

Signal-Guided Dual-Predictor Learning for Ultra-Short-Term PV Power Forecasting

Shumei Zhang¹, Zhiwen Wang^{1,2,*}, Haoxu Wang¹, Wannian Wei¹, Wei Qi¹ and Hao Liu¹

¹School of Automation and Electrical Engineering, Lanzhou University of Technology, Lanzhou, China

²Key Laboratory of Gansu Advanced Control for Industrial Processes, Lanzhou University of Technology, Lanzhou, China.

*Corresponding author Email: 3097818815@qq.com

Abstract: To address the insufficient adaptability of a unified predictor caused by the non-stationarity of photovoltaic (PV) power series, pronounced multi-scale fluctuations, and differences in the predictable structures of decomposed subcomponents, this paper proposes an ultra-short-term PV power forecasting method based on CEEMDAN-VMD and signal-property-guided dual predictors. First, Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) is employed to decompose the original power series, and the resulting Intrinsic Mode Functions (IMFs) are reconstructed into high-, medium-, and low-frequency sequences using K-means clustering. Subsequently, Variational Mode Decomposition (VMD) is applied only to the high-frequency component for secondary decomposition. Furthermore, Sample Entropy (SE), Central Frequency (CF), Energy Proportion (EP), First-order Autocorrelation Coefficient (Lag1ACF), and Fluctuation Index (FI) are used to characterize the properties of the final components, based on which differentiated modeling is performed. Specifically, H1, H2, and Low are modeled using the Regularity Predictor (RP), whereas H3–H6 and Mid are modeled using the Dynamic Predictor (DP), which integrates a multi-scale dilated Temporal Convolutional Network (TCN) with a Gated Recurrent Unit (GRU). The forecasting results of all components are directly summed to obtain the final PV power forecast. Experiments conducted on seasonal datasets from a PV power station in Northwest China showed that, compared with several representative deep learning models, the proposed method reduced the Mean Absolute Error (MAE) by 72.49%, 61.54%, 61.79%, and 44.10% in spring, summer, autumn, and winter, respectively, and reduced the Root Mean Square Error (RMSE) by 79.91%, 68.61%, 77.33%, and 63.16%, respectively, with the coefficient of determination (R^2) ranging from 0.9948 to 0.9991. The Wilcoxon test indicated that the performance improvements were statistically significant. On the independent Alice Springs dataset in Australia, the MAE, RMSE, and coefficient of determination (R^2) were 0.0793, 0.1308, and 0.9988, respectively, further validating the generalization capability of the model. The results demonstrate that component-level differentiated forecasting based on signal properties can effectively improve the accuracy and model adaptability of ultra-short-term PV power forecasting.

Keywords: Renewable Energy; Deep Learning; Dual-predictor Learning; Multi-scale TCN; GRU.

1. Introduction

With the continuously increasing penetration of renewable energy in power systems, PV power generation has become an important technological pathway for advancing the low-carbon transition of the energy sector. However, rapid variations in solar irradiance, cloud movement, and local meteorological conditions cause PV power to exhibit pronounced randomness, intermittency, and multi-timescale fluctuations. Particularly in minute-level ultra-short-term forecasting scenarios, PV power curves contain not only relatively stable trend evolution but also superimposed rapid oscillations and local abrupt changes, making it difficult to reliably extract dynamic patterns from a limited historical window. Accurate characterization of short-term future power variations is of great significance for reducing the impact of renewable energy uncertainty on power grid scheduling, improving reserve resource allocation, and enhancing operational reliability under high PV penetration. Recent studies have shown that factors such as forecasting model type, climatic conditions, and forecasting horizon can all affect PV power forecasting performance. Therefore, developing forecasting models that simultaneously account for accuracy, stability, and adaptability to different signal structures remains an important research direction in this field [1]

With the development of computer technology, deep

learning has gradually become an important approach for PV power forecasting. Recurrent neural networks can characterize temporal dependencies through gated states, convolutional networks can efficiently extract local variation patterns, and architectures such as Transformers further enhance the capability of modeling long-range dependencies. Limouni et al. [2] combined Long Short-Term Memory (LSTM) with a Temporal Convolutional Network (TCN) for ultra-short-term PV forecasting, improving the accuracy of both single-step and multi-step forecasting through the complementarity of temporal memory and temporal convolution. Wang et al. [3] developed an efficient shrinkage TCN, in which residual shrinkage and adaptive thresholds were employed to enhance the suppression of noisy information. Dai et al. [4] established a hybrid forecasting model incorporating a Convolutional Neural Network (CNN), an attention mechanism, and a Bidirectional Gated Recurrent Unit (BiGRU) to improve forecasting performance across different seasons. Xiong et al. [5] employed a Transformer with Gaussian process approximation for probabilistic solar forecasting. These studies indicate that combining multiple deep neural networks can enhance the representation of complex temporal relationships. However, continuously increasing network depth or introducing additional attention modules and feature extraction structures also leads to larger parameter sizes and higher modeling complexity. More importantly, when the original power series simultaneously

contains multi-scale components with distinctly different properties, relying solely on a unified complex network cannot fundamentally eliminate the mutual interference among information at different timescales.

In recent years, hybrid frameworks combining signal decomposition with deep forecasting models have attracted extensive attention. The basic idea is to first reduce the non-stationarity of the original PV series and then perform targeted modeling of the resulting subsequences. Liu et al. [6] employed Variational Mode Decomposition (VMD) to decompose PV power at different timescales and combined it with a convolutional network for multi-horizon forecasting. Wang et al. [7] used Ensemble Empirical Mode Decomposition (EEMD) to reconstruct the original PV time series into subsequences with different frequencies and subsequently established an optimized LSTM-based forecasting model. Wu et al. [8] further combined VMD and CEEMDAN to construct a dual-decomposition structure to enhance the separation of complex PV power series. More recently, Udenze et al. employed CEEMDAN together with a secondary decomposition strategy for ultra-short-term solar power forecasting and indicated that two-stage signal processing can effectively reduce the forecasting uncertainty caused by fluctuations at different frequencies. Liu et al. [9] also constructed IMF features by jointly applying wavelet denoising and VMD, further demonstrating the improvement in forecasting performance brought by decomposition information. These studies show that shifting from direct forecasting of the original series to a decomposition-forecasting-reconstruction framework can effectively reduce the learning difficulty associated with complex non-stationary series.

However, an issue in existing decomposition-based forecasting frameworks still deserves further investigation: after a signal is decomposed, should different subcomponents still be processed by the same forecasting model. Most existing methods focus on optimizing decomposition algorithms, deepening network architectures, or tuning hyperparameters, while insufficient attention has been paid to differences in the intrinsic predictable structures of individual components. In recent years, some studies have begun to explore heterogeneous modeling strategies. For example, Zhou et al. [10] employed linear AR and nonlinear DCC-BiLSTM models in parallel to exploit the complementarity between linear and nonlinear structures. Xue and Li [11] further assigned different forecasting models to high- and low-frequency components. These studies suggest that configuring predictors with different capacities according to different data structures has research value. However, existing mappings are still largely based on frequency ranges or predefined groupings. In fact, high frequency does not necessarily imply greater forecasting difficulty, nor does low frequency necessarily mean that a simple model is sufficient. Some high-frequency components may still exhibit strong regularity if they possess stable negative autocorrelation or periodic alternating patterns, whereas some lower-frequency components may exhibit higher complexity and stronger local fluctuations. Therefore, using frequency position alone as the criterion for determining predictor complexity is insufficient to reveal the true predictable structure of individual subcomponents.

