Energy 341 (2025) 139396

Contents lists available at ScienceDirect

Energy

journal homepage: www.elsevier.com/locate/energy

Spectral dynamic aggregation transformer and fitted swing-door algorithm for wind power monitoring

Hao Xu [a,1], Zhenhao Zhu [a,1], Hongbing Liu [a,*], Enrico Zio [b], Xiaolong Qiu [a], Yuchen Lu [a], Xianqiang Qu [a]

[a] Yantai Research Institute, Harbin Engineering University, Yantai, China
[b] Ecole de Mines, Politecnico di Milao, Milao, Italy

ARTICLE INFO

Keywords:
Wind power
Power prediction
Spectral dynamic aggregation
Dual early warning with fitted swing-door algorithm
Transformer

ABSTRACT

Current research on wind energy monitoring predominantly focuses on power prediction while often overlooking the advanced warning of sudden operational anomalies. To this end, we propose a wind power monitoring model based on a spectral dynamic aggregation transformer integrated with a fitted swing gate algorithm. First, the integration of spectral and dynamic aggregation blocks within the Transformer framework yields an accurate wind power prediction model that effectively alleviates the impact of data fluctuations. On this basis, the MI method is utilized to quantify the nonlinear relationships between multi-source meteorological variables and wind power output. By integrating STL for residual analysis to extract salient features, the proposed approach not only enhances the input quality of the prediction model but also provides a physically interpretable foundation for the early warning module. This facilitates seamless integration of prediction and early warning at the feature level. Furthermore, the prediction results and key features jointly drive a two-tier early warning framework: a wind power early warning system is constructed based on the random forest algorithm and the swing door algorithm, followed by joint calibration with the predictive model. By leveraging multi-source data, the model is capable of detecting anomalous power changes and ramp events, thereby ensuring efficient anomaly identification and advanced warning. Through the case study, it is demonstrated that the proposed model can achieve wind power prediction and advanced warning functions, thereby providing robust support for flexible grid scheduling and efficient wind power integration.

1. Introduction

Excessive exploitation and utilization of traditional fossil fuels have exacerbated environmental degradation and intensified global warming trends. As a clean and renewable alternative, wind energy has gained widespread attention. As witnessed, the expansion of the wind power market has been accompanied by a significant increase in installed capacity [1].Offshore wind power installations in particular have exceeded 35 GW, with an expected annual growth rate surpassing 30 % by 2025 [2]. However, Wind power generation exhibits notable randomness and intermittency, posing significant challenges to the control and prediction of active power [3]. Furthermore, the integration of such intermittent energy sources into the power grid increases the complexity of system reliability analysis [4]. The unpredictability of wind generates, as in fluctuating power output, leading to a significant increase in the operational risk of large-scale grid integration and raising maintenance costs for wind turbines [5].In recent years, indeed, the connection of large offshore wind farms to inland power grids has had a considerable impact on grid stability [6].Thus, enhancing the ability to provide high-quality wind energy and ensuring timely monitoring of wind power performance are strategic for the development of the wind power industry.

To achieve accurate monitoring of wind power performance, various approaches have been developed.These approaches primarily focused on wind power forecasting can be broadly categorized into three traditional methods: physical models, statistical fitting and artificial intelligence models [7].In physical models, numerical weather prediction (NWP) data [8] and relevant geographical information of wind farms are used to formulate wind speed prediction equations. One application is wind power forecasting based on physical mathematical functions,

* Corresponding author.
E-mail address: hb_liu@hrbeu.edu.cn (H. Liu).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.energy.2025.139396
Received 14 January 2025; Received in revised form 30 October 2025; Accepted 20 November 2025
Available online 26 November 2025
0360-5442/© 2025 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

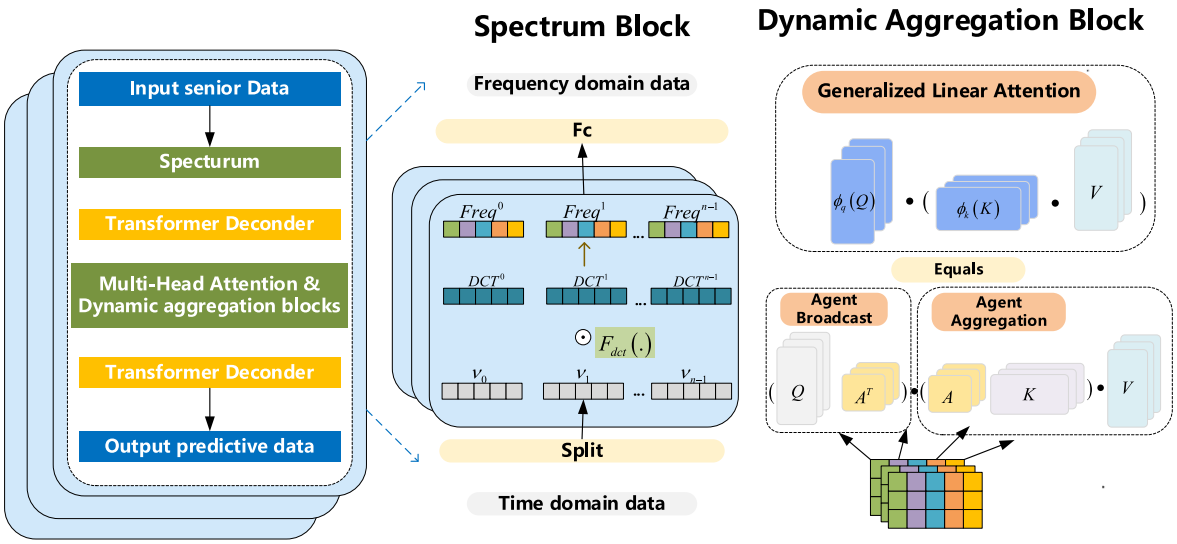


Fig. 1. Schematic of the Transformer-based model.

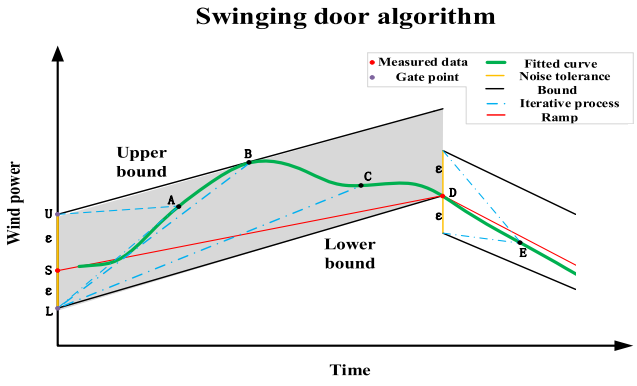


Fig. 2. Working principle of the Swing-Door Algorithm.

particularly for monitoring newly constructed wind farms for which field data are not yet available. However, the approach suffers from low prediction accuracy and weak robustness [9]. Statistical models, on the other hand, rely on historical data and are primarily used for time-series-based forecasting of wind power. Classical short-term monitoring methods include autoregressive (AR) models [10], exponential smoothing methods (ESM) [11], autoregressive moving average (ARMA) [12], and Kalman filters (KF) [13]. For instance, Liu et al. [14] utilized a seasonal ARIMA model for short-term offshore wind speed prediction, Naik et al. [15] applied robust multi-kernel ridge regression for forecasting wind speed and wind power. However, as the autocorrelation of wind power sequences significantly decreases with longer prediction intervals, these simple statistical approaches are typically effective only for short-term wind power monitoring tasks.

With the rapid advancement of artificial intelligence (AI), AI models are increasingly demonstrating their potential in wind power monitoring [16].Examples of AI methods applied to wind power forecasting include Support Vector Machines (SVM) [17], Support Vector Regression (SVR) [18], Decision Trees [19], Artificial Neural Networks [20] and Wavelet Neural Networks (WNN) [21].These methods allow properly handling the nonlinear characteristics of wind power, but may suffer from overfitting the data and may result inefficiency when dealing with large numbers of features and large volumes of observations.

Relatively recently, Deep learning (DL) has brought a breakthrough in artificial intelligence for monitoring applications, offering the capability of extracting complex nonlinear relationships from large datasets [22]. Models like GRU, DNN [23] and LSTM [24] introduce unique gating mechanisms that allow early-sequence information to propagate to later steps [25], overcoming issues of gradient explosion or vanishing, inherent in traditional Recurrent Neural Networks (RNNs). With the development of backpropagation, Neshat et al. [26] applied an evolutionary decomposition approach offshore wind speed data from offshore wind farms in the Baltic Sea to develop a Bi-Directional Long Short-Term Memory (Bi-LSTM) model for wind speed forecasting.

Zhao et al. [27] introduced a bi-directional mechanism model for short-term wind speed prediction, based on Extreme Learning Machine (ELM) to accelerate network training [28]. For long-term monitoring needs, Transformer-based models have become dominant in sequence-based deep learning [29]. Transformers do not rely on recurrence or convolution, making them well-suited for learning long-term dependencies. Li et al. [30] explored a Transformer-based architecture that proposed using convolutional layers for local processing and sparse attention mechanisms to expand the receptive field during processing. Subsequent researchers focused on reducing the complexity of self-attention for long-term time-series tasks. Informer [31] extended self-attention using a KL-divergence criterion to select dominant queries, while Reformer [32] introduced the Locality-Sensitive Hashing (LSH) mechanism to estimate attention via similar query assignments. These models not only reduce model complexity but also enhance its efficiency and accuracy for time-series monitoring.

Autoformer [33] incorporated decomposition blocks into a canonical structure and designed self-correlations to capture series connections. Pyraformer [34] introduced the Pyramid Attention Module (PAM) to capture time dependencies at various hierarchical levels. These advancements continue to push the boundaries of performance in long-term wind power monitoring.

However, for complex problems involving multivariate and spatio-temporal features, the accuracy of a single deep learning (DL) model may be limited, whereas hybrid models demonstrate superior performance [35].Since the original feature space of wind power comprises a superposition of different frequency components, effectively identifying and extracting these components is crucial to reducing interference from multicollinearity and enhancing monitoring adaptability [36]. Attention mechanisms address sequence correlation modeling challenges by guiding the model to focus on the most critical features [37].Compared to traditional models, hybrid models incorporating attention mechanisms or improved attention methods can achieve better scalability [38] and robustness [39].To enhance the accuracy of short-term wind power

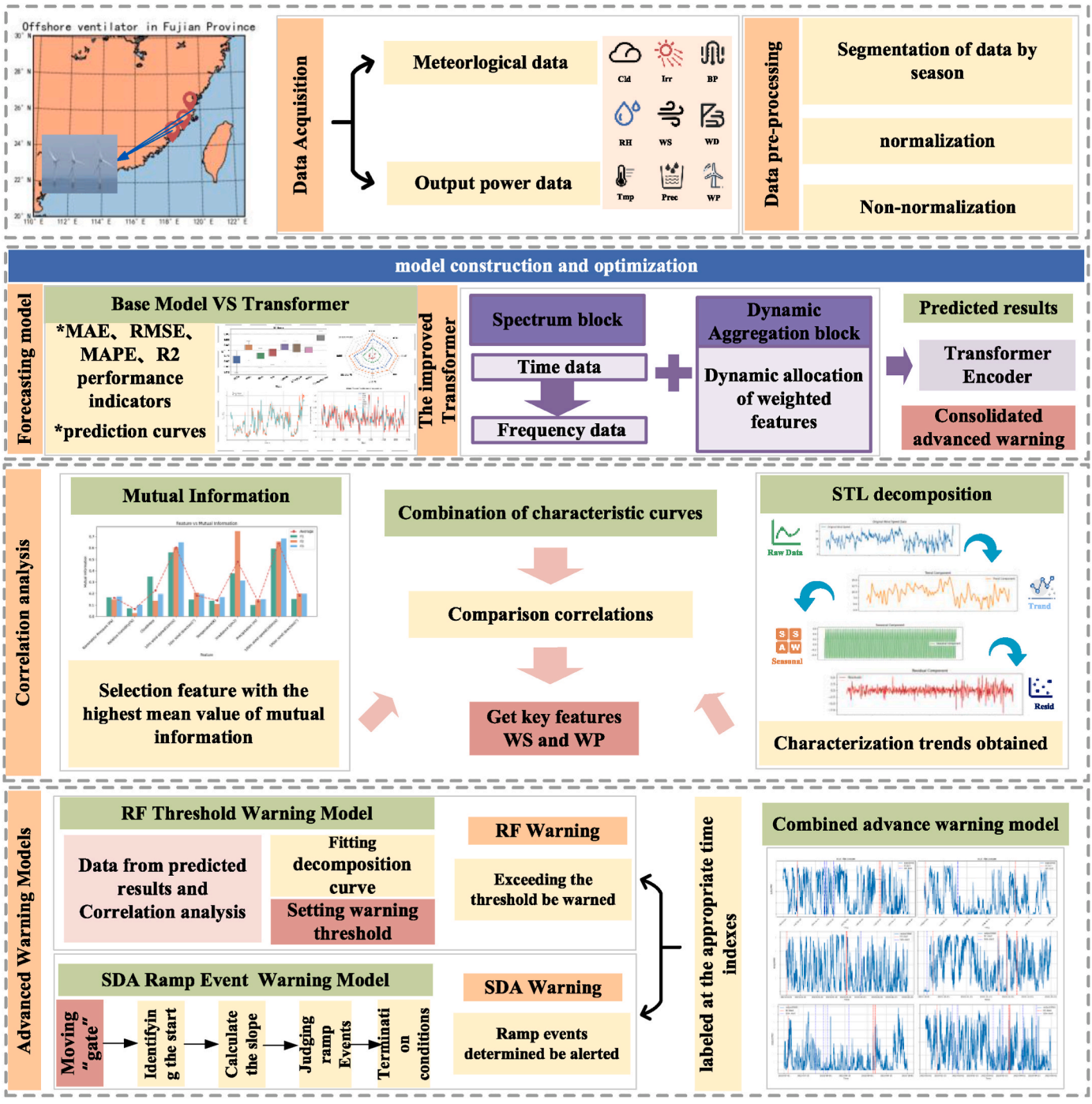


Fig. 3. Flowchart of the proposed method.

forecasting (WPF), integrating attention mechanisms has proven both reliable and effective. For instance, Zhen et al. [40] combined Singular Spectrum Analysis (SSA) with a Temporal Convolutional Attention Network (TCA) to differentiate and identify the trend, periodic and noise components of raw wind power sequences, enabling more precise wind power monitoring.

