Physics-guided multiscale fusion network for ultra-short-term wind power forecasting

Citation for the original published paper (version of record):

Guo, F., Tao, S., Zhang, R. et al (2026). Physics-guided multiscale fusion network for ultra-short-term wind power forecasting. Computers and Electrical Engineering, 139.
http://dx.doi.org/10.1016/j.compeleceng.2026.111367

(article starts on next page)

Computers and Electrical Engineering 139 (2026) 111367

Contents lists available at ScienceDirect

Computers and Electrical Engineering

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

Physics-guided multiscale fusion network for ultra-short-term wind power forecasting ☆

Fusen Guo [a,b], Shengyu Tao [c,b,*], Rui Zhang [d], Hailing Zhou [e], Jun Zhang [a], Huadong Mo [b,**]

[a] School of Science, Computing and Emerging Technologies, Swinburne University of Technology, Hawthorn, VIC 3122, Australia
[b] School of Systems and Computing, University of New South Wales (UNSW), Canberra, ACT 2600, Australia
[c] Department of Electrical Engineering, Chalmers University of Technology, 41296 Gothenburg, Sweden
[d] School of Electrical Engineering and Telecommunications, University of New South Wales (UNSW), Sydney, NSW 2033, Australia
[e] School of Engineering, Swinburne University of Technology, Hawthorn, VIC 3122, Australia

ARTICLE INFO

Dataset link: https://doi.org/10.5281/zenodo.5516543

Keywords:
Wind power forecasting
Multiscale modeling
Fourier transform
Physics-guided learning
Wake effect

ABSTRACT

With the increasing penetration of wind power in modern power systems, ultra-short-term wind power forecasting plays an important role in maintaining grid security and operational stability. However, wind power sequences are characterized by strong nonlinearity and nonstationarity, and are further affected by turbine wake coupling, making it difficult for purely data-driven methods to simultaneously achieve high forecasting accuracy and physical consistency. To address this issue, this paper proposes a Physics-Guided Multiscale Fusion Network (PGMFNet) for ultra-short-term wind power forecasting. Specifically, the proposed model employs a State-Space Model (SSM) to capture long-term dependencies and periodic evolutionary patterns in wind power sequences, thereby enhancing the modeling capability for global temporal dynamics. In addition, a Multi-Scale Fourier Transform Network (MSFNet) is constructed to improve the representation of local dynamics, short-term variation patterns, and abrupt fluctuations through frequency-domain encoding under multiple temporal resolutions. A Channel–Spatial Attention Module (CBAM) is further introduced to adaptively emphasize critical variables and informative time steps. Finally, a physics-constrained loss derived from Jensen's wake theory is incorporated to enforce consistency between the forecasting results and turbine aerodynamic behavior. Experimental results on real-world inland and offshore wind farm datasets in the United States show that the proposed model achieves the best overall performance in terms of the coefficient of determination (R^2), root mean squared error (RMSE), mean absolute error (MAE), and mean absolute scaled error (MASE) among the compared models. The results further indicate that the proposed framework can improve forecasting robustness and encourage wake-consistent prediction behavior under different wind farm conditions. Overall, PGMFNet provides an effective framework for accurate and physically consistent real-time wind power forecasting, offering practical support for intelligent wind farm operation and power grid dispatch.

☆ This project is partially supported by the ARC Research Hub for Resilient and Intelligent Infrastructure Systems (IH210100048), the Trailblazer for Recycling and Clean Energy under the project *Delivering net-zero energy buildings: A next-generation hybrid energy system with demand-side management* and the Australian Government through the ARC. Dr. Rui Zhang is supported by an ARC Discovery Early Career Researcher Award (DE220101277).

* Corresponding author at: Department of Electrical Engineering, Chalmers University of Technology, 41296 Gothenburg, Sweden.
** Corresponding author.
E-mail addresses: shengyu.tao@chalmers.se (S. Tao), huadong.mo@unsw.edu.au (H. Mo).

https://doi.org/10.1016/j.compeleceng.2026.111367
Received 31 March 2026; Received in revised form 28 June 2026; Accepted 29 June 2026
Available online 10 July 2026

1. Introduction

Driven by the global energy transition and Australia's commitment to achieving net-zero emissions by 2050, wind energy has experienced rapid development as a clean, renewable, and sustainable energy resource [1]. Owing to its abundant reserves, wide availability, technological maturity, and environmental friendliness, wind power has become an important component in optimizing the national energy structure and supporting the transition toward low-carbon power systems [2]. However, the inherently intermittent, stochastic, and fluctuating nature of wind energy poses significant challenges to wind farm operation, power dispatch, and grid stability. In particular, the increasing penetration of wind power introduces greater uncertainty into power system scheduling and real-time balancing. Therefore, achieving reliable and stable integration of high-penetration wind energy into the power grid has become a key issue in both academic research and industrial applications [3,4].

Wind power forecasting has been widely recognized as an effective approach to mitigating these challenges. According to the prediction horizon, wind power forecasting can generally be categorized into four types: ultra-short-term forecasting, ranging from seconds to 4 h [5]; short-term forecasting, ranging from 4 to 24 h; medium-term forecasting, ranging from 1 to 7 days; and long-term forecasting, extending beyond 7 days [6]. Among these categories, ultra-short-term wind power forecasting is particularly important for frequency regulation, spinning reserve allocation, real-time turbine control, and power system dispatch. Improving the accuracy and robustness of ultra-short-term forecasting can reduce energy losses, enhance economic dispatch, and strengthen the operational flexibility and reliability of power systems.

Existing wind power forecasting methods can be broadly divided into statistical models, machine learning models, and hybrid models. Statistical approaches, such as Autoregressive Integrated Moving Average (ARIMA) [7] and Generalized Linear Models (GLM) [8], establish mathematical relationships based on historical time-series data and offer advantages in simplicity, computational efficiency, and interpretability. However, these methods often rely on linear assumptions and may perform inadequately when dealing with nonlinear, non-stationary, and highly fluctuating wind power signals. Machine learning and deep learning models, including Long Short-Term Memory (LSTM) networks [9], Convolutional Neural Networks (CNNs), and Transformer-based architectures, have demonstrated strong capability in capturing nonlinear temporal dependencies and complex feature interactions. Nevertheless, purely data-driven models usually neglect the physical mechanisms underlying wind power generation, which may limit their generalization ability and interpretability under complex operating conditions.

In recent years, hybrid forecasting models have attracted increasing attention by integrating multisource meteorological variables, attention mechanisms, frequency-domain representations, spatial dependency modeling, and physical prior knowledge [10]. Transformer-based models have been widely adopted to capture long-range temporal dependencies in wind power sequences [11]. Some studies combine Transformers with recurrent neural networks, such as LSTM, to improve contextual sequence modeling, while others redesign the input structure to better represent multivariate dependencies [12]. Meanwhile, spatio-temporal forecasting methods based on graph neural networks (GNNs) and Transformer–GNN hybrids have been developed to model turbine topology, spatial correlations, and wake-related dependencies within wind farms. Beyond basic graph convolutional networks (GCNs), variants such as graph attention networks (GATs), spatio-temporal graph convolutional networks (STGCNs), and dynamic graph-based models provide effective tools for representing non-Euclidean spatial coupling and complex wake propagation across multiple turbines. These methods are particularly suitable for large-scale wind farms with rich spatial interactions [13,14].

In addition, frequency-domain and decomposition-based methods provide another effective way to represent periodic components, short-term fluctuations, and multiscale variations in wind power data [15]. For example, Jiang and Liu [16] proposed an EEMD-PSO-LSTM model, where ensemble empirical mode decomposition was used to decompose the original wind power sequence into multiple subseries, and particle swarm optimization was adopted to tune the hyperparameters of LSTM for each decomposed component. Chen et al. [17] further developed an ultra-short-term wind power prediction method based on quadratic variational mode decomposition and multi-model fusion, in which optimized VMD was used to decompose high-frequency non-stationary components and different prediction models were assigned to different subcomponents. These studies indicate that decomposition-assisted forecasting can improve prediction accuracy by reducing the non-stationarity, randomness, and complexity of wind power series. Compared with purely time-domain models, such approaches are often more effective in characterizing oscillatory patterns and local perturbations in highly non-stationary wind power sequences. However, most decomposition-based models still rely on explicit signal decomposition and separated prediction modules, and they do not jointly consider multiresolution spectral representation, attention-based feature refinement, and wake-related physical consistency within a unified forecasting framework.

Wake-aware and physics-guided forecasting methods have also been investigated to improve the physical consistency of wind farm-level prediction. Liu et al. [6] proposed a joint wake–power prediction method by combining large eddy simulation and convolutional neural networks. Zhu et al. [14] incorporated wake correction mechanisms into a non-stationary Transformer framework for large-scale wind farm forecasting. Giraldo Echeverri et al. [5] introduced turbine geometric relationships and wind direction information into a spatio-temporal Transformer model to characterize multi-turbine wake effects. Liu et al. [18] developed a spatio-temporal physics-guided graph neural network, where wake coupling was represented through directed local subgraphs and physics-induced attention. Romero-Barrera et al. [15] further combined analytical wake models with a variational mode decomposition-enhanced deep learning architecture. These studies indicate that wake-related physical information can provide useful guidance for improving the plausibility, robustness, and interpretability of wind power forecasting models. However, most existing studies mainly focus on wake-aware correction, spatial dependency modeling, or physical-prior enhancement separately, while a unified framework that jointly accounts for long-range temporal evolution, multiresolution frequency-domain perturbations, and physics-consistent regularization remains insufficiently explored.

Despite these advances, several limitations still remain. First, in many existing physics-guided forecasting models, physical knowledge is incorporated mainly through auxiliary loss functions, posterior correction strategies, or weak regularization terms, which may provide insufficient constraints on wake-induced turbine dynamics [14]. Second, most existing models are primarily constructed in the time domain and therefore struggle to jointly capture long-term evolutionary patterns, short-term abrupt fluctuations, and periodic components in wind power sequences [15]. Third, wind power generation is highly sensitive to instantaneous atmospheric conditions, complex terrain, meteorological variations, and wake interactions among turbines, leading to pronounced temporal variability and multiscale fluctuations [19]. However, a unified framework that effectively integrates multiscale feature extraction, key feature enhancement, and physical consistency constraints remains insufficiently explored [20]. These limitations motivate the development of a forecasting model that can jointly exploit temporal dynamics, frequency-domain characteristics, and wake-related physical knowledge.