Based on the above analysis, this paper proposes an ultra-short-term PV power forecasting method based on CEEMDAN-VMD and signal-property-guided dual-

predictor learning. Instead of directly equating component frequency with forecasting difficulty, the proposed method characterizes different subsequences from multiple dimensions, including complexity, dominant timescale, energy contribution, dependence between adjacent samples, and local fluctuation intensity, and accordingly matches model capacity with the predictable structure of each signal component. The main contributions of this paper are summarized as follows:

(1) To address the problem that a single decomposition may be insufficient to fully characterize complex high-frequency fluctuations, whereas repeatedly applying secondary decomposition to all components may lead to over-decomposition, a selective secondary decomposition strategy based on CEEMDAN-VMD is proposed. First, CEEMDAN is used to obtain multi-scale IMFs, and SE and CF are employed as features to reconstruct the IMFs into high-, medium-, and low-frequency sequences through K-means clustering. Subsequently, VMD-based secondary decomposition is applied only to the reconstructed high-frequency sequence that still exhibits complex and rapid fluctuations. In this way, high-frequency dynamics are further refined while the complete structures of the medium- and low-frequency components are preserved.

(2) To address the model-capacity mismatch caused by assigning a unified predictor to different subcomponents in conventional decomposition-based forecasting methods, a dual-predictor mechanism guided by multidimensional signal properties is constructed. SE, CF, ER, Lag1ACF, and FI are jointly used to characterize the final components, and differentiated modeling is implemented according to the corresponding predictor-component matching relationships. H1, H2, and Low, which exhibit stable local regularities, are modeled using a lightweight Regularity Predictor (RP) based on local-level normalization. In contrast, H3-H6 and Mid are modeled using a multi-scale dilated TCN-GRU Dynamic Predictor (DP) to extract complex local variations and temporal dependencies, thereby avoiding the indiscriminate use of complex networks for all components.

(3) The effectiveness of the proposed framework is validated under multiple scenarios. In addition to independent forecasting experiments across the four seasons and comparisons with several representative deep learning models, ablation experiments on the decomposition strategy and predictor configuration are further conducted to analyze the contribution of each module. Daily forecasting errors, the Wilcoxon signed-rank test, and Holm correction for multiple comparisons are also employed to evaluate the statistical reliability of the performance differences. In addition, PV data from Alice Springs, Australia, are used for cross-dataset validation to examine the generalization capability of the proposed model.

The remainder of this paper is organized as follows. Section 2 introduces the CEEMDAN-VMD selective secondary decomposition strategy, signal-property characterization, the Regularity Predictor, the Dynamic Predictor, and the overall dual-predictor forecasting framework. Section 3 describes the experimental data, parameter settings, evaluation metrics, and experimental results, followed by analyses of decomposition performance, ablation experiments, model comparisons, statistical significance, and generalization performance. Finally, the conclusions are presented and directions for future research are discussed.

2. Experimental Methods

2.1. Data Decomposition

2.1.1. CEEMDAN

CEEMDAN is an adaptive signal decomposition method developed from EMD-based methods [12]. Conventional EMD can decompose a non-stationary series into several IMFs according to the local-scale characteristics of the data itself. However, when processing signals with complex fluctuations, mode mixing can easily occur, whereby information at different timescales coexists within the same IMF or information with similar scales is distributed across multiple IMFs. By introducing adaptive noise at different decomposition stages and averaging the results of multiple ensemble decompositions, CEEMDAN can alleviate mode mixing while retaining the data-driven characteristics of the decomposition process, thereby improving the separability of components at different scales. Therefore, this method is suitable for non-stationary time series such as PV power, which simultaneously contain trend variations, periodic fluctuations, and random disturbances.

Let $w^{(r)}(t)$ denote the zero-mean white-noise sequence added in the r -th realization, R denote the number of ensemble decompositions, and $\mathcal{E}_j(\cdot)$ denote the j -th IMF obtained from the input signal by EMD. CEEMDAN first extracts the first IMF from multiple noise-added sequences and then performs ensemble averaging to obtain the first mode:

$$c_1(t) = \frac{1}{R} \sum_{r=1}^R \mathcal{E}_1 [x(t) + \varepsilon_0 w^{(r)}(t)] \quad (1)$$

After the first IMF is obtained, the first-order residual is defined as $r_1(t) = x(t) - c_1(t)$. For the ℓ -th decomposition stage, adaptive noise at the corresponding scale is added to the residual signal obtained from the previous stage, and its first IMF is then extracted. The general form is expressed as:

$$c_\ell(t) = \frac{1}{R} \sum_{r=1}^R \mathcal{E}_1 \left[r_{\ell-1}(t) + \varepsilon_{\ell-1} \mathcal{E}_{\ell-1}(w^{(r)}(t)) \right], \quad \ell \geq 2 \quad (2)$$

where $\varepsilon_{\ell-1}$ denotes the adaptive noise amplitude corresponding to the $\ell - 1$ -th decomposition stage. After each new IMF is obtained, the residual is updated as:

$$r_\ell(t) = r_{\ell-1}(t) - c_\ell(t) \quad (3)$$

When the residual signal can no longer generate an IMF satisfying the decomposition conditions, the decomposition process terminates. The original series can then be expressed as

$$x(t) = \sum_{\ell=1}^M c_\ell(t) + r_M(t) \quad (4)$$

where M denotes the total number of IMFs finally obtained, and $r_M(t)$ denotes the final residual.

2.1.2. VMD

After reconstruction using CEEMDAN and K-means clustering, the low- and medium-frequency components respectively represent relatively slow trend variations and medium-scale dynamics, whereas the high-frequency component may still contain rapid fluctuations at multiple timescales. If secondary decomposition is further applied to all three reconstructed sequences, the number of final components will not only increase, but relatively stable components may also be excessively refined. Therefore, a selective secondary decomposition strategy is adopted in this paper, in which VMD is applied only to the reconstructed high-frequency sequence $x_H(t)$, thereby refining complex high-frequency fluctuations while preserving the complete

dynamic structures of the low- and medium-frequency components.

VMD represents the input signal as a set of narrow-band modes with finite bandwidths distributed around different center frequencies. Suppose that the high-frequency sequence $x_H(t)$ is decomposed into K modes $u_k(t)$, with the corresponding center angular frequencies denoted by ω_k . First, the analytic signal of each mode is constructed using the Hilbert transform. The spectrum is then shifted to the baseband by multiplication with a complex exponential, and the mode bandwidth is characterized by estimating the squared norm of the gradient of the demodulated signal. The basic constrained variational model of VMD is expressed as:

$$\min_{\{u_k\}, \{\omega_k\}} \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2, \quad \text{s.t.} \quad \sum_{k=1}^K u_k(t) = x_H(t) \quad (5)$$

To solve the above constrained optimization problem, a quadratic penalty term and a Lagrange multiplier are introduced. The corresponding augmented Lagrangian can be expressed as:

$$\mathcal{L} = \alpha \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 + \left\| x_H(t) - \sum_{k=1}^K u_k(t) \right\|_2^2 + \langle \lambda(t), x_H(t) - \sum_{k=1}^K u_k(t) \rangle \quad (6)$$

where α is the penalty factor associated with the bandwidth constraint, and $\lambda(t)$ is the Lagrange multiplier. By alternately updating u_k , ω_k and λ , narrow-band modes with different center frequencies can be progressively obtained.

2.1.3. Multidimensional Signal Characterization of the Final Components

Assigning forecasting models solely according to high-, medium-, and low-frequency labels essentially still determines the forecasting difficulty of different components mainly on the basis of their frequency positions. However, frequency level alone cannot fully reflect the predictable structure of a time series. For example, although high-frequency components vary rapidly, they may still exhibit strong local regularity if they possess stable positive-negative alternating relationships. In contrast, some medium-frequency components, despite having relatively lower frequencies, may contain complex nonlinear dynamics. Therefore, after obtaining the final eight components, this paper further quantitatively characterizes them from five dimensions: complexity, dominant timescale, energy contribution, correlation between adjacent time instants, and local fluctuation intensity.