The aforementioned studies primarily focus on monitoring wind power performance through forecasting methods. However, the sudden changes in wind power pose significant threats to the stability and safety of power systems, an issue that cannot be overlooked. There remains a considerable gap in existing research regarding early warning mechanisms for wind power fluctuations. Some studies have addressed extreme events, such as Chen et al. [41] proposed a novel hybrid model integrating TCN, DANet, and a sparse Transformer architecture for offshore wind power forecasting, specifically designed to capture and integrate spatiotemporal coupling characteristics across multiple wind farms. Dong et al. [42]developed the Wind-Mambaformer model to enhance ultra-short-term wind power prediction at the turbine level by leveraging long-range dependency modeling and computational efficiency. Wang et al. [43]introduced a dual-path prediction framework that combines the Transformer architecture with a BiGRU equipped with an attention mechanism, enabling separate modeling of high-frequency and low-frequency components in wind power time series. Zhang et al. [44]constructed a short-term multi-step forecasting model based on the Pt-Transformer neural network, which incorporates a sparse attention mechanism to improve predictive accuracy while reducing

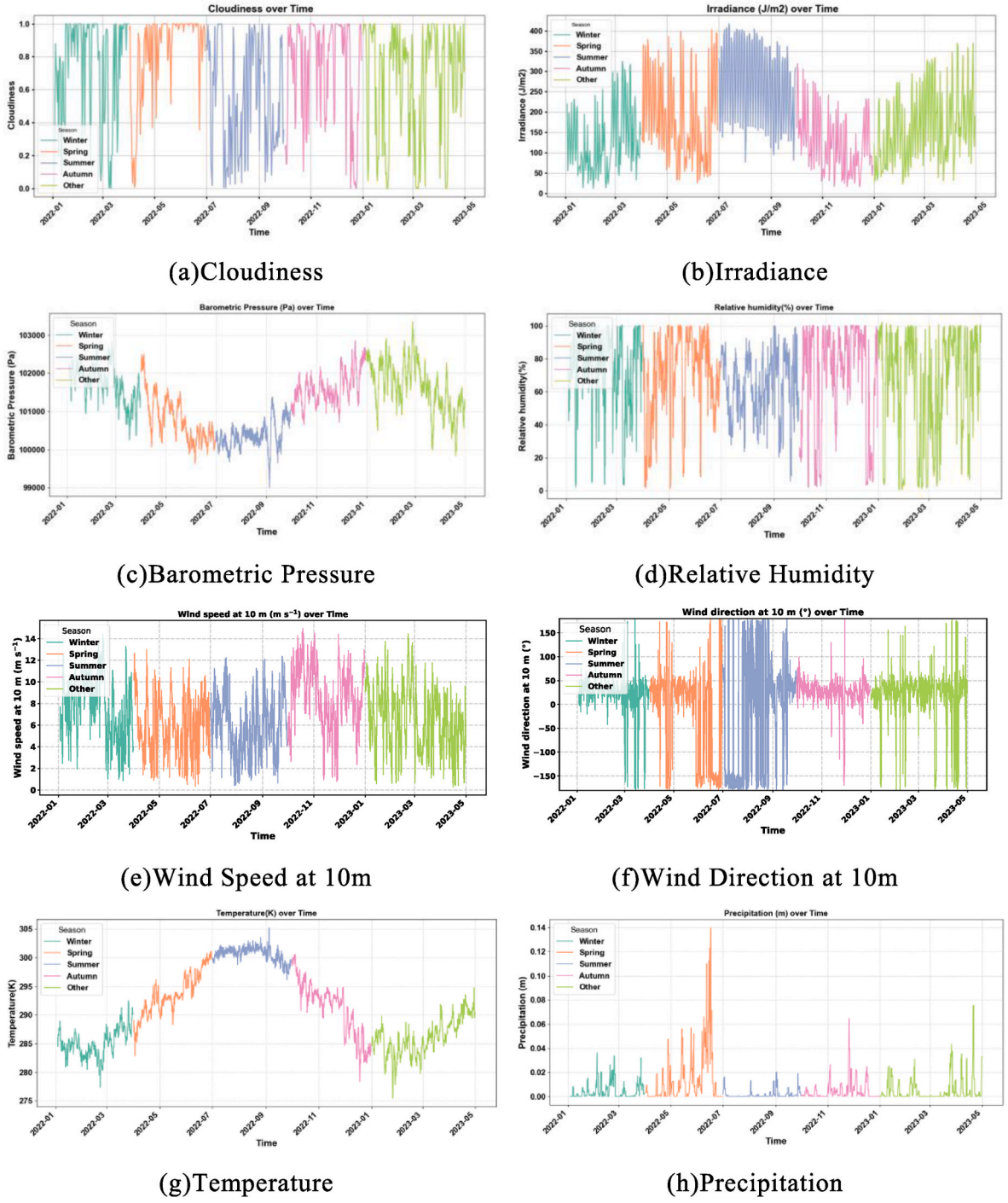
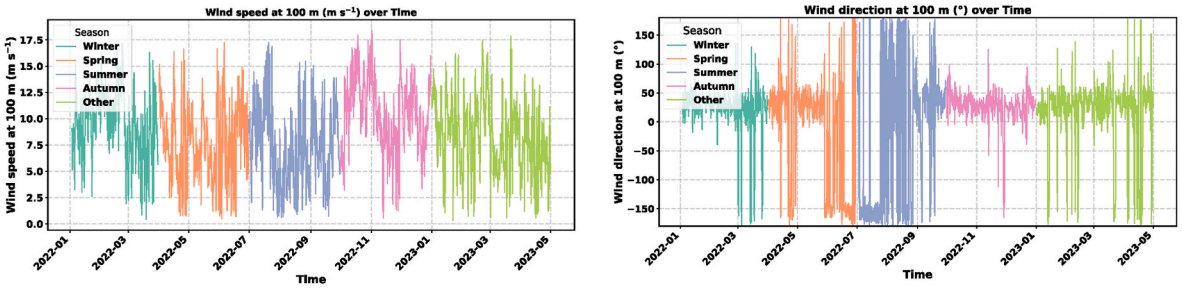


Fig. 4. Trends in turbine 1 start data.

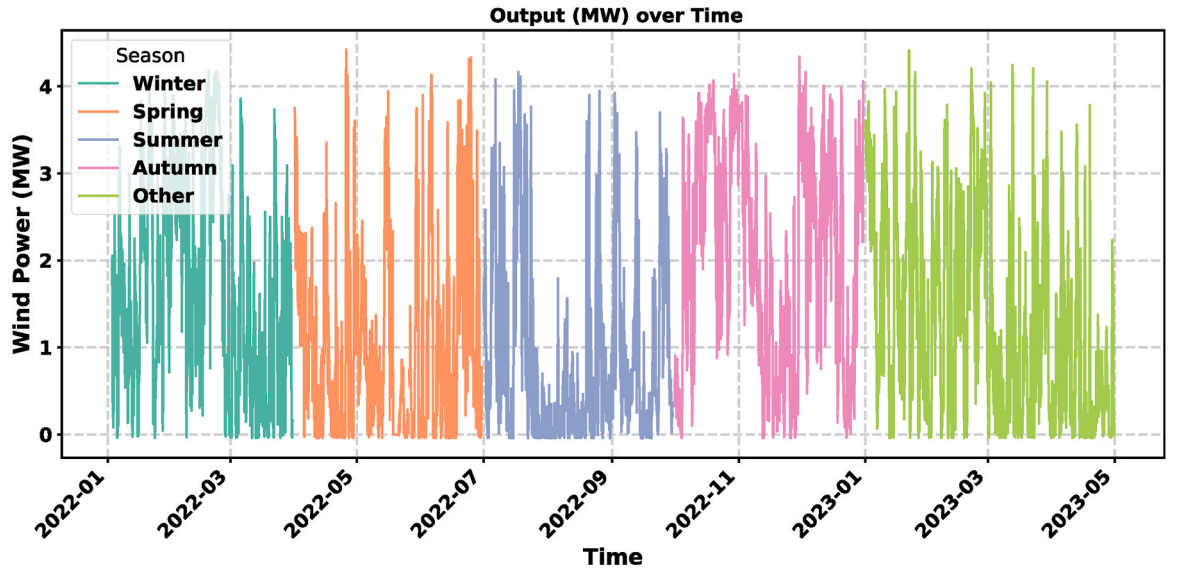
computational complexity. Li et al. [45] proposed a hybrid prediction approach for ultra-short-term wind power generation to address the challenges posed by extreme weather events, with the aim of supporting early warning systems under atypical meteorological conditions. Yu et al. [46]developed a segmented forecasting strategy tailored to extreme weather scenarios, designed to improve the responsiveness and adaptability of wind power predictions to abrupt meteorological fluctuations.

The present study focuses on developing early warning strategies for wind power fluctuations to further enhance the stability and safety of power systems in response to wind power variability. To improve the performance of wind power monitoring, this study aims at developing high-performance wind power monitoring models and corresponding early warning mechanisms. Six foundational models for wind power forecasting have been systematically compared, leading to the selection of a Transformer-based architecture to improve the accuracy of wind power forecasting. To further improve model performance, an integrated method incorporating spectral blocks and dynamic aggregation blocks is introduced. The integration of the two modules significantly improves the model's ability to predict wind power fluctuations. Additionally, a multi-source data early warning model is developed, integrating regression fitting techniques and the adaptive dynamic



(i)Wind Speed at 100m

(j)Wind Direction at 100m



(k)Output

Fig. 4. (*continued*).

algorithm (SDA) to effectively issue warnings for extreme weather conditions or system failures. Compared to other wind power forecasting models, the proposed hybrid dual-warning model—featuring an attention-enhanced Transformer with integrated spectral and dynamic aggregation blocks—exhibits superior generalization and robustness. This advancement promises strong support for reliable wind power monitoring.

2. Methodology

2.1. Spectral dynamic aggregation transformer

This paper proposes a time-series forecasting framework for wind power. The framework is based on an enhanced Transformer architecture that integrates spectral blocks and dynamic aggregation blocks, which are designed to capture multi-level features of power data from both the frequency and temporal domains, thereby improving the accuracy of wind power predictions. The novel multimodal spatiotemporal model based on Transformer is illustrated in Fig. 1.

The self-attention mechanism serves as the core of the Transformer model [47] with its main steps outlined as follows.

Each position in the input sequence is mapped to three vectors: *query*(Q), *key*(K) and *value*(V). The generation of queries, keys and values is achieved through linear transformations:

$$Q = XW^Q \; K = XW^K \; V = XW^V \tag{1}$$

where X is the input sequence, and W^Q, W^K, W^V are learnable weight matrices.

For each pair of *query*(Q) and *key*(K), similarity is computed using a dot product to measure the correlation between different positions in the input sequence. This similarity is,then, scaled to prevent excessively large dot product values, as given by the following formula:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{2}$$

where d_k represents the dimensionality of the key vectors, ensuring numerical stability in the similarity values.

In the Transformer model, self-attention is extended to multi-head attention by splitting the input sequence into h subspaces, where each subspace corresponds to a head, each head independently performs self-attention computations. The outputs from all heads are,then, concatenated and passed through a linear transformation to generate the final multi-head attention output:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \ldots, \text{head}_h)W^O \tag{3}$$

where W^O represents the linear transformation matrix. Although the multi-head attention mechanism involves h heads, the number of parameters and computational complexity do not exceed that of single-head self-attention.

Table 1
Performance metrics of the prediction models for F1.

Dataset	Model	MAE	RMSE	MAPE	R^2
F1_1	ELM	5.260	6.669	2.505	0.701
	SVR	4.139	5.460	1.852	0.799
	RBF	5.043	6.515	2.221	0.714
	BP	4.573	5.957	1.821	0.763
	LSTM	4.130	5.336	2.056	0.808
	Bi-LSTM	4.178	5.464	1.985	0.800
	GRU	4.653	5.919	2.499	0.764
	Transformer	2.581	3.249	1.763	**0.891**
F1_2	ELM	5.454	7.204	6.894	0.542
	SVR	4.354	6.239	4.226	0.656
	RBF	5.085	6.815	6.160	0.590
	BP	4.900	6.561	6.178	0.620
	LSTM	4.881	6.361	0.714	0.643
	Bi-LSTM	4.258	5.956	4.425	0.687
	GRU	4.621	6.365	4.905	0.642
	Transformer	3.358	4.730	1.614	**0.844**
F1_3	ELM	4.840	6.402	6.679	0.632
	SVR	3.263	4.709	3.337	0.801
	RBF	4.019	5.511	5.036	0.727
	BP	3.865	5.269	5.092	0.751
	LSTM	3.277	4.733	2.598	0.799
	Bi-LSTM	3.201	4.561	2.773	0.813
	GRU	3.464	4.971	3.227	0.778
	Transformer	4.555	6.706	6.734	0.629
F1_4	ELM	4.922	6.318	1.462	0.743
	SVR	3.735	5.067	0.901	0.835
	RBF	4.640	5.992	1.381	0.769
	BP	4.378	5.668	1.390	0.793
	LSTM	3.841	5.091	0.975	0.833
	Bi-LSTM	3.732	5.041	1.069	0.836
	GRU	3.997	5.287	1.265	0.820
	Transformer	3.027	3.496	2.449	**0.929**
F1_5	ELM	5.409	6.986	5.526	0.644
	SVR	4.180	5.790	3.345	0.755
	RBF	5.125	6.706	5.224	0.672
	BP	4.852	6.469	4.801	0.695
	LSTM	4.329	5.783	4.338	0.756
	Bi-LSTM	4.416	6.043	3.560	0.733
	GRU	4.608	6.109	4.657	0.728
	Transformer	1.493	1.943	8.139	**0.870**

2.1.1. Spectrum blocks

To address the variability inherent in the data, the Transformer model incorporates spectral blocks into its structure. These spectral blocks convert input data from the time domain to the frequency domain, enabling the extraction of critical frequency-domain features such as periodicity and trends. In our model, spectral blocks are positioned immediately after the original time-series data input layer, allowing each subsequent encoder layer to extract and optimize these frequency-domain features. The main working principle is as follows.

The basis function for one-dimensional Discrete Cosine Transform (DCT) is used.

$$B_l^i = \cos\left(-\frac{\pi l}{L_s}\left(i+\frac{1}{2}\right)\right) \tag{4}$$

where B_l^i represents the value of the l-th basis function at the i-th sample point, l being the index of the basis function, indicating the specific basis function being used, and i denoting the index of the sample point in the input signal. L_s represents the length of the input signal, i.e., the total number of samples in the signal.

In the Discrete Cosine Transform, the range of the basis function index is $0 \leqslant k \leqslant L-1$, where L denotes the length of the input signal S. This implies that for an input time series of length L, the application of the DCT yields a frequency-domain sequence S of the same length L, comprising L frequency components ranging from S_0 (the DC component) to $S\{L-1\}$ (the highest frequency component).

The standard formula for the DCT-II transform is as follows:

$$S_k = \alpha(k)\sum_{i=0}^{L-1} s_i \cos\left(\frac{\pi}{L}k(i+0.5)\right) \tag{5}$$

The normalization constant $\alpha(k)$ is defined as follows:

$$\alpha(k) = \begin{cases} \sqrt{1/L} & \text{if } k=0 \\ \sqrt{2/L} & \text{if } k>0 \end{cases} \tag{6}$$

The one-dimensional DCT can be expressed as:

$$f_1^l = \sum_{i=0}^{L_s-1} x_i^{1d} B_l^i \tag{7}$$

where $l \in \{0,1,\ldots,L_s-1\}$; $f_1^l \in \mathbb{R}^L$ represents the one-dimensional DCT frequency spectrum, $x^{1d} \in \mathbb{R}^L$ is the input signal, L is the length of the signal x^{1d}.

The fundamental concept of the DCT is to compute the weighted summation of the time-domain signal with respect to orthogonal basis functions, thereby yielding its frequency-domain representation. Accordingly, the formula can be explicitly expressed as:

$$f_l^1 = \sum_{i=0}^{L-1} x_i \cos\left(\frac{-\pi l\left(i+\frac{1}{2}\right)}{L}\right) \tag{8}$$

The basis functions of the inverse one-dimensional DCT are identical to those in the forward transformation, thereby preserving the symmetry between the time-domain and frequency-domain representations. Accordingly, the inverse one-dimensional DCT can be written as:

$$x^{1d}x_i^{1d} = \sum_{l=0}^{L_s-1} f_1^l B_l^i \tag{9}$$

where x_i^{1d} represents the value of the i-th sample point recovered through inverse DCT, with range of values falling within $i \in \{0,1,\ldots,L_s-1\}, f_1^l \in \mathbb{R}^L$.

Through the weighted summation using the same basis functions, the frequency-domain signal is reconstructed into the time-domain signal. The specific formula is presented as follows:

$$x_i^{1d} = \sum_{l=0}^{L-1} f_l^1 \cos\left(\frac{-\pi l\left(i+\frac{1}{2}\right)}{L}\right) \tag{10}$$

Note that in Equations (5) and (7), certain constant normalization factors have been omitted for the sake of simplicity in expression and will not impact the outcomes of the study.

