Journal of Zhejiang University-SCIENCE A 2026 27(8):885-896
www.jzus.zju.edu.cn; www.springer.com/journal/11582
E-mail: jzus_a@zju.edu.cn

Research Article
https://doi.org/10.1631/jzus.A2500582

Physics-informed deep learning for data-efficient and robust photovoltaic power forecasting

Chang HUANG[1], Xuanbin HUANG[1], Jinmin GUO[2], Zhuo CAO[1], Ting HE[1]✉, Wentao SHANG[1]✉

[1]*Energy and Electricity Research Center, Jinan University, Zhuhai 519070, China*
[2]*Beijing Tsintergy Technology Co., Ltd., Guangzhou 510510, China*

Abstract: Photovoltaic (PV) power forecasting is challenged by its inherent variability. Pure data-driven models struggle with generalization under data scarcity and complex weather conditions. In this paper, we introduce a physics-informed deep learning hybrid model (PIDL-HM) that systematically generates physically grounded input features (e.g., plane-of-array irradiance and module temperature) for a convolutional neural network–long short-term memory (CNN–LSTM), establishing a principled integration framework beyond simple ensemble methods. Rigorously validated across multiple PV power plants in China and Australia for 15-min, 4-h, and 24-h forecasting, our approach demonstrates superior performance, with a reduction of up to 8.93% in root mean square error compared to a purely data-driven baseline. Crucially, the model shows remarkable data efficiency, maintaining high accuracy with only three months of training data, and exceptional robustness, providing a 6.87% improvement in performance under strong cross-seasonal data distribution shifts. This work provides a reliable and data-efficient forecasting solution, establishing the PIDL-HM as a foundational element for next-generation forecasting systems.

Key words: Photovoltaic (PV) power forecasting; Physics-informed deep learning; Various forecasting horizons; Data efficiency; Physically grounded features

1 Introduction

1.1 Research background

The global energy transition is accelerating, with photovoltaic (PV) power becoming a key renewable energy source. The share of wind and solar PV in global electricity generation is projected to increase substantially by 2030 (IEA, 2024), but their intermittency and volatility challenge grid stability. High-accuracy forecasting is thus critical for enhancing renewable energy integration and reducing operational costs, supporting a sustainable energy future.

1.2 Current status of PV power forecasting research

PV forecasting techniques can be categorized into physical, data-driven, and hybrid methods. Physical forecasting methods simulate PV power generation using radiation transfer equations and component characteristics, integrating meteorological data and system parameters. They avoid the need for large historical datasets but require accurate parameters, limiting their generalizability and robustness in practical applications due to sensitivity to external disturbances. For instance, Mayer and Gróf (2021) constructed a model chain based on seven key processes of PV power generation, systematically comparing 32400 physical model combinations and validating them against one year of production data from 16 actual power plants. Their study revealed that models performing well experimentally are not necessarily superior in practical applications, highlighting the robustness limitations of purely physical models in engineering scenarios.

AI techniques have been widely adopted in PV forecasting owing to their ability to accurately capture complex nonlinear mapping relationships and long-term temporal dependencies between inputs and outputs. For instance, Rao et al. (2024) integrated a gated recurrent unit (GRU) with a temporal convolutional network (TCN) and attention mechanisms to model complex input–output mappings, significantly improving

✉ Ting HE, heting@jnu.edu.cn
Wentao SHANG, wtshang@jnu.edu.cn
Chang HUANG, https://orcid.org/0000-0002-4987-9840
Ting HE, https://orcid.org/0000-0003-1718-9906
Wentao SHANG, https://orcid.org/0000-0002-5168-7696

Received Nov. 10, 2025; Revision accepted Feb. 9, 2026; Crosschecked Apr. 8, 2026; Online first May 28, 2026

forecasting accuracy over traditional methods. Other widely used AI methods include support vector regression (SVR) (Das et al., 2017), extreme gradient boosting (XGBoost) (Zhu CY et al., 2025), backpropagation neural networks (Liu et al., 2017), convolutional neural networks (CNNs) (Agga et al., 2021), and long short-term memory (LSTM) networks (Sun et al., 2025). Beyond model construction, scholars have explored forecasting across different time scales. For example, Zhu JB et al. (2023) proposed a multi-scale fluctuation characteristic extraction method that integrates meteorological data, historical PV power, and temporal dynamics into an XGBoost-based model, demonstrating enhanced forecasting accuracy across both minute-level and hour-level scales.

