(t)| \\ \vec{X}(t+1) &= \vec{X}_p(t) - \vec{A}\vec{D} \end{aligned} \quad (15)$$

where  $\vec{X}_p$  represents the position of the prey,  $\vec{X}(t)$  denotes the position of the gray wolf at time  $t$ ,  $\vec{X}(t+1)$  indicates the position of the gray wolf at time  $t+1$ , and  $\vec{D}$  is the distance between the gray wolf and the prey.  $\vec{A}$  and  $\vec{C}$  are coefficients determined as follows:  $\vec{A} = 2a(t)\vec{r}_1 - a(t)$ ,  $\vec{C} = 2\vec{r}_2$ , where  $\vec{r}_1$  and  $\vec{r}_2$  are uniformly distributed in the interval  $[0, 1]$ . This paper discusses the inertia weight setting in the particle swarm optimization (PSO) algorithm and proposes a strategy for nonlinear adjustment of the convergence factor, specifically given by:  $a(t) = a_{initial} - (a_{initial} - a_{final}) \times \log\left(1 + (e + 1) \times \frac{t}{Max_{iter}}\right)$  where  $t$  denotes the current iteration number,  $Max_{iter}$  represents the maximum iteration number, and  $a_{initial}$  and  $a_{final}$  are 2 and 0, respectively.

(3) Furthermore, the gray wolf  $\omega$  adjusts its position based on  $\alpha$ ,  $\beta$ , and  $\delta$  to effectively pursue the prey. Specifically, the distances  $\vec{D}_\alpha$ ,  $\vec{D}_\beta$ ,  $\vec{D}_\delta$  between  $\omega$  and  $\alpha$ ,  $\beta$ ,  $\delta$  are calculated in Eq. (16):

$$\begin{aligned} \vec{D}_\alpha &= |\vec{C}_1 * \vec{X}_\alpha(t) - \vec{X}(t)| \\ \vec{D}_\beta &= |\vec{C}_2 * \vec{X}_\beta(t) - \vec{X}(t)| \\ \vec{D}_\delta &= |\vec{C}_3 * \vec{X}_\delta(t) - \vec{X}(t)| \end{aligned} \quad (16)$$

where  $\vec{C}_1$ ,  $\vec{C}_2$ ,  $\vec{C}_3$  are random numbers between 0 and 1, and the  $\omega$  individual adjusts its position according to the distance, as shown in Eq. (17):

$$\begin{aligned} \vec{X}_1 &= |\vec{X}_\alpha - \vec{A}_1 * \vec{D}_\alpha| \\ \vec{X}_2 &= |\vec{X}_\beta - \vec{A}_2 * \vec{D}_\beta| \\ \vec{X}_3 &= |\vec{X}_\delta - \vec{A}_3 * \vec{D}_\delta| \end{aligned} \quad (17)$$

(4) Based on the particle swarm algorithm, this process enables individuals to concurrently learn the global optimal position and their individual historical optimal positions, proposing an enhanced position update rule. Once the prey stops moving, the wolf pack initiates its attack behavior, as defined in Eq. (18):

$$\begin{aligned} X(t+1) &= w \cdot \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3} + C_1 \cdot r_1 \cdot \\ &\quad (X_{(pbest)} - X) + C_2 \cdot r_2 \cdot (X_{(1)} - X) \\ w &= \frac{Max_{iter} - t}{Max_{iter}} \times (w_{initial} - w_{final}) + w_{final} \end{aligned} \quad (18)$$

where  $t$  denotes the current iteration number,  $Max_{iter}$  is the maximum number of iterations,  $\vec{r}_1$  and  $\vec{r}_2$  are random numbers uniformly distributed over  $[0, 1]$ ,  $C_1$  and  $C_2$  are constants set to 0.5. Besides,  $w$  represents the inertia weight set to 0.5,  $X_{(pbest)}$  denotes the global optimal position and  $X_{(1)}$  represents the individual optimal position. Based on the gray wolves  $\alpha$ ,  $\beta$ , and  $\delta$ , along with their historical optimal positions, the new fitness value are updated and the parameters are adjusted until reaching the maximum number of iterations.

The steps of optimizing the parameters of our model based on improved GWO are shown below:

1. Initialize the wolves parameters that are hyperparameters of the model;

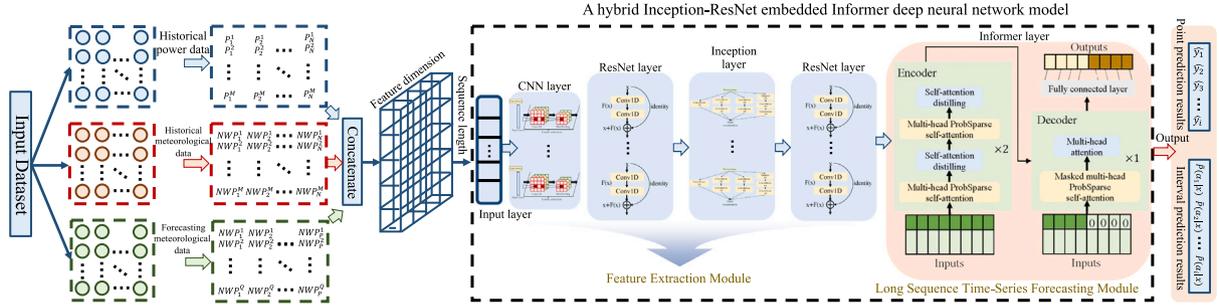


Fig. 6. A hybrid Inception-ResNet deep neural network embedded with Informer.

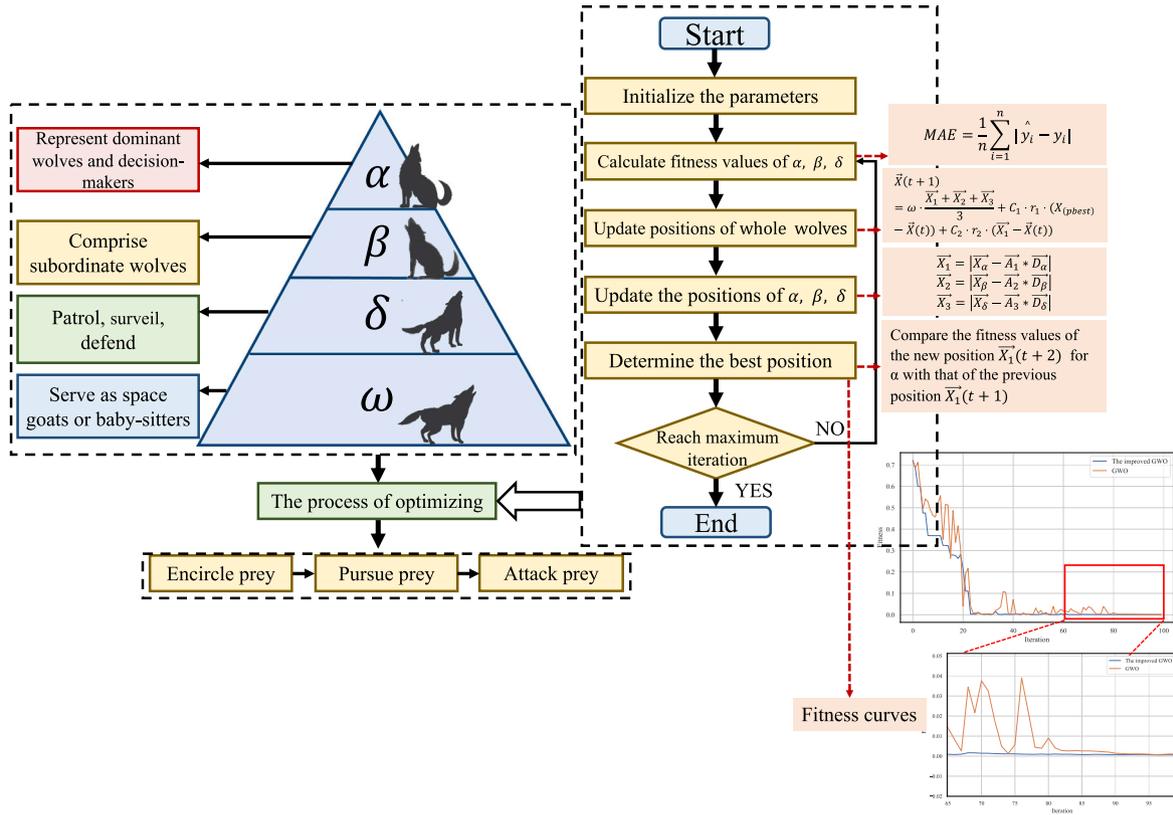


Fig. 7. The flow chart of GWO optimization.

2. Calculate the fitness and mark the top three gray  $\alpha$ ,  $\beta$ , and  $\delta$  wolves with the best fitness, which is the forecasting error of the model;
3. Update the position of wolves based on  $\alpha$ ,  $\beta$ , and  $\delta$  and update the position of wolves  $\alpha$ ,  $\beta$ , and  $\delta$ ;
4. If the end condition is satisfied, end the calculation; otherwise, proceeding 2;
5. Determine the best position based on the position of  $\alpha$  and save the model for forecasting.

