

Spatiotemporal Joint Power Forecasting for Hydro-Wind-Solar Complementary Systems via a Hierarchical Complementary-Prior Regularized MSTD-HGFormer

Weijie Zhao^{1,2}, Yi Jiang^{1,2*}, Yuxin Xie^{1,2}, Panfeng Guo⁵, Lingfei Li^{2,3}, Jianxiang Huang^{1,2}, Zhixiong Huang⁴

¹Kunming Bureau of CSG EHV Transmission Company, Kunming 650217, China

²Yunnan Provincial UHVDC Smart O&M and Safe Operation Engineering Research Center, Kunming 650217, China

³Electric Power Research Institute, CSG, Guangzhou 510663, China

⁴Guangdong Tianguang Energy Technology Development Co., Ltd., Guangzhou 510620, China

⁵School of Electrical Engineering and Renewable Energy, China Three Gorges University, Yichang 443002, China

Abstract

Accurate joint forecasting of hydro-wind-photovoltaic bases is difficult because hydropower regulation, wind randomness and photovoltaic intermittency interact through both temporal complementarity and spatial-electrical coupling. This study aims to improve source-level and bundled-power forecasting by embedding the complementary mechanism of heterogeneous renewable sources into both representation learning and model training. A hierarchical complementary-prior regularized multi-scale seasonal-trend decoupled heterogeneous graph Transformer, denoted MSTD-HGFormer-HCPR, is proposed. The model first separates the historical multi-source sequence into low-frequency trend and high-frequency fluctuation components through learnable multi-scale moving-average kernels. The two components are then encoded by parallel heterogeneous graph Transformer branches using temporal patch attention and a static-dynamic fused adjacency that combines physical topology, historical correlation and learnable latent relations. A gated fusion module produces the final forecast, while a composite loss introduces capacity-weighted system-level and correlation-aware topology-level complementary-prior constraints. Experiments on a real 17-node hydro-wind-photovoltaic base show that the proposed method reduces rRMSE by 12.9–50.6% against seven baselines, improves the correlation coefficient by 0.011–0.013 over the strongest baseline, and keeps the forecasting accuracy at or above 96.2% for hydro, wind, photovoltaic and aggregate outputs. A second, 23-node cascade-radial base with five-year records is further examined under a rolling-origin and season-stratified protocol; the aggregate rRMSE is reduced by 13.2–44.0%, and the pooled accuracies remain 96.0% for hydropower, 96.0% for wind, 96.4% for photovoltaic and 96.2% for aggregate outputs. The proposed regularized joint forecaster provides stable and accurate spatiotemporal prediction for complementary renewable bases and reduces the accumulation of source-level errors in bundled output forecasting.

Keywords: hydro-wind-solar complementary system; joint power forecasting; hierarchical complementary-prior regularization; multi-scale seasonal-trend decoupling; heterogeneous graph Transformer

Received on 27 June 2026; accepted on 21 September 2026; published on 23 October 2026

Copyright © 2026 Weijie Zhao *et al.*, licensed to EAI. This is an open access article distributed under the terms of the [CC BY-NC-SA 4.0](#), which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

doi:10.4108/ew.15713

*Corresponding author. Email: 296472604@qq.com

1. Introduction

High renewable penetration is changing the operating pattern of modern power systems. Wind and photovoltaic (PV) generation increase the proportion of weather-dependent output, while the retirement or reduced dispatch of synchronous units weakens inertia and flexibility margins [1, 2]. In this context, source-load and renewable-output forecasting under high renewable penetration has become an important prerequisite for secure dispatch [3]. A hydro-wind-solar complementary base is a representative engineering form for renewable accommodation: hydropower provides regulation, wind and PV supply large-scale clean energy, and the three resources are usually exported through a shared transmission corridor [4–6]. For system operators, the relevant scheduling object is therefore not only the power of an individual station, but also the bundled output of the whole base.

The forecasting problem is challenging because the three sources are heterogeneous in physics, time scale and operating constraint. Hydropower is partly dispatchable and often exhibits low-frequency regulation. Wind power is highly stochastic and responds rapidly to meteorological variation. PV output is dominated by diurnal periodicity, irradiance fluctuation and cloud-induced intermittency. When stations are geographically close, they may be affected by related weather processes but respond with different amplitudes, delays and smoothing effects, as illustrated in Figure 1. Meanwhile, these stations are connected through a non-trivial spatial-electrical topology. The resulting sequence is nonlinear, non-stationary and graph-coupled; errors in source-level forecasts can also be superposed in the aggregate power exported by the base.

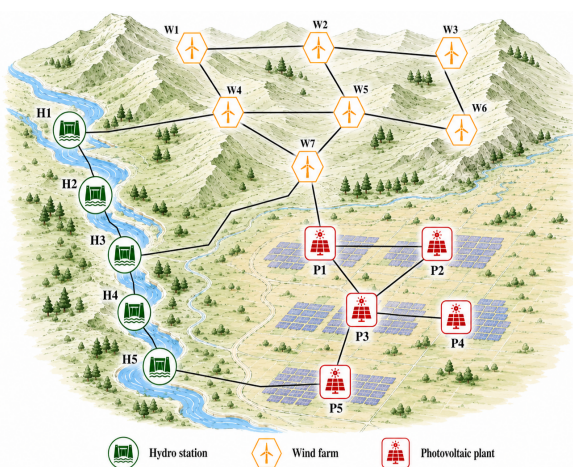


Figure 1. Spatiotemporal network topology and temporal complementarity of the hydro-wind-solar base

Existing studies have made substantial progress in single-source forecasting and homogeneous multi-station forecasting. However, many multi-source joint models still concatenate hydro, wind and PV sequences or map them into a shared latent space and then optimize a generic point-wise error loss. This strategy improves data fusion, but it tends to weaken the distinctions among source scale, regulating mechanism and fluctuation mode. In particular, a high correlation between heterogeneous nodes does not necessarily imply similar output curves. The desired relation is often complementary: hydropower may smooth wind-PV volatility at the system level, while neighbouring heterogeneous stations should remain distinguishable at the topology level.

To address this issue, this paper proposes a hierarchical complementary-prior regularized multi-scale seasonal-trend decoupled heterogeneous graph Transformer, abbreviated as MSTD-HGFormer-HCPR. The overall architecture is shown in Figure 2. At the input level, learnable multi-scale seasonal-trend decoupling separates each multi-source sequence into a low-frequency trend component and a high-frequency fluctuation component. At the structural level, a heterogeneous graph Transformer encodes temporal patches and station relations over a fused adjacency that integrates physical electrical connections, historical correlation and learnable latent coupling. At the optimization level, hierarchical complementary-prior regularization (HCPR) augments the mean square error (MSE) with a capacity-weighted system-level constraint and a correlation-aware topology-level constraint. In this way, complementarity is treated as an explicit training prior rather than as an implicit pattern to be inferred only from data.

The main contributions are summarized as follows.

First, a dual-branch MSTD-HGFormer backbone is developed for hydro-wind-solar joint forecasting. It models trend and fluctuation components separately and represents heterogeneous station relations by a static-dynamic fused graph attention mechanism.

Second, a hierarchical complementary-prior composite loss is designed. The capacity-weighted system-level term imposes cross-source smoothing on the aggregate output, while the correlation-aware topology-level term suppresses over-smoothing between connected heterogeneous nodes.

Third, the method is validated on a real 17-node hydro-wind-solar base and on a second 23-node base with a five-year record. Compared with seven temporal and spatiotemporal baselines, it improves both source-level and aggregate forecasting accuracy: rRMSE falls by 12.9–50.6% on the 17-node base and by 13.2–44.0% on the 23-node base. An ablation study isolates the joint graph backbone and the two complementary-prior