To address the above issues, this paper proposes a Physics-Guided Multiscale Fusion Network (PGMFNet) for ultra-short-term wind power forecasting. The proposed framework consists of an embedding layer, a state-space modeling (SSM) module, a Multi-Scale Fourier Transform Network (MSFNet), a channel–spatial attention module (CBAM), a fully connected prediction layer, and a physics-constrained module based on Jensen's wake theory. Specifically, the SSM module is employed to capture long-range temporal dependencies and global sequence evolution in wind power data. The MSFNet is designed to extract periodic, multiscale, and high-frequency fluctuation patterns through Fourier-domain encoding under multiple temporal resolutions. The CBAM module adaptively enhances critical variables and informative time steps by integrating channel attention and spatial attention. Furthermore, the Jensen-based physical constraint is incorporated into the learning framework to improve the consistency between forecasting results and turbine wake dynamics, thereby enhancing the physical plausibility and interpretability of the model. Through the coordinated design of these components, PGMFNet achieves complementary fusion across temporal, frequency, attention-enhanced, and physical-prior domains, enabling more robust and physically consistent wind power forecasting.

The main contributions of this paper are summarized as follows:

1. A novel Physics-Guided Multiscale Fusion Network, termed PGMFNet, is proposed for ultra-short-term wind power forecasting. The framework integrates data-driven temporal representation learning with wake-related physical prior knowledge, thereby improving forecasting robustness, physical plausibility, and interpretability.
2. A Multi-Scale Fourier Transform-based neural modeling mechanism, namely MSFNet, is introduced to enhance frequency-domain feature extraction. By replacing conventional query–key–value attention with Fourier-domain operations, MSFNet provides an efficient way to model periodic components, trend information, and multiscale fluctuations in wind power sequences.
3. A CBAM-based channel–spatial attention module is developed to strengthen key feature identification. By combining channel attention and spatial attention, the proposed module dynamically assigns attention weights to critical input variables and informative time steps, thereby improving the representation of local abrupt variations and important temporal patterns.
4. A physics-based constraint derived from the one-dimensional Jensen wake model is incorporated into the learning framework. A dynamic weighting schedule is employed to balance predictive accuracy and physical consistency, which improves the generalization ability and interpretability of the proposed forecasting model.

The remainder of this paper is organized as follows. Section 2 presents the detailed methodology of the proposed approach. Section 3 describes the experimental settings and evaluation metrics. Section 4 provides comprehensive comparison and ablation experiments. Finally, Section 5 summarizes the main findings and discusses potential directions for future research.

2. Methods

2.1. State-Space Modeling (SSM)

To effectively capture the non-stationary and temporally dependent characteristics inherent in wind power sequences, this work incorporates an SSM module into the PGMFNet framework. The SSM provides a compact and mathematically principled approach for modeling latent temporal dynamics, making it particularly suitable for learning both short-term fluctuations and long-range dependencies in time series data.

In discrete form, the SSM can be expressed as:

$$
\begin{aligned}
h_t &= Ah_{t-1} + Bx_t \\
y_t &= Ch_t + Dx_t
\end{aligned}
\tag{1}
$$

where $x_t \in \mathbb{R}^d$ is the input at time step t, $h_t \in \mathbb{R}^{d_h}$ is the latent hidden state, and y_t is the output. The matrices A, B, C, and D are trainable parameters that govern state transitions and output projection.

To enhance efficiency and scalability in long sequence modeling, we adopt a convolutional parameterization of the transition functions, following recent advances in structured state-space sequence models. This design enables the model to maintain stable gradients while capturing both local variations and long-range temporal structures.

By embedding the SSM module within the prediction pipeline, PGMFNet gains the ability to model wind power fluctuations under changing atmospheric and operational conditions, without relying on explicit recurrent mechanisms. The resulting structure serves as a lightweight temporal modeling component within the proposed forecasting framework.

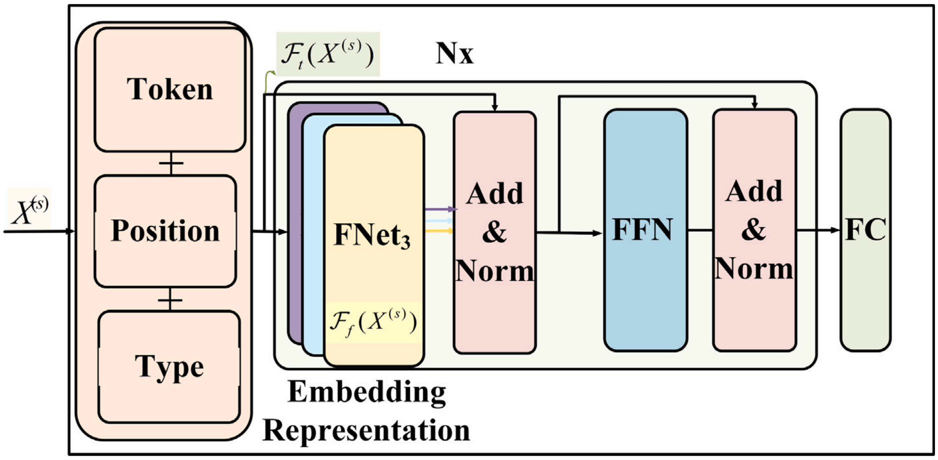

Fig. 1. Overall architecture of the MSFNet.

Algorithm 1 Forward Procedure of the Proposed MSFNet

Require: $X \in \mathbb{R}^{B\times L\times D}$; $\mathcal{S} = \{1, 2, 4\}$
Ensure: X_{out}
for all $s \in \mathcal{S}$ **do**
  $X^{(s)} \leftarrow \text{AvgPool1D}(X,\ \text{stride} = s)$
end for
for all $s \in \mathcal{S}$ **do**
  $X_t^{(s)} \leftarrow \mathcal{F}_t(X^{(s)})$
  $X_f^{(s)} \leftarrow \mathcal{F}_f(X^{(s)})$
  $X_{\text{freq}}^{(s)} \leftarrow \phi^{(s)}\left(\text{Concat}\left[X_t^{(s)}, X_f^{(s)}\right]\right)$
  $\widetilde{X}^{(s)} \leftarrow \mathcal{F}^{-1}\left(X_{\text{freq}}^{(s)}\right)$
end for
for all $s \in \mathcal{S}$ **do**
  $\hat{X}^{(s)} \leftarrow \text{Upsample}(\widetilde{X}^{(s)}, L)$
end for
$X_{\text{msf}} \leftarrow \text{Concat}[\ \hat{X}^{(1)}, \hat{X}^{(2)}, \hat{X}^{(4)}\ ]$
$X_{\text{out}} \leftarrow W_f X_{\text{msf}}$
return X_{out}

2.2. Multi-Scale Fourier Transform Network (MSFNet)

To enhance the model's capability of perceiving local dynamics, short-term variation patterns, and abrupt fluctuations in wind power sequences, we propose a Multi-Scale Fourier Transform Network (MSFNet). Built upon the standard FNet, the proposed MSFNet constructs parallel FNet-based encoding branches under multiple temporal resolutions and achieves multiscale feature fusion through temporal alignment followed by linear projection. Different from TimesNet, which relies on explicit periodic reshaping, and FRETS, which emphasizes frequency reconstruction, the proposed MSFNet focuses on hierarchical perception of local spectral perturbations through parallel Fourier encoding and multiresolution feature aggregation. The architecture is illustrated in Fig. 1, and the pseudocode is provided in Algorithm 1.

2.2.1. Motivation and multiscale design

Conventional FNet applies Fourier Transform at a single temporal resolution and mainly captures dominant frequency patterns. However, wind power sequences usually contain rapid short-term fluctuations, as well as local abrupt variations induced by gusts, turbulence, and wake interactions. A single-resolution frequency encoder is therefore insufficient to fully characterize such multiscale dynamics.

Given the input sequence $X \in \mathbb{R}^{B\times L\times D}$, where B, L, and D denote the batch size, sequence length, and feature dimension, respectively, we define a scale set $\mathcal{S} = \{1, 2, 4\}$. Under the current sampling setting, these three scales correspond to temporal resolutions of 15 min, 30 min, and 1 h, respectively. For each scale $s \in \mathcal{S}$, the input sequence is first downsampled by average pooling to obtain the representation at the corresponding temporal granularity:

$$X^{(s)} = \text{AvgPool1D}(X, \text{stride} = s), \quad s \in \mathcal{S}. \tag{2}$$

Here, $s = 1$ preserves fine-grained local variations at the original temporal resolution, while $s = 2$ and $s = 4$ progressively enlarge the temporal receptive field and smooth local noise. In this way, different branches are able to perceive local dynamics, short-term variation patterns, and abrupt fluctuations under different temporal granularities. Such a coarse-to-fine temporal hierarchy also provides a practical trade-off between multiscale coverage and computational complexity.

2.2.2. Frequency-domain encoding and reconstruction

For each scale s, the scaled input $X^{(s)}$ is processed by an independent FNet-based branch. The temporal Fourier representation and the feature-wise Fourier representation are first extracted separately:

$$X_t^{(s)} = \mathcal{F}_t(X^{(s)}), \qquad X_f^{(s)} = \mathcal{F}_f(X^{(s)}), \tag{3}$$

where $\mathcal{F}_t(\cdot)$ and $\mathcal{F}_f(\cdot)$ denote FFT operations along the temporal and feature dimensions, respectively.

Here, $X_t^{(s)}$ mainly captures variation patterns under the current temporal resolution, whereas $X_f^{(s)}$ reflects cross-feature interactions among multivariate inputs. These two spectral representations are then fused in the representation space through concatenation followed by a learnable linear projection:

$$X_{\text{freq}}^{(s)} = \phi^{(s)}\left(\text{Concat}\left[X_t^{(s)}, X_f^{(s)}\right]\right), \tag{4}$$

where $\phi^{(s)}(\cdot)$ denotes a scale-specific linear projection used to integrate temporal spectral information and cross-feature spectral dependencies. In implementation, the complex-valued FFT outputs are converted into real-valued spectral embeddings before the learnable projection. The inverse FFT is then applied along the temporal dimension, and the real component of the reconstructed representation is retained for subsequent fusion. The feature-wise Fourier operation is used as an FNet-inspired feature-mixing mechanism rather than a strict physical spectral decomposition of heterogeneous variables. In this way, the fusion is implemented as a learnable representation-level aggregation of heterogeneous spectral features.

The fused spectral representation is then transformed back into the time domain:

$$\widetilde{X}^{(s)} = \mathcal{F}^{-1}(X_{\text{freq}}^{(s)}), \tag{5}$$

where $\mathcal{F}^{-1}(\cdot)$ denotes the inverse FFT operation along the temporal dimension, which restores the frequency-enhanced representation to the time-domain feature space.

