Advances in Computer, Signals and Systems (2026)
Clausius Scientific Press, Canada

DOI: 10.23977/acss.2026.100220
ISSN 2371-8838 Vol. 10 Num. 2

Day-Ahead Photovoltaic Power Forecasting Based on Pseudo-Label Contrastive Learning and Weather-Conditioned Gated Fusion of Physical Bases

Bihui Lv*, Fang Liu, Xuri Fan, Wei Chen

China Huadian Overseas Investment Co., Ltd., Beijing, China
** Corresponding author: bhl_chdoi@163.com*

Keywords: Day-Ahead Photovoltaic Power Forecasting; Weather Pseudo-Labels; Supervised Contrastive Learning; Physical Bases; Gated Fusion

Abstract: To address biases in numerical weather prediction (NWP) for day-ahead photovoltaic power forecasting and the lack of physical constraints in purely data-driven end-to-end regression, which can produce nonzero nighttime output or power exceeding physical limits, this paper proposes a forecasting method based on weather pseudo-label contrastive learning and gated fusion of multiple physical bases. First, a clear-sky power benchmark is established for each station using solar geometry and a clear-sky irradiance model. A clear-sky index is constructed from historical power measurements to generate weather pseudo-labels, and supervised contrastive pretraining is used to enhance the weather representation learned by the historical encoder. Three physical bases are then constructed from NWP irradiance, the clear-sky index at the same time on the previous day, and the most recent valid clear-sky index. A gating network performs a time-step-wise convex combination of these bases according to weather conditions, with Softmax-normalized weights ensuring nonnegative predictions, compliance with capacity limits, and zero nighttime output by construction. Experiments on ten stations in the PVOD dataset show that the proposed method generally outperforms multiple baselines in MAE, RMSE, and R^2. Ablation experiments verify the contribution of each physical basis, while cross-station adaptation experiments further demonstrate that parameter-efficient fine-tuning achieves accuracy close to full fine-tuning at a substantially lower training cost.

1. Introduction

As photovoltaic installed capacity continues to grow, the uncertainty of photovoltaic generation imposes increasingly stringent requirements on the secure operation of power systems and the integration of renewable energy. Photovoltaic power exhibits pronounced intermittency, volatility, and nonstationarity due to solar irradiance, cloud movement, ambient temperature, and the diurnal cycle. Accurate short-term photovoltaic power forecasting provides an important basis for grid dispatch, reserve-capacity allocation, renewable-energy integration, and electricity-market trading, and therefore remains a key research problem in renewable-energy forecasting.

Existing photovoltaic power forecasting methods can be broadly divided into physical and data-driven approaches. Physical methods establish mappings using NWP and photovoltaic conversion mechanisms. They are highly interpretable, but their performance is susceptible to NWP biases. Data-driven methods, represented by deep networks such as LSTM and Transformer models, offer strong fitting capabilities but require large amounts of data. Moreover, end-to-end direct regression lacks explicit physical constraints and can readily produce nonzero nighttime output or power beyond physical limits, while a single fixed architecture struggles to adapt to varying weather conditions.

To address these issues, this paper proposes a day-ahead photovoltaic power forecasting method that fuses multiple physical bases through weather-conditioned gating. A clear-sky power benchmark is first established and weather pseudo-labels are constructed, after which supervised contrastive pretraining enhances the historical encoder. Three physical bases are then constructed from NWP, same-time previous-day persistence, and most-recent-state persistence. A gating network fuses these bases at each time step according to weather conditions, while structural constraints such as convex combination ensure the physical consistency of the predictions.

The main contributions of this paper are as follows:

(1) Multiple physical forecasting bases with a unified physical convention are constructed. Three deterministic power candidates are generated using NWP irradiance, the clear-sky index at the same time on the previous day, and the most recent valid clear-sky index. All three share the same clear-sky geometry and station-scale calibration, avoiding scale inconsistencies caused by independently calibrating different physical pathways.