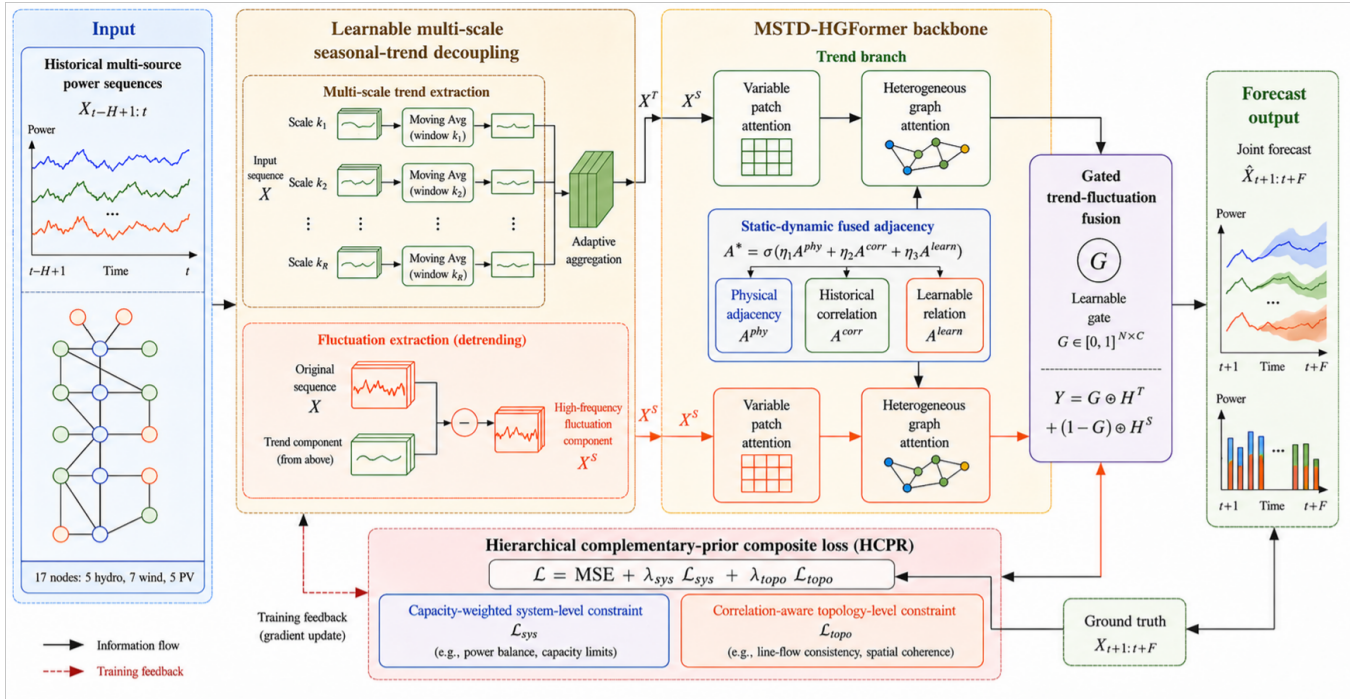


Figure 2. Overall framework of MSTD-HGFormer-HCPR

terms. A rolling-origin, season-stratified protocol then tests whether the gain persists across years and seasons.

The remainder of this paper is organized as follows. Section 2 reviews related studies. Section 3 formulates the forecasting task and describes the proposed model. Section 4 presents the datasets, baselines, evaluation metrics and experimental results, including the extended multi-year case. Section 5 concludes the paper.

2. Related Work

2.1. Single-source power forecasting

Single-source forecasting is the basis of joint renewable prediction. For hydropower generation forecasting, recent studies have incorporated hydrometeorological variables, hybrid deep networks, spatiotemporal correlations and operational constraints to characterize the nonlinear mapping from water-resource conditions to power output [7, 8]. For wind and PV, deep learning has become a dominant approach. Convolutional and recurrent networks have been used for regional probabilistic wind forecasting and short-term solar irradiance forecasting [9, 10]. Signal decomposition is also widely used to reduce non-stationarity before learning; representative studies include multi-temporal-scale wind forecasting, quadratic decomposition with multi-objective optimization and adaptive PV forecasting with error compensation [11–13].

These methods are effective in extracting temporal patterns within one energy type. Their limitation is

that spatial relations among co-located stations and complementary behavior among different sources are usually outside the modeling scope. Consequently, they are less suited to forecasting the bundled output of a hydro-wind-solar base, where the value of the system depends on cross-source compensation rather than on isolated source behavior.

2.2. Spatiotemporal deep learning for renewable forecasting

Spatiotemporal deep learning treats stations as graph nodes and combines temporal encoding with spatial information propagation. For wind power, multi-task spatiotemporal attention [14], correlation-constrained vector autoregression [15], graph convolution [16], dynamic-graph modeling [17], blockage-aware temporal-spatial graphs [18] and graph-enhanced KANformer structures [19] have all improved multi-site prediction. For distributed PV and behind-the-meter resources, deep graph dictionary learning [20], multi-site graph neural networks [21] and dual-attention LSTM models [22] show that spatial correlation can reduce forecast error.

Most of these models are designed for homogeneous station groups. Their graph construction and feature propagation are usually performed within one energy source, and node attributes are assumed to have comparable physical meaning. Directly extending them to a hydro-wind-solar base may confuse correlation with complementarity, because strongly connected

heterogeneous nodes should not always be smoothed toward similar trajectories.

2.3. Multi-source joint forecasting

Joint forecasting has been explored to exploit the complementary potential among hydro, wind and PV. Existing work includes deep neural networks for integrated hydro-wind-solar prediction [23], joint hydro-wind-solar forecasting that accounts for the spatiotemporal delay of weather processes [4], integrated ultra-short-term regional hydro-wind-solar forecasting [5], and hybrid machine-learning models for regional hydro-wind-solar power forecasting [6]. These studies confirm that multi-source information is useful for renewable prediction.

The remaining issue is how to encode heterogeneity. Many joint models concatenate, decompose or pool multi-source features and then use deeper networks to represent coupling. This strategy may improve fitting accuracy, but it does not explicitly express the hierarchy of complementarity: source-level temporal smoothing, graph-level spatial differentiation and aggregate-level error control. The present study addresses this gap by coupling a dual-branch heterogeneous graph forecasting backbone with a hierarchical complementary-prior objective.

3. Methodology

The proposed method contains three main components: learnable multi-scale seasonal-trend decoupling, heterogeneous graph spatiotemporal encoding and hierarchical complementary-prior regularization. As shown in Figure 2, the input sequence is first decomposed; the trend and fluctuation components are then processed by two parallel HGFormer branches, and the two branch forecasts are combined through a learnable gate. During training, the HCPR loss constrains both system-wide co-smoothing and cross-source topological differentiation.

3.1. Problem formulation

A hydro-wind-solar base is formulated as a heterogeneous graph $G = (V, E)$, where $V = \{v_1, v_2, \dots, v_N\}$ denotes the set of station nodes and E denotes the set of electrical interconnections or physically meaningful spatial associations among stations. Each node is assigned an energy-type attribute corresponding to hydropower, wind power or PV power, thereby preserving source heterogeneity in the graph-based spatiotemporal representation. The physical adjacency matrix $A = [a_{ij}]$ is defined as

$$a_{ij} = \begin{cases} 1, & (v_i, v_j) \in E \text{ or } i = j, \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

Self-connections are retained to preserve node-specific information during graph propagation. Given a historical tensor $X_{t-H+1:t} \in \mathbb{R}^{N \times H \times D}$, where H is the input length and D is the feature dimension, the forecasting task is to learn the mapping

$$\hat{X}_{t+1:t+F} = f_{\Theta}(X_{t-H+1:t}, A) \quad (2)$$

where F is the forecast horizon and Θ denotes all learnable parameters. The target is to approximate $X_{t+1:t+F}$ at both node and aggregate levels while preserving the complementary relation among hydropower, wind power and PV power.

3.2. Learnable multi-scale seasonal-trend decoupling

Multi-source renewable sequences contain both slow regulation and fast fluctuation. Following the decomposition idea of forecasting Transformers [24, 25], the input is separated into a low-frequency trend component and a high-frequency fluctuation component before graph-temporal encoding. Unlike a fixed single-window moving average, the proposed module uses several learnable scales, as shown in Figure 3:

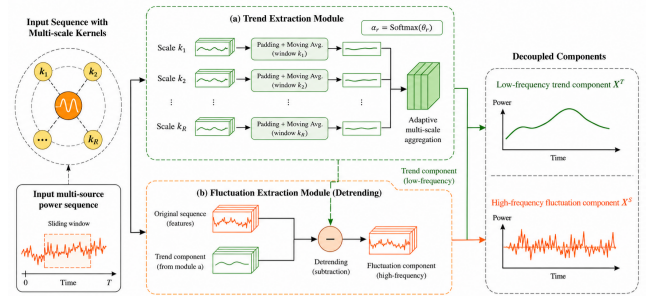


Figure 3. Learnable multi-scale seasonal-trend decoupling module

$$\begin{cases} X_{t-H+1:t}^T = \sum_{r=1}^R \alpha_r \text{AvgPool}_{k_r}(\text{Padding}(X_{t-H+1:t})) \\ \alpha_r = \frac{\exp(\theta_r)}{\sum_{q=1}^R \exp(\theta_q)} \end{cases} \quad (3)$$

The fluctuation component is obtained by detrending:

$$X_{t-H+1:t}^S = X_{t-H+1:t} - X_{t-H+1:t}^T \quad (4)$$

Here R is the number of scales, k_r is the moving-average window at scale r , θ_r is a learnable scale parameter and α_r is its Softmax-normalized weight. The trend component captures low-pass patterns caused by hydropower regulation and persistent resource conditions, while the fluctuation component retains high-pass wind-PV volatility and short-term

disturbances. This decomposition avoids the need for a single encoder to represent incompatible temporal frequencies.

3.3. Heterogeneous graph spatiotemporal encoding and gated fusion

The backbone in Figure 4 first partitions each input sequence into temporal patches. At layer l , a sequence of length H_l is divided into $P_l = H_l/S_l$ patches, where S_l is the patch stride. For node i and patch p , a learnable pseudo timestamp $T_p^{i,l} \in \mathbb{R}^{1 \times d}$ is used as the query. The variable patch attention is

$$\begin{aligned} \rho(T_p^{i,l}) &= \text{patt}(T_p^{i,l}, \chi_p^{i,l}) \\ &= \varphi \left(\frac{T_p^{i,l} (\chi_p^{i,l} W_K^{i,l})^T}{\sqrt{d}} \right) \\ &\quad \times (\chi_p^{i,l} W_V^{i,l}) \end{aligned} \quad (5)$$

where $\chi_p^{i,l}$ denotes the patch input, d is the hidden dimension, $\varphi(\cdot)$ is the Softmax function, and $W_K^{i,l}$ and $W_V^{i,l}$ are learnable temporal projections. This design follows the variate-oriented attention principle used in multivariate forecasting [26], while preserving local temporal context through patch partition.