2.2.3. Upsampling and feature fusion

Since the output lengths differ across scales, each reconstructed sequence $\widetilde{X}^{(s)}$ is upsampled via linear interpolation to match the original length L:

$$\hat{X}^{(s)} = \text{Upsample}(\widetilde{X}^{(s)}, L). \tag{6}$$

Finally, the multiscale features are concatenated and fused through a linear projection:

$$X_{\text{msf}} = \text{Concat}\left[\hat{X}^{(1)}, \hat{X}^{(2)}, \hat{X}^{(4)}\right], \quad X_{\text{out}} = W_f X_{\text{msf}}, \tag{7}$$

where $W_f \in \mathbb{R}^{(D\times|S|)\times D}$ is the fusion weight matrix. This operation enables multiscale frequency-domain feature fusion through linear projection, thereby yielding a unified representation of local dynamics, short-term variation patterns, and abrupt fluctuations in wind power sequences.

2.3. Convolutional Block Attention Module (CBAM)

To enhance the model's ability to focus on important variables and informative temporal patterns in high-dimensional multivariate inputs, a Convolutional Block Attention Module (CBAM) is introduced. This module dynamically allocates attention weights along both the feature and spatial dimensions, enabling adaptive enhancement of critical features and suppression of irrelevant ones. The architecture is illustrated in Fig. 2.

2.3.1. Channel Attention (CA)

The channel attention mechanism aims to assess the global contribution of each feature variable to the prediction task. Let $F \in \mathbb{R}^{T\times C}$ denote the fused input feature map, where T is the sequence length and C is the number of channels.

Two descriptors are obtained via global average pooling (GAP) and global max pooling (GMP) along the temporal axis:

$$f_{\text{avg}}(j) = \frac{1}{T}\sum_{t=1}^{T} F(t,j), \quad f_{\text{max}}(j) = \max_{1\le t\le T} F(t,j) \tag{8}$$

These descriptors are then passed through a shared two-layer multilayer perceptron (MLP) with ReLU and sigmoid activations to obtain the channel attention vector:

$$u = \sigma(W_2\delta(W_1 f_{\text{avg}} + W_1 f_{\text{max}})) \tag{9}$$

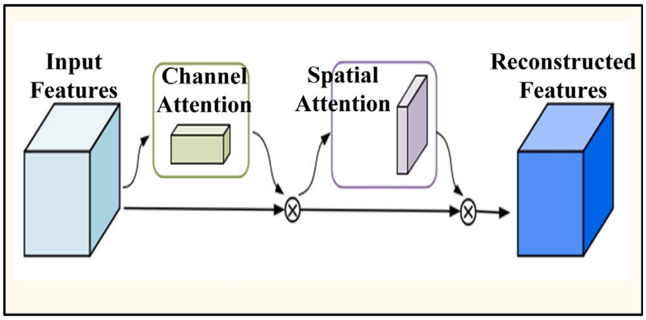

Fig. 2. Overall architecture of the CBAM.

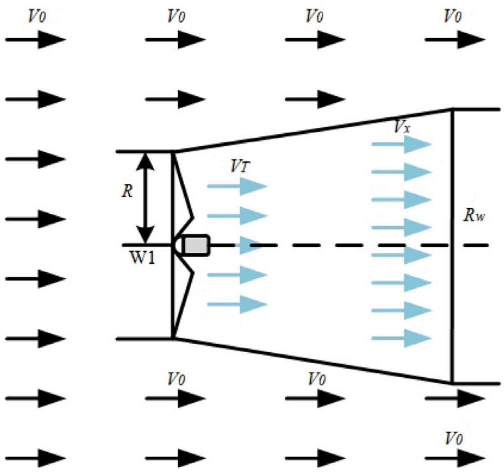

Fig. 3. Schematic diagram of Jensen's wake model.

where W_1 and W_2 are trainable weight matrices, $\delta(\cdot)$ denotes the ReLU activation, and $\sigma(\cdot)$ denotes the sigmoid activation. The resulting vector $u \in [0,1]^C$ represents the importance of each channel and is broadcast back to the original feature map:

$$F'(t,j) = u(j) \cdot F(t,j), \quad \forall t \in [1,T], j \in [1,C] \tag{10}$$

This operation amplifies the channels most relevant to the output while suppressing redundant inputs, enabling adaptive selection of important feature dimensions.

2.3.2. Spatial Attention (SA)

To further capture key temporal points such as wind fluctuations or power spikes, spatial attention is applied over the time axis. From the enhanced feature map F', global average pooling and global max pooling are applied across the channel dimension:

$$g_{\text{avg}}(t) = \frac{1}{C}\sum_{j=1}^{C} F'(t,j), \quad g_{\max}(t) = \max_{1\le j\le C} F'(t,j) \tag{11}$$

These two descriptors are concatenated into a matrix $G \in \mathbb{R}^{2\times T}$ and passed through a one-dimensional convolution layer:

$$s(t) = \sigma(w * [g_{\text{avg}}; g_{\max}]) \tag{12}$$

where w is the convolution kernel and $*$ denotes 1D convolution. The resulting sequence $s(t)$ provides temporal attention weights, which are then applied to the feature map along the time dimension:

$$F''(t,j) = s(t) \cdot F'(t,j), \quad \forall t \in [1,T], j \in [1,C] \tag{13}$$

In this way, the network emphasizes critical moments associated with rapid power shifts or meteorological disturbances, while reducing the influence of relatively stable periods.

2.4. Physics-constrained learning based on Jensen's wake model

After modeling the multiscale temporal and frequency-domain features of wind power sequences, the forecasting results are still required to satisfy the fundamental physical laws governing turbine operation. Due to the significant wake effect among wind turbines, purely data-driven models may produce predictions that are inconsistent with the actual aerodynamic mechanisms under certain wind conditions. To improve the physical consistency and reliability of the forecasting results, this study introduces a physics-constrained mechanism based on Jensen's wake model to regularize the model outputs.

When an upstream turbine operates, its rotor blades generate a downstream velocity-deficit region, namely the wake region. The existence of this wake reduces the incoming wind speed of downstream turbines, thereby decreasing their power output. Based on the principle of momentum conservation, Jensen's wake model assumes that the wake cross-section expands linearly along the downstream direction and provides an analytical expression for the wind speed at any downstream location, as illustrated in Fig. 3. For a downstream position located at a distance x from the upstream turbine, the normalized wind speed given by Jensen's model is expressed as

$$\frac{U_d(x)}{U_0} = 1 - \left(1 - \sqrt{1 - C_T}\right)\left(\frac{D}{D + 2kx}\right)^2, \tag{14}$$

where U_0 denotes the free-stream wind speed, $U_d(x)$ is the downstream wind speed at distance x, D is the rotor diameter, C_T is the thrust coefficient, and k is the wake expansion coefficient. As the downstream distance increases, $U_d(x)$ gradually recovers, but it remains lower than the free-stream wind speed U_0 within a finite range. Therefore, the degree of wake attenuation is jointly determined by the turbine resistance characteristics and the downstream distance.

It should be noted that Jensen's wake model is adopted in this study as a lightweight physical prior rather than a high-fidelity aerodynamic simulation model. This model is mainly applicable to operating scenarios with relatively regular turbine layouts and dominant wake effects, while more complex unsteady wake superposition, strong turbulence, and aerodynamic coupling under complex terrain are not explicitly modeled. Therefore, the proposed physical constraint is essentially a soft-constrained learning mechanism based on physical priors, whose purpose is to enhance the consistency between the forecasting results and wake dynamics, rather than to fully replace detailed physical modeling.

Since the turbine power output is approximately proportional to the cube of wind speed, i.e., $P \propto U^3$, the above wind speed attenuation law determines the physical upper bound of the downstream turbine power under given wind conditions and turbine layouts. If this constraint is ignored, the forecasting model may produce results that are inconsistent with the actual wind farm dynamics, for example, predicting an abnormal increase in downstream turbine output even when the upstream turbine is already operating close to its rated condition.

To avoid such unreasonable predictions, this study introduces a one-dimensional Jensen wake model to compute the physical prediction value and uses it as an important reference for data-driven forecasting. Specifically, by combining the actual wind farm layout and meteorological conditions, the Jensen model is used to calculate the physical expected power $P^{\text{phys}}(t)$ at each time step. For adjacent turbines, the measured wind speed and power information of the upstream turbine can be used to estimate the theoretical upper bound of the downstream turbine power at time t. Subsequently, the mean squared error (MSE) between the data-driven prediction and the physical prediction is formulated as a physics-constrained term to guide model training in a soft-constrained manner. The physics-constrained loss is defined as

$$L_{\text{phys}} = \frac{1}{T}\sum_{t=1}^{T}\left(\hat{P}(t) - P^{\text{phys}}(t)\right)^2, \tag{15}$$

where T denotes the prediction horizon. By minimizing this loss, the model is encouraged to follow wake dynamics while reducing forecasting errors, thereby improving the physical plausibility and reliability of the forecasting results.

2.5. Loss function and optimization strategy

To jointly consider forecasting accuracy and physical consistency, the overall training objective of the proposed model is defined as a weighted combination of the data-driven error term and the physics-constrained term, i.e.,

$$L_{\text{total}} = L_{\text{data}} + \lambda L_{\text{phys}}, \tag{16}$$

where L_{data} denotes the data-driven prediction loss measuring the discrepancy between the predicted power $\hat{P}(t)$ and the ground-truth power $P(t)$, L_{phys} denotes the physics-constrained term constructed from Jensen's wake model, and λ is a non-negative coefficient used to balance the contributions of the two terms. In this study, the mean squared error (MSE) is adopted as the data-driven loss, which is formulated as

$$L_{\text{data}} = \frac{1}{T}\sum_{t=1}^{T}\left(\hat{P}(t) - P(t)\right)^2, \tag{17}$$

where T is the prediction horizon.

It should be emphasized that the physical constraint adopted in this work is essentially a regularized learning mechanism based on physical priors. The data-driven term ensures the model's fitting capability to historical statistical patterns, whereas the physical constraint term limits the deviation of the forecasting results from wake dynamics, thereby enhancing the physical plausibility of the model outputs. To avoid a significant scale mismatch between the two loss components, both the data-driven predictions and the physical reference values are normalized in advance, ensuring that L_{data} and L_{phys} are numerically comparable. This makes the weighted summation reasonable and stable during optimization.

