

A selective memory attention mechanism for chaotic wind speed time series prediction with auxiliary variable

Ke Fu, Shengli Chen[✉], Zhengru Ren^{✉*}

Institute for Ocean Engineering, Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, China

ARTICLE INFO

Keywords:

Wind speed prediction
Chaotic time series
Attention mechanism
Selective SSM
Wind energy

ABSTRACT

Wind speed prediction is crucial for enhancing wind energy utilization and optimizing grid integration of wind power. Its chaotic nature and the lack of correlated variables make accurate prediction difficult. Most studies rely solely on past wind speed, limiting accuracy improvements. While wind power is highly correlated with wind speed, this correlation is reversely causal. The key challenge is effectively leveraging this reverse causality between wind power and wind speed to enhance prediction precision. This study proposed SMAMnet to address the challenge mentioned, a model that establishes its backbone network via proposed new attention mechanism. The convolution operation is employed to restructure features, besides, the frequency-domain transformation and selective state space model (SSM) serves for attention weights. The novelty of SMAMnet is characterized by the development of an adaptive frequency-domain selected attention weight operator to adaptively parse meaningful information in different frequency domain intervals. Taking 15 min and 1-hour mean absolute error as the standard, the actual wind speed prediction error is reduced by 68% and 49% compared with the classic LSTM algorithm. The feasibility of mining reverse causality to improve prediction accuracy was verified.

1. Introduction

The over-exploitation and combustion of fossil fuels [1] have precipitated a host of environmental issues, propelling the utilize of renewable energy into the forefront of research [2]. Wind energy stands out as a sustainable and clean alternative [3], with wind power generation being an important application [4]. Nevertheless, wind speed is characterized by its intermittency, volatility, and instability [5,6], leading to correspondingly fluctuant wind power output [7]. This characteristic imposes challenges on the power system, requiring redundant backup capacity and consequently increasing construction and operational costs [8,9].

To address these challenges, accurately predicting wind speed offers a promising solution [10,11]. Such predictions provide foresight for the operation of wind turbines and the scheduling of wind power, thereby enhancing the availability and reliability of wind energy [12]. Consequently, enhancing the precision of wind speed prediction and the temporal scope of these predictions are crucial for the widespread adoption of wind power generation [13]. Nonetheless, the non-stationary nature of wind poses significant difficulties in achieving high-precision wind speed prediction [14,15].

In early research, four primary methodologies were employed [16]: physical modeling, statistics [17], traditional machine learning, and

deep learning [18,19]. Physical modeling uses parameters such as numerical weather prediction (NWP), topographical data, and meteorological variables to construct complex mathematical and physical models for wind speed prediction [20,21]. This method is lauded for its interpretability [22] but necessitates extensive data and computational resources. The accuracy of NWP, however, is compromised by uncertainties in model equations, physical approximations, and initial and boundary conditions [23], potentially leading to significant prediction errors from minor NWP deviations [24]. Statistical methods, on the other hand, involve fitting functions based on the correlations between wind speed data and associated factors [25]. Notably, autoregressive moving average (ARMA) [26] and autoregressive integrated moving average (ARIMA) [27] are frequently utilized statistical models. These methods, however, struggle with capturing the nonlinear relationships among meteorological elements, resulting in limited prediction accuracy that often falls short of the requirements for wind farm operations [28].

Traditional machine learning approaches for wind speed prediction include support vector machines (SVM) [29], least squares support vector machines (LSSVM) [30], Gaussian process regression [31], and various hybrid models [32–35], etc. Despite their utility, these methods are limited in their ability to extract temporal dependencies and deep

* Corresponding author.

E-mail address: zhengru.ren@sz.tsinghua.edu.cn (Z. Ren).

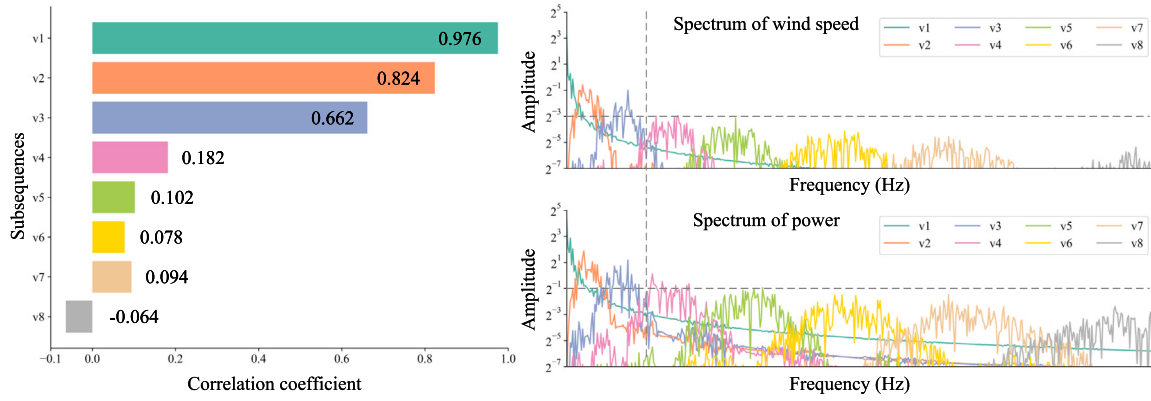


Fig. 1. Graphical representation of the correlation between wind speed and wind power for different frequency intervals. The low-frequency component is dominant in amplitude and shows strong correlation, as evidenced by both parts of the analysis.

nonlinear features from wind speed data [36]. In response to these limitations, numerous deep learning architectures have been applied to wind speed prediction [37,38], including long short-term memory networks (LSTM) [39] and time-frequency recurrent networks [40]. The sequential nature of recurrent neural network (RNN), however, limits parallel computation, prompting researchers to explore models based on attention mechanisms, such as the transformer [41] and multi-resolution interactive transformer [42]. To address the quadratic computational complexity inherent in transformer models, alternatives like Fastformer [43], which employs linear attention, and Mamba [44], grounded in the state space model (SSM), have been introduced. While deep learning models are well-suited for mining patterns in multivariate, nonlinear data, the challenge of wind speed prediction is often complicated by a scarcity of relevant predictive variables.

A primary challenge stems from the scarcity of effective auxiliary variables in wind speed prediction. Many potential variables show minor correlation with wind speed [45,46], forcing models to largely bypass multivariate analysis. This often results in a reliance on historical wind speed data [47,48], a significant constraint. Most studies only use a single variable of wind speed, which limits the further improvement of prediction accuracy. Data analysis has revealed a strong correlation between wind turbine power output and wind speed. However, this relationship is unidirectional: wind speed influences wind turbine power, but not vice-versa, especially the difference of cut-in and cutoff wind speeds in wind turbine controllers. This suggests a reverse causal relationship, which makes it difficult to use wind power as an auxiliary variable.

The conceptualization of reverse causality in predictive models is challenging, where the physical meaning is difficult to explain. To the best of the author's knowledge, this approach has not been explored in wind speed research. Consequently, it is worth considering whether incorporating wind turbine power as an auxiliary variable could enhance the precision of wind speed predictions, further, choosing a modeling strategy to effectively utilize the wind power information [49]. In other words, the interesting question arises: how can we optimally leverage wind turbine power data to improve wind speed prediction?

The motivation for this research arises from the observed similarity between wind turbine power and wind speed data across various frequencies. Fig. 1 illustrates the frequency spectrum of each variable in wind turbine power and wind speed data following variational mode decomposition (VMD) decomposition. Drawing inspiration from this observation, a frequency domain selection attention operator was developed that adaptively identifies the mapping relationships between related frequencies. Then, the memory effect subsequent to the frequency domain selection by incorporating hidden states was considered. Recognizing the inconsistent influence of various variables, a convolution operator was employed to reconstruct the original data based on their respective weights. The comprehensive solution was

then modeled within the attention mechanism framework that can be computed in parallel.

Our contributions are as follows:

- The SMAMnet was proposed, incorporating a novel selective memory attention mechanism specifically designed to leverage the reverse causality inherent in wind speed prediction tasks.
- The utility of employing wind turbine power as an auxiliary variable for wind speed prediction was validated, thereby mitigating the challenge of limited data availability inherent to this field.
- An innovative strategy for calculating attention weights was developed, integrating the memory effect characteristic of state space models with frequency-domain selection capabilities. This approach enables the adaptive parsing of salient information across distinct frequency domain intervals.
- The proposed method demonstrates not only high prediction accuracy but is also engineered for parallel computation, significantly enhancing its practical applicability.

The accuracy of the proposed method was verified through real-world data, and its potential for engineering applications was demonstrated by high-precision predictions for 15 min and 1 h.

The rest of the paper is organized as follows: Section 2 outlines the problem and the method related to this research. Section 3 describes the proposed SMAMnet model. Section 4 elaborates on the experimental design and the dataset utilized. Section 5 presents an analysis of the experimental results. Finally, Section 6 concludes the paper and discusses future research plans.

2. Preliminary

2.1. Variable exploration

The C-C method was used to calculate the optimal delay time τ of wind speed, $\tau=13$. Then, the Lyapunov-Wolf method was used to calculate that the maximum Lyapunov exponent of the wind speed time series is 0.064(positive), which means that the wind speed time series belongs to a chaotic time series.

Assume that the original wind speed time series is $X_S(t_k)$ and the corresponding wind power time series is $X_P(t_k)$, where t_k represents time step. Through the VMD method, we decompose these two time series into eight subsequences, labeled $S_{vi}(t_k)$ and $P_{vi}(t_k)$, respectively, where $i = 1, 2, \dots, 8$.

Please note that VMD is applied solely during the data analysis stage to illustrate the data characteristics and is not incorporated into the actual modeling process. And the division of subsequences is not rigidly fixed at 8 groups. In practice, the number of subsequences can be adaptively modified to optimize matching accuracy as required.

For each subsequence $S_{vi}(t_k)$ and $P_{vi}(t_k)$, we calculate its Fourier transform $\mathcal{F}[S_{vi}(t_k)]$ and $\mathcal{F}[P_{vi}(t_k)]$ to obtain the spectrum. The spectrum can be expressed as:

$$S_S(f_i) = |\mathcal{F}[S_{vi}(t_k)]|, \quad S_P(f_i) = |\mathcal{F}[P_{vi}(t_k)]|, \quad (1)$$

where f_i is the frequency, $S_S(f_i)$ and $S_P(f_i)$ are the spectra of the wind speed and wind power subsequences, respectively.