For spatial encoding, the physical adjacency alone is insufficient because station coupling also depends on historical co-movement and latent operating relations. A fused adjacency is therefore constructed as

$$A^* = \sigma(\eta_1 A^{\text{phy}} + \eta_2 A^{\text{corr}} + \eta_3 A^{\text{learn}}) \quad (6)$$

where $A^{\text{phy}} = A$ is the fixed physical adjacency, A^{corr} is the historical-correlation adjacency, A^{learn} is a learnable relation matrix, η_1, η_2 and η_3 are non-negative fusion coefficients, and $\sigma(\cdot)$ denotes a normalizing nonlinearity. The correlation adjacency is computed as

$$A_{ij}^{\text{corr}} = |\text{corr}(X_i^{\text{hist}}, X_j^{\text{hist}})| \quad (7)$$

Only the training data or the historical window preceding the prediction instant is used when calculating A^{corr} , so that information from the forecast horizon is not leaked. The learnable relation matrix is generated by node embeddings:

$$A^{\text{learn}} = \text{Softmax}(\text{ReLU}(E_1 E_2^T)) \quad (8)$$

where $E_1, E_2 \in \mathbb{R}^{N \times d_e}$ and d_e is the node-embedding dimension. The fused adjacency simultaneously preserves known electrical structure, historical dependence and latent heterogeneous coupling.

Let $\rho[T_p^l] = [\rho(T_p^{1,l}), \dots, \rho(T_p^{N,l})] \in \mathbb{R}^{N \times d}$. The heterogeneous graph attention output for patch p is

$$\begin{aligned} \zeta(\rho[T_p^l]) &= \text{hgatt}(\rho[T_p^l], \rho[T_p^l]) \\ &= \varphi \left(\frac{(\rho[T_p^l] U_Q^{p,l})(\rho[T_p^l] U_K^{p,l})^T}{\sqrt{d}} \odot A^* \right) \\ &\quad \times (\rho[T_p^l] U_V^{p,l}) \end{aligned} \quad (9)$$

where $U_Q^{p,l}$, $U_K^{p,l}$ and $U_V^{p,l}$ are learnable spatial projections, and $\odot$ denotes the Hadamard product. After L spatiotemporal blocks, cross-layer skip fusion gives

$$O = \sum_{l=1}^L W_l O_l \quad (10)$$

where $O_l = [\zeta(\rho[T_1^l]), \zeta(\rho[T_2^l]), \dots, \zeta(\rho[T_{P_l}^l])]$ and W_l is a learnable layer-fusion parameter. A two-layer prediction head maps O to the forecast space:

$$\hat{X}_{t+1:t+F}^* = (\text{ReLU}(W_1 O)) W_2 \quad (11)$$

The trend and fluctuation components are processed by two parallel branches:

$$\hat{X}_{t+1:t+F}^T = \text{HGFormer}(X_{t-H+1:t}^T, A^*) \quad (12)$$

$$\hat{X}_{t+1:t+F}^S = \text{HGFormer}(X_{t-H+1:t}^S, A^*) \quad (13)$$

Because the relative importance of trend and fluctuation differs by source type, operating condition and forecast step, the two branch outputs are combined by a gated fusion mechanism:

$$G = \sigma \left(W_g [\hat{X}_{t+1:t+F}^T; \hat{X}_{t+1:t+F}^S] + b_g \right) \quad (14)$$

$$\hat{X}_{t+1:t+F} = G \odot \hat{X}_{t+1:t+F}^T + (1 - G) \odot \hat{X}_{t+1:t+F}^S \quad (15)$$

where G is the element-wise gate, $[\cdot; \cdot]$ denotes concatenation, and W_g and b_g are learnable parameters. The gate assigns larger weights to the trend branch during hydropower-dominated periods and retains more fluctuation information during volatile wind-PV periods.

3.4. Hierarchical complementary-prior composite loss

MSE measures point-wise numerical error but does not encode the physical complementarity of a hydro-wind-solar base. Under a plain error objective, cross-source attention may over-smooth heterogeneous neighbours or ignore the co-smoothing effect that matters for aggregate dispatch. Therefore, HCPR augments the base loss with system-level and topology-level priors, as indicated in the lower part of Figure 2:

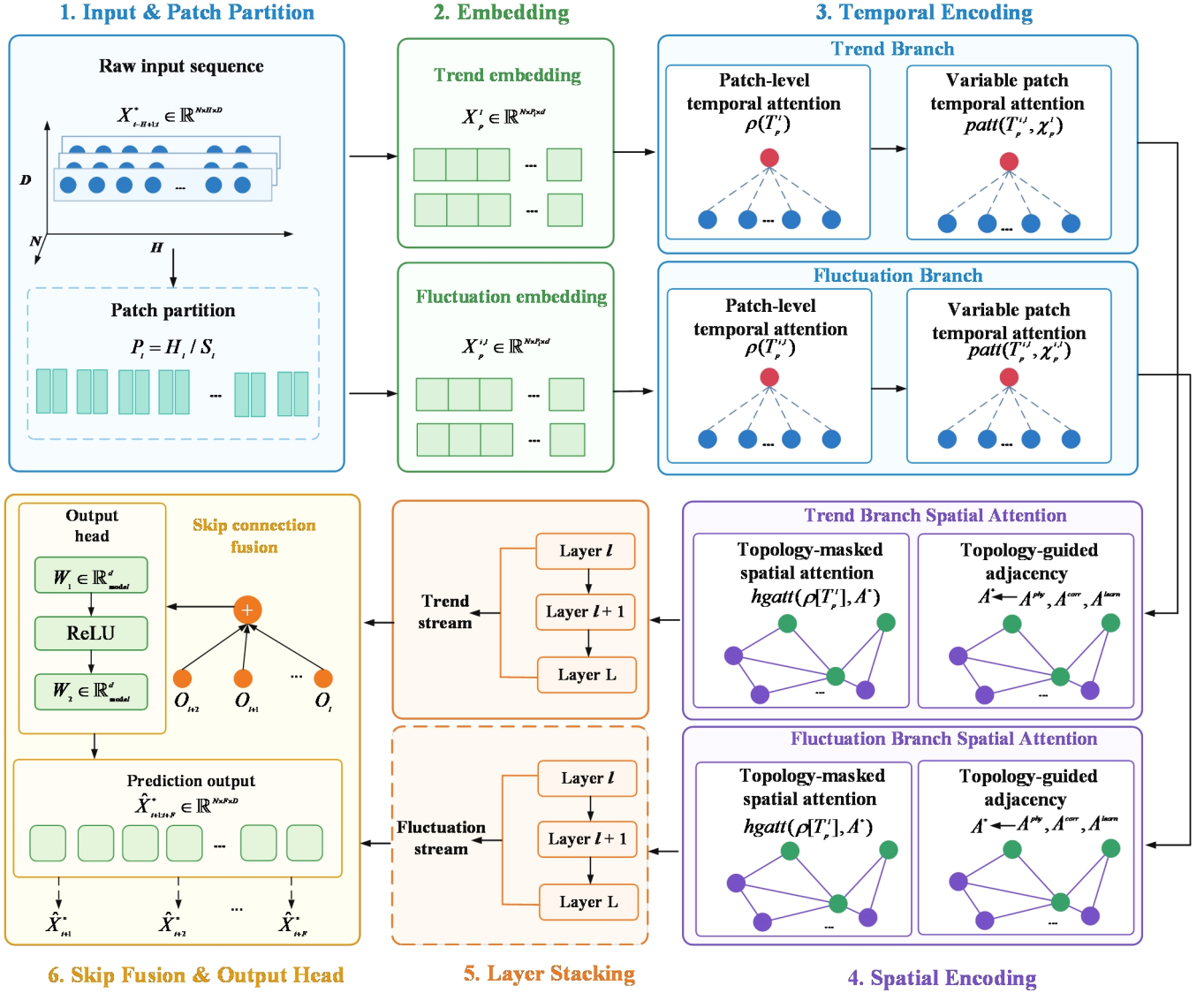


Figure 4. MSTD-HGFormer backbone for dual-branch spatiotemporal encoding and fusion

$$L = \text{MSE}(\hat{X}_{t+1:t+F}, X_{t+1:t+F}) + \lambda_{\text{sys}} L_{\text{sys}} + \lambda_{\text{topo}} L_{\text{topo}} \quad (16)$$

where L_{sys} is the capacity-weighted system-level term, L_{topo} is the correlation-aware topology-level term, and λ_{sys} and λ_{topo} are hyperparameters. The two weights are selected on the validation set from $\{0.001, 0.005, 0.01, 0.05, 0.1, 0.5\}$ using total-power rRMSE as the criterion; both are set to 0.05 in the experiments.