At the early stage of training, the model mainly focuses on learning the statistical characteristics of wind power sequences. If the physical constraint is assigned an excessively large weight at this stage, it may interfere with the fitting of the dominant temporal patterns. Therefore, a relatively small weight is assigned to the physical term at the beginning of training, allowing the model to prioritize the learning of fundamental statistical features from the wind power data. As training proceeds and the model becomes able to capture the underlying data distribution more effectively, the influence of the physical constraint is gradually strengthened so that the forecasting results become increasingly consistent with wake dynamics.

To implement this training balance in a stable and simple manner, a dynamic weighting schedule is adopted for λ. Specifically, the initial value is set as $\lambda_0 = 0.001$, and a linear increasing schedule is employed to progressively raise the weight of the physics-constrained term as the training epochs increase. In this way, the contribution of the physical constraint to the total loss is adjusted from low to high during training. Such a schedule allows the model to focus on data-driven feature learning in the early training stage, while gradually strengthening the physical regularization effect in later epochs, thereby improving the consistency between prediction accuracy and wake-aware physical guidance.

It should also be noted that the current weighting strategy is a deterministic scheduling mechanism rather than a fully adaptive weighting method driven by uncertainty estimation or gradient balancing. Its advantage lies in implementation simplicity and stable optimization behavior under the present experimental setting. More advanced dynamic weighting strategies, such as uncertainty-based weighting or gradient-aware adaptive balancing, will be further investigated in future work.

2.6. Overall architecture of PGMFNet

To effectively capture the multiscale characteristics of wind power sequences while improving the physical consistency of forecasting results, this paper proposes a PGMFNet framework. The proposed framework jointly integrates temporal modeling, multiresolution frequency-domain representation, adaptive attention refinement, and Jensen-based physical regularization.

The proposed PGMFNet architecture consists of six main components. First, an embedding layer transforms the raw meteorological and power inputs into a unified latent space. Then, an SSM branch is employed to model long-term dependencies and periodic evolutionary patterns in the sequence. In parallel, the MSFNet extracts local dynamics, short-term variation patterns, and abrupt fluctuations under multiple temporal resolutions. The representations from both branches are further refined by a CBAM, which adaptively identifies the most influential input variables and critical time steps. The fused features are then passed through a fully connected (FC) layer to generate the final wind power forecast. Finally, a physics-constrained learning module based on Jensen's wake theory is incorporated into the training process to improve the consistency of the forecasting results with turbine aerodynamic behavior. The complete model structure is shown in Fig. 4.

Although PGMFNet integrates several existing modeling ideas, these components are organized through a problem-driven information flow rather than being used as independent predictors. The SSM branch first captures global temporal evolution, while the MSFNet branch provides a complementary multiresolution spectral representation of local fluctuations. These two representations are fused before being passed to the CBAM, so that the attention module refines a joint temporal–spectral feature space and adaptively emphasizes important variables and time steps. The refined representation is then mapped to the final forecast by the FC layer, while the Jensen-based physical constraint regularizes the prediction output toward wake-consistent behavior. Therefore, the proposed framework is designed as a coordinated temporal–spectral–attention–physical learning pipeline, rather than a simple module stacking strategy.

3. Experimental setup and evaluation protocol

3.1. Dataset description

This study employs real-world measurements from a publicly available hybrid onshore–offshore wind farm dataset in the United States. Two benchmark forecasting tasks are constructed to evaluate the proposed method: wind power forecasting for WT1 in the inland wind farm and wind power forecasting for WT3 in the offshore wind farm. To support wake-aware physical modeling, two pairs of adjacent wind turbines, W01 & W02 in the onshore wind farm and W03 & W04 in the offshore wind farm, are selected as the experimental subjects. In each pair, W02 and W04 are regarded as upstream turbines, while W01 and W03 are regarded as downstream turbines. This setting provides a physical basis for incorporating Jensen's wake theory into the forecasting framework.

The dataset covers the period from January 1, 2010 to December 31, 2010. The original measurements include active power outputs and concurrent meteorological variables, including wind speed, wind direction, air density, turbulence intensity, wind shear, and normalized wind power. For each benchmark forecasting task, the full-year chronological sequence is divided into training, validation, and test sets with a ratio of 7:1:2. Specifically, the first 70% of the samples are used for model training, the subsequent 10% are used for validation and hyperparameter selection, and the remaining 20% are reserved for final testing. No random shuffling is applied during data partitioning, thereby preserving the temporal order of the wind power sequence and preventing future information leakage.

The main statistical properties of the key variables for WT1 and WT2 in the onshore wind farm are summarized in Table 1, while the temporal variation patterns of wind speed and normalized wind power are illustrated in Fig. 5. Specifically, the variables include wind speed (V), wind direction (D), air density (ρ), turbulence intensity (I), wind shear (S), and normalized wind power (NormPW).

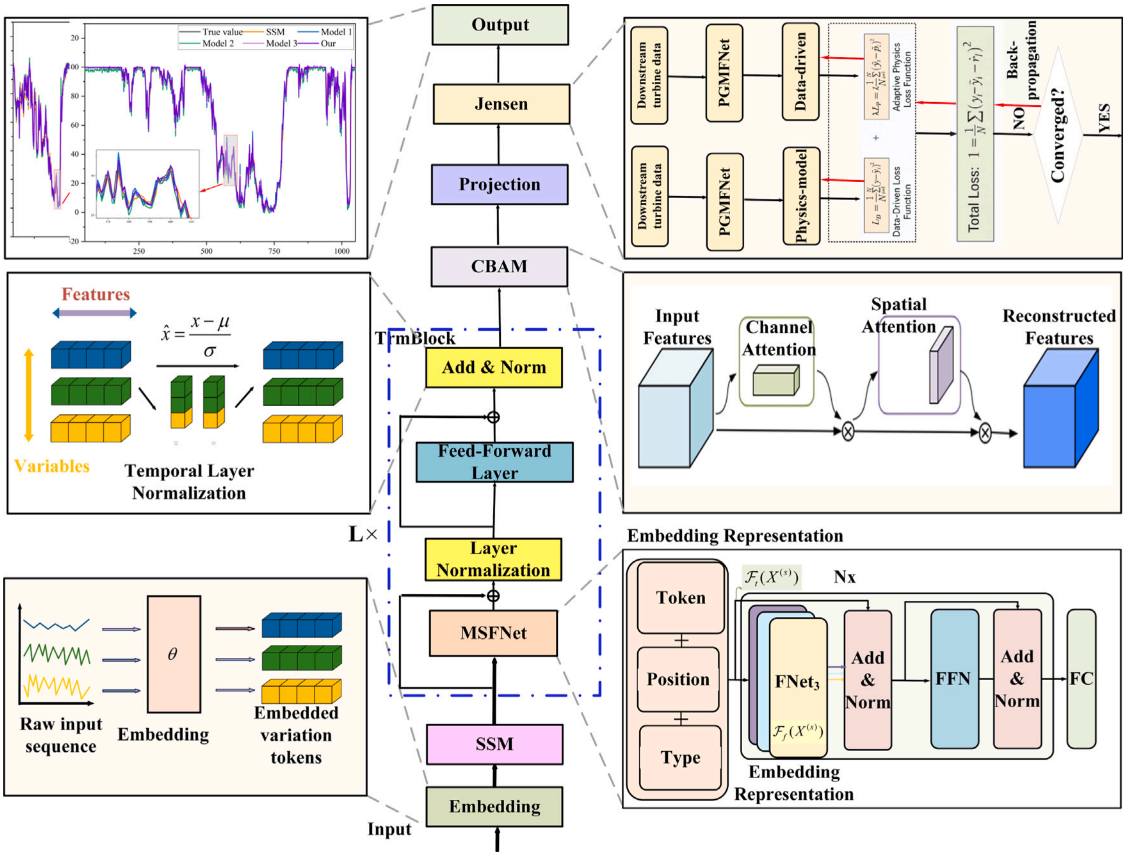

Fig. 4. Overall architecture of the proposed PGMFNet model.

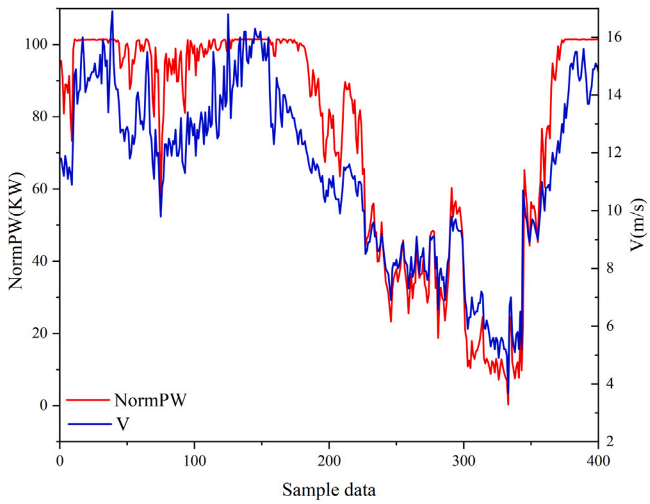

Fig. 5. Visualization of wind speed and normalized wind power in the WT1 dataset of the inland wind farm.

3.2. Data preprocessing

Data preprocessing includes outlier removal and normalization. All input features are scaled using min–max normalization to eliminate unit inconsistencies. Both cleaned meteorological variables and active power output are used as model inputs. To avoid future information leakage, the normalization parameters are fitted only on the training set and then applied to the validation

Table 1
Statistical summary of key variables for WT1 and WT2 in the onshore wind farm.

Feature	WT1			WT2		
	Mean	Median	Std	Mean	Median	Std
V	9.2032	9.0000	3.0244	8.7303	8.6000	2.9415
D	208.0250	205.1000	88.2483	208.6195	205.6000	87.1929
ρ	1.2036	1.1981	0.0498	1.2050	1.1989	0.0510
I	0.1115	0.0877	0.1181	0.1037	0.0797	0.0990
S	0.3033	0.2616	0.2097	0.3053	0.2616	0.2131
NormPW	52.8688	51.4848	33.5405	53.2255	51.5152	34.0575

Table 2
Key hyperparameter settings of the proposed PGMFNet.

Module	Parameter	Description	Search range	Selected value
SSM	d_{conv}	SSM kernel size	[1, 12]	4
	γ_{exp}	SSM expansion factor	[1, 12]	2
MSFNet	$\mathcal{S}$	Fourier scales	–	[1, 2, 4]
	d_{ff}	FNet FFN dim.	[1, 128]	32
CBAM	r_{red}	Channel reduction	[1, 64]	32
	k_{spa}	Spatial kernel size	[1, 12]	7

and testing sets. This preprocessing ensures sample diversity and data integrity, serving as a robust foundation for evaluating the proposed PGMFNet.

