(Received 13 April 2026, Accepted 9 May 2026, Scheduled 20 May 2026)

A Photovoltaic Power Forecasting Method Based on Improved TimeMixer

Chao Wang[1], Xinyuan Xie[2], Fengsheng Chen[2,*], Pengyi Fan[1], Zhengning Pan[1], Tao Yu[1], and Zhongan Yu[2]

[1] School of Electric Power Engineering, South China University of Technology, Guangzhou 510000, China
[2] School of Electrical Engineering and Automation, Jiangxi University of Science and Technology, Ganzhou 341000, China

ABSTRACT: Photovoltaic (PV) power sequences are highly susceptible to high-frequency stochastic noise under complex micro-meteorological conditions. Furthermore, existing forecasting models struggle with isolated multi-scale physical features and insufficient nonlinear mapping capabilities. To address these limitations, this paper proposes an improved TimeMixer-based PV power forecasting method. First, the macroscopic trend and microscopic seasonal components are extracted using a past-decomposable-mixing architecture. Second, an adaptive gated feature fusion mechanism is introduced as a physically motivated feature-level filter to attenuate high-frequency noise channels through dynamic attention masks, effectively blocking the cross-scale propagation of invalid meteorological interference. Finally, a cross-scale joint nonlinear network is constructed to capture nonlinear interactions among multi-band components through state matrix aggregation and activation operators. Case studies utilizing operational data from a 50 MW PV power plant, in Xinjiang, China, demonstrate that the proposed architecture effectively overcomes smoothing degradation and phase lag under complex scenarios, such as abrupt cloud cover. Compared with the original baseline, the proposed method reduces the forecasting mean squared error by 10.80%, significantly enhancing both global fitting accuracy and dynamic extreme-value tracking capability.

1. INTRODUCTION

The widespread deployment of renewable energy is a vital strategy for addressing energy crises and environmental challenges, with the integration of high-penetration renewable energy sources emerging as a defining characteristic of future power systems [1]. However, power generation systems, particularly PV systems, are directly influenced by meteorological variables, such as solar irradiance, temperature, and cloud shading, resulting in significant stochasticity, intermittency, and non-stationarity in their power output [2]. Accurate PV power forecasting is therefore a critical prerequisite for ensuring secure and stable grid operation, optimizing energy storage configurations, and facilitating day-ahead economic dispatch.

In recent years, extensive research has been conducted by scholars worldwide on PV power forecasting, resulting in the development of various predictive models. Early studies predominantly utilized traditional machine learning algorithms, such as Support Vector Machines (SVM) and Support Vector Regression (SVR), achieving satisfactory results in stationary series forecasting and mid-to-long-term trend fitting [3, 4]. With the advancement of deep learning, Long Short-Term Memory (LSTM) networks integrated with attention mechanisms have been widely adopted, effectively enhancing the nonlinear feature extraction capability for time-series data [5]. To further mitigate the limitations of long-sequence modeling, Transformer architectures and their variants, such as Informer and Autoformer, have been introduced into PV forecasting. Their internal sequence decomposition mechanisms help capture temporal periodicity and trend patterns, thereby improving forecasting accuracy [6–9]. Furthermore, the application of multi-time-scale feature fusion techniques, combined with methods such as empirical mode decomposition, has augmented the capacity of models to process complex multivariate time-series data [10, 11]. Nevertheless, there remains room for further optimization when handling complex PV sequences: traditional deep networks are prone to error accumulation in long-range forecasting, while existing multi-scale and Transformer-based models incur significant computational overhead during feature decomposition and self-attention operations, and often struggle to fully exploit the complementary advantages of microscopic and macroscopic information across different sampling scales. Consequently, exploring novel architectures capable of efficiently decoupling and fusing multi-scale temporal features is of great significance.

The recently proposed TimeMixer model offers a novel perspective for addressing the aforementioned multi-scale time-series forecasting issues [12]. Distinct from the mainstream pure decomposition paradigm, this model constructs a Past-Decomposable-Mixing (PDM) architecture to extract microscopic seasonal fluctuations and macroscopic trend information at fine and coarse scales, respectively, demonstrating favorable forecasting performance across multiple time-series benchmarks. Currently, this architecture and its derivatives have been progressively applied to domains, such as joint photovoltaic-load scenario generation [13] and power load forecasting [14]. Nevertheless, when confronted with com-

* Corresponding authors: Fengsheng Chen (894060860@qq.com).

plex micro-meteorological conditions, such as sudden cloud cover, rapid irradiance fluctuations, and abrupt weather transitions, models based on multi-scale mixing still encounter several challenges [15, 16]. First, transient high-frequency disturbances induced by cloudy weather often dominate fine-grained sequences. Deficient in adaptive noise reduction alongside feature screening mechanisms, these low-confidence disordered noises are transmitted to downstream networks, frequently rendering local distortions within the final forecasting curve. Concurrently, existing models typically employ independent linear predictors at the output stage to generate multi-scale results prior to equal-weight superposition. This ensemble approach partially neglects the deep physical cross-correlations among distinct frequency components, necessitating further exploration to accurately characterize the nonlinear dynamic coupling regularities between macroscopic trend evolutions alongside microscopic fluctuations under complex meteorological environments [17].

In response to the aforementioned limitations, this paper proposes a PV power forecasting method based on an improved TimeMixer model. First, building upon the foundational architecture of TimeMixer, the trend and seasonal components are extracted under multi-resolution observation views to comprehensively capture the macroscopic and microscopic regularities of historical time series. Second, an adaptive gated feature fusion mechanism is introduced during the multi-scale feature interaction stage. By generating data-driven dynamic attention masks, this mechanism effectively attenuates and suppresses the cross-scale propagation of high-frequency meteorological noise, thereby enhancing the signal-to-noise ratio (SNR) of the hidden-layer features at the fundamental physical level. Finally, constructing a joint cross-scale nonlinear forecasting network at the model's output end optimizes the original independent linear superposition module. Utilizing state matrix aggregation alongside nonlinear activation operators facilitates the deep cross-fusion of multi-band components. The proposed method aims to improve robustness against high-frequency noise and strengthen the interaction among multi-scale features under complex micro-meteorological conditions.