(1) Sample Entropy. SE is used to measure the complexity and irregularity of a time series [13]. For a sequence of length N , with embedding dimension m and tolerance r , the sample entropy is expressed as:

$$SE(m, r, N) = \ln[A^m(r)/B^m(r)] \quad (7)$$

Where σ denotes the standard deviation of the corresponding component. A larger SE indicates a more complex local structure and fewer repetitive patterns in the sequence.

(2) Central Frequency. The Fast Fourier Transform (FFT) is applied to the detrended component, and the spectral centroid is calculated based on the one-sided power spectrum $P(f)$ as:

$$CF = [\sum_f f \cdot P(f)] / [\sum_f P(f)] \quad (8)$$

CF reflects the dominant timescale at which the main energy of the component is concentrated.

(3) Energy Ratio. Let the energy of the i -th component be $E_i = \sum_t x_i^2(t)$, then:

$$ER_i = E_i / \sum_j E_j \quad (9)$$

ER is used to measure the contribution of each component to the total signal energy.

(4) First-order Autocorrelation Coefficient. In this paper, ACF1 is calculated in the form of the Pearson correlation coefficient between the sequences at two adjacent time instants:

$$ACF1 = \frac{\sum_t (x_t - \bar{x}_1)(x_{t+1} - \bar{x}_2)}{\sqrt{[\sum_t (x_t - \bar{x}_1)^2][\sum_t (x_{t+1} - \bar{x}_2)^2]}} \quad (10)$$

The closer $|ACF1|$ is to 1, the stronger the stable relationship between adjacent samples. A positive value indicates strong continuity, whereas a negative value indicates an evident alternating variation pattern.

(5) Fluctuation Index. To quantify the magnitude of changes between adjacent time instants relative to the fluctuation level of the sequence itself, the FI is defined as:

$$FI = \text{std}(\Delta x) / [\text{std}(x) + \varepsilon] \quad (11)$$

A larger FI indicates more intense local variations, whereas a smaller FI indicates a smoother component.

2.2. Dual-Predictor Forecasting Model

2.2.1. Regularity Predictor

RP is mainly designed for subcomponents exhibiting strong linear regularity, pronounced positive or negative autocorrelation, or slowly varying trends. For such sequences, the state at the next time step is usually closely related to the relative variation pattern over the most recent historical period. Therefore, it is unnecessary to introduce a large number of nonlinear parameters for feature transformation. Based on this characteristic, this paper adopts a local-level normalization strategy similar to that used in NLinear. Specifically, the last observation in the input window is taken as a reference, so that the learning target is transformed from the absolute level of the sequence to its relative variation. The network architecture is shown in Fig. 1.

For the q -th component, a standardized historical window of length L is constructed at time t as:

$$s_{q,t} = [s_q(t-L+1) \quad s_q(t-L+2) \quad \cdots \quad s_q(t)]^T \quad (12)$$

where $s_q(t)$ denotes the last observation in the current historical window. Taking this value as the local-level baseline, all historical samples in the window are shifted as:

$$s_{q,t} = s_{q,t} - s_q(t) \mathbf{1}_L \quad (13)$$

where $\mathbf{1}_L$ denotes an all-ones vector of length L . After this transformation, the last element of the window becomes zero, whereas the remaining elements represent the variation amplitudes of the corresponding historical observations relative to the current time. Therefore, the model no longer focuses on learning the absolute level of the component itself, but instead concentrates on extracting relative variation relationships within the historical window.

The RP employs only a single fully connected layer to predict the increment at the next time step relative to the current local level:

$$\Delta \hat{s}_q(t+1) = \mathbf{w}_q^T s_{q,t} + b_q \quad (14)$$

Where $\mathbf{w}_q \in \mathbb{R}^L$ and b_q denote the trainable weight vector and bias, respectively. The predicted increment is then added back to the last observation in the historical window:

$$\hat{s}_q^{RE}(t+1) = s_q(t) + \Delta \hat{s}_q(t+1) \quad (15)$$

Finally, inverse normalization is performed using the normalization parameters determined from the development set:

$$\hat{z}_q^{RE}(t+1) = \sigma_q \hat{s}_q^{RE}(t+1) + \mu_q \quad (16)$$

This structure contains very few parameters, which can effectively reduce model redundancy for regular components, while remaining suitable for describing slowly varying continuous trends and stable positive-negative alternating patterns.

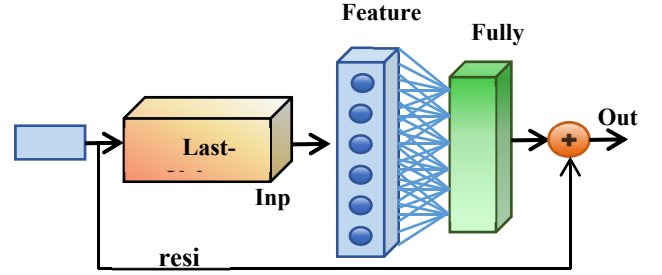


Fig. 1 RP Network Architecture

2.2.2. Dynamic Predictor

Although the non-stationarity of the medium- and high-frequency components is reduced after signal decomposition, these components still contain pronounced local abrupt variations, multi-scale fluctuations, and temporal dependencies. The receptive field of single-scale convolution is relatively fixed, making it difficult to simultaneously extract dynamic patterns with different durations, whereas purely linear models have limited capability in representing complex nonlinear variations. Therefore, this paper constructs a DP composed of a multi-scale dilated TCN and a GRU, thereby combining local multi-scale feature extraction with temporal state modeling.

Let the initial input sequence be $S_{q,t}$. The output of the Stem layer is expressed as:

$$(\cdot) F_{q,0} = \phi \text{ Conv}_{1,1}^{\text{causal}}(S_{q,t}) \quad (17)$$

where $\phi(\cdot)$ denotes the ReLU activation function.

To simultaneously extract dynamic patterns at different timescales, three parallel TCN branches are constructed after the Stem layer. For an arbitrary input feature F , the one-dimensional dilated causal convolution can be expressed as:

$$[\cdot] \text{ Conv}_{k,d}(F)_{c,t} = b_c + \sum_{c'=1}^{C_{in}} \sum_{j=0}^{k-1} W_{c,c',j} F_{c',t-dj} \quad (18)$$

where k denotes the convolution kernel size, d denotes the dilation factor, C_{in} denotes the number of input channels, and $W_{c,c',j}$ denotes the convolution kernel parameters.

The outputs of the three branches are subsequently concatenated along the feature-channel dimension:

$$F_{q,\text{cat}} = F_{q,1} \oplus F_{q,2} \oplus F_{q,3} \quad (19)$$

where $\oplus$ denotes concatenation along the channel dimension. Since each branch contains 16 feature channels, the concatenated output forms a 48-dimensional multi-scale feature representation.

To enable interaction among information at different scales and compress the feature channels, a 1×1 convolution is further employed to remap the concatenated features into a 32-dimensional feature space:

$$[\cdot] F_{q,\text{fus}} = \phi \text{ Conv}_{1,1}(F_{q,\text{cat}}) \quad (20)$$

Relying solely on multi-scale convolutional fusion may weaken the original local representations during successive feature transformations. To preserve the basic temporal information contained in the Stem layer, a residual connection is established between the fused features and the initial projected features:

$$F_{q,R} = F_{q,\text{fus}} + F_{q,0} \quad (21)$$

The residual connection preserves the original low-level features while introducing multi-scale dynamic information, thereby improving the complementarity among information at different scales.

The multi-scale TCN is mainly responsible for extracting local dynamic patterns, whereas the final forecasting process still requires the integration of state dependencies among different temporal positions within the entire historical window. Therefore, the residual-fused feature sequence is further fed into the GRU, and the hidden state at the final time step is used as the integrated temporal representation of the current window:

$$\mathbf{h}_{q,t} = \text{GRU}(\mathbf{F}_{q,R}(:, t), \mathbf{h}_{q,t-1}) \quad (22)$$

The hidden state corresponding to the final time step is then taken as:

$$\mathbf{h}_q = \mathbf{h}_{q,L} \quad (23)$$

and is used as the temporal feature for subsequent regression forecasting. By introducing the GRU after the TCN, sequential dependencies within the historical window can be further integrated on the basis of multi-scale local pattern extraction, thereby enhancing the model's capability to characterize the temporal behavior of complex dynamic components.