As shown in the right part of Fig. 1, for all i , $f_1 < f_2 < \dots < f_8$, that is, the subsequences are arranged from low to high frequency. Due to the large difference in spectrum amplitude, the vertical axis is represented by logarithmic coordinates. Since the sampling interval is 15 min, the frequency is generally low, so the frequency on the horizontal axis represents a relative value.

To quantify the relationship between the wind speed and wind power subsequences, the Pearson correlation coefficient for each corresponding pair is computed. The left panel of Fig. 1 reveals that in the low-frequency range, the correlation coefficients for the top three pairs of wind speed and wind power subsequences are 0.976, 0.824, and 0.662, respectively, signifying a strong correlation. Conversely, in the high-frequency range, the correlation coefficients for the remaining five pairs are all below 0.2, suggesting that the data in these frequency ranges are largely uncorrelated.

This results indicate that the wind power data contains the variation pattern of wind speed, which worth further investigation. Thus, frequency domain adaptive selection emerges as a promising approach.

2.2. Selective state space models

State space models have emerged as a class of promising methods for sequence modeling. For example, the recently proposed Mamba [44] model is a selective state space-based sequence modeling method with linear time complexity.

In the Mamba architecture, compared with the traditional state space model, an interesting design is the selection mechanism to filter out irrelevant information and remember relevant information indefinitely. First, we introduce the state space model. The inspiration for the state space model comes from mapping a system through its hidden state $h(t) \in \mathbb{R}^N$ to a function or sequence $x(t) \in \mathbb{R} \mapsto y(t) \in \mathbb{R}$. The four parameters (Δ, A, B, C) in the model are defined:

$$h_t = \bar{A}h_{t-1} + \bar{B}x_t, \quad (2)$$

$$y_t = Ch_t, \quad (3)$$

$$\bar{K} = (C\bar{B}, C\bar{A}\bar{B}, \dots, C\bar{A}^{k-1}\bar{B}, \dots), \quad (4)$$

$$y = x * \bar{K}, \quad (5)$$

where k represents the discretized time step. The calculation method of Δ is $\Delta_t = \text{softplus}(\text{Linear}(x_t))$. The continuous parameter (Δ, A, B) can be mapped to the discrete parameter ($\bar{A}, \bar{B}$) by the fixed formulas:

$$\bar{A} = \exp(\Delta A), \quad (6)$$

$$\bar{B} = (\Delta A)^{-1}(\exp(\Delta A) - I) \cdot \Delta B. \quad (7)$$

Then, the selective mechanism used in the Mamba model is introduced. The background of this method is derived from the thinking of sequence modeling, that is, to selectively compress the necessary context information into a limited state. In this way, the trade-off between the efficiency and effectiveness of the sequence model depends on how to compress the context information. The selectivity mechanism algorithm of the model is described in detail in the literature [44]. The idea of this selectivity mechanism is to make the parameters that affect sequence interactions depend on the input.

2.3. Attention mechanism

The attention mechanism realizes global context modeling and can be calculated in parallel. Therefore, it is popular in time series modeling. The Transformer first proposed the attention mechanism and has been tested for a long time. For example, the underlying architecture of the large language model ChatGPT uses Transformer. In order to solve the problem of quadratic computational complexity, Fastformer was proposed. Furthermore, the Mamba model also has similar ideas to the attention mechanism. Coincidentally, the SSM in the previous section originated from the Mamba model. Therefore, in Fig. 2, the above three types of models are described, and they are compared and analyzed below.

(1) Transformer

As shown in Fig. 3 (a), the architecture of Transformer is based on simulating the pairwise relationship between each element and other elements. Transformer adopts a multi-head attention mechanism, where h is the number of scaled dot product attentions, which are then concatenated together and mapped to the output through the feed forward network. Specifically, the scaled dot product attention can be expressed as

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \quad (8)$$

where Q , K , and V represent query, key, and value after linear projection transformation, that is, the expansion and transformation of the input sequence. In addition, d_k is the data dimension.

The calculation process involved in this formula is consistent with Fig. 3 (a), and the multiplication calculation uses matrix dot multiplication. First, the query and key are dot-product and then divided by $\sqrt{d_k}$, where $\sqrt{d_k}$ represents the scale. After that, the softmax function is used to calculate the attention weight, which is multiplied by the value to get the output of the scaled dot-product attention.

(2) Fastformer

Although Transformer is a pioneer of the attention mechanism, the matrix dot product of pairwise relationships brings quadratic computational complexity, which has attracted many researchers to simplify it. For example, the basic idea of Fastformer is to use additive attention to replace the multiplicative attention in Transformer. Therefore, the difference between Fastformer and Transformer two mainly focuses on the calculation of attention weights, as shown in Fig. 3 (b).

$$\alpha_i = \frac{\exp(\mathbf{w}_q^T \mathbf{q}_i / \sqrt{d})}{\sum_{j=1}^N \exp(\mathbf{w}_q^T \mathbf{q}_j / \sqrt{d})}, \quad \mathbf{q} = \sum_{i=1}^N \alpha_i \mathbf{q}_i, \quad (9)$$

$$\beta_i = \frac{\exp(\mathbf{w}_k^T \mathbf{p}_i / \sqrt{d})}{\sum_{j=1}^N \exp(\mathbf{w}_k^T \mathbf{p}_j / \sqrt{d})}, \quad \mathbf{k} = \sum_{i=1}^N \beta_i \mathbf{p}_i. \quad (10)$$

The above two formulas are the attention calculation process of Fastformer that is different from Transformer, where W_q and W_k represent the matrices after query and key transformation respectively.

(3) Mamba

The architectural design of the Mamba network is also related to the attention mechanism. This is because the Mamba architecture is a variant of the H3 architecture [50] and Gated MLP, as shown in Fig. 3 (c). The H3 architecture integrates the state space model into the attention mechanism.

3. SMAMnet

In this study, SMAMnet is proposed to solve the problem of wind speed prediction. The proposed model not only improves the prediction accuracy but also expands the range of available variables in wind speed prediction by exploiting reverse causality.

Consider the input $X \in \mathbb{R}^{n \times d}$, where n is the number of time steps and d is the variable dimension. As shown in Fig. 4, the proposed

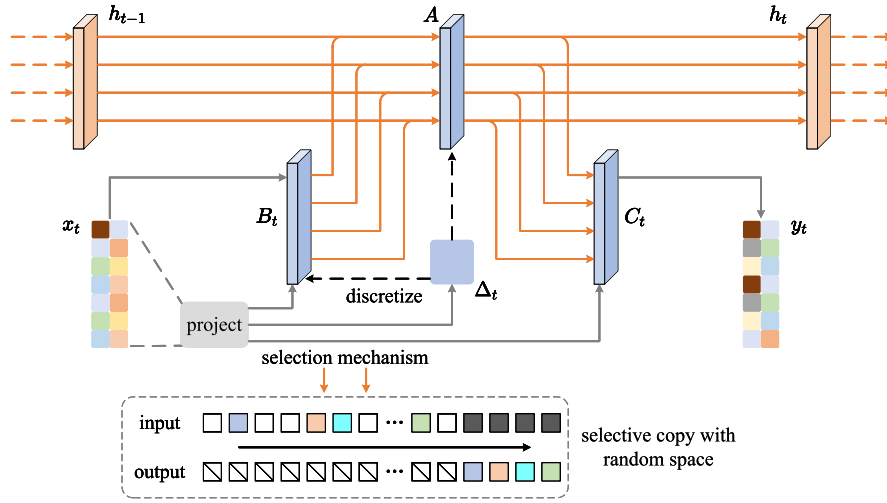


Fig. 2. The architecture of a state space model with selection mechanisms.

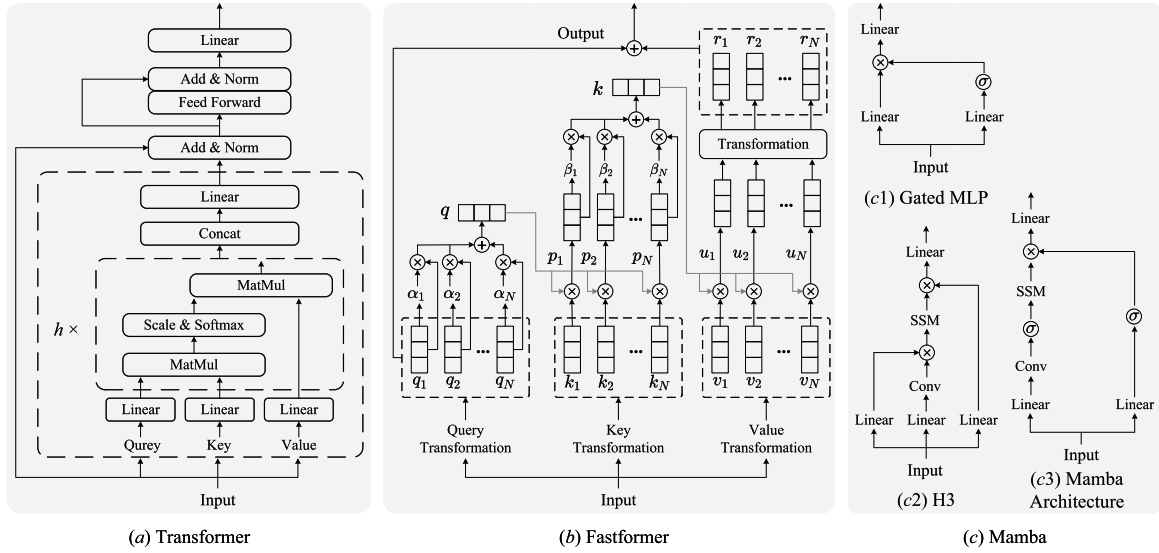


Fig. 3. (a) and (b) are the architectures of representative attention mechanisms, including Transformer and Fastformer. (c) describes the relationship between Mamba and the attention mechanism.

model accepts X and then normalizes it. Subsequently, the model automatically checks whether X is one-dimension, in other words, checks $d = 1$. If $d = 1$, X is input into the proposed attention model with improved weight operator, and then Flatten the tensor. After inverse normalization, the prediction result is output. If $d \neq 1$, the difference from the previous steps is that there is a feature reconstruction module prior to the attention model. The process can be described as

$$X_{out} = \text{Flatten}(\text{Attention}(X_{in})), \quad \text{if } d = 1, \quad (11)$$

$$X_{out} = \text{Flatten}(\text{Attention}(\text{Convfr}(X_{in}))), \quad \text{if } d \neq 1. \quad (12)$$

This feature reconstruction module is designed to alleviate the weight differences between different variables. Specifically, 1×1 convolution is used to transform different variables separately, and each variable uses a set of single-channel convolution kernel. That is

$$X'_{in} = \text{Concatenate}(\text{Conv}_{1 \times 1}(X_{in}[:, 1]), \text{Conv}_{1 \times 1}(X_{in}[:, 2]), \dots, \text{Conv}_{1 \times 1}(X_{in}[:, d])). \quad (13)$$

The proposed method centers on an attention mechanism equipped with a novel weight operator, and its architectural foundation resembles the Fastformer. Upon receiving the input features, the method

initiates processing with a linear transformation via a fully connected layer. It is important to note that the fully connected layer contains three times the number of neurons as the required number of channels. Expressed mathematically, assume the input feature is $X_1 \in \mathbb{R}^{n \times F}$, the tensor after linear mapping is $Z \in \mathbb{R}^{n \times 3F}$, where F is the feature dimension. That is

$$Z = W_z \cdot X_1 + b_z, \quad W \in \mathbb{R}^{3F \times F}, b \in \mathbb{R}^{3F}. \quad (14)$$

Then the obtained tensor Z is divided into query, key, and value tensors according to the number of channels, where $Q = Z[:, : F]$, $K = Z[:, F : 2F]$, $V = Z[:, 2F : 3F]$.