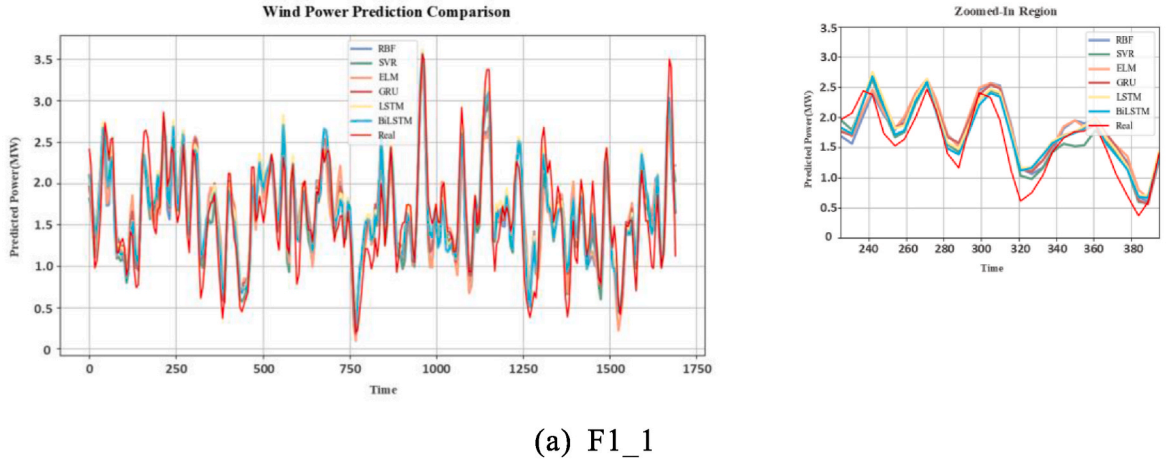
When the processed frequency-domain information is fed into the Transformer encoder, it enables the model to generate more representative high-dimensional representations. This enhances the model's ability to understand and predict fluctuations in wind power, improving its effectiveness in capturing complex dynamic variations.

2.1.2. Dynamic aggregation blocks

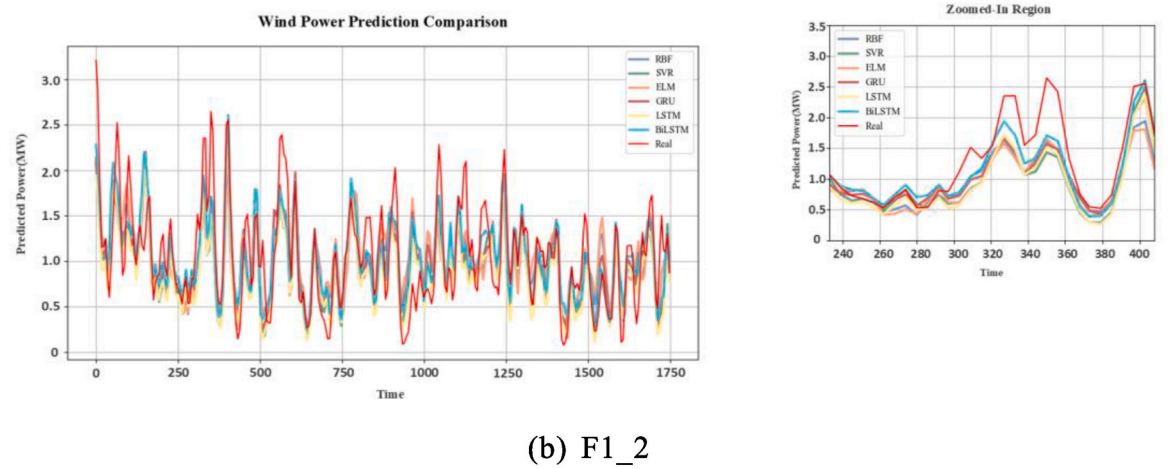
The dynamic aggregation block is an innovative attention mechanism designed to effectively capture and utilize spatiotemporal relationships in data. It combines the strengths of self-attention and cross-attention mechanisms by introducing proxy tokens, extending the traditional attention mechanism to a quadruplet form (Q,A,K,V). This approach enhances the model's representation capability while significantly reducing computational overhead.

Based on both Softmax attention and linear attention, Agent Tokens are introduced to facilitate the aggregation and dissemination of information. Below, these two approaches will be elaborated.

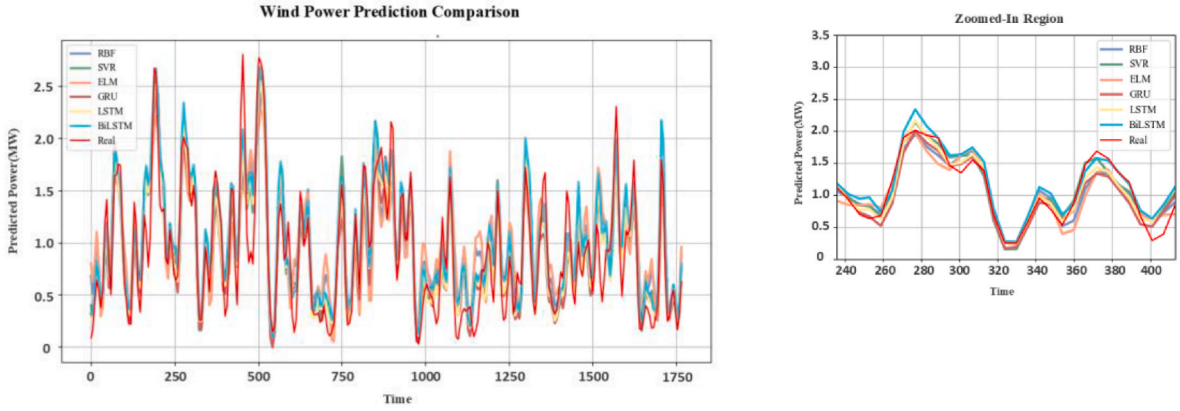
The fundamental formula of the Softmax attention mechanism is presented as follows:



(a) F1_1



(b) F1_2



(c) F1_3

Fig. 5. F1 Forecasted actual values of wind power.

$$O_S = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{11}$$

Linear attention modifies the order of query and key computations by applying the mapping function $\phi(\cdot)$, thereby reducing computational complexity. The formula is as follows:

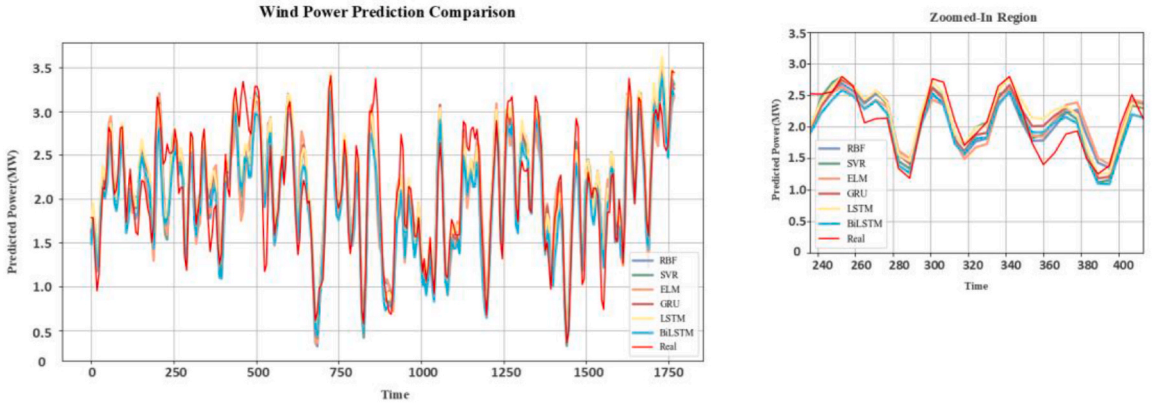
$$O_\phi = \phi(Q)\left(\phi(K)^T V\right) \tag{12}$$

Where, $\phi(Q)$ and $\phi(K)$ are the mapping functions for queries and keys, respectively.

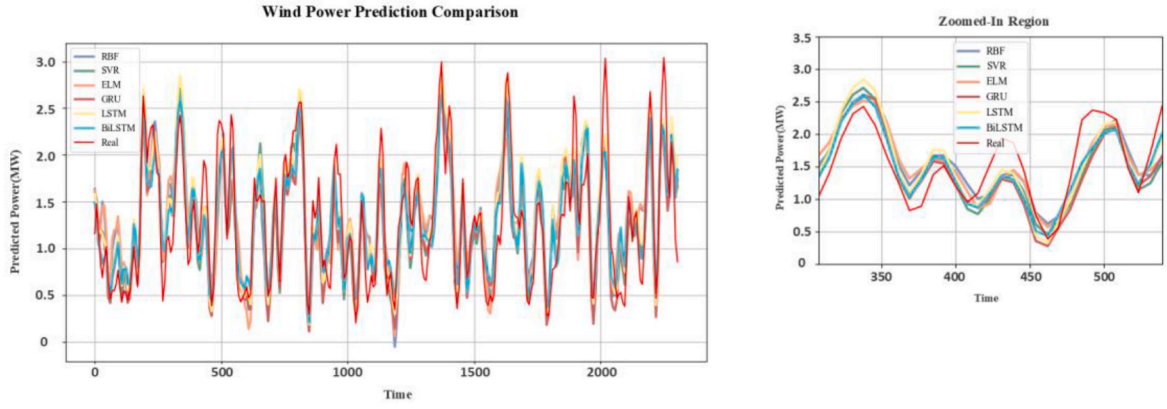
During the process of information aggregation and integration, this module divides the attention mechanism into two steps: proxy aggregation and proxy broadcasting.

For proxy aggregation, the proxy token A first acts as the query, interacting with the key K and value V through a traditional soft attention operation:

$$VA = \text{Softmax}\left(AK^T\right)V \tag{13}$$



(d) F1_4



(e) F1_5

Fig. 5. (*continued*).

where the proxy token A effectively aggregates information from the value token V. Each proxy token dynamically calculates attention weights, allowing it to emphasize important features effectively. This enhances the model's sensitivity to the global context.

For proxy broadcasting, proxy token A serves as a key to perform a secondary soft attention operation with the original query token Q, distributing aggregated information to each query token.

$$O = \mathrm{Softmax}\left(QA^T\right)VA \tag{14}$$

This mechanism avoids direct interaction between queries and the original keys, while leveraging a dynamically updated weight allocation to enhance the model's adaptability to complex data structures. It significantly reduces computational complexity while ensuring efficient information propagation.

The final formula of the proxy attention mechanism combines the two steps of proxy token aggregation and broadcasting. In this process, the proxy token serves not only as an aggregator of information but also efficiently disseminates the aggregated information back to the query space via broadcasting, thereby enhancing the expressiveness of the model.

This step makes full use of the role of proxy tokens in different stages to achieve efficient flow and processing of information in the following form:

$$O_A = \mathrm{Softmax}\left(QA^T\right)\mathrm{Softmax}\left(AK^T\right)V \tag{15}$$

To further enhance the model's representational capacity and computational efficiency, proxy bias and depth convolution steps are introduced within the dynamic aggregation block. These additions enrich feature diversity and improve positional sensitivity, often incorporating bias terms in attention computation:

$$O = \mathrm{Softmax}\left(QA^T + B_2\right)\mathrm{Softmax}\left(AK^T + B_1\right)V \tag{16}$$

where B_1 and B_2 provide rich positional encoding information, enabling the proxy tokens to focus on distinct feature regions, thereby further enhancing the model's performance.

The proxy token is a learnable tensor with dimensionality (C, d). Here, C denotes the number of proxy tokens, a hyperparameter that is significantly smaller than the input sequence length L (i.e., $\mathrm{C} \ll \mathrm{L}$). The variable d denotes the feature dimension, which is identical to those of the Query, Key and Value vectors in the model.

The proxy biases are introduced to enrich the model's positional encoding capabilities by incorporating additional learnable positional information. In our implementation, these two biased terms are incorporated as learnable parameter tensors within the model architecture. These parameters are typically set using standard random initialization schemes and subsequently optimized in an end-to-end manner via the backpropagation algorithm throughout the model training process.

The dynamic aggregation block leverages the proxy token aggregation and broadcasting mechanism to enhance the processing and fusion of complex data features, enabling more flexible handling of high-dimensional data. Since the number of proxy tokens can be designed to be significantly smaller than that of query tokens, this module reduces the computational complexity of input sequences from $O(N^2)$ to $O(N)$ for a token count of N, This achieves efficient global context modeling, providing substantial support for wind power prediction models.

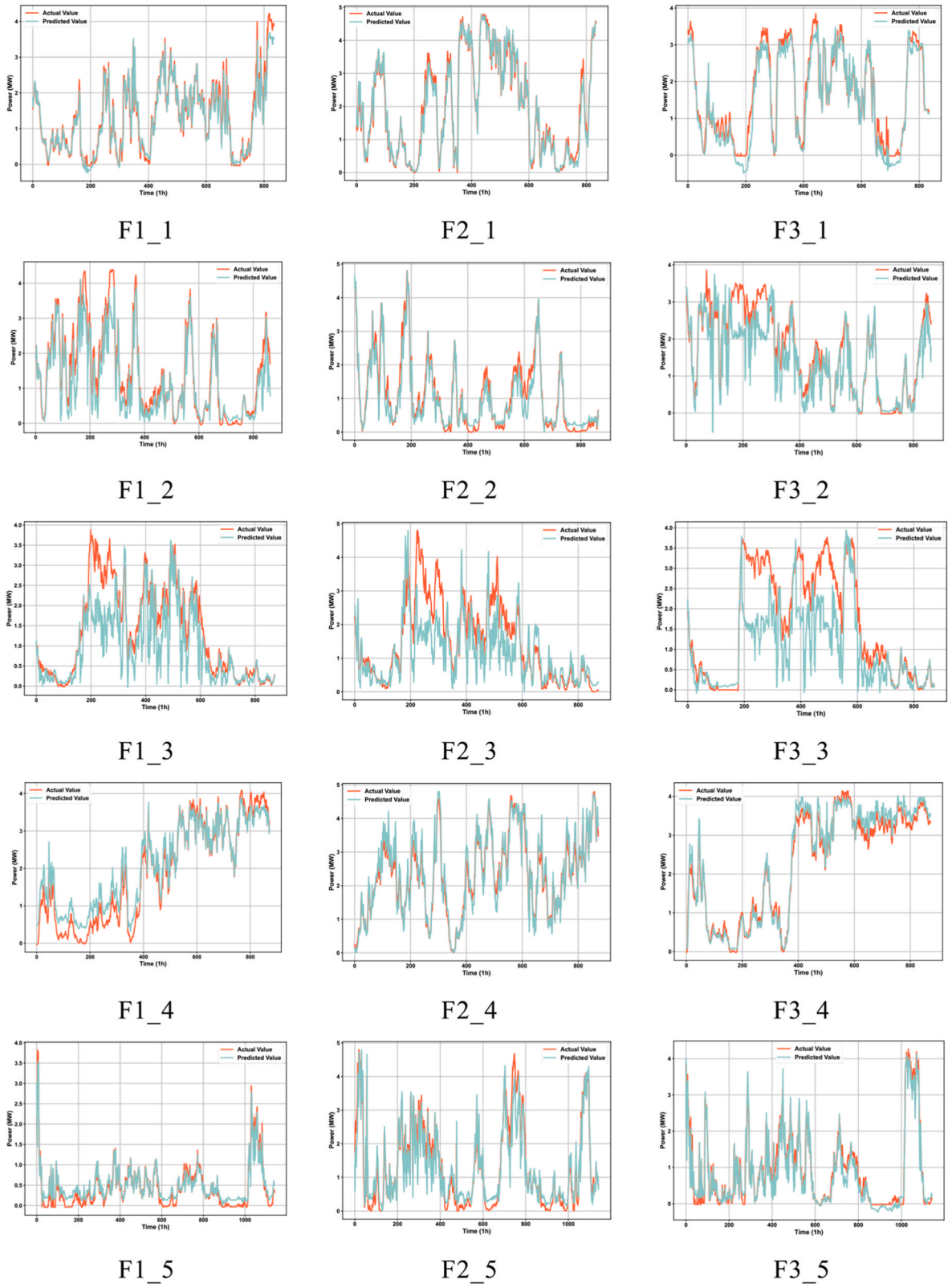


Fig. 6. Transformer Predicted vs Actual Wind Power.

2.2. *Mutual information (MI) method*

In information theory, mutual information (MI) is used to evaluate arbitrary dependencies between random variables [48]. For two discrete variables X and Y, their mutual information measures the amount of information about X provided by Y. If the two variables are independent, their mutual information equals zero. The MI between the two variables is defined as:

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) = H(X) + H(Y) - H(X;Y) \tag{17}$$

Table 2
Evaluation Metrics of the enhanced Transformer Model.