In recent years, researchers have begun to embed physical mechanisms into data-driven frameworks to improve the accuracy and robustness of deterministic PV power prediction. For instance, Ogliari et al. (2017) introduced a clear sky irradiation model into a neural network; Niccolai et al. (2021) coupled a component performance model with deep learning; Mayer (2022) systematically compared physical, data-driven, and hybrid irradiance-to-power conversion methods, highlighting the importance of physical model chain optimization and consistency in forecast directives. Their work established a valuable benchmark for hybrid modeling, particularly in scenarios with sufficient historical data. However, these studies mainly adopt a shallow integration strategy, where physical models are used to generate predictions or constraints that are later processed by machine learning models. Moreover, their robustness under data scarcity and distribution shifts remains unproven, and they lack adaptability to diverse forecasting paradigms and regional climates. This constitutes a critical research gap that our study aims to address.

1.3 Research gap and contributions

Current hybrid methods rely mainly on shallow integration, where machine learning merely corrects physical model outputs. As noted by Mayer (2022), such approaches require extensive historical data (1–2 years) to outperform purely physical baselines and lack robustness under data scarcity or distribution shifts. To bridge this gap, we developed a physics-informed deep learning hybrid model (PIDL-HM) that reshapes the feature space from the ground up. By leveraging physical models to generate high-fidelity, stable input features and using deep learning to capture residual dynamics, the framework ensures high accuracy and robustness in data-scarce and non-stationary environments.

The main contributions of this paper are summarized as follows:

1. Adaptive PIDL framework: a principled methodology for adaptive feature engineering that optimizes inputs based on forecasting horizons (sequence- vs. point-based) and regional climates.

2. Multi-horizon validation: rigorous validation across various forecasting horizons (15-min, 4-h, and 24-h) using diverse datasets from China and Australia to demonstrate broad applicability.

3. Robustness and data efficiency: empirical evidence showing superior performance under data-scarce scenarios (three-month training) and cross-seasonal distribution shifts.

4. Synergistic performance enhancement: demonstration that PIDL-HM establishes a physically constrained baseline that synergistically enhances performance when fused with models such as XGBoost.

2 PIDL-HM

PIDL-HM integrates white-box physical modeling with black-box data-driven learning (Fig. 1). Physical models simulate mechanistic processes to transform raw meteorological data into stable, high-fidelity features. These interpretable inputs guide and regularize the deep learning network, which captures complex nonlinear temporal dependencies and residual dynamics that are difficult to model purely physically.

2.1 Physics-based feature generation

The PIDL-HM framework begins with a physics-based front-end. This module translates fundamental input data into intermediate physical variables, such as plane-of-array irradiance and module temperature, which directly govern the PV power conversion process.

2.1.1 Solar position calculation

The solar position is defined by several angles, calculated based on the location’s latitude, longitude, and time.