(2) A supervised contrastive pretraining strategy based on weather pseudo-labels is proposed. Weather pseudo-labels are constructed from statistical features of the historical clear-sky index, including its mean, variability, and ramping behavior. A supervised contrastive loss brings historical-window representations from the same weather state closer together and separates those from different weather states, without using future measured power.

(3) A weather-conditioned, time-step-wise gated fusion mechanism is designed. The gating network integrates the historical-window representation, NWP clear-sky index, historical persistence state, solar position, and temporal information. It outputs convex-combination weights for the three bases through Softmax, with structural constraints guaranteeing nonnegativity, compliance with capacity limits, and zero nighttime output.

(4) Experimental validation is conducted in terms of both forecasting accuracy and deployment efficiency. Experiments on ten PVOD stations evaluate the method through comparisons with multiple baselines, physical-basis ablations, and analyses of forecasting curves and gating behavior. The cross-station deployment costs of full fine-tuning and parameter-efficient adaptation methods, including Adapter and LoRA, are also compared.

The remainder of this paper is organized as follows. Section 2 reviews related work on clear-sky physical modeling, contrastive representation learning, time-series encoding, and parameter-efficient adaptation. Section 3 details the proposed method. Section 4 presents the experimental design and data analysis. Finally, Section 5 concludes the paper.

2. Related Work

2.1. Photovoltaic Power Forecasting and Clear-Sky Physical Modeling

Photovoltaic power forecasting methods can be categorized into physical approaches, statistical and machine-learning approaches, and hybrid approaches [1]. Recurrent networks such as LSTM can model temporal dependencies [2], while recent time-series models such as DLinear [3] and TimesNet

[4] further improve data-driven forecasting accuracy. Related computations rely on open-source toolchains such as pvlib [5].

2.2. Contrastive Representation Learning and Time-Series Encoding

Photovoltaic curves under different weather conditions differ markedly in their mean, variability, and ramping characteristics. Supervised contrastive learning uses labels to organize positive and negative samples for learning discriminative representations [6], while the Transformer self-attention mechanism provides a foundational architecture for time-series encoding [7].

2.3. Parameter-Efficient Cross-Station Adaptation

Cross-station deployment generally requires retraining for the target station. Adapters introduce a small number of trainable parameters through bottleneck projections [8], whereas LoRA represents weight increments using low-rank matrices [9]. Both offer feasible approaches to reducing the cost of cross-station adaptation.

3. Methodology

This paper addresses day-ahead photovoltaic power forecasting, in which historical information from the preceding 24 h is used to forecast photovoltaic output over the next 24 h. At forecast issuance, historical measured power and future NWP are available, but future measured meteorological variables and actual power are not. NWP provides future weather information but contains forecast errors, whereas historical power reflects the recent operating state of the station and weather persistence. Therefore, rather than directly mapping input features to future power using a neural network, the proposed method first constructs multiple candidate forecasting bases with explicit physical meanings and then adaptively fuses them at each time step using a weather-conditioned gating network.

The overall method comprises four main stages. First, a clear-sky power benchmark is established using solar geometry and a clear-sky irradiance model, and a clear-sky index is constructed from historical actual power. Second, weather pseudo-labels are generated from statistical features of the historical clear-sky index, and supervised contrastive learning enhances the historical-window representation. Third, an NWP physical basis, a same-time previous-day persistence basis, and a most-recent-state persistence basis are constructed. Finally, fusion weights for the three bases are generated from the historical state and future weather conditions, and the final prediction is obtained through convex combination. All internal model computations use capacity-normalized power. After forecasting, the outputs are converted back to MW according to the installed capacity of each station for experimental evaluation.