Dataset	MAE	RMSE	MAPE	R^2
F1_1	2.044	2.930	0.727	0.912
F1_2	2.695	3.636	1.011	0.908
F1_3	1.531	6.975	3.100	0.964
F1_4	3.108	6.312	1.369	0.929
F1_5	1.069	7.863	2.889	0.917
F2_1	4.433	5.726	0.458	0.948
F2_2	2.479	3.353	1.613	0.932
F2_3	2.965	4.462	0.282	0.942
F2_4	4.679	5.535	0.158	0.904
F2_5	2.812	4.186	2.237	0.949
F3_1	1.666	2.224	1.980	0.965
F3_2	2.297	2.827	0.366	0.934
F3_3	2.108	2.457	0.956	0.965
F3_4	3.157	3.572	0.967	0.926
F3_5	1.604	2.273	0.821	0.951

where H represents entropy, $H(X|Y)$ and $H(Y|X)$ denote conditional entropy, and $H(X;Y)$ is the joint entropy of X and Y, defined as:

$$H(X) = -\int_x p_X(x)\log p_X(X)dx \tag{18}$$

$$H(Y) = -\int_y p_Y(y)\log p_Y(y)dy \tag{19}$$

$$H(X;Y) = -\int_x\int_y p_{X,Y}(x,y)\log p_{X,Y}(x,y)dxdy \tag{20}$$

$$H(X|Y) = -\int_y\int_x p(x,y)\log\frac{p(x,y)}{p(y)}dxdy \tag{21}$$

where $P_{XY}(x,y)$ represents the joint probability density function, whereas $P_X(x)$ and $P_Y(y)$ are the marginal functions of X and Y, respectively the marginal functions are:

$$p_X(x) = \int_y p_{X,Y}(x,y)dy \tag{22}$$

$$p_Y(y) = \int_x p_{X,Y}(x,y)dx \tag{23}$$

By substituting Eqs. (18)–(20) into Eq. (17), the MI equation becomes:

$$I(X;Y) = \int_x\int_y p_{X,Y}(x,y)\log\frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)}dxdy \tag{24}$$

In the discrete form, integration is replaced by summation over all possible values occurring in the data. Therefore, estimating $P_{XY}(x,y)$ is sufficient to compute the MI between X and Y.

To compute the mutual information between continuous variables, the data are first discretized using the Equal-Frequency Binning strategy. This method sorts the feature values and partitions them into k bins, each containing approximately the same number of samples, thereby providing robustness against uneven data distributions or the presence of outliers. Following discretization, the joint probability $P_{XY}(x,y)$ and marginal probabilities $P_X(x)$ and $P_Y(y)$, which are required for mutual information computation, are estimated non-parametrically by counting the frequency of samples within each bin.

2.3. Seasonal trend residual decomposition (STL)

STL decomposition is a time series analysis method that breaks down time series data into three components: Trend, Seasonal and Residual [49]. The implementation is entirely based on numerical methods, making it suitable for processing large volumes of time series data [50].

The STL decomposition model is represented in an additive form as:

$$Y_t = T_t + S_t + R_t \tag{25}$$

where Y_t represents the observed value, T_t denotes the trend component, S_t is the seasonal component, R_t is the residual component and the index t denotes time. This decomposition method utilizes Loess (locally weighted regression) to smooth and remove seasonal patterns at the periodic level, for extracting the trend curve [51]. Loess employs the bi-cubic formula shown in Eq.(26) as the weight function for the i-th observation. The Loess regression function is defined as follows:

$$v_i(t) = \begin{cases} \left[1 - \left(\frac{|t_i - t|}{\delta_\gamma(t)}\right)^3\right]^3 & \text{for } 0 \le \frac{|t_i - t|}{\delta_\gamma(t)} \le 1 \\ 0 & \text{for } \frac{|t_i - t|}{\delta_\gamma(t)} \ge 1 \end{cases} \tag{26}$$

where $v_i(t)$ represents the weight of the i-th observation at time point t and $\delta_\gamma(t)$ denotes the distance from the farthest t_i to the current t.

2.4. Swing-door algorithm (SDA)

The working principle of SDA is illustrated in Fig. 2. SDA selectively retains data points by moving a "gate" along the data sequence. This "gate" swings with the changes in the data, recording a data point only when it exceeds a predefined threshold. The process continues until non-overlapping upper and lower bounds for future data are reached, completing one iteration of the algorithm and achieving data compression [40].

Let us assume the starting point of the SDA is S, with the measured data points and gate points connected by line segments. The line formed by the upper gate point U and the peak point B defines the upper bound, and the line connecting the lower gate point L to the measured data point D defines the lower bound. The segment connecting point D to the starting point S is identified as the compression segment, and point D becomes the starting point for the next compression iteration. The gate width ε is the only adjustable parameter, directly influencing the sensitivity to data fluctuations. The termination condition for the iteration is expressed as follows:

$$k_{UB} \ge k_{LD} \tag{27}$$

$$k_{UB} = \frac{P_B - (P_S + \varepsilon)}{t_B - t_S} \tag{28}$$

$$k_{LD} = \frac{P_D - (P_S - \varepsilon)}{t_D - t_S} \tag{29}$$

where k_{UB} and k_{LD} represent the slopes of the upper and lower bounds, respectively; P_S is the wind power value at the starting point S and t_S is the time at the starting point S. Similarly, P_D and t_D denote the power value and time of the connecting point D whereas P_B and t_B correspond to the power value and time of the peak point B.

2.5. Evaluation metrics

To evaluate the performance of the proposed wind power prediction model, various error metrics have been employed: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Coefficient of Determination (R^2)and Mean Absolute Percentage Error (MAPE).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2} \tag{30}$$

$$MAPE = \sum_{i=1}^{n}\left|\frac{y_i - x_i}{y_i}\right| \times \frac{100}{n} \tag{31}$$

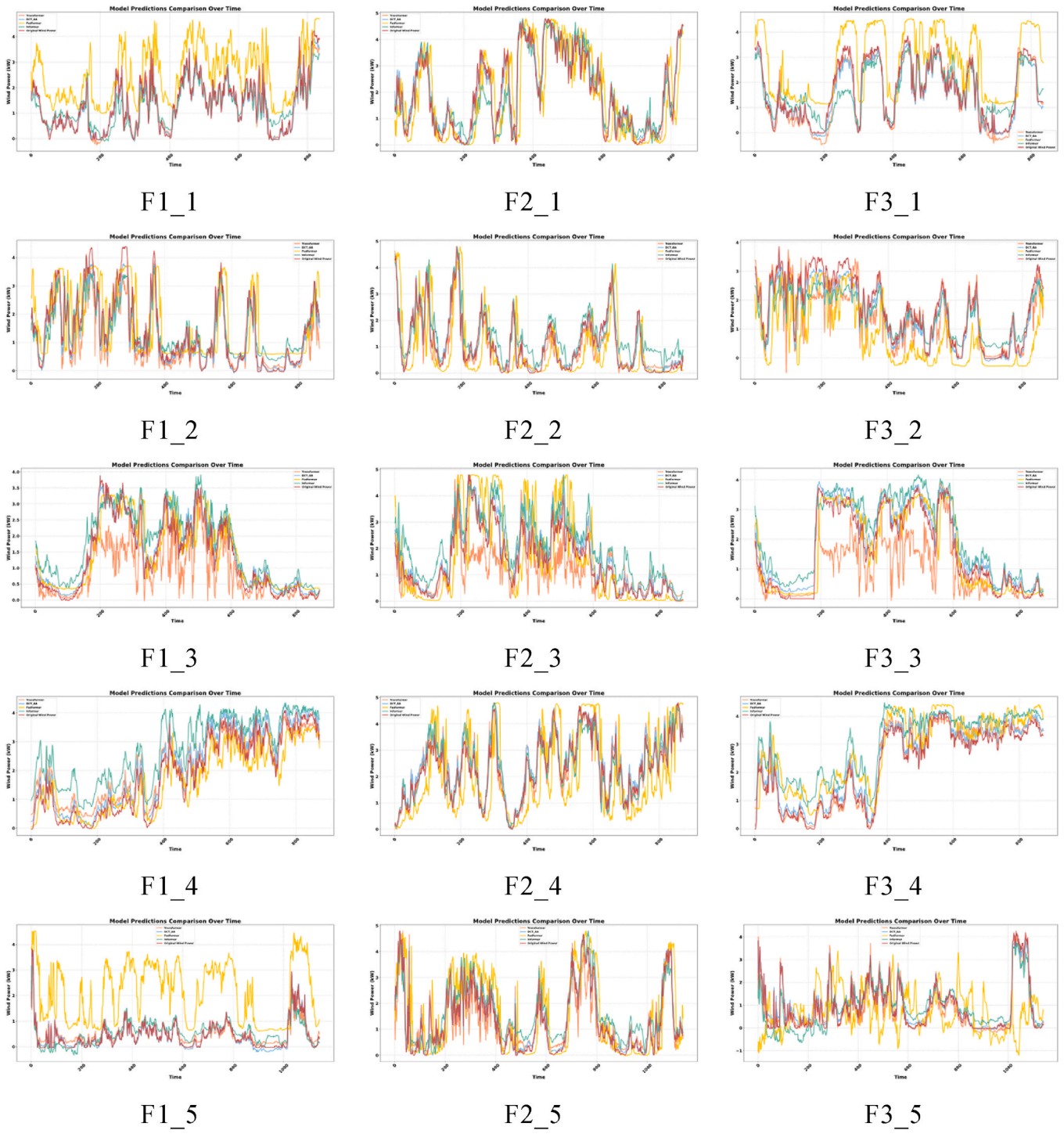


Fig. 7. Comparison of the Model Before and After Enhancement by the Spectrum and Dynamic blocks.

$$R^2 = 1 - \frac{\sum_{t=1}^{n} |y_t - \hat{y}_t|}{\sum_{t=1}^{n} |y_t - \bar{y}_t|} \tag{32}$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - x_i| \tag{33}$$

where y_i represents the actual output of the i-th training sample, x_i is the predicted value of the i-th sample, $\overline{x_i}$ is the mean of the observed samples $\overline{y_i}$ is the mean of the simulated samples, and n is the total number of samples.

For the evaluation of the early warning system, through the PI interval [52], the adopted metrics are Prediction Interval Coverage Probability (PICP) and Prediction Interval Nominal Confidence (PINC).

$$PINC = 1 - P(N(t) > \gamma | Y(t)) \tag{34}$$

where $N(t)$ represents the noise component, γ is the predefined threshold and $Y(t)$ denotes the observed signal.

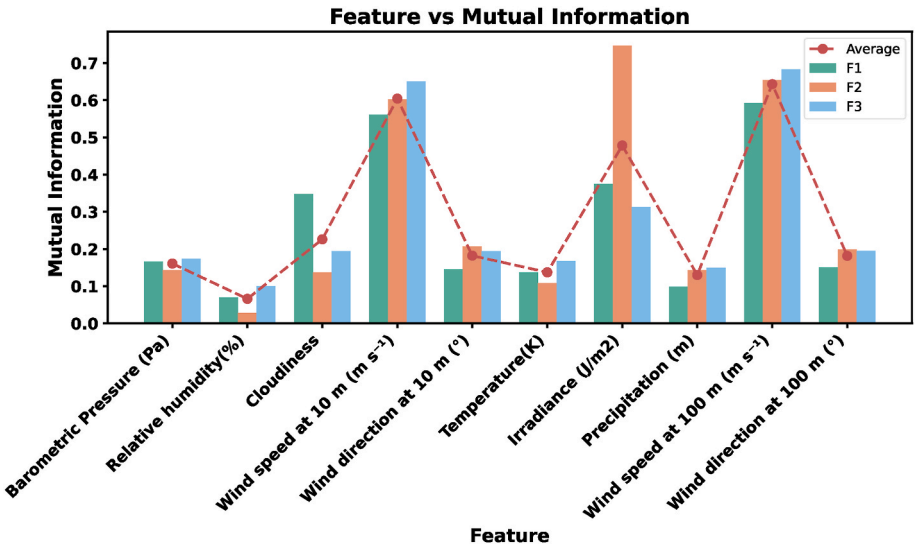


Fig. 8. MI analysis results.

$$PICP = \frac{1}{N}\sum_{i=1}^{N} a_i \tag{35}$$

where, N represents the number of test data samples, and a_i is a binary value defined by the following formula:

$$a_i = \begin{cases} 1, & y_i \in \left[y_{i,l}, y_{i,u}\right] \\ 0, & y_i \notin \left[y_{i,l}, y_{i,u}\right] \end{cases} \tag{36}$$

In Equation (36), y_i is the value of the test sample i, while $y_{i,l}$ and $y_{i,u}$ represent the estimated upper and lower bounds, respectively.

2.6. Proposed monitoring model

The modeling process of the designed spectrum-based dynamic aggregation Transformer for prediction and fitting in the rotary gate monitoring model is illustrated in Fig. 3. The main steps are as follows:

Step 1: Initialize the Transformer model parameters, spectrum block parameters, and the number of proxy tokens in the dynamic aggregation block.
Step 2: Preprocess the raw time series data, including normalization to scale the data to the range [0,1], and split it into training, validation, and test sets in appropriate proportions.
Step 3: Apply Discrete Cosine Transform (DCT) to the data for feature extraction, converting time-domain data into frequency-domain data to extract frequency components.
Step 4: Input the extracted frequency features into the spectrum block for further processing. Utilize the attention mechanism within the spectrum block for compression and weight adjustment.
Step 5: Utilize the dynamic aggregation block, where proxy tokens act as queries (Q) to aggregate information from keys (K) and values (V) and broadcast it to the query tokens.
Step 6: Input the features extracted by the spectrum block and the dynamically weighted features from the dynamic aggregation block into the Transformer encoder. Use the multi-head attention mechanism for deep encoding of the input features, further extracting temporal dependencies.
Step 7: Pass the high-level features output by the Transformer encoder to the Transformer decoder, which generates the final prediction results.
Step 8: Use the MI method to quantify the nonlinear dependencies between input features and wind power output, identifying key variables.
Step 9: Apply STL decomposition to the selected key features to separate different frequency components.
Step 10: Use the relationship between the decomposed features and wind power to perform fitting with a random forest model, and set early warning thresholds based on the results.
Step 11: Identify ramp events using the SDA algorithm, detect and issue warning signals, achieving dual early warning functionality.

3. Case study

3.1. Data set partitioning and processing

The dataset encompasses data collected from three wind turbines, each with a capacity of 4.8 MW, located in the China Fujian region. Measurements were recorded at 15-min intervals, including wind power output, wind speed, wind direction, cloud coverage, temperature, humidity and atmospheric pressure. Given the turbine hub height considerations, wind speed and direction were measured at both 10m and 100m elevations. The data collection period spans from January 3, 2022, to April 30, 2023. Wind energy-related parameters exhibit pronounced seasonal characteristics, and the intensity of fluctuations varies significantly across different seasons. Ignoring these season-specific patterns of change may result in substantial prediction errors in the model during specific times of the year. Meanwhile, in the operation and maintenance of actual wind farms, seasonal modeling can more effectively align with the temporal planning of grid scheduling and turbine maintenance. By taking into account the modeling characteristics and generalization performance of models across different seasons, this paper partitions the entire dataset into five quarterly subsets in chronological order. Fig. 4 illustrates the fluctuation trends observed in the raw data from all three turbines.