2. MULTI-SCALE CHARACTERISTICS OF PV POWER

2.1. Macro-Micro Features of PV Power

PV power plants exhibit strong stochasticity, volatility, and non-stationarity. From the perspective of time-series analysis and signal processing, real-world active PV power is not a single-scale stationary signal. Instead, it consists of multiple components with different dynamic response characteristics.

1) Macroscopic low-frequency trend component

The diurnal cycle of solar irradiance primarily dictates the macroscopic low-frequency trend of PV power. Barring extreme meteorological interference, the ideal PV output curve presents a smooth, continuous profile featuring pronounced diurnal periodicity. This low-frequency component reflects the overall envelope trajectory alongside the baseline output level of PV power over daily or extended temporal windows. Within the multi-scale space, coarse-grained observation data can effectively filter transient noise, optimally preserving this low-frequency trend information to exhibit exceptionally high autocorrelation coupled with predictability.

2) Microscopic high-frequency disturbance component

During actual grid-connected operation, subject to highly stochastic factors like transient cloud shading alongside sudden aerosol concentration variations, PV output frequently experiences drastic power ramp events. Manifesting on the power curve as high-frequency oscillations coupled with local distortions superimposed upon the low-frequency baseline, these abrupt disturbances constitute the microscopic high-frequency disturbance component within the time series. Appearing within the fine-grained space as high-amplitude broadband noise featuring significant intermittency alongside nonlinear characteristics, this component serves as the core catalyst for the sharp amplification of PV power forecasting errors under complex meteorological scenarios (e.g., cloudy, rainy, snowy conditions).

From a fundamental physical perspective, these multi-scale characteristics are the manifestation of distinct natural processes. The macroscopic trend is inherently dictated by the deterministic astronomical cycle (i.e., the diurnal solar path), while the microscopic disturbances reflect the stochastic atmospheric modulation—specifically, the transient scattering and absorption of solar radiation by clouds and aerosols. Isolating these components via multi-scale decomposition is therefore a prerequisite for modeling the complex superposition of these deterministic and stochastic physical drivers.

2.2. Necessity of Multi-Scale Decomposition

Given these characteristics, traditional single-scale PV power forecasting models exhibit inherent limitations. Modeling exclusively at the original high-sampling rate exacerbates overfitting to high-frequency random disturbances, inducing severe, irrational forecasting oscillations. Conversely, excessive data smoothing introduces pronounced phase lags, precluding the sensitive tracking of transient power jumps.

Consequently, the multi-scale decomposition coupled with the representation of PV power sequences assumes critical engineering necessity. Constructing multi-resolution observation views facilitates the effective disentanglement of smooth low-frequency trends from random high-frequency disturbances. Fig. 1 illustrates the measured power curve of an operational PV power plant under typical cloudy conditions alongside its multi-scale time-domain representations.

As evidenced by Fig. 1, PV power sequences across varying scales encapsulate markedly distinct dynamic information under complex meteorological conditions. This observation furnishes a robust empirical foundation coupled with theoretical underpinning for subsequently introducing time-series forecasting networks equipped with multi-scale fusion capabilities. Adopting differentiated feature fusion strategies tailored to the sequence characteristics of distinct frequency bands represents the pivotal approach to enhancing the tracking precision of local power fluctuations, concurrently preserving the global smoothness of the forecasting curve.

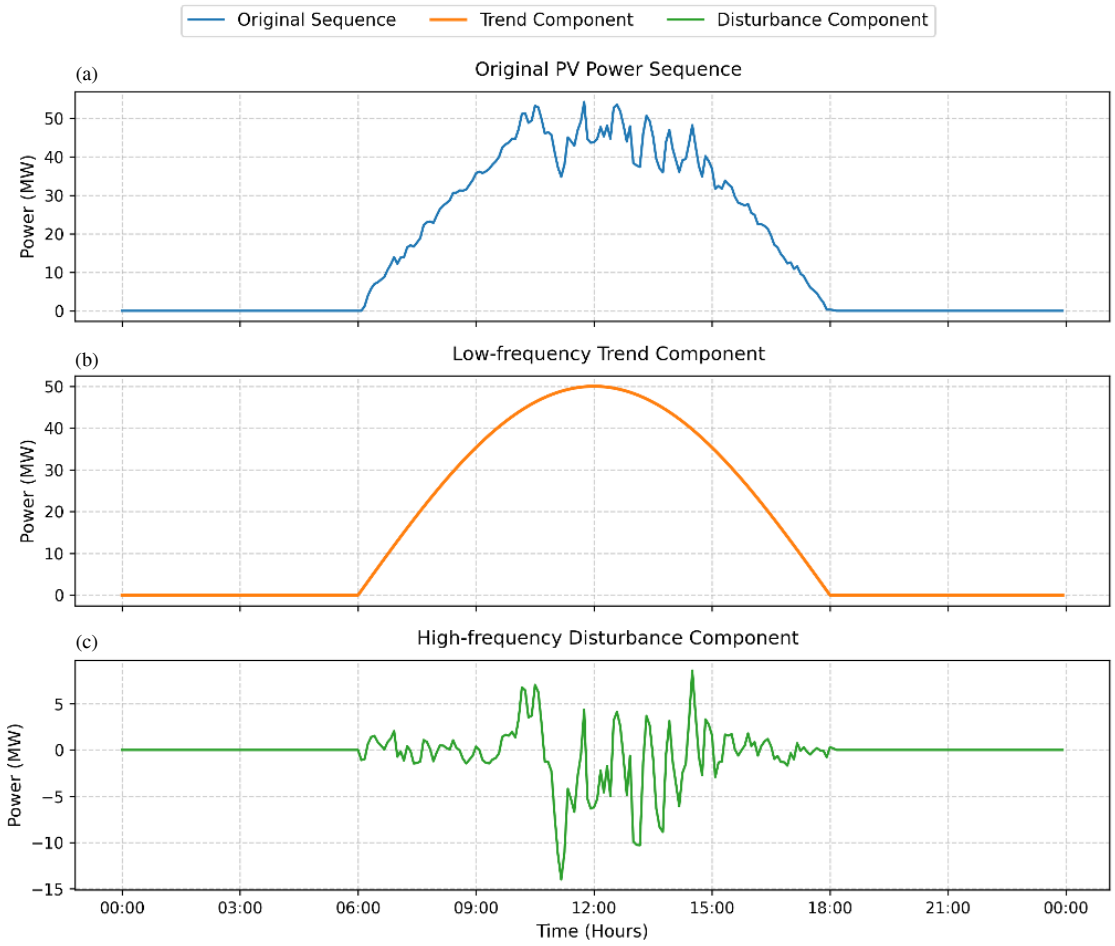

FIGURE 1. Multi-scale time-domain characteristics of PV power under typical complex weather conditions.

