POWER AND ENERGY FUTURE

Open Access
Research Article

Considering Meteorological Spatio-temporal Propagation and Multi-Factor Coupling Short-Term Wind Power Forecasting

Mao Yang, Yuxi Jiang, and Tao Huang

***Abstract*—Accurate short term power forecasting across multiple wind farms under extreme weather remains challenging because of temporal misalignment between meteorological disturbances and power responses, heterogeneous coupling among meteorological factors, and the limited adaptability of fixed graph topologies. To address these challenges, this paper proposes a spatiotemporal forecasting framework that incorporates physical constraints, meteorological propagation, and multifactor coupling. First, physical operating constraints and temporal trend similarity are integrated to identify abnormal power periods and generate reliable anomaly labels. Second, an anomaly guided self attention module is developed to quantify dynamic interactions among meteorological variables and transform the high dimensional meteorological matrix into a unified meteorological impact feature. Third, time delay attention and a dynamic graph derived from the physical diffusion process are incorporated into a spatiotemporal graph neural network to capture delayed power responses and changing spatial dependencies. Experiments were conducted on 100 wind farms in Western Inner Mongolia. The proposed method achieves a normalized mean absolute error (NMAE) of 5.785% and a normalized root mean square error (NRMSE) of 7.542%, outperforming the comparison models. The results confirm its effectiveness and robustness under extreme weather conditions.**

***Index Terms*—Extreme weather, wind power forecasting, anomaly detection, physical constraints, multi-meteorological-factor coupling, spatiotemporal graph neural network.**

ABBREVIATION

AE	autoencoder
ARIMA	autoregressive integrated moving average
CDDAN	cross domain differential attention network
GCDA	graph contrastive domain adaptation
CNN	convolutional neural network
DBSCAN	density-based spatial clustering of applications with noise
DNN	deep neural network
DTW	dynamic time warping
GAT	graph attention network
GCN	graph convolutional network
GNN	graph neural network
GRU	gated recurrent unit
IQR	interquartile range
ITT	interpretable feature-time transformer
LOF	local outlier factor
LSTM	long short-term memory
MAPE	mean absolute percentage error
NMAE	normalized mean absolute error
NRMSE	normalized root mean square error
NWP	numerical weather prediction
OCSVM	one-class support vector machine
PE-STGNN	physics-enhanced spatiotemporal graph neural network
PINN	physics-informed neural network
r	Pearson correlation coefficient
R^2	coefficient of determination
RS	residual stationarity
SHAP	Shapley additive explanations
SMAPE	symmetric mean absolute percentage error
STGNN	spatiotemporal graph neural network
SVR	support vector regression
TCN	temporal convolutional network
TFT	temporal fusion transformer
WPF	wind power forecasting
XGBoost	extreme gradient boosting

Manuscript received 30 May, 2026; accepted 17 September, 2026. *(Corresponding author: Yuxi Jiang, e-mail: 1202500204@neepu.edu.cn)*

M. Yang, Y. Jiang are with the Key Laboratory of Modern Power System Simulation and Control & Renewable Energy Technology, Ministry of Education, Northeast Electric Power University, Jilin 132012, China.

T. Huang is with the Department of Energy, Politecnico di Torino, Corso Duca degli Abruzzi, 24, Turin, Turin 10129, Italy.

Color versions of one or more of the figures in this paper are available online at https://doi.org/10.26599/PEF.2026.9650019.

Digital Object Identifier 10.26599/PEF.2026.9650019

I. INTRODUCTION

Wind power has become a major source of low carbon electricity, and its growing penetration has increased the operational value of accurate power forecasting [1]. Fore-

© The author(s) 2026. Published by Tsinghua University Press.
This is an open access article under the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/).

cast errors affect reserve scheduling, congestion management, and frequency regulation [2]. These effects become more pronounced during extreme weather, when rapid changes in wind speed, temperature, humidity, and pressure can produce sharp power ramps, turbine derating, or protective shutdowns. Reliable forecasting over short horizons for geographically distributed wind farms therefore requires models that capture temporal dependence within each farm, changing meteorological effects, and weather propagation across the farm cluster.

Wind power forecasting has progressed from statistical and shallow machine learning methods to deep neural networks [3]. Statistical models are efficient but provide limited representations of nonlinear and nonstationary power dynamics. Support vector regression, tree ensembles, and other machine learning methods improve nonlinear fitting, but their static mappings do not fully describe long temporal dependence or interactions among wind farms. Multivariate sequence learning with multitask correction has improved multistep wind power prediction [4], while large time series models provide a new route for transferring general temporal knowledge to wind forecasting tasks [5]. Recurrent networks, temporal convolutional networks, and Transformers also learn multiscale temporal patterns. However, sequence models alone cannot represent the spatial dependence induced by shared weather systems.

Graph neural networks address this issue by treating wind farms as nodes and their spatial relations as edges. Spatiotemporal attention and multigraph convolution have been used to combine power, weather, and station relations [6], [7]. Adaptive graph convolution [8] and globally aware dynamic graph learning [9] further improve topology estimation. Condition dependent modeling of prediction errors has also been shown to improve probabilistic renewable power forecasting by describing the joint effects of space, time, and operating conditions [10]. Although these methods perform well under ordinary conditions, many graphs are derived from geographic distance or long term statistical correlation. Such edges change slowly and may not describe the direction, speed, and attenuation of a rapidly evolving weather system.

Extreme weather also complicates anomaly identification. A large power change may indicate an actual equipment or operational anomaly, but it may also be a physically reasonable response to a rapid change in wind conditions. Existing studies have examined icing losses and turbine operating states using weather clustering and deep temporal models [11], [12], while physics based and energy aware methods have improved the reliability of anomaly detection [13], [14]. Related progress in industrial condition monitoring provides additional insight into model robustness across operating domains. Qin et al. proposed a cross domain differential attention network (CDDAN) that compares multiple known operating domains to learn transferable fault features under unknown conditions [15]. Dai et al. developed a digital twin assisted graph contrastive domain adaptation (GCDA) network that aligns simulated and measured representations for small sample diagnosis [16]. These approaches mainly reduce cross domain feature differences, data distribution shifts, and label scarcity. They do not, however, explicitly distinguish physically plausible wind power ramps from true anomalies, align weather changes with delayed power responses, or model weather propagation among geographically distributed wind farms.

The effects of multiple meteorological variables also vary with weather conditions and operating states. Interpretable temporal models have been developed to extract coupled weather features [17]. Fluctuation classification and feature factor extraction have also improved very short term photovoltaic forecasting [18], indicating the value of condition specific feature construction for variable renewable generation. Shapley additive explanations (SHAP) based analysis [19] and weather informed scenario generation [20] further demonstrate the value of meteorological context and feature interpretation. Nevertheless, most feature weights are learned as global mappings or are estimated independently of abnormal power states. Directly feeding a high dimensional meteorological matrix into the forecasting model can preserve redundant information and weaken generalization when the dominant drivers change rapidly. A compact feature representation that is guided by anomaly states and updated over time is therefore required.

Recent dynamic graph methods have started to represent the direction and lag of wind propagation [21]. Dynamic dual graph learning has improved low output event prediction by updating heterogeneous relations within wind farm clusters [22], while bidirectional interaction between homogeneous and heterogeneous information strengthens cluster level representation [23]. Physics based forecasting and physics informed neural networks also provide constraints that improve operational plausibility and uncertainty awareness [24], [25]. However, dynamic topology learning, meteorological response delay, and turbine operating limits are often considered separately. Consequently, existing models may produce spatial links that lack physical meaning or forecasts that violate feasible operating boundaries during severe weather transitions.

In summary, existing studies provide effective tools for temporal learning, graph modeling, feature interpretation, domain adaptation, and physical regularization, but three issues remain unresolved for multiple wind farms under extreme weather. First, normal large power ramps and true abnormal operation are difficult to separate. Second, the influence of meteorological factors changes across weather states, while weather variations and power responses are temporally misaligned. Third, fixed or weakly constrained graph topologies cannot represent the evolving propagation of meteorological disturbances across wind farms.

To address these issues, the proposed framework establishes an explicit correspondence between each problem and model component. The physics constrained anomaly identification module separates true anomalies from large but physi-

cally plausible power variations. The meteorological impact feature extractor adjusts the importance of weather variables across operating conditions and suppresses redundant inputs. The time delay attention module aligns meteorological changes with delayed power responses. The dynamic physical graph updates inter-farm affinities according to distance attenuation and time-varying meteorological impact features. Finally, the physics constrained loss penalizes forecasts that exceed feasible turbine operating boundaries. Together, these components target complementary error sources within a unified forecasting model.

On this basis, this paper proposes a wind power forecasting framework that considers meteorological spatiotemporal propagation and the coupling of multiple factors. The main contributions are as follows:

(1) A physics constrained anomaly identification method combines turbine operating characteristics with temporal trend similarity. It distinguishes abnormal disturbances from physically reasonable power ramps and generates reliable anomaly labels for subsequent feature learning.

(2) An anomaly guided self attention model dynamically quantifies the effects of multiple meteorological variables. It converts the high dimensional weather matrix into a compact meteorological impact feature, thereby reducing redundancy while retaining condition dependent information.

(3) A physics enhanced spatiotemporal graph neural network, termed PE-STGNN, integrates time delay attention, a dynamic physical graph, and a physics constrained loss. The model jointly represents delayed power responses, meteorological diffusion among wind farms, and feasible operating boundaries, improving forecast accuracy during rapid weather transitions.

II. Anomaly Identification of Power Generation Based on Physical Constraints and Time Series Similarity

By integrating the operational physical mechanisms of wind turbines with the correlation characteristics of time series data, a distributed wind power anomaly identification method combining physical constraint rules with temporal sequence similarity verification is proposed. First, multi-dimensional physical constraint rules are established by leveraging characteristics of power generation and cluster power patterns, enabling preliminary screening of anomalous segments from multiple perspectives. To mitigate the misclassification tendency of single physical rules, window continuity verification and wind speed-power trend similarity checks are subsequently introduced to refine the preliminary screening results and eliminate spurious anomalies. Finally, anomaly visualization across all wind farm stations is completed, and two categories of anomaly label datasets are saved to support subsequent analysis. The proposed method possesses strong physical interpretability, circumvents the inherent deficiencies of purely data-driven models, and demonstrates adaptability across diverse wind farm scenarios.

A. Preliminary anomaly detection with physical constraints

Based on the operational physical mechanisms of wind turbines, multi-dimensional physical constraint conditions are established to filter normal power fluctuations that conform to physical principles, reduce misclassification by data-driven models, and achieve preliminary identification of anomalous intervals. By leveraging multi-dimensional operational parameters including wind speed, active power, temperature, and humidity, multi-dimensional joint physical constraint rules are constructed to capture abnormal power generation states from multiple perspectives.

(a) Power curve constraint: Based on the physical law of wind turbine power curve, a reasonable interval of power-wind speed mapping is constructed

$$P_{\text{valid}}(v) = [P_{\text{curve}}(v) - \Delta P,\ P_{\text{curve}}(v) + \Delta P] \tag{1}$$

where P_{curve} is the theoretical power corresponding to wind speed v, and ΔP is the width of adaptive fluctuation interval. Based on the historical normal operation data of the station, it is adaptively generated by kernel density estimation and dynamically adjusted with wind speed to fit the power fluctuation characteristics of different wind speed segments. $P_{\text{valid}}(v)$ is the reasonable range of power corresponding to wind speed v. When the actual power $P_{\text{actual}} \in P_{\text{valid}}(v)$, it is judged to be normal fluctuation; and when $P_{\text{actual}} \notin P_{\text{valid}}(v)$, the candidate abnormal samples are included.

In order to further quantify the deviation degree of the power curve, the absolute deviation ε_v between the normalized actual power and the theoretical power is calculated.

$$\varepsilon_v = \frac{|P_{\text{M}}(v) - P_{\text{curve}}(v)|}{P_{\text{cap}}} \tag{2}$$

where P_{cap} is the total rated power and P_{M} is the measured power of the wind turbine. Based on the historical normal deviation data, the normal distribution interval of the deviation is determined by kernel density estimation. When the real-time deviation exceeds the interval, it is judged as abnormal power operation.