3.1. Problem Definition

Let the installed capacity of a station be C, the actual photovoltaic power at time t be P_t, and its capacity-normalized form be defined in Equation (1):

$$p_t = \frac{P_t}{C} \quad (1)$$

The temporal resolution of the data is 15 min, and the historical window L_h and forecasting window L_f lengths satisfy $L_h = L_f = 96$, so that the model uses the previous 96 time steps, corresponding to 24 h of historical information, to forecast photovoltaic power over the next 96 time steps, corresponding to the next 24 h.

For forecast origin t, the historical input is represented as $\mathbf{X}_t^h = \left[\mathbf{x}_{t-L_h+1}^h, \ldots, \mathbf{x}_t^h\right] \in \mathbb{R}^{L_h \times d_h}$. The input features at each historical time step include capacity-normalized measured power, seven NWP variables, and cyclical time encodings. The seven NWP variables are global irradiance, direct irradiance, temperature, relative humidity, wind speed, wind direction, and pressure.

The covariates available over the future forecasting horizon are denoted by $\mathbf{X}_t^f = \left[\mathbf{x}_{t+1}^f, \ldots, \mathbf{x}_{t+L_f}^f\right] \in \mathbb{R}^{L_f \times d_f}$. These include NWP for the target period, cyclical time encodings, and deterministic solar-geometry features, such as solar zenith angle and solar elevation angle, derived from the station location and time.

Neither future measured power nor LMD is used as a model input. Because future measured information is unavailable at forecast issuance, no future-information leakage occurs.

3.2. Clear-Sky Power Benchmark and Auxiliary Weather Representation

3.2.1. Clear-Sky Power Benchmark

Solar position is calculated using pvlib from the latitude and longitude of each station, temporal information, and array tilt. The Ineichen clear-sky model is then used to obtain global horizontal irradiance, direct normal irradiance, and diffuse horizontal irradiance under clear-sky conditions. Subsequently, the Liu–Jordan isotropic-sky model is used for plane-of-array irradiance transposition, yielding the clear-sky plane-of-array global irradiance $G_{\mathrm{POA},t}^{\mathrm{cs}}$.

Rather than developing a complete module-level photovoltaic conversion model incorporating module temperature, inverter efficiency, and other factors, this paper converts plane-of-array irradiance into a simplified capacity-normalized power proxy, as defined in Equation (2):

$$p_{\mathrm{cs},t}^{(0)} = \mathrm{clip}\left(\frac{G_{\mathrm{POA},t}^{\mathrm{cs}}}{1000}, 0, 2\right) \quad (2)$$

The station calibration coefficient is defined in Equation (3) as the 90th percentile of the ratio of actual power to the raw clear-sky power proxy over valid samples in the training segment:

$$\alpha = Q_{0.9}\left(\left\{\frac{p_t}{p_{\mathrm{cs},t}^{(0)}} : t \in \mathcal{S}\right\}\right) \quad (3)$$

The final clear-sky power benchmark is then given by Equation (4):

$$p_{\mathrm{clear},t} = \max\left(\alpha p_{\mathrm{cs},t}^{(0)}, 0\right) \quad (4)$$

3.2.2. Clear-Sky Index

When the sun is close to the horizon and the clear-sky power benchmark is very small, even a small absolute power error can produce large ratio fluctuations. To avoid numerical instability under low-irradiance conditions around sunrise and sunset, the clear-sky index is clipped as shown in Equation (5):

$$k_{\mathrm{cs},t} = \begin{cases} \mathrm{clip}\left(\frac{p_t}{p_{\mathrm{clear},t}}, 0, 1.5\right), & p_{\mathrm{clear},t} > 0.02, \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

3.2.3. Construction of Weather Pseudo-Labels

To extract weather-state information from historical windows, auxiliary weather statistics are constructed from the clear-sky index. Daytime clear-sky samples within a historical window are treated as valid samples. If the number of valid samples in the window is $n \geq 3$, the following four statistics are extracted: the mean, describing the overall output level; the standard deviation, describing weather variability; the maximum adjacent difference, representing the largest short-term change within the window; and the ramp proportion, describing the frequency of large power changes. If fewer than three valid clear-sky indices are available in a window, all four statistical features are set to zero.