The subsequent calculation process of query includes two steps. First, calculate a set of weights α of query through the softmax function, and use α to weighted summation to get the tensor q as in Eq. (9). Second, calculate q with the selective memory weight operator to obtain W_q . Specifically for the proposed selective memory weight operator, the process begins by applying FFT (fast fourier transform) to the input tensor, transitioning it from the time domain to the frequency domain. Subsequently, a fully connected layer is employed to assign weights to the various frequency domain tensors, adaptively selecting appropriate frequency domain information. Following this, the weighted frequency

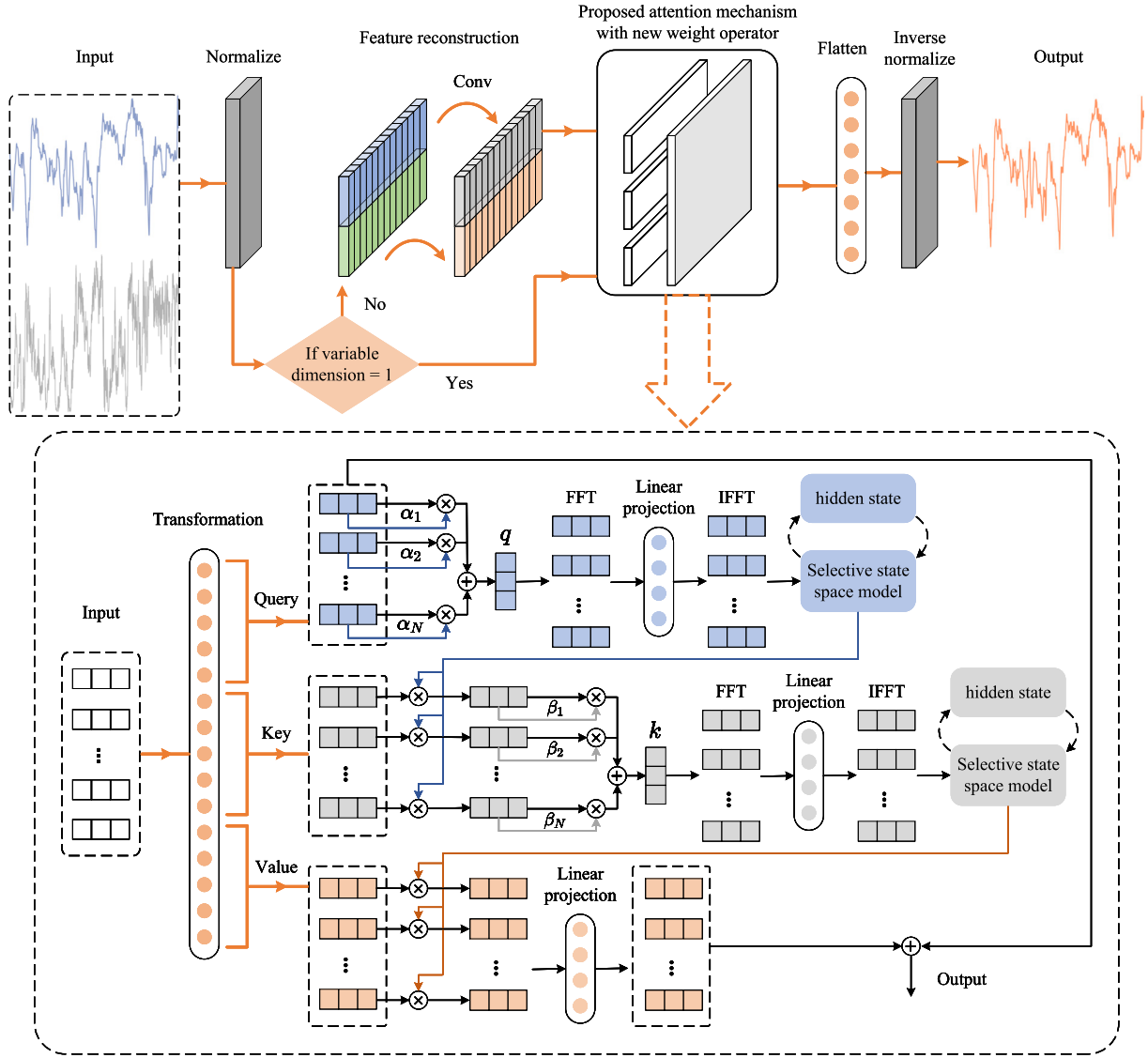


Fig. 4. Graphical illustration of the proposed SMAMnet architecture. The primary elements include a feature reconstruction module for handling variable dimensions and an innovative proposed attention mechanism with new weight operator.

domain data is inverted back to the time domain via IFFT (inverse fast fourier transform). The resulting time domain information is then fed into the selective state space model (Selective SSM), where the embedded selectivity mechanism facilitates the storage of adaptively selected information within the hidden state. That is

$$W_q = \text{Selective SSM}(\text{IFFT}(\text{NN}_{fc}(\text{FFT}(q)))) \quad (15)$$

The subsequent calculation process of key includes three steps. First, multiply key with W_q . Second, calculate β just like α , and get the tensor $\mathbf{k}$ as in Eq. (10). Third, calculate $\mathbf{k}$ with the selective memory weight operator to obtain W_k . The vector $\mathbf{k}$ is transformed into the frequency domain via FFT. Subsequently, it undergoes a linear projection using a fully connected neural network. After that, it is converted back to the time domain through IFFT. Finally, the resulting vector is computed as W_k using the Selective SSM, as described in Section 2.2. That is

$$W_k = \text{Selective SSM}(\text{IFFT}(\text{NN}_{fc}(\text{FFT}(\mathbf{k})))) \quad (16)$$

The subsequent calculation process of value includes three steps. First, multiply value with W_k . Second, use the fully connected layer to perform linear projection calculation. Third, the linear transformation

result is added to the query and output. That is

$$\text{Attention}(X_1) = \text{Add}(\text{NN}_{fc}(W_k \cdot V), Q). \quad (17)$$

All product operations throughout this process are conducted using element-wise multiplication. The selective state space model employed is consistent with that used in recently proposed Mamba, enabling parallel computation.

The Fourier transform [51] can be used to convert the time series from the time domain to the frequency domain. The discrete Fourier transform (DFT) is described by the following formula:

$$X_k = \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi j}{N} nk}, \quad k = 0, 1, \dots, N-1, \quad (18)$$

where x_n , X_k are the time-domain series and frequency-domain series, respectively. In addition, N is the length of x_n , and j is the imaginary unit.

The fast fourier transform (FFT) [52] is often used to calculate the DFT. This study used built-in functions in TensorFlow to calculate FFT and IFFT (inverse fast fourier transform).

In this paper, the Adam optimizer was used to update the network parameters during training, and the mean square error (MSE) loss

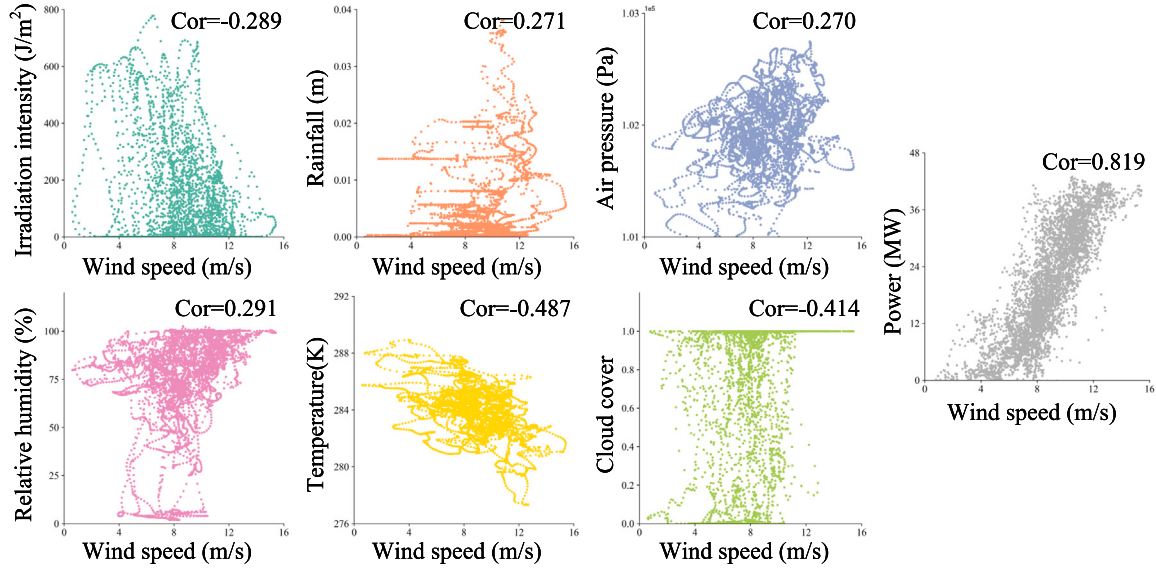


Fig. 5. Graphical illustration of the correlation between different variables and wind speed. The Cor represent Pearson correlation coefficient.

function was used for optimization, which is defined as:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (t_i - p_i)^2, \quad (19)$$

where t_i is the true value of the time series, p_i is the predicted value of the time series, and n is the length of the time series.