(b) Mechanical property constraints: For the abnormal change of environmental parameters, when the environmental parameters such as temperature and humidity change sharply in a short time, it is easy to cause the fluctuation of the operating state of the equipment. Therefore, it is necessary to couple the sudden change of environmental parameters with the abnormal power output to broaden the recognition dimension of environmental induced anomalies.

When the ambient temperature reaches or falls below −10 °C, the permissible fluctuation range of wind turbine power output is broadened to ±10% of the rated power, attributable to the effects of increased lubricant viscosity and low-temperature material properties. If the actual output power fluctuation under such conditions exceeds this ±10% boundary, the event is classified as a low-temperature-induced power anomaly.

Under extreme high-wind conditions, the threshold of 20 m/s is used as a high-wind risk trigger in this study rather than a universal turbine cut-out speed. For the investigated data, this operating range is associated with an increased likelihood of protective derating or shutdown. If the actual power under such conditions remains above 5% of the rated power, the event is classified as an anomaly, where 5% of the power threshold is derived from statistical analysis of the measured dataset comprising 100 wind farm stations in this study.

During the startup and shutdown phases of wind turbine operation, power fluctuations must conform to specific criteria. The startup and shutdown phases refer to the process in which power rises from zero to rated power or decreases from rated power to zero, respectively. Rapid power fluctuations are permissible during these phases; however, the fluctuation amplitude must remain within 50% of the rated power per 15-min interval. This fluctuation amplitude threshold is derived from statistical analysis of historical startup and shutdown data from the wind farm stations. Any exceedance of this limit is classified as an anomaly.

(c) Meteorological correlation constraints: With respect to meteorological-power response anomalies, this study identifies turbine response failure anomalies by monitoring the temporal abrupt change characteristics of wind speed. The primary operating conditions of concern include scenarios where wind speed undergoes severe fluctuations without a synchronous power response, or where the directional trends of wind speed and power changes are entirely contradictory. A strong correlation exists between abrupt changes in meteorological factors and power anomalies, on the basis of which meteorological-power response anomaly constraint conditions are constructed. The thresholds for meteorological factor abrupt change amplitudes are derived from statistical analysis of historical meteorological and power correlation data from the wind farm stations, and are specifically defined as a wind speed change exceeding 10 m/s within 15 min and a temperature change exceeding 15 °C within 1 h. When the abrupt change amplitude of meteorological factors in the region of the ith wind farm exceeds the aforementioned thresholds, if the turbine power fluctuation fails to respond synchronously within a reasonable time window, that is, wind speed surges without a corresponding power increase, or temperature drops sharply without a commensurate power decrease, the event is classified as a power response anomaly. Conversely, if the power fluctuation remains synchronized with the meteorological abrupt change and the fluctuation amplitude falls within the physically permissible range, the event is classified as normal fluctuation.

B. Timing similarity

Physical constraint-based preliminary screening enables only the initial identification of anomalies and may still retain spurious anomaly samples. Accordingly, a temporal sequence similarity verification mechanism is introduced, which leverages the temporal continuity of anomalies and the cooperative output characteristics among wind turbines to perform refined filtering and noise reduction on the candidate anomaly samples obtained from preliminary screening. This achieves a complementary integration of physical rules and data temporal features, thereby enhancing the accuracy of anomaly identification.

A sliding neighborhood window is adopted as the basic unit to compute local anomaly density, reinforcing the temporal continuity of anomalous segments in a manner consistent with the practical engineering characteristic that wind power anomalies predominantly manifest as consecutive time intervals. The sliding window size is set to W with a sliding step of one time step, and the anomaly density within the window is defined as follows

$$\rho = K/W \tag{3}$$

where ρ denotes the anomaly density within the window, and K represents the number of preliminarily screened anomalous points flagged within the window. W is set to 10 in this study based on the characteristics of the time series data. When the anomaly density falls below the threshold, the anomalous points within the window are considered to exist in isolation, and the segment is classified as a spurious anomaly attributable to random sensor noise or instantaneous disturbances and is consequently removed. Conversely, when the anomaly density meets or exceeds the threshold, the anomalous points within the window are deemed to exhibit temporal continuity and are retained as valid anomalous segments.

The Pearson correlation coefficient (r) is introduced to quantify the trend similarity between wind speed and power within local sliding windows, enabling the distinction between natural operating condition fluctuations and genuine anomalies. If wind speed and power changes within a flagged anomalous interval exhibit a high degree of synchronization and strong trend coupling, the power fluctuation is determined to be essentially a normal physical response driven by meteorological factors, and the anomaly label is accordingly revoked. Only instances in which power is decoupled from meteorological parameters and the operational pattern severely deviates from the normal range are retained as genuine anomalies. The Pearson correlation coefficient is computed at the wind farm station level to perform an overall trend verification.

The Pearson correlation coefficient between the average wind speed sequence $\{v_1, v_2, \cdots, v_W\}$ and the aggregated power sequence $\{P_{\text{M},1}, P_{\text{M},2}, \cdots, P_{\text{M},W}\}$ in the sliding window is calculated. The formula is as follows

$$r_t = \frac{\sum_{t=1}^{W}(v_t - \overline{v})(P_{\text{M},t} - \overline{P}_{\text{M}})}{\sqrt{\sum_{t=1}^{W}(v_t - \overline{v})^2}\sqrt{\sum_{t=1}^{W}(P_{\text{M},t} - \overline{P}_{\text{M}})^2}} \tag{4}$$

where $\overline{v}$ and $\overline{P}_{\text{M}}$ denote the mean wind speed averaged across

清華大學出版社 Tsinghua University Press | SciOpen

all stations within the window and the mean aggregated power of the wind farm stations within the window, respectively. The similarity determination interval is adaptively established based on the distribution characteristics of r_t under historical normal operating conditions. When r_t falls within the normal similarity interval, the overall output of the wind farm stations is considered to be highly synchronized with variations in mean wind speed, and the event is classified as a natural operating condition fluctuation, whereupon the anomaly label is removed. When r_t falls within the anomalous similarity interval, the anomaly label is retained and the anomalous range is further confirmed in conjunction with the results of the unit cooperative anomaly density verification.

Through the dual optimization of local window continuity verification and temporal sequence similarity verification, spurious anomaly samples are precisely eliminated: Isolated anomalous points and spurious anomalies attributable to natural operating condition fluctuations are removed, while genuine anomalies characterized by temporal continuity and decoupling between power output and meteorological parameters are retained. The refined and optimized anomaly label dataset A is ultimately output, with the anomalous conditions of each wind turbine clearly annotated, thereby providing comprehensive and reliable data support for subsequent anomaly traceability and mechanism investigation.

III. Self-Attention-Based Extraction for Influence Features of Meteorological Factors

To precisely capture the differential influences of various meteorological factors on wind farm power generation while comprehensively accounting for multi-faceted coupling effects, this study proposes a self-attention-based multi-meteorological factor influence degree feature extraction method. By constructing multi-dimensional meteorological features, partitioning weather scenarios, computing power deficit degrees, and leveraging self-attention mechanisms to automatically learn the importance weights of individual meteorological factors, a meteorological influence degree score is ultimately derived for each time step, providing interpretable and strongly correlated features for subsequent power forecasting models.

A. Correlation modeling of multi-source meteorology and power generation

The meteorological influence feature is a dimensionless composite descriptor of meteorological conditions and associated deviations in wind power output. One scalar is constructed for each wind farm at each time step by combining an attention-based representation of meteorological histories with a power deficit indicator. Its physical motivation comes from the selected weather variables, operating constraints used for anomaly identification, and deviations from a reference power trajectory.

(a) Weather scenario partitioning: There are significant differences in the changes of meteorological factors and their effects on power generation under different weather scenarios. To precisely quantify the influence weights of meteorological factors under each scenario, three binary weather-condition indicators are constructed and mapped to a categorical scenario label $S_t \in \{I_1, I_2, I_3, I_4\}$ are constructed based on real-time data and dynamic variation characteristics of wind speed, temperature, and humidity, providing a scenario partitioning basis for subsequent dynamic weight learning

$$I_{1,i,t} = I(v_{i,t} \geq 20 \cup \Delta v_{i,t} \geq 8) \tag{5}$$

$$I_{2,i,t} = 2I(T_{i,t} \leq -10)(1 - I_{1,i,t}) \tag{6}$$

$$\begin{aligned} I_{3,i,t} = 3I(&H_{i,t} \geq 70\% \cap \Delta v_{i,t} \geq 2.5 \cap \\ &\Delta T_{i,t} \geq 1.2 \cap v_{i,t} \geq 3)(1 - I_{1,i,t}) \end{aligned} \tag{7}$$

where $I(\cdot)$ is a binary indicator function that equals 1 when its condition is satisfied and 0 otherwise; v, T, and H represent wind speed, temperature, and humidity, respectively.

(1) Extreme strong wind scenario $I_{1,i,t} = 1$: When the real-time wind speed is greater than or equal to 20 m/s, or the instantaneous change of wind speed is greater than or equal to 8 m/s, it is judged to be an extreme strong wind scenario. Under such scenarios, wind speed exerts the most pronounced influence on power, and is prone to inducing turbine shutdown or abrupt power drops.

(2) Extreme low-temperature scenario $I_{2,i,t} = 2$: When the real-time temperature falls below −10 °C and the conditions for the extreme high-wind scenario are not satisfied, the scenario is classified as an extreme low-temperature scenario. Under such conditions, low temperatures may cause blade icing, adversely affecting turbine output efficiency.

(3) Convective weather scenario $I_{3,i,t} = 3$: When the humidity is greater than or equal to 70%, the instantaneous wind speed variation exceeds 2.5 m/s, the instantaneous temperature variation exceeds 1.2 °C, the real-time wind speed exceeds 3 m/s, and the conditions for the extreme high-wind scenario are not satisfied, the scenario is classified as a convective weather scenario. Under such conditions, multiple meteorological factors undergo cooperative variations, resulting in complex and compound influences on power generation.

(4) Normal weather scenario $I_{4,i,t} = 0$: All conditions that do not satisfy the criteria of the aforementioned three scenario categories are classified as normal weather scenarios. Under such conditions, meteorological factors undergo gradual and moderate variations, and their influence on power generation remains relatively stable.

Through weather scenario partitioning, the dominant roles of meteorological factors under different scenarios can be clearly delineated, providing a basis for weight allocation in the subsequent computation of power generation influence losses and ensuring the specificity of influence degree quantification.

(b) Power deviation and anomaly indicators: The reference power shortfall and power deficit indicator associate the meteorological scenario with deviations in power output and identified operating anomalies. They provide operational context for feature extraction without assigning the deviation exclusively to meteorological causes.

(1) Reference power shortfall: A reference trajectory is obtained by uniform filtering of observed power. The positive relative shortfall is weighted by weather scenario and clipped to [0,1] as follows

$$L_t(t) = \mathrm{clip}\left(\lambda_{s(t)} \cdot \mathrm{clip}\left(\frac{P_{\mathrm{normal}}(t) - P_{\mathrm{M}}(t)}{P_{\mathrm{normal}}(t)}, 0, 1\right), 0, 1\right) \quad (8)$$

where $\lambda_{s(t)}$ is the weight assigned to weather scenario $s(t)$, and $P_{\mathrm{normal}}(t)$ denotes a reference power trajectory obtained by uniform filtering of the measured power series. The operator clip(·,0,1) restricts the result to [0,1]. Thus, $L_t(t)$ measures the weighted power shortfall relative to this statistical reference. The reference does not necessarily represent the theoretical power output in the absence of meteorological disturbances.

(2) Power deficit indicator: The power shortfall index is combined with the anomaly label to characterize the defined operating state. The resulting dimensionless indicator is calculated as follows

$$L_{\mathrm{def}}(t) = \mathrm{clip}\left(\omega \cdot L_t(t) + (1-\omega) \cdot A(t), 0, 1\right) \quad (9)$$

where $\omega \in [0,1]$ controls the relative contributions of $L_t(t)$ and the binary anomaly label $A(t)$. Here, $A(t) = 1$ indicates an identified anomaly, and $A(t) = 0$ indicates no identified anomaly. The resulting $L_{\mathrm{def}}(t)$ summarizes the reference power deficit and anomaly information. Its interpretation depends on the reference trajectory, scenario weights, and labeling procedure; it does not directly quantify the power loss caused by meteorological factors.