Only the raw data trends from Turbine No. 1 are presented here, additional data are provided in the Appendix. The turbines are designated as F1, F2 and F3, corresponding to Turbines No. 1, 2,and 3, respectively, to facilitate problem and results presentation. Although all three turbines are situated within the same geographic region, their parameter data exhibit marked variations among them. These variations serve as indication of the operational characteristics under diverse meteorological conditions in the wind farms where they are located and directly influence wind power generation. In this study, the model is

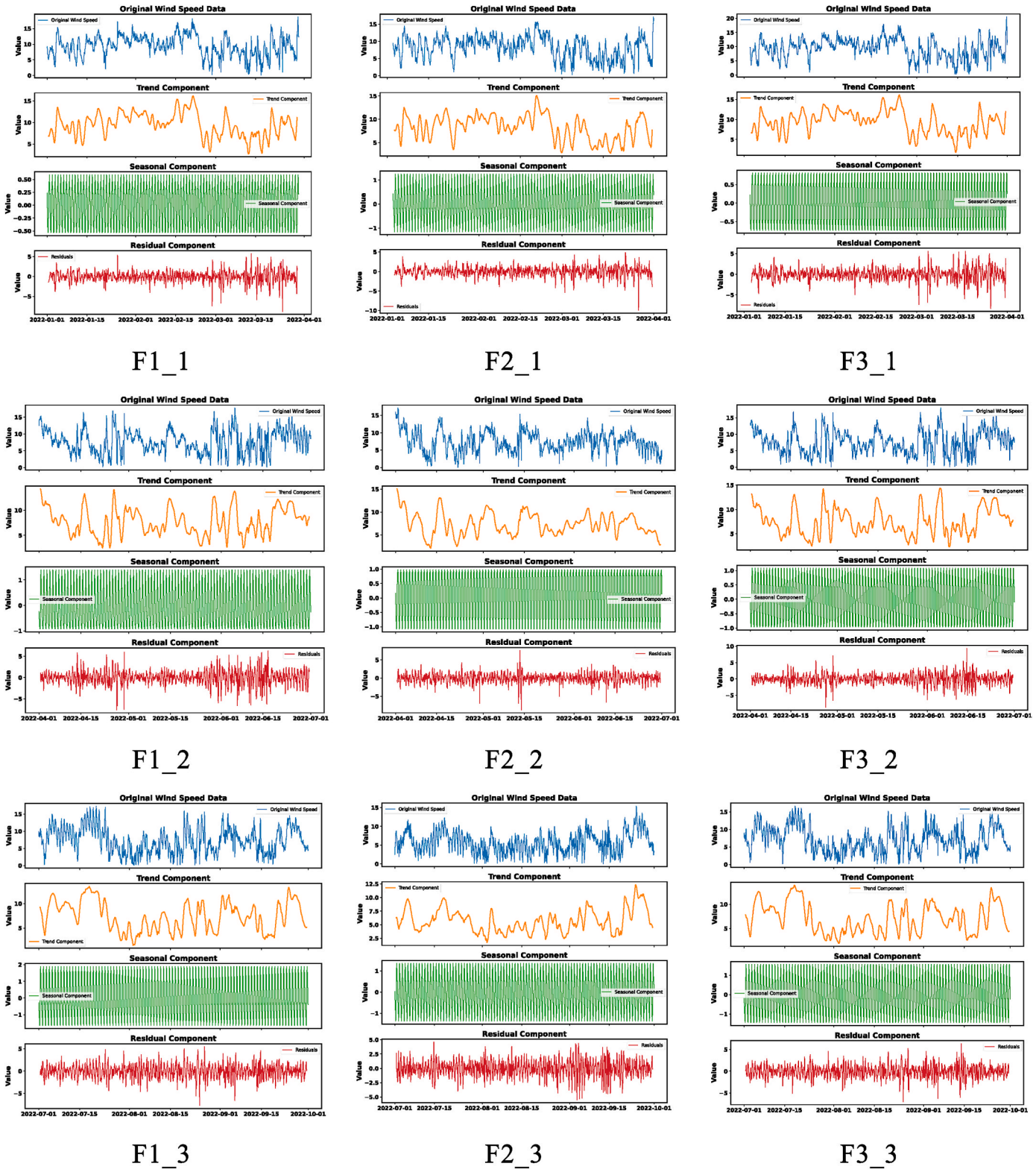


Fig. 9. STL decomposition models.

globally constructed to avoid the introduction of noise arising from differences in turbine characteristics during sub-turbine modeling, thereby enhancing the overall generalization capability of the model. At the same time, the adoption of single-turbine modeling significantly enhances the efficiency and precision of grid scheduling operations, wind turbine maintenance, and advanced warning systems. Preliminary data preprocessing was needed to optimize the model's predictive capabilities. The dataset was partitioned into training, validation and testing sets with a ratio of 7:2:1. To eliminate dimensional inconsistencies that could impact model predictions, all data values were

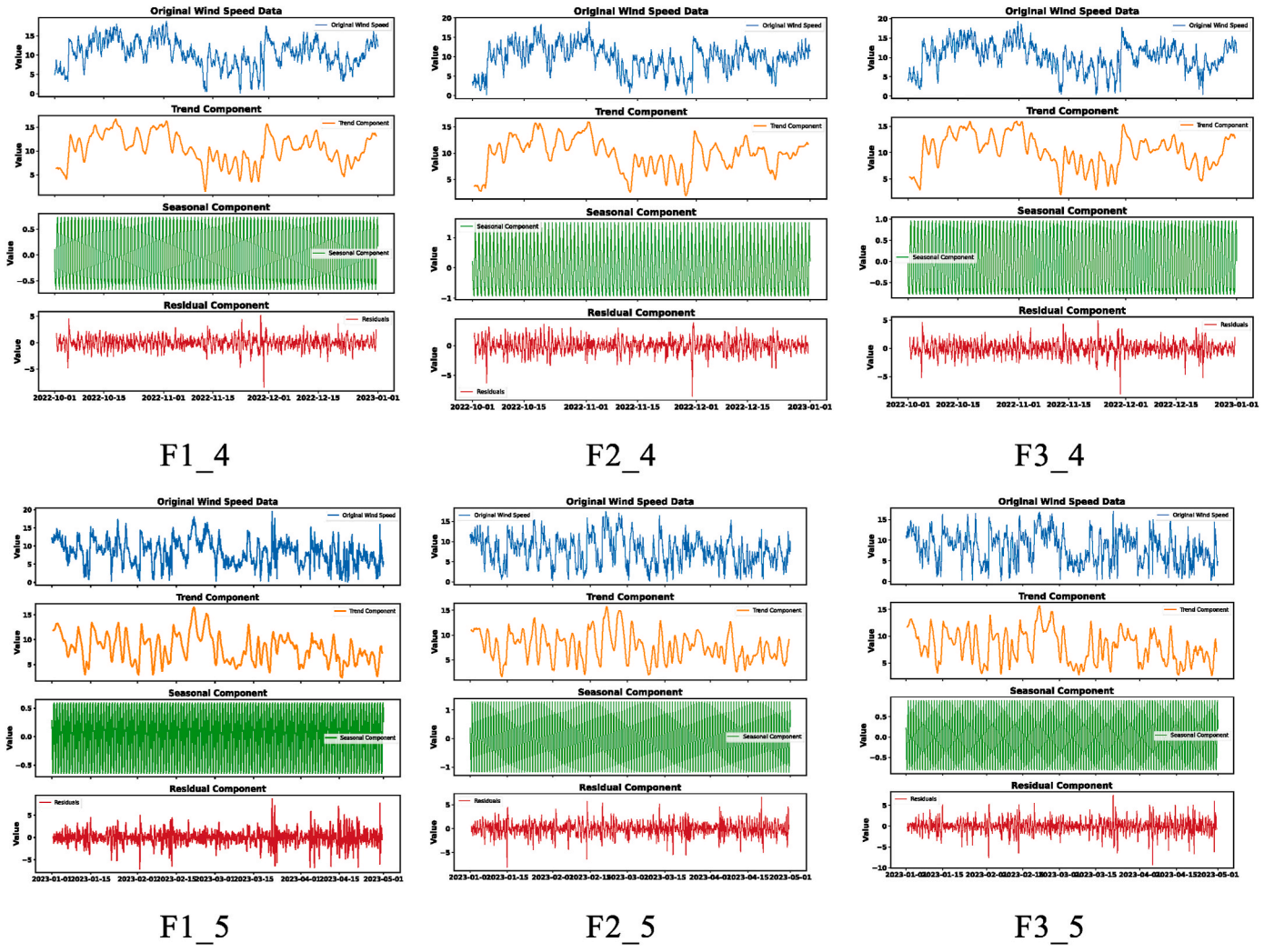


Fig. 9. (*continued*).

normalized and scaled to the range of [0,1] as follows:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{37}$$

where x represents the original sequence, x' denotes the normalized sequence, $x_{\max}$ and $x_{\min}$ represent the maximum and minimum values of the sequence, respectively.

To obtain the actual wind power sequence, the wind power data must be denormalized to produce the final wind power sequence.

$$\hat{y} = (y_{\max} - y_{\min}) \times \hat{y}' + y_{\min} \tag{38}$$

$\hat{y}'$ represents the normalized wind power prediction sequence in the test set; $\hat{y}$ represents the wind power prediction sequence in the test set after denormalization; $y_{\max}$ and $y_{\min}$ represent the maximum and minimum values of the wind power sequence in the training set, respectively.

3.2. Offshore wind power prediction

3.2.1. Prediction based on the transformer base model

To validate the feasibility of the proposed methodology, several baseline models were selected for preliminary analysis. The comparative study was conducted by developing eight models: BP, LSTM, Bi-LSTM, ELM, SVR, RBF, GRU and Transformer. These models were independently applied to predict the wind power output of F1, F2 and F3. The datasets were partitioned according to the previously established ratio to ensure prediction accuracy. The performance of each model was systematically evaluated through multiple metrics, including MAE, RMSE, MAPE and R^2, these metrics were employed to comprehensively assess the predictive capabilities of the respective models (See Table 1).

Based on the performance metrics for F1(The F2 and F3 datasets are shown in the Appendix), the Transformer model demonstrates superior performance in key indicators including MAE, RMSE, and R^2. This is particularly evidenced in F1_1, F1_4 and F1_5 (where the notation F1_2 denotes Turbine 1 in the second quarter), exhibiting minimal errors and optimal fitting capabilities. SVR and LSTM models are ranked second in performance, maintaining consistent reliability across quarters, with notably exceptional results in F1_3, where R^2 values are improved by approximately 5–10 %. The Bi-LSTM model also exhibits commendable performance in RMSE and MAPE metrics. In contrast, ELM and RBF models are characterized by comparatively higher error rates and suboptimal fitting performance, particularly evident in F1_2 and F1_3, where MAE and MAPE increase by over 20 %, accompanied by significant R^2 degradation. Whereas the GRU model maintains relatively stable performance overall, a slight increase in error rates is observed in F1_2. In conclusion, the Transformer model consistently outperforms other models across all quarters, followed closely by SVR and LSTM.

To further illustrate the performance difference among models across different seasons, radar charts and box plots are shown in Appendix for multidimensional analysis of key reference metrics, as well as the parameter settings for the base models.

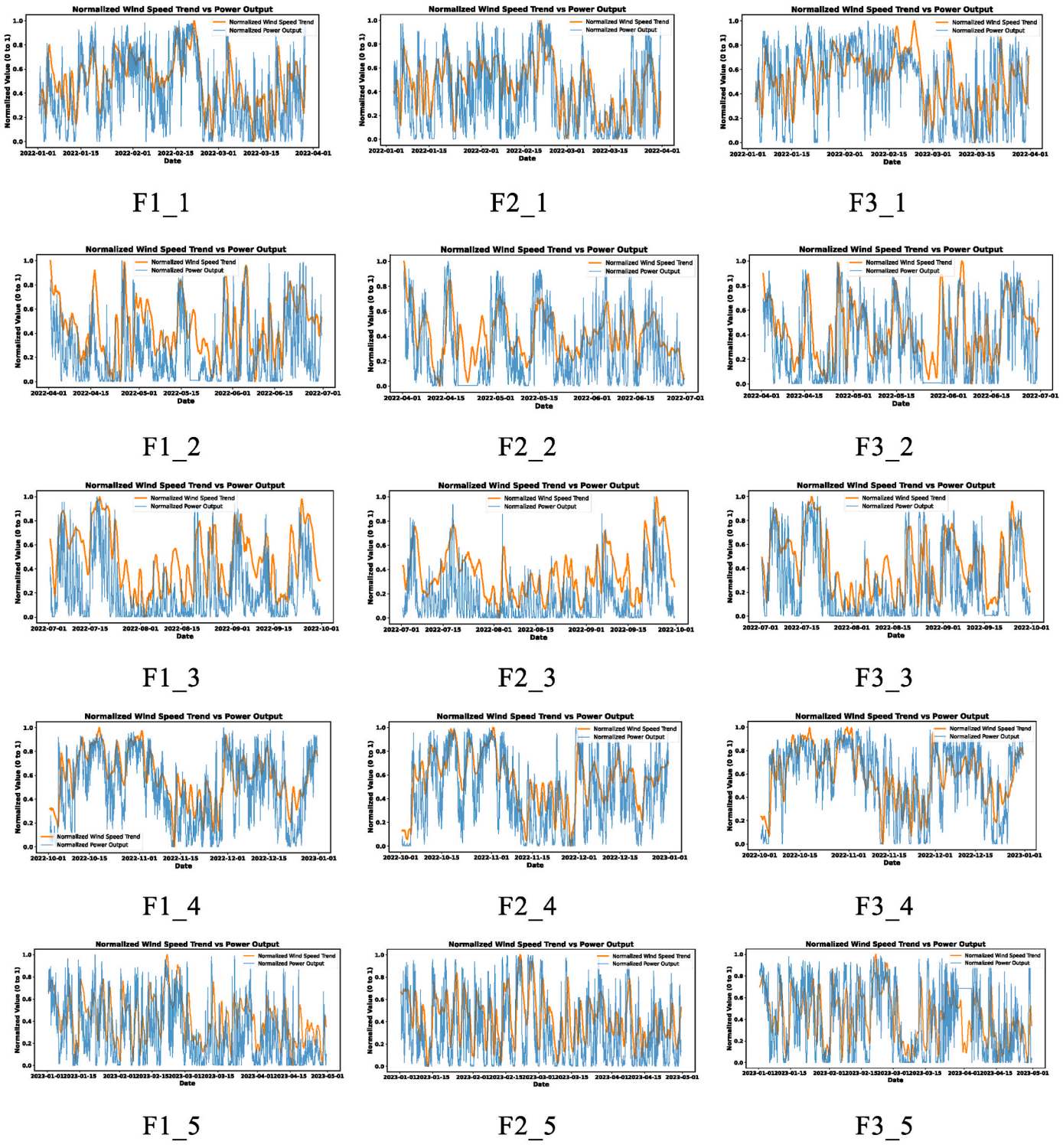


Fig. 10. Decomposed wind speed trends and wind power.

Table 3
Degree of overlap.