To ensure temporal causality, all datasets are divided into training, validation, and testing subsets in chronological order without random shuffling. A rolling time-series split strategy is adopted during the forecasting process. Specifically, the model is trained using historical observations and evaluated on subsequent unseen samples, while the input window is moved forward step by step along the time axis. This setting better simulates practical wind power forecasting scenarios, where only past and currently available information can be used for future prediction.

3.3. Model configuration and training parameters

The proposed PGMFNet integrates temporal modeling, frequency-domain representation learning, and physics-constrained optimization within a unified forecasting framework. The main components of the model include an embedding layer, a SSM, an MSFNet-based frequency-domain modeling module, a CBAM module, a fully connected prediction layer, and a physics-constrained learning module.

During training, grid search is employed to optimize the key hyperparameters of the proposed model, and the final selected configurations are summarized in Table 2. The Adam optimizer is adopted to iteratively update the network parameters, with the initial learning rate set to 1×10^{-3}. In addition, an adaptive learning rate decay strategy based on the validation error is applied to accelerate convergence. In terms of the loss design, the Mean Squared Error (MSE) is used to measure the discrepancy between the predicted value $\hat{y}$ and the ground-truth value y, and serves as the primary optimization objective. Meanwhile, a physics-constrained loss term is further introduced to ensure the physical plausibility of the forecasting results, as described in Section 2.4.

To ensure fair comparison, all models were trained and evaluated on the same datasets, using the same chronological train/validation/test partitioning, input–output setting, preprocessing procedure, and evaluation metrics. Hyperparameter tuning was conducted based only on the validation set, while the test set was used exclusively for final performance evaluation. For each model, candidate configurations were evaluated under the same training protocol, and the configuration with the best validation performance was selected for final testing.

The key hyperparameters of PGMFNet are listed in Table 2. For the SOTA sequence forecasting baselines, the main tuned configurations used in the final comparison are summarized in Table 3. The key hyperparameter settings of the commonly used baseline models are summarized in Table 4.

3.4. Evaluation metrics

To objectively evaluate the prediction accuracy of different models in wind power forecasting, the following commonly used metrics are adopted:

- **Coefficient of Determination** (R^2):

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left(f_r(x_i) - f_n(x_i)\right)^2}{\sum_{i=1}^{n} \left(f_r(x_i) - \bar{f}_r\right)^2} \tag{18}$$

Table 3
Unified default configurations of the SOTA forecasting models.

Parameter	Description	Value	Parameter	Description	Value
L_{seq}	Input length	10	L_{label}	Label length	5
L_{pred}	Horizon	1	B	Batch size	64
N_{epoch}	Epochs	100	r_{stop}	Stop ratio	0.2
d_{dec}	Decoder dim.	d_{data}	d_{enc}	Encoder dim.	d_{data}
d_{out}	Output dim.	1	d_{model}	Hidden dim.	64
N_{head}	Attention heads	8	p_{drop}	Dropout	0.1
N_{enc}	Encoder layers	2	N_{dec}	Decoder layers	1
d_{ff}	FFN dim.	32	λ_{attn}	Attn. factor	5
σ	Activation	GELU	k_{top}	Top-k	3
N_{kernel}	Kernel number	6	δ_{distil}	Distillation	1

Table 4
Key hyperparameter settings of commonly used baseline models.

Model	Hyperparameter	Search range	Selected value
XGBoost	Number of trees	[1, 200]	100
	Learning rate	[1e−4, 1e−1]	1e−1
	Maximum depth	[1, 100]	10
RF	Number of trees	[1, 200]	100
	Maximum depth	[1, 100]	10
	Minimum samples split	[2, 10]	2
	Minimum samples leaf	[1, 10]	1
LSTM	Number of layers	[1, 10]	2
	Units per layer	[1, 512]	64, 128
	Dropout rate	–	0.2
	Learning rate	[0.000, 0.1]	0.001
CNN	Number of layers	[1, 10]	3
	Number of filters	[1, 100]	16
	Kernel size	[1, 12]	4
	Pooling size	[1, 12]	2
GRU	Number of layers	[1, 10]	2
	Units per layer	[1, 512]	64, 128
	Dropout rate	–	0.2
	Learning rate	[0.000, 0.1]	0.001

- **Mean Absolute Error (MAE):**

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|f_n(x_i) - f_r(x_i)\right| \tag{19}$$

- **Root Mean Squared Error (RMSE):**

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(f_n(x_i) - f_r(x_i)\right)^2} \tag{20}$$

- **Mean Absolute Scaled Error (MASE):**

$$\text{MASE} = \frac{\frac{1}{n}\sum_{i=1}^{n}\left|f_n(x_i) - f_r(x_i)\right|}{\frac{1}{n-1}\sum_{i=2}^{n}\left|f_r(x_i) - f_r(x_{i-1})\right|} \tag{21}$$

where $f_n(x_i)$ and $f_r(x_i)$ denote the predicted and true values at time step i, respectively, and $\bar{f}_r$ represents the sample mean of the true values. MAE, RMSE, and MASE measure the absolute error, squared error, and scaled error, respectively, while R^2 evaluates the goodness of fit between the predicted and actual values.

4. Experimental results and discussion

This section presents the experimental results, including comparative and ablation studies, to comprehensively evaluate the proposed model.

Table 5
Prediction performance comparison on the WT1 dataset of the inland wind farm and the WT3 dataset of the offshore wind farm.

Model	WT1 of Inland wind farm				WT3 of Offshore wind farm			
	R^2 ↑	RMSE ↓	MAE ↓	MASE ↓	R^2 ↑	RMSE ↓	MAE ↓	MASE ↓
Informer [22]	0.9765	4.9963	3.2785	1.0623	0.9705	6.3682	3.5383	1.2925
Autoformer [23]	0.9462	7.5547	5.1798	1.6783	0.9531	8.0320	4.6679	1.7051
iTransformer [21]	0.9813	4.4589	2.6785	0.8678	0.9859	4.4101	2.3534	0.8597
PatchTST [24]	0.9825	4.3046	2.6867	0.8705	0.9797	5.2838	3.1780	1.1609
TimesNet [25]	0.9523	7.1151	4.3956	1.4242	0.9752	5.8416	3.1831	1.1627
FRETS [26]	0.9560	6.8274	4.2102	1.3641	0.9742	5.9587	3.1865	1.1640
XGBoost [29]	0.8324	13.3332	11.6412	3.7718	0.8463	14.5452	12.6169	4.6088
RF [30]	0.9782	4.8134	2.9986	0.9715	0.9695	6.4772	3.5036	1.2798
GRU [28]	0.8648	11.9741	8.4744	2.7457	0.9363	9.3662	6.1990	2.2644
CNN [27]	0.8719	11.6566	8.7000	2.8188	0.9275	9.9869	6.8459	2.5007
PGMFNet (Proposed)	**0.9846**	**4.0395**	**2.4308**	**0.7876**	**0.9871**	**4.2068**	**2.2327**	**0.8156**

4.1. Comparative experiments

To verify the effectiveness of the proposed hybrid framework, multiple comparative experiments were conducted based on the settings commonly adopted in recent studies [21]. The proposed PGMFNet was compared against both traditional time-series forecasting models and recent state-of-the-art (SOTA) deep learning models, including Informer [22], Autoformer [23], iTransformer [21], PatchTST [24], TimesNet [25], FRETS [26], CNN [27], GRU [28], XGBoost [29], and Random Forest(RF) [30]. These models represent the most widely adopted and competitive approaches for wind power prediction in recent years. It should be noted that all numerical results reported for the comparison models were implemented, trained, and evaluated by the authors under the same experimental protocol. The results were not directly copied from previous publications. Specifically, all baseline models were tested using the same dataset partitioning, input–output settings, preprocessing procedure, and evaluation metrics as the proposed PGMFNet, thereby ensuring a fair and consistent comparison.

Table 5 summarizes the forecasting performance on the Inland_Wind_Farm_2010_WT1 and Offshore_Wind_Farm_2010_WT3 datasets, respectively, while the detailed comparative results are further illustrated in Fig. 6.

On the WT1 dataset of the inland wind farm, PGMFNet achieves the best performance across all four metrics. Compared with the best competing result for each metric, the proposed model improves R^2 by about 0.21%, while reducing RMSE, MAE, and MASE by approximately 6.16%, 9.25%, and 9.24%, respectively. Compared with iTransformer, the RMSE, MAE, and MASE are reduced by about 9.41%, 9.25%, and 9.24%, respectively, while compared with PatchTST, the corresponding reductions are approximately 6.16%, 9.52%, and 9.52%. These results indicate that, in the inland wind farm scenario, the proposed model not only provides better overall fitting capability, but also achieves more stable error control than existing strong Transformer-based baselines. This advantage mainly arises from the complementary design of the SSM and MSFNet: the SSM captures long-range dependencies and periodic evolutionary patterns, whereas the MSFNet enhances the perception of local dynamics, short-term variations, and abrupt fluctuations through multiresolution frequency-domain encoding. As a result, the proposed framework can better characterize the multiscale dynamics of inland wind power sequences.

On the WT3 dataset of the offshore wind farm, PGMFNet also achieves the best overall performance among all compared models. Compared with the strongest competing baseline, iTransformer, the proposed model improves R^2 by about 0.12%, while reducing RMSE, MAE, and MASE by approximately 4.61%, 5.13%, and 5.13%, respectively. Compared with PatchTST, the proposed model further improves R^2 by about 0.76%, while reducing RMSE, MAE, and MASE by approximately 20.38%, 29.75%, and 29.74%, respectively. In addition, compared with TimesNet, the RMSE, MAE, and MASE are reduced by approximately 27.99%, 29.86%, and 29.85%, respectively. These results suggest that, under the more stochastic and highly fluctuating offshore wind conditions, PGMFNet is more effective at suppressing larger prediction deviations and maintaining overall fitting accuracy. This improvement is closely related to the coordinated modeling strategy of the proposed framework: the SSM is responsible for capturing long-term temporal evolution, while the MSFNet strengthens the representation of local abrupt changes and short-term nonstationary perturbations, enabling the model to remain robust under stronger wake interactions and more complex atmospheric disturbances.