3. IMPROVED TIMEMIXER PV POWER FORECASTING MODEL

3.1. TimeMixer Model

The core architecture of TimeMixer primarily comprises three modules: multi-scale sequence construction, PDM, and Future-Multipredictor-Mixing (FMM).

1) Multi-scale sequence construction

Given the input sequence of historical PV output coupled with associated meteorological features $x \in \mathbb{R}^{P \times C}$ (where P denotes the time steps, and C represents the feature dimension), TimeMixer initially generates M sub-sequences featuring varying resolutions along the temporal dimension via average pooling operations. Designating the base sequence as x_0, the down-sampled sequence x_m at the m-th scale is formulated as:

$$x_m = \text{AvgPool}(x_{m-1}) \quad m \in \{0, 1, \ldots, M\} \tag{1}$$

where AvgPool$(\cdot)$ denotes the average pooling operation.

Subsequently, an embedding layer maps these sequences into a deep feature set X^0. This multi-scale representation empowers the model to capture short-term high-frequency PV output fluctuations at higher resolutions, concurrently extracting long-term low-frequency trends at lower resolutions.

2) Past-decomposable-mixing

To facilitate comprehensive multi-scale historical information interaction, the model stacks l layers of PDM modules. Within the l-th layer, a sequence decomposition module decomposes the multi-scale features outputted by the preceding layer into a seasonal component s_m^l and a trend component t_m^l:

$$s_m^l, t_m^l = \text{Decomp}(x_m^{l-1}) \quad m \in \{0, \ldots, M\} \tag{2}$$

where Decomp$(\cdot)$ denotes the decomposition block network.

Given the markedly distinct temporal characteristics of the seasonal and trend components, the PDM module adopts differentiated interaction strategies for these two components.

Seasonal mixing: Employing a bottom-up residual mixing approach, this strategy supplements coarse-scale seasonal modeling utilizing microscopic high-frequency information extracted from fine scales.

$$\text{For } m: 1 \to M, \quad s_m^l = s_m^l + \text{Mix}_\text{s}\left(s_{m-1}^l\right) \tag{3}$$

where $\text{Mix}_\text{s}(\cdot)$ denotes the bottom-up information fusion network for periodic components.

Trend mixing: Adopting a top-down mixing configuration, this paradigm leverages macroscopic prior knowledge derived from coarse scales to guide fine-scale trend modeling, thereby filtering local noise interference.

$$\text{For } m: (M-1) \to 0, \quad t_m^l = t_m^l + \text{Mix}_\text{T}(t_{m+1}^l) \tag{4}$$

where $\text{Mix}_{\text{T}}(\cdot)$ denotes the trend mixing network facilitating top-down information fusion.

3) Future-multipredictor-mixing

Capitalizing on the complementary advantages of varying-scale sequences in forecasting long-term versus short-term future variations, TimeMixer incorporates the FMM module during the final stage. Based on the multi-scale historical features x_m^L extracted through L layers of PDM, the FMM allocates an independent predictor to each scale, generating the ultimate result via an ensemble strategy:

$$\tilde{x}_m = \text{Predictor}_m(x_m^L) \quad m \in \{0, \ldots, M\} \tag{5}$$

$$\hat{x} = \sum_{m=0}^{M} \tilde{x}_m \tag{6}$$

where $\tilde{x}_m$ denotes the forecasted value of the m-th predictor, Predictorm$(\cdot)$ the m-th forecasting module, and $\hat{x}$ the PV power forecasted value.

3.2. Limitations of the Original TimeMixer

Despite achieving effective multi-scale historical feature extraction via the PDM module, the TimeMixer model exhibits pronounced structural limitations within its output-stage FMM module when processing complex PV power sequences.

The foundational logic of the FMM module entails allocating completely independent linear predictors to each temporal scale, subsequently generating the ultimate forecasting sequence via simple equal-weight linear superposition. Within PV power forecasting tasks characterized by profound stochasticity coupled with non-stationarity, this ensemble mechanism exposes two critical defects.

Initially, the equal-weight linear superposition strategy presumes a constant physical contribution from multi-scale features to the forecasting results, lacking dynamic evaluation alongside adaptive screening capabilities regarding feature confidence. Under complex meteorological conditions, such as cloudy or showery weather, fine-grained PV output sequences are frequently pervaded by extensive high-frequency disordered noise induced by micro-meteorological disturbances. Absent effective filtering mechanisms, these low-confidence noise features are directly transmitted into the ultimate forecasting results, rendering abnormal local oscillations within the output power forecasting curve, concurrently introducing conflicting gradient signals during the model's backpropagation optimization process. Accumulating such invalid gradients severely disrupts the smoothness of the error surface within the high-dimensional parameter space, subsequently restricting further enhancements in model convergence speed, coupled with forecasting accuracy.

Furthermore, using completely independent linear predictors may weaken the modeling of cross-scale physical correlations of PV time-series signals across varying scales. The PV output evolution under complex meteorological conditions essentially constitutes the manifestation of deep nonlinear coupling between macroscopic weather trends and microscopic local disturbances. For instance, the overarching decline in macroscopic irradiance typically accompanies drastic evolutions in microscopic fluctuation amplitudes. Executing superficial aggregation following the independent completion of linear mappings for each scale, the original model essentially obstructs the cross-scale feature interaction pathways. This purely linear independent inference, coupled with a shallow integration paradigm, struggles to accurately characterize the complex high-dimensional nonlinear dynamic mapping regularities of PV output under extreme meteorological conditions.