Dataset	Overlap	Dataset	Overlap	Dataset	Overlap
F1_1	0.903	F2_1	0.891	F3_1	0.901
F1_2	0.869	F2_2	0.889	F3_2	0.830
F1_3	0.873	F2_3	0.865	F3_3	**0.917**
F1_4	**0.925**	F2_4	**0.923**	F3_4	0.914
F1_5	0.858	F2_5	0.903	F3_5	0.867

The analysis of its performance metrics across its baseline models reveals the Transformer's superior performance across most datasets, particularly with respect to MAE, RMSE, and R^2 metrics, demonstrating robust predictive capabilities in handling high-dimensional, complex wind power forecasting tasks. Notable limitations were observed in MAPE performance for specific datasets, particularly during F1_3 and F1_5 quarters, indicating seasonal vulnerability to extreme data patterns. SVR, LSTM and Bi-LSTM models have demonstrated commendable stability, achieving low MAE, RMSE and high R^2 values across

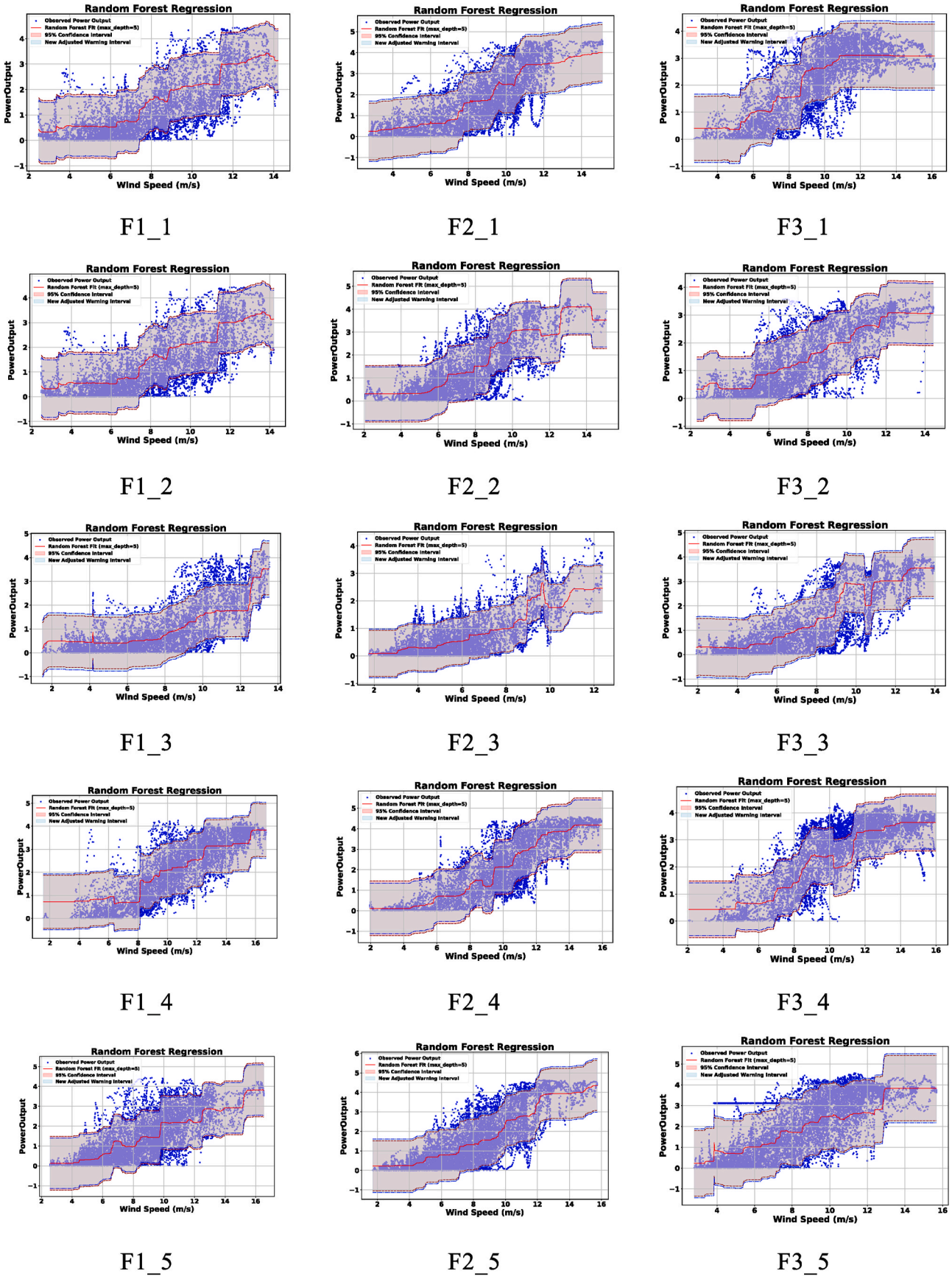


Fig. 11. RF chart.

Table 4
Interval coverage probability.

Dataset	PICP1	PICP2	Dataset	PICP1	PICP2	Dataset	PICP1	PICP2
F1_1	93.32 %	92.79 %	F2_1	94.19 %	**94.48** %	F3_1	**94.49** %	**94.81** %
F1_2	93.32 %	92.67 %	F2_2	93.82 %	93.70 %	F3_2	94.17 %	93.95 %
F1_3	93.07 %	93.75 %	F2_3	93.51 %	93.68 %	F3_3	93.41 %	93.58 %
F1_4	94.16 %	**94.25** %	F2_4	**94.37** %	93.75 %	F3_4	94.32 %	93.85 %
F1_5	**94.39** %	94.14 %	F2_5	93.96 %	94.41 %	F3_5	93.41 %	93.44 %

multiple datasets. BP performed similarly to GRU, showing moderate overall performance with acceptable fitting in some instances but substantial MAE and RMSE errors. ELM and RBF models consistently underperformed, exhibiting significant prediction errors and poor fitting capabilities, particularly evident in MAE and RMSE metrics.

Prediction curves comparing baseline models output predictions against actual values are presented in Fig. 5.

The comparative analysis of predicted versus actual wind power curves across a specific quarter (exclusively showing F1 turbine's baseline models, with other turbine data available in the Appendix) reveals substantial deviations between all baseline models and the actual values (depicted by the red curve). These discrepancies become increasingly pronounced in the latter half of the timeline, highlighting particular challenges in long-term prediction capabilities. The magnified sections provide detailed focus on model predictions during specific time intervals, clearly demonstrating the baseline models' inferior accuracy. These limitations are particularly evident during sudden fluctuations, where prediction deviations become notably pronounced due to model sensitivity issues. None of the aforementioned baseline models demonstrate substantially good in wind power prediction accuracy.

To highlight the good performance of Transformer as the base model for prediction, its visualization is presented in Fig. 6.

The Transformer model's predictions slow proof overlap with the actual values across various time intervals demonstrating relatively accurate tracking of rapid fluctuations in both ascending and descending trends, whereas the other models are considered to exhibit notable lag or fail to capture peak values.

Indeed, although predictions closely approximate actual values across most time periods, some limitations are observed also for the Transformer model in handling sudden anomalies.

The superior performance of the Transformer model primarily stems from its self-attention mechanism, which simultaneously captures feature information across all time steps in the input sequence, enabling more flexible feature extraction. Furthermore, the model's multi-head attention mechanism enables multi-perspective focus on key features within input data, enhancing long-term information retention capabilities. These characteristics enable the Transformer to effectively enhance prediction accuracy when processing complex time-series data for wind power forecasting, demonstrating robust adaptability and reliability. Consequently, the Transformer has been selected as the foundation for subsequent wind power prediction model development.

3.2.2. Prediction access after incorporating spectral and dynamic aggregation blocks into the transformer model

The Transformer model architecture is enhanced by incorporating spectral blocks and dynamic aggregation block attention mechanisms, for improving model reference stability.

DCT is applied to input data for frequency domain analysis of temporal features, enabling enhanced extraction of periodic patterns and long-term dependencies within time-series data. This transformation facilitates improved comprehension of frequency components, particularly crucial for wind power data interpretation.

The integration of attention mechanisms, specifically dynamic aggregation blocks, enables interactive feature processing through weighted attention to significant feature combinations. In multi-feature inputs, relationships between individual features potentially influence final output results. The dynamic aggregation blocks assign varied weights to these feature dependencies, facilitating optimal information integration across features.

The enhanced architecture is designed to address the base Transformer's limitations while maintaining its fundamental advantages in temporal sequence processing. The evaluation metrics for the Transformer-based model indeed seem to improve with the addition of a dual-attention mechanism as reported in Table 2.

The reference metrics demonstrate notably low prediction errors, with the consistently high R^2 values reflecting the model's robust fitting capabilities. For a clear comparison of the model's improvement post-enhancement, the performance metrics of the base Transformer and the enhanced model with the additional spectrum and dynamic blocks are visually compared in Appendix.

In order to clearly demonstrate the superiority of the improved model predictions, the Informer and FEDformer models are introduced for comparison, with their corresponding parameter settings provided in the Appendix, as specifically illustrated in Fig. 7.

It is evident from the figure that, apart from the improved Transformer model proposed in this study (blue curve), the other enhanced models exhibit substantial deviations when fitting the actual values (red curve), particularly in regions with rapid data fluctuations. In contrast, the model proposed in this study exhibits superior fitting capability in response to drastic changes and more precisely captures the oscillatory trend of actual values, thereby validating its predictive accuracy and stability in complex dynamic scenarios.

The integration of spectral blocks and dynamic aggregation blocks enables more refined feature interpretation, resulting in enhanced prediction accuracy. The spectral blocks facilitate feature extraction in the frequency domain, whereas the dynamic aggregation blocks identify the most representative components among all features, ensuring the model's access to comprehensive temporal and frequency information. This dual mechanism significantly enhances the model's predictive capabilities. The enhanced model demonstrates, superior capability in capturing local details, improved processing of complex temporal variations, enhanced performance in handling high volatility and uncertainty inherent in wind power generation.

3.3. Correlation between relevant features and wind power

3.3.1. Correlation analysis using MI method

To investigate the influence of different factors on wind power, MI analysis has been employed to evaluate feature correlation for turbines F1, F2 and F3. The mutual information method objectively quantifies arbitrary nonlinear correlations between variables, thereby avoiding subjective empirical biases. By providing precise inputs for subsequent warning systems, it effectively reduces computational complexity [53, 54] and significantly enhances the timeliness of responses. The results are illustrated in Fig. 8.

The Figures illustrate the MI between wind power features across all three turbines. Average wind speed measurements at 100 m height have consistently high MI values across all three turbines, highlighting its reference to wind power prediction. In contrast, other features such as relative humidity and precipitation have notably lower MI values, indicating their relatively minor contribution to wind power prediction.

Given the importance of 100 m wind speed measurements across all

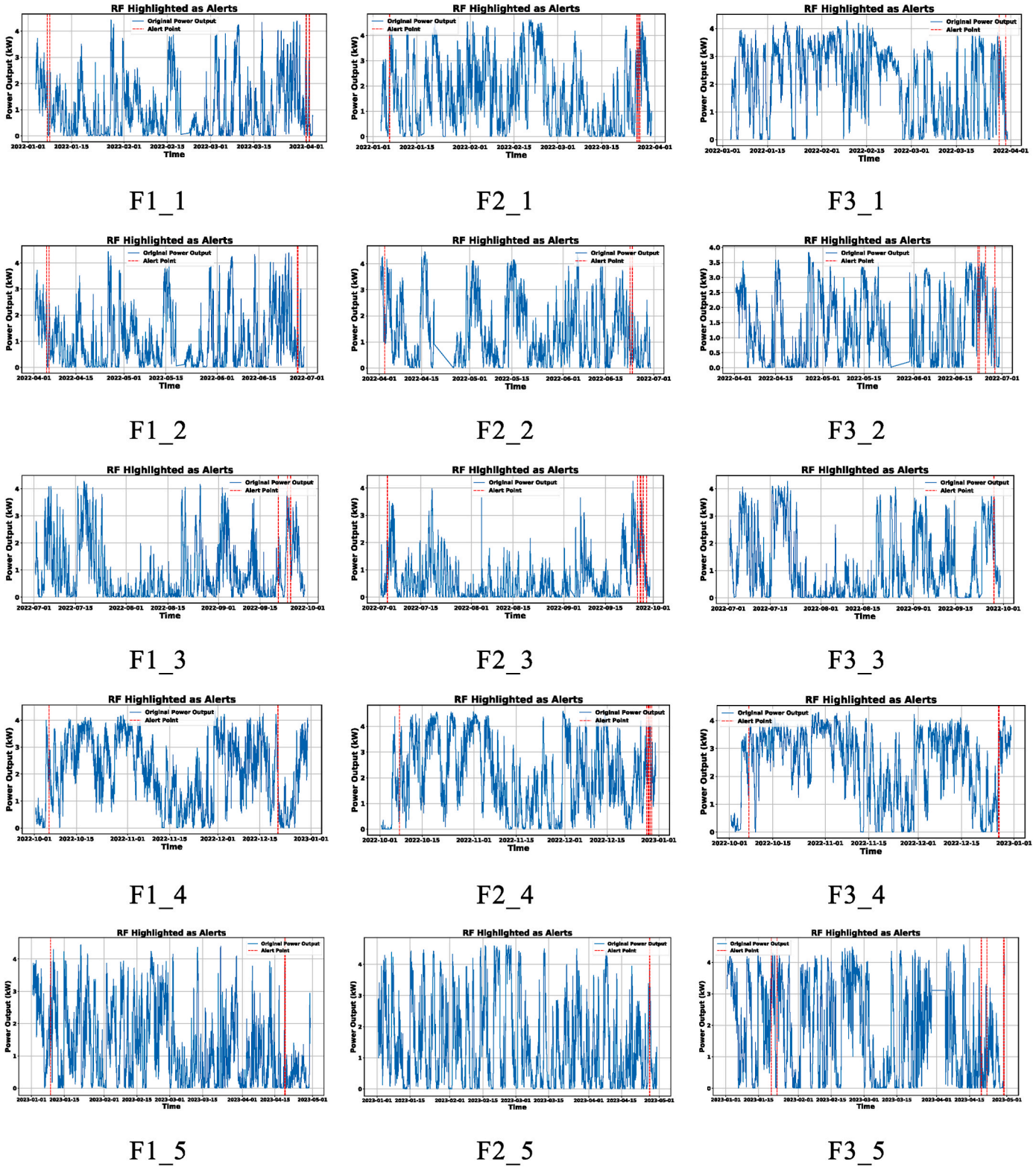


Fig. 12. RF advanced warning.

three turbines, this feature has been selected for subsequent detailed decomposition and analysis.

3.3.2. STL decomposition

With reference to the key feature of wind speed at 100 m height, Seasonal and Trend decomposition using Loess (STL)has been reformed to decompose data across different quarters for each turbine, the results are known in Fig. 9.

Four distinct curves illustrate the decomposition of wind speed data, original wind speed data, wind speed trend curve, seasonal trend curve, residual curve. The STL decomposition effectively handles non-linear relationships by extracting trends and seasonal components, processing diverse seasonal patterns and clearly separating distinct components of wind speed data. The trend component shows the long-term wind

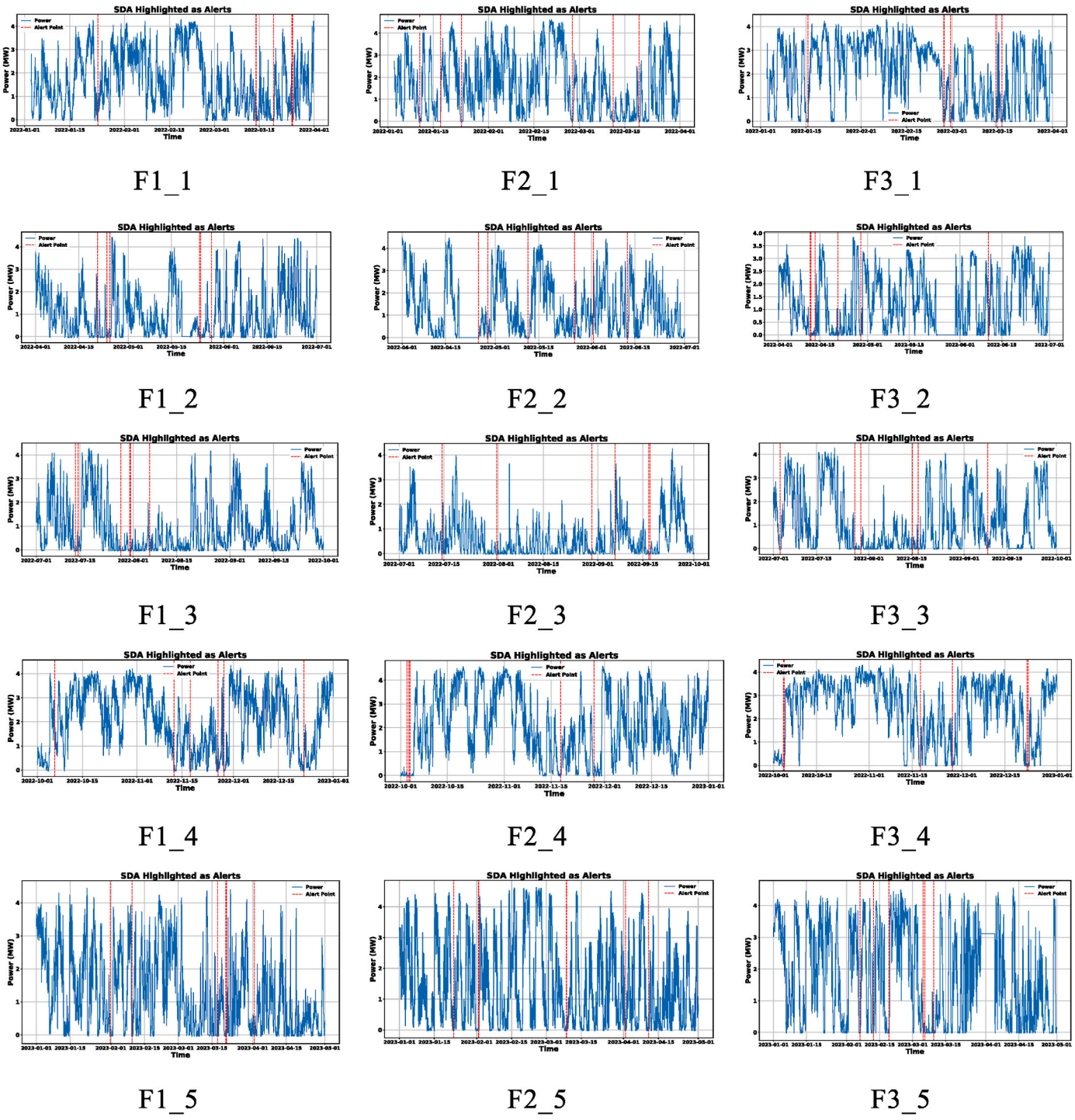


Fig. 13. SDA advanced warning.

speed variations, the seasonal component demonstrates cyclical patterns, the residual component reflects unexplained random fluctuations. This comprehensive decomposition establishes a robust foundation for subsequent analysis, enabling through understanding of wind speed behavior patterns and their temporal variations.