(c) Meteorological feature construction: Eight inputs describe meteorological states, temporal changes, trends, and their coincidence with power anomalies. Their construction is detailed below:

(1) Static meteorological features: comprising real-time wind speed, real-time temperature, and real-time humidity, these features directly reflect the fundamental state of current meteorological conditions and constitute the core baseline features influencing power generation output.

(2) Dynamic variation features: comprising the first-order difference of wind speed and the first-order difference of temperature, these features characterize the instantaneous variation intensity of meteorological factors and reflect the influence of dynamic fluctuations in meteorological conditions on power generation output.

(3) Trend features: comprising the one-dimensional uniform filter value of wind speed with a window size of 5 and the absolute deviation between real-time wind speed and the filtered value, these features capture the long-term variation trend of wind speed and the degree of deviation between instantaneous fluctuations and the underlying trend, conforming to the response pattern of wind power output to wind speed trends.

(4) Synergistic features: constructed based on the cooperative relationship between instantaneous wind speed variations and anomaly labels, these features take a value of 1 when the instantaneous wind speed variation exceeds 1.5 m/s and the operating condition is classified as anomalous, and 0 otherwise, serving to capture the synergistic interaction between abrupt meteorological changes and power generation anomalies.

The resulting meteorological feature matrix $X \in R^{T\times N\times M}$ contains T time steps, N wind farms, and M input features. It supplies the weather state and temporal variability information for the attention-based feature extraction model.

B. Multi-meteorological factor coupling influence extraction model based on self-attention

The extraction model uses anomaly guided self attention to encode dependencies in the multivariate meteorological time series. Its scalar projection is combined with the power deficit indicator to obtain a compact feature associated with the defined operating state.

Multi-source meteorological feature matrix $X \in R^{T\times N\times M}$: The meteorological feature matrix obtained from the preceding screening process, encompassing instantaneous features, temporal sequence features, extreme condition features, and anomaly-correlated features, where M denotes the feature dimensionality and T denotes the number of time steps, reflecting the multi-meteorological factor states across different time steps.

Anomaly correlated vector $Y \in R^{T\times 2}$: The anomaly labels $A(t)$ and power deficit indicator $L_{\mathrm{def},t}$ are combined as $Y_t = [A_t, L_{\mathrm{def},t}]^T$. The vector Y supplies operating state information to the attention layer, linking the meteorological representation to identified power anomalies.

(a) Self-attention feature extraction layer: The layer aggregates information across time using meteorological similarities and anomaly guidance. Its coefficients determine how temporal information is combined in the representation.

(1) Anomaly guided attention: The anomaly correlated vector Y enters the attention score so that the encoded temporal associations can reflect the identified operating anomalies.

The fused input matrix X_{in} is mapped into query matrix $Q \in R^{T\times d}$, key matrix, and value matrix $V \in R^{T\times d}$ through three independent linear transformation matrices W_{Q}, W_{K}, and W_{V}, respectively, where d denotes the feature dimensionality after mapping, achieving linear transformation and dimensionality unification of the features.

The weight factor β of the anomaly-correlated vector Y is introduced to incorporate anomaly information into the attention score computation, correcting the limitation of conventional self-attention mechanisms that focus solely on the features themselves. The calculation formula for the attention score matrix S is as follows

$$S = \frac{QK^T}{\sqrt{d}} + \beta \cdot YY^T \tag{10}$$

where β denotes the anomaly guidance coefficient, whose value is determined through optimization on the training set; YY^T denotes the autocorrelation matrix of the anomaly-correlated vector, employed to reinforce the associations among anomaly information across different time steps and guide the model to focus on meteorological features during anomalous intervals; and $\sqrt{d}$ denotes the scaling factor, where d is the feature mapping dimensionality. This factor is introduced to mitigate the dimensionality curse of attention scores, ensuring the reasonableness of attention scores and preventing saturation of the Softmax function caused by excessively large score values.

The Softmax function normalizes each row of S to obtain the attention matrix $A \in R^{T\times T}$. Each entry $A(i, j)$ weights information from time j when constructing the representation at time i. These coefficients describe information aggregation within the model and do not directly measure the physical or causal contribution of an individual meteorological variable.

(2) Multi-scale feature fusion: To account for the influences of meteorological features at different scales, including instantaneous abrupt changes and long-term fluctuations, two self-attention layers are constructed in this study. Batch normalization and residual connections are incorporated after each self-attention layer to prevent model overfitting and reinforce the multi-scale feature fusion capability. Specifically, the first self-attention layer focuses on capturing the associations between instantaneous meteorological features and power anomalies, while the second layer concentrates on the long-term dependency relationships between temporal meteorological features and power anomalies. The features from both layers are subsequently effectively integrated through residual connections, thereby enhancing the model's comprehensive perceptual capability with respect to multi-scale meteorological influences and yielding the multi-scale self-attention feature $Z \in R^{T\times d}$.

(b) Design of feature fusion layer and output layer: The self-attention feature Z and the power deficit degree vector $\text{Loss}_{\text{deficit}}$ are fused. First, the self-attention feature Z is reduced to one dimension through a fully connected layer, yielding the preliminary meteorological influence feature $Z_1 \in R^{(T\times 1)}$. A fixed fusion coefficient α_t is subsequently introduced to perform weighted fusion of Z_1 and L_{def}, producing the fused feature $F_{\text{final}} \in R^{(T\times 1)}$. The calculation formula is as follows

$$F_{\text{final}} = \text{MinMaxNorm}\left(\alpha_t \cdot Z_{1,t} + (1-\alpha_t)\cdot L_{\text{def},t}\right) \tag{11}$$

where α_t is the fusion coefficient obtained from the row mean of the normalized attention matrix A. For a fixed sequence length T, this coefficient equals $1/T$. The projected feature $Z_{1,t}$ encodes meteorological dependencies, whereas $L_{\text{def},t}$ incorporates reference power shortfalls and identified anomalies. For each farm, $F_{\text{final}} \in [0, 1]^{T\times 1}$ contains one dimensionless scalar per time step. It summarizes joint weather and operating conditions relative to the chosen reference trajectory and labeling rules. Normalization provides a common numerical range; the

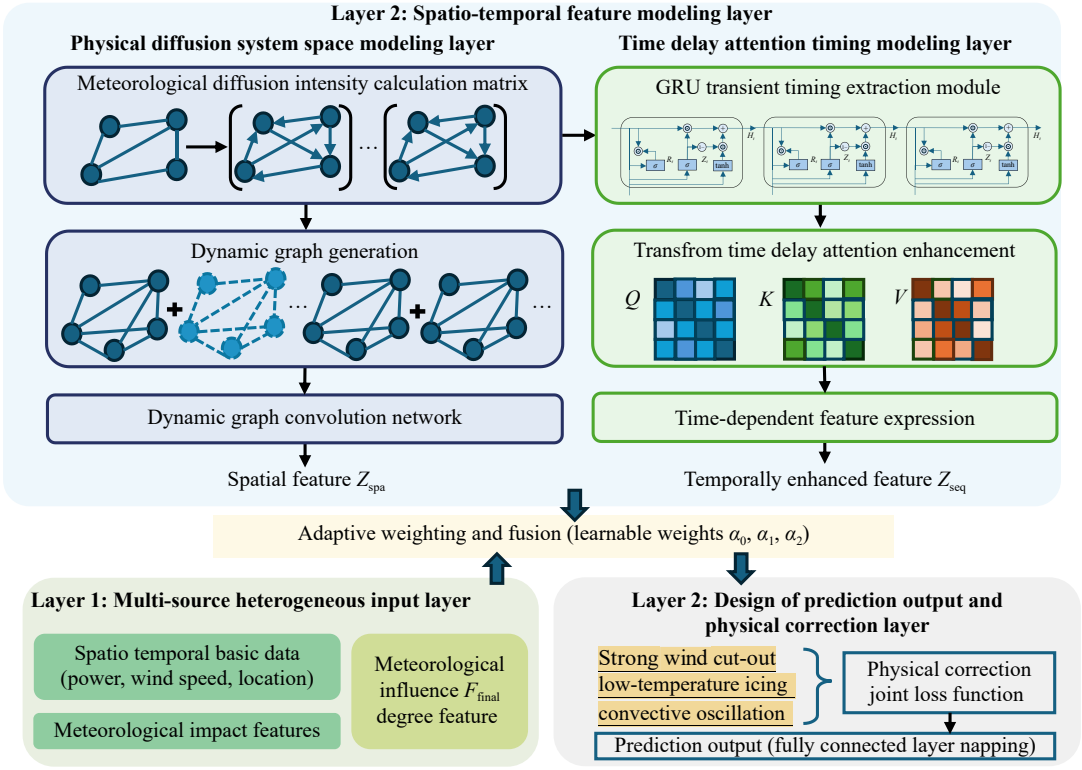

Fig. 1. The structure of forecasting model.

value is neither a calibrated anomaly probability nor an actual weather related loss percentage. A larger score need not imply a greater causal effect or a monotonically increasing power deficit. The feature supplements wind speed in forecasting and modulates spatial graph edge weights.

IV. Forecasting Model

Under extreme weather conditions, wind power output exhibits strong fluctuation, nonlinearity, and complex spatiotemporal correlation characteristics attributable to the coupled influences of multiple meteorological factors, making it difficult for conventional models to accurately capture the underlying relationships, thereby leading to degraded forecasting accuracy. To this end, this chapter designs a PE-STGNN for wind power forecasting under extreme weather conditions, the architecture of which is illustrated in Fig. 1. The model comprehensively exploits the spatial correlations, temporal dependencies, and meteorological coupling influence mechanisms of wind power output, synergistically combining the advantages of physical mechanisms and data-driven approaches to accommodate multi-station extreme weather forecasting scenarios. With the meteorological influence features extracted in the preceding chapter as its core input, the model incorporates anomaly labels and wind farm spatiotemporal data, employs a temporal delay attention layer to capture dynamic time delay effects, constructs a dynamic graph structure through a physical diffusion system to extract spatial diffusion characteristics, achieves deep fusion of temporal sequences, spatial information, meteorological influences, and anomaly labels through a feature fusion layer, and ultimately outputs power forecasting results via a fully connected layer, ensuring that the model conforms to physical reality while remaining adaptive to the dynamic variations induced by extreme weather conditions.

A. Physical diffusion system space modeling layer

Based on the spatial diffusion physical principles of meteorological factors, a dynamic graph structure is constructed to transcend the limitations of conventional fixed graph structures, enabling precise capture of spatial correlations and meteorological diffusion pathways among wind farms and extraction of high-quality spatial features.

(a) Dynamic graph structure construction: The graph structure consists of nodes and edges, where each node represents an individual wind farm and the edge weights are determined by the physical diffusion intensity. Based on the meteorological influence features and the wind farm distance matrix, the meteorological diffusion intensity $W(n, m)$ between any two wind farms n and m is computed according to the following formula

$$W(n,m) = \exp(-D(n,m)/L_{\mathrm{d}}) \times \frac{F_{\mathrm{final}}(t,n) + F_{\mathrm{final}}(t,m)}{2} \quad (12)$$

where $D(n, m)$ denotes the geographical distance between farms n and m, and L_{d} controls the distance decay. The term $\exp(-D(n, m)/L_{\mathrm{d}})$ gives greater weight to nearby farms, while $(F_{\mathrm{final}}(t, n) + F_{\mathrm{final}}(t, m))/2$ adjusts the weight according to their mean meteorological influence feature. Thus, $W(n, m)$ is a dimensionless affinity coefficient for graph message passing, rather than a measurement of physical meteorological flux.

As F_{final} changes over time, the edge weights $W(n, m)$ are updated to form $G_t = (V, E_t)$, where V and E_t denote the nodes and the edges at time t, respectively. This construction represents changing inter-farm affinities under distance attenuation. The symmetric edge weight in Eq. (12) does not independently identify propagation direction or speed.