The standardizer for the four-dimensional statistical features is fitted using only training-segment data. After standardization, K-means clustering is applied with the number of clusters set to three. Because the cluster identifiers have no inherent meteorological semantics, the original numerical cluster labels are always used during model training. Semantic mapping is performed only when interpreting the results: the cluster with the lowest mean clear-sky index is labeled "overcast"; among the remaining two clusters, the one with the higher standard deviation of the clear-sky index is labeled "partly cloudy"; and the final cluster is labeled "clear."

Therefore, the clear, partly cloudy, and overcast categories used here are weather pseudo-labels derived solely from historical power statistics, rather than weather classes assigned from actual meteorological observations.

3.2.4. Historical Encoding and Supervised Contrastive Pretraining

In addition to the low-dimensional statistics described above, a Transformer encoder is used to learn representations of the complete historical window. Let the historical encoder be defined in Equation (6):

$$\mathbf{H} = E_\theta(\mathbf{X}^h) \quad (6)$$

where $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_{L_h}]$ denotes the hidden states at all historical time steps.

For sample i in a training batch, let its weather pseudo-label be y_i^w. The other samples with the same pseudo-label form the positive set defined in Equation (7):

$$P(i) = \{p : p \neq i,\ y_p^w = y_i^w\} \quad (7)$$

The supervised contrastive loss in Equation (8) is adopted:

$$\mathcal{L}_{\text{con}} = -\frac{1}{|\mathcal{V}|}\sum_{i\in\mathcal{V}} \frac{1}{|P(i)|}\sum_{p\in P(i)} \log \frac{\exp(\mathbf{z}_i^\top \mathbf{z}_p/\tau)}{\sum_{a\neq i}\exp(\mathbf{z}_i^\top \mathbf{z}_a/\tau)} \quad (8)$$

where τ is the temperature coefficient and $\mathcal{V}$ denotes the set of valid anchors with at least one positive sample in the current batch. When $\mathcal{V}$ is empty or the batch contains fewer than two samples, the contrastive loss for that batch is set to zero.

3.3. Construction of Multiple Physical Bases

Historical power alone cannot reflect future weather changes, whereas NWP alone is susceptible to meteorological forecast errors. To incorporate both future weather information and the persistence of historical states, three deterministic forecasting bases with the same capacity-normalized power convention are constructed to represent future NWP information, same-time previous-day persistence, and most-recent-weather-state persistence, respectively.

3.3.1. NWP Physical Basis

Let the NWP global irradiance at the target time be G_t^{nwp} and the direct irradiance be D_t^{nwp}. In this implementation, the NWP direct-irradiance field in the dataset is treated as direct normal irradiance (DNI). Under this convention, diffuse horizontal irradiance is inferred from global and direct irradiance using Equation (9):

$$G_{\mathrm{dhi},t}^{\mathrm{nwp}} = \max\left[G_t^{\mathrm{nwp}} - D_t^{\mathrm{nwp}}\cos\theta_{z,t}, 0\right] \quad (9)$$

where $\theta_{z,t}$ is the solar zenith angle. The same plane-of-array transposition method used for the clear-sky benchmark is then applied to obtain $G_{\mathrm{POA},t}^{\mathrm{nwp}}$.

After normalization and scaling by the same station coefficient α used for clear-sky power, the NWP physical basis is obtained as Equation (10):

$$b_t^{\mathrm{nwp}} = \mathrm{clip}\left[M_t\alpha\frac{G_{\mathrm{POA},t}^{\mathrm{nwp}}}{1000}, 0,1\right] \quad (10)$$

where $M_t = \mathbb{I}(\beta_t > 0^\circ)$ is the solar-elevation-angle mask.

3.3.2. Same-Time Previous-Day Persistence Basis

Photovoltaic power exhibits a pronounced diurnal cycle. When weather conditions change only slightly, the clear-sky index at the same time on the day before forecast issuance can serve as a historical reference for the current weather state.