### 2.5. Performance metrics of forecasting model

This paper adopts the quantile regression method (QR) to perform interval prediction of wind and photovoltaic power generation by constructing the relationship between the quantile of a random variable

and a series of related factors  $x$ , the QR model can be expressed as:

$$\tilde{P}(\alpha|x) = \beta(\alpha)x \quad (19)$$

where  $\tilde{P}(\alpha|x)$  is the  $\alpha$ -th conditional quantile corresponding to the quantile level  $\alpha$  with ranging from 0 to 1,  $x$  is the input variable and  $\beta(\alpha)$  is the regression coefficient. The problem of solving the regression coefficients  $\beta(\alpha)$  for different quantile points can be transformed into minimizing the loss function:

$$L = \sum_{i=1}^n (\gamma_{\alpha}(\tilde{P}(\alpha|x_i) - P_i)) \quad (20)$$

$$\gamma_{\alpha}(s) = \begin{cases} \alpha s, & s \geq 0 \\ (\alpha - 1)s, & s < 0 \end{cases}$$

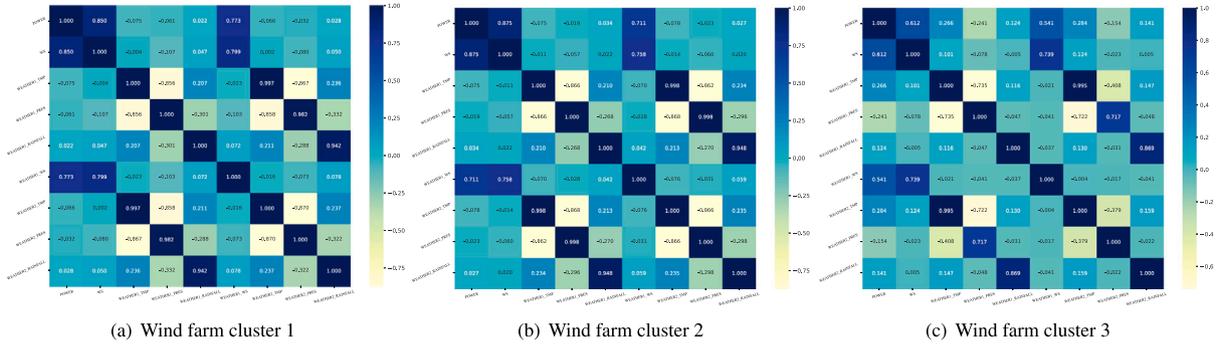


Fig. 8. The heat map of wind power generation and influencing factors based on Pearson correlation coefficient.

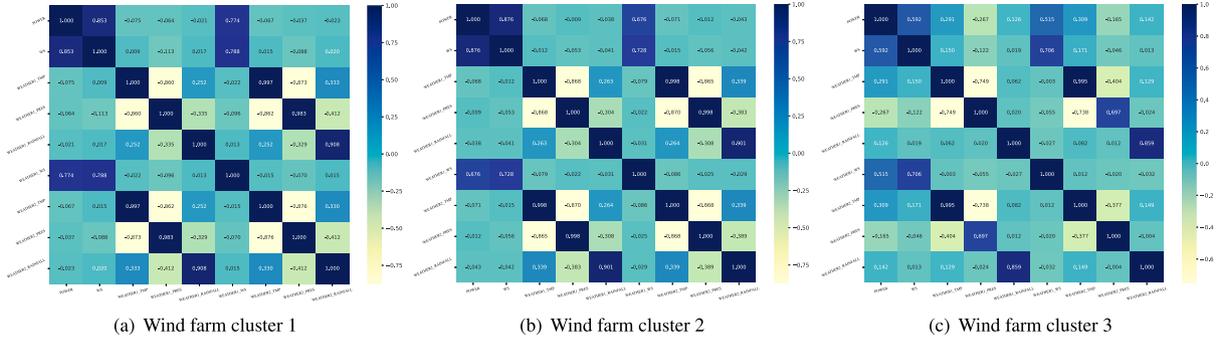


Fig. 9. The heat map of wind power generation and influencing factors based on Spearman rank correlation coefficient on copula function.

where  $P_i$  is the actual power value,  $n$  is the number of power points, and  $\gamma_\alpha$  represents an asymmetric function.

Under a certain confidence interval  $1 - \tau$ , the quantile level  $\alpha$  as  $\tau/2$  and  $1 - \tau/2$  respectively and the regression coefficient  $\beta(\alpha)$  are set, which assigned the upper and lower bounds to  $\tilde{P}(\alpha|x)$  of this confidence interval.

To evaluate the effectiveness and feasibility of the proposed model, eight evaluation indices are selected: mean absolute percentage error (MAE), mean square error (MSE), mean absolute percentage error (MAPE), root mean square error (RMSE), coefficient of determination ( $R^2$ ), prediction interval normalized average width (PINAW), corrected prediction interval accuracy(CPIA), and total calculation time ( $T_{total}(s)$ ), which assess both point prediction and interval prediction results.

$$\begin{aligned}
 FitnessFunction &= MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \\
 MAE &= \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \\
 MAPE &= \frac{1}{n} \sum_{i=1}^n \frac{|\hat{y}_i - y_i|}{y_i} \times 100\% \\
 RMSE &= \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \\
 R^2 &= 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (\bar{y}_i - y_i)^2} \\
 T_{total} &= T_{opt} + T_{tra} + T_{pre} \\
 PINAW &= \frac{1}{NC} \sum_{i=1}^N (U_i\tau - L_i\tau) \\
 CPIA &= \frac{1}{T} (1 - PINAW) \sum_{i=1}^N C_i
 \end{aligned} \tag{21}$$

where  $\hat{y}_i$  denotes the forecasting value,  $y_i$  denotes the actual value,  $\bar{y}_i$  represents the average of the actual values, and  $T_{opt}$ ,  $T_{tra}$ ,  $T_{pre}$

are the operation time of the optimization, clustering, and prediction algorithms, respectively.  $L_i\tau$  and  $U_i\tau$  represent the upper and lower bounds under the  $1 - \tau$  confidence interval,  $N$  denotes the number of samples,  $C$  is the difference between the maximum and minimum actual power, and  $C_i$  stands for the precision coefficient where a smaller PINAW and CPIA close to 1 indicates a better interval prediction effect of the model.

### 3. Case studies

#### 3.1. Data preprocessing and correlation analysis

This study presents the results of short-term wind and photovoltaic power forecasting using the wind and photovoltaic dataset in western China from January 8, 2020, to January 20, 2021, which includes ten wind farms dataset containing 11 variables and 67,105 samples and ten photovoltaic sites converting 15 variables and 25,104 samples during the time range from 6:00 to 18:00. The data are preprocessed to handle missing values and outliers and Pearson and Spearman correlation analysis based on Copula function are applied to select key influencing factors shown in Figs. 8 and 9.

According to the correlation classification of Tables 1 and Tables 2, WS (m/s) (actual wind speed), WEATHER1\_WS (m/s) (wind speed of meteorological source 1) and POWER (KW) (wind power) are related to wind power, POAI (W/m<sup>2</sup>) (actual horizontal irradiation), WEATHER1\_IR (W/m<sup>2</sup>) (horizontal irradiation of meteorological source 1), WEATHER2\_IR (W/m<sup>2</sup>) (horizontal irradiation of meteorological source 2) and POWER (KW) (photovoltaic generation) are strongly correlated with solar power. Therefore, the above variables are selected as the characteristic variables of power sites for clustering section shown in Figs. 10–12.

#### 3.2. Power sites clustering

This paper employs the FCM algorithm with the fuzzy parameter  $m$  set to 2 and the termination criterion error of 0.005. Ten wind

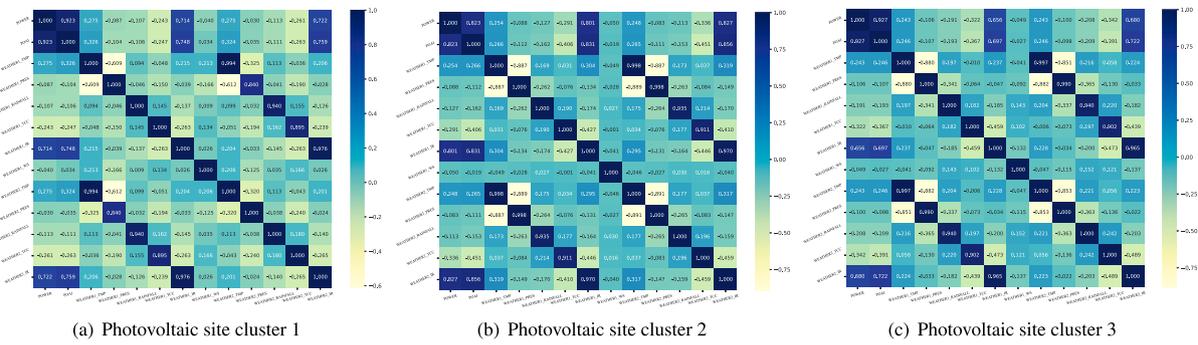


Fig. 10. The heat map of photovoltaic generation and influencing factors based on Pearson correlation coefficient.

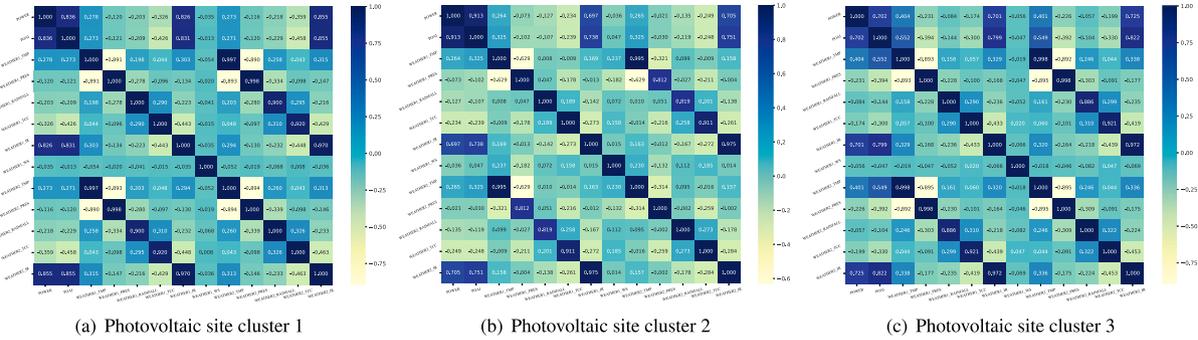


Fig. 11. The heat map of photovoltaic generation and influencing factors based on Spearman rank correlation coefficient on copula function.