The system-level term extends the hydro-wind-solar complementarity metric in [27]. Let $P_w(t)$, $P_h(t)$ and $P_s(t)$ denote the capacity-normalized predicted aggregate wind, hydro and PV outputs at time t within the training horizon. Their installed-capacity weights are

$$\begin{cases} \alpha = \frac{IC_w}{IC_w + IC_h + IC_s} \\ \beta = \frac{IC_h}{IC_w + IC_h + IC_s} \\ \gamma = \frac{IC_s}{IC_w + IC_h + IC_s} \end{cases} \quad (17)$$

where IC_w , IC_h and IC_s are installed capacities. For each source against the other two-source group, capacity-weighted aggregate series are defined as

$$\begin{cases} P_{w-h,s} = \alpha P_w + (1 - \alpha)(P_h + P_s) \\ P_{h-w,s} = \beta P_h + (1 - \beta)(P_w + P_s) \\ P_{s-w,h} = \gamma P_s + (1 - \gamma)(P_w + P_h) \end{cases} \quad (18)$$

For any pair of time steps (i, j) with $j > i$, define $\Delta P_x = P_x(j) - P_x(i)$. The capacity-weighted system fluctuation

complementarity is

$$C_{\text{time}} = \frac{3}{T(T-1)/2} \sum_{i < j} \left[a \left(1 - \frac{|\Delta P_{w-h,s}|}{\alpha |\Delta P_w| + (1-\alpha) |\Delta(P_h + P_s)| + \varepsilon} \right) + b \left(1 - \frac{|\Delta P_{h-w,s}|}{\beta |\Delta P_h| + (1-\beta) |\Delta(P_w + P_s)| + \varepsilon} \right) + c \left(1 - \frac{|\Delta P_{s-w,h}|}{\gamma |\Delta P_s| + (1-\gamma) |\Delta(P_w + P_h)| + \varepsilon} \right) \right] \quad (19)$$

where T is the horizon length, a , b and c are capacity-scale correction coefficients, and ε avoids zero denominators. A larger C_{time} indicates stronger compensation of source-level increments. The corresponding penalty is

$$L_{\text{sys}} = 1 - C_{\text{time}} \quad (20)$$

This term penalizes in-phase fluctuation in the predicted multi-source output and promotes the smoothing pattern expected from a complementary base.

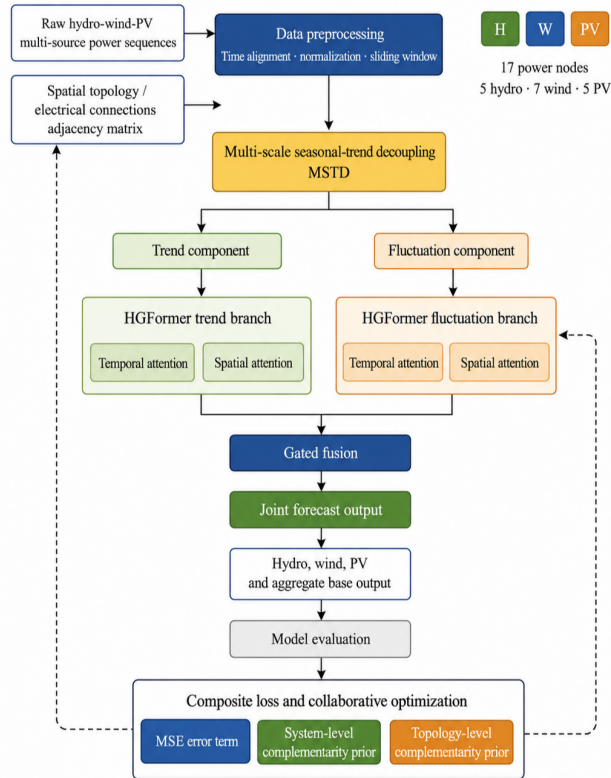


Figure 5. Flowchart for joint power prediction of hydro-wind-solar hybrid generation

The topology-level term acts on neighbouring cross-source node pairs. For predicted node sequences $\hat{x}_i, \hat{x}_j \in \mathbb{R}^F$, the mean-centred cosine similarity is

$$\cos_{ij} = \frac{(\hat{x}_i - \bar{\hat{x}}_i \mathbf{1})^T (\hat{x}_j - \bar{\hat{x}}_j \mathbf{1})}{\|\hat{x}_i - \bar{\hat{x}}_i \mathbf{1}\|_2 \|\hat{x}_j - \bar{\hat{x}}_j \mathbf{1}\|_2 + \varepsilon} \quad (21)$$

A margin penalty is then defined as

$$\ell_{ij}^{\text{pred}} = [\max(0, \cos_{ij} - \tau)]^2 \quad (22)$$

where $\tau \in [0, 1]$ is the maximum admissible similarity for heterogeneous neighbouring predictions. The adaptive edge weight is

$$\tilde{w}_{ij} = \frac{w_{ij}^{\text{hist}} w_{ij}^{\text{edge}}}{\sum_{(u,v) \in E_{\text{cross}}} w_{uv}^{\text{hist}} w_{uv}^{\text{edge}} + \varepsilon} \quad (23)$$

where w_{ij}^{edge} is obtained from the adjacency matrix, w_{ij}^{hist} is derived from historical correlation and a learnable gate, and E_{cross} is the set of connected node pairs with different source types. The topology-level constraint is

$$L_{\text{topo}} = \sum_{(i,j) \in E_{\text{cross}}} \tilde{w}_{ij} \ell_{ij}^{\text{pred}} \quad (24)$$

This term penalizes excessive similarity between heterogeneous neighbours while leaving same-source neighbours unconstrained. It therefore distinguishes local node behaviours without destroying useful graph information.

3.5. Overall forecasting procedure

The complete procedure is summarized in Figure 5. First, hydro, wind and PV power records are time-aligned, normalized and converted into sliding-window tensors. Second, the physical adjacency matrix is built from the station topology. Third, the input tensor is decomposed into trend and fluctuation components by the learnable multi-scale module. Fourth, the two components are encoded by parallel HGFormer branches using temporal patch attention and fused heterogeneous graph attention. Fifth, the gated fusion module produces the final node-level and aggregate forecasts. During training, MSE, L_{sys} and L_{topo} are jointly optimized; during testing, the trained model directly outputs $\hat{X}_{t+1:t+F}$.

4. Experiments and Results

4.1. Study area and dataset

The experiments are first conducted on a real 17-node hydro-wind-solar generation base in China, comprising 5 hydropower stations, 7 wind farms and 5 PV plants. To specify the physical graph used by the proposed model, a station-level spatial-electrical topology is constructed according to the electrical interconnections

and spatial associations among these nodes, as shown in Figure 6.

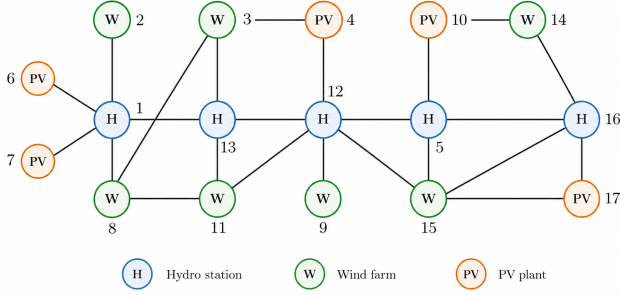


Figure 6. Station-level spatial-electrical topology of the 17-node base

Operational active-power records from 1 January to 31 December 2018 were collected at a 15-min interval.

The history length and forecast horizon are both set to $H = F = 384$, corresponding to four days. The dataset is split chronologically into training, validation and test subsets in a 7:2:1 ratio. Validation and test periods occur strictly after the training period, and all normalization statistics and correlation matrices are estimated without using future target values. This split is retained for the 17-node base. A second, 23-node base with a five-year record is examined under a rolling-origin protocol in Section 4.7.

4.2. Baselines and implementation details

Seven baselines are selected to cover decomposition-based Transformers, efficient attention models, recurrent networks and graph-based spatiotemporal models: Autoformer [24], Informer [28], iTransformer [26], BiLSTM, xLSTM, STGCN [29] and GCN-GRU [30]. All eight models share the same input channels, history length, forecast horizon, chronological partition, per-node min-max normalization and sliding-window construction. The baselines are trained with MSE. MSTD-HGFormer-HCPR is trained with the composite loss in Section 3.4.

A single hyperparameter protocol is used for every model. The shared search space takes the number of layers from $\{2, 3, 4, 5\}$, the hidden dimension from $\{64, 96, 128, 160\}$, dropout from $\{0.05, 0.10, 0.20, 0.25, 0.30\}$, the learning rate from $\{10^{-4}, 2 \times 10^{-4}, 5 \times 10^{-4}, 8 \times 10^{-4}, 10^{-3}, 1.5 \times 10^{-3}, 2 \times 10^{-3}\}$, and the batch size from $\{32, 64, 128\}$. Forty random-search trials are allocated to each model. Early stopping monitors validation total-power rRMSE with a patience of 12 epochs. Architecture-specific hyperparameters that have no counterpart in another family (decomposition scales, patch stride, node-embedding dimension, the HCPR margin τ , and the graph-fusion weights η_1, η_2, η_3) are included

in the same 40-trial budget. The configurations retained by the search are reported in Table 1. Each configuration is evaluated with five random seeds {2018, 2021, 2022, 2023, 42}. Trainable parameter counts lie between 1.38 M and 2.63 M.

Experiments are implemented in Python 3.10 on a workstation with an Intel Core i7-13620H processor and 16 GB memory.

4.3. Evaluation metrics

Four metrics are used: relative root mean square error (rRMSE), relative mean absolute error (rMAE), correlation coefficient (R), and accuracy (Acc). They are defined as

$$\begin{cases} \text{rRMSE} = \sqrt{\frac{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}{\bar{y}}} \\ \text{rMAE} = \frac{\frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|}{\bar{y}} \\ R = \frac{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sigma_y \sigma_{\hat{y}}} \\ \text{Acc} = \left(1 - \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}}{y_{\text{cap}}} \right) \times 100\% \end{cases} \quad (25)$$

where n is the number of samples, y_i and $\hat{y}_i$ are actual and predicted values, $\bar{y}$ and $\bar{\hat{y}}$ are their means, σ_y and $\sigma_{\hat{y}}$ are their standard deviations, and y_{cap} is the corresponding rated capacity. Lower rRMSE and rMAE indicate lower error, while higher R and Acc indicate better waveform tracking and capacity-normalized accuracy.