(b) Spatial feature extraction: Dynamic graph convolution is adopted as the spatial feature extraction unit. Based on the dynamic graph structure G_t, spatial feature extraction is performed on the temporally enhanced features Z_{spa}. Through a graph attention layer, the attention weights between each node and its neighboring nodes are computed to capture the spatial cooperative relationships among wind farms, and the temporal features of each node are weighted and fused with the features of neighboring nodes to yield the spatial feature $Z_{\mathrm{spa}} \in R^{(T\times N\times D)}$.

B. Time delay attention timing modeling layer

By integrating temporal delay attention mechanisms with gated recurrent unit (GRU) gated temporal modeling, the proposed approach enables precise capture of the time delay effects and temporal fluctuation patterns of wind power output under extreme weather conditions, effectively addressing the challenges faced by conventional temporal modeling approaches in handling long-duration delay dependencies and instantaneous abrupt changes. To dynamically characterize the time delay relationship between meteorological propagation and power response, the model transcends the limitations of conventional fixed delay step sizes by adaptively learning delay weights across different time steps and different wind farms, thereby accurately capturing temporal correlations.

(a) Transient temporal extraction module: The GRU module serves as the foundation for temporal modeling, employed to preliminarily capture the temporal dependency relationships of power output, particularly the abrupt change characteristics and long-term fluctuation patterns of power under extreme weather conditions. Through its gating mechanism, the module controls the transmission and forgetting of features, accommodating the modeling requirements of time delay effects:

(1) Reset gate and update gate computation: The current input features and the hidden state from the previous time step are concatenated and transformed through a linear layer, and the reset gate and update gate are obtained via the Sigmoid activation function, as formulated below

$$ru = \sigma(\mathrm{ff}(\mathrm{cat}([x, h]))) \quad (13)$$

$$r, u = \mathrm{chunk}(ru, 2, -1) \quad (14)$$

清華大學出版社 Tsinghua University Press | SciOpen

where σ denotes the Sigmoid activation function, ff denotes the linear transformation layer, cat($[x, h]$ denotes the concatenation of the current input feature x and the hidden state from the previous time step h, and the cat($[x, h]$ function is employed to decompose the transformed result into the reset gate r_t and the update gate z_t. The update gate controls the degree of retention of the hidden state from the previous time step, adapting to the influence weights of different time delays.

(2) Candidate hidden state computation: Combining the reset gate with the current input feature and the hidden state at the previous moment, the candidate hidden state is obtained through the linear layer and the activation function, and the time delay correlation between the current moment and the historical moment is captured. The formula is as follows

$$\widehat{h} = \tanh(z(\text{cat}([x, r \times h]))) \tag{15}$$

where tanh is the activation function, z is the linear transformation layer, $r \times h$ is the adjustment of the reset gate to the hidden state of the previous moment, and the historical time delay characteristics related to the current moment are screened out.

(3) Hidden state update: The hidden state of the current moment is obtained by fusing the hidden state of the previous moment and the candidate hidden state through the update gate

$$h = u \times \widehat{h} + (1 - u) \times h \tag{16}$$

where u is the update gate weight, which is used for the contribution of the current time feature and the historical time delay feature to realize the transmission and accumulation of time delay dependence.

(b) Time-delayed attention enhancement

On the basis of GRU time series modeling, the time delay attention mechanism is introduced to further strengthen the capture ability of time delay dependence. At the same time, combined with the meteorological attenuation mechanism, the time series feature changes under extreme weather are adapted.

(1) Position coding injection: The sine and cosine position information is injected into the feature through the position coding matrix to ensure the order of the model recognition time.

The position coding is generated based on sine and cosine functions. The even number dimension adopts sine function, and the odd number dimension adopts cosine function. The formula is

$$pe[:, 0 :: 2] = \sin(\text{pos} \times \text{div}), pe[:, 1 :: 2] = \cos(\text{pos} \times \text{div}) \tag{17}$$

$$\text{div} = \exp\left(n \cdot \frac{-\ln(100000)}{F}\right) \tag{18}$$

where pos denotes the time step position, div denotes the decay coefficient, F denotes the feature dimensionality, and $n = [0, 2, 4, \cdots, F - 2]$ is a sequential vector. The positional encoding differentiates between different time steps, providing a positional basis for the computation of time delay associations.

(2) Delay-aware attention: Calculate the scale dot product attention score scores, and introduce the meteorological attenuation factor to correct the score.

Through the linear layer, the position-encoded features are divided into a query matrix, a key matrix, and a value matrix. The query matrix corresponds to the current moment feature, and the key matrix corresponds to the historical moment feature, which is used to calculate the delay correlation weights of different time steps. The formula is as follows

$$Q, K, V = \text{chunk}(qkv(x_{\text{pe}}), 3, \text{dim} = -1) \tag{19}$$

where x_{ep} denotes the features after positional encoding, qkv denotes the linear transformation layer, and the chunk function is employed to split the transformed features along the last dimension into the query matrix Q, key matrix K, and value matrix V.

The attention score is calculated by scaling dot product attention, and the scaling factor is used to alleviate the dimension disaster. The attention score directly reflects the time delay correlation strength between different historical time steps and the current moment.

$$\text{scores} = \frac{Q \times K^T}{\sqrt{D}} \tag{20}$$

where K^T denotes the transpose of the key matrix, D denotes the feature dimensionality, and the attention score matrix scores contains elements corresponding to the association strength between different time steps.

By using the meteorological-guided delayed attention adjustment, the meteorological impact factor is introduced to generate the attenuation coefficient, which is multiplied by the attention score. The formula is as follows

$$\text{met}_{\text{decay}} = \exp(-|F_{\text{final}}|) \cdot I_{\text{d}} \tag{21}$$

$$\text{scores}_{\text{adj}} = \text{scores} \times \text{met}_{\text{decay}} \tag{22}$$

where F_{final} denotes the meteorological influence factor, $\exp(-|F_{\text{final}}|)$ denotes the meteorological attenuation factor, I_{d} is employed to adjust the dimensionality to match that of the attention scores, and $\text{scores}_{\text{adj}}$ denotes the adjusted temporal delay attention scores.

The attention score is normalized by the softmax function, and the attention output is obtained by multiplying with the value matrix. Combined with residual connection, layer normalization and feedforward network, the stability of the model is improved, and over-fitting is avoided. Finally, the time series characteristics after strengthening the time delay dependence are output. The formula is as follows

$$\text{attn} = \text{softmax}(\text{scores}_{\text{adj}}, \text{dim} = -1) \tag{23}$$

$$\text{out}_{\text{attn}} = \text{attn} \times V \tag{24}$$

$$x_1 = \ln1(x_{\text{pe}} + \text{out}_{\text{attn}}) \tag{25}$$

$$\text{seq} = \ln2(x_1 + \text{ff}(x_1)) \tag{26}$$

where softmax is applied to normalize the attention scores to obtain the attention weights attn, out_{attn} denotes the attention output, and ln1 and ln2 denote layer normalization operations. Finally, through softmax normalization and residual connections, the temporally enhanced feature Z_{seq} that reinforces time delay dependencies is output.

C. Design of prediction output and physical correction layer

The meteorological influence factor is mapped into multi-dimensional features through a linear layer to align its dimensionality with that of the temporal delay features and spatial diffusion features. On this basis, learnable weight parameters are introduced to perform weighted fusion of the three categories of features, yielding the fused feature. The specific formula is as follows

$$\text{fused} = \alpha_0 \cdot Z_{\text{seq}} + \alpha_1 \cdot Z_{\text{spa}} + \alpha_2 \cdot \text{met}_{\text{emb}} \tag{27}$$

where Z_{spa} denotes the spatial correlation features output by the physical diffusion dynamic graph convolution, Z_{seq} denotes the temporal delay features output by the temporal delay attention layer, met_{emb} denotes the linear embedding features of the meteorological influence degree feature F_{final}, and α_0, α_1, α_2 are learnable weighting coefficients employed to adaptively regulate the respective contributions of the temporal delay features, spatial diffusion features, and meteorological features, respectively, achieving optimal fusion of multi-source features and highlighting the comprehensive influences of spatiotemporal synergy and meteorological guidance under extreme weather conditions. A preliminary forecast is subsequently obtained through linear mapping

$$P_{\text{p}} = \text{Linear}(\text{fused}) \tag{28}$$

The supplementary module to improve the rationality of the model prediction is used to avoid the situation that the prediction results violate the physical laws of wind power operation, especially in extreme weather, to ensure that the prediction results are in line with the actual operation scenarios.

$$\text{loss} = \text{MSE}(P_{\text{p}}, P_{\text{M}}) + \lambda_{\text{phy}}.\text{phy}_{\text{scene}} \tag{29}$$

where P_{p} denotes the model-predicted power values, P_{M} denotes the actual power values, and MSE( ) is employed to measure the deviation between predicted and actual values. λ_{phy} denotes the physical constraint coefficient, used to regulate the weight of the physical constraint term. $\text{phy}_{\text{scene}}$ denotes the physical constraint term, which incorporates the weather scenario label $S_t \in \{0, 1, 2, 3\}$corresponds to normal weather conditions. Through scenario masking, the cooperative evolution of multiple meteorological factors is transformed into multi-level loss weighting:

$$\text{phy}_{\text{scene}} = \sum_{k=1}^{3} I(S_t = k) \cdot w_k \cdot f_k(P_{\text{p}}) \tag{30}$$

where w_1, w_2, and w_3 denote the physical weights assigned to the three scenarios within the scenario penalty term $\text{phy}_{\text{scence}}$ respectively; f_1, f_2 and f_3 denote the strong wind cut-out protection function, the low-temperature deviation constraint function, and the convective dynamic oscillation limitation function, respectively; and P_{p} denotes the predicted output power of the ith wind farm station at time step t as produced by the model.

$$f_1(P_{\text{p}}) = I(v_t \geq 25) \cdot \left[\text{ReLU}\left(P_{\text{p}} - 0.05P_{\text{cap}}\right)\right]^2 + I(\Delta v_t \geq 8) \cdot \left[\text{ReLU}\left(\Delta P_t - 0.05P_{\text{cap}}\right)\right]^2 \tag{31}$$

$$f_2(P_{\text{p}}) = I(T_t \leq -10) \cdot \left[\text{ReLU}\left(|P_{\text{p}} - P_{\text{M}}| - 0.1P_{\text{cap}}\right)\right]^2 \tag{32}$$

$$f_3(P_{\text{p}}) = \left[\text{ReLU}\left(|\Delta P_t| - 0.3P_{\text{cap}}\right)\right]^2 \tag{33}$$

where Δ denotes the first-order forward temporal difference operator and P_{cap} denotes the rated capacity of the wind turbine.

The strong wind cut-out protection function is activated when the real-time wind speed exceeds the cut-out threshold; if the predicted power under such conditions exceeds 5% of the rated capacity, a quadratic penalty is applied. When the instantaneous wind speed increment exceeds the prescribed threshold, the function penalizes positive power increments exceeding 5% of the rated capacity. The low-temperature deviation constraint function monitors the absolute deviation of the current predicted power from the corresponding measured power; once the deviation amplitude exceeds 0.1 times the rated capacity, the function is activated and generates a penalty gradient. The convective dynamic oscillation limitation function constrains the short-term transient power oscillations induced by abrupt changes in atmospheric pressure and humidity under severe convective weather conditions; based on aerodynamic physical characteristics, the rate of dynamic abrupt change in temporal sequences should not be excessively large. When the power variation amplitude $|\Delta P_t|$ between adjacent temporal sampling points exceeds 30% of the rated capacity, the function triggers a penalty.

V. Overall Method

Aiming at the technical problems of short-term wind power prediction of multi-stations in extreme weather, such as meteorological-power timing dislocation, spatio-temporal heterogeneity of meteorological factor coupling driving mechanism, and solidification of spatial topology modeling, this paper constructs a multi-factor coupling short-term wind power prediction framework integrating meteorological spatio-temporal propagation mechanism. The overall method is shown in Fig. 2.

(1) Wind power anomaly identification method integrating physical constraints and temporal similarity.