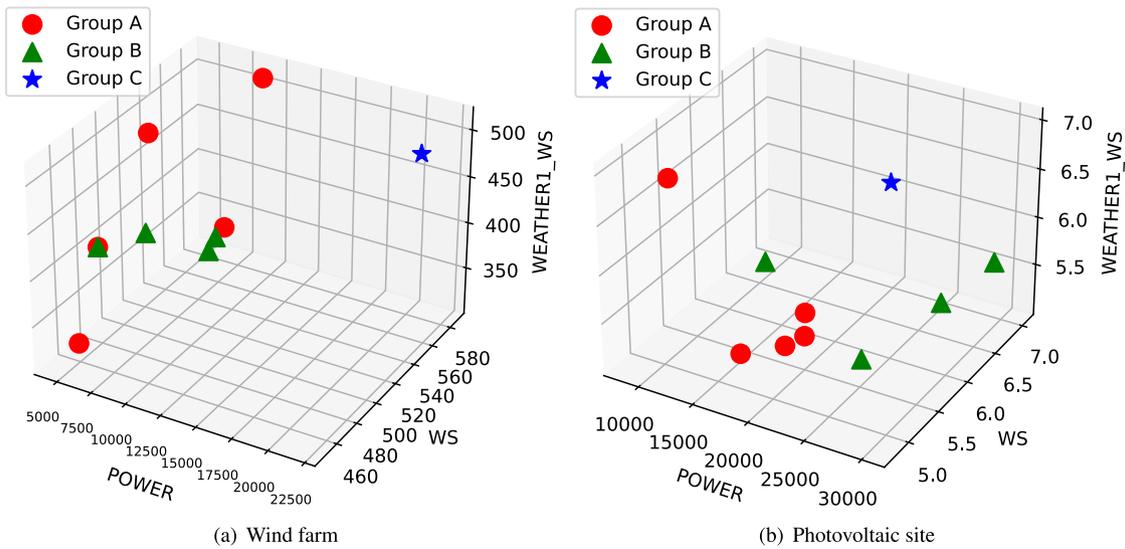


Fig. 12. The clustering results of the different algorithms.

farms and ten photovoltaic sites are clustered and compared with four algorithms: K-Means, K-Means++, DBSCAN, and BIRCH, whose numbers of clusters ( $N = 3, 4, 5$ ) with  $M$  denoting the number of power generation sites in every group.

The results of different clustering algorithms are displayed in Fig. 13, where elements of the same color are in the same group and the selected representative sites are (representative sites) RT. As shown in Table 3,  $N$  is the clustering number,  $M$  is the number of wind turbines in each group, RT represents the representative site in each group, and SC represents the average silhouette coefficient whose increases demonstrates that it enhances forecasting accuracy.

The clustering results of various algorithms and selected representative sites are listed in Table 3, where three cases with different cluster

numbers ( $N = 3, 4, \text{ and } 5$ ) are selected for the clustering and the average silhouette coefficient (SC) obtained through FCM clustering is superior to other algorithms, suggesting the selected sites can better represent regional site clusters.

### 3.3. Enhanced GWO algorithm

It is necessary for hybrid Inception-ResNet deep neural network embedded with Informer to optimize parameters and consequently this paper improved GWO algorithm whose fitness function, max number of iterations, range and number of wolves are listed in Table 5. The optimization parameters consist of kernel size1, kernel filter1, kernel stride1 in the first CNN layer; Res kernel1 and Res kernel2 in the

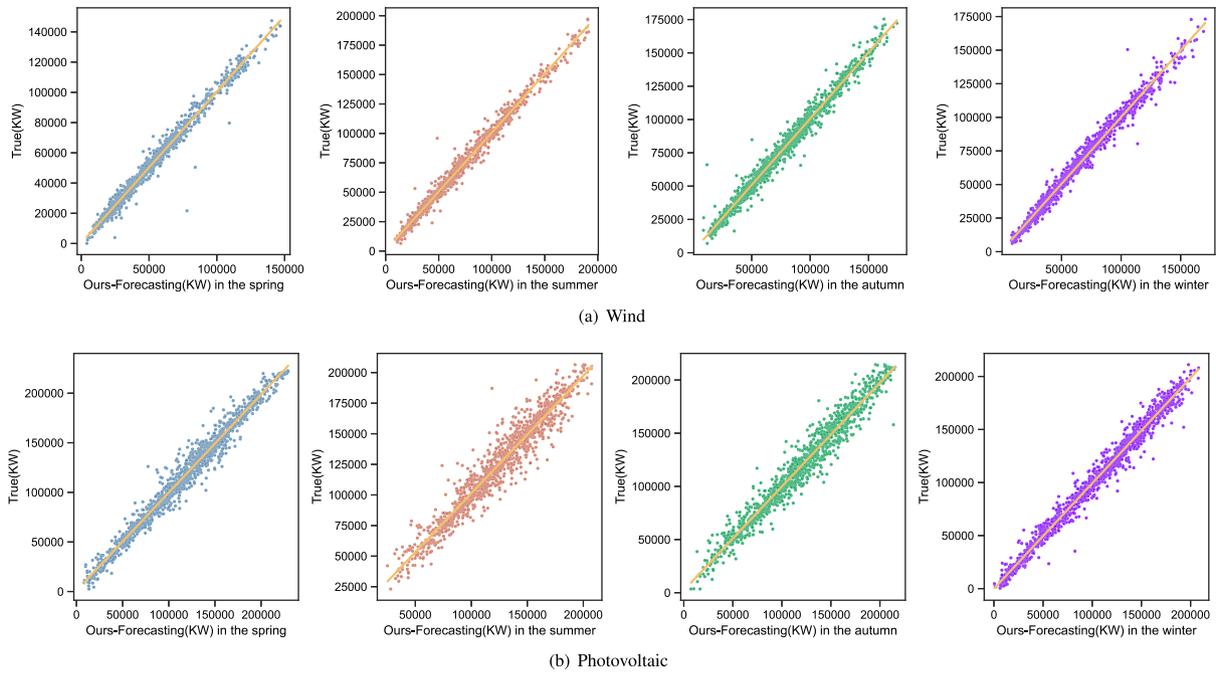


Fig. 13. The scatter plot of generation forecasting for ours.

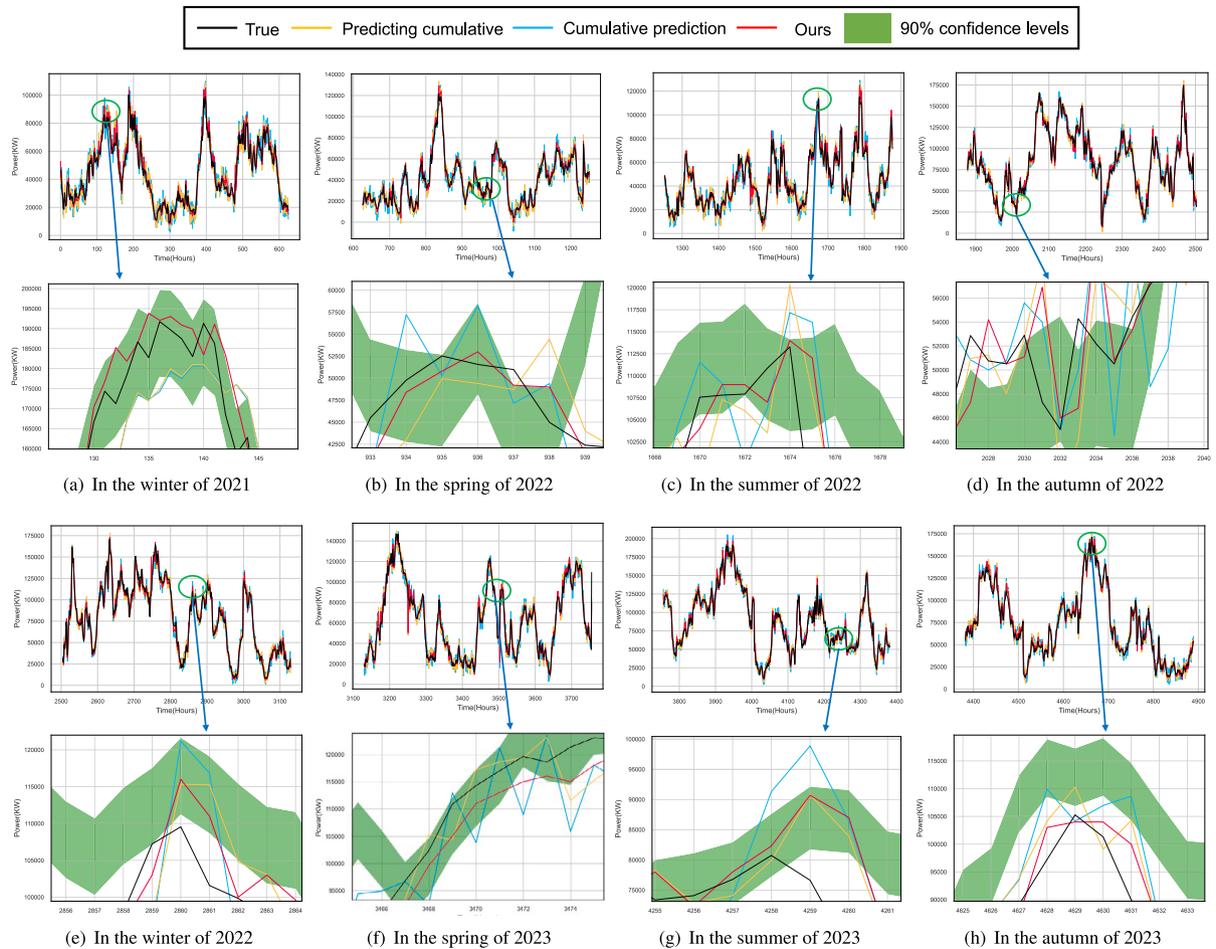


Fig. 14. Comparison with three forecasting frameworks on wind.