By comparing the decomposed wind speed trend curves with the variations in the original wind power data, the correlation between the changes is validated. The comparison results are shown in Fig. 10.

From a comprehensive perspective, the wind speed patterns (depicted by the orange curve) exhibit a strong correlation with wind power output (illustrated by the blue curve) across most temporal intervals. This correlation is particularly pronounced during peak and trough periods, where wind speed fluctuations accurately reflect corresponding changes in wind power generation. As wind velocities increase, there is a proportional rise in power output; conversely, declining wind speeds result in diminished power generation. This synchronicity demonstrates the fundamental physical relationship between wind velocity and power production, wherein greater wind speeds enable wind turbines to generate progressively higher power outputs.

To provide a more intuitive and accurate assessment of the overlap between wind speed variation curves and wind power output, the overlap statistics for wind speed decomposition curves and wind power

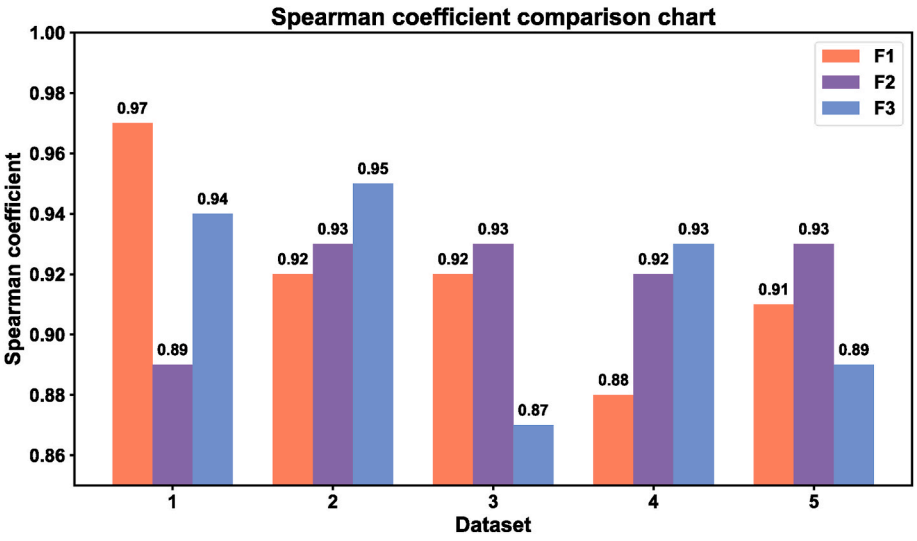


Fig. 14. Spearman coefficient comparison chart.

across different seasons were compiled. The detailed results are presented in Table 3.

From the tabulated results, it can be seen that there is a robust correlation between wind speed patterns and power output throughout most temporal intervals, with particularly pronounced synchronization during peak and trough periods.This synchronicity exemplifies the underlying physical relationship between wind velocity and power generation, where ascending wind speeds consistently correspond with enhanced power output, whereas decreasing velocities result in proportional power reduction. The coherence coefficients across turbines, consistently exceeding 0.86 and predominantly approaching or surpassing 0.90, further substantiate the strong interdependence between wind speed trends and power fluctuations.

3.4. Multi-source data advanced warning

The multi-source early warning system proposed integrates the strengths of Random Forest and SDA, facilitating synchronized anomaly detection through both wind velocity and power output dimensions. Random Forest algorithms model wind-power relationships to monitor long-term trend anomalies, effectively identifying power deviations caused by environmental fluctuations. Simultaneously, SDA conducts direct power output analysis, precisely capturing abrupt ramp events within short intervals, such as those triggered by extreme weather conditions or equipment malfunctions. At the same time, according to the prediction error performance of the spectrum dynamic aggregation Transformer model under different datasets, an organic linkage with the early warning mechanism is established. This dual-alert mechanism, combining external and internal factor analysis, comprehensively enhances the accuracy and reliability of wind power anomaly detection.

3.4.1. RF advanced warning

The previous analysis reveals that variations in wind power are closely correlated with changes in 100 m wind speed. Thus, monitoring wind speed fluctuations can effectively assess the state of wind power variations.

The relationship between wind velocity and power output exhibits nonlinear characteristics, with complex interactions and inherent variability. To effectively address these irregular patterns, Random Forest regression models generate power output predictions with 95 % confidence intervals establishing alert thresholds. The average prediction accuracy is computed based on the historical prediction accuracies of individual wind turbines and serves as a benchmark for dynamic threshold partitioning. When the prediction accuracy is high, the width of the confidence interval is moderately expanded to reduce false alarms arising from normal fluctuations. In contrast, when prediction accuracy is lower and data volatility is higher, the confidence interval is tightened to enhance warning sensitivity and to avoid underreporting of potentially anomalous events. The calibration process results in a new alarm threshold, which is determined using the following formula:

$$\text{AlarmZone} = \text{Base Alarm Zone} \times \alpha \times \left(\left(R^2 - \Delta R^2\right) + 1\right) \tag{39}$$

where α is the adjustment factor, and ΔR^2 represents the average prediction accuracy across all turbines.

Deviations beyond these intervals trigger anomaly alerts. The resultant step-wise power curves effectively capture localized variation trends while minimizing discrepancies. The fitted curves for individual turbines across quarterly intervals demonstrate the seasonal adaptability of this modeling approach, as illustrated in Fig. 11.

Fig. 11 demonstrates the Random Forest power output modeling efficacy, with actual power values predominantly contained within the confidence intervals (depicted by shaded regions). The Random Forest fitting curve (red line) illustrates segmented modeling across wind speed intervals, exhibiting incremental power increases with rising wind velocities rather than continuous smooth progression.

As depicted in the figure, the light red region (95 % confidence interval) represents the default range of predictive uncertainty generated by the model, without accounting for variations in predictive accuracy. The light blue region (calibrated warning intervals) has been dynamically adjusted by modifying the warning thresholds based on prediction accuracy, thereby enhancing the flexibility of the warning intervals. Although the uncalibrated region can cover most actual observations, some data points still fall outside the interval. The calibrated region further refines the warning thresholds, enabling a greater proportion of actual observations to fall within the intervals while minimizing unnecessary warning triggers.

The model's fitting accuracy can be quantified using the Prediction Interval Coverage Probability (PICP), which measures the proportion of actual power output data points falling within a specified confidence interval. A higher proportion of data points within the confidence interval indicates greater model reliability. Table 4 presents the calculated coverage probabilities for different wind speed intervals before and after calibration.

The calibrated threshold intervals differ from the original threshold intervals, suggesting that joint calibration has a significant impact on advanced warning. The PICP values exceed 93 % across all turbines, demonstrating model stability and reliability across diverse operational

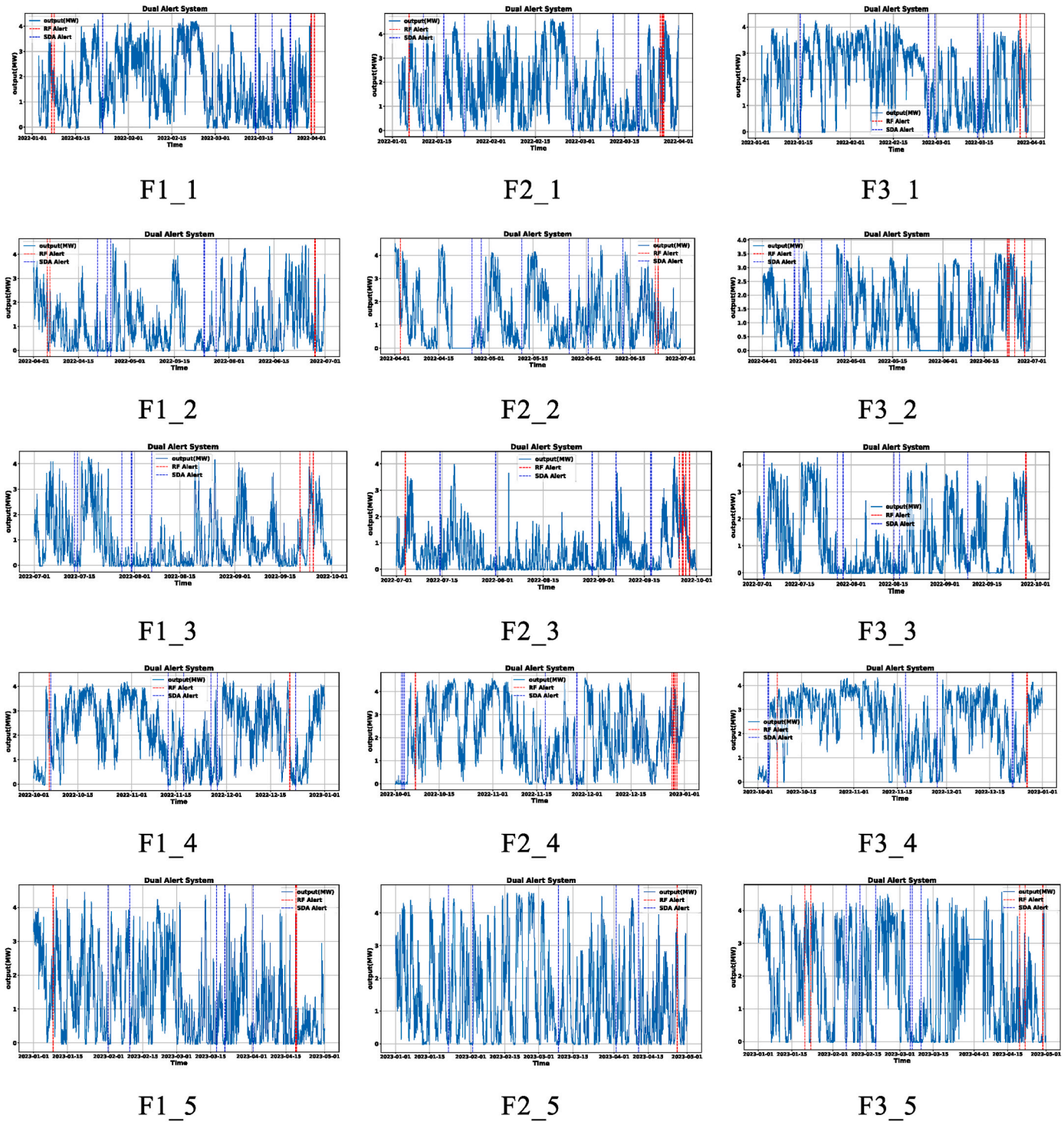


Fig. 15. Multi-source data warning model.

periods. The minimal variation in coverage probabilities among turbines, with PICP values concentrated between 93 % and 94.5 %, indicates consistent fitting performance.

Anomalous events exceeding the confidence interval are flagged for early warning, and periods with a high frequency of anomalies within a specific timeframe are also highlighted, as shown in Fig. 12.

The red vertical lines indicating alert events predominantly coincide with significant power output fluctuations, signifying substantial deviations between actual and fitted values. These alerts may arise from asynchronous relationships between wind speed and power output variations, potentially influenced by generator performance constraints, such as rated power limitations or diminished power growth beyond certain wind speed thresholds. Excursions beyond confidence intervals may also indicate meteorological anomalies (sudden wind shifts or extreme weather events) deviating from modeled baseline conditions. Each alert warrants contextual analysis and specific operational investigation.

3.4.2. *SDA advanced warning*

In wind power forecasting and anomaly detection, ramp events

constitute critical phenomena characterized by rapid, substantial power output fluctuations within brief intervals. These events significantly impact operational management and grid stability, necessitating prompt detection and warning systems. SDA has been implemented to effectively identify and alert these abrupt, high-magnitude power variations within compressed timeframes.

SDA is capable of detecting rapid temporal variations (slope changes) without requiring prior assumptions about data distribution or patterns, demonstrating remarkable adaptability to complex time series data. Prediction accuracy is similarly incorporated as an adjustment factor for threshold width. For turbine operation periods characterized by large prediction errors and high uncertainty, the SDA threshold value is moderately increased to filter out noise interference and prevent the frequent triggering of invalid alarms. In contrast, during phases of enhanced predictive stability, the threshold value is reduced to enable the system to more sensitively capture steep changes in the power curve, thereby improving the identification of abrupt events. The calibration process generates a new threshold warning parameter, which is determined using the following equation:

$$\delta_i = \delta_{base} \times b \times \left(1 + \left(R^2 - \Delta R^2\right)\right) \tag{40}$$

where δ represents the threshold width, and b denotes the adjustment factor. ΔR^2 represents the average of the prediction accuracies for all turbines. For turbines exhibiting low prediction accuracy, the SDA threshold is expanded to filter out noise perturbations, while the warning threshold is reduced to enable more accurate warning capture.

The algorithm monitors local gradient variations with high sensitivity and real-time capability, enabling prompt ramp event detection. Fig. 13 illustrates anomalous events detected by SDA over a temporal sequence.

The SDA alert visualization displays power output variations represented by the blue curve, with red vertical dashed lines marking triggered ramp event instances. Each alert coincides with critical power transition points, signifying significant wind velocity fluctuations' impact on power generation. These abrupt transitions potentially indicate equipment stress or grid stability risks, warranting heightened operational vigilance.

To provide a more comprehensive evaluation, we incorporated the Spearman correlation coefficient to quantify the temporal similarity of warning events. In this study, we systematically analyzed the timing of warning events obtained using the RF and SDA methods. Specifically, we calculated the time difference between the SDA-derived warnings and the RF-derived warnings and identified the optimal temporal alignment between the two systems based on the Spearman correlation coefficient. This approach allowed us to assess the temporal consistency between the RF warning system and the SDA warning system. Fig. 14 presents a quantitative visualization of the overlap between the two warning times, as assessed using the Spearman system.

The evaluation of Spearman's rank correlation coefficients indicates that the consistency across different turbines exhibits a high degree of correlation, with correlation coefficients exceeding 0.87 for all turbines. This demonstrates that the warning times of the two warning systems, namely RF and SDA, exhibit a good match and high consistency across different turbines. Higher consistency coefficients indicate that the temporal difference between predictions of the two advanced warning systems is minimal and their correlation is strong, thereby supporting the effective synergy of the two systems within the wind turbine advanced warning framework.

3.4.3. Multi-Source Data Warning Model

The integration of dual detection mechanisms on the power output curve, as illustrated in Fig. 15, demonstrates that singular model implementation proves insufficient for comprehensive anomaly detection. The synthesized approach combines Random Forest alerts with SDA detection events, providing exhaustive coverage of power output anomalies. This consolidated visualization superimposes both alert types onto the power generation curve, enabling comprehensive monitoring of distinct anomaly patterns.

As shown in Fig. 15, Random Forest alerts (red dashed lines) highlight deviations between actual and predicted power outputs, triggering when values exceed confidence intervals and primarily detecting anomalies in long-term trends such as equipment malfunctions or turbine performance degradation. SDA alerts (blue dashed lines) focus on identifying severe gradient changes, specifically targeting rapid power fluctuations within brief intervals, particularly effective for capturing wind-speed-induced ramp events. Regions where both RF and SDA alerts are triggered simultaneously (dual alerts) indicate critically severe anomalies or equipment instability that warrant immediate attention.