The same-time previous-day persistence basis is defined in Equation (11):

$$b_{t+j}^{\mathrm{day}} = \mathrm{clip}\left[M_{t+j}p_{\mathrm{clear},t+j}k_{t+j}^{\mathrm{day}}, 0,1\right] \quad (11)$$

3.3.3. Most-Recent-Valid-State Persistence Basis

The same-time previous-day state captures the diurnal cycle but cannot fully describe the latest weather state near the forecast origin. A persistence basis based on the most recent valid clear-sky index is therefore constructed.

Within the historical window, let r denote the time closest to the forecast origin that also satisfies $p_{\mathrm{clear},r} > 0.02$, and define $k^{\mathrm{recent}} = k_{\mathrm{cs},r}$.

This clear-sky index is used as the most-recent-state reference over the entire future forecasting horizon, as defined in Equation (12):

$$b_{t+j}^{\mathrm{recent}} = \mathrm{clip}\left[M_{t+j}p_{\mathrm{clear},t+j}k^{\mathrm{recent}}, 0,1\right] \quad (12)$$

If no valid time satisfying the condition exists in the historical window, the persistence state is set to zero.

3.4. Weather-Conditioned Gated Fusion and Physical Consistency

The three physical bases are applicable under different conditions. The NWP basis contains future weather information and should generally serve as the primary basis for day-ahead forecasting, although its accuracy is affected by NWP biases. The same-time previous-day basis provides a stable diurnal reference, whereas the most-recent-valid-state basis places greater emphasis on weather persistence near the forecast origin. Because fixed-proportion fusion cannot adapt to varying weather conditions, a time-step-wise weather-conditioned gating network is designed to learn the proportions of the three bases adaptively.

3.4.1. Gating Inputs

After processing by the historical Transformer encoder, the historical-window-level representation $\bar{\mathbf{h}}$ is obtained. For future time step j, the historical-window representation is combined with the future information available at that time step to form the gating feature in Equation (13):

$$\mathbf{g}_{t+j} = \mathrm{Concat}\left(\bar{\mathbf{h}}, \tilde{\mathbf{x}}_{t+j}^{f}, k_{t+j}^{\mathrm{nwp}}, k_{t+j}^{\mathrm{day}}, k^{\mathrm{recent}}\right) \quad (13)$$

where $\tilde{\mathbf{x}}_{t+j}^{f}$ contains only the normalized solar elevation angle and the sine and cosine encodings of the hour of day.

3.4.2. Time-Step-Wise Convex Combination

The gating feature is fed into a multilayer perceptron to obtain three unnormalized gating scores, as defined in Equation (14):

$$\mathbf{a}_{t+j} = \mathrm{MLP}_g\left(\mathbf{g}_{t+j}\right) \in \mathbb{R}^3 \quad (14)$$

Softmax is then applied to obtain the weights corresponding to the three physical bases in Equation (15):

$$\mathbf{w}_{t+j} = \mathrm{Softmax}\left(\mathbf{a}_{t+j}\right) \quad (15)$$

The final prediction is given by Equation (16):

$$\hat{p}_{t+j} = w_{t+j}^{\mathrm{nwp}} b_{t+j}^{\mathrm{nwp}} + w_{t+j}^{\mathrm{day}} b_{t+j}^{\mathrm{day}} + w_{t+j}^{\mathrm{recent}} b_{t+j}^{\mathrm{recent}} \quad (16)$$

Because the gating weights are nonnegative and sum to one, the convex combination necessarily satisfies the physical constraint in Equation (17):

$$0 \le \hat{p}_{t+j} \le 1 \quad (17)$$

3.5. Two-Stage Training Strategy

Model training is divided into two stages: auxiliary representation pretraining and power-forecasting training.