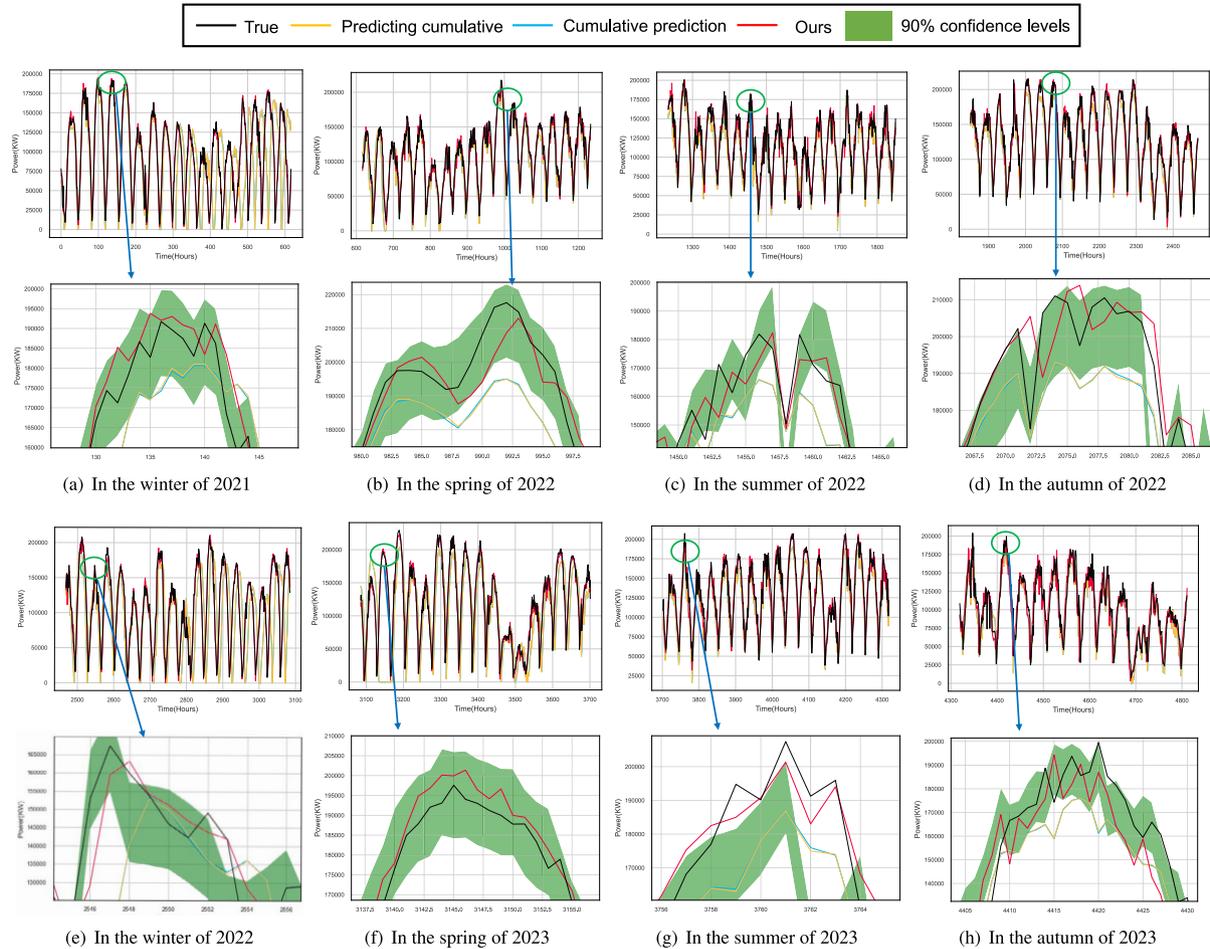


Fig. 15. Comparison with three forecasting frameworks on photovoltaic.

first and second ResNet layers; kernel size2, kernel filter2, kernel stride2 in the second CNN layer; Res kernel3 and Res kernel4 in the third and fourth ResNet layers; kernel size3, kernel filter3, kernel stride3 in the third CNN layer; Res kernel5 in the fifth ResNet layer; kernel size4, kernel filter4, kernel stride4 in the fourth CNN layer; Res kernel6 in the sixth ResNet layer; Inc convk1 and Inc convf in the Inception layer branch1;

Inc convk21 and Inc convk22 in branch2 CNN layers; Inc convk31 and Inc convk32 in branch3 CNN layers; pool size, Inc convk41, and Inc convk42 in branch4 pooling and CNN layers in Table 4.

The enhanced GWO algorithm locates the prey through the positions of the wolves  $\alpha$ ,  $\beta$  and  $\delta$  to gradually reduce the distance from the prey and ultimately capture it.

In addition to the accuracy, the convergence speed is also a crucial indicator. Fig. 7 shows the MSE convergence curves with the line during the training process represents the loss curve for wind and photovoltaic forecasting, the line with orange represents the loss curve of the enhanced GWO and the line with blue represents the loss curve for the original GWO. It is clear that that our model exhibits a faster convergence speed, with smaller differences in loss between different tasks the model is essentially converged around Epoch 50, and the total loss is minimized around Epoch 80. Combined with Table 5, it is evident that our improved outperformed GWO both in convergence speed and solution accuracy according to the optimization results and the feasibility and validity of the enhanced GWO are verified.

### 3.4. Results analysis and error comparison

The datasets with 15-minute sampling interval are into different seasons, where the spring is from March 1 to May 31, the summer

is from June 1 to August 30, the autumn is from September 1 to November 30 and the winter is from December 1 to February 28. The division ratio of the training set, validation set, and test set is 8:1:1 and seven days in each month are randomly selected as the test set.

The experimental simulation for this study utilized a Xeon(R) Platinum 8255C CPU and an NVIDIA V100-SXM2-32 GB GPU, operating Windows 11. Furthermore, the hybrid Inception-ResNet deep neural network embedded with Informer for wind and photovoltaic 24-hour power generation is constructed with the Tensorflow in Python and the detailed information of our model are given in Table 5. This paper assesses the training process and evaluates model performance at the same time as implements the validation set and determines prediction accuracy on the test set to verify model performance.

#### 3.4.1. Analysis of our model

The forecasting results proposed in this paper are shown in Table 6 and it can be seen that the average MAPE is reduced to 6.980%, the average  $R^2$  is increased to 0.969 and the average  $T_{total}$  is 340.94 s, indicating that its excellent performance and prominent efficiency of short-term wind and photovoltaic power generation forecasting.

Meanwhile, Fig. 13 suggests the Inception layer extracting high-dimensional features can concentrate the points along the diagonal with fewer outliers and capturing the characteristics of wind and photovoltaic power generation. These findings highlight that the clustering algorithm accumulates the forecasting results of each cluster to complement errors each other, which can contribute to improving prediction accuracy.

**Table 3**  
Clustering results and selection of representative sites.

Wind farm																
N	Group	FCM			K-Means			K-Means++			BIRCH			BSCAN		
		M	RT	SC	M	RT	SC	M	RT	SC	M	RT	SC	M	RT	SC
N = 3	Group = A	4	8	<b>0.881</b>	5	1	0.475	4	10	0.875	2	6	0.333	-	-	-
N = 3	Group = B	2	2	<b>0.881</b>	4	10	0.475	1	7	0.875	3	5	0.333	-	-	-
N = 3	Group = C	4	5	<b>0.881</b>	1	7	0.475	5	1	0.875	5	9	0.333	-	-	-
N = 4	Group = A	1	7	0.623	1	6	0.223	5	1	-0.026	-	-	-	-	-	-
N = 4	Group = B	2	2	0.623	5	1	0.223	1	7	-0.026	-	-	-	-	-	-
N = 4	Group = C	4	8	0.623	3	10	0.223	2	8	-0.026	-	-	-	-	-	-
N = 4	Group = D	3	1	0.623	1	7	0.223	2	6	0.556	-	-	-	-	-	-
N = 5	Group = A	2	8	0.378	2	2	0.556	4	5	0.556	-	-	-	2	7	0.156
N = 5	Group = B	1	7	0.378	4	5	0.556	2	2	0.556	-	-	-	3	9	0.156
N = 5	Group = C	2	3	0.378	1	8	0.556	1	7	0.556	-	-	-	2	4	0.156
N = 5	Group = D	2	4	0.378	2	7	0.556	1	6	0.556	-	-	-	2	3	0.156
N = 5	Group = E	2	10	0.378	1	6	0.556	2	8	0.556	-	-	-	1	1	0.156
Photovoltaic site																
N	Group	FCM			K-Means			K-Means++			BIRCH			BSCAN		
		M	RT	SC	M	RT	SC	M	RT	SC	M	RT	SC	M	RT	SC
N = 3	Group = A	1	10	<b>0.830</b>	5	5	0.381	2	7	0.382	3	7	0.210	-	-	-
N = 3	Group = B	4	4	<b>0.830</b>	2	7	0.381	5	5	0.382	5	4	0.210	-	-	-
N = 3	Group = C	5	5	<b>0.830</b>	3	4	0.381	3	4	0.382	2	10	0.210	-	-	-
N = 4	Group = A	2	9	0.650	1	10	0.442	3	4	0.441	-	-	-	3	5	0.190
N = 4	Group = B	4	5	0.650	3	4	0.442	1	1	0.441	-	-	-	2	2	0.190
N = 4	Group = C	3	1	0.650	1	7	0.442	1	7	0.441	-	-	-	4	9	0.190
N = 4	Group = D	1	10	0.650	5	5	0.442	5	5	0.441	-	-	-	1	1	0.190
N = 5	Group = A	5	5	0.625	2	1	0.556	5	5	0.506	-	-	-	-	-	-
N = 5	Group = B	1	9	0.500	1	10	0.556	1	10	0.506	-	-	-	-	-	-
N = 5	Group = C	1	1	0.500	1	7	0.556	2	1	0.506	-	-	-	-	-	-
N = 5	Group = D	2	2	0.500	5	5	0.556	1	2	0.506	-	-	-	-	-	-
N = 5	Group = E	1	10	0.500	1	2	0.556	1	7	0.506	-	-	-	-	-	-

**Table 4**  
Optimal parameters of our model.

Parameters of GWO	Parameters of ours	Range	Results	Parameters of ours	Range	Range	
The gray wolf population	27	kernel size1	[1, 7]	7	kernel filter1	[16, 256]	16
		kernel stride1	[1, 2]	1	Res kernel1	[1, 7]	7
		Res kernel2	[1, 7]	7	kernel size2	[1, 7]	4
		kernel filter2	[16, 265]	18	kernel stride2	[1, 2]	2
		Res kernel3	[1, 7]	5	Res kernel4	[1, 7]	7
		kernel size3	[1, 7]	3	kernel filter3	[16, 256]	16
Fitness function	MAE	kernel stride3	[1, 2]	1	Res kernel5	[1, 7]	5
		kernel size4	[1, 7]	5	kernel filter4	[16, 156]	32
		kernel stride4	[1, 2]	2	Res kernel6	[1, 7]	1
		Inc convk1	[1, 7]	7	Inc convf	[16, 156]	16
Max number of iterations	100	Inc convk21	[1, 7]	3	Inc convk22	[1, 7]	7
		Inc convk31	[1, 7]	6	Inc convk32	[1, 7]	7
		pool size	[1, 7]	1	Inc convk41	[1, 7]	7
		Inc convk42	[1, 7]	4			

3.4.2. Comparison with three frameworks of power sites generation

In order to compare the performance of three frameworks, namely, cumulative prediction, predicting cumulatively, and clustering prediction, this section conducts the contrasting experiments based on the regional wind and photovoltaic power generation dataset.