Addressing these structural limitations necessitates a profound reconstruction of the feature fusion alongside forecasting generation stages at the model's output end. This paper overcomes the forecasting bottlenecks inherent within the original model via the introduction of an adaptive gated feature fusion mechanism coupled with a joint cross-scale nonlinear forecasting mechanism.

3.3. Adaptive Gated Feature Fusion Mechanism

Following extraction via multiple PDM layers, the model acquires $M+1$ deep hidden-layer features featuring varying temporal lengths. Preserving the high-dimensional representational capacity of hidden-layer features to facilitate precise feature screening, this paper initially introduces a temporal projection layer. Utilizing linear transformations, this layer uniformly aligns the temporal lengths of multi-scale features to the forecasting horizon, concurrently maintaining the feature channel dimension invariant, denoted as $\bar{x}_m$. Subsequently, tailored to each aligned feature scale, the model computes its dynamic attention mask under the current data state via an independent linear transformation coupled with a Sigmoid activation function:

$$G_m = \sigma\left(W_m^{\text{G}} \bar{x}_m + b_m^{\text{G}}\right) \tag{7}$$

where G_m denotes the dynamic attention mask at the m-th scale; W_m^{G} and b_m^{G} represent the weight matrix coupled with the bias term of the m-th scale gating network, respectively; $\sigma(\cdot)$ designates the Sigmoid activation function, serving to strictly constrain the feature confidence within the probability interval of $(0, 1)$.

Ultimately, executing a Hadamard product operation yields the feature-filtered forecasting sequence:

$$\tilde{x}_m^{\text{G}} = G_m \odot \bar{x}_m \tag{8}$$

where $\tilde{x}_m^{\text{G}}$ designates the m-th sequence following feature filtering.

Employing this methodology, the gating mechanism can be interpreted as a physically motivated feature-level adaptive filter motivated by the physical behavior of PV power under micro-meteorological disturbances. It is designed to evaluate the confidence of features based on the current state of micro-meteorological turbulence. By dynamically attenuating channels dominated by high-frequency 'pseudo-fluctuations' (e.g., chaotic irradiance spikes caused by passing clouds), the mechanism ensures that only physically robust signals are propagated to the subsequent fusion stage, effectively enhancing the signal-to-noise ratio at the feature level.

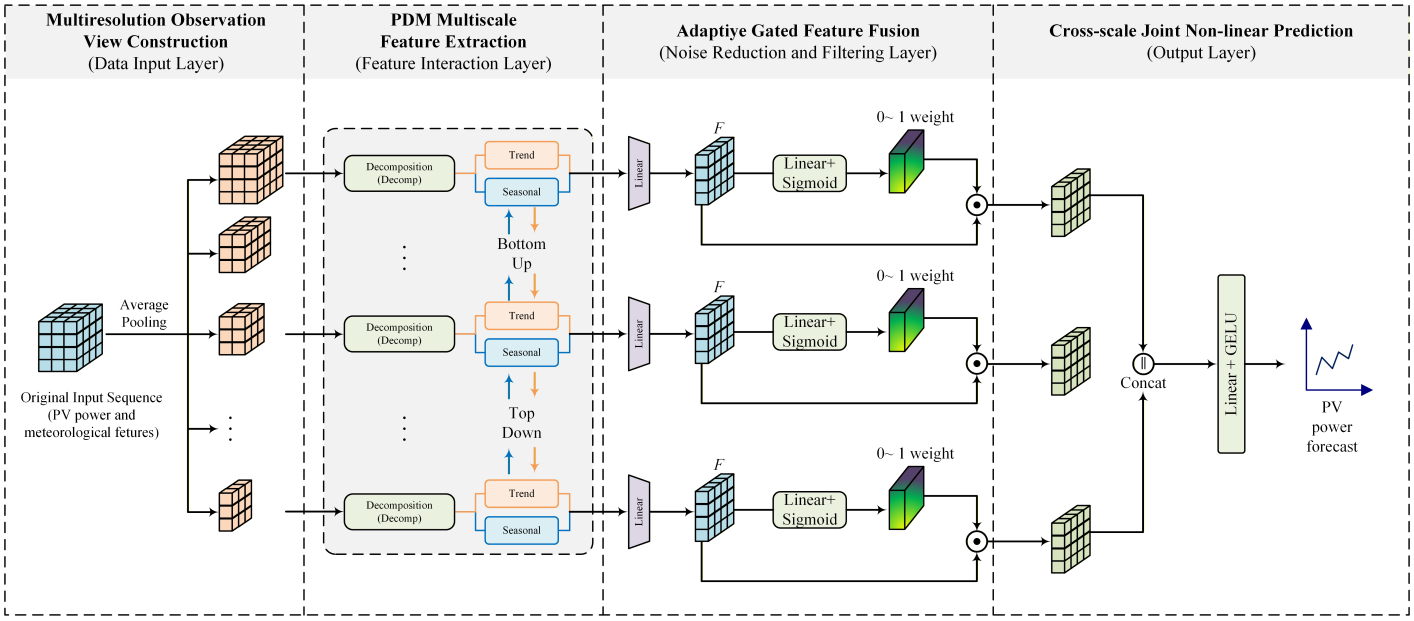

FIGURE 2. Architecture of the improved TimeMixer model.

3.4. Cross-Scale Joint Nonlinear Forecasting Mechanism

Having acquired high-quality multi-scale feature sets filtered via the gating mechanism, adhering to the original independent linear forecasting paradigm precludes the accurate characterization of deep physical coupling relationships among distinct frequency bands. Consequently, this paper designs a joint cross-scale nonlinear forecasting mechanism, facilitating the deep cross-fusion of multi-scale temporal states.

Aggregating individual scale components along the feature dimension constructs the cross-scale state matrix:

$$\tilde{x} = \tilde{x}_0^{\mathrm{G}} \parallel \tilde{x}_1^{\mathrm{G}} \parallel \cdots \parallel \tilde{x}_m^{\mathrm{G}} \tag{9}$$

where $\tilde{x}$ denotes the cross-scale state matrix, with $\parallel$ representing the concatenation operation along the feature channel dimension.