4. Experiment

4.1. Experimental configuration

All deep learning experiments were implemented in Tensorflow 2.10.0 platform and Python 3.12, while machine learning models used Sklearn 1.0.2. The computer configuration used includes Intel i7-13700k CPU, Nvidia Quadro A2000 GPU with CUDA 11.2 and 32 GB RAM.

The number of channels for all deep learning models is set to 24 to ensure fair comparison, which is consistent with the size of the sliding window, because 24 sampling points represent 6 h of data fluctuation in this study. In addition, the learning rate η of the Adam optimizer is 0.001, the batch size is 4096 during training, and the epochs is 2000. It is worth noting that the above key hyperparameters are applicable to all compared deep learning models to ensure fairness.

4.2. Description of dataset

The data used in this study comes from a national offshore wind power research and testing base in Fujian, China. The dataset includes: basic information of wind farms, meteorological variable data and actual power data. Meteorological variable data include air pressure, relative humidity, cloud cover, wind speed, wind direction, temperature, radiation intensity, and precipitation. The actual power data describes the power output of each wind farm. All data are collected from January 2022 to January 2024, with a sampling interval of 15 min. The initial 50 days data of a 48MW wind farm is used in this study, and the training set, validation set, and test set are 35 days, 5 days, 10 days respectively. In order to eliminate unnecessary rapid random fluctuations and reflect the data patterns, this study used the Savitzky–Golay filter to smooth the data. Fig. 5 illustrates the correlation of each variable with the wind speed time series, revealing a low correlation between wind speed and most variables, except for wind power.

4.3. Baseline

This study establishes a comprehensive baseline for comparison by employing a diverse array of models. The machine learning approaches include support vector machines (SVM), Gaussian processes (GP), and decision trees (DT). The deep learning models include the classic LSTM, gate recurrent unit (GRU), temporal convolutional network (TCN), Transformer, as well as hybrid models CNN-LSTM, CNN-GRU, BiLSTM, and ConvLSTM. Furthermore, the recently proposed Fnet, Fastformer, SCINet and Mamba are also involved in the comparison.

Owing to the constraints of manuscript length, we limit our detailed graphical presentation to three representative models beyond proposed model: GRU (as an RNN variant), Fastformer (as an attention mechanism-based model), and Mamba (as a state space model).

4.4. Ablation experiments

The ablation experiments consisted of four groups, the feature reconstruction module is removed, labeled SMAMnet-1; the selective state space model is removed, labeled SMAMnet-2; the adaptive frequency domain selection module is removed, labeled SMAMnet-3; and the selective state space model and the adaptive frequency domain selection module are removed, labeled SMAMnet-4.

4.5. Performance metrics

This study uses six evaluation metrics [53,54] to measure the accuracy of model prediction, including mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), symmetric mean absolute percentage error (SMAPE), Spearman correlation coefficient (ρ), and coefficient of determination (R^2). In addition, the following formulas are given.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |t_i - p_i|, \quad \text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (t_i - p_i)^2}, \quad (20)$$

$$\text{MAPE} = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{t_i - p_i}{t_i} \right|, \quad \text{SMAPE} = \frac{100\%}{n} \sum_{i=1}^n \frac{|p_i - t_i|}{(|p_i| + |t_i|)/2}, \quad (21)$$

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}, \quad R^2 = 1 - \frac{\sum_{i=1}^n (t_i - p_i)^2}{\sum_{i=1}^n (t_i - \bar{t})^2}, \quad (22)$$

Table 1

The 15-min wind speed prediction results using a single variable between the proposed model and the baseline models.

Category	Model	MAE	RMSE	MAPE	SMAPE	ρ	R^2
Model proposed	SMAMnet	0.045832	0.060014	0.481401	0.482051	0.999613	0.999317
Existing model	SVM	0.719856	0.957071	6.809890	6.958154	0.979452	0.826402
	GP	0.110704	0.319723	0.908459	0.934778	0.992877	0.980627
	DT	0.124754	0.307280	1.099670	1.121439	0.998134	0.982105
	LSTM	0.100540	0.125200	1.076634	1.081565	0.998648	0.997029
	GRU	0.104025	0.128383	1.130230	1.132501	0.998435	0.996876
	TCN	0.224602	0.253442	2.241757	2.263318	0.997639	0.987826
	CNN-LSTM	0.112546	0.146967	1.186666	1.191111	0.998337	0.995906
	CNN-GRU	0.097823	0.120621	1.057205	1.058958	0.998654	0.997243
	BiLSTM	0.192823	0.237996	2.074023	2.082059	0.994703	0.989265
	ConvLSTM	0.122559	0.160361	1.267177	1.269126	0.998111	0.995126
	Transformer	0.349071	0.436068	3.719302	3.728290	0.982536	0.963962
	Fnet	0.337561	0.414525	3.632974	3.646384	0.983150	0.967434
	Fastformer	0.189622	0.233974	2.051442	2.057123	0.994796	0.989625
	SCINet	0.381709	0.465931	4.028569	4.043189	0.981799	0.958857
	Mamba	0.350890	0.438179	3.738423	3.747411	0.982354	0.963612
Ablation test	SMAMnet-1	0.045832	0.060014	0.481401	0.482051	0.999613	0.999317
	SMAMnet-2	0.049225	0.066823	0.520520	0.521259	0.999486	0.999154
	SMAMnet-3	0.099172	0.127624	1.049177	1.053343	0.998701	0.996913
	SMAMnet-4	0.114046	0.149306	1.218474	1.223810	0.998074	0.995775

where t_i and p_i represent the true value and the predicted value of the time series, respectively. Besides, n is the length of the time series, d_i is the difference between the actual and predicted values, and $\bar{t}$ is the mean value of the time series. The values of RMSE, MAE, MAPE and SMAPE are close to 0, and the values of ρ and R^2 are close to 1, indicating that the model has high accuracy. In addition, the training and inference times of the proposed model are not the longest and within a reasonable range, so they are not discussed later.

5. Results and discussion

5.1. The 15 min wind speed prediction using a single variable

Table 1 summarizes the performance differences of 15 min wind speed prediction between the proposed model and various baseline models using a single variable. Among the machine learning methods, GP and DT have advantages and disadvantages under different prediction error evaluation metrics. Specifically, GP demonstrates a slight advantage in MAE, MAPE, and SMAPE, whereas DT outperforms slightly in RMSE, ρ , and R^2 . RMSE weights the error and is therefore sensitive to some large deviations, which indicates that large deviations are relatively rare in the prediction error of DT. In deep learning methods, CNN-GRU achieves the best performance generally, the inharmonious is that the MAPE and SMAPE of the GP are lower. Nevertheless, since evaluation metrics have different focuses and weighting rules, no evaluation metric is completely perfect, which does not affect the conclusion. Compared with the optimal value of each evaluation metric of the existing model, the prediction accuracy of the proposed SMAMnet is reduced by 53.15%, 50.25%, 47.01% and 48.43% in MAE, RMSE, MAPE and SMAPE respectively, and the deviations of ρ and R^2 relative to the perfect prediction are reduced by 71.25% and 75.23% respectively.

The model architecture analysis reveals that GP and DT within the machine learning category exhibit higher prediction accuracy, surpassing even some deep learning models. This phenomenon may be alleviated by increasing the number of training epochs or channels in deep learning models. Among deep learning models, those RNN variants (e.g., LSTM, GRU, CNN-GRU) generally outperform attention mechanism and state space models, aligning with previous research. However, the lack of parallel computing support in RNN constrains their appeal for research. Consequently, despite RNN seems to perform superior in sequence modeling tasks, there has been a growing of research on attention mechanisms and state space models.

Ablation experiments. In the ablation experiment, the prediction results of SMAMnet-1 are exactly the same as those of the proposed model, which is because the multivariate feature reconstruction module is inactive in univariate prediction scenarios. Across all prediction evaluation metrics, the performance of SMAMnet-2, SMAMnet-3, and SMAMnet-4 deteriorates sequentially, suggesting that both SSM and frequency domain selection modules are important and significantly contribute to the predictive accuracy of the model.

Graphical prediction results. Fig. 6 shows the prediction results of 15 min wind speed using a single variable for four representative models, GRU, Fastformer, Mamba and SMAMnet. According to the comparison between the predicted value and the actual value, GRU, Fastformer and the proposed SMAMnet can all predict the wind speed curve well, while the prediction accuracy of Mamba is visually poor. The above view is also proved by the prediction error curve. Mamba has the largest prediction error, and the prediction error of SMAMnet is significantly lower than that of the other three models. The scatter plot illustrate that the distribution of the predicted value and the actual value of SMAMnet is almost the same, followed by GRU, Fastformer appears loose, and Mamba appears significantly scattered, with a large deviation from the theoretical fitting line. In addition, The probability distribution of the error show that GRU, Fastformer and SMAMnet deviate less from the center value, and SMAMnet has the smallest distribution variance, which is 0.057. In contrast, the distribution variances of GRU, Fastformer and Mamba increased by 122.81%, 308.77% and 664.91%, respectively. The prediction error of SMAMnet is close to a normal distribution with notably concentrated pattern. The above graphical analysis shows that SMAMnet has excellent prediction performance.