The comparison of power sites generation forecasting among different frameworks is presented in Table 7. In the wind power dataset, hybrid Inception-ResNet deep neural network embedded with Informer model achieves a high R<sup>2</sup> score of 0.969, 0.980 and 0.979 in spring, summer and winter, indicating its strong predictive performance. However, the our model also demonstrates good accuracy and improves the average CPIA by 0.033 and reduces the average MAPE on wind power by 5.186 compared to the rest framework. Moving to Photovoltaic power generation forecasting, which involves datasets from western China, hybrid Inception-ResNet embedded with Informer continued to showcase its superiority, achieving R<sup>2</sup> of 0.981, 0.949 and 0.982 in spring, summer and autumn, respectively. The other methods, namely, cumulative prediction, predicting cumulatively, exhibits slightly lower

R<sup>2</sup> values in these cases. Moreover, our model achieves a 10.227% lower MAPE than cumulative prediction in spring than and a 9.328% lower than predicting cumulatively in autumn. Furthermore, the clustering prediction achieves the lowest MAPE among three models in every season and consumes less operation time T<sub>total</sub> compared to predicting cumulatively and is not different of the cumulative prediction. Overall, the clustering prediction framework demonstrates consistently high R<sup>2</sup> scores and low calculation time and cost across the different datasets and various seasons, making it a favorable choice for accurate PV and wind power forecasting.

As demonstrated in Table 7 and Figs. 14–15, our clustering prediction achieves higher accuracy and consumes lower computational cost and time, which verifies the effectiveness of the data clustering method. The reason is separated two section: for one thing, the clustering prediction can improve curve predictability by superimposing curves from different sites corresponding to clustering for several site groups. For another, This also offsets errors in day-ahead power generation forecasting across different clusters, which contributes to decreasing

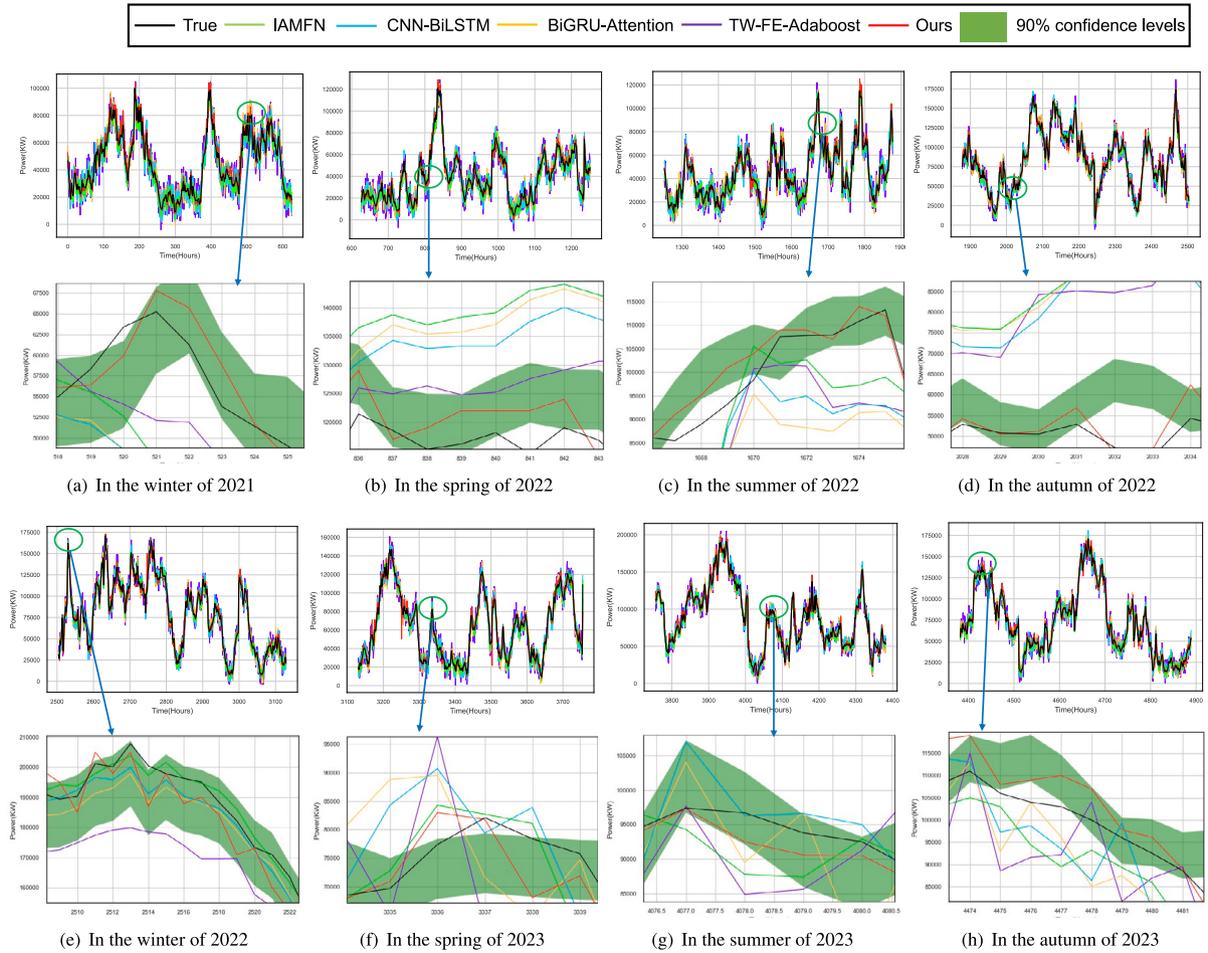


Fig. 16. The forecasting results of state-of-the-art models on wind.

Table 5  
Detailed description of the hybrid Inception-ResNet deep neural network embedded with Informer.

Model parameters	Model layer	Model parameters	Model layer	Model parameters
Epoch = 200, Batch_size = 32, validation_split = 0.3, Learning rate = 0.001, Optimizer = Adam	Input		Inception	
	Conv1D	kernel size = 7, kernel filter = 16, kernel stride = 1	branch1 Conv1D	kernel size = 2, kernel filter = 16
	ResNet	kernel size = 7, kernel filter = 16	branch2 Conv1D	kernel size = 3, kernel filter = 16
	ResNet	kernel size = 7, kernel filter = 16	Conv1D	kernel size = 7, kernel filter = 16
	Conv1D	kernel size = 4, kernel filter = 18, kernel stride = 2	branch3 Conv1D	kernel size = 6, kernel filter = 16
	ResNet	kernel size = 5, kernel filter = 18	Conv1D	kernel size = 7, kernel filter = 16
	ResNet	kernel size = 7, kernel filter = 18	branch4 MaxPooling1D	pool size = 1, kernel size = 7, kernel filter = 16
	Conv1D	kernel size = 3, kernel filter = 16, kernel stride = 1	Conv1D	kernel size = 4, kernel filter = 16
	ResNet	kernel size = 5, kernel filter = 16	Concatenate	
	Conv1D	kernel size = 5, kernel filter = 32, kernel stride = 2	Informer	d_model = 512, n_heads = 8, e_layers = 3, d_layers = 2, dropout=0.01, attN = 'prob', embed = 'fixed', activation = 'gelu', factor = 5, freq = 'h'
ResNet	kernel size = 1, kernel filter = 32	Output		