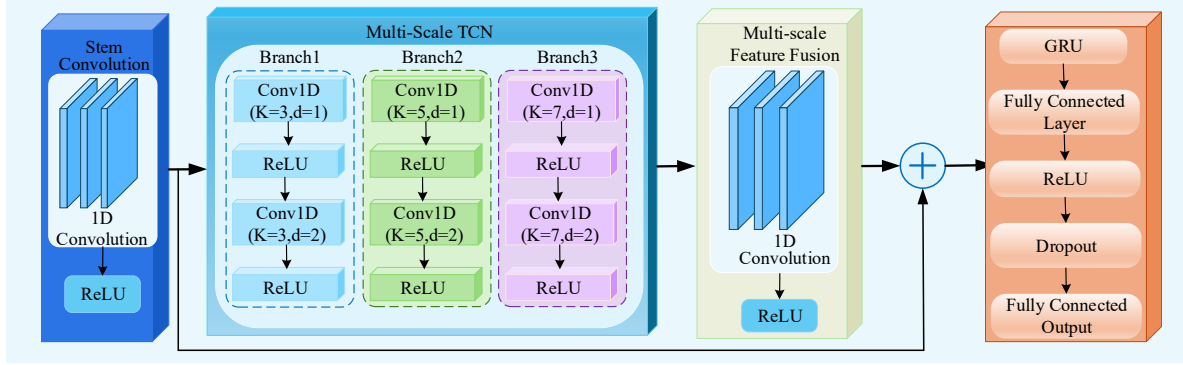


Fig. 2 Architecture of the DP Network

2.2.3. Secondary Decomposition and RP–DP Dual-Predictor Forecasting Framework

To address the pronounced nonlinearity and non-stationarity of PV power series, as well as the substantial differences in the predictable structures of components at different timescales, this study proposes an ultra-short-term PV power forecasting framework based on CEEMDAN–VMD and signal-property-guided dual predictors. The framework first employs CEEMDAN to adaptively decompose the original PV power series and then performs clustering-based reconstruction according to the SE and CF of each IMF. Subsequently, VMD-based secondary decomposition is applied only to the reconstructed high-frequency sequence with more complex dynamic variations to further refine rapid fluctuations at different timescales. On this basis, the signal properties of the final components are characterized from multiple dimensions, including complexity, frequency, energy, autocorrelation, and fluctuation intensity. According to the matching relationship between the components and predictors, different components are then assigned to the RP and DP for differentiated forecasting. Finally, the predicted values of all components are directly summed to reconstruct the final PV power forecast. The overall procedure of the proposed model is illustrated in Fig. 3, and the specific steps are described as follows.

Step 1: Data partitioning and preprocessing. The PV power series is constructed in chronological order. During the parameter and model-architecture selection stage, the dataset is sequentially divided into 70% training data, 15% validation data, and 15% held-out test data. After the model architecture and parameters are determined, the first 85% of the data are recombined to form the development set, which is used for the final fitting of the decomposition rules and model training, while the remaining 15% are retained as an independent test set for final performance evaluation.

Step 2: CEEMDAN–VMD selective secondary decomposition. First, CEEMDAN is employed to decompose

the original PV power series into several IMFs with different timescales. The SE and CF of each IMF are then calculated, and the two types of features are standardized before being input into the K-means clustering algorithm. According to the average center frequency of the IMFs contained in each cluster, the three clustering results are sequentially defined as the high-, medium-, and low-frequency groups, and mode reconstruction is performed for each group. Subsequently, VMD-based secondary decomposition is applied only to the high-frequency component, whereas the original structures of the reconstructed medium- and low-frequency components are retained.

Step 3: Signal-property analysis and determination of predictor mapping. For each of the eight final components, SE, CF, ER, ACF1, and FI are calculated to characterize its dynamic properties from five dimensions: complexity, dominant timescale, energy contribution, regularity between adjacent time instants, and local fluctuation intensity. The fixed mapping relationship between components and predictors is further determined by considering the forecasting suitability of the two predictors for different components. Ultimately, H1, H2, and Low are modeled using the RP, whereas H3–H6 and Mid are modeled using the DP.

Step 4: Differentiated forecasting using the dual predictors. For components exhibiting strong local regularity, stable positive or negative autocorrelation, or slowly varying characteristics, the RP is employed for forecasting. Based on an NLinear-like local-level normalization strategy, the RP removes the local level and establishes a lightweight linear incremental forecasting model, thereby extracting stable variation patterns with relatively few parameters. For components exhibiting stronger multi-scale dynamics and complex temporal dependencies, the DP is employed. The DP first performs feature projection through causal convolution and then uses parallel multi-scale TCN branches to extract local dynamic features at different timescales. The multi-scale features are subsequently concatenated and fused through convolution, while a residual connection is introduced to

preserve the basic temporal information. A GRU is then used to further model temporal dependencies within the fused features, and a fully connected regression layer finally outputs the predicted value of the component at the next time step. Through this differentiated modeling strategy, the model is matched to the predictable structures of different components, reducing unnecessary complex modeling for regular components while enhancing the feature extraction capability for complex dynamic components.

Step 5: Forecast reconstruction and performance evaluation. After obtaining the one-step-ahead forecasting

results of the eight subcomponents, the predicted values of all components are directly summed according to the additive relationship of signal decomposition to obtain the final PV power forecast. The MAE, RMSE, and R² are employed to evaluate the forecasting performance of the model. Experiments across different seasons, comparisons with different models, and ablation experiments are further conducted to analyze the effects of the proposed selective secondary decomposition strategy, dual-predictor architecture, and differentiated forecasting mechanism on ultra-short-term PV power forecasting performance.

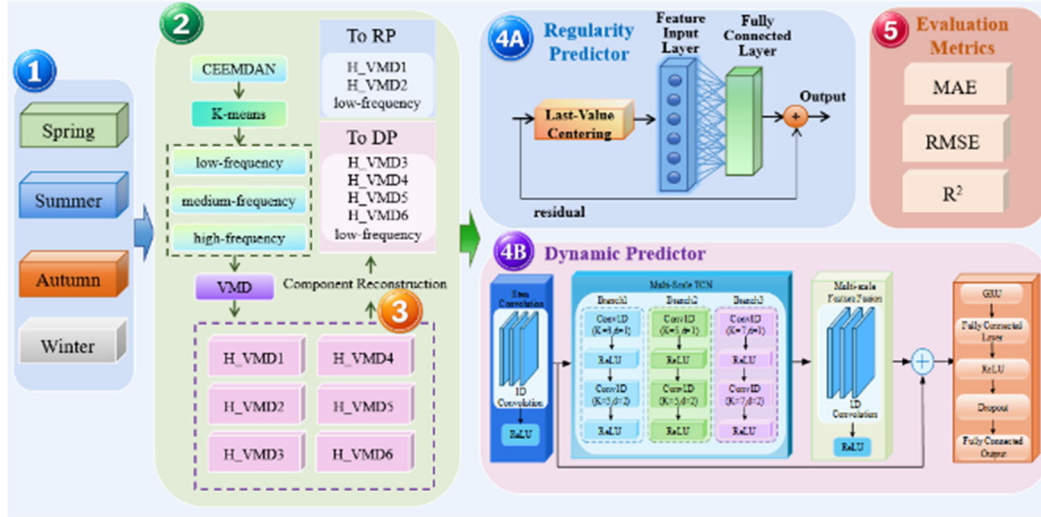


Fig. 3 Overall framework of the proposed dual-predictor model for PV power forecasting.