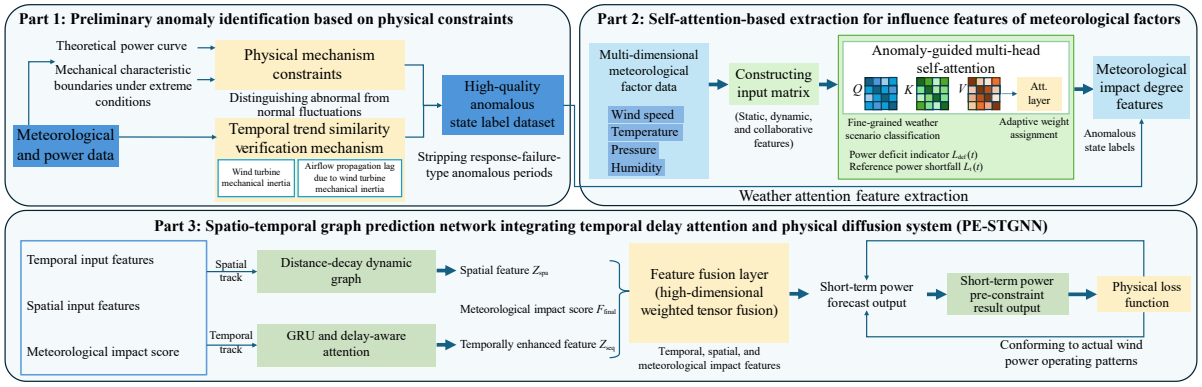

Fig. 2. Schematic of the proposed method.

Aiming at the problem that it is difficult to distinguish the true and false fluctuations of wind power under extreme conditions, and the timing lag is easy to cause label misjudgment, this paper constructs a collaborative verification mechanism of physical mechanism and timing characteristics. Based on the power characteristics of the wind turbine and the extreme operating boundary, the physical constraints are constructed to distinguish the abnormal meteorological disturbance from the normal output fluctuation of the unit. The timing phase deviation between weather and power is corrected by timing similarity. The pseudo-abnormal samples are eliminated by double verification, and a high-confidence abnormal data set is constructed to provide high-quality data support for wind power prediction modeling in extreme scenarios.

(2) Meteorological coupling feature extraction model based on multi-head self-attention.

Aiming at the problem that the coupling relationship of multi-meteorological factors is complex and the fixed weight cannot adapt to the difference of dynamic scenes, this paper constructs an unsupervised multi-head self-attention feature extraction model. The multi-dimensional meteorological time series data are fused to adaptively learn the dynamic influence weights of each meteorological factor in different periods and extreme conditions, and the coupling evolution law of multi-meteorological factors is excavated, and the dynamic refined meteorological coupling characteristics are output to provide a high-adapted meteorological prior representation for the spatio-temporal prediction model.

(3) Dynamic spatio-temporal graph prediction network combining time-delay attention and physical diffusion mechanism.

Aiming at the defects that the traditional static graph model cannot adapt to the spatio-temporal diffusion of extreme weather and is difficult to capture the time-series hysteresis characteristics, this paper constructs a physically enhanced dynamic spatio-temporal prediction network. At the temporal level, the gated unit and time delay attention are combined to accurately capture the transient characteristics of power sudden change in extreme weather and align the hysteresis effect of meteorological propagation. At the spatial level, combined with the meteorological diffusion mechanism, the topological structure is dynamically updated according to the location of the station and the real-time meteorological influence, and the spatial correlation adaptive modeling is realized. The model integrates multi-dimensional features and introduces physical constraint loss to correct the deviation, which is suitable for wind power operation.

VI. Results And Discussion

A. Dataset introduction and evaluation metrics

The experiments were conducted using synchronized meteorological and operational data collected from 100 onshore wind farms in Western Inner Mongolia. The dataset covers the period from January 1 to December 29, 2020, with a temporal resolution of 15 min, corresponding to 34,944 sampling times before data quality screening. The meteorological inputs include numerical weather prediction (NWP) wind speed, relative humidity, air temperature, and measured wind speed, while measured active power is used as the prediction target. All variables were aligned according to their timestamps, and records containing invalid or missing values were excluded. To prevent information leakage, the valid data were divided chronologically into training and test subsets at a ratio of 7:3. The first 70% of the records were used for model training, and the remaining 30% were used for testing. This study is a deterministic forecast, which is evaluated by criteria including NMAE, NRMSE, mean absolute percentage error (MAPE), symmetric mean absolute percentage error (SMAPE), coefficient of determination (R^2), and r

$$E_{\mathrm{RMSE}} = \frac{1}{P_{\mathrm{cap}}}\sqrt{\frac{1}{N}\sum_{i=1}^{N}(P_{\mathrm{M}i}-P_{\mathrm{P}i})^2} \tag{34}$$

$$E_{\mathrm{MAE}} = \frac{1}{NP_{\mathrm{cap}}}\sqrt{\frac{1}{N}\sum_{i=1}^{N}(P_{\mathrm{M}i}-P_{\mathrm{P}i})^2} \tag{35}$$

$$\mathrm{MAPE} = \frac{100\%}{N}\sum_{i=1}^{N}\left|\frac{P_{\mathrm{M}i}-P_{\mathrm{P}i}}{P_{\mathrm{M}i}}\right| \tag{36}$$

$$E_{\mathrm{SMAPE}} = \frac{100\%}{N}\sum_{i=1}^{N}\frac{|P_{\mathrm{M}i}-P_{\mathrm{P}i}|}{(|P_{\mathrm{M}i}|+|P_{\mathrm{P}i}|)/2} \tag{37}$$

$$R^2 = 1-\frac{\sum_{i=1}^{n}(P_{\mathrm{M}i}-P_{\mathrm{P}i})^2}{\sum_{i=1}^{n}\left(P_{\mathrm{M}i}-\overline{P}_{\mathrm{M}}\right)^2} \tag{38}$$

$$r = \frac{\sum_{i=1}^{n}\left(P_{\mathrm{M}i}-\overline{P}_{\mathrm{M}}\right)\left(P_{\mathrm{P}i}-\overline{P}_{\mathrm{P}}\right)}{\sqrt{\sum_{i=1}^{n}\left(P_{\mathrm{M}i}-\overline{P}_{\mathrm{M}}\right)^2}\sqrt{\sum_{i=1}^{n}\left(P_{\mathrm{P}i}-\overline{P}_{\mathrm{P}}\right)^2}} \tag{39}$$

where N denotes the total number of predicted sampling points, P_{cap} denotes the installed capacity, $P_{\mathrm{M}i}$ denotes the actual power at time step i, $P_{\mathrm{P}i}$ denotes the predicted power at time step i, and $\overline{P}_{\mathrm{M}}$ and $\overline{P}_{\mathrm{P}}$ denote the mean values of the actual power and predicted power over the evaluation samples, respectively.

A high-confidence reference subset was constructed using measured meteorological variables, power-curve constraints, and operating-state characteristics. Ambiguous samples were excluded from the quantitative evaluation. The selected intervals cover both normal operation and typical abnormal power-response processes. The independently verified labels were used only for quantitative evaluation and were not involved in the formulation of the physical constraints, threshold determination, or anomaly-label refinement. Based on these reference labels, precision, recall, and $F1$-score are calculated as:

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \tag{40}$$

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{41}$$

$$F1 = \frac{2\mathrm{Precision}\times\mathrm{Recall}}{\mathrm{Precision}+\mathrm{Recall}} \tag{42}$$

where TP, FP, TN, and FN denote true-positive, false-positive, true-negative, and false-negative samples, respectively.

For reproducibility, the five indicators used in the radar chart are explicitly defined according to the implemented evaluation procedure. Let $A_t \in \{0, 1\}$ denote the identified anomaly label, P_t^{norm} the normalized measured power, and P_t^{theo} the normalized theoretical power.

(1) Physical consistency measures the proportion of identified anomalies located outside the theoretical power envelope

$$\mathrm{PC} = \frac{\sum_{t=1}^{T}A_t\mathbb{I}\left(\tilde{P}_t < 0.75\tilde{P}_t^{\mathrm{theo}} \text{ or } \tilde{P}_t > 1.25\tilde{P}_t^{\mathrm{theo}}\right)}{\sum_{t=1}^{T}A_t+\varepsilon} \tag{43}$$

(2) Extreme-condition coverage measures the proportion of predefined extreme-condition samples identified as anomalous

$$\mathrm{EC} = \frac{\sum_{t=1}^{T}A_tE_t}{\sum_{t=1}^{T}E_t+\varepsilon} \tag{44}$$

where $E_t = 1$ when the wind speed is no less than 20 m/s or the temperature is no greater than −10 °C, and $E_t = 0$ otherwise.

(3) Pseudo-anomaly removal rate evaluates the temporal continuity of anomaly identification

$$\mathrm{PRR} = 1-\frac{N_{\mathrm{iso}}}{N_{\mathrm{A}}+\varepsilon} \tag{45}$$

where N_{iso} denotes the number of isolated one-step anomaly points whose adjacent samples are both normal, and N_{A} is the total number of identified anomalies.

(4) Trend decoupling intensity measures the reduction in wind-speed–power correlation between normal and anomalous samples

$$\mathrm{TDI} = \min\{1, \max(0, \rho_N-\rho_A)\} \tag{46}$$

where ρ_{N} and ρ_{A} are the Pearson correlation coefficients between wind speed and power for normal and anomalous samples, respectively.

(5) Residual stationarity evaluates the dispersion of the power-curve residuals remaining after anomaly removal

$$\mathrm{RS} = \min\left\{1, \frac{1}{10[\sigma(D_t|A_t=0)+\varepsilon]}\right\} \tag{47}$$

where $D_t = \left|P_t^{\mathrm{norm}}-P_t^{\mathrm{theo}}\right|$ and $\sigma(\cdot)$ denotes the standard deviation. A larger RS indicates a more concentrated residual distribution for the retained normal samples.

All five indicators are dimensionless and scaled to [0, 1], providing complementary diagnostic information on the anomaly identification results.

One sample is defined as one valid record from one wind farm at one sampling time. According to the weather classification criteria defined in Section III-A, the dataset contains 6754 extreme strong wind samples, 330,201 extreme low temperature samples, and 68,531 severe convective weather samples. Their distributions in the training and test subsets are summarized in Table 1. For the deterministic forecasting task, the evaluation metrics employed in this study include NMAE, NRMSE, SMAPE, R^2, and r.

The experimental parameter settings of this study are configured as follows. Wind speed and the meteorological influence degree feature F_{final} constitute a two-dimensional input, from which a temporal tensor of shape (B, 9, 100, 2) is constructed via a sliding window, where $B = 32$ denotes the batch size, $T = 9$ denotes the number of time steps, and $N = 100$ denotes the number of wind farms. The anomaly identification module employs a sliding window of size $W = 10$, with kernel density estimation adopted to determine the physical

TABLE 1
SAMPLE COUNTS UNDER DIFFERENT WEATHER SCENARIOS

Weather scenario	Training samples	Test samples	Total samples
Extreme strong wind	4086	2668	6754
Extreme low temperature	17,405	312,796	330,201
Severe convective weather	53,385	15,146	68,531

constraint thresholds. The meteorological influence quantification module constructs an eight-dimensional multi-scenario meteorological feature vector and employs two layers of anomaly-guided self-attention mechanisms with a feature mapping dimensionality of $d = 32$. In the PE-STGNN forecasting model, the hidden layer dimensionality of both the graph convolutional layer and the GRU layer is set to 32, the number of attention heads is set to 4, the diffusion decay coefficient is set to $L_d = 0.5$, and the physical constraint coefficient is set to $\lambda_{phy} = 1.0$. The training process employs the Adam optimizer with a learning rate of 10^{-3}.

B. Verification of generation power anomaly identification

To validate the effectiveness and superiority of the wind power anomaly identification method integrating physical constraints and temporal sequence similarity proposed in this study, a comprehensive comparative experiment is designed. Traditional statistical methods, including the 3σ criterion, interquartile range (IQR), local outlier factor (LOF), one-class support vector machine (OCSVM), density-based spatial clustering of applications with noise (DBSCAN), sliding 3σ, and time series autoencoder (AE) are selected as baseline methods.

As shown in Fig. 3, the proposed method shows competitive performance across the five diagnostic indicators of pseudo-anomaly rejection rate, trend decoupling intensity, residual stability, physical consistency, and extreme condition coverage. The comprehensive performance is better than other comparison algorithms. It illustrates the diagnostic characteristics of this method in both physical mechanism constraints and meteorological-driven anomaly identification, which can effectively reduce misjudgment and provide high-quality data support for subsequent extreme weather power prediction.