Subsequently, feeding the joint state matrix into a deep nonlinear mapping network facilitates excavating complex correlation patterns among multi-scale components. Leveraging a hidden-layer projection operator incorporating the GELU activation function, this network induces cross-coupling between features across distinct frequency bands, yielding the fused hidden state:

$$z = \mathrm{GELU}(W_1 \tilde{x} + b_1) \tag{10}$$

where z designates the fused hidden state matrix, with W_1 and b_1 representing the hidden-layer projection matrix coupled with the bias term, respectively.

Ultimately, an output mapping matrix projects the fused hidden state into the target physical space, yielding the final univariate PV active power forecasting sequence:

$$\hat{x} = W_2 z + b_2 \tag{11}$$

where W_2 and b_2 denote the weight matrix coupled with the bias term of the output layer, respectively.

This architectural enhancement is physically motivated by the fact that PV energy conversion is a highly nonlinear process. The impact of macroscopic trends (e.g., ambient temperature and overall irradiance levels) on PV output is deeply coupled with microscopic fluctuations (e.g., rapid irradiance drops). For instance, the photoelectric conversion efficiency is nonlinearly influenced by the cross-interaction of these variables across different time scales. By utilizing state matrix aggregation and nonlinear activation, the model transcends simple linear superposition, allowing it to more accurately characterize these deep physical cross-correlations.

3.5. Overall Architecture of the Improved TimeMixer Model

Figure 2 delineates the overarching architecture characterizing the improved TimeMixer PV power forecasting model proposed herein, fundamentally anchored upon gated feature fusion coupled with joint nonlinearity.

During the forward inference data flow, the execution logic of each module is delineated as follows:

1) Multi-scale sequence construction: The data input layer receives the original input sequence comprising historical PV active power alongside meteorological covariates. Executing average pooling down-sampling operations along the temporal dimension generates a sequence set covering multiple time scales. This stage furnishes multi-granular foundational data views, facilitating the subsequent segregation of macroscopic low-frequency trends from microscopic high-frequency PV output disturbances.

2) PDM multi-scale feature extraction: Within this layer, individual scale sequences undergo decoupling into trend alongside seasonal components via decomposition modules. Building upon this, the model executes top-down trend information transmission concurrently with bottom-up seasonal feature fusion, thereby deeply excavating and extracting the inherent evo-

lutionary regularities of PV temporal signals across varying scales.

3) Adaptive gated feature fusion: Individual scale features initially undergo alignment along the temporal dimension via linear projections. Subsequently, leveraging a linear layer coupled with a Sigmoid activation function facilitates the joint computation of dynamic attention masks under the current data state, yielding feature confidence weights constrained within $(0, 1)$. Ultimately, executing Hadamard product operations yields the element-wise multiplication of these weights with the original feature tensors. This mechanism adaptively attenuates alongside suppressing high-frequency meteorological noise intermingled within fine-grained features, providing pure, high-SNR features for downstream networks.

4) Joint cross-scale nonlinear forecasting: Gated noise-reduced multi-scale feature components ultimately converge at the output layer. Diverging from the independent scale-wise forecasting ensemble approach of the original model, this layer initially executes a concatenation operation to uniformly aggregate multi-band features along the channel dimension. Subsequently, feeding the aggregated state tensors into a deep projection network incorporating GELU activation operators facilitates deep cross-mapping among multi-scale features, ultimately outputting the univariate PV active power forecasting sequence.

4. CASE STUDY ANALYSIS

4.1. Experimental Data and Preprocessing

To validate the effectiveness alongside the generalization capability of the proposed model, this study utilizes the open-source dataset from the 2021 State Grid Corporation of China New Energy Power Forecasting Competition. Sourced from the Supervisory Control and Data Acquisition system of an operational PV power plant, this dataset encompasses continuous PV operational and meteorological records spanning 2019 to 2020. Specifically, the monitored facility is located in Xinjiang, China, featuring an installed capacity of 50 MW. The temporal sampling interval is maintained at 15 minutes, yielding 96 discrete data points daily.

Regarding input feature construction, meteorological variables exhibiting high correlation with PV output are selected as auxiliary features according to Pearson correlation analysis, comprising: global solar irradiance, direct normal irradiance, global horizontal irradiance, ambient temperature, atmospheric pressure, alongside relative humidity. These meteorological features, coupled with historical PV output sequences, constitute the high-dimensional input feature matrix.

During the data preprocessing stage, linear interpolation serves to fill a minor portion of missing time steps. Conversely, long-term data loss or out-of-bounds outliers resulting from equipment failures undergo removal. To mitigate the impact of varying dimensions on training, a standardization method maps all input variables into a standard normal distribution space. Ultimately, the processed dataset undergoes partitioning into training, validation, and testing sets according to a 7:1:2 ratio following chronological order.

4.2. Evaluation Metrics and Experimental Settings

To quantify forecasting performance comprehensively and objectively, this paper selects Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and the Coefficient of Determination (R^2) as evaluation metrics. The formulation for each metric is delineated as follows:

$$E_{\text{MSE}} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{12}$$

$$E_{\text{RMSE}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{13}$$

$$E_{\text{MAE}} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{14}$$

$$E_{\text{MAPE}} = \frac{100\%}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{15}$$

$$E_R = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{16}$$

where N denotes the total number of samples within the testing set, y_i the ground-truth value, $\hat{y}_i$ the forecasted value, featuring $\bar{y}$ as the mean of the ground-truth value, E_{MSE} the MSE value, E_{RMSE} the RMSE value, E_{MAE} the MAE value, and E_R the coefficient of determination.

To verify the superiority of the improved architecture across multiple dimensions, this paper conducts comprehensive comparative experiments involving five algorithms spanning traditional machine learning, classical deep learning, cutting-edge Transformer architectures, and the baseline model:

SVM: Representing traditional shallow machine learning, this model serves to verify the necessity of deep networks for high-dimensional meteorological feature extraction.

LSTM: Representing classical recurrent neural networks, this model evaluates the foundational performance in handling long-term temporal dependencies.

Informer: Incorporating the ProbSparse self-attention mechanism, this model serves as a highly representative Transformer architecture within the long-term forecasting domain.

Autoformer: Innovatively introducing a sequence decomposition mechanism into the self-attention framework, this model possesses significant benchmarking value against the temporal decomposition logic utilized herein.