3. Literature References

3.1. Dataset Description

The experiments were conducted using a dataset collected from a PV power station in Northwest China in 2019, with a sampling interval of 15 min. The data were divided chronologically into four seasonal subsets corresponding to spring, summer, autumn, and winter. The dataset was initially

split into 70% training data, 15% validation data, and 15% test data. After the model architecture and hyperparameters were determined, the final model was retrained using the first 85% of the samples as the development set, and the final results were obtained only on the remaining 15% independent test set.

3.2. Parameter Settings

Table 1. Main parameter settings of different models

Models	Parameters	Value
Proposed Model	CEEMDAN Nstd	0.2
	Number of K-means Clusters	3
	Number of VMD Modes	6
	VMD Penalty Factor	1000
	Dynamic Expert Kernel Size	3, 5, 7
	Dropout	0.10
	Number of hidden units	32
LSTM	Number of hidden units	128
TCN	Number of Convolutional Layers	4
	Kernel Size	3
iTransformer	Number of Attention Heads	4
TCN-BiLSTM-Attention	Number of TCN Convolutional Layers	4
	Kernel Size	3
	Number of BiLSTM Hidden Units	32
	Number of Attention Heads	4
CNN-LSTM-Attention	Number of Convolutional Filters	64
	Number of Convolutional Filters	3, 5
	Number of hidden units	128
	Number of Attention Heads	4

Reasonable parameter configuration is an important basis for model training and performance evaluation. To reduce the interference of different experimental conditions on the

comparison of forecasting results and to ensure that all models were trained and tested under consistent conditions, the experimental environment and major training parameters

were uniformly configured in this study. The main parameter settings of each forecasting model are listed in Table 1.

All experiments were conducted on the same hardware and software platform. In terms of hardware, the computer was equipped with an Intel i5-10400F processor with a clock frequency of 2.90 GHz and 16 GB of memory. In terms of software, model construction, training, and testing were all implemented in MATLAB R2024b. During training, the Adam optimizer was uniformly adopted, with an initial learning rate of 0.001, a maximum of 100 training epochs, a batch size of 128, and a random seed of 1.

3.3. Parameter Settings

In this study, MAE, RMSE, and R2 are selected to evaluate the forecasting results. These three metrics assess the forecasting performance of the model from the perspectives of prediction accuracy, error sensitivity, and goodness of fit, respectively. Let the actual value be y_i , the predicted value be $\hat{y}_i$, the sample mean be $\bar{y}$, and the number of samples be N . The corresponding expressions are given as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (24)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (25)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (26)$$

Lower MAE and RMSE values indicate smaller deviations between the predicted and actual power values, whereas an R2 value closer to 1 indicates a higher degree of agreement between the predicted results and the actual variations in PV power.

Table 2. Statistical characteristics of the signal properties of each component

Component	SampleEntropy	CenterFrequency	EnergyRatio	Lag1ACF	FluctuationIndex
H_VMD1	0.15576	1.9061	0.0002968	-0.98344	1.9925
H_VMD2	0.26289	1.5374	0.00034446	-0.74245	1.8675
H_VMD3	0.19469	1.2448	0.00037609	-0.37332	1.6579
H_VMD4	0.10883	1.036	0.00044182	-0.056349	1.4534
H_VMD5	0.10244	0.85649	0.00044334	0.22272	1.2472
H_VMD6	0.27088	0.60069	0.0001664	0.58182	0.91485
Mid	0.49749	0.4925	0.0037866	0.69973	0.77465
Low	0.040243	0.061739	0.99414	0.99334	0.11547

3.5. Ablation Experiments

To verify the contributions of the CEEMDAN-VMD decomposition strategy and the differentiated forecasting mechanism to model performance, five groups of ablation experiments were designed in this study, as detailed in Table 3. MD1 and MD2 do not employ data decomposition and directly forecast the original PV power using the Regularity Predictor and the Dynamic Predictor, respectively. Based on CEEMDAN-VMD decomposition, MD3 and MD4 use a single RP and a single DP, respectively, to uniformly model all subcomponents. MD5 is the complete model proposed in this paper, in which the RP and DP are assigned to different subcomponents according to their signal properties.

As can be seen from Table 4 and Fig. 4, data decomposition leads to a significant improvement in forecasting performance. Compared with MD1 without decomposition, MD3 with CEEMDAN-VMD reduces the MAE in spring, summer, autumn, and winter by 67.45%, 49.97%, 58.65%, and 42.67%, respectively, and reduces the RMSE by 77.03%, 61.60%, 75.33%, and 64.32%, respectively. Similarly, compared with

3.4. Data Decomposition and Reconstruction

PV power series exhibit pronounced nonlinearity, non-stationarity, and multi-timescale fluctuations. Directly applying a unified model to the original series may lead to mutual interference among information at different frequencies. To reduce the forecasting difficulty, CEEMDAN is first employed to decompose the original PV power series into 14 IMF components. SE and CF are then selected as clustering features, and K-means is used to reconstruct the IMFs into high-, medium-, and low-frequency components. Among them, the high-frequency component exhibits faster variation scales and stronger fluctuations, whereas the medium- and low-frequency components mainly characterize relatively continuous dynamic variations and the dominant variation trend.

Since the reconstructed high-frequency signal still contains rapid variations at different scales, VMD is further applied to decompose it into multiple subcomponents, which are denoted as H1–H6 in descending order of center frequency. Consequently, a total of eight forecasting components, namely H1–H6, Mid, and Low, are obtained. Subsequently, the dynamic properties of each subcomponent are analyzed using Sample Entropy, Central Frequency, Energy Ratio, the First-order Autocorrelation Coefficient, and the Fluctuation Index. The results, as shown in Table 2, indicate that the frequency level of a component cannot fully reflect its forecasting regularity. Therefore, according to temporal regularity and dynamic complexity, H1, H2, and Low are further classified as regular components, whereas H3–H6 and Mid are classified as complex dynamic components, and are modeled using the RP and DP, respectively.

MD2, MD4 also shows clear decreases in MAE and RMSE across the four seasons, and the R² values are also improved. These results indicate that, regardless of whether the RP or the DP is used, decomposing the original non-stationary PV power series can effectively reduce the difficulty of subsequent forecasting.

On this basis, further comparison among MD3, MD4, and MD5 shows that differentiated modeling can further improve model performance. In spring, compared with MD3, MD5 reduces the MAE and RMSE by 13.28% and 15.17%, respectively. Compared with MD4, which uniformly uses the DP, MD5 also achieves lower MAE and RMSE in spring, summer, and autumn. In particular, the MAE in summer is reduced by 36.83%, indicating that applying a unified forecasting structure to subcomponents with different properties makes it difficult to fully exploit the adaptability of each model, whereas selecting the appropriate predictor according to component characteristics can further improve forecasting accuracy.

It is worth noting that, in winter, MD4 outperforms MD5,

indicating that the advantage of the differentiated forecasting strategy is influenced by seasonal variations in signal characteristics. However, from the overall results, MD5 achieves the best performance in spring, summer, and autumn, with average MAE and RMSE values of 0.4433 and 0.6642, respectively, across the four seasons. Therefore,

CEEMDAN-VMD decomposition can effectively reduce forecasting difficulty, and on this basis, the use of differentiated predictors according to component properties can further enhance the model's adaptability to different dynamic characteristics.

Table 3. Description of the ablation models

Abbreviation	Model	Description
MD1	Regularity Predictor	Direct prediction of the original PV power using the regularity predictor
MD2	Dynamic Predictor	Direct prediction of the original PV power using the dynamic predictor
MD3	CEEMDAN-VMD + Regularity Predictor	All decomposed components are uniformly predicted by the regularity predictor.
MD4	CEEMDAN-VMD + Dynamic Predictor	All decomposed components are uniformly predicted by the dynamic predictor.
MD5	Proposed	Different components are assigned to the regularity or dynamic predictor according to their signal characteristics.