4.4. Forecasting-curve comparison

The prediction curves on the 17-node base are shown in Figs 7–10.

For hydropower output (Figure 7), Autoformer and Informer capture the general periodic trend but oscillate around peaks. BiLSTM and GCN-GRU produce smoother curves and show phase lag during ramping periods. xLSTM, STGCN and iTransformer reduce the lag, but still deviate near turning points. The proposed model tracks both peak and valley evolution more closely.

For wind-power output (Figure 8), all baselines are affected by the strong stochasticity of wind power. Informer and BiLSTM attenuate the amplitude, xLSTM fails to reproduce several sharp transitions, and GCN-GRU over-smooths local changes. Autoformer and iTransformer preserve part of the overall trend but remain unstable around troughs and rapid ramps. MSTD-HGFormer-HCPR gives the closest trajectory, especially in volatile periods.

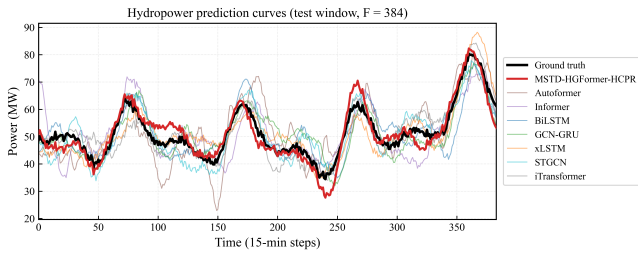


Figure 7. Hydropower prediction curves on the 17-node base

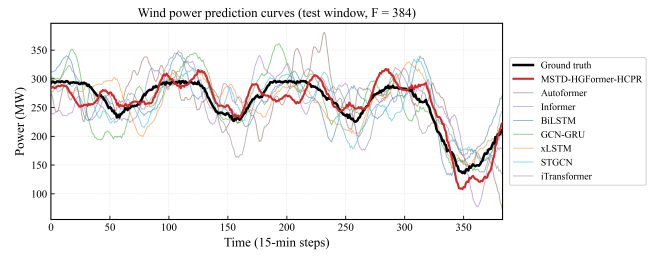


Figure 8. Wind power prediction curves on the 17-node base

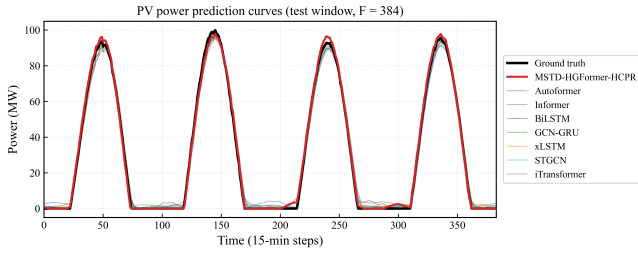


Figure 9. Photovoltaic prediction curves on the 17-node base

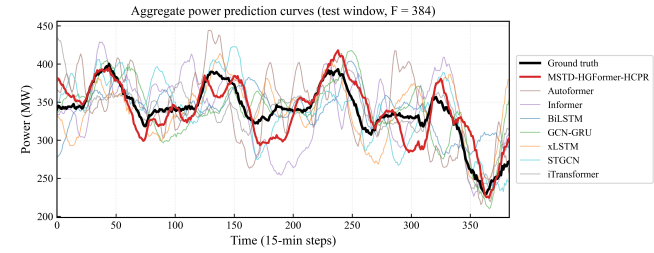


Figure 10. Aggregate power prediction curves on the 17-node base

Table 1. Selected training configuration after the unified hyperparameter search

Model	Epochs	Layers	Hidden dim	Dropout	Optimizer	Batch	Learning rate	Params (M)
Autoformer	120	3	128	0.05	Adam	64	0.0001	2.18
Informer	120	3	128	0.10	Adam	64	0.0015	2.41
iTransformer	120	3	128	0.10	Adam	64	0.0010	1.86
BiLSTM	120	2	160	0.25	Adam	128	0.0002	1.52
xLSTM	120	2	128	0.10	Adam	64	0.0008	1.94
STGCN	120	2	96	0.30	Adam	64	0.0010	1.47
GCN-GRU	120	2	128	0.20	Adam	128	0.0005	1.38
MSTD-HGFormer-HCPR	120	4	128	0.25	Adam	64	0.0008	2.63

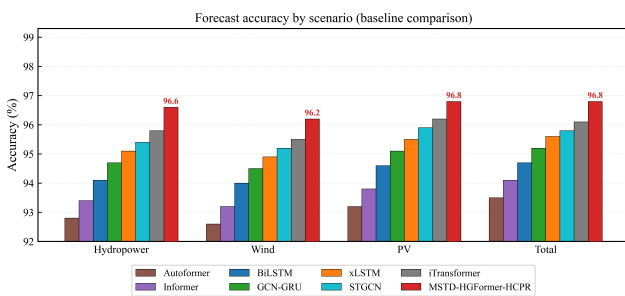


Figure 11. Forecast accuracy by output category in the baseline comparison on the 17-node base

For PV output (Figure 9), Informer, xLSTM and BiLSTM underestimate clear-sky peaks, while STGCN and GCN-GRU occasionally overshoot and decay slowly at sunset edges. iTransformer gives a stronger fit but still shows bias near extreme points. The proposed method better reproduces both daily periodicity and cloud-induced deviations. For aggregate output (Figure 10), errors from all source types are superposed.

Several baselines underfit high-output intervals or drift during rapidly changing periods, whereas the proposed model retains the best alignment with the total power curve.

4.5. Quantitative error and accuracy comparison

Tables 2–5 report the numerical results on the 17-node base for aggregate power, hydropower, wind power and PV power. Each entry is the mean over five random seeds {2018, 2021, 2022, 2023, 42}; seed-wise standard deviations are deferred to Section 4.7. Aggregate output is presented first because it corresponds to the bundled outbound power scheduled by the grid.

For aggregate output (Table 2), MSTD-HGFormer-HCPR obtains an rRMSE of 0.115, an rMAE of 0.088, an R of 0.978 and an accuracy of 96.8%. Relative to the baselines, rRMSE is reduced by 12.9–44.7% and rMAE by 16.2–48.5%. This result shows that the proposed method suppresses the accumulation of source-level errors in the total output.

Table 2. Forecasting errors and accuracy in the aggregate (bundled-power) scenario on the 17-node base

Model	rRMSE	rMAE	R	Acc (%)
Autoformer	0.208	0.171	0.932	93.5
Informer	0.197	0.163	0.939	94.1
BiLSTM	0.166	0.138	0.948	94.7
GCN-GRU	0.159	0.129	0.953	95.2
xLSTM	0.147	0.118	0.958	95.6
STGCN	0.141	0.112	0.962	95.8
iTransformer	0.132	0.105	0.967	96.1
MSTD-HGFormer-HCPR	0.115	0.088	0.978	96.8

Table 3. Forecasting errors and accuracy in the hydropower scenario on the 17-node base

Model	rRMSE	rMAE	R	Acc (%)
Autoformer	0.241	0.202	0.921	92.8
Informer	0.228	0.191	0.927	93.4
BiLSTM	0.192	0.157	0.938	94.1
GCN-GRU	0.176	0.145	0.946	94.7
xLSTM	0.163	0.131	0.952	95.1
STGCN	0.157	0.124	0.957	95.4
iTransformer	0.148	0.115	0.961	95.8
MSTD-HGFormer-HCPR	0.119	0.094	0.972	96.6

For hydropower (Table 3), the proposed method reduces rRMSE by 19.6-50.6% and rMAE by 18.3-53.5%, reaching 96.6% accuracy.

Table 4. Forecasting errors and accuracy in the wind scenario on the 17-node base

Model	rRMSE	rMAE	R	Acc (%)
Autoformer	0.273	0.228	0.912	92.6
Informer	0.258	0.211	0.919	93.2
BiLSTM	0.224	0.181	0.931	94.0
GCN-GRU	0.209	0.168	0.94	94.5
xLSTM	0.198	0.157	0.945	94.9
STGCN	0.191	0.149	0.951	95.2
iTransformer	0.182	0.141	0.956	95.5
MSTD-HGFormer-HCPR	0.145	0.118	0.969	96.2

For wind (Table 4), which is the most stochastic scenario, rRMSE and rMAE are reduced by 20.3-46.9% and 16.3-48.2%, respectively, and accuracy remains 96.2%. For PV (Table 5), rRMSE and rMAE are reduced by 20.0-49.1% and 21.2-53.4%, and the accuracy reaches 96.8%. Across all output categories, the correlation coefficient is improved by 0.011-0.013 compared with the strongest baseline in each case.