TimeMixer: Serving as the foundational baseline prior to the proposed improvements, this model directly demonstrates the precision enhancement benefits derived from the introduced gated noise reduction alongside joint nonlinear mechanisms.

The experimental framework establishes a historical look-back window of 96 combined with a forecasting horizon of 96, facilitating the verification of model robustness under long-term prediction scenarios. Optimization is executed via the Adam optimizer utilizing an initial learning rate of 0.001.

Model training is capped at 1000 epochs, preserving the optimal weights evaluated on the validation set for final testing. To ensure strict reproducibility, all experiments are conducted within a unified computing environment equipped with an Intel Core i5-12400 processor alongside an NVIDIA GeForce RTX 3060 GPU, maintaining a training batch size of 256. Furthermore, the key hyperparameters delineated in Table 1 are determined through rigorous empirical tuning, utilizing the optimal validation performance as the ultimate selection criterion.

TABLE 1. Key hyperparameters of the models.

Model	Key Hyperparameters
Improved TimeMixer	$L = 3$, $M = 3$, $d_{model} = 64$
TimeMixer	$L = 3$, $M = 3$, $d_{model} = 64$
SVM	RBF kernel, $c = 4$, $\varepsilon = 0.01$
LSTM	4 layers, $d_{model} = 64$
Informer	2 encoder layers, 1 decoder layer, 4 attention heads, $d_{model} = 64$, factor = 5
Autoformer	2 encoder layers, 1 decoder layer, $\text{moving}_{\text{avg}} = 25$, 4 attention heads, $d_{model} = 64$

4.3. Comparison and Analysis of Forecasting Results

To comprehensively validate the forecasting performance of the proposed improved TimeMixer model, this paper conducts a comparative analysis across six distinct models utilizing the testing set. This section unfolds the discussion through two dimensions: overall evaluation metrics and dynamic tracking capabilities under typical meteorological conditions.

4.3.1. Overall Forecasting Performance Evaluation

The error evaluation results of each model on the test set are presented in Table 2.

TABLE 2. Comparison of overall forecasting errors of different models on the test set.

Model	MSE	MAE	MAPE	R^2
Improved TimeMixer	**25.0848**	**2.6110**	**37.66**	**0.8767**
TimeMixer	28.1227	2.9389	41.76	0.8549
SVM	40.4785	2.8595	47.63	0.7757
LSTM	34.7548	3.2342	54.14	0.8074
Informer	31.1594	2.7529	46.69	0.8273
Autoformer	30.2690	2.9411	47.94	0.8323

As shown in Table 2, the traditional machine learning model SVM manifests the highest MSE alongside an R^2 of merely 0.7757. This observation underscores the model's limitation in fitting solely the mean trend of PV output, incurring substantial outlier errors when processing local drastic fluctuations. Similarly, the LSTM model exhibits restricted overall forecasting precision, primarily attributed to error accumulation issues within long-sequence forecasting scenarios.

In contrast, Transformer-based architectures demonstrate superior performance. Specifically, Autoformer outperforms Informer regarding MSE, validating the effectiveness of sequence decomposition mechanisms for PV temporal feature extraction. The original TimeMixer model, benefiting from multi-scale feature interactions, further reduces the MSE to 28.1227. The improved TimeMixer model proposed herein achieves the optimal values across all evaluation metrics. Compared to the original model, the proposed architecture yields a 10.80% reduction in MSE alongside an 11.16% decrease in MAE, featuring an R^2 enhancement to 0.8767.

This performance leap demonstrates that incorporating gated feature fusion coupled with cross-scale nonlinear mapping mechanisms at the output stage effectively facilitates the suppression of cross-layer interference from high-frequency meteorological noise. Consequently, this enhancement strengthens the deep coupling capabilities of multi-scale features, thereby significantly elevating forecasting precision.

4.3.2. Analysis of Forecasting Results Under Typical Meteorological Conditions

To further evaluate the local feature fitting capability of the model across varying boundary conditions, typical alongside complex scenarios within the testing set undergo selection for validation. The resulting forecasting curves are delineated in Fig. 3 and Fig. 4, respectively.

Under the steady-state conditions illustrated in Fig. 3, PV output manifests a relatively regular bell-shaped curve, with each model demonstrating proficient foundational fitting capabilities. Leveraged by its inherent smoothing characteristics, the SVM model achieves a superior R^2 of 0.9725. Nevertheless, during the power ramp-up alongside ramp-down phases of dawn and dusk, most baseline models exhibit a certain degree of temporal lag. The improved model proposed herein (possessing an R^2 of 0.9738) maintains high consistency with the ground-truth power curve throughout the entire duration. Effectively mitigating phase shifts, this consistency validates the model's robustness under steady-state operating conditions.

Under the complex operating conditions illustrated in Fig. 4, PV output exhibits high-frequency, non-stationary, alongside drastic fluctuations, primarily driven by micro-meteorological abruptions, such as cloud shading. Within this scenario, the forecasting performance of the SVM model deteriorates significantly, with its results degenerating into smooth envelopes incapable of characterizing instantaneous power surges. Meanwhile, Informer and TimeMixer both exhibit pronounced oversmoothing and under-forecasting phenomena during power valley intervals.

The improved model proposed herein maintains a superior R^2 of 0.9099 under these complex conditions, achieving the highest forecasting precision. Benefiting from the filtration of redundant meteorological noise via the adaptive gated mechanism, alongside the recombinational mapping of complex states through the joint nonlinear layer, the improved model effectively suppresses pseudo-fluctuation interference. Furthermore, it achieves effective fitting of extreme power troughs during sharp PV output drops, validating its dynamic response