Table 4. Performance of different models in the four seasons in the ablation experiments

Season	Metric	Model				
		MD1	MD2	MD3	MD4	MD5
Spring	MAE	1.0918	1.1050	0.3554	0.3624	0.3082
	RMSE	2.1955	2.0891	0.5042	0.4750	0.4277
	R2	0.9782	0.9803	0.9988	0.9989	0.9991
Summer	MAE	1.2993	1.3248	0.6500	0.8301	0.5244
	RMSE	2.7466	2.7785	1.0546	1.0329	0.8361
	R2	0.9644	0.9636	0.9947	0.9949	0.9967
Autumn	MAE	0.8957	0.9786	0.3704	0.4961	0.3416
	RMSE	2.2749	2.0627	0.5612	0.6152	0.4947
	R2	0.9756	0.9799	0.9985	0.9982	0.9988
Winter	MAE	1.0087	1.1213	0.5783	0.4245	0.5899
	RMSE	2.5094	2.3041	0.8954	0.6820	0.8984
	R2	0.9596	0.9660	0.9948	0.9970	0.9948

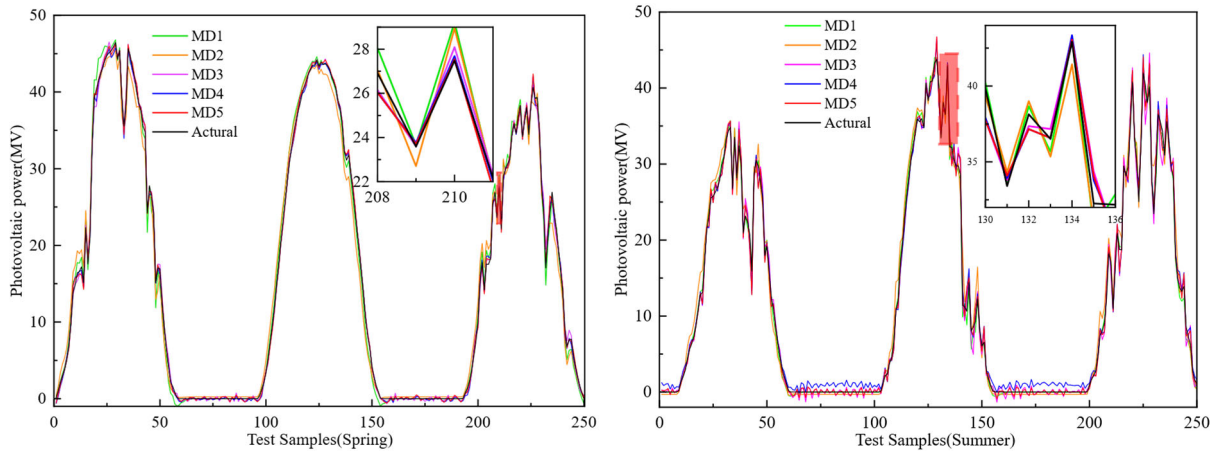


Fig. 4 Performance comparison of different models in the ablation experiments

3.6. Comparative Experiments

To evaluate the forecasting performance of the proposed model, LSTM [14], TCN [15], iTransformer [16], TCN-LSTM-Attention [17], and CNN-LSTM-Attention [18] are selected as benchmark models and denoted as M1–M5, respectively, while the proposed model is denoted as M6. These models cover representative architectures, including recurrent neural networks, temporal convolutional networks, Transformer-based models, and hybrid deep neural networks. The abbreviations of the comparative models are summarized in Table 5.

According to the experimental results shown in Table 6 and Fig. 5, the comparative models exhibit certain performance differences across different seasons. Among them, LSTM and

TCN show relatively stable overall forecasting performance, whereas iTransformer and the two hybrid models exhibit relatively large errors in some seasons, indicating that more complex network architectures do not necessarily lead to stable performance improvements. In contrast, the proposed model achieves the lowest MAE and RMSE and the highest R2 in all four seasons, demonstrating more stable adaptability across different seasonal conditions.

Compared with the best-performing benchmark model in each season, the proposed model reduces the MAE by 72.49%, 61.54%, 61.79%, and 44.10% in spring, summer, autumn, and winter, respectively, and reduces the RMSE by 79.91%, 68.61%, 77.33%, and 63.16%, respectively. Moreover, the R2 values in all four seasons are higher than 0.994. These results indicate that CEEMDAN-VMD can effectively separate the

multi-scale fluctuations contained in the original PV power series, while differentiated forecasting based on component properties can further improve the modeling adaptability to

different dynamic characteristics, thereby yielding more accurate and stable forecasting results.

Table 5. Abbreviations of the comparative models

Abbreviation	M1	M2	M3	M4	M5	M6
Model	LSTM	TCN	iTransformer	TCN-BiLSTM-Attention	CNN-LSTM-Attention	Proposed

Table 6. Forecasting performance of the comparative models

Season	Metric	Model					
		M1	M2	M3	M4	M5	M6
Spring	MAE	1.1203	1.1485	1.9774	2.4980	1.7285	0.3082
	RMSE	2.1445	2.1293	2.6325	3.1113	2.7038	0.4277
	R2	0.9789	0.9795	0.9686	0.9562	0.9668	0.9991
Summer	MAE	1.3635	1.3879	2.0861	3.8230	1.9145	0.5244
	RMSE	2.7208	2.6640	3.2644	4.5853	3.0586	0.8361
	R2	0.9649	0.9665	0.9498	0.9010	0.9561	0.9967
Autumn	MAE	0.9559	0.8939	1.5553	2.1113	2.2700	0.3416
	RMSE	2.1818	2.1824	2.5486	3.0547	2.8763	0.4947
	R2	0.9770	0.9775	0.9693	0.9560	0.9611	0.9988
Winter	MAE	1.0553	1.1533	1.7031	1.2660	2.3512	0.5899
	RMSE	2.4385	2.5148	3.1931	2.8405	3.1667	0.8984
	R2	0.9616	0.9595	0.9347	0.9483	0.9352	0.9948