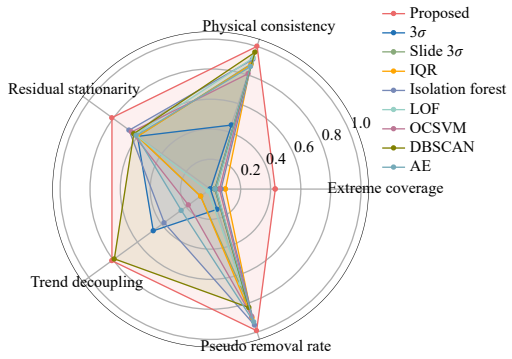

Fig. 3. Comprehensive performance radar charts of different anomaly recognition algorithms.

As shown in Table 2, the proposed method achieves the highest recall and $F1$-score of 0.413 and 0.519, respectively. Although isolation forest obtains a higher precision of 0.834, its recall is only 0.279, indicating that a considerable proportion of anomalous samples are missed. In comparison, the proposed method achieves a better balance between false-alarm suppression and anomaly detection capability, resulting in the best overall $F1$-score among all comparison methods.

TABLE 2
QUANTITATIVE PERFORMANCE OF ANOMALY IDENTIFICATION METHODS

Method	Precision	Recall	$F1$-score
3σ	0.0000	0.0000	0.0000
Sliding 3σ	0.1013	0.0363	0.0534
IQR	0.1071	0.0811	0.0923
Isolation forest	0.8342	0.2794	0.4186
LOF	0.2111	0.0715	0.1069
OCSVM	0.6500	0.1943	0.2992
DBSCAN	0.6637	0.1335	0.2222
AE	0.2219	0.0737	0.1106
Proposed	0.6981	0.4132	0.5191

As shown in Fig. 4, the wind power anomaly identification results based on physical power envelope constraints are presented in a wind speed-power scatter plot. The cyan curve in the figure represents the theoretical power, reflecting the distribution law of the upper and lower bounds of the power during the normal operation of the wind farm. The grey point is the normal sample falling in the envelope band, and the red point is the abnormal sample beyond the envelope band. The abnormal points are mainly concentrated in the power below or above the reference envelope section in the low wind speed region, the power deviation section in the medium speed region, and the power limit or shutdown section in the high wind speed region, which intuitively verifies the effective identification ability of the physical constraint method for power anomalies in different wind speed regions.

As shown in Fig. 5, it is the distribution comparison diagram of the power residual before and after the anomaly iden-

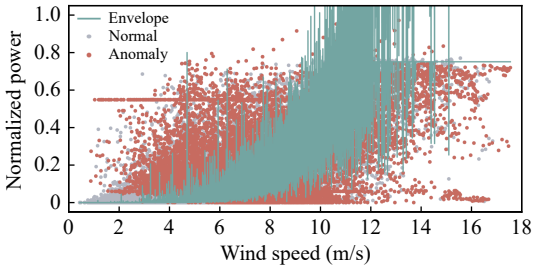

Fig. 4. Scatter plot of wind power anomaly identification results.

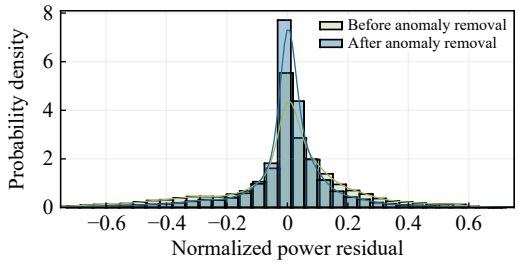

Fig. 5. Residual distribution map.

tification. The dispersion of power residual distribution before anomaly identification is high, and the tail extension is wide, showing obvious non-Gaussian dispersion characteristics, indicating that the original data contains a large number of abnormal points deviating from the physical law. After eliminating abnormal samples by the proposed method, the residual distribution is significantly narrowed, the peak value is higher, the symmetry is stronger, the retained residuals are more concentrated around zero, and the tails become less pronounced. These results show that removing the samples identified as anomalous by the proposed method narrows the residual distribution of the retained observations.

As illustrated in Fig. 6, the multivariate time series curves under extreme low-temperature operating conditions demonstrate the coupling relationships among temperature, wind speed, and power sequences. In the figure, the blue dashed line represents the −10 °C temperature threshold, the orange curve represents temperature variations, and the shaded regions correspond to extreme low-temperature intervals satisfying the condition of temperature below −10 °C. It is evident that wind speed remains within the normal range during the low-temperature intervals, thereby excluding the interference of strong wind-induced shutdowns. In the power subplot, the blue solid line represents the actual power, the green dashed line represents the theoretical power, and the orange and red points denote the anomalous points identified by physical constraints and their corrected results, respectively. Within the low-temperature shaded intervals, the turbine output deviates markedly from the theoretical curve, with severe fluctuations and frequent zero-power observations. The refined anomaly identification method effectively filters spurious anomalies and identifies power-response deviations during the selected low-temperature intervals.

As illustrated in Fig. 7, the multivariate time series curves under convective weather operating conditions demonstrate the cooperative variation characteristics among humidity, wind speed, and power sequences. In the figure, the green dashed line represents the 70% humidity threshold, the purple curve represents humidity variations, and the shaded regions correspond to severe convective weather intervals satisfying the condition of relative humidity no less than 70%, together with significant wind-speed and temperature variations and the absence of extreme strong-wind conditions. It is evident that wind speed undergoes severe fluctuations during the convective intervals, with multiple instances of abrupt increases and decreases. In the power subplot, the deviation between the actual power curve and the theoretical power increases substantially within the convective shaded intervals, with a considerable number of anomalous points appearing outside the envelope band. The physical constraint-identified orange anomalous points are relatively numerous, whereas the corrected red anomalous points effectively eliminate a portion of the spurious anomalies, precisely capturing the power anoma-

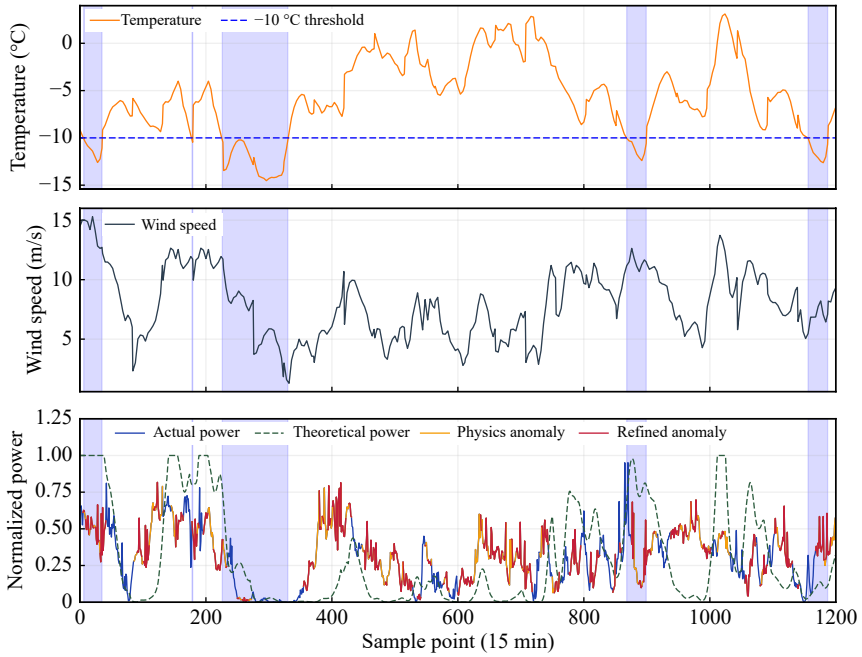

Fig. 6. Time series characteristic analysis of extreme low temperature.

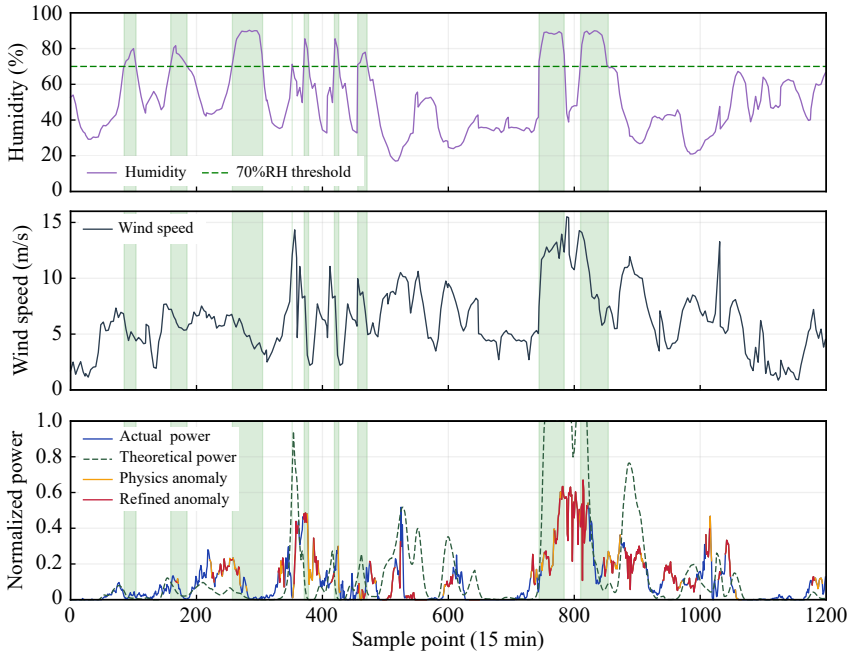

Fig. 7. Time series characteristic analysis of convective weather.

lies induced by the cooperative variations of multiple meteorological factors, thereby validating the effectiveness of the proposed correction method under complex convective operating conditions.

C. Effectiveness verification of multi-meteorological factor feature extraction

To validate the rationality of feature selection by the proposed self-attention mechanism across different meteorological scenarios, visualization experiments of weight distributions are conducted by constructing four categories of typical weather operating conditions, namely, normal stable weather, extreme high winds, extreme low temperatures, and severe convective weather.

Figure 8 presents the self-attention weight distributions of eight meteorological features across 100 wind farm stations under four categories of typical weather scenarios. The model is capable of adaptively adjusting feature importance allocation in accordance with different operating conditions, and the weight distributions exhibit scenario-dependent patterns in the displayed feature scores governing wind power responses.

Under normal stable weather conditions, the attention weights of individual meteorological features are distributed relatively uniformly without a single dominant factor, indicating that turbine output is subject to the cooperative influence of multiple meteorological factors with no pronounced disturbance source. In the extreme high-wind scenario, the model significantly elevates the weights assigned to temperature and humidity features, reflecting the heightened susceptibility of wind turbine output to variations in ambient temperature and humidity under extreme wind conditions, rather than being solely determined by static wind speed. Under the extreme low-temperature scenario, the weights of wind speed variation and wind speed fluctuation features increase markedly, reflecting the sensitivity of turbines to short-term abrupt wind speed changes and dynamic fluctuations in low-temperature environments. In the severe convective weather scenario, the weights of multiple features including wind speed, temperature, humidity, and wind speed variation are simultaneously enhanced, forming a multi-factor cooperative attention pattern that conforms to the coupled influence of intense meteorological disturbances on power output under severe convective weather conditions.

The proposed self-attention mechanism abandons the deficiencies of conventional fixed-weight or manually assigned weighting approaches, enabling dynamic and autonomous importance allocation among multiple meteorological factors in accordance with real-time weather types and anomalous states, thereby providing a compact meteorological representation for wind power forecasting under extreme operating conditions.

D. Comparative analysis of prediction models

To comprehensively validate the effectiveness of the PE-STGNN model proposed in this study, a comparative evaluation is conducted against three categories of mainstream models: (1) traditional machine learning models: extreme gradient boosting (XGBoost); (2) purely deep learning temporal models: long short-term memory (LSTM) and Transformer; and (3) spatio-temporal graph models: graph convolutional network (GCN) and STGNN.