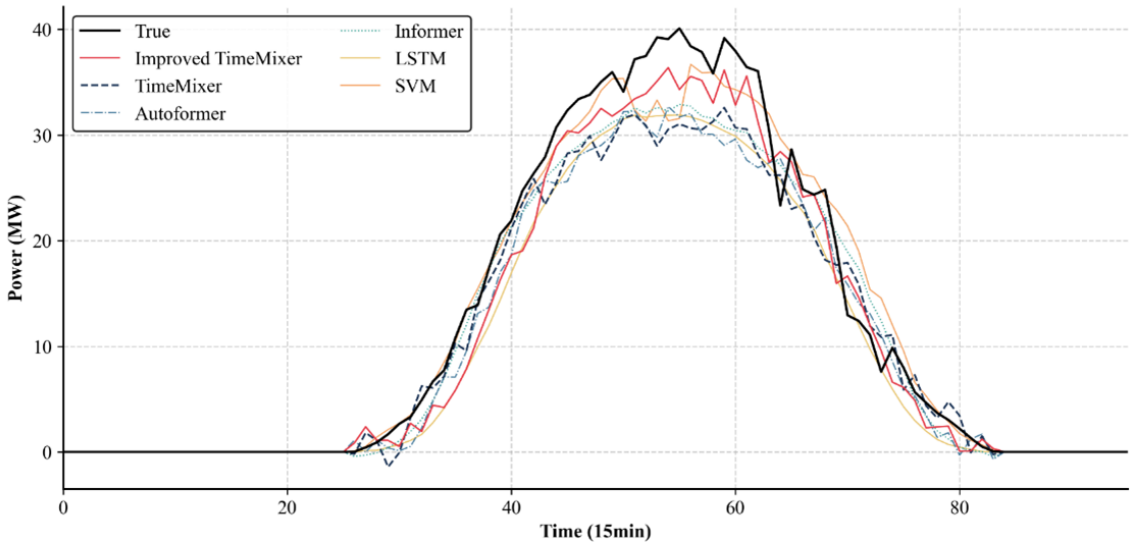

FIGURE 3. PV power forecasting curves under typical sunny day conditions.

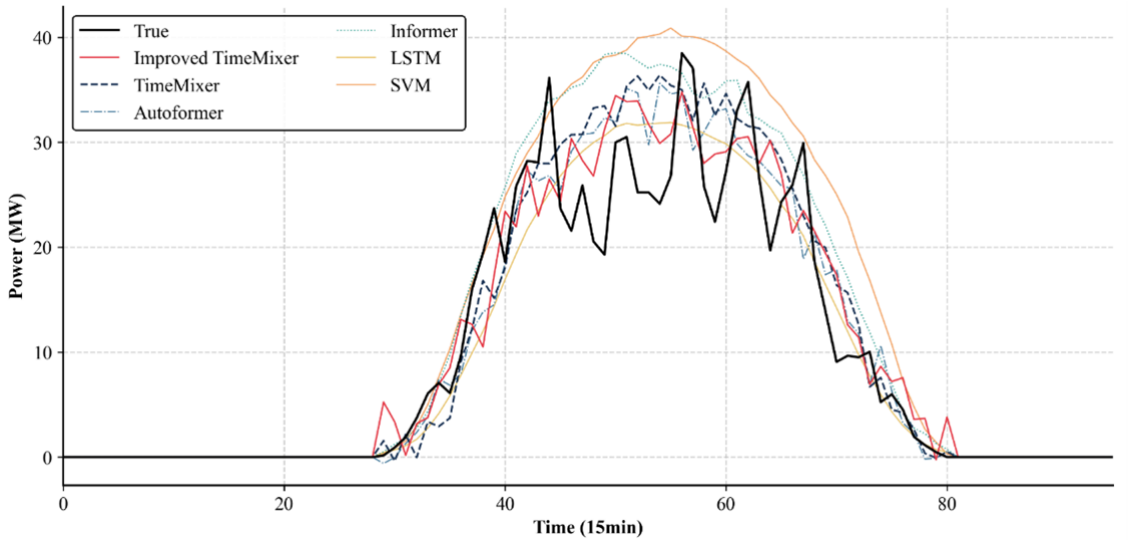

FIGURE 4. PV power forecasting curves under abrupt cloudy transition conditions.

TABLE 3. Ablation study evaluation results.

Model	Module Configuration	MSE	MAE	MAPE	R^2
A	Baseline TimeMixer	28.1227	2.9389	41.76	0.8549
B	Gated Mechanism	26.8532	2.8015	39.82	0.8632
C	Nonlinear Forecasting	26.2105	2.7344	38.95	0.8688
D	**Gated Mechanism+Nonlinear Forecasting**	**25.0848**	**2.6110**	**37.66**	**0.8767**

capability coupled with robustness when confronting extreme meteorological abruptions.

4.3.3. Ablation Study

To validate the effectiveness of the proposed adaptive gated feature fusion mechanism and cross-scale joint nonlinear forecasting mechanism, four groups of control-variable ablation experiments undergo design under unified experimental parameter configurations coupled with dataset partitioning rules. The definitions of each model variant are delineated as follows:

Model A: The original TimeMixer model (lacking integrated gating mechanisms alongside joint nonlinear layers, utilizing independent linear outputs).

Model B: Building upon Model A, incorporating exclusively the adaptive gated feature fusion mechanism at the output stage.

Model C: Building upon Model A, incorporating exclusively the joint cross-scale nonlinear forecasting mechanism at the output stage.

Model D: Simultaneously integrating both the gating mechanism and the joint nonlinear forecasting architecture.

The error evaluation results of each model variant on the test set are presented in Table 3.

According to the results in Table 3, incorporating different enhancement modules consistently reduces the forecasting errors to varying degrees:

1) Comparing Model A and Model B, the exclusive introduction of the gating mechanism reduces the MSE from 28.1227 to

26.8532, representing a 4.51% decrease. This observation underscores that adaptively computed attention masks effectively suppress the cross-layer transmission of high-frequency meteorological noise, subsequently enhancing the purity of features during the spatio-temporal fusion stage.

2) Comparing Model A and Model C, the standalone incorporation of the joint nonlinear layer yields a 6.80% reduction in MSE. This result indicates that aggregating multi-scale features and introducing the GELU activation function can reduce the isolation among different frequency components and enhance the nonlinear mapping capability of the network.

3) Synthesizing the results of the four experimental groups, Model D achieves the minimum values across all error metrics. This outcome demonstrates a synergistic effect between the gated noise reduction mechanism and the joint nonlinear network regarding their underlying mechanisms: the former facilitates filtering redundant interference beforehand to provide high-SNR state tensors for the latter, which subsequently executes deep feature recombination. The cascading combination of these two components effectively overcomes the forecasting fitting challenges under complex meteorological conditions.