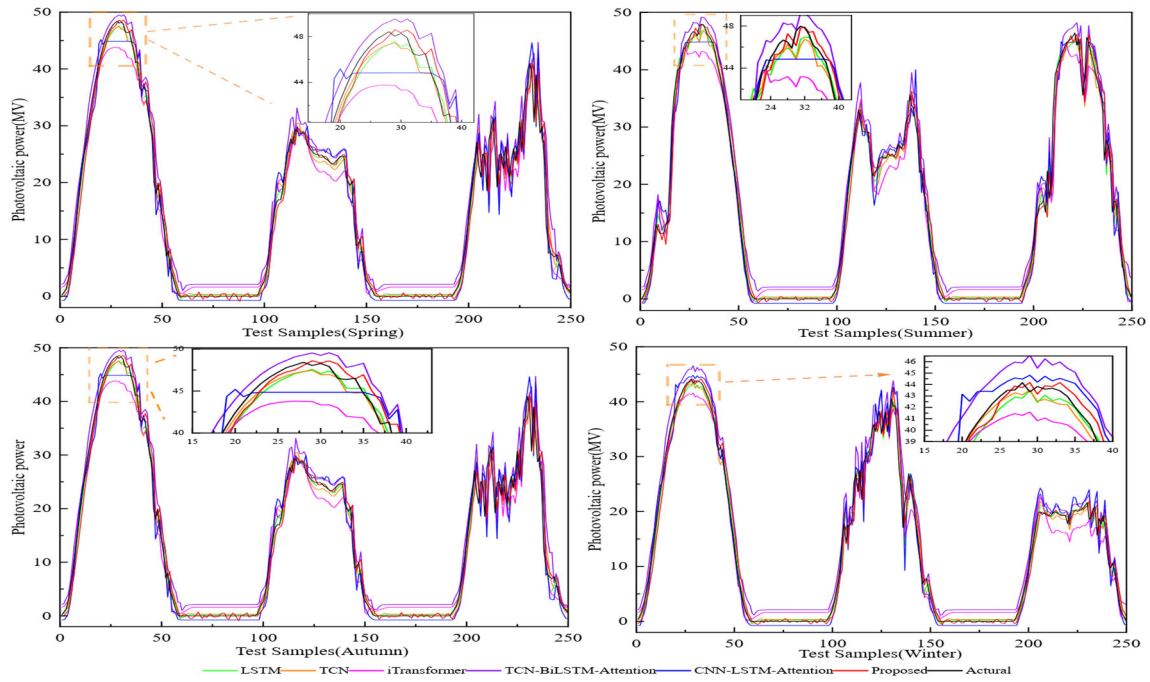


Fig. 5 Forecasting performance of different models for selected local samples

3.7. Statistical Significance and Error Distribution Analysis

To further examine the statistical reliability of the differences in forecasting performance among different models, the test sequence was divided into 24-h time blocks, each containing 96 consecutive sampling points, and the MAE and RMSE were calculated separately for each block. The boxplots of the daily forecasting errors are shown in Fig. 6. On this basis, the Wilcoxon signed-rank test was employed to compare the forecasting errors of the proposed model with those of the benchmark models, and the Holm method was used for multiple-comparison correction. As shown in Table 7, compared with M1–M5, the proposed model reduces the mean daily MAE by 61.58%, 62.16%, 79.15%, 86.37%, and

72.72%, respectively, and the mean daily RMSE by 69.07%, 67.60%, 75.88%, 83.35%, and 73.73%, respectively. The Holm-adjusted p -values for all comparisons are 6.1035×10^{-4} , which are substantially lower than 0.05, and the rank-biserial effect sizes are all equal to 1, indicating that the performance improvements are statistically significant and highly consistent.

Further analysis of the error distributions shows that the standard deviation of the errors and the 95th percentile of the absolute errors for the proposed model are 0.8356 and 1.7880, respectively, both of which are markedly lower than those of the other benchmark models. Although its median absolute error is not the lowest, the smaller error dispersion and lower tail error indicate that the proposed model can effectively suppress large forecasting deviations and provide more stable

forecasting performance.

Table 7. Errors over 24-h time windows and statistical significance test results for different models

Model	MAE (Mean $\pm$ Std)	MAE (Mean $\pm$ Std)	Error Standard Deviation	95th Percentile of Absolute Error
M1	1.3570 $\pm$ 0.5280	0.5213 $\pm$ 0.2153	2.7210	5.8365
M2	1.3776 $\pm$ 0.6551	2.3355 $\pm$ 1.2921	2.6362	5.9893
M3	2.4995 $\pm$ 0.5102	3.1372 $\pm$ 0.8225	3.1264	5.7832
M4	2.4995 $\pm$ 0.5102	4.5432 $\pm$ 0.6091	2.8812	8.3964
M5	2.4995 $\pm$ 0.5102	2.8803 $\pm$ 1.0288	3.0299	6.5098
M6	0.5213 $\pm$ 0.2153	0.7566 $\pm$ 0.3578	0.8356	1.7880

Std.: Standard Deviation.

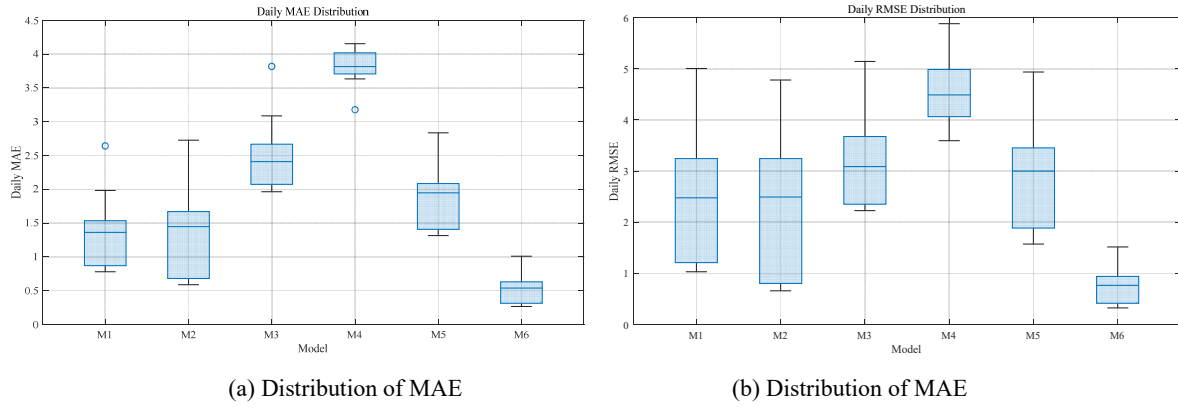


Fig. 6 Boxplots of daily forecasting errors for different models

3.8. Generalization Experiment

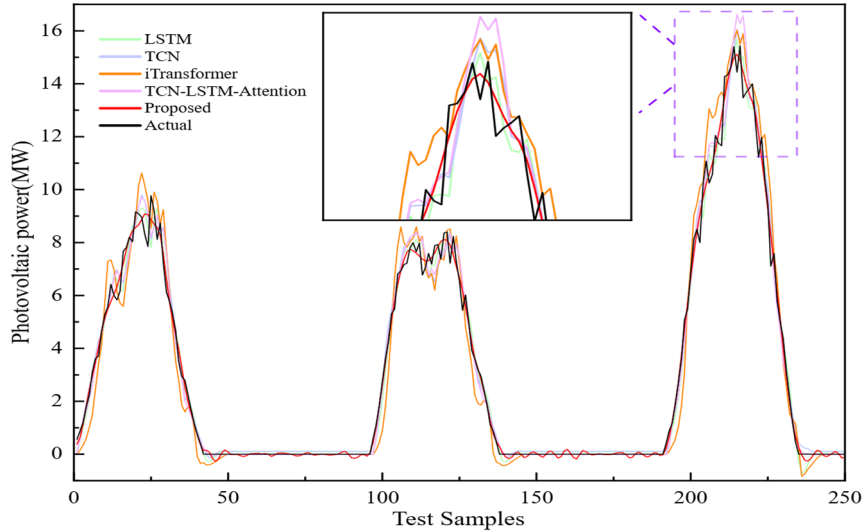


Fig. 7 Forecasting results of different models in the generalization experiment

Table 8. Forecasting performance of different models in the generalization experiment

Metric	LSTM	TCN	iTransformer	TCN-LSTM-Attention	Proposed
MAE	0.2918	0.3005	0.4316	0.2528	0.0793
RMSE	0.5126	0.5114	0.7630	0.5332	0.1308
R2	0.9826	0.9826	0.9614	0.9811	0.9988

To further evaluate the applicability and stability of the proposed model under different data conditions, an independent validation experiment was conducted using 4,800 samples from the Alice Springs PV dataset in Australia collected in January 2019. The experimental results, as shown in Table 8 and Fig. 7, indicate that the proposed model

achieves the best performance across all three evaluation metrics, with an MAE of 0.0793, an RMSE of 0.1308, and an R2 of 0.9988. Compared with LSTM, TCN, iTransformer, and TCN-LSTM-Attention, the proposed model reduces the MAE by 72.82%, 73.61%, 81.63%, and 68.63%, respectively, and reduces the RMSE by 74.48%, 74.42%, 82.86%, and 75.47%,

respectively. Among the benchmark models, iTransformer exhibits relatively large forecasting errors on this dataset, whereas LSTM, TCN, and TCN-LSTM-Attention show relatively good fitting performance but still achieve lower overall forecasting accuracy than the proposed model.