Compared with conventional machine learning models and shallow neural networks, the superiority of PGMFNet is even more significant. On both datasets, the proposed model consistently yields substantially lower RMSE, MAE, and MASE than XGBoost, RF, GRU, and CNN, which confirms that simple nonlinear mapping or shallow temporal modeling is insufficient for representing the complex coupling of long-term evolution, local abrupt fluctuations, and wake-related physical effects in wind power forecasting. By jointly integrating the SSM, MSFNet, CBAM, and Jensen-based physical constraint, PGMFNet achieves complementary fusion across temporal modeling, frequency-domain representation, and physical-prior guidance. This coordinated design enables the proposed framework to deliver more accurate, robust, and physically consistent ultra-short-term wind power forecasting for both inland and offshore wind farm scenarios.

4.2. Ablation study

To investigate the contribution of each key component to the overall model performance, an ablation study was conducted using the encoder backbone containing the SSM as the baseline. The experiment quantitatively evaluates the effects of the MSFNet, the

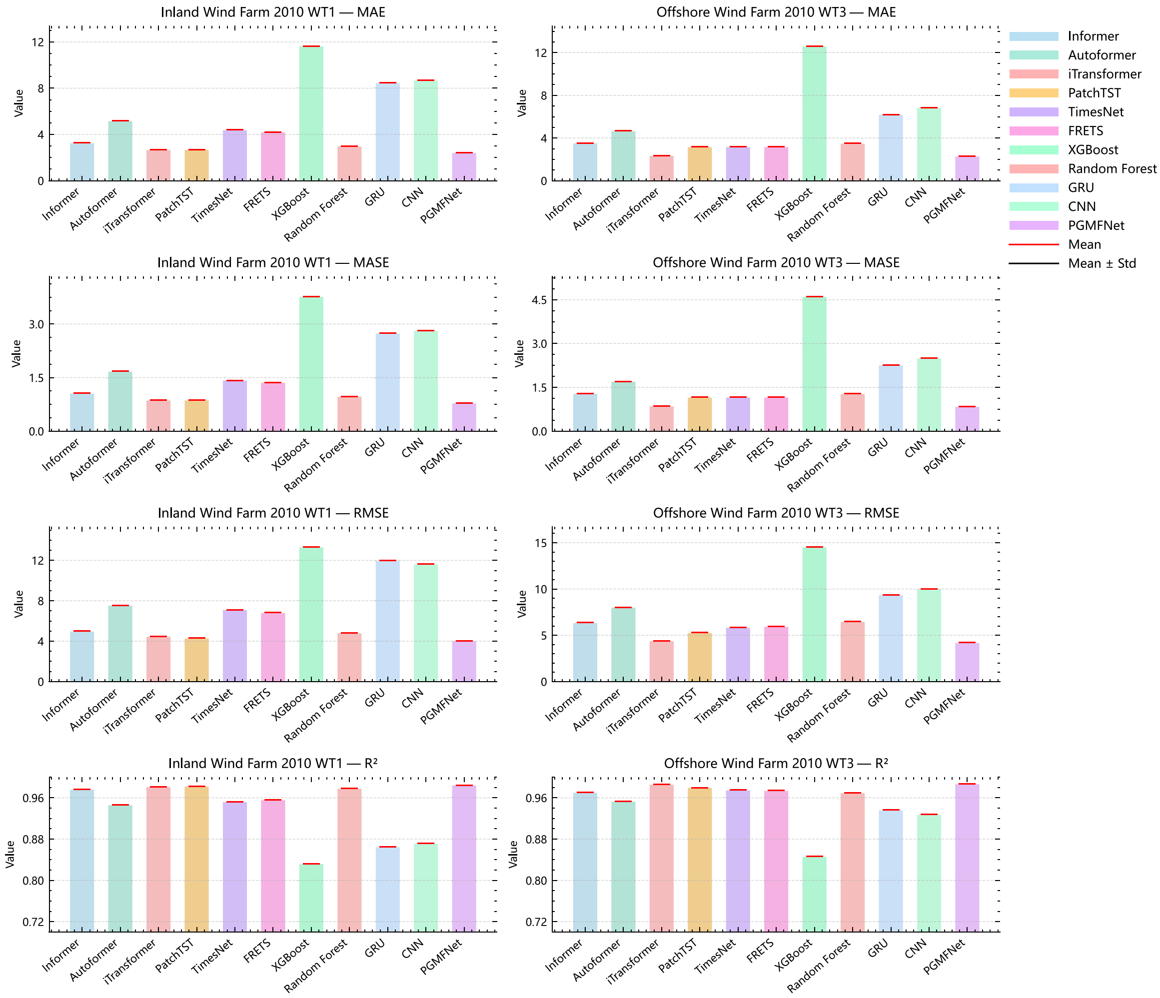

Fig. 6. Comparative performance of multiple forecasting models on two wind farm datasets across four evaluation metrics.

Table 6
Ablation study results of the proposed PGMFNet on two wind farm datasets.

Name	Module combination			WT1 of Inland wind farm				WT3 of Offshore wind farm			
	MSFNet	CBAM	Physics	R^2 ↑	RMSE ↓	MAE ↓	MASE ↓	R^2 ↑	RMSE ↓	MAE ↓	MASE ↓
SSM				0.9818	4.3874	2.9986	0.9715	0.9782	5.4807	3.1831	1.1627
Model 1		✓		0.9826	4.2985	3.2785	1.0623	0.9828	4.8635	2.7213	0.9941
Model 2	✓			0.9835	4.1784	2.6785	0.8678	0.9798	5.2780	3.4428	1.2576
Model 3	✓	✓		0.9842	4.0926	2.6867	0.8705	0.9869	4.2384	2.3069	0.8427
Ours	✓	✓	✓	**0.9846**	**4.0395**	**2.4308**	**0.7876**	**0.9871**	**4.2068**	**2.2327**	**0.8156**

CBAM, and the physics-constrained component based on Jensen's wake model. A controlled-variable approach was adopted, where each experiment removes or replaces a single module of the PGMFNet while keeping all other settings identical. This design allows a clear assessment of the impact of each component on the model's predictive accuracy and robustness.

As shown in Table 6 and illustrated in Fig. 7, the contribution of each component is clearly demonstrated in the *Inland Wind Farm* environment.

On the WT1 dataset of the inland wind farm, the MSFNet contributes more consistently to improving prediction accuracy. When only the MSFNet is introduced (Model 2), the RMSE, MAE, and MASE are reduced by approximately 4.76%, 10.67%, and 10.67%, respectively, compared with the baseline SSM. This indicates that multiscale frequency-domain modeling can effectively enhance the perception of local dynamics, short-term variations, and abrupt fluctuations. By contrast, introducing CBAM alone (Model 1) shows a mixed effect on WT1. Although it slightly improves the R^2 and reduces the RMSE by approximately 2.03%, the MAE and MASE increase by about 9.33% and 9.35%, respectively. This suggests that attention enhancement alone may not be sufficient to improve

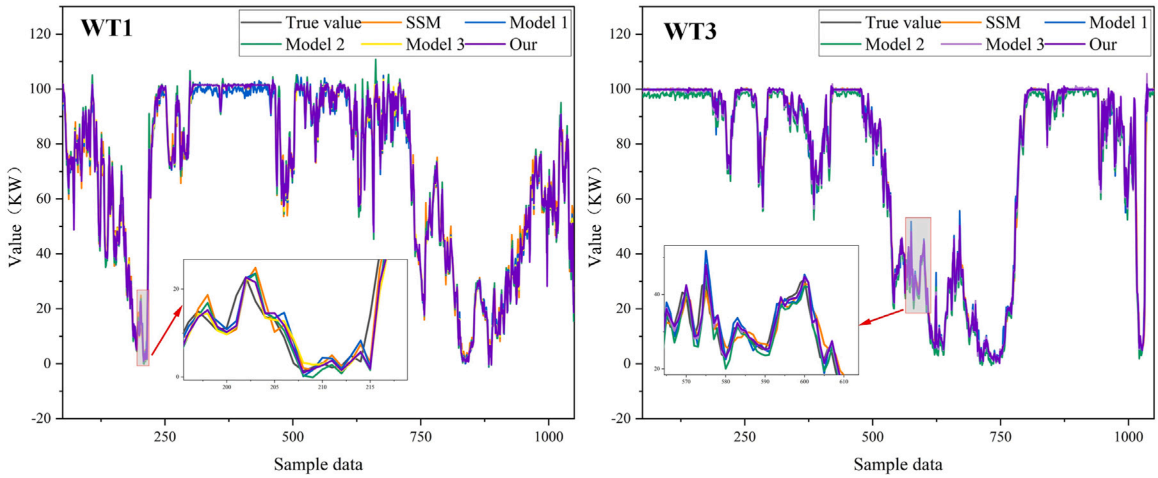

Fig. 7. Prediction comparison of ablation variants and the proposed PGMFNet on WT1 and WT3.

point-wise prediction accuracy in the relatively stable inland scenario without adequate multiscale dynamic representation. When MSFNet and CBAM are jointly incorporated (Model 3), the RMSE, MAE, and MASE are reduced by approximately 6.72%, 10.40%, and 10.40%, respectively, compared with the baseline SSM, confirming that the two modules can form a complementary effect once multiscale frequency-domain features are adequately extracted. After further introducing the Jensen-based physics-constrained component, the complete PGMFNet achieves the best overall performance, with an absolute improvement of 0.0028 in R^2 and reductions of 7.93%, 18.94%, and 18.93% in RMSE, MAE, and MASE, respectively, compared with the baseline SSM. Even relative to Model 3, the complete model further reduces RMSE, MAE, and MASE by approximately 1.30%, 9.52%, and 9.52%, respectively. These results indicate that the physics-guided component provides additional improvement in prediction consistency and point-wise error reduction on top of the temporal–spectral representation.

In the WT3 dataset of the offshore wind farm, the contributions of the individual modules become more pronounced and exhibit different characteristics, reflecting the greater complexity of offshore wind conditions. After introducing CBAM alone (Model 1), the RMSE, MAE, and MASE decrease by approximately 11.26%, 14.51%, and 14.50%, respectively, compared with the baseline SSM. This indicates that the attention mechanism can effectively identify critical variables and informative time steps under stronger stochasticity and turbulence. When only the MSFNet is adopted (Model 2), the RMSE is reduced by approximately 3.70%, while both MAE and MASE become worse than those of the baseline. This suggests that frequency-domain modeling alone is insufficient to maintain stable point-wise prediction under highly nonstationary offshore conditions. When the MSFNet and CBAM are combined (Model 3), the RMSE, MAE, and MASE are reduced by approximately 22.67%, 27.53%, and 27.52%, respectively, substantially outperforming the single-module settings and demonstrating strong complementarity between multiscale frequency modeling and attention-based feature refinement. After further incorporating the physics-constrained component, the complete PGMFNet achieves the highest R^2 and the lowest RMSE, MAE, and MASE on this dataset. Compared with the baseline SSM, the full model improves R^2 by an absolute value of 0.0089, while reducing RMSE, MAE, and MASE by approximately 23.24%, 29.86%, and 29.85%, respectively. Even relative to Model 3, the complete model still achieves additional reductions of approximately 0.75%, 3.22%, and 3.22% in RMSE, MAE, and MASE, respectively. This confirms that the Jensen-based physical guidance is particularly beneficial in offshore scenarios, where stronger wake interactions and more complex atmospheric disturbances require not only accurate temporal representation but also physically consistent prediction behavior.