Figure 11 summarizes the accuracy comparison, and Figure 12 gives the rRMSE and rMAE statistics by output category. On the 17-node hold-out the proposed method remains at or above 96.2% in all four output

Table 5. Forecasting errors and accuracy in the PV scenario on the 17-node base

Model	rRMSE	rMAE	R	Acc (%)
Autoformer	0.212	0.176	0.929	93.2
Informer	0.198	0.165	0.938	93.8
BiLSTM	0.171	0.139	0.951	94.6
GCN-GRU	0.158	0.125	0.958	95.1
xLSTM	0.151	0.118	0.964	95.5
STGCN	0.142	0.111	0.968	95.9
iTransformer	0.135	0.104	0.971	96.2
MSTD-HGFormer-HCPR	0.108	0.082	0.982	96.8

categories and has the smallest variation across them. These results indicate that the gain is not limited to one source type; it is achieved consistently for regulated, stochastic, periodic and aggregate outputs.

4.6. Ablation study

An ablation study on the 17-node base is conducted to isolate the effect of joint graph modeling and the two complementary-prior terms. Four variants are compared with the full model. MSTD-HGFormer-S keeps the backbone but diagonalizes the spatial adjacency and removes HCPR, thereby disabling cross-source graph interaction. MSTD-HGFormer-NG removes the system-level constraint and keeps only the topology-level term. MSTD-HGFormer-NL removes the topology-level constraint and keeps only the system-level term. MSTD-HGFormer-L keeps the full graph backbone but trains only with MSE.

The results are shown in Table 6. Compared with MSTD-HGFormer-S, the full model lowers rRMSE by 30.0-38.6% and rMAE by 29.8-38.8% across the four scenarios. This verifies the importance of joint spatiotemporal modeling. Compared with MSTD-HGFormer-L, rRMSE and rMAE are reduced by 40.7-42.0% and 36.2-44.2%, respectively, showing that the complementary prior is essential even when the graph backbone is retained.

The two partial-prior variants show different strengths. MSTD-HGFormer-NG performs better on wind and PV than MSTD-HGFormer-NL, indicating that the topology-level term is useful for distinguishing neighbouring heterogeneous nodes under high volatility. MSTD-HGFormer-NL performs better on hydropower and aggregate output, indicating that the system-level term contributes more to source-level co-smoothing and total-power stability. The full model combines both advantages and achieves the best values in all four scenarios on the 17-node base. Figure 13 further confirms that, on this hold-out, its accuracy is 96.6% for hydropower, 96.2% for wind, 96.8% for PV and 96.8% for aggregate power.

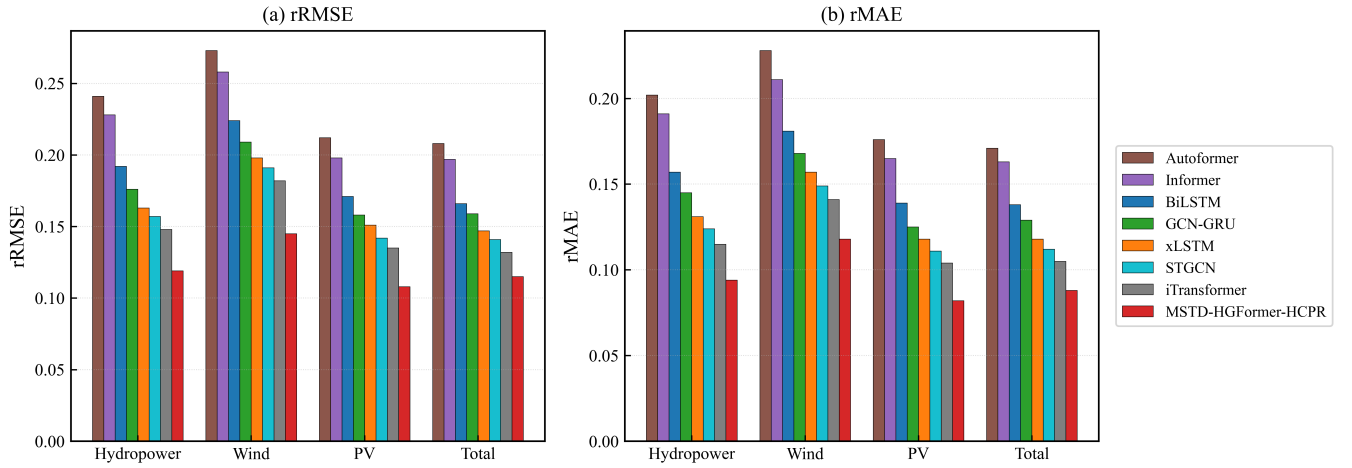


Figure 12. Baseline rRMSE and rMAE statistics by scenario on the 17-node base

Table 6. Forecasting errors of the ablation variants across the four scenarios on the 17-node base

Model	Hydropower			Wind power			PV power			Aggregate power		
	rRMSE	rMAE	R	rRMSE	rMAE	R	rRMSE	rMAE	R	rRMSE	rMAE	R
MSTD-HGFormer-S	0.183	0.141	0.958	0.207	0.168	0.950	0.176	0.134	0.964	0.179	0.142	0.961
MSTD-HGFormer-NG	0.168	0.129	0.961	0.181	0.149	0.956	0.131	0.105	0.973	0.154	0.119	0.969
MSTD-HGFormer-NL	0.141	0.112	0.965	0.211	0.165	0.948	0.148	0.122	0.967	0.136	0.104	0.972
MSTD-HGFormer-L	0.205	0.154	0.951	0.246	0.185	0.942	0.183	0.147	0.959	0.194	0.151	0.955
MSTD-HGFormer-HCPR	0.119	0.094	0.972	0.145	0.118	0.969	0.108	0.082	0.982	0.115	0.088	0.978

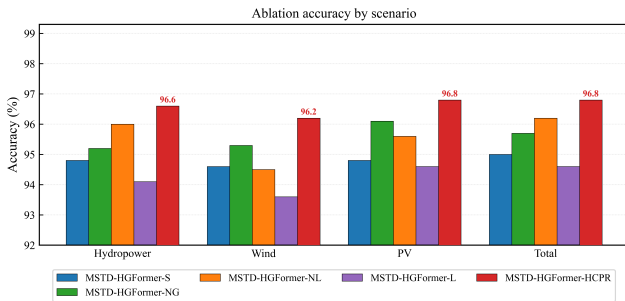


Figure 13. Ablation accuracy by scenario on the 17-node base

4.7. Extended validation on a 23-node base with cross-year and season-stratified testing

In order to examine the scalability of the proposed method, a 23-node hydro-wind-solar complementary base is further employed to verify the model ranking and the source-level accuracy. Compared with the 17-node base, this base has a different physical topology and provides a five-year record, and it is evaluated under a rolling-origin protocol [31] with season-stratified metrics.

As listed in Table 7, the 23-node base comprises 6 cascade hydropower stations, 9 wind farms and 8 PV plants, whose installed capacities are 486 MW, 337.5 MW and 441.5 MW, respectively, giving a total of 1,265

MW. Unlike the meshed multi-cluster graph of the 17-node base, the topology of this base is cascade-radial: the hydropower stations form a river chain, whereas the wind farms and PV plants are connected to the nearest cascade node through collector-style links and share one 500 kV outbound corridor. It can be seen from Table 7 that the mean degree is almost unchanged (2.94 versus 2.96), while the 23-node graph has a larger diameter (6 versus 5), more cross-source edges (17 versus 11), a higher PV share, and a five-year rather than one-year record. Therefore, the fused adjacency A^* is applied to a genuinely different interconnection topology, not merely to a longer sample of the same graph.

The active-power records of the 23-node base cover 1 January 2019 to 31 December 2023 at the same 15-min interval, and the history and forecast lengths remain $H = F = 384$. On this basis, three rolling-origin folds are constructed. Fold 1 is trained on 2019–2020, validated on the first half of 2021 and tested on the whole of 2021; Fold 2 is trained on 2019–2021 and tested on 2022; Fold 3 is trained on 2019–2022 and tested on 2023. In this way, each test year is unseen. It is worth noting that the normalization statistics and A^{corr} are recomputed inside the training portion of the corresponding fold. Once the three folds have been evaluated, every test timestamp is assigned

Table 7. Comparison of the two study bases
The test-window size of the 23-node base is the pooled length of the three full-year test folds

Item	17-node (Section 4.1)	23-node (this subsection)
Nodes (hydro / wind / PV)	17 (5 / 7 / 5)	23 (6 / 9 / 8)
Installed capacity (MW)	322.6 / 296.7 / 173.8	486 / 337.5 / 441.5
Graph class	meshed multi-cluster	cascade-radial
Undirected edges / mean degree / diameter	25 / 2.94 / 5	34 / 2.96 / 6
Cross-source neighbouring pairs	11	17
Record	2018, 15 min (35,040 steps)	2019–2023, 15 min (175,296 steps)
Evaluation	one chronological 7:2:1 split	three rolling-origin folds
Test steps (days)	3,504 (36.5)	105,120 (1,095)

to a meteorological season, namely MAM (March–May), JJA (June–August), SON (September–November) and DJF (December–February), in which JJA and DJF correspond to the wet and dry seasons of the study region, respectively. Moreover, the same unified search is reused, and the hyperparameters listed in Table 1 are not retuned on the 23-node base. Training is performed on the Intel Core i7-13620H workstation with 16 GB memory. One fold of the proposed model required about 6.4 h of wall-clock time, and the full three-fold, five-seed protocol required about 96 h of sequential CPU time.