As shown in Table 3, the proposed method achieves lower errors on the evaluated dataset than other comparison models in terms of various evaluation indicators and time series curve

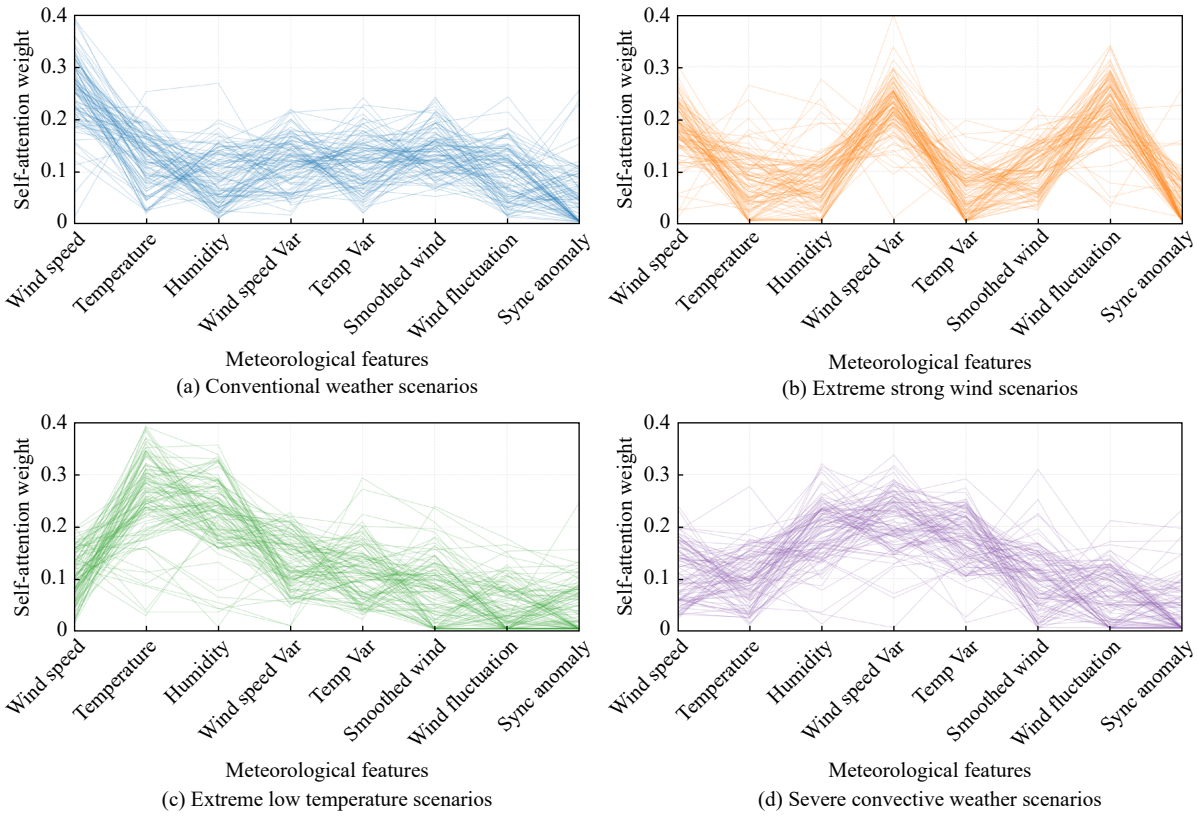

Fig. 8. Analysis of self-attention weight distribution.

fitting effects. The NMAE, NRMSE, MAPE, and SMAPE of the proposed method are reduced to 5.785%, 7.542%, 27.401%, and 24.362%, respectively, which are 30.3%, 26.6%, 29.9%, and 20.7% lower than the STGNN baseline, respectively. At the same time, the R^2 and r reach 0.839 and 0.925, respectively. The prediction error is significantly reduced and the goodness of fit is greatly improved.

As shown in Fig. 9, the comparison of time series curves shows that the XGBoost and GCN baselines are difficult to capture the non-stationary fluctuations of power series. LSTM and Transformer are prone to lag or over-smoothing problems. Although STGNN has the ability of spatio-temporal modeling, there are still details lost during power peaks, troughs and sudden changes in strong convection. In contrast, the prediction curve of the proposed method has the highest degree of fit with the measured sequence, which can accurately track the violent fluctuation and mutation process under extreme conditions, effectively overcome the deviation and distortion of the baseline model, and illustrate its forecasting performance and advantages in multi-meteorological coupling scenarios.

As shown in Table 4, the proposed method achieves the lowest NMAE and NRMSE under all three high-impact weather scenarios. Under extreme strong-wind conditions, its NMAE and NRMSE are 6.123% and 7.590%, respectively. Under extreme low-temperature conditions, the corresponding values are 6.380% and 7.924%, while under severe convective weather they reach 5.684% and 7.783%. These results demonstrate that the proposed method maintains stable forecasting advantages across different high-impact weather conditions, particularly during rapidly varying meteorological processes.

On the whole, the model in this paper realizes the coupling

TABLE 3
OVERALL FORECASTING PERFORMANCE OF DIFFERENT MODELS

Model	NMAE (%)	NRMSE (%)	MAPE (%)	SMAPE (%)	R^2	r
XGBoost	9.209	11.916	43.998	34.607	0.599	0.801
LSTM	9.532	12.137	40.022	35.687	0.584	0.795
Transformer	9.107	11.039	57.047	35.872	0.656	0.815
GCN	8.934	11.449	39.058	33.237	0.630	0.820
STGNN	8.303	10.275	39.060	30.705	0.702	0.872
Proposed method	5.785	7.542	27.401	24.362	0.839	0.925

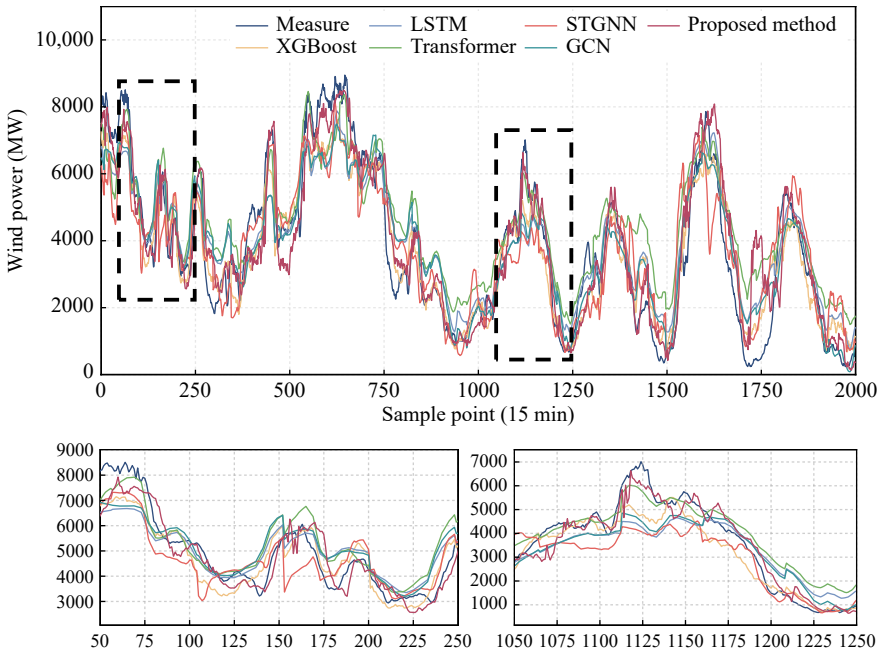

Fig. 9. The prediction results of different models.

TABLE 4
FORECASTING PERFORMANCE UNDER DIFFERENT WEATHER SCENARIOS

Scenarios	Strong wind		Low temperature		Convection	
Model	NMAE (%)	NRMSE (%)	NMAE (%)	NRMSE (%)	NMAE (%)	NRMSE (%)
XGBoost	9.268	11.732	9.601	12.400	10.062	12.712
LSTM	10.888	13.620	10.084	12.816	9.433	11.905
Transformer	9.335	11.527	9.390	11.342	9.475	11.468
GCN	9.977	11.987	9.053	11.777	9.052	11.435
STGNN	11.493	12.890	8.961	11.068	7.613	9.425
Proposed method	6.123	7.590	6.380	7.924	5.684	7.783

modeling of meteorology and power by embedding the physical diffusion mechanism. It not only inherits the feature extraction advantages of the spatio-temporal graph network, but also improves performance on the evaluated test set through physical constraints. Compared with the traditional model and the pure deep learning model, it has achieved improvements in prediction accuracy and adaptability to complex working conditions.

E. Ablation experiment

In order to verify the gain effect of physical constraint anomaly identification, single factor meteorological feature extraction, multi-factor meteorological coupling feature, and PE-STGNN model combined with the three on prediction performance, five groups of ablation models are set up in this study.

Method 1: STGNN, without any innovative modules.

Method 2: On the basis of STGNN, only the physical constraint anomaly label is added to optimize the abnormal samples, eliminate the false anomalies and reduce the noise interference, and test the improvement effect of the anomaly identification module on the training quality.

Method 3: Only introduces self-attention single-factor meteorological features, adaptively extracts dominant meteorological factors in different scenarios, and verifies the ability of single-factor dynamic feature extraction.

Method 4: Only the multi-factor coupling characteristics of weather attention are added to mine the linkage disturbance law of multiple meteorological factors to capture the non-stationary change of power.

Method 5: PE-STGNN model, which combines physical constraints, single factor features, and coupling features.

As presented in Table 5, the ablation experiment forecasting results and evaluation metrics demonstrate that the proposed method achieves the best overall results among the evaluated variants in Table 5, with NMAE, NRMSE, MAPE, and SMAPE as low as 5.785%, 7.542%, 27.401%, and 24.362%, respectively, while the coefficient of determination R^2 and

TABLE 5
OVERALL FORECASTING PERFORMANCE OF DIFFERENT ABLATION VARIANTS

Method	NMAE (%)	NRMSE (%)	MAPE (%)	SMAPE (%)	R^2	r
Method 1	8.303	10.275	39.060	30.705	0.702	0.872
Method 2	6.427	8.133	30.581	25.273	0.813	0.921
Method 3	7.413	9.524	29.919	29.347	0.744	0.892
Method 4	6.658	8.403	33.166	25.713	0.806	0.896
Proposed method	5.785	7.542	27.401	24.362	0.839	0.925

Pearson correlation coefficient attain values of 0.839 and 0.925, respectively. Compared with the baseline spatiotemporal model (method 1), the NMAE and NRMSE are reduced by 30.3% and 26.6%, respectively, demonstrating substantial superiority over models incorporating only single meteorological features or anomaly identification alone, thereby supporting the benefit of the complete configuration on forecasting accuracy under extreme weather conditions.

Table 6 further shows that the contributions of individual modules are scenario dependent. Under strong-wind conditions, the complete model achieves the lowest NMAE and NRMSE. Under low-temperature conditions, method 4 yields slightly lower errors than the complete model, showing that this variant performs better in the low-temperature scenario. Under severe convective weather, the proposed method achieves the lowest NMAE, whereas method 2 obtains a slightly lower NRMSE. Overall, the complete framework provides the most balanced performance across the three high-impact weather scenarios.

As shown in Fig. 10, methods 1 and 3 exhibit pronounced over-smoothing and detail loss issues during power peak, trough, and severe fluctuation intervals, rendering them incapable of reproducing the dynamic variations of the measured sequences. Methods 2 and 4 produce forecasting curves that more closely approximate the measured values, yet still exhibit certain residual deviations. In contrast, the proposed method achieves the highest degree of conformity between the forecasted and measured power curves, enabling precise tracking of the overall power trends, short-term fluctuations, and abrupt changes under extreme operating conditions, while showing closer agreement with the measured curves in the displayed intervals, thereby validating the overall effectiveness of the complete configuration.

VII. CONCLUSION

This study focuses on the problem of accurate prediction of multi-station wind power in extreme weather, and combines anomaly identification, adaptive meteorological feature extraction and dynamic spatio-temporal modeling methods to carry out systematic research. The following conclusions are drawn:

(1) This study proposes a wind farm power anomaly identification method that combines physical constraints and time series similarity verification. The core is to solve the problem of high misjudgment rate and weak generalization of a single method. Taking into account the physical interpretability and data-driven accuracy, adapting to complex field station scenarios, avoiding pure data-driven model defects, and outputting optimized abnormal label data sets, laying the foundation for subsequent research.