3.5.1. Auxiliary Representation Pretraining

In the first stage, the weather pseudo-labels and supervised contrastive loss $\mathcal{L}_{\mathrm{con}}$ defined in Section 3.2 are used to train the historical Transformer encoder and projection head, enabling the encoder to learn historical-window representations that discriminate among weather states.

3.5.2. Power-Forecasting Training

After auxiliary pretraining, the historical-encoder parameters are retained, the projection head used only for contrastive learning is removed, and the historical encoder is connected to the gated forecasting network.

During forecasting training, future actual capacity-normalized power serves as the supervision target, and the mean squared error in Equation (18) is adopted:

$$\mathcal{L}_{\mathrm{pred}} = \frac{1}{L_f} \sum_{j=1}^{L_f} \left(\hat{p}_{t+j} - p_{t+j}\right)^2 \quad (18)$$

Both stages use the AdamW optimizer with gradient-norm clipping. Before each training epoch, only the reading order of the training-window samples is shuffled; the temporal order within each window and the correspondence between historical inputs and future labels remain unchanged. The validation set is used for early stopping, and the test set is not involved in training or model selection.

3.6. Parameter-Efficient Cross-Station Adaptation

Parameter-efficient adaptation is investigated as an extended experiment to examine the trade-offs among forecasting performance, the number of trainable parameters, and memory overhead under different parameter-update schemes when target-station training data are limited.

3.6.1. Adapter Adaptation

For an intermediate Transformer representation $\mathbf{h}$, a bottleneck Adapter is added as defined in Equation (19):

$$\mathrm{Adapter}(\mathbf{h}) = \mathbf{h} + \mathbf{W}_{\mathrm{up}}\sigma(\mathbf{W}_{\mathrm{down}}\mathbf{h} + \mathbf{b}_{\mathrm{down}}) + \mathbf{b}_{\mathrm{up}}, \quad (19)$$

3.6.2. LoRA Adaptation

In addition to Adapter, Low-Rank Adaptation (LoRA) is evaluated. For a linear projection matrix $\mathbf{W}$ in a Transformer attention module, LoRA does not directly update the original weights but represents the parameter increment as the product of two low-rank matrices, as shown in Equations (20) and (21):

$$\mathbf{W}' = \mathbf{W} + \Delta\mathbf{W}, \quad (20)$$

$$\Delta\mathbf{W} = \lambda\mathbf{B}\mathbf{A}, \quad (21)$$

LoRA is applied to the Q, K, and V projections of the attention module. The original backbone parameters are frozen, and only the low-rank increment parameters and the parameters of the gated forecasting network are optimized.

4. Experiments and Results

4.1. Dataset and Experimental Settings

Experiments are conducted using the public PVOD dataset [10], which contains power, NWP, LMD, and metadata for ten photovoltaic stations in China at a temporal resolution of 15 min. For each station, the data are chronologically divided into training, validation, and test segments comprising 65%, 15%, and 20% of the data, respectively, with a 192-time-step gap between adjacent segments. All timestamps are converted to the Asia/Shanghai time zone.

Both the historical and forecasting windows contain 96 steps, and the window stride is four for all data segments. The encoder dimension and window-representation dimension are both 32. The encoder comprises one Transformer layer, four attention heads, and a 128-dimensional feed-forward layer. The contrastive projection dimension is 32, and the temperature coefficient is $\tau = 0.1$. The gating network has a time-step feature dimension of 6, a hidden dimension of 64, an output dimension of 3, and a dropout rate of 0.1. The Adapter bottleneck dimension is 8. In the main experiments, auxiliary contrastive pretraining is fixed at five epochs, and gated regression is trained for at most 20 epochs, with an early-stopping patience of 8 and a minimum improvement threshold of 10^{-7}. In the extended transfer experiments, LoRA uses $rank = 4$ and $alpha = 8$, and the proportions of target-station training windows are set to 1%, 10%, and 100%. Both stages use AdamW with a

learning rate of 10^{-3}, a weight decay of 10^{-4}, a batch size of 64, and a gradient-norm clipping threshold of 1.0. Forecasting performance is evaluated using MAE, RMSE, and the standard coefficient of determination R^2.