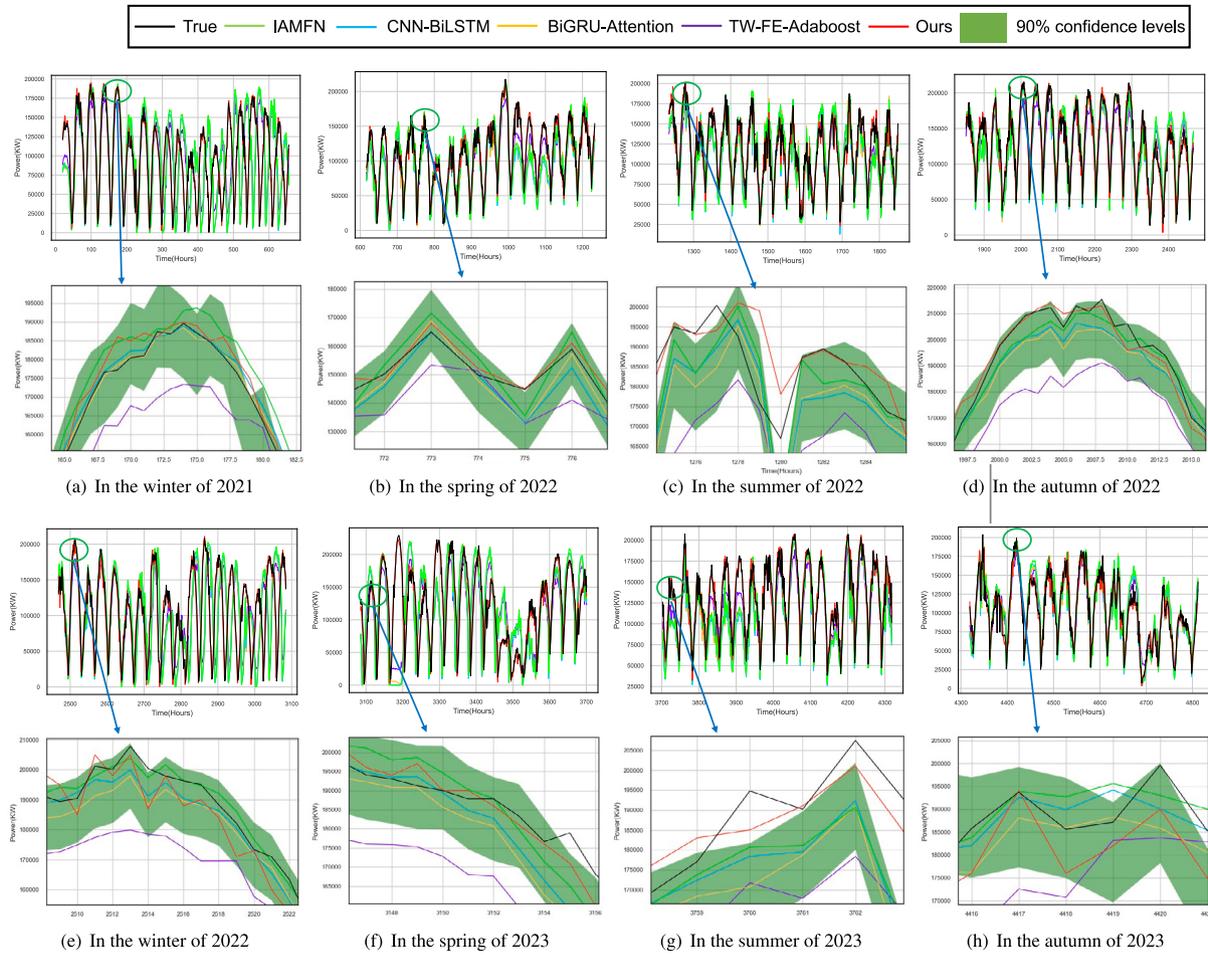


Fig. 17. The forecasting results of state-of-the-art models on photovoltaic.

Table 6

The forecasting results of ours.

Season	Type	Evaluation index					
		R <sup>2</sup>	MAE(KW)	MAPE(%)	T <sub>total</sub> (s)	PINAW	CPIA
Spring	Wind	0.969	2677.121	6.430	301.04	0.120	0.983
Summer	Wind	0.980	2263.515	7.110	269.14	0.090	0.963
Autumn	Wind	0.966	2667.857	6.470	344.97	0.098	0.952
Winter	Wind	0.979	3615.907	6.110	358.99	0.069	0.942
Spring	Photovoltaic	0.981	4930.973	8.810	339.81	0.121	0.963
Summer	Photovoltaic	0.951	6889.463	8.170	452.86	0.114	0.973
Autumn	Photovoltaic	0.982	5625.590	6.940	332.79	0.106	0.952
Winter	Photovoltaic	0.941	6846.052	5.800	327.99	0.134	0.922

the numbers of calculation and supporting high accuracy in contrast with forecasting each site individually. These results testify the strong performance of the hybrid Inception-ResNet embedded with Informer in short-term wind and photovoltaic power generation forecasting.

### 3.4.3. Comparison with state-of-the-art forecasting models

To further verify the superiority of our model, this paper selects five state-of-the-art models to confirm ours on day-ahead generation forecasting including IEDN-RNET proposed by Ling et al. [27], IAMFN developed by Zhao et al. [28], CNN-BiLSTM utilized by Liu et al. [17], BiGRU-Attention noted by Liu et al. [29], and TW-FE-Adaboost obtained by Xiao et al. [30]. Based on the above mentioned, error performances and prediction curves of each model are evaluated across the seasons converting spring, summer, autumn and winter.

The forecasting results curves of various models are illustrated in Figs. 16–17 where the lines with black, red, green, blue, yellow, and purple represent the true and predictions of the above models, namely IAMFN, CNN-BiLSTM, BiGRU-Attention, TW-FE-Adaboost and ours respectively. It is evident that the power curve of IAMFN, BiGRU-Attention, and our model are closer to the actual value indicating that the prediction errors of these are gradually declined compared with the rest.

In order to quantitatively evaluate the prediction performance of our model, various evaluation index, for instance, R<sup>2</sup> and MAPE along with T<sub>total</sub> are calculated to evaluate point prediction error and estimate the operation time and cost. For simplicity, the PINAW and CPIA are reported for interval prediction error evaluation. The error statistics for different models are enumerated in Table 8 and it is clear that the MAPE and PINAW of ours are smaller, the value of R<sup>2</sup> is larger than others while the T<sub>total</sub> differs not much from other models.

(1) Fig. 16 presents the results of various models in short-term wind power generation forecasting and in this case study ours achieves the lowest MAPE (6.110), the lowest PINAW (0.069), and the highest R<sup>2</sup> (0.979) on the winter dataset while the MAPE of BiGRU-Attention is 7.970, and the PINAW is 0.098, proving that the prediction values align well with the actual values. Furthermore, the MAPE of ours achieves 1.24 lower and the PINAW is 0.003 lower than BiGRU-Attention among the comparison models in summer. Our model reduces the MAPE by 2.09 and boosts 2.09 contrast to CNN-BiLSTM in autumn. These results demonstrate that our model achieves lower prediction errors and a stronger fit, which promotes the effectiveness and feasibility for wind generation forecasting.

**Table 7**  
The forecasting results of three forecasting frameworks.

Wind power generation							
Forecasting framework	Season	Evaluation index					
		R <sup>2</sup>	MAE(KW)	MAPE(%)	T <sub>total</sub> (s)	PINAW	CPIA
Cumulative prediction	Spring	<b>0.973</b>	3730.809	11.424	339.80	<b>0.091</b>	0.929
	Summer	0.914	4517.496	9.877	427.99	<b>0.068</b>	0.933
	Autumn	<b>0.982</b>	3299.321	12.655	<b>329.14</b>	0.072	0.946
	Winter	0.938	3038.657	12.905	<b>351.92</b>	0.080	0.915
Predicting cumulatively	Spring	0.922	3364.523	9.650	1126.85	0.096	0.951
	Summer	0.907	2933.918	14.430	1075.10	0.091	0.961
	Autumn	0.903	3739.685	13.840	1189.71	0.105	0.949
	Winter	0.899	4747.395	11.230	1089.52	0.095	0.941
Ours clustering prediction	Spring	0.969	<b>2677.121</b>	<b>6.430</b>	<b>301.04</b>	0.120	<b>0.983</b>
	Summer	<b>0.980</b>	<b>2263.515</b>	<b>7.110</b>	<b>269.14</b>	0.090	<b>0.963</b>
	Autumn	0.966	<b>2667.857</b>	<b>6.470</b>	344.97	<b>0.098</b>	<b>0.952</b>
	Winter	<b>0.979</b>	<b>2615.907</b>	<b>6.110</b>	358.99	<b>0.069</b>	<b>0.942</b>
Photovoltaic power generation							
Forecasting framework	Season	Evaluation index					
		R <sup>2</sup>	MAE(KW)	MAPE(%)	T <sub>total</sub> (s)	PINAW	CPIA
Cumulative prediction	Spring	0.956	10813.690	19.037	418.63	0.130	0.944
	Summer	0.865	11944.353	9.682	480.75	0.151	0.915
	Autumn	0.910	11215.267	9.656	421.91	0.134	0.916
	Winter	0.941	7130.944	20.190	429.14	0.143	<b>0.923</b>
Predicting cumulatively	Spring	0.882	7438.988	11.427	1765.84	0.182	0.916
	Summer	0.911	8240.810	11.317	1245.86	0.186	0.911
	Autumn	0.922	8247.107	16.268	1863.79	0.231	0.909
	Winter	0.902	8977.478	10.782	1201.05	0.173	0.917
Ours clustering prediction	Spring	<b>0.981</b>	<b>4930.973</b>	<b>8.810</b>	<b>339.81</b>	<b>0.121</b>	<b>0.963</b>
	Summer	<b>0.949</b>	<b>6889.463</b>	<b>8.170</b>	<b>452.86</b>	<b>0.114</b>	<b>0.973</b>
	Autumn	<b>0.982</b>	<b>5625.590</b>	<b>6.940</b>	<b>332.79</b>	<b>0.106</b>	<b>0.952</b>
	Winter	<b>0.941</b>	<b>6846.052</b>	<b>5.800</b>	<b>327.99</b>	<b>0.104</b>	0.922

(2) The short-term photovoltaic forecasting results for solar sites are obtained and it is provided in Table 9 that our model outperforms all of its competitors with R<sup>2</sup> of 0.982, MAPE of 6.949, and PINAW of 0.106, achieving a strong correlation between the prediction and actual values. However, IAMFN, CNN-BiLSTM, and BiGRU-Attention algorithms gain the MAPE of 9.055, 8.640 and 8.641 and PINAW of 0.119, 0.118, and 0.109, indicating obvious forecasting errors. Specifically, the performance of ours in summer is slightly worse than that in autumn, considering that power generation are more volatile due to significant variations in sunshine intensity during summer. Nonetheless, it demonstrates our superior performance compared to the state-of-the-art model the R<sup>2</sup> of 0.951, the MAPE of 8.170, and the PINAW of 0.114. Among the algorithms evaluated, IAMFN acquires 5.225 higher of MAPE, 0.002 higher in PINAW which highlights the reliable and effective improvement of ours. On balance, these comparative experiments verifies the strong performance of our model in short-term PV forecasting.