5. CONCLUSIONS

Addressing the susceptibility of PV power forecasting to high-frequency noise interference and the physical isolation of multi-scale features under complex weather conditions, this paper proposes an improved TimeMixer model. This architecture is fundamentally anchored in adaptive gated feature fusion and joint cross-scale nonlinearity. Through comparative experiments and ablation studies utilizing real-world PV plant data, the primary conclusions are derived as follows:

1) Incorporating an adaptive gating mechanism during feature extraction effectively suppresses dynamic meteorological noise. By computing dynamic attention masks, the gating network functions as a physical filter that adaptively attenuates feature channels dominated by severe micro-meteorological noise. This mechanism obstructs the cross-layer propagation of invalid information, fundamentally enhancing the signal-to-noise ratio of multi-scale hidden features.

2) The cross-scale joint nonlinear network reconstructed at the output stage significantly strengthens the feature coupling capability. Diverging from the traditional independent linear superposition paradigm, the combination of state matrix aggregation and nonlinear activation shatters the isolation of individual scale features. This enhancement precisely maps the deep nonlinear correlations between macroscopic meteorological evolution and microscopic local fluctuations.

3) The proposed architecture demonstrates superior performance in both global fitting precision and local extreme-value tracking. Case studies indicate that the improved model achieves a 10.80% reduction in MSE compared to the baseline TimeMixer. Specifically under complex scenarios featuring abrupt cloudy variations, the model overcomes the smoothing degradation and phase lag defects inherent in traditional algorithms, maintaining robust tracking precision during drastically fluctuating intervals.

In future research, the proposed forecasting framework could be further extended to other highly stochastic renewable energy forecasting tasks, such as wind power generation, and applied to joint multi-energy load forecasting within complex microgrid systems to further verify its generalization capability.

ACKNOWLEDGEMENT

This work was supported in part by the National Natural Science Foundation of China (Project No. U24B6010), the Guangdong Basic and Applied Basic Research Foundation (No. 2025A1515010118), alongside the Guangdong Basic and Applied Basic Research Foundation (No. 2026A1515011200).

REFERENCES

[1] Kang, C. and L. Yao, "Key scientific issues and theoretical research framework for power systems with high proportion of renewable energy," *Automation of Electric Power Systems*, Vol. 41, No. 9, 1–11, 2017.

[2] Antonanzas, J., N. Osorio, R. Escobar, R. Urraca, F. J. Martinez-de Pison, and F. Antonanzas-Torres, "Review of photovoltaic power forecasting," *Solar Energy*, Vol. 136, 78–111, 2016.

[3] Zhu, Y. and J. Tian, "Application of least square support vector machine in photovoltaic power forecasting," *Power System Technology*, Vol. 35, No. 7, 54–59, 2011.

[4] Polo, A., "A two-step learning-by-examples method for photovoltaic power forecasting," *Progress In Electromagnetics Research C*, Vol. 125, 35–49, 2022.

[5] Zhou, H., Y. Zhang, L. Yang, Q. Liu, K. Yan, and Y. Du, "Short-term photovoltaic power forecasting based on long short term memory neural network and attention mechanism," *IEEE Access*, Vol. 7, 78 063–78 074, 2019.

[6] Zhou, H., S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, "Informer: Beyond efficient transformer for long sequence time-series forecasting," in *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35, No. 12, 11 106–11 115, 2021.

[7] Wu, H., J. Xu, J. Wang, and M. Long, "Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting," *Advances in Neural Information Processing Systems*, Vol. 34, 22 419–22 430, 2021.

[8] Cai, Z., Z. Tan, Y. Zhao, Y. Zhang, X. Liu, and H. Zhang, "Multiformer-TSA-based photovoltaic power forecasting method," *Journal of Electric Power Science and Technology*, Vol. 41, No. 1, 130–139, 2026.

[9] Bak, S., S. Choi, D. Yang, D. Kim, H. Rho, and K. Lee, "Transfer learning for photovoltaic power forecasting across regions using large-scale datasets," *IEEE Access*, Vol. 13, 136 175–136 190, 2025.

[10] Li, Z., Y. Wang, R. Zhang, *et al.*, "A multi-timescale photovoltaic power prediction method based on SE-CNN-BiLSTM and improved Transformer," *Zhejiang Electric Power*, Vol. 45, No. 3, 120–130, 2026.

[11] Lin, H., L. Gao, M. Cui, H. Liu, C. Li, and M. Yu, "Short-term distributed photovoltaic power prediction based on temporal self-attention mechanism and advanced signal decomposition techniques with feature fusion," *Energy*, Vol. 315, 134395, 2025.

[12] Wang, S., H. Wu, X. Shi, T. Hu, H. Luo, L. Ma, J. Y. Zhang, and J. Zhou, "Timemixer: Decomposable multiscale mixing for time series forecasting," *arXiv preprint arXiv:2405.14616*, 2024.

[13] Li, Z., X. Wang, C. Fu, *et al.*, "Planning-state PV-load joint scenario generation considering decomposable multi-scale temporal feature fusion," *Power System Protection and Control*, Vol. 53, No. 12, 152–164, 2025.

[14] Qiu, Z., C. Fu, B. Liang, and C. Cao, "Power load forecasting based on the improved TimeMixer," in *International Conference on Image, Signal Processing, and Pattern Recognition (ISPP 2025)*, Vol. 13664, 1094–1100, Nanjing, China, 2025.

[15] Gao, S., C. Li, Z. Li, *et al.*, "Fault detection of offshore floating photovoltaic power system considering wave motion and time-sequence characteristics," *Transactions of China Electrotechnical Society*, Vol. 41, No. 3, 999–1011, 2026.

[16] Piantadosi, G., S. Dutto, A. Galli, S. D. Vito, C. Sansone, and G. D. Francia, "Photovoltaic power forecasting: A Transformer based framework," *Energy and AI*, Vol. 18, 100444, 2024.

[17] Huang, X., X. Ding, Y. Han, Q. Sima, X. Li, and Y. Bao, "Day-ahead photovoltaic power forecasting based on SN-transformer-bimixer," *Energies*, Vol. 18, No. 16, 4406, 2025.