All-day metrics include a large number of zero values at night, whereas daytime metrics are calculated only for time steps with a solar elevation angle greater than zero. Cross-station macro averages are computed by first calculating the metrics separately for each station and then averaging them.

4.2. Comparative Experiments

Weather-supervised contrastive pretraining and power-forecasting training are conducted separately for each station. Models are selected on the validation set, all metrics are computed on the test set, and the results are compared with those of multiple baseline methods.

Table 1: Comparison of Averaged Metrics Across Ten Stations.

Method	Daytime MAE (MW)	Daytime RMSE (MW)	Daytime R^2	All-Day MAE (MW)	All-Day RMSE (MW)	All-Day R^2
Ours	**1.5022**	**2.3659**	**0.7758**	**0.8829**	**1.8158**	**0.8634**
NWP	1.6036	2.5038	0.7514	0.9430	1.9218	0.8483
GRU	1.9564	2.7282	0.6941	1.1623	2.1011	0.8104
TimesNet-style	1.8951	2.6538	0.7173	1.1251	2.0446	0.8247
DLinear	2.2823	3.1915	0.4675	1.3478	2.4531	0.6686
Clear-Sky Persistence	2.5677	4.1137	0.3195	1.5089	3.1540	0.5888

Note: Lower MAE and RMSE values indicate better performance, whereas R^2 values closer to 1 are preferable. Each column reports the macro average across the ten stations, and boldface denotes the best result in that column.

As shown in Table 1, the proposed method achieves the best results for all three macro-averaged metrics in both the daytime and all-day evaluations. Compared with NWP, the second-best method, the proposed method reduces daytime MAE and RMSE by 6.32% and 5.51%, respectively, while R^2 increases from 0.7514 to 0.7758. Its advantages over TimesNet-style, GRU, and DLinear are even more pronounced. These results indicate that the physical basis directly using NWP is already a strong forecasting pathway, and that incorporating historical persistence bases and gated fusion further reduces forecasting errors.

4.3. Core Ablation of Physical Bases

To directly examine the core structure of the proposed method, the ablation compares only three settings: the full model, removal of the NWP physical basis, and removal of both persistence bases. Preprocessing is fitted independently and training is conducted on station01 for this experiment. Therefore, the Full result in Table 2 is intended only for comparison within this ablation group and should not be conflated with the station01 results from the main experiment.

Table 2: Daytime Physical-Basis Ablation Results for station01.

Variant	MAE (MW)	RMSE (MW)	R^2
Full	**1.5489**	**2.5197**	**0.8243**
w/o NWP basis	2.4461	4.0434	0.5476
w/o persistence bases	1.6298	2.5396	0.8215

After the NWP physical basis is removed, the model can fuse only the previous-day and most-recent-valid clear-sky-index persistence bases. MAE increases from 1.5489 MW to 2.4461 MW, RMSE increases from 2.5197 MW to 4.0434 MW, and R^2 decreases from 0.8243 to 0.5476. This pronounced degradation demonstrates that the NWP physical basis, which contains future meteorological information, is the primary information source for day-ahead forecasting. It also explains the advantage of the full model over purely historical extrapolation methods.

After both persistence bases are removed, the model degenerates to a single NWP physical basis. MAE increases to 1.6298 MW, while RMSE and R^2 also deteriorate slightly. These changes are much smaller than those caused by removing the NWP basis, indicating that the persistence bases primarily provide bias correction and supplementary stable information rather than replacing NWP. Together, the two ablations support the core design of using the NWP physical basis as the primary component, supplemented by historical persistence bases.