5.2. The 15 min wind speed prediction by adding auxiliary variable

Table 2 describes the performance differences of 15 min wind speed prediction between the proposed model and the baseline models by adding auxiliary variable. Among the machine learning methods, DT has the best prediction accuracy, followed by GP, and SVM is the worst. Compared with the single variable prediction, a noteworthy phenomenon has emerged, that is, the prediction accuracy of the three machine learning methods generally decreases when auxiliary variables are added. For example, the R^2 of adding auxiliary variables is 0.978462, 0.956802, and 0.817865 from high to low, while the R^2 of single variables is 0.982105, 0.980627, and 0.826402. This shows that the ability of machine learning methods to process multivariate information in wind speed prediction tasks is limited, which means

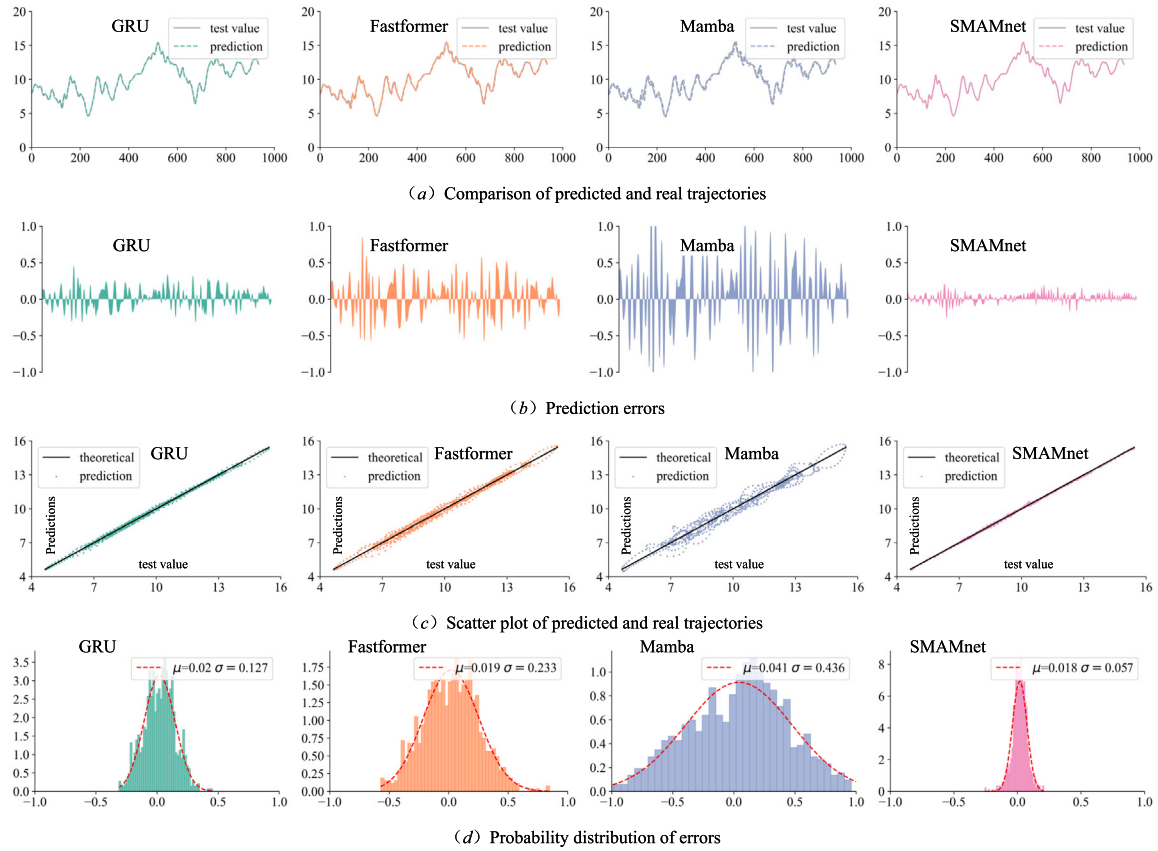


Fig. 6. Graphical comparison of 15 min wind speed prediction results using a single variable in different models.

Table 2

The 15-min wind speed prediction results by adding auxiliary variable between the proposed model and the baseline models.

Category	Model	MAE	RMSE	MAPE	SMAPE	ρ	R^2
Model proposed	SMAMnet	0.033604	0.044915	0.354349	0.354671	0.999779	0.999618
Existing model	SVM	0.785422	0.980321	7.552849	7.781525	0.959856	0.817865
	GP	0.194187	0.477423	1.648043	1.708957	0.987306	0.956802
	DT	0.136873	0.337109	1.207227	1.233719	0.997697	0.978462
	LSTM	0.105797	0.140930	1.145893	1.149802	0.998057	0.996236
	GRU	0.107044	0.133096	1.145653	1.147560	0.998356	0.996643
	TCN	0.210137	0.265455	2.245054	2.257188	0.993908	0.986645
	CNN-LSTM	0.125372	0.161987	1.341898	1.343509	0.997701	0.995027
	CNN-GRU	0.134849	0.165665	1.443147	1.445695	0.997330	0.994799
	BiLSTM	0.173065	0.219454	1.831705	1.830260	0.995711	0.990873
	ConvLSTM	0.135590	0.170779	1.454889	1.460994	0.997408	0.994473
	Transformer	0.366707	0.486179	3.909944	3.926478	0.977702	0.955203
	Fnet	0.260477	0.351785	2.764175	2.758765	0.987999	0.976546
	Fastformer	0.194086	0.243137	2.088560	2.094277	0.994395	0.988796
Ablation test	SCINet	0.252809	0.326598	2.769273	2.763505	0.989176	0.979784
	Mamba	0.423232	0.557355	4.511092	4.529485	0.971747	0.941126
	SMAMnet-1	0.041287	0.052635	0.430525	0.431339	0.999713	0.999475
	SMAMnet-2	0.038257	0.049716	0.397457	0.397990	0.999724	0.999532
	SMAMnet-3	0.089080	0.115667	0.944838	0.947261	0.998851	0.997464
	SMAMnet-4	0.146545	0.202981	1.510757	1.511925	0.997133	0.992191

that it is not easy to deeply explore the correlation between multiple variables. In deep learning methods, GRU has the best performance overall, the inconsistency is that the MAE of LSTM is lower. Nevertheless, compared with the MAE (0.134849) of CNN-GRU, which has the highest prediction accuracy for a single variable, the difference in MAE between LSTM and GRU is minimal.

Compared with each optimal metric of baseline models, the prediction accuracy of the proposed SMAMnet is reduced by 68.24%, 66.25%, 69.07% and 69.09% in MAE, RMSE, MAPE and SMAPE respectively,

and the deviations of ρ and R^2 relative to the perfect prediction are reduced by 86.56% and 88.62% respectively. Moreover, the prediction accuracy of the proposed model is further improved after adding auxiliary variables. For instance, in the case of a single variable, the MAE and RMSE of the proposed model are 0.045832 and 0.060014 respectively. After adding auxiliary variables, the MAE and RMSE are reduced by 26.68% and 25.16% respectively.

Compared with the best results obtained by machine learning models, the prediction accuracy of five groups of deep learning models

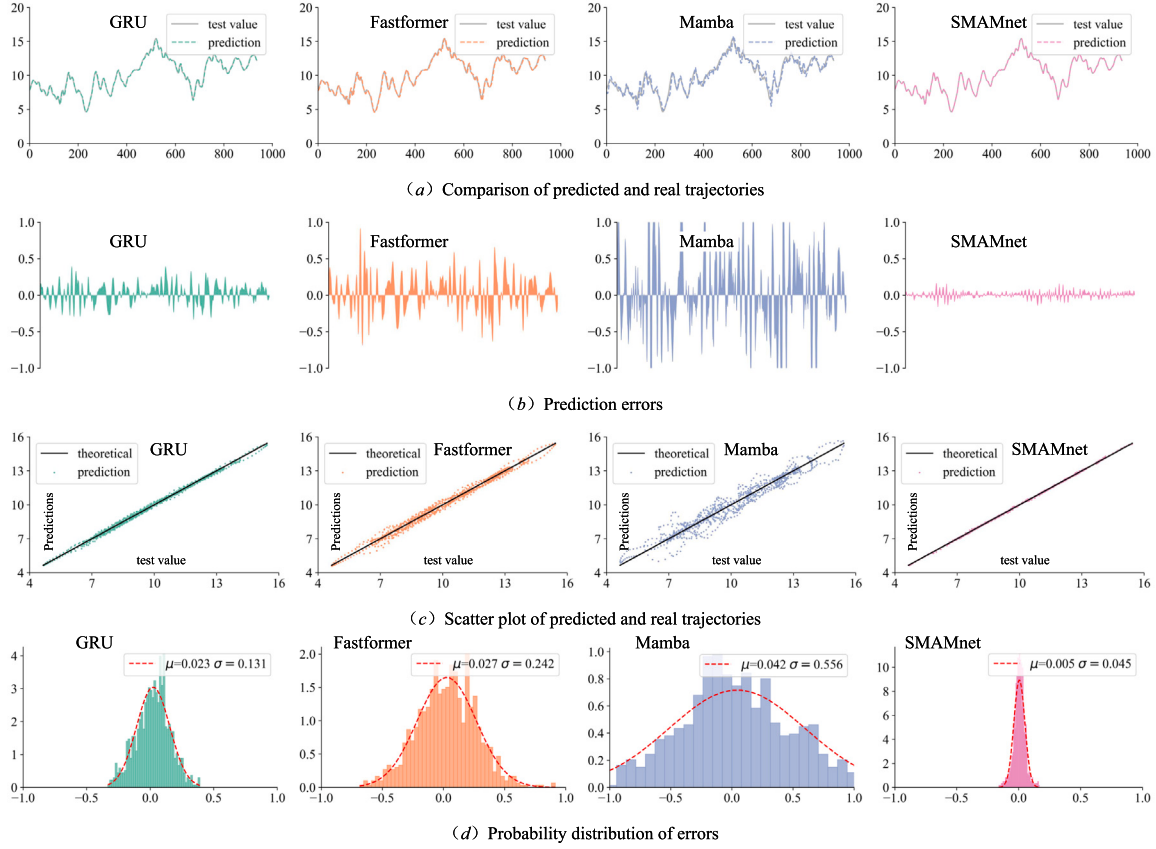


Fig. 7. Graphical comparison of 15 min wind speed prediction results by adding auxiliary variable in different models.

Table 3

The 1-h wind speed prediction results using a single variable between the proposed model and the baseline models.