In the end, the integrated alert system, combining Random Forest modeling and SDA, enables comprehensive anomaly monitoring in wind power generation. Random Forest detects deviations from predicted values, specializing in long-term power anomaly surveillance, SDA identifies rapid power fluctuations, particularly ramp events. Meanwhile, the aforementioned research design fully encompasses the core content of analogous ablation experiments and substantiates the substantial improvement of the dual warning mechanism over a single mechanism in terms of warning accuracy and robustness from the perspective of multi-source data fusion.

4. Conclusion

The Transformer prediction model incorporating spectral dynamic aggregation and rotary gating mechanisms is capable of addressing limitations of traditional forecasting models, which typically capture general power output trends while lacking sensitivity to abrupt changes. The proposed enhanced modeling framework allows comprehensive monitoring capabilities of simultaneously tracking power fluctuation patterns while enabling rapid response to sudden events through integrated surveillance mechanisms.

The proposed Transformer model has been demonstrated superior across MAE, RMSE, MAPE, and R^2 metrics when benchmarked against BP, LSTM, Bi-LSTM, ELM, SVR, RBF and GRU model architectures. Incorporating spectral blocks specifically designed for oscillatory offshore wind data characteristics, the enhanced Transformer framework mitigates analytical instabilities inherent in conventional models. The integration of dynamic aggregation modules enables the detections of rapid mutation while effectively minimizing latency phenomena.

Analysis of cross-correlation information metrics across various offshore turbines reveals strong correlation between the values of 100 m wind velocity and power output. Furthermore, seasonal trend-residual decomposition of 100 m wind speeds demonstrates high correlation between decomposed velocity trends and power output patterns. Consequently, 100 m wind velocity has been identified as the critical parameter for the power output alert system.

Thirdly, in the study of a multi-source data warning system based on a fitted swing-gate algorithm, the first tier of warnings involves analyzing various trend curves obtained from the decomposition of 100 m wind speed data. The random forest method was employed to fit the wind power curve, and the accuracy index of the joint prediction model was calibrated to determine the warning threshold range. The RF alert is triggered when the spectral dynamics aggregation Transformer model predicts a wind power value exceeding the specified range. The second layer of warnings employs a swing-gate algorithm to assess whether the predicted wind power indicates slope events, while simultaneously co-calibrating the threshold. If a slope event is detected, a slope alert is triggered. The results of the study demonstrate that the multi-source data early warning system synergistically identifies anomalies from both wind speed and wind power perspectives. By integrating a dual-warning mechanism that accounts for both external and internal factors, it provides timely warning information, thereby effectively supporting the objective of achieving a flexible and reliable energy supply.

5. Discussion

The wind power prediction model proposed in this study, which integrates multi-source data fusion, spectral blocks, and a dynamic aggregation mechanism, exhibits excellent prediction performance under routine operating conditions. Furthermore, by incorporating random forest and SDA algorithms, the model establishes a robust advanced warning mechanism. However, under extreme weather or low wind speed conditions, where wind power output tends toward zero, the model's performance is constrained by the physical operating mechanisms of the wind turbine and is not attributable solely to modeling errors. There remains a scarcity of sufficient field data for conducting a comprehensive evaluation of model performance across various extreme meteorological environments. The adaptability of the model under extreme conditions is anticipated to further improve with subsequent data accumulation.

One of the primary objectives of the "prediction–early warning" linkage mechanism proposed in this study is to enhance the operational reliability of wind power systems. A comprehensive management framework can be conceptualized as comprising three sequential components: "prediction," "early warning," and "reliability." The high-precision spectral dynamic aggregation Transformer model developed in this work provides an accurate forecast of future wind power output, which directly and significantly influences the outcomes of reliability assessments.

Furthermore, the dual early warning mechanism—integrating Random Forest for detecting trend anomalies and SDA for identifying ramp events—enables timely detection of deviations from normal operational behavior. By converting the outputs of the "prediction–early warning" process into dynamic and quantifiable reliability metrics, the framework facilitates both the identification of system vulnerabilities and the enhancement of overall system reliability. This enables a fully integrated closed-loop workflow, progressing from "state awareness " to "risk identification," and ultimately to "decision support."

In addition, the issue of data fusion and alignment among multiple wind turbines plays a crucial role in power planning at the wind farm level. This study focuses on the independent modeling and warning analysis of three wind turbines, with the aim of verifying the generalization capability of the proposed spectral dynamic aggregation model across different units. If it is extended to power planning scenarios, further exploration of effective pathways for collaborative modeling among multiple units will be required.

To address these limitations, future research may prioritize the following directions:

(1) Integration of physical mechanisms and data-driven modeling: By incorporating turbine operational parameters, leveraging zonal modeling, and employing integrated learning strategies, this approach aims to enhance the predictive stability and risk identification capability of the model in extreme wind speed intervals.
(2) Enhanced generalization capability for extreme conditions: Extending the training set by leveraging physical priors and synthetic data, Optimize the adaptability and robustness of the model for atypical wind speed-power relationships.
(3) Constructing a unified spatiotemporal framework and optimizing a collaborative Multi-Turbine modeling structure: A unified spatiotemporal coordinate system is established while preserving the individual differences among wind turbines, and data alignment is accomplished through sliding-window, interpolation, and normalization strategies. Simultaneously, a turbine embedding vector is introduced to design a shared-parameter structure, which captures individual operational characteristics while extracting group patterns across units, thereby supporting the regional-level wind power prediction task.
(4) Development of a collaborative Multi-Wind turbine advanced warning and risk linkage mechanism: Introduction of cross-wind turbine event resonance discrimination rules, establishment of a group anomaly detection system, and realization of collaborative advanced warning and risk response among multiple units, thereby enhancing the dispatching efficiency of wind farms and ensuring the safe operation of the power system.

CRediT authorship contribution statement

Hao Xu: Writing – original draft, Visualization, Software, Methodology. **Zhenhao Zhu:** Writing – review & editing, Visualization, Resources. **Hongbing Liu:** Supervision, Funding acquisition. **Enrico Zio:** Writing – review & editing, Investigation. **Xiaolong Qiu:** Validation. **Yuchen Lu:** Formal analysis. **Xianqiang Qu:** Investigation.

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.

Acknowledgment

This research is funded by the Shandong Provincial Natural Science Foundation (ZR2022QE091), Key R&D Program of Shandong Province, China (2023CXGC010407), the Special fund for Taishan Industry Leading Talent Project (tsls20230605).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.energy.2025.139396.

Data availability

Data will be made available on request.

References

[1] Li L L, Chang Y B, Tseng M L, et al. Wind power prediction using a novel model on wavelet decomposition-support vector machines-improved atomic search algorithm. J Clean Prod 2020;270:121817.
[2] Council GWE. GWEC| global wind report 2021. 2021.
[3] Li Y, Su Y, Xia L, et al. Reliability evaluation of wind power systems by integrating granularity-related Latin hypercube sampling with LSTM-based prediction. Comput Ind 2025;173:104365.
[4] Al-Muhaini M, Bizrah A, Heydt G, et al. Impact of wind speed modelling on the predictive reliability assessment of wind-based microgrids. IET Renew Power Gener 2019;13(15):2947–56.
[5] Hu H, Wang L, Tao R. Wind speed forecasting based on variational mode decomposition and improved echo state network. Renew Energy 2021;164:729–51.
[6] Shao H, Henriques R, Morais H, et al. Power quality monitoring in electric grid integrating offshore wind energy: a review. Renew Sustain Energy Rev 2024;191:114094.
[7] Jiang P, Wang Y, Wang J. Short-term wind speed forecasting using a hybrid model. Energy 2017;119:561–77.
[8] Peng X, Wang H, Lang J, et al. EALSTM-QR: interval wind-power prediction model based on numerical weather prediction and deep learning. Energy 2021;220:119692.
[9] Zhang Y, Li Y, Zhang G. Short-term wind power forecasting approach based on Seq2Seq model using NWP data. Energy 2020;213:118371.
[10] Tang Y, Yang K, Zhang S, et al. Wind power forecasting: a hybrid forecasting model and multi-task learning-based framework. Energy 2023;278:127864.
[11] Cadenas E, Jaramillo O A, Rivera W. Analysis and forecasting of wind velocity in chetumal, quintana roo, using the single exponential smoothing method. Renew Energy 2010;35(5):925–30.
[12] Cunha J L R N, Pereira C M N A. A hybrid model based on STL with simple exponential smoothing and ARMA for wind forecast in a Brazilian nuclear power plant site. Nucl Eng Des 2024;421:113026.
[13] Monjazeb M R, Amiri H, Movahedi A. Wholesale electricity price forecasting by quantile regression and Kalman filter method. Energy 2024;290:129925.
[14] Liu X, Lin Z, Feng Z. Short-term offshore wind speed forecast by seasonal ARIMA-A comparison against GRU and LSTM. Energy 2021;227:120492.

[15] Naik J, Dash P K, Dhar S. A multi-objective wind speed and wind power prediction interval forecasting using variational modes decomposition based Multi-kernel robust ridge regression. Renew Energy 2019;136:701–31.
[16] Wang Y, Zou R, Liu F, et al. A review of wind speed and wind power forecasting with deep neural networks. Appl Energy 2021;304:117766.
[17] Li L L, Zhao X, Tseng M L, et al. Short-term wind power forecasting based on support vector machine with improved dragonfly algorithm. J Clean Prod 2020; 242:118447.
[18] Jørgensen K L, Shaker H R. Wind power forecasting using machine learning: state of the art, trends and challenges[C]. In: 2020 IEEE 8th international conference on smart energy grid engineering (SEGE). IEEE; 2020. p. 44–50.
[19] Ramos D, Faria P, Morais A, et al. Using decision tree to select forecasting algorithms in distinct electricity consumption context of an office building. Energy Rep 2022;8:417–22.
[20] Ak R, Vitelli V, Zio E. An interval-valued neural network approach for uncertainty quantification in short-term wind speed prediction. IEEE Transact Neural Networks Learn Syst 2015;26(11):2787–800.
[21] Aly H H H. A hybrid optimized model of adaptive neuro-fuzzy inference system, recurrent Kalman filter and neuro-wavelet for wind power forecasting driven by DFIG. Energy 2022;239:122367.
[22] Yang B, Zhong L, Wang J, et al. State-of-the-art one-stop handbook on wind forecasting technologies: an overview of classifications, methodologies, and analysis. J Clean Prod 2021;283:124628.
[23] Chen Z, Zhou D, Zio E, et al. A deep learning feature fusion based health index construction method for prognostics using multiobjective optimization. IEEE Trans Reliab 2022;72(3):1038–52.
[24] Shahid F, Zameer A, Muneeb M. A novel genetic LSTM model for wind power forecast. Energy 2021;223:120069.
[25] Hao Z, Di Maio F, Zio E. Monte carlo tree search-based deep reinforcement learning for flexible operation & maintenance optimization of a nuclear power plant. J Saf and Sustain 2024;1(1):4–13.
[26] Neshat M, Nezhad M M, Abbasnejad E, et al. A deep learning-based evolutionary model for short-term wind speed forecasting: a case study of the Lillgrund offshore wind farm. Energy Convers Manag 2021;236:114002.
[27] Zhao Y, Ye L, Li Z, et al. A novel bidirectional mechanism based on time series model for wind power forecasting. Appl Energy 2016;177:793–803.
[28] González-Sopeña J M, Pakrashi V, Ghosh B. An overview of performance evaluation metrics for short-term statistical wind power forecasting. Renew Sustain Energy Rev 2021;138:110515.
[29] Bentsen L Ø, Warakagoda N D, Stenbro R, et al. Spatio-temporal wind speed forecasting using graph networks and novel transformer architectures. Appl Energy 2023;333:120565.
[30] Li S, Jin X, Xuan Y, et al. Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting. Adv Neural Inf Process Syst 2019;32.
[31] Zhou H, Zhang S, Peng J, et al. Informer: beyond efficient transformer for long sequence time-series forecasting[C]. In: Proceedings of the AAAI conference on artificial intelligence, vol. 35; 2021. p. 11106–15. 12.
[32] Kitaev N, Kaiser Ł, Levskaya A. Reformer: the efficient transformer [C]. In: International conference on learning representations; 2020.
[33] Wu H, Xu J, Wang J, et al. Autoformer: decomposition transformers with auto-correlation for long-term series forecasting. Adv Neural Inf Process Syst 2021;34: 22419–30.
[34] Liu S, Yu H, Liao C, et al. Pyraformer: low-complexity pyramidal attention for long-range time series modeling and forecasting [C]. International conference on learning representations. 2022.
[35] Xiao Y, Wu S, He C, et al. An effective hybrid wind power forecasting model based on" decomposition-reconstruction-ensemble" strategy and wind resource matching. Sustain Energy, Grids and Networks 2024;38:101293.
[36] Valcamonico D, Baraldi P, Amigoni F, et al. A framework based on natural language processing and machine learning for the classification of the severity of road accidents from reports. Proc Inst Mech Eng O J Risk Reliab 2024;238(5): 957–71.
[37] Niu Z, Yu Z, Tang W, et al. Wind power forecasting using attention-based gated recurrent unit network. Energy 2020;196:117081.
[38] Zhu Z, Zheng Q, Liu H, et al. Prediction model for pipeline pitting corrosion based on multiple feature selection and residual correction. J Mar Sci Appl 2025;24(4): 805–15.
[39] Zhang S, Chen Y, Xiao J, et al. Hybrid wind speed forecasting model based on multivariate data secondary decomposition approach and deep learning algorithm with attention mechanism. Renew Energy 2021;174:688–704.
[40] Shao Z, Han J, Zhao W, et al. Hybrid model for short-term wind power forecasting based on singular spectrum analysis and a temporal convolutional attention network with an adaptive receptive field. Energy Convers Manag 2022;269: 116138.
[41] Chen J, Fu X, Zhang L, et al. A novel offshore wind power prediction model based on TCN-DANet-sparse transformer and considering spatio-temporal coupling in multiple wind farms. Energy 2024;308:132899.
[42] Dong Z, Zhao Y, Wang A, et al. Wind-mambaformer: ultra-short-term wind turbine power forecasting based on advanced transformer and mamba models. Energies 2025;18(5).
[43] Wang S, Shi J, Yang W, et al. High and low frequency wind power prediction based on transformer and BiGRU-Attention. Energy 2024;288:129753.
[44] Zhang Z, Sun Z, Guo X, et al. Short-term multi-step wind power prediction model based on Pt-Transformer neural network integrating spatio-temporal feature and sparse attention. Elec Power Syst Res 2025;248:111970.
[45] Li W, Yang T, Yang J, et al. Predicting ultra-short-term wind power combinations under extreme weather conditions. IEEE Access 2025;13:26575.
[46] Yu G Z, Lu L, Tang B, et al. Ultra-short-term wind power subsection forecasting method based on extreme weather. IEEE Trans Power Syst 2022;38(6):5045–56.
[47] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. Adv Neural Inf Process Syst 2017:30.
[48] Amiri F, Yousefi M M R, Lucas C, et al. Mutual information-based feature selection for intrusion detection systems. J Netw Comput Appl 2011;34(4):1184–99.
[49] Chang K, Kim S, Chang D, et al. Uncertainty analysis for target SIL determination in the offshore industry. J Loss Prev Process Ind 2015;34:151–62.
[50] Trull O, García-Díaz J C, Peiró-Signes A. Multiple seasonal STL decomposition with discrete-interval moving seasonalities. Appl Math Comput 2022;433:127398.
[51] Hu S, Wang Y, Cai W, et al. A combined method for short-term load forecasting considering the characteristics of components of seasonal and trend decomposition using local regression. Appl Sci 2024;14(6):2286.
[52] Ak R, Fink O, Zio E. Two machine learning approaches for short-term wind speed time-series prediction. IEEE Transact Neural Networks Learn Syst 2015;27(8): 1734–47.
[53] Zheng Q, Wang C, Zhu Z. Research on the prediction of mine water inrush disasters based on multi-factor spatial game reconstruction. Geomechanics and Geophysics for Geo-Energy and Geo-Resources 2024;10(1):41.
[54] Liu H, Zhu Z, Zhang J, et al. Submarine pipeline corrosion rate prediction model based on high-dimensional mapping augmentation and residual update gradient forest. Appl Ocean Res 2025;155:104432.