Overall, the ablation results demonstrate that the effectiveness of each module varies across different wind farm environments. In the inland case, MSFNet contributes more consistently to error reduction, while CBAM alone shows a mixed effect. In the offshore case, CBAM provides stronger standalone benefits, whereas MSFNet alone is insufficient to reduce point-wise errors. The combination of MSFNet and CBAM achieves clear complementary gains in both datasets, and the Jensen-based physics-constrained component further improves prediction consistency, particularly in terms of MAE and MASE. These results verify that the proposed PGMFNet benefits from the coordinated integration of multiscale frequency modeling, attention-based feature refinement, and physics-guided regularization, leading to improved prediction accuracy, robustness, and physical interpretability in both inland and offshore wind farm environments.

4.3. Diebold–Mariano test

To further verify the statistical significance of the forecasting improvement achieved by the proposed PGMFNet, the Diebold–Mariano (DM) test is conducted on the WT1 dataset of the inland wind farm. The DM test is a widely used statistical method for comparing the predictive accuracy of two forecasting models. In this study, PGMFNet is regarded as the benchmark model, and the squared prediction error is used to construct the loss differential. The null hypothesis assumes that the compared model and

Table 7
DM test results between PGMFNet and comparison models on the WT1 dataset of the inland wind farm.

Proposed model	Compared model	DM statistic	p-value
PGMFNet	Informer [22]	6.0758	<0.001
	Autoformer [23]	14.1693	<0.001
	iTransformer [21]	5.7523	<0.001
	PatchTST [24]	3.3856	0.0007
	TimesNet [25]	11.8251	<0.001
	FRETS [26]	10.9724	<0.001
	XGBoost [29]	55.7383	<0.001
	Random Forest [30]	6.3026	<0.001
	GRU [28]	19.1266	<0.001
	CNN [27]	23.8021	<0.001

Table 8
Multi-horizon forecasting performance of PGMFNet on two wind farm datasets.

Horizon	WT1 of Inland wind farm				WT3 of Offshore wind farm			
	R^2 ↑	RMSE ↓	MAE ↓	MASE ↓	R^2 ↑	RMSE ↓	MAE ↓	MASE ↓
2-step	**0.9762**	**4.7263**	**3.0749**	**1.0214**	**0.9784**	**5.1819**	3.5331	1.2786
4-step	0.9660	6.0072	4.3853	1.4208	0.9761	5.7401	**3.4077**	**1.2448**
6-step	0.9345	8.3328	6.4197	2.0800	0.9398	9.1045	6.9787	2.5492
8-step	0.8992	10.3392	7.7051	2.4965	0.8885	12.3876	8.8960	3.2496

PGMFNet have equal predictive accuracy, while the alternative hypothesis indicates that there is a significant difference in their forecasting performance. The significance level is set to $\alpha = 0.05$.

As shown in Table 7, all comparison models obtain positive DM statistics, including Informer, Autoformer, iTransformer, PatchTST, TimesNet, FRETS, XGBoost, RF, GRU, and CNN. This indicates that, under the adopted loss criterion, the prediction losses of these models are larger than those of PGMFNet. Moreover, all corresponding p-values are smaller than 0.05, and most of them are lower than 0.001. Therefore, the null hypothesis of equal predictive accuracy can be rejected for all model comparisons. These results demonstrate that the performance improvement of PGMFNet over the comparison models is statistically significant rather than caused by random fluctuations in the test samples.

Specifically, PGMFNet shows significant advantages over Transformer-based forecasting models, including Informer, Autoformer, iTransformer, PatchTST, and TimesNet. This verifies that the proposed multiscale spectral modeling and physics-constrained learning mechanism can more effectively capture the complex temporal fluctuations and nonlinear patterns of wind power series. Compared with traditional machine learning models such as XGBoost and RF, PGMFNet also achieves statistically significant superiority, indicating its stronger nonlinear representation capability and better adaptability to nonstationary wind power data. In addition, the significant DM results against GRU and CNN further confirm that the proposed model benefits from the joint modeling of temporal dependencies, frequency-domain characteristics, and physical constraints.

Overall, the DM test results provide further statistical evidence for the effectiveness and robustness of PGMFNet. Combined with the quantitative evaluation results, it can be concluded that the proposed model achieves not only lower prediction errors but also statistically significant forecasting improvements on the WT1 dataset of the inland wind farm.

4.4. Multi-horizon forecasting analysis

To further evaluate the predictive performance of the proposed model under different forecasting horizon settings, 2-step, 4-step, 6-step, and 8-step forecasting experiments were conducted on the WT1 of Inland Wind Farm and the WT3 of Offshore Wind Farm datasets, so as to systematically analyze the adaptability and stability of the model in multi-horizon scenarios. The corresponding results are presented in Table 8.

As shown in Table 8, as the forecasting horizon gradually increases from 2 steps to 8 steps, the R^2 of PGMFNet on the WT1 of Inland Wind Farm dataset continuously decreases, while RMSE, MAE, and MASE exhibit an overall upward trend. This indicates that, as the prediction range expands, the error accumulation effect becomes increasingly significant and the forecasting task becomes more challenging. Nevertheless, under the 2-step and 4-step forecasting settings, the proposed model still maintains relatively high fitting accuracy and low error levels, demonstrating that it can effectively characterize the long-range dependencies and local dynamic patterns of inland wind power sequences over short horizons. Even under the 6-step and 8-step forecasting settings, although the prediction errors increase, the overall performance degradation remains relatively smooth, indicating that PGMFNet possesses good adaptability to different forecasting horizons and a certain resistance to error accumulation.

On the WT3 of Offshore Wind Farm dataset, the performance degradation becomes more pronounced as the forecasting horizon increases, especially under the 6-step and 8-step settings, where RMSE, MAE, and MASE increase considerably and the decline in R^2 is more significant. This suggests that long-horizon forecasting is more challenging in offshore wind farms due to stronger wind stochasticity and more complex wake propagation effects. However, under the 2-step and 4-step forecasting settings, PGMFNet

Table 9
Comparison of the proposed PGMFNet with other SOTA models on two wind farm datasets.

Model	WT1 of Inland wind farm				WT3 of Offshore wind farm			
	R^2 ↑	RMSE ↓	MAE ↓	MASE ↓	R^2 ↑	RMSE ↓	MAE ↓	MASE ↓
Transformer [31]	0.9667	5.9401	4.1154	1.3334	0.9765	5.6859	3.7426	1.3671
DLinear [32]	0.9737	5.2778	3.6235	1.1740	0.9525	8.0894	4.6367	1.6937
MLPM [33]	0.9434	7.7467	6.5483	2.1217	0.9837	4.7333	3.1384	1.1464
EDESN [34]	0.9794	4.6792	3.1508	1.0209	0.9784	5.4501	3.8913	1.4214
RVFL [35]	0.9778	4.8569	3.5131	1.1383	0.9840	4.6887	2.9877	1.0914
PGMFNet	**0.9846**	**4.0395**	**2.4308**	**0.7876**	**0.9871**	**4.2068**	**2.2327**	**0.8156**

still maintains relatively good forecasting accuracy, reflecting its strong stability in complex offshore scenarios. Overall, as the forecasting horizon increases, both the inland and offshore datasets exhibit a certain degree of performance degradation. Such a trend is consistent with the general characteristics of multi-step forecasting tasks and further demonstrates that the proposed model has good practical value in short-horizon ultra-short-term wind power forecasting.

4.5. Comparison with other SOTA models

To further validate the effectiveness and advancement of the proposed model on the same time-series forecasting task, several representative state-of-the-art models were additionally introduced as baselines on top of the original comparative experiments, including Transformer [31], DLinear [32], MLP-Mixer (MLPM) [33], ensemble deep ESN (EDESN) [34], and RVFL [35]. These models were selected to comprehensively evaluate the predictive performance of the proposed method under different modeling paradigms. To ensure fairness of comparison, all competing models were trained and tested under the same data split, input–output settings, and evaluation criteria as those described in Section 4.1. The corresponding experimental results are reported in Table 9.

On the WT1 of Inland Wind Farm dataset, PGMFNet achieves further improvements over the best competing model in this group, with an increase of approximately 0.53% in R^2, and reductions of approximately 13.67%, 22.85%, and 22.85% in RMSE, MAE, and MASE, respectively. Compared with Transformer, the RMSE, MAE, and MASE are reduced by about 32.00%, 40.93%, and 40.93%, respectively. Compared with DLinear, the corresponding reductions are approximately 23.46%, 32.92%, and 32.91%. These results indicate that, in the inland wind farm scenario, the proposed model not only provides better overall trend fitting, but also yields significantly stronger error control than both typical attention-based models and strong linear baselines. The underlying reason lies in the fact that the SSM can more effectively characterize long-range dependencies and periodic evolutionary patterns in wind power sequences, while the MSFNet enhances the perception of local dynamics, short-term variation patterns, and abrupt fluctuations through frequency-domain encoding under multiple temporal resolutions, thereby compensating for the limitations of conventional time-domain-dominated models in complex multiscale fluctuation modeling.

On the WT3 of Offshore Wind Farm dataset, PGMFNet also achieves superior performance over the best competing model, with an improvement of approximately 0.32% in R^2, and reductions of approximately 10.28%, 25.27%, and 25.27% in RMSE, MAE, and MASE, respectively. Compared with Transformer and DLinear, the RMSE is reduced by approximately 26.01% and 48.00%, the MAE is reduced by approximately 40.34% and 51.85%, and the MASE is reduced by approximately 40.34% and 51.85%, respectively. These results suggest that, as wind conditions become more stochastic and wake propagation effects become more pronounced in offshore wind farms, the advantage of PGMFNet becomes even more evident in complex scenarios. Its performance improvement stems not only from the coordinated modeling of long-term evolutionary patterns and local abrupt fluctuations by the SSM and MSFNet, but also from the enhanced representation of critical features and the improved physical consistency introduced by the overall framework.