Fig. 18 shows the average MAPE and R<sup>2</sup> for the all cases. Additionally, detailed information on the case study results can be found in Table 8. In general, the proposed model, hybrid Inception-ResNet deep neural network embedded with Informer, is evaluated for wind and PV power forecasting using four different seasons, with spring and winter specifically related to wind power, and autumn and winter focused on PV. As shown in Fig. 18, ours exhibits superior performance compared to others in wind forecasting, which achieves a 5% lower MAE and a 6% lower MAPE than IAMFN. Furthermore, ours outperforms the CNN-BiLSTM model by 3% in terms of MAPE. Additionally, the R<sup>2</sup> of ours is 6% higher than that of the CNN-BiLSTM model. Moreover, hybrid Inception-ResNet embedded with Informer demonstrates a 4% higher R-square value than IAMFN and approximately a 5% improvement in R-square compared to both the CNN-BiLSTM and BiGRU-Attention models. In case of PV forecasting, ours continues to exhibit superior performance compared to IAMFN, which achieves a 6% lower MAPE than IAMFN. Furthermore, the R-square value of ours reaches 0.972,

which is 3.6% higher than that of the CNN-BiLSTM model. Additionally, ours is approximately 6% higher than that of BiGRU-Attention and TW-FE-Adaboost in R<sup>2</sup>, which demonstrates a very high correlation coefficient for both wind and PV power forecasting.

#### 3.4.4. Generalization experiment

To verify the robustness and generalization of our model, data from the European Center for Medium-range Weather Forecasts (ECMWF) [31] comprising of seven wind farm datasets of the same region in Europe and PVOOD [32] dataset of 10 photovoltaic power plants located in the Hebei Province of China are used to test our model.

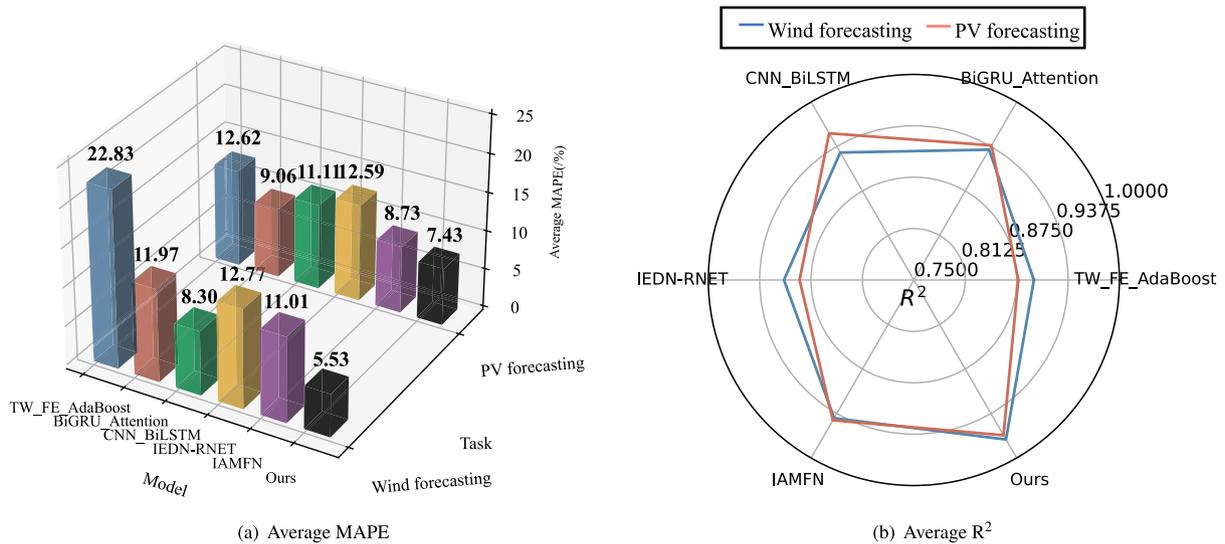
Fig. 19 shows the prediction results for all cases, and detailed information on the results of the case studies can be found in Table 9. In general, our model achieves R<sup>2</sup> of 0.985, MAPE of 8.465% and PINAW of 0.041 for wind power prediction and R<sup>2</sup> of 0.971, MAPE of 4.927% and PINAW of 0.028 for PV forecasting, which shows that it has small relative errors and good stability in the other scenarios, demonstrating the robustness and generalization of the model.

## 4. Conclusion

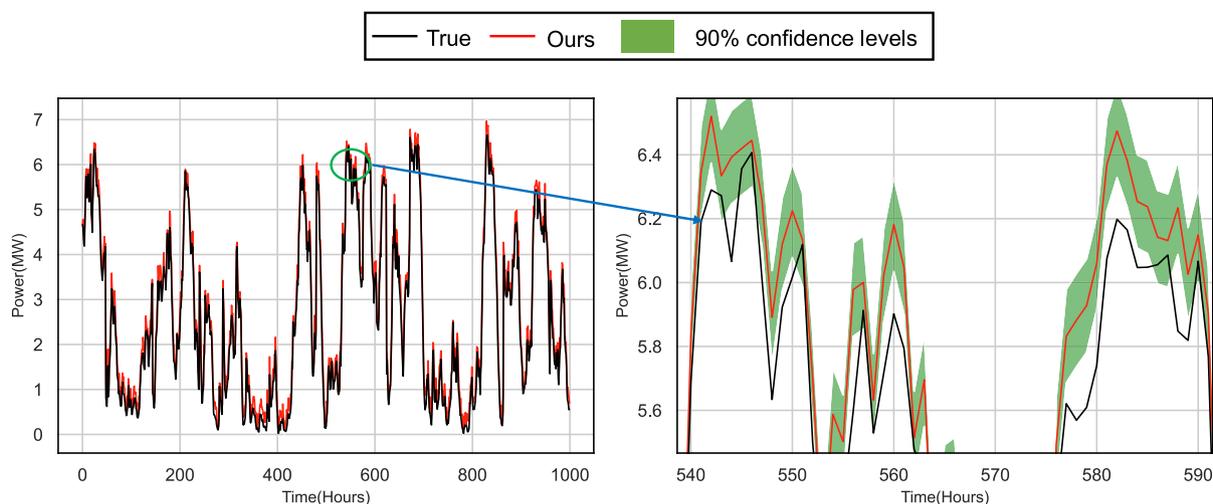
Our model proposed in this paper based on FCM clustering and hybrid Inception-ResNet demonstrates superior performance compared to other algorithms, which exhibits high accuracy and low computing time costs on wind and photovoltaic generation datasets. Initially, the FCM clustering algorithm is employed to partition multiple power generation sites into clusters and the clustering prediction approach ensures that meteorological and power data within each cluster are specific and consistent, which contributes to minimizing prediction errors and reducing computing time and costs compared to the frameworks of cumulative prediction and predicting cumulatively. These strikes a balance between prediction accuracy and calculation efficiency. Secondly, the Pearson correlation coefficient and Spearman rank correlation coefficient on Copula function are utilized to analyze

**Table 8**  
The forecasting results of state-of-the-art models.

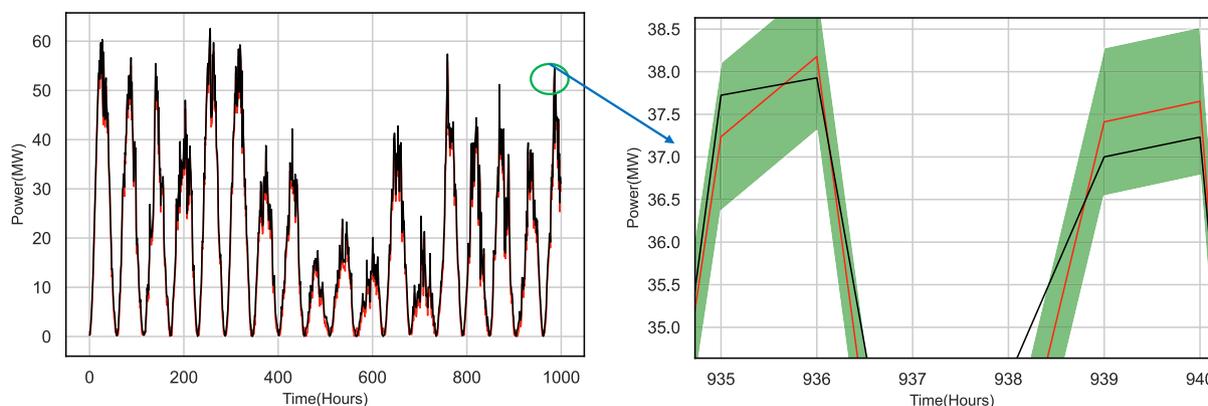
Wind power							
Season	Evaluation index	Ours	IEDN-RNET	IAMFN	CNN-BiLSTM	BiGRU-Attention	TW-FE-AdaBoost
Spring	R <sup>2</sup>	<b>0.969</b>	0.865	0.956	0.921	0.927	0.854
	MAPE(%)	<b>6.430</b>	13.070	9.911	7.664	8.517	16.416
	T <sub>total</sub> (s)	301.04	290.17	603.41	<b>286.86</b>	489.18	311.89
	PINAW	0.067	0.165	<b>0.058</b>	0.155	0.071	0.196
Summer	R <sup>2</sup>	<b>0.980</b>	0.901	0.924	0.945	0.968	0.925
	MAPE(%)	<b>7.110</b>	10.810	10.280	9.115	8.350	14.400
	T <sub>total</sub> (s)	269.14	234.49	599.40	288.17	576.07	<b>201.22</b>
	PINAW	<b>0.090</b>	0.164	0.096	0.094	0.093	0.139
Autumn	R <sup>2</sup>	<b>0.966</b>	0.928	0.963	0.941	0.928	0.949
	MAPE(%)	<b>6.470</b>	16.824	12.120	8.560	10.700	32.850
	T <sub>total</sub> (s)	344.97	<b>228.63</b>	578.16	278.16	579.10	280.91
	PINAW	<b>0.069</b>	0.164	0.076	0.086	0.077	0.142
Winter	R <sup>2</sup>	<b>0.979</b>	0.913	0.952	0.925	0.956	0.859
	MAPE(%)	<b>6.110</b>	10.370	11.720	7.877	7.970	27.660
	T <sub>total</sub> (s)	358.99	317.81	600.39	280.75	633.04	<b>201.98</b>
	PINAW	<b>0.069</b>	0.089	0.072	0.087	0.098	0.168
Photovoltaic generation							
Season	Evaluation index	Ours	IEDN-RNET	IAMFN	CNN-BiLSTM	BiGRU-Attention	TW-FE-AdaBoost
Spring	R <sup>2</sup>	<b>0.981</b>	0.902	0.952	0.967	0.940	0.911
	MAPE(%)	<b>8.810</b>	12.282	9.650	9.985	9.576	10.393
	T <sub>total</sub> (s)	339.81	338.75	291.16	289.11	499.68	<b>198.78</b>
	PINAW	<b>0.121</b>	0.207	0.178	0.138	0.129	0.266
Summer	R <sup>2</sup>	<b>0.951</b>	0.890	0.948	0.933	0.942	0.849
	MAPE(%)	<b>8.170</b>	11.920	9.230	13.365	8.695	15.770
	T <sub>total</sub> (s)	452.86	297.88	211.60	238.84	533.12	<b>199.67</b>
	PINAW	<b>0.114</b>	0.231	0.116	0.118	0.133	0.248
Autumn	R <sup>2</sup>	<b>0.982</b>	0.884	0.969	0.961	0.959	0.893
	MAPE(%)	<b>6.949</b>	12.880	9.055	8.640	8.461	11.330
	T <sub>total</sub> (s)	332.79	241.78	240.92	276.81	520.63	<b>201.32</b>
	PINAW	<b>0.106</b>	0.169	0.119	0.118	0.109	0.288
Winter	R <sup>2</sup>	0.941	0.898	0.942	<b>0.946</b>	0.933	0.893
	MAPE(%)	<b>5.800</b>	13.260	6.919	12.436	9.513	12.995
	T <sub>total</sub> (s)	332.79	263.89	335.29	254.95	473.53	<b>200.87</b>
	PINAW	<b>0.104</b>	0.156	0.123	0.112	0.132	0.256