(2) Aiming at the problem of nonlinear coupling of multiple meteorological factors and difficulty in quantifying dynamic effects, the weather attention model can adaptively mine the differential weight and coupling evolution law of each meteorological element under different extreme conditions, and get rid of the disadvantages of fixed weight modeling. Without a large amount of labeled data, it can generate fine meteorological features with physical interpretability, which greatly improves the scene adaptability of meteorological characterization.

(3) Based on the physical law of extreme meteorological factors spreading from high gradient to low gradient, the PE-

TABLE 6
ABLATION PERFORMANCE UNDER DIFFERENT HIGH-IMPACT WEATHER SCENARIOS

Scenarios	Strong wind		Low temperature		Convection	
Method	NMAE (%)	NRMSE (%)	NMAE (%)	NRMSE (%)	NMAE (%)	NRMSE (%)
Method 1	11.493	12.890	8.961	11.068	7.613	9.425
Method 2	7.610	9.162	6.787	8.661	6.206	7.682
Method 3	8.741	10.646	7.774	9.910	7.322	9.469
Method 4	9.071	10.950	6.188	7.752	6.962	8.688
Proposed method	6.123	7.590	6.380	7.924	5.684	7.783

清華大學出版社 Tsinghua University Press | SciOpen

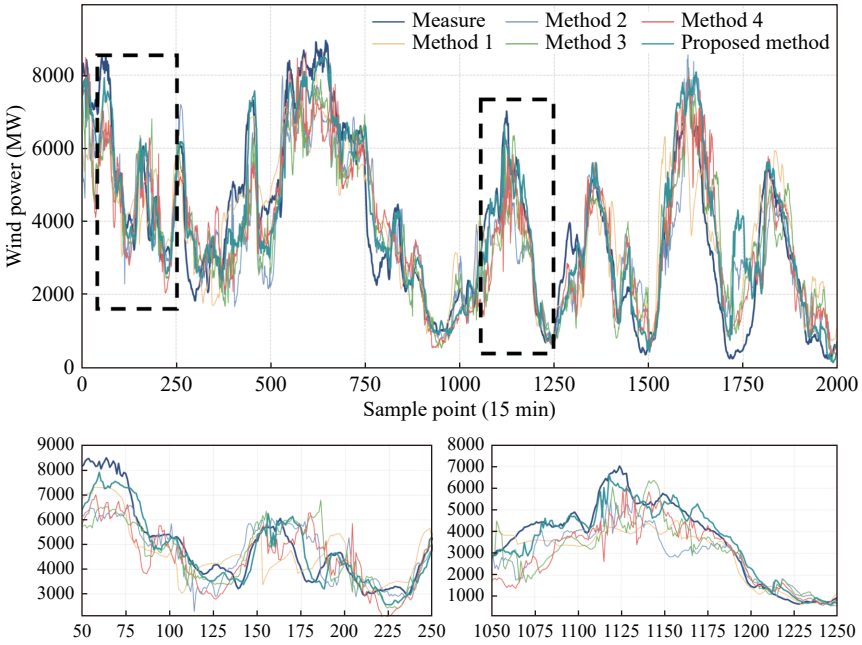

Fig. 10. The predicted results of ablation experiments.

STGNN physical enhanced spatio-temporal prediction model is constructed by combining the geographical distribution of wind farms and real-time meteorological influence characteristics to construct dynamic graph edge weights, so as to realize the adaptive update of network topology and accurately fit the actual physical process of meteorological diffusion. At the same time, the time-delay attention is combined to capture the time-delay characteristics, supplemented by multi-scenario physical constraints to correct the prediction deviation, which breaks through the defects of traditional static graph model modeling solidification, effectively fits the actual operation mechanism of the fan, and significantly improves the accuracy of wind power prediction at multiple stations in extreme weather.

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.

Acknowledgements

This work is carried out without the support of any funding agency.

Use of AI statement

During the preparation of this manuscript, the authors used AI-assisted tools for language polishing only. The authors have reviewed and edited the content and take full responsibility for the final version of the manuscript.

REFERENCES

[1] M. O'Malley, H. Holttinen, N. Cutululis, T. K. Vrana, J. King, V. Gevorgian, X. F. Wang, F. Rajaei-Najafabadi, and A. Hadjileonidas, "Grand challenges of wind energy science—meeting the needs and services of the power system," *Wind Energy Sci.*, vol. 9, no. 11, pp. 2087–2112, Nov. 2024.

[2] M. Yang, Y. T. Huang, C. Y. Xu, C. Y. Liu, and B. Z. Dai, "Review of several key processes in wind power forecasting: mathematical formulations, scientific problems, and logical relations," *Appl. Energy*, vol. 377, pp. 124631, Jan. 2025.

[3] X. J. Guo, P. L. Zeng, X. Xiong, G. W. Wang, and Y. Cui, "Short-term wind power forecasting methods based on machine learning: a review and case study," *Energy Rep.*, vol. 14, pp. 3753–3782, Dec. 2025.

[4] M. Yang, X. Y. Li, Z. F. Han, Y. Sun, and X. Q. Chang, "Ultra-short-term multi-step prediction of wind power fusing multivariate time-series relationships and multi-task correction," *IEEE Trans. Ind. Appl.*, vol. 62, no. 3, pp. 4286–4296, May/Jun. 2026.

[5] J. Yan, Y. J. Li, H. Wang, S. Han, W. L. Shang, and Y. Q. Liu, "Wind power forecasting based on large time series model," *Engineering*, to be published.

[6] Y. N. Zhao, H. H. Liao, S. J. Pan, and Y. Zhao, "Interpretable multi-graph convolution network integrating spatio-temporal attention and dynamic combination for wind power forecasting," *Expert Syst. Appl.*, vol. 255, pp. 124766, Dec. 2024.

[7] G. L. Hou, Q. W. Li, and C. Z. Huang, "Spatiotemporal forecasting using multi-graph neural network assisted dual domain transformer for wind power," *Energy Convers. Manage.*, vol. 325, pp. 119393, Feb. 2025.

[8] L. Wang, J. H. Gao, Y. C. Li, and D. Wang, "A method for ultra-short-term wind power forecasting of large-scale wind farms based on adaptive spatiotemporal graph convolution," *Renew. Energy*, vol. 249, pp. 123321, Aug. 2025.

[9] Q. Y. Zhang, K. Mei, Q. Zhou, X. Z. Ding, Z. C. Dong, M. Tian, D. H. Liu, W. He, Z. Zeng, and M. L. Li, "A wind farm cluster power prediction method based on globally aware dynamic spatiotemporal graph continual learning framework," *Energy*, vol. 348, pp. 140591, Apr. 2026.

[10] M. Yang, K. X. Wang, X. Su, M. M. Ma, G. Wu, and D. W. Huang, "Short-term photovoltaic output probability prediction method

considering spatio-temporal-condition dependence of prediction error," *CSEE J. Power Energy Syst.*, vol. 12, no. 3, pp. 1386–1399, May 2026.

[11] L. Y. Gao, T. Dasari, and J. R. Hong, "Wind farm icing loss forecast pertinent to winter extremes," *Sustain. Energy Technol. Assess.*, vol. 50, pp. 101872, Mar. 2022.

[12] J. Z. Dai, M. Rotea, and N. Kehtarnavaz, "An ensemble network for high-accuracy and long-term forecasting of icing on wind turbines," *Sensors*, vol. 24, no. 24, pp. 8167, Dec. 2024.

[13] C. Y. Lee and E. D. C. Maceren, "Physics-informed anomaly and fault detection for wind energy systems using deep CNN and adaptive elite PSO-XGBoost," *IET Gener. Transm. Distrib.*, vol. 19, no. 1, pp. e13289, Jan./Dec. 2025.

[14] A. V. Ciocan, S. Hamrioui, P. Lorenz, V. Courboulay, and A. Ciocan, "Energy-aware anomaly detection in wind turbine SCADA systems," in *Proceedings of International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, Split, Croatia, 2025, pp. 1–6.

[15] B. F. Qin, D. Q. He, Z. Z. Jin, S. Zhang, X. W. Li, J. X. Wu, H. M. Sun, and Y. Zhuang, "Robust open-circuit fault diagnosis for PMSM drives under unknown operating conditions," *IEEE Trans. Instrum. Meas.*, vol. 75, pp. 3503313, Jan. 2026.

[16] C. Y. Dai, D. Q. He, Z. Z. Jin, X. W. Zhang, G. Chen, J. X. Wu, J. Y. Zhao, and Y. Xu, "Digital twin-assisted graph contrastive domain adaptation for small-sample bearing fault diagnosis," *Struct. Health Monit.*, to be published.

[17] L. Liu, X. Y. Wang, X. Dong, K. Chen, Q. J. Chen, and B. Li, "Interpretable feature-temporal transformer for short-term wind power forecasting with multivariate time series," *Appl. Energy*, vol. 374, pp. 124035, Nov. 2024.

[18] M. Yang, X. X. Shen, D. W. Huang, and X. Su, "Fluctuation classification and feature factor extraction to forecast very short-term photovoltaic output powers," *CSEE J. Power Energy Syst.*, vol. 11, no. 2, pp. 661–670, Mar. 2025.

[19] C. Cakiroglu, S. Demir, M. H. Ozdemir, B. L. Aylak, G. Sariisik, and L. Abualigah, "Data-driven interpretable ensemble learning methods for the prediction of wind turbine power incorporating SHAP analysis," *Expert Syst. Appl.*, vol. 237, pp. 121464, Mar. 2024.

[20] H. Y. Zhang, R. Zandehshahvar, M. Tanneau, and P. Van Hentenryck, "Weather-informed probabilistic forecasting and scenario generation in power systems," *Appl. Energy*, vol. 384, pp. 125369, Apr. 2025.

[21] Y. Z. Cai and Y. T. Li, "Short-term wind speed forecast based on dynamic spatio-temporal directed graph attention network," *Appl. Energy*, vol. 375, pp. 124124, Dec. 2024.

[22] M. Yang, Y. F. Guo, P. Colella, Y. Jiang, J. S. Chen, J. Yin, Y. Li, and T. Huang, "Wind power low-output event identification and accurate prediction strategy based on global heterogeneous information dynamic fusion with dynamic dual graph neural networks," *Energy*, vol. 349, pp. 140620, Apr. 2026.

[23] Z. Zhang, B. Wang, M. Yang, Z. Wang, W. Zhang, and T. Huang, "Short-term prediction of wind farm cluster power based on bidirectional interactive fusion of homogeneous and heterogeneous information," *Prot. Control Mod. Power Syst.*, vol. 11, no. 2, pp. 107–125, Mar. 2026.

[24] R. T. A and C. C. Chinnappan, "Integrating data-driven and physics-based approaches for robust wind power prediction: a comprehensive ML-PINN-Simulink framework," *Sci. Rep.*, vol. 15, no. 1, pp. 29102, Aug. 2025.

[25] A. Gijón, A. Pujana-Goitia, E. Perea, M. Molina-Solana, and J. Gómez-Romero, "Prediction of wind turbines power with physics-informed neural networks and evidential uncertainty quantification," *Eng. Appl. Artif. Intell.*, vol. 164, pp. 113331, Jan. 2026.



Mao Yang received his Ph.D. degree in control theory and control engineering from Jilin University, Changchun, China, in 2010. Since 2018, he has been a Professor with the School of Electrical Engineering, Northeast Electric Power University, Jilin, China. His research interests include the areas of load forecasting and micro-grid operation management with an emphasis on wind power/photovoltaic power prediction.



Yuxi Jiang received her B.Eng. degree in Electrical Engineering from Northeast Electric Power University, China, in 2023. She is currently pursuing a Ph.D. degree in Electrical Engineering at Northeast Electric Power University, China. Her research interest is wind/photovoltaic power prediction.



Tao Huang received his B.S. and M.E. degrees from Shanghai Jiao Tong University, Shanghai, China, in 2002 and 2007, respectively, and his Ph.D. degree from Politecnico di Torino, Italy, in 2011. He is currently an Associate Professor in the Department of Energy at Politecnico di Torino. His research interests focus on electricity market, smart grids, power system operation, and optimization.