4.6. Experimental discussion

The experimental results reveal several important insights into the effectiveness of the proposed framework. First, the consistent performance improvement across both inland and offshore datasets suggests that capturing multi-scale temporal dynamics is essential for modeling wind power. In particular, the ability of the MSFNet to represent both high-frequency fluctuations and low-frequency trends explains its advantage over conventional time-domain models, which tend to overlook rapid variations. Besides, the experimental results highlight that frequency-domain modeling alone is insufficient under highly nonstationary conditions. The observed performance variation of MSFNet in offshore scenarios indicates that accurate forecasting requires not only spectral decomposition but also adaptive feature selection and contextual modeling. This explains the complementary role of the CBAM module, which enhances robustness by focusing on relevant variables and suppressing noise. Furthermore, the incorporation of physics-guided constraints plays a crucial role in stabilizing model behavior under complex wind conditions. Rather than merely improving accuracy, the physics-based loss enforces consistency with wake dynamics, which contributes to improved generalization across different environments. This is particularly important in offshore settings, where strong turbulence and wake interactions amplify uncertainty.

5. Conclusion and future work

This paper proposed a Physics-Guided Multiscale Fusion Network (PGMFNet) for ultra-short-term wind power forecasting. The proposed framework integrates state-space temporal modeling, multi-resolution frequency-domain representation, attention-based feature refinement, and Jensen wake model-based physical regularization. Through this coordinated temporal–spectral–attention–physical learning pipeline, PGMFNet aims to enhance the representation of multiscale wind power fluctuations while improving the physical consistency of forecasting results.

Experimental results on both inland and offshore wind farm datasets demonstrate that PGMFNet achieves superior forecasting performance in terms of R^2, RMSE, MAE, and MASE compared with traditional machine learning methods, recurrent and convolutional neural networks, and recent state-of-the-art time-series forecasting models. The ablation study confirms the effectiveness of the MSFNet, CBAM, and Jensen-based physical constraint, showing that each component contributes to improved prediction accuracy, robustness, or physical consistency. In addition, the Diebold–Mariano test on the WT1 inland wind farm dataset verifies that the forecasting improvements achieved by PGMFNet are statistically significant under the adopted squared-error criterion.

From a computational perspective, PGMFNet introduces additional modeling components compared with conventional shallow baselines. However, its architecture is designed to remain computationally manageable. The SSM branch provides efficient sequence modeling without relying on the quadratic self-attention mechanism used in standard Transformer-style models. The MSFNet branch mainly depends on Fourier-domain operations, whose computational cost increases approximately in a quasi-linear manner with sequence length, while the CBAM module introduces only lightweight channel-wise and temporal attention operations. Therefore, the proposed framework provides a practical balance among forecasting accuracy, multiscale representation capability, physical consistency, and computational efficiency.

Future work will focus on further extending the proposed framework in three directions. First, graph-based or diffusion-based mechanisms will be incorporated to model turbine-level spatial correlations and wake propagation for spatio-temporal wind power forecasting. Second, more advanced physical priors and adaptive loss-weighting strategies will be explored to improve physical realism, training flexibility, and robustness under complex operating conditions. Third, additional public benchmark datasets and real-world wind farm datasets with detailed turbine layout and wake-interaction information will be used to further evaluate the generalization ability and practical applicability of the proposed method.

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

The dataset used in this study is publicly available from Zenodo: Ying Ding, Wind Spatio-Temporal Dataset, September 19, 2021, https://doi.org/10.5281/zenodo.5516543.

References

[1] Zhao Y, Liao H, Zhao Y, Pan S. Data-augmented trend-fluctuation representations by interpretable contrastive learning for wind power forecasting. Appl Energy 2025;380:125052.
[2] Wang D, Yang M, Zhang W, Ma C, Su X. Short-term power prediction method of wind farm cluster based on deep spatiotemporal correlation mining. Appl Energy 2025;380:125102.
[3] Tian H, Xing W, Leng Y. Tri-DCNNet: A three-branch network with deep compensation for wind power forecasting. Comput Electr Eng 2026;134:111104.
[4] Zhao X, Liu HP, Jin H, Cao S, Tang G. An improved hybrid model for wind power forecasting through fusion of deep learning and adaptive online learning. Comput Electr Eng 2024;120:109768.
[5] Alaeddini A, Echeverri GAG, Adesunkanmi R, Tamayo CR, Bhaganagar K. Spatio-temporal transformer-based modeling of multi-turbine wake effects for wind farm power forecasting. Energy AI 2026;100715.
[6] Songyue L, Qiusheng L, Bin L, Junyi H. Prediction of offshore wind turbine wake and output power using large eddy simulation and convolutional neural network. Energy Convers Manage 2025;324:119326.
[7] Liu M-D, Ding L, Bai Y-L. Application of hybrid model based on empirical mode decomposition, novel recurrent neural networks and the ARIMA to wind speed prediction. Energy Convers Manage 2021;233:113917.
[8] Liu S, Xu T, Du X, Zhang Y, Wu J. A hybrid deep learning model based on parallel architecture TCN-LSTM with Savitzky-Golay filter for wind power prediction. Energy Convers Manage 2024;302:118122.
[9] Xiang L, Liu J, Yang X, Hu A, Su H. Ultra-short term wind power prediction applying a novel model named SATCN-LSTM. Energy Convers Manage 2022;252:115036.
[10] Zehtabiyan-Rezaie N, Iosifidis A, Abkar M. Physics-guided machine learning for wind-farm power prediction: Toward interpretability and generalizability. PRX Energy 2023;2(1):013009.
[11] Wang Y, Yang F, Qi Y, Liu G, Gao J, Wang X, Wang Z. A transformer-LSTM network enhanced by EEMD for ultra-short-term wind power forecasting. Energy AI 2026;100682.
[12] Alaeddini A, Echeverri GAG, Adesunkanmi R, Tamayo CR, Bhaganagar K. Spatio-temporal transformer-based modeling of multi-turbine wake effects for wind farm power forecasting. Energy AI 2026;100715.
[13] Daenens S, Verstraeten T, Daems P-J, Nowé A, Helsen J. Spatio-temporal graph neural networks for power prediction in offshore wind farms using SCADA data. Wind Energy Sci 2025;10(6):1137–52.

[14] Zhu G, Jia W, Cheng L, Xiang L, Hu A. A non-stationary transformer model for power forecasting with dynamic data distillation and wake effect correction suitable for large wind farms. Energy Convers Manage 2025;324:119292.

[15] Romero-Barrera AJ, Sipols AE, Paricio-Garcia A, Lopez-Carmona MA. Short-term wind power forecasting integrating wake effect modeling with variational mode decomposition enhanced deep learning architectures. Energy Convers Manage 2026;348:120738.

[16] Jiang T, Liu Y. A short-term wind power prediction approach based on ensemble empirical mode decomposition and improved long short-term memory. Comput Electr Eng 2023;110:108830.

[17] Chen C, Li S, Wen M, Yu Z. Ultra-short term wind power prediction based on quadratic variational mode decomposition and multi-model fusion of deep learning. Comput Electr Eng 2024;116:109157.

[18] Liu Y, Qiu Y, Feng Y, Zhou H, Xiong X-L. Spatiotemporal physics-guided graph neural network for wind farm power prediction. Eng Appl Artif Intell 2026;171:114301.

[19] Machlev R, Heistrene L, Perl M, Levy KY, Belikov J, Mannor S, Levron Y. Explainable artificial intelligence (XAI) techniques for energy and power systems: Review, challenges and opportunities. Energy AI 2022;9:100169.

[20] Jia Q, Kang K, Wang B, Wu Y, Zhang Y, Guo F. Research on wind power prediction with secondary decomposition and multi-algorithm fusion for complex nonlinear time series. Comput Electr Eng 2025;128:110688.

[21] Zhao X, Liu HP, Jin H, Shen X, Ren W. STE-HOLNet: A new method for wind power prediction by integrating spatio-temporal features, dynamic concept drift detection and adaptive correction. Energy Convers Manage 2025;344:120138.

[22] Ye L, Peng Y, Li Y, Li Z. A novel informer-time-series generative adversarial networks for day-ahead scenario generation of wind power. Appl Energy 2024;364:123182.

[23] Wu H, Xu J, Wang J, Long M. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. Adv Neural Inf Process Syst 2021;34:22419–30.

[24] Lin P, Zhang X, Gong L, Lin J, Zhang J, Cheng S. Multi-timescale short-term urban water demand forecasting based on an improved PatchTST model. J Hydrol 2025;651:132599.

[25] Yu S, He B, Fang L. Multi-step short-term forecasting of photovoltaic power utilizing TimesNet with enhanced feature extraction and a novel loss function. Appl Energy 2025;388:125645.

[26] Yi K, Zhang Q, Fan W, Wang S, Wang P, He H, An N, Lian D, Cao L, Niu Z. Frequency-domain mlps are more effective learners in time series forecasting. Adv Neural Inf Process Syst 2023;36:76656–79.

[27] LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proc IEEE 1998;86(11):2278–324.

[28] Cho K, Van Merriënboer B, Gulçehre Ç, Bahdanau D, Bougares F, Schwenk H, Bengio Y. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In: Proceedings of the 2014 conference on empirical methods in natural language processing. EMNLP, 2014, p. 1724–34.

[29] Chen T, Guestrin C. Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. 2016, p. 785–94.

[30] Breiman L. Random forests. Mach Learn 2001;45(1):5–32.

[31] Han K, Xiao A, Wu E, Guo J, Xu C, Wang Y. Transformer in transformer. Adv Neural Inf Process Syst 2021;34:15908–19.

[32] Su C. Dlinear makes efficient long-term predictions. In: Proceedings of ACM conference (baidu KDD cup). 2022.

[33] Tolstikhin IO, Houlsby N, Kolesnikov A, Beyer L, Zhai X, Unterthiner T, Yung J, Steiner A, Keysers D, Uszkoreit J, et al. Mlp-mixer: An all-mlp architecture for vision. Adv Neural Inf Process Syst 2021;34:24261–72.

[34] Suganthan PN, Katuwal R. On the origins of randomization-based feedforward neural networks. Appl Soft Comput 2021;105:107239.

[35] Ren Y, Suganthan PN, Srikanth N, Amaratunga G. Random vector functional link network for short-term electricity load demand forecasting. Inform Sci 2016;367:1078–93.