Table 8 reports the aggregate-power metrics on each fold together with the pooled mean, in which all values are given as means $\pm$ standard deviations over the five seeds. As shown in Table 8, MSTD-HGFormer-HCPR attains a pooled rRMSE of 0.131 ± 0.004 , an rMAE of 0.101 ± 0.002 , an R of 0.973 ± 0.001 and an accuracy of $96.2 \pm 0.12\%$, and the aggregate rRMSE is thereby reduced by 13.2–44.0% relative to the seven baselines. It can be observed that the ranking remains stable across the three test years. Fold 3 (2023) yields the largest error for every model, which is consistent with a more anomalous inflow and wind year; nevertheless, the proposed model still retains the lowest error. Correspondingly, the pooled source-level accuracies of the proposed model are 96.0% for hydropower, 96.0% for wind, 96.4% for PV and 96.2% for aggregate power.

The season-wise results are presented in Tables 9 and 10 and in Figure 14. Table 9 compares the proposed model with iTransformer, the strongest baseline on the 17-node base, in every season and every output category, while Table 10 lists the aggregate rRMSE of all eight models. It can be found that the proposed model attains the lowest error in every season and every output category. The largest residual occurs for JJA wind, with an rRMSE of 0.182 ± 0.006 and an accuracy of 95.5%. This is because summer in this basin is wet, windier days are fewer, and the remaining wind events are short and poorly correlated across the nine wind farms. Although the topology-level prior still separates neighbouring hydropower

and wind nodes, the fluctuation component contains fewer repeated patterns that can be used for prediction. By contrast, DJF wind is predicted more accurately (96.4%), which is consistent with a smoother seasonal wind peak. Hydropower is most accurate in JJA (96.4%) and least accurate in DJF (95.7%), which follows the wet-to-dry contrast in reservoir inflow, whereas PV is slightly less accurate in JJA than in SON (96.0% versus 96.6%), which is consistent with more frequent cloud transients in the wet season. It should be noted that none of these seasonal gaps reverses the ranking against iTransformer.

As summarized in Tables 8–10, the ranking is unchanged in every fold and every season: the proposed model consistently retains the lowest error, and the largest margin relative to the seven baselines occurs in the wet season.

Based on the above simulation results and analysis, the proposed method is verified on the new 23-node case. Although some indices are slightly lower than those obtained on the 17-node base, they remain within an acceptable range, and the proposed method still exhibits excellent performance. With the same backbone and the same HCPR weights, the model retains the lowest error under a different interconnection topology, over three unseen years and in all four seasons. The lowest seasonal accuracies on the 23-node base are 95.5% for JJA wind and 95.7% for DJF hydropower. The residual error remains larger for brief, weakly coupled wind events and for winter hydropower, whose output is governed mainly by upstream release rather than by inflow.

5. Conclusion

This paper proposed MSTD-HGFormer-HCPR for spatiotemporal joint power forecasting of hydro-wind-solar complementary systems. The model integrates multi-scale seasonal-trend decoupling, heterogeneous graph spatiotemporal encoding and hierarchical complementary-prior regularization. The learnable decomposition separates low-frequency trend from high-frequency fluctuation; the fused graph attention

Table 8. Aggregate-power metrics on the 23-node base by rolling-origin fold (mean $\pm$ s.d. over five seeds)
 Pooled values are averaged over the three test years

Model	Fold 1 (2021)		Fold 2 (2022)		Fold 3 (2023)		Pooled	
	rRMSE	Acc (%)	rRMSE	Acc (%)	rRMSE	Acc (%)	rRMSE	Acc (%)
Autoformer	0.224 $\pm$ 0.010	93.6	0.232 $\pm$ 0.008	93.2	0.246 $\pm$ 0.013	93.0	0.234 $\pm$ 0.010	93.3
Informer	0.213 $\pm$ 0.007	93.9	0.218 $\pm$ 0.010	93.7	0.233 $\pm$ 0.011	93.4	0.221 $\pm$ 0.009	93.7
BiLSTM	0.179 $\pm$ 0.006	94.8	0.191 $\pm$ 0.009	94.5	0.199 $\pm$ 0.010	94.4	0.190 $\pm$ 0.008	94.6
GCN-GRU	0.175 $\pm$ 0.007	95.0	0.179 $\pm$ 0.006	94.8	0.191 $\pm$ 0.008	94.6	0.182 $\pm$ 0.007	94.8
xLSTM	0.163 $\pm$ 0.005	95.3	0.171 $\pm$ 0.006	95.1	0.178 $\pm$ 0.008	94.9	0.171 $\pm$ 0.006	95.1
STGCN	0.159 $\pm$ 0.006	95.5	0.161 $\pm$ 0.005	95.3	0.170 $\pm$ 0.007	95.1	0.163 $\pm$ 0.006	95.3
iTransformer	0.145 $\pm$ 0.004	95.8	0.149 $\pm$ 0.006	95.7	0.159 $\pm$ 0.006	95.4	0.151 $\pm$ 0.005	95.6
MSTD-HGFormer-HCPR	0.125 $\pm$ 0.003	96.4	0.130 $\pm$ 0.004	96.2	0.138 $\pm$ 0.005	96.1	0.131 $\pm$ 0.004	96.2

 Table 9. Season-stratified metrics on the 23-node base for the proposed model and iTransformer
 (mean $\pm$ s.d. over five seeds, pooled over folds)

Season	Model	Hydropower		Wind		PV		Aggregate	
		rRMSE	Acc (%)	rRMSE	Acc (%)	rRMSE	Acc (%)	rRMSE	Acc (%)
MAM	iTransformer	0.160 $\pm$ 0.005	95.4	0.180 $\pm$ 0.007	95.4	0.139 $\pm$ 0.004	96.0	0.152 $\pm$ 0.004	95.7
	Proposed	0.140 $\pm$ 0.004	96.0	0.152 $\pm$ 0.005	96.1	0.122 $\pm$ 0.003	96.5	0.130 $\pm$ 0.003	96.3
JJA	iTransformer	0.145 $\pm$ 0.004	95.7	0.211 $\pm$ 0.009	94.7	0.167 $\pm$ 0.007	95.4	0.165 $\pm$ 0.006	95.3
	Proposed	0.121 $\pm$ 0.003	96.4	0.182 $\pm$ 0.006	95.5	0.144 $\pm$ 0.004	96.0	0.139 $\pm$ 0.005	96.0
SON	iTransformer	0.156 $\pm$ 0.006	95.4	0.184 $\pm$ 0.005	95.4	0.133 $\pm$ 0.004	96.1	0.146 $\pm$ 0.004	95.7
	Proposed	0.136 $\pm$ 0.004	96.0	0.161 $\pm$ 0.005	95.9	0.115 $\pm$ 0.003	96.6	0.127 $\pm$ 0.003	96.3
DJF	iTransformer	0.171 $\pm$ 0.006	95.1	0.160 $\pm$ 0.006	96.0	0.144 $\pm$ 0.005	95.9	0.148 $\pm$ 0.005	95.7
	Proposed	0.150 $\pm$ 0.005	95.7	0.141 $\pm$ 0.004	96.4	0.122 $\pm$ 0.004	96.5	0.127 $\pm$ 0.004	96.3

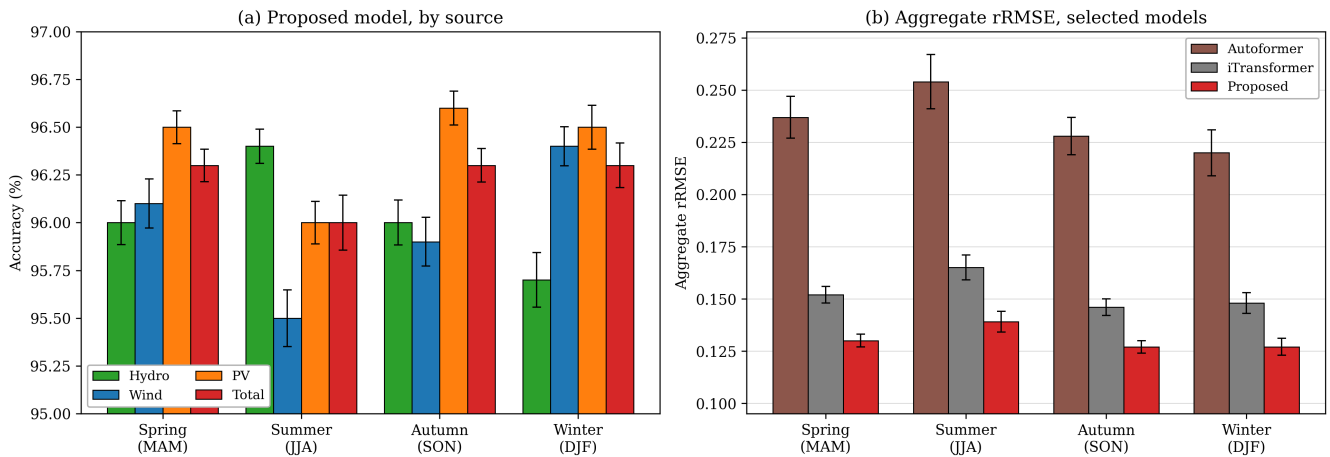


Figure 14. Season-stratified results on the 23-node base

(a) Accuracy of the proposed model by source. (b) Aggregate rRMSE of Autoformer, iTransformer and the proposed model
 Error bars are standard deviations over five seeds.

combines physical topology, historical correlation and learnable latent relations; and the composite loss constrains both system-level smoothing and cross-source topological differentiation.