Category	Model	MAE	RMSE	MAPE	SMAPE	ρ	R^2
Model proposed	SMAMnet	0.190326	0.253144	2.064179	2.068763	0.993562	0.987865
Existing model	SVM	0.759405	1.025385	7.000471	7.231761	0.949208	0.800894
	GP	0.529790	0.803059	5.302893	5.368856	0.948468	0.877875
	DT	0.439441	0.666314	4.306066	4.392537	0.972694	0.915925
	LSTM	0.285795	0.353177	3.049839	3.086714	0.989832	0.976379
	GRU	0.289832	0.359208	3.101312	3.123040	0.989020	0.975566
	TCN	0.332334	0.400444	3.373073	3.429566	0.991919	0.969634
	CNN-LSTM	0.272447	0.343030	2.924772	2.948266	0.989764	0.977717
	CNN-GRU	0.263778	0.323311	2.844467	2.856069	0.990663	0.980205
	BiLSTM	0.397341	0.482546	4.289181	4.317809	0.978893	0.955905
	ConvLSTM	0.303737	0.375626	3.278175	3.302283	0.988367	0.973281
	Transformer	0.537199	0.654601	5.735056	5.759396	0.959886	0.918855
	Fnet	0.531048	0.645670	5.713853	5.746755	0.959899	0.921054
	Fastformer	0.404414	0.491915	4.366010	4.387177	0.977626	0.954176
	SCINet	0.578818	0.699685	6.150427	6.212588	0.956261	0.907292
	Mamba	0.539039	0.656759	5.754175	5.778478	0.959570	0.918319
Ablation test	SMAMnet-1	0.190326	0.253144	2.064179	2.068763	0.993562	0.987865
	SMAMnet-2	0.210412	0.279591	2.284014	2.285861	0.992048	0.985197
	SMAMnet-3	0.262528	0.333769	2.809857	2.833339	0.990949	0.978904
	SMAMnet-4	0.295383	0.383006	3.123366	3.152468	0.988214	0.972221

(LSTM, GRU, CNN-LSTM, CNN-GRU, ConvLSTM) is outstanding. Notably, the top five deep learning models in terms of prediction accuracy are all RNN-based variants. Despite RNN perform well, their inability to compute in parallel constrains their potential for future advancement. Consequently, the researches are increasingly turning to attention mechanisms and state space models to enable parallel computation. Unfortunately, the prediction accuracy of these two types of parallel calculation methods is not ideal. In contrast, the proposed

model does not contain RNN operators and achieves the highest prediction accuracy. This demonstrates the superiority of the proposed model from the architecture perspective.

Ablation experiments. In the ablation experiments, the prediction accuracy of all ablation models decreased, which proves the usefulness of each component in the proposed model. Specifically, taking MAE as an example, SMAMnet-2 and SMAMnet-1 have small changes relative to the proposed model, increasing by 13.85% and 22.86% respectively. However, SMAMnet-3 and SMAMnet-4 have large changes, increasing

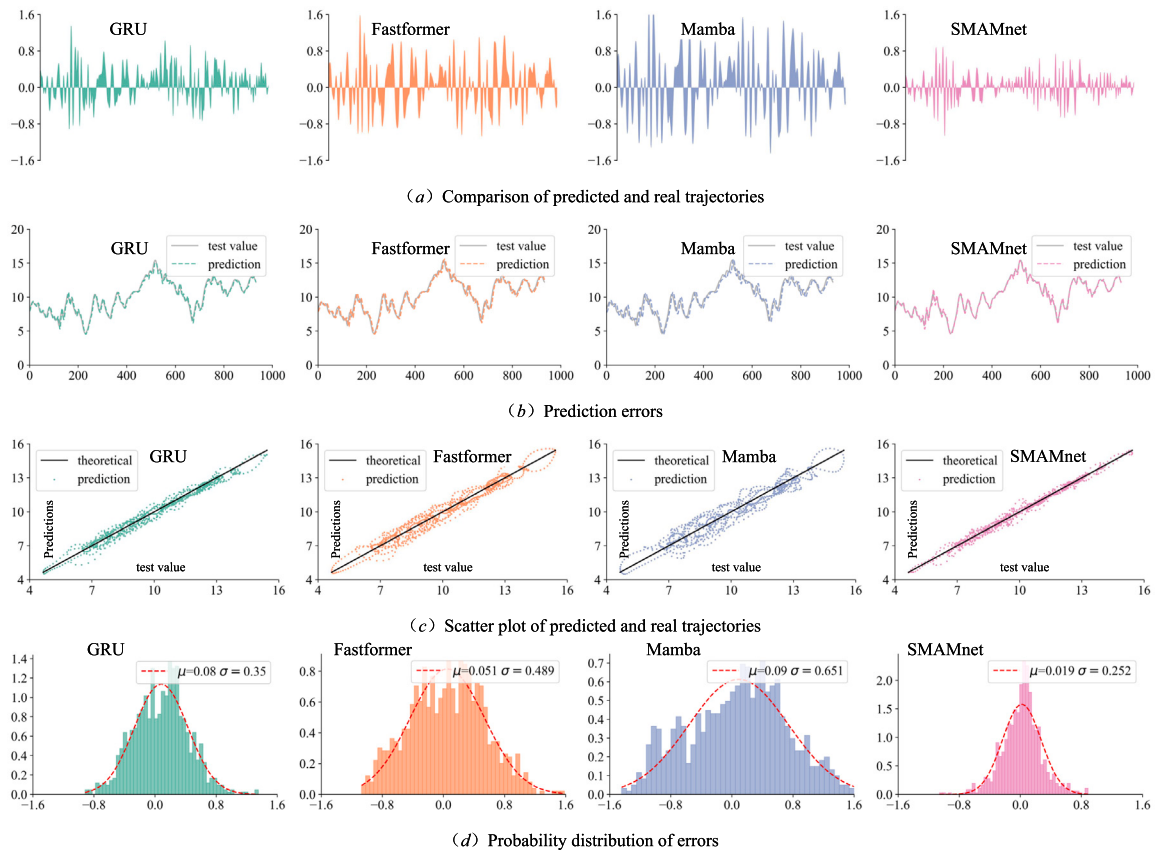


Fig. 8. Graphical comparison of 1-hour wind speed prediction results using a single variable in different models.

Table 4

The 1-h wind speed prediction results by adding auxiliary variable between the proposed model and the baseline models.

Category	Model	MAE	RMSE	MAPE	SMAPE	ρ	R^2
Model proposed	SMAMnet	0.154367	0.206250	1.649216	1.649762	0.996310	0.991944
Existing model	SVM	0.815214	1.003712	8.033816	8.207184	0.945140	0.809222
	GP	0.616778	1.002791	6.020964	6.173434	0.913522	0.809572
	DT	0.465248	0.649173	4.756994	4.840495	0.969723	0.920195
	LSTM	0.302813	0.390367	3.223344	3.264587	0.987458	0.971143
	GRU	0.274793	0.340171	2.918523	2.936640	0.989540	0.978087
	TCN	0.356084	0.460734	3.869168	3.840443	0.980193	0.959801
	CNN-LSTM	0.318929	0.405780	3.421159	3.365676	0.987898	0.968819
	CNN-GRU	0.327253	0.401569	3.501740	3.515513	0.985019	0.969463
	BiLSTM	0.378610	0.471890	4.030753	4.057913	0.980238	0.957831
	ConvLSTM	0.335659	0.431632	3.525391	3.563393	0.985522	0.964719
	Transformer	0.542404	0.690397	5.766334	5.816745	0.956109	0.909737
	Fnet	0.441364	0.552441	4.705709	4.712579	0.971386	0.942206
Ablation test	Fastformer	0.398020	0.494798	4.280821	4.302853	0.977427	0.953638
	SCINet	0.424737	0.530531	4.602512	4.619109	0.973341	0.946699
	Mamba	0.587806	0.748860	6.203721	6.262095	0.949525	0.893803
	SMAMnet-1	0.190256	0.248256	2.046860	2.055217	0.994191	0.988329
	SMAMnet-2	0.193219	0.254240	2.063848	2.075137	0.994146	0.987760
	SMAMnet-3	0.249464	0.321775	2.679598	2.693408	0.991159	0.980393
	SMAMnet-4	0.320363	0.418456	3.359455	3.369798	0.987033	0.966840

by 165.09% and 336.09% respectively. Note that the MAE of SMAMnet-2 is lower than that of SMAMnet-1, which means that multivariate feature reconstruction significantly benefits information mining. Compared with the univariate prediction accuracy, both SMAMnet-1 and SMAMnet-2 exceed the univariate prediction accuracy, which means that the frequency domain selection module is beneficial to multivariate information mining, even for reverse causality.

Graphical prediction results. Fig. 7 shows the results of 15 min wind speed prediction after adding auxiliary variables. Based on the comparison between the predicted trajectory and the actual trajectory,

only the prediction accuracy of Mamba is visibly poor, showing a large number of burrs, and other models can predict the wind speed curve well. The prediction error curve is consistent with the above view. Mamba has the largest prediction error, and the prediction error curve of the proposed model is significantly lower than that of the other models. In the scatter plot, the prediction value of the Mamba model deviates greatly from the theoretical fitting line, and other rules are consistent with the 15 min prediction diagram. In the error probability distribution, SMAMnet has the smallest deviation from the center value, which is 0.005, and the other models are more than three times higher

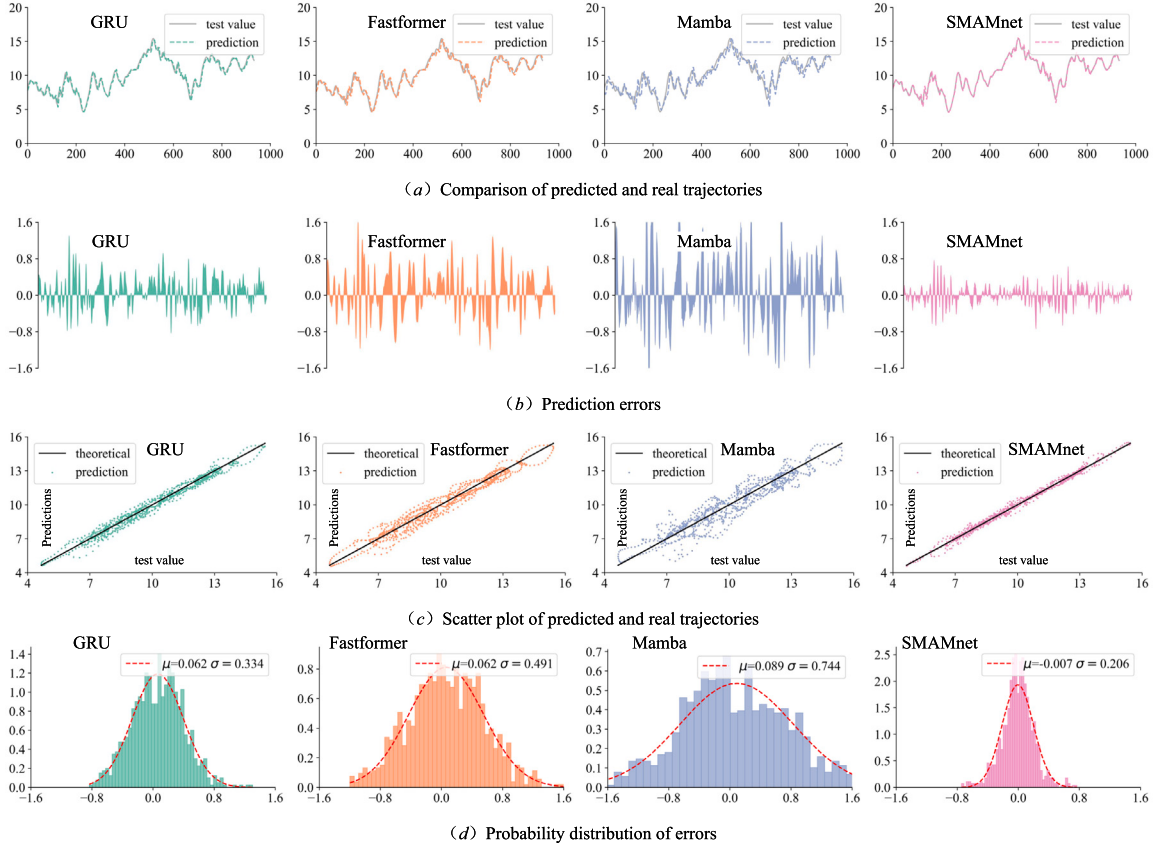


Fig. 9. Graphical comparison of 1-hour wind speed prediction results by adding auxiliary variable in different models.

than it. The distribution variance of SMAMnet is also the smallest, which is 0.045, and the other models are much higher than this value. This shows that the prediction error of the proposed model is small and the most compact. The above graphical analysis shows that SMAMnet exhibits superior prediction performance.