**Fig. 18.** Evaluation index of state-of-the-art forecasting results.



(a) Our forecasting results in ECMWF



(b) Our forecasting results in PVOD

Fig. 19. Our forecasting in different dataset.

Table 9

Our forecasting in different dataset.

Our forecasting results in ECMWF						
Type	R <sup>2</sup>	MAE(MW)	MAPE(%)	T <sub>total</sub> (s)	PINAW	CPIA
Wind forecasting	0.985	0.198	8.465	325.19	0.041	0.971
Our forecasting results in PVOD						
Type	R <sup>2</sup>	MAE(MW)	MAPE(%)	T <sub>total</sub> (s)	PINAW	CPIA
PV forecasting	0.971	1.439	4.927	278.51	0.028	0.988

the linear and nonlinear correlation between power generation and their influencing factors and strong features are selected to establish the input set of our forecasting model. Additionally, hybrid Inception-ResNet deep neural network embedded with Informer consisting of CNN module, ResNet blocks, Inception modules, and Informer, which can effectively capture long-term dependency and nonlinear mapping relationship within generation sequences. With its outstanding multi-level feature extraction capability and the temporal features extracting capacity, our model demonstrates excellent performance in short-term wind and photovoltaic power prediction tasks. Finally, the GWO optimization algorithm is enhanced to improve the efficiency of model optimization to obtain optimal parameters, thereby strengthening the forecasting performance.

Overall, our focus will explore the lightweight application of Informer to reduce training time and conduct case studies on large-scale

clusters in different regions to emphasize research in large-scale power site clusters prediction.

### CRediT authorship contribution statement

**Daogang Peng:** Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Yu Liu:** Writing – original draft, Software, Resources, Project administration. **Danhao Wang:** Supervision, Software. **Ling Luo:** Validation, Supervision. **Huirong Zhao:** Validation. **Bogang Qu:** Visualization, Validation, Supervision.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Data availability

Data will be made available on request.

### Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 62373241) and the State Grid Shanghai Electric Power Company Science and Technology Project (No. 52094024001A).

## References

- [1] Qays MO, Ahmad I, Habibi D, Aziz A, Mahmoud T. System strength shortfall challenges for renewable energy-based power systems: A review. *Renew Sustain Energy Rev* 2023;113447.
- [2] Huang Y, Peng D, Yao J, Zhang H, Yu H. Ultra-short-term photovoltaic power forecast of TD-BP neural network based on SSA K-means. *Acta Energetica Sinica* 2021;42:229–38, China.
- [3] Jo HH, Kim J, Kim S. Enhancing the power generation performance of photovoltaic system: Impact of environmental and system factors. *Appl Therm Eng* 2022;240:122221.
- [4] Couto A, Estanqueiro A. Wind power plants hybridised with solar power: A generation forecast perspective. *J Clean Prod* 2023;423:138793.
- [5] Wang H, Zhang N, Du E, Yan J, Han S, Liu Y. A comprehensive review for wind, solar, and electrical load forecasting methods. *Glob Energy Interconnect* 2022;5(1):9–30.
- [6] Almazrouee AI, Almeshal AM, Almutairi AS, Alenezi MR, Alhajeri SN, Alshammari FM. Forecasting of electrical generation using prophet and multiple seasonality of holt-winters models: A case study of Kuwait. *Appl Sci* 2020;10(23):8412.
- [7] Abou Houran M, Bukhari SMS, Zafar MH, Mansoor M, Chen W. COA-CNN-LSTM: Coati optimization algorithm-based hybrid deep learning model for PV/wind power forecasting in smart grid applications. *Appl Energy* 2023;349:121638.
- [8] Hu Y, Liu H, Wu S, Zhao Y, Wang Z, Liu X. Temporal collaborative attention for wind power forecasting. *Appl Energy* 2024;357:122502.
- [9] Liu Q, Li Y, Jiang H, Chen Y, Zhang J. Short-term photovoltaic power forecasting based on multiple mode decomposition and parallel bidirectional long short term combined with convolutional neural networks. *Energy* 2024;286(C).
- [10] Mirza AF, Mansoor M, Usman M, Ling Q. Hybrid inception-embedded deep neural network ResNet for short and medium-term PV-wind forecasting. *Energy Convers Manage* 2023;294:117574.
- [11] Mansoor M, Mirza AF, Usman M, Ling Q. Hybrid forecasting models for wind-PV systems in diverse geographical locations: Performance and power potential analysis. *Energy Convers Manage* 2023;287:117080.
- [12] Che J, Yuan F, Deng D, Jiang Z. Ultra-short-term probabilistic wind power forecasting with spatial-temporal multi-scale features and K-FSDW based weight. *Appl Energy* 2023;331:120479.
- [13] Yang M, Huang Y, Guo Y, Zhang W, Wang B. Ultra-short-term wind farm cluster power prediction based on FC-GCN and trend-aware switching mechanism. *Energy* 2024;290:130238.
- [14] Liu Y, Liu Y, Cai H, Zhang J. An innovative short-term multihorizon photovoltaic power output forecasting method based on variational mode decomposition and a capsule convolutional neural network. *Appl Energy* 2023;343:121139.
- [15] Tyralis H, Papacharalampous G. A review of predictive uncertainty estimation with machine learning. *Artif Intell Rev* 2024;57(4):94.
- [16] Hou G, Wang J, Fan Y. Wind power forecasting method of large-scale wind turbine clusters based on DBSCAN clustering and an enhanced hunter-prey optimization algorithm. *Energy Convers Manage* 2024;307:118341.
- [17] Liu Q, Li Y, Jiang H, Chen Y, Zhang J. Short-term photovoltaic power forecasting based on multiple mode decomposition and parallel bidirectional long short term combined with convolutional neural networks. *Energy* 2024;286(C).
- [18] Lai Z, Wu T, Fei X, Ling Q. BERT4ST: Fine-tuning pre-trained large language model for wind power forecasting. *Energy Convers Manage* 2024;307:118331.
- [19] Li K, Mu Y, Yang F, Wang H, Yan Y, Zhang C. Joint forecasting of source-load-price for integrated energy system based on multi-task learning and hybrid attention mechanism. *Appl Energy* 2024;360:122821.
- [20] Liu D, Ding W, Dong ZS, Pedrycz W. Optimizing deep neural networks to predict the effect of social distancing on COVID-19 spread. *Comput Ind Eng* 2022;166:107970.
- [21] Ebrahimi SR, Rahimiyan M, Assili M, Hajizadeh A. Home energy management under correlated uncertainties: A statistical analysis through copula. *Appl Energy* 2022;305:117753.
- [22] Yet ZR, Masseran N, Ariff NM. Analyzing solar radiation characteristics using copula and solar power potential in peninsular Malaysia. *Arab J Geosci* 2022;15(23):1736.
- [23] Chen Y, Shen C, Gao L, Chen Z, Tao Q, Wang Y, Lin W, Zheng W, Yang Z, Chen Z, et al. Faraday angle prediction based on 1-D CNN on the J-TEXT tokamak. *Fusion Eng Des* 2024;199:114101.
- [24] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *IEEE*; 2016.
- [25] Wu T, Ling Q. Mixformer: Mixture transformer with hierarchical context for spatio-temporal wind speed forecasting. *Energy Convers Manage* 2024;299:117896.
- [26] Szegedy C, Liu W, Jia Y, Sermanet P, Rabinovich A. Going deeper with convolutions. *IEEE Computer Society*; 2014.
- [27] Mirza AF, Mansoor M, Usman M, Ling Q. Hybrid inception-embedded deep neural network ResNet for short and medium-term PV-wind forecasting. *Energy Convers Manage* 2023;294:117574.
- [28] Yin L, Zhao M. Inception-embedded attention memory fully-connected network for short-term wind power prediction. *Appl Soft Comput* 2023;141:110279.
- [29] Wang S, Shi J, Yang W, Yin Q. High and low frequency wind power prediction based on transformer and BiGRU-attention. *Energy* 2024;288:129753.
- [30] Xiao L, Dong Y, Dong Y. An improved combination approach based on Adaboost algorithm for wind speed time series forecasting. *Energy Convers Manage* 2018;160(3):273–88.
- [31] Zameer A, Arshad J, Khan A, Raja MAZ. Intelligent and robust prediction of short term wind power using genetic programming based ensemble of neural networks. *Energy Convers Manage* 2017;134:361–72.
- [32] Yao T, Wang J, Wu H, Zhang P, Li S, Wang Y, Chi X, Shi M. A photovoltaic power output dataset: Multi-source photovoltaic power output dataset with Python toolkit. *Sol Energy* 2021;230:122–30.