4. Conclusion

To address the pronounced non-stationarity and multi-scale fluctuations of ultra-short-term PV power series, as well as the substantial differences in the predictable structures of decomposed subcomponents, this paper proposes a forecasting method based on CEEMDAN–VMD selective secondary decomposition and signal-property-guided dual predictors. Unlike the conventional strategy of feeding all components into the same forecasting model, this study further considers the regularity and dynamic complexity of individual subcomponents after decomposition and accordingly assigns predictors with different capacities, so that the structural complexity of the forecasting models can better adapt to the predictable characteristics of different components.

(1) The selective secondary decomposition strategy can achieve a favorable balance between refining complex fluctuations and avoiding over-decomposition. CEEMDAN is first employed to perform multi-scale decomposition of the original PV power series, and the resulting IMFs are clustered and reconstructed according to Sample Entropy and Central Frequency, so that rapid fluctuations, medium-scale dynamics, and the dominant trend in the original series are preliminarily separated. On this basis, VMD-based secondary decomposition is applied only to the reconstructed high-frequency component that still contains complex and rapid variations, while the original dynamic structures of the medium- and low-frequency components are preserved.

(2) Predictor matching based on multidimensional signal properties can more reasonably accommodate the regularity and dynamic complexity of different components. The ablation results show that this differentiated strategy improves forecasting performance in spring, summer, and autumn, while the winter results also indicate that the fixed component–predictor mapping still has certain limitations in seasonal adaptability.

(3) Experiments across the four seasons, statistical tests, and cross-dataset validation demonstrate that the proposed method achieves favorable forecasting accuracy, stability, and generalization capability. Compared with the better-performing benchmark model in each season, the proposed method reduces the MAE by 44.10%–72.49% and the RMSE by 63.16%–79.91%, while all R2 values exceed 0.994. The statistical tests further confirm the significance of the performance improvements. On the Alice Springs dataset, the model achieves an MAE of 0.0793, an RMSE of 0.1308, and an R2 of 0.9988, indicating good adaptability to different data environments.

This study mainly focuses on single-site, single-step ultra-short-term PV power forecasting. The component–predictor mapping remains fixed once the model is determined, whereas the signal structure of actual PV power may vary with season, weather conditions, and operating environments. In particular, the winter ablation results indicate that the fixed mapping strategy still has room for further improvement. Future research can be conducted in two directions. On the one hand, spatial correlation information from neighboring PV stations can be introduced to extend component-property

characterization to multisource spatiotemporal information. On the other hand, dynamic predictor allocation and online updating mechanisms for streaming data can be investigated, enabling different components to adaptively select or adjust forecasting models according to real-time signal states, thereby further enhancing the adaptability and practical deployment value of the model under seasonal transitions, abrupt weather changes, and cross-regional application scenarios.

Acknowledgements

The research is supported by the National Natural Science Foundation of China (Grant No. 62263019) and the Natural Science Foundation of Gansu Province, China (Grant No. 26JRRA507).

References

- [1] Sabadus, A., Blaga, R., Hategan, S. M., & Pop, N. (2024). A cross-sectional survey of deterministic PV power forecasting: Progress and limitations in current approaches. *Renewable Energy*, 226, Article 120385. <https://doi.org/10.1016/j.renene.2024.120385>
- [2] Limouni, T., Yaagoubi, R., Bouziane, K., Guissi, K., & Baali, E. H. (2023). Accurate one step and multistep forecasting of very short-term PV power using LSTM-TCN model. *Renewable Energy*, 205, 1010-1024. <https://doi.org/10.1016/j.renene.2023.01.079>
- [3] Wang, M., Rao, C. J., Xiao, X. P., Zhang, Y., & Li, H. (2024). Efficient shrinkage temporal convolutional network model for photovoltaic power prediction. *Energy*, 297, Article 131295. <https://doi.org/10.1016/j.energy.2024.131295>
- [4] Dai, Y. M., Yu, W. J., & Leng, M. M. (2024). A hybrid ensemble optimized BiGRU method for short-term photovoltaic generation forecasting. *Energy*, 299, Article 131458. <https://doi.org/10.1016/j.energy.2024.131458>
- [5] Xiong, B. Y., Chen, Y. T., Chen, D. L., Wang, H., & Li, J. (2025). Deep probabilistic solar power forecasting with Transformer and Gaussian process approximation. *Applied Energy*, 382, Article 125294. <https://doi.org/10.1016/j.apenergy.2025.125294>
- [6] Liu, Y. F., Liu, Y., Cai, H. H., Zhang, L., & Wang, Q. (2023). An innovative short-term multihorizon photovoltaic power output forecasting method based on variational mode decomposition and a capsule convolutional neural network. *Applied Energy*, 343, Article 121139. <https://doi.org/10.1016/j.apenergy.2023.121139>
- [7] Wang, L. N., Mao, M. X., Xie, J. L., Lv, Z., & Zhang, H. (2023). Accurate solar PV power prediction interval method based on frequency-domain decomposition and LSTM model. *Energy*, 262, Article 125592. <https://doi.org/10.1016/j.energy.2022.125592>
- [8] Wu, T., Hu, R. F., Zhu, H. Y., Chen, Y., & Liu, X. (2024). Combined IXGBoost-KELM short-term photovoltaic power prediction model based on multidimensional similar day clustering and dual decomposition. *Energy*, 288, Article 129770. <https://doi.org/10.1016/j.energy.2023.129770>
- [9] Liu, L., Zhang, J. Q., & Xue, S. B. (2025). Photovoltaic power forecasting: Using wavelet threshold denoising combined with VMD. *Renewable Energy*, 249, Article 123152. <https://doi.org/10.1016/j.renene.2025.123152>
- [10] Zhou, Z. D., Dai, Y. M., & Leng, M. M. (2025). A photovoltaic power forecasting framework based on attention mechanism and parallel prediction architecture. *Applied Energy*, 391, Article 125869. <https://doi.org/10.1016/j.apenergy.2025.125869>

- [11] Xue, S. Y., & Li, L. (2026). Photovoltaic power forecasting based on secondary decomposition strategy and hybrid model. *Scientific Reports*, 16, Article 12915. <https://doi.org/10.1038/s41598-026-12915-8>
- [12] Yang, Z. H., Li, J. D., Wang, H. T., & Zhang, Y. (2025). An Informer model for very short-term power load forecasting. *Energies*, 18(6), Article 1523. <https://doi.org/10.3390/en18061523>
- [13] Wang, S., Yan, S., Li, H. J., & Zhang, X. (2024). Short-term prediction of photovoltaic power based on quadratic decomposition and residual correction. *Electric Power Systems Research*, 236, Article 110948. <https://doi.org/10.1016/j.epsr.2024.110948>
- [14] Wang, L. S., Liu, Y. H., Li, T. S., & Xie, X. (2020). Short-term PV power prediction based on optimized VMD and LSTM. *IEEE Access*, 8, 165849-165862. <https://doi.org/10.1109/ACCESS.2020.3022632>
- [15] Li, Y., Song, L., Zhang, S., Wang, Y., & Chen, H. (2023). A TCN-based hybrid forecasting framework for hours-ahead utility-scale PV forecasting. *IEEE Transactions on Smart Grid*, 14(5), 4058-4071. <https://doi.org/10.1109/TSG.2023.3261984>
- [16] Pei, J. X., Liu, N., Shi, J. Q., Wang, H., & Li, Y. (2025). Tackling the duck curve in renewable power system: A multi-task learning model with iTransformer for net-load forecasting. *Energy Conversion and Management*, 326, Article 119458. <https://doi.org/10.1016/j.enconman.2025.119458>
- [17] Li, M. K., Sun, L. T., & Sun, Y. T. (2026). Long-term and short-term photovoltaic power generation forecasting using a multi-scale fusion MHA-BiLSTM model. *Energies*, 19(2), Article 411. <https://doi.org/10.3390/en19020411>
- [18] Soltani, I., & Shahnia, F. (2026). Hybrid CNN-LSTM modeling with attention mechanism for solar power forecasting. *Discover Applied Sciences*, 8, Article 145. <https://doi.org/10.1007/s42452-026-07341-2>