5.3. The 1-hour wind speed prediction using a single variable

Table 3 summarizes the performance differences of 1-hour wind speed prediction between the proposed model and the baseline models using a single variable. Machine learning methods have lagged far behind deep learning methods. For example, the RMSE of DT, GP, and SVM are 0.666314, 0.803059, and 1.025385, respectively, while only SCINet has an RMSE of 0.699685 for deep learning methods, and the RMSE of other models are all lower than 0.666314. Furthermore, the RMSE of CNN-GRU is as low as 0.323311, which is 51.48% lower than the optimal result of machine learning methods (0.666314). This shows that when it comes to information association of long-distance correlation, the availability of machine learning models is no longer comparable to that of deep learning models. Focusing on deep learning models, CNN-GRU has the best overall performance, the inconsistency is that TCN has a higher ρ . TCN and CNN-GRU are the only two models with ρ exceeding 0.99.

Compared with optimal metric of each baseline models, the prediction accuracy of the proposed SMAMnet is reduced by 27.85%, 21.70%, 27.43% and 27.57% in MAE, RMSE, MAPE and SMAPE respectively, and the deviations of ρ and R^2 relative to the perfect prediction are reduced by 20.33% and 38.70% respectively. Compared with the 15 min prediction results, the 1-hour prediction accuracy improvement percentage of the proposed model seems smaller. This is because the prediction error of 1 h is higher than that of 15 min.

Ablation experiments. In the ablation experiment, SMAMnet-1 also has no change, which is consistent with the 15 min prediction results. The prediction accuracy of SMAMnet-2, SMAMnet-3, and SMAMnet-4 is getting worse in order, indicating that each component of the proposed model is meaningful and beneficial to the 1-hour prediction.

Graphical prediction results. Fig. 8 shows the result of 1-hour wind speed prediction using a single variable. Compared with the 15 min wind speed, the situation does not seem optimistic. From the comparison between the predicted value and the actual value, the proposed model can predict the wind speed curve well, but some prediction results of GRU have burrs, and there is a deviation similar to translation between the predicted value and the actual value of Fastformer and Mamba. From the prediction error curve, SMAMnet has the smallest error. Turning to the scatter plot, the fitting curves of the predicted value and the true value of all models are deviated, and the distribution of the predicted value and the true value of the proposed model is relatively close. Through the quantification of the error probability distribution, SMAMnet has the smallest deviation from the center value and the smallest distribution variance, and the other models have significant increases. Therefore, all these evidences show that the proposed model has excellent prediction accuracy.

5.4. The 1-hour wind speed prediction by adding auxiliary variable

Table 4 describes the performance differences of 1-hour wind speed prediction between the proposed model and the baseline models by adding auxiliary variable. Similar to the 15 min prediction results, the prediction accuracy of the machine learning method is worse after adding auxiliary variables. Therefore, the machine learning method is no longer analyzed in detail. GRU has the best prediction accuracy,

and no discordant prediction metrics appear. Only some deep learning models (such as GRU, BiLSTM, and Fnet) have improved their prediction accuracy after adding auxiliary variables. However, please note that the prediction accuracy of Fnet and SCINet has increased relatively significantly after adding auxiliary variables. The MAE of Fnet has dropped from 0.531048 to 0.441364, and the MAE of SCINet has dropped from 0.578818 to 0.424737.

Compared with the optimal value of each evaluation metric of the existing model, the prediction accuracy of the proposed SMAMnet is reduced by 43.82%, 39.37%, 43.49% and 43.82% in MAE, RMSE, MAPE and SMAPE respectively, and the deviations of ρ and R^2 relative to the perfect prediction are reduced by 64.72% and 63.24% respectively. The 1-hour prediction is not as good as the 15 min prediction because of the higher uncertainty.

Combined with the 1-hour prediction results of a single variable, the percentage increase in prediction accuracy is significant after adding auxiliary variables. This is because the baseline model performs ordinary after adding auxiliary variables, while the proposed model achieves higher accuracy. In contrast, the MAE of the proposed model's single variable prediction is 0.190326, and after adding auxiliary variables, the MAE drops to 0.154367, that is, a decrease of 18.89%. However, the single variable prediction accuracy of the proposed model is already far ahead, which means that the proposed model is not only good at mining variable correlations, but the architecture itself is worthy of attention.

Ablation experiments. In the ablation experiment, the prediction accuracy of SMAMnet-1, SMAMnet-2, SMAMnet-3, and SMAMnet-4 deteriorated in turn. Taking MAE as an example, they increased by 23.25%, 25.17%, 61.60%, and 107.53% respectively. This means that the importance of the modules involved in the ablation model increases in turn. It is worth noting that, unlike the prediction results of adding auxiliary variables in 15 min, the prediction effect of SMAMnet-2 is not as good as that of SMAMnet-1, which shows that SSM plays an important role in relatively long-term predictions. In addition, the measured data is often heterogeneous, so the data itself will also have an impact on the modeling process.

Graphical prediction results. Fig. 9 describes the results of adding auxiliary variables for 1-hour wind speed prediction. Through the comparison of four types of graphical methods, including comparison of predicted values with actual values, prediction error curve, scatter plot and error probability distribution, only SMAMnet meets good prediction expectations. This shows that the proposed model is most suitable for 1-hour wind speed prediction tasks.

6. Conclusion

This paper addresses limited variables in wind speed prediction by proposing the use of reverse causality, incorporating auxiliary variables like wind power. A high correlation (0.976) was found between low-frequency wind speed and wind power components. Based on this, we proposed SMAMnet, a multivariate prediction model featuring an improved attention mechanism and a feature reconstruction module. Its novelty is characterized by the development of an adaptive frequency-domain selected attention weight operator to adaptively parse meaningful information in different frequency domain intervals. Experimental results indicate that with the inclusion of auxiliary variables, SMAMnet significantly enhances the accuracy of both 15 min and 1-hour wind speed predictions, outperforming the best results of compared baseline that do not utilize auxiliary variables.

The research is helpful for the wind energy utilization. The limitation is that the universality of adding extra real or simulated auxiliary variables to univariate time series prediction tasks has not been proven. In the future, we intend to explore the adaptability of this approach to other renewable energy fields and widely deployed [55] in wind farms.

CRedit authorship contribution statement

Ke Fu: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Shengli Chen:** Writing – review & editing. **Zhengru Ren:** Writing – review & editing, Supervision, Resources, Project administration.

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.

References

- [1] Yu Huang, Bingzhe Zhang, Huizhen Pang, Biao Wang, Kwang Y Lee, Jiale Xie, Yupeng Jin, Spatio-temporal wind speed prediction based on clayton copula function with deep learning fusion, *Renew. Energy* 192 (2022) 526–536.
- [2] Chengqing Yu, Guangxi Yan, Chengming Yu, Xiwei Mi, Attention mechanism is useful in spatio-temporal wind speed prediction: Evidence from China, *Appl. Soft Comput.* 148 (2023) 110864.
- [3] Qiang Wu, Hongling Zheng, Xiaozhu Guo, Guangqiang Liu, Promoting wind energy for sustainable development by precise wind speed prediction based on graph neural networks, *Renew. Energy* 199 (2022) 977–992.
- [4] Yunbo Niu, Jianzhou Wang, Ziyuan Zhang, Yannan Yu, Jingjiang Liu, A combined interval prediction system based on fuzzy strategy and neural network for wind speed, *Appl. Soft Comput.* 155 (2024) 111408.
- [5] Yagang Zhang, Xue Kong, Jingchao Wang, Siqi Wang, Zheng Zhao, Fei Wang, A comprehensive wind speed prediction system based on intelligent optimized deep neural network and error analysis, *Eng. Appl. Artif. Intell.* 128 (2024) 107479.
- [6] Qingguo Zhou, Chen Wang, Gaofeng Zhang, A combined forecasting system based on modified multi-objective optimization and sub-model selection strategy for short-term wind speed, *Appl. Soft Comput.* 94 (2020) 106463.
- [7] Xiong Xiong, Ruilin Zou, Tao Sheng, Weilin Zeng, Xiaoling Ye, An ultra-short-term wind speed correction method based on the fluctuation characteristics of wind speed, *Energy* 283 (2023) 129012.
- [8] Zhongmei Tian, Wei Shi, Xin Li, Yonghui Park, Zhiyu Jiang, Ji Wu, Numerical simulations of floating offshore wind turbines with shared mooring under current-only conditions, *Renew. Energy* 238 (2025) 121918.
- [9] Fei Wang, Shuang Tong, Yiqian Sun, Yongsheng Xie, Zhao Zhen, Guoqing Li, Chunmei Cao, Neven Duić, Dagui Liu, Wind process pattern forecasting based ultra-short-term wind speed hybrid prediction, *Energy* 255 (2022) 124509.
- [10] Yizhuo Cai, Yanting Li, Short-term wind speed forecast based on dynamic spatio-temporal directed graph attention network, *Appl. Energy* 375 (2024) 124124.
- [11] Zhuochun Wu, Liye Xiao, A secondary decomposition based hybrid structure with meteorological analysis for deterministic and probabilistic wind speed forecasting, *Appl. Soft Comput.* 85 (2019) 105799.
- [12] Qiuling Yang, Changhong Deng, Xiqiang Chang, Ultra-short-term/short-term wind speed prediction based on improved singular spectrum analysis, *Renew. Energy* 184 (2022) 36–44.
- [13] Mei Yu, Boan Tao, Xuwei Li, Zhiqiang Liu, Wei Xiong, Local and long-range convolutional LSTM network: A novel multi-step wind speed prediction approach for modeling local and long-range spatial correlations based on ConvLSTM, *Eng. Appl. Artif. Intell.* 130 (2024) 107613.
- [14] Ziwen Gu, Yatao Shen, Zijian Wang, Yaqun Jiang, Chun Huang, Peng Li, A dynamic-inner manifold broad learning system with coupled time-frequency domain for wind speed prediction, *IEEE Trans. Ind. Inform.* (2024).
- [15] Wei Sun, Bin Tan, Qiqi Wang, Multi-step wind speed forecasting based on secondary decomposition algorithm and optimized back propagation neural network, *Appl. Soft Comput.* 113 (2021) 107894.
- [16] Ji Jin, Jinyu Tian, Min Yu, Yong Wu, Yuanan Tang, A novel ultra-short-term wind speed prediction method based on dynamic adaptive continued fraction, *Chaos Solitons Fractals* 180 (2024) 114532.
- [17] Hang Chen, Shanbi Wei, Wei Yang, Shanchao Liu, Input wind speed forecasting for wind turbines based on spatio-temporal correlation, *Renew. Energy* 216 (2023) 119075.
- [18] Mao Yang, Da Wang, Wei Zhang, A novel ultra short-term wind power prediction model based on double model coordination switching mechanism, *Energy* 289 (2024) 130075.