On the 17-node base the proposed method outperforms seven baselines. Across hydro, wind, PV and aggregate scenarios the accuracy remains no lower than 96.2%, and the aggregate-power accuracy reaches 96.8%. On the 23-node base the ranking is unchanged:

Table 10. Season-stratified aggregate rRMSE on the 23-node base for all compared models (mean $\pm$ s.d. over five seeds)

Model	MAM	JJA	SON	DJF
Autoformer	0.237 $\pm$ 0.010	0.254 $\pm$ 0.013	0.228 $\pm$ 0.009	0.220 $\pm$ 0.011
Informer	0.227 $\pm$ 0.009	0.236 $\pm$ 0.012	0.212 $\pm$ 0.007	0.210 $\pm$ 0.008
BiLSTM	0.190 $\pm$ 0.008	0.208 $\pm$ 0.010	0.184 $\pm$ 0.006	0.180 $\pm$ 0.006
GCN-GRU	0.180 $\pm$ 0.006	0.198 $\pm$ 0.009	0.175 $\pm$ 0.006	0.177 $\pm$ 0.007
xLSTM	0.169 $\pm$ 0.005	0.183 $\pm$ 0.008	0.164 $\pm$ 0.005	0.162 $\pm$ 0.004
STGCN	0.159 $\pm$ 0.006	0.177 $\pm$ 0.008	0.158 $\pm$ 0.004	0.162 $\pm$ 0.006
iTransformer	0.152 $\pm$ 0.004	0.165 $\pm$ 0.006	0.146 $\pm$ 0.004	0.148 $\pm$ 0.005
MSTD-HGFormer-HCPR	0.130 $\pm$ 0.003	0.139 $\pm$ 0.005	0.127 $\pm$ 0.003	0.127 $\pm$ 0.004

pooled accuracies are 96.0% (hydro), 96.0% (wind), 96.4% (PV) and 96.2% (aggregate), and the aggregate rRMSE falls by 13.2–44.0% relative to the baselines. The lowest seasonal accuracies are 95.5% for JJA wind and 95.7% for DJF hydropower. The ablation results show that joint graph modeling reduces error relative to independent prediction, while HCPR further improves both waveform tracking and capacity-normalized accuracy. These findings indicate that complementarity should be represented not only in the model architecture but also in the optimization objective.

This study has several limitations. The forecasts remain deterministic point estimates. Operational value is judged by forecasting metrics rather than by unit commitment, reserve allocation or curtailment. Interval or scenario products and a dispatch-level evaluation are left to future work. Dynamic topology evolution caused by source expansion or reconfiguration will be addressed by graph incremental learning and transfer learning.

Acknowledgements

This work was supported by the Key Science and Technology Projects of Yunnan Province (No. 202503AA080001).

References

- [1] Rahman MM, Dadon SH, He M, et al. An overview of power system flexibility: High renewable energy penetration scenarios. *Energies*, 2024;17(24):6393.
- [2] Song J, Zhou X, Zhou Z, et al. Review of low inertia in power systems caused by high proportion of renewable energy grid integration. *Energies*, 2023;16(16):6042.
- [3] Zhao W, Duan Y, Xu Y, et al. A Hybrid Source-Load Power Forecasting Model for Distribution Networks with High Penetration of Renewable Energy. *EAI Endorsed Transactions on Energy Web*, 2026;13.
- [4] Ge C, Yan J, Zhang H, et al. Joint short-term power forecasting of hydro-wind-photovoltaic considering spatiotemporal delay of weather processes. *Renewable Energy*, 2024;237:121679.
- [5] Dong L, Li Y, Xiu X, et al. An integrated ultra short term power forecasting method for regional wind-pv-hydro. *Energy Reports*, 2023;9: 1531-1540.
- [6] Tang Y, Zhou Y, Liu P, et al. Revolutionizing solar-hydro-wind power forecasts in regional power systems with hybrid machine learning models. *Solar Energy*, 2025;291:113391.
- [7] Peng W, Jiang Z, Yao H, et al. Short-term hydropower generation prediction model for run-of-river hydropower plants considering hydrometeorological factors. *Renewable Energy*, 2025;254:123790.
- [8] Zhang J, Feng Z, Niu W, et al. An interpretable physics-informed artificial intelligence model for daily hydropower generation prediction in cascaded reservoirs incorporating spatiotemporal correlations. *Renewable Energy*, 2026;270:125952.
- [9] Jonkers J, Avendano D N, Van Wallendael G, et al. A novel day-ahead regional and probabilistic wind power forecasting framework using deep CNNs and conformalized regression forests. *Applied Energy*, 2024;361:122900.
- [10] Michael N E, Hasan S, Al-Durra A, et al. Short-term solar irradiance forecasting based on a novel Bayesian optimized deep Long Short-Term Memory neural network. *Applied Energy*, 2022;324:119727.
- [11] Gao Q. Multi-temporal Scale Wind Power Forecasting Based on Lasso-CNN-LSTM-LightGBM. *EAI Endorsed Transactions on the Energy Web*, 2024;11(1).
- [12] Chen H, Zhang Z, Tong S, et al. An Ultra-Short-Term Wind Power Prediction Method Based on Quadratic Decomposition and Multi-Objective Optimization. *EAI Endorsed Transactions on the Energy Web*, 2024;11(1).
- [13] Liu S, Gong B, Cui C, et al. Photovoltaic Power Adaptive Hybrid Forecasting Model Integrated with Multi-Dimensional Error Compensation. *EAI Endorsed Transactions on Energy Web*, 2026;12.
- [14] Wang L, He Y. M2STAN: Multi-modal multi-task spatiotemporal attention network for multi-location ultra-short-term wind power multi-step predictions. *Applied Energy*, 2022;324: 119672.
- [15] Zhao Y, Ye L, Pinson P, et al. Correlation-constrained and sparsity-controlled vector autoregressive model for spatio-temporal wind power forecasting. *IEEE Transactions on Power Systems*, 2018;33(5): 5029-5040.
- [16] Song Y, Tang D, Yu J, et al. Short-term forecasting based on graph convolution networks and multiresolution convolution neural networks for wind power. *IEEE Transactions on Industrial Informatics*, 2022; 19(2):1691-1702.
- [17] Li B, Ding B, Pang W, et al. Triple-Flow Dynamic Graph Convolutional Network for Wind Power Forecasting. *Symmetry*, 2025;17(12):2026.
- [18] Qiu H, Shi K, Wang R, et al. A novel temporal-spatial graph neural network for wind power forecasting considering blockage effects. *Renewable Energy*, 2024;227:120499.
- [19] Shi Z, Zhang G, Lu C, et al. Dynamic Spatio-Temporal Graph-Enhanced KANformer for high-fidelity ultra-short-term wind power forecasting. *Energy*, 2025;337:138590.
- [20] Khodayar M, Liu G, Wang J, et al. Spatiotemporal behind-the-meter load and PV power forecasting via deep graph dictionary learning. *IEEE transactions on neural networks and learning systems*, 2020;32(10):4713-4727.
- [21] Simeunovic J, Schubnel B, Alet P J, et al. Spatio-temporal graph neural networks for multi-site PV power forecasting. *IEEE Transactions on Sustainable Energy*, 2021;13(2):1210-1220.
- [22] Zhang Y, Zhang Y, Pan D, et al. Distributed photovoltaic power prediction considering spatiotemporal correlation and dual Attention-LSTM. *EAI Endorsed Transactions on Energy Web*, 2025;12.
- [23] Ni C, Chen Z, Tang Y, et al. Integrated Wind-Solar-Hydro Power Prediction Method Based on Deep Neural Network. *Proceedings of the 2022 4th International Conference on Electrical Engineering and Control Technologies (CEEET)*, Shanghai, China, 16-18 December 2022. *IEEE*, 2022; 210-215.
- [24] Wu H, Xu J, Wang J, et al. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Advances in neural information processing systems*, 2021;34:22419-22430.
- [25] Wang Y, Zhu J, Kang R. Destformer: A transformer based on explicit seasonal-trend decomposition for long-term series forecasting. *Applied Sciences*, 2023;13(18):10505.
- [26] Liu Y, Hu T, Zhang H, et al. itransformer: Inverted transformers are effective for time series forecasting[C]//International

- conference on learning representations. 2024;2024:11116-11140.
- [27] Xu H, Zhang J, Cheng C, et al. A novel metric for evaluating hydro-wind-solar energy complementarity. *Applied Energy*, 2024;373:123926.
- [28] Zhou H, Zhang S, Peng J, et al. Informer: Beyond efficient transformer for long sequence time-series forecasting[C]//Proceedings of the AAAI conference on artificial intelligence. 2021;35(12):11106-11115.
- [29] Wang H, Zheng X, Lai W. Short-term power prediction of regional distributed photovoltaic power plants based on spatiotemporal information fusion. *Journal of Renewable and Sustainable Energy*, 2025;17(4):046103.
- [30] Lu J, Xin Z, Chi W, et al. A Hybrid GCN-GRU Model with Attention Mechanism for Ultra-Short-Term Photovoltaic Power Forecasting. *Proceedings of the 2025 37th Chinese Control and Decision Conference (CCDC)*. Xiamen, China, 16-19 May 2025; IEEE, 2025;2650-2656.
- [31] Bergmeir C, Benítez JM. On the use of cross-validation for time series predictor evaluation. *Information Sciences*, 2012;191:192-213. doi:10.1016/j.ins.2011.12.028.