4.4. Cross-Station Transferability Experiment

Transfer experiments are conducted to analyze the cross-station adaptation capability of the model. Station01 is used as the target station, and contrastive pretraining is performed on the training sets of the other nine stations. Adaptation is then conducted using 1%, 10%, and 100% of the training and validation windows from the target station. Target-station samples are not used in source-station pretraining. Table 3 reports the daytime MAE for the three adaptation methods.

Table 3: Daytime MAE (MW) for Cross-Station Adaptation on station01.

Adaptation Method	1% Data	10% Data	100% Data	Trainable Parameters (%)
Adapter	**1.8899**	1.5972	1.4970	**14.20%**
LoRA	1.8900	1.6026	1.5905	15.15%
Full Fine-Tuning	1.9028	**1.5819**	**1.4755**	100%

With 1% of the data, the difference in MAE between Adapter and full fine-tuning is small. With 10% and 100% of the data, full fine-tuning achieves slightly higher accuracy. However, the parameter-efficient methods update only a small proportion of the parameters and therefore provide a more favorable trade-off between training cost and performance.

5. Conclusions

To address forecast errors in future meteorological information and the information gap between historical observations and future weather in day-ahead photovoltaic power forecasting, this paper proposes a method that fuses multiple physical bases through weather-conditioned gating. Three deterministic physical bases are constructed from NWP, the clear-sky index at the same time on the previous day, and the most recent valid clear-sky index. A Transformer historical encoder is pretrained using weather pseudo-labels and supervised contrastive learning to enhance the representation of historical weather states. On this basis, a time-step-wise gating network adaptively learns fusion weights for the different physical bases, allowing the model to integrate future weather priors with historical persistence information. The architecture also guarantees zero nighttime output, nonnegativity, and compliance with the capacity limit. Experimental results demonstrate the effectiveness of the proposed method.

The current method remains relatively dependent on NWP quality, and the scale of the cross-station experiments is limited. Future research may investigate NWP bias correction, multisource meteorological-information fusion, probabilistic forecasting, and parameter-efficient adaptation at a larger scale to improve model robustness and practical deployability.

References

[1] Antonanzas, J., Osorio, N., Escobar, R., et al. (2016) Review of Photovoltaic Power Forecasting. Solar Energy, 136, 78-111.
[2] Hochreiter, S. and Schmidhuber, J. (1997) Long Short-Term Memory. Neural Computation, 9(8), 1735-1780.
[3] Zeng, A., Chen, M., Zhang, L. and Xu, Q. (2023) Are Transformers Effective for Time Series Forecasting? Proceedings of the AAAI Conference on Artificial Intelligence, 37(9), 11121-11128.
[4] Wu, H., Hu, T., Liu, Y., et al. (2023) TimesNet: Temporal 2D-Variation Modeling for General Time Series Analysis. International Conference on Learning Representations.
[5] Holmgren, W.F., Hansen, C.W. and Mikofski, M.A. (2018) pvlib python: A Python Package for Modeling Solar Energy Systems. Journal of Open Source Software, 3(29), 884.
[6] Khosla, P., Teterwak, P., Wang, C., et al. (2020) Supervised Contrastive Learning. Advances in Neural Information Processing Systems, 33, 18661-18673.
[7] Vaswani, A., Shazeer, N., Parmar, N., et al. (2017) Attention Is All You Need. Advances in Neural Information Processing Systems, 5998-6008.
[8] Houlsby, N., Giurgiu, A., Jastrzebski, S., et al. (2019) Parameter-Efficient Transfer Learning for NLP. Proceedings of the 36th International Conference on Machine Learning, 2790-2799.
[9] Hu, E.J., Shen, Y., Wallis, P., et al. (2022) LoRA: Low-Rank Adaptation of Large Language Models. International Conference on Learning Representations.
[10] Yao, T., Wang, J., Wu, H., et al. (2021) A Photovoltaic Power Output Dataset: Multi-Source Photovoltaic Power Output Dataset with Python Toolkit. Solar Energy, 230, 122-130.