- [19] Wen Ding, Fanyong Meng, Point and interval forecasting for wind speed based on linear component extraction, *Appl. Soft Comput.* 93 (2020) 106350.
- [20] Victoria Hoolohan, Alison S. Tomlin, Timothy Cockerill, Improved near surface wind speed predictions using Gaussian process regression combined with numerical weather predictions and observed meteorological data, *Renew. Energy* 126 (2018) 1043–1054.
- [21] Jing Zhao, Zhen-Hai Guo, Zhong-Yue Su, Zhi-Yuan Zhao, Xia Xiao, Feng Liu, An improved multi-step forecasting model based on WRF ensembles and creative fuzzy systems for wind speed, *Appl. Energy* 162 (2016) 808–826.
- [22] Jiyang Wang, Zhiwu Li, Wind speed interval prediction based on multidimensional time series of convolutional neural networks, *Eng. Appl. Artif. Intell.* 121 (2023) 105987.
- [23] Sultan Al-Yahyai, Yassine Charabi, Abdullah Al-Badi, Adel Gastli, Nested ensemble NWP approach for wind energy assessment, *Renew. Energy* 37 (1) (2012) 150–160.
- [24] Han Wang, Shuang Han, Yongqian Liu, Jie Yan, Li Li, Sequence transfer correction algorithm for numerical weather prediction wind speed and its application in a wind power forecasting system, *Appl. Energy* 237 (2019) 1–10.
- [25] Chu Zhang, Chunlei Ji, Lei Hua, Huixin Ma, Muhammad Shahzad Nazir, Tian Peng, Evolutionary quantile regression gated recurrent unit network based on variational mode decomposition, improved whale optimization algorithm for probabilistic short-term wind speed prediction, *Renew. Energy* 197 (2022) 668–682.
- [26] Yagang Zhang, Yuan Zhao, Chunhui Kong, Bing Chen, A new prediction method based on VMD-PRBF-ARMA-E model considering wind speed characteristic, *Energy Convers. Manage.* 203 (2020) 112254.
- [27] S.N. Singh, Abhejeet Mohapatra, et al., Repeated wavelet transform based ARIMA model for very short-term wind speed forecasting, *Renew. Energy* 136 (2019) 758–768.
- [28] Jingjiang Liu, Jianzhou Wang, Yunbo Niu, Boqian Ji, Lei Gu, A point-interval wind speed forecasting system based on fuzzy theory and neural networks architecture searching strategy, *Eng. Appl. Artif. Intell.* 132 (2024) 107906.
- [29] Ling-Ling Li, Xue Zhao, Ming-Lang Tseng, Raymond R. Tan, Short-term wind power forecasting based on support vector machine with improved dragonfly algorithm, *J. Clean. Prod.* 242 (2020) 118447.
- [30] Yagang Zhang, Ruixuan Li, Short term wind energy prediction model based on data decomposition and optimized LSSVM, *Sustain. Energy Technol. Assess.* 52 (2022) 102025.
- [31] Jianming Hu, Jianzhou Wang, Short-term wind speed prediction using empirical wavelet transform and Gaussian process regression, *Energy* 93 (2015) 1456–1466.
- [32] Zhenkun Liu, Ping Jiang, Lifang Zhang, Xinsong Niu, A combined forecasting model for time series: Application to short-term wind speed forecasting, *Appl. Energy* 259 (2020) 114137.
- [33] Zishu Cheng, Jiyang Wang, A new combined model based on multi-objective salp swarm optimization for wind speed forecasting, *Appl. Soft Comput.* 92 (2020) 106294.
- [34] Jujie Wang, Yaning Li, An innovative hybrid approach for multi-step ahead wind speed prediction, *Appl. Soft Comput.* 78 (2019) 296–309.
- [35] Rana Muhammad Adnan, Reham R Mostafa, Mo Wang, Kulwinder Singh Parmar, Ozgur Kisi, Mohammad Zounemat-Kermani, Improved random vector functional link network with an enhanced remora optimization algorithm for predicting monthly streamflow, *J. Hydrol.* 650 (2025) 132496.
- [36] Fouzi Harrou, Ahmed Saidi, Ying Sun, Wind power prediction using bootstrap aggregating trees approach to enabling sustainable wind power integration in a smart grid, *Energy Convers. Manage.* 201 (2019) 112077.
- [37] Xuefang Xu, Shiting Hu, Peiming Shi, Huaishuang Shao, Ruixiong Li, Zhi Li, Natural phase space reconstruction-based broad learning system for short-term wind speed prediction: Case studies of an offshore wind farm, *Energy* 262 (2023) 125342.
- [38] Rana Muhammad Adnan, Ahmed A Ewees, Mo Wang, Ozgur Kisi, Salim Heddami, Kulwinder Singh Parmar, Mohammad Zounemat-Kermani, Enhancing BOD5 forecasting accuracy with the ANN-enhanced Runge Kutta model, *J. Environ. Chem. Eng.* 13 (2) (2025) 115430.
- [39] Xiaohui Yuan, Chen Chen, Min Jiang, Yanbin Yuan, Prediction interval of wind power using parameter optimized beta distribution based LSTM model, *Appl. Soft Comput.* 82 (2019) 105550.
- [40] Chuanjin Yu, Yongle Li, Liyang Zhao, Qian Chen, Yuxing Xun, A novel time-frequency recurrent network and its advanced version for short-term wind speed predictions, *Energy* 262 (2023) 125556.
- [41] Binrong Wu, Lin Wang, Two-stage decomposition and temporal fusion transformers for interpretable wind speed forecasting, *Energy* 288 (2024) 129728.
- [42] Chengqing Yu, Guangxi Yan, Chengming Yu, Xinwei Liu, Xiwei Mi, MRIformer: A multi-resolution interactive transformer for wind speed multi-step prediction, *Inform. Sci.* 661 (2024) 120150.
- [43] Chuhan Wu, Fangzhao Wu, Tao Qi, Yongfeng Huang, Xing Xie, Fastformer: Additive attention can be all you need, 2021, *arXiv preprint arXiv:2108.09084*.
- [44] Albert Gu, Tri Dao, Mamba: Linear-time sequence modeling with selective state spaces, 2023, *arXiv preprint arXiv:2312.00752*.
- [45] Yu Huang, Zongshi Zhang, Xuxin Li, Jiale Xie, Kwang Y Lee, Layered-vine copula-based wind speed prediction using spatial correlation and meteorological influence, *IEEE Trans. Instrum. Meas.* (2023).
- [46] Jianzhou Wang, Mengzheng Lv, Zhiwu Li, Bo Zeng, Multivariate selection-combination short-term wind speed forecasting system based on convolution-recurrent network and multi-objective chameleon swarm algorithm, *Expert Syst. Appl.* 214 (2023) 119129.
- [47] Haipeng Zhang, Jianzhou Wang, Yuansheng Qian, Qiwei Li, Point and interval wind speed forecasting of multivariate time series based on dual-layer LSTM, *Energy* 294 (2024) 130875.
- [48] Zhongda Tian, Modes decomposition forecasting approach for ultra-short-term wind speed, *Appl. Soft Comput.* 105 (2021) 107303.
- [49] Binrong Wu, Lin Wang, Yu-Rong Zeng, Interpretable wind speed prediction with multivariate time series and temporal fusion transformers, *Energy* 252 (2022) 123990.
- [50] Daniel Y Fu, Tri Dao, Khaled K Saab, Armin W Thomas, Atri Rudra, Christopher Ré, Hungry hungry hippos: Towards language modeling with state space models, 2022, *arXiv preprint arXiv:2212.14052*.
- [51] Mu-Yen Chen, Bo-Tsuen Chen, Online fuzzy time series analysis based on entropy discretization and a fast Fourier transform, *Appl. Soft Comput.* 14 (2014) 156–166.
- [52] Subhashini Narayan, E. Sathiyamoorthy, A novel recommender system based on FFT with machine learning for predicting and identifying heart diseases, *Neural Comput. Appl.* 31 (2019) 93–102.
- [53] Rana Muhammad Adnan, Zhongmin Liang, Slavisa Trajkovic, Mohammad Zounemat-Kermani, Binqian Li, Ozgur Kisi, Daily streamflow prediction using optimally pruned extreme learning machine, *J. Hydrol.* 577 (2019) 123981.
- [54] Rana Muhammad Adnan, Amin Mirboluki, Mojtaba Mehraein, Anurag Malik, Salim Heddami, Ozgur Kisi, Improved prediction of monthly streamflow in a mountainous region by metaheuristic-enhanced deep learning and machine learning models using hydroclimatic data, *Theor. Appl. Climatol.* 155 (1) (2024) 205–228.
- [55] Yang Chen, Yuemin Ding, Zhen-Zhong Hu, Zhengru Ren, Geometrized task scheduling and adaptive resource allocation for large-scale edge computing in smart cities, *IEEE Internet Things J.* (2025).