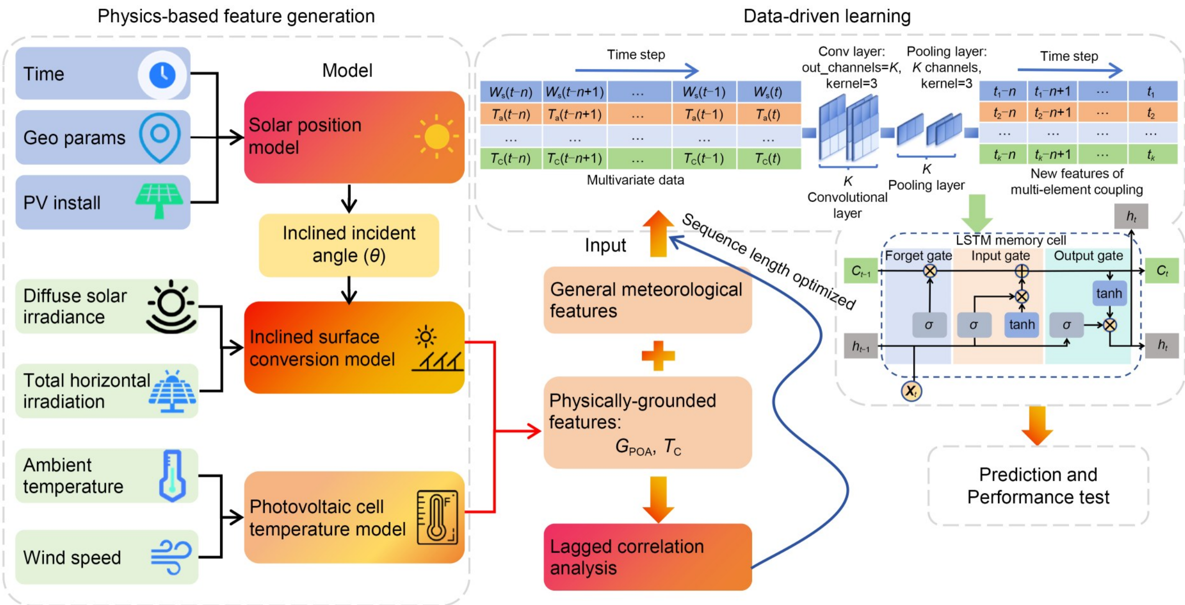

Fig. 1 Workflow of PIDL-HM. ***θ*****: angle of incidence;** W_s**: wind speed;** T_a**: ambient temperature;** T_C**: PV module temperature;** G_{POA}**: total irradiance on a tilted surface;** X_t**: input vector at time step** ***t*****;** h_{t-1} **and** h_t**: hidden states at the previous and current time steps, respectively;** C_{t-1} **and** C_t**: cell states at the previous and current time steps, respectively;** ***σ*****: sigmoid activation function; tanh: hyperbolic tangent activation function**

Solar time T_{sun} is the local time adjusted for the equation of time and longitude correction.

$$T_{sun}=T_{local}+\frac{T_E}{60}-C_L-D, \tag{1}$$

where T_{local} is the local time (h), T_E is the equation of time (min), C_L is the longitude correction (h), and D is the daylight-saving time indicator (1 if active, 0 otherwise). The solar hour angle ω is derived from T_{sun}.

T_E corrects for the eccentricity of Earth's orbit and axial tilt:

$$T_E=9.87\sin 2B-7.53\cos B-1.5\sin B, \tag{2}$$

where 9.87, 7.53, and 1.5 are empirically fitted coefficients. $B=\frac{360}{365}(N_{day}-81)$ and N_{day} is the day of the year, ranging from 1 to 365.

The declination angle δ is the angle between the rays of the sun and the plane of the Earth's equator:

$$\delta=23.45°\sin\left[\frac{360}{365}(284+N_{day})\right], \tag{3}$$

where 23.45° denotes the approximate obliquity of the ecliptic.

The solar altitude angle α is the angle between the sun's rays and the horizontal plane:

$$\sin\alpha=\sin\phi\sin\delta+\cos\phi\cos\delta\cos\omega, \tag{4}$$

where ϕ denotes the geographic latitude.

The solar azimuth angle γ_s is the compass direction from which the sunlight arrives, measured from true north:

$$\cos\gamma_s=\frac{\sin\alpha\sin\phi-\sin\delta}{\cos\alpha\cos\phi}. \tag{5}$$

The angle of incidence θ is the angle between the sun's rays and the normal to the PV panel surface. Consider a fixed panel with tilt angle β and azimuth angle γ_p, the relationship is expressed as

$$\cos\theta=\sin\alpha\cos\beta+\cos\alpha\sin\beta\cos(\gamma_s-\gamma_p). \tag{6}$$

2.1.2 Tilted plane-of-array (POA) irradiance model

The total irradiance on a tilted surface G_{POA} is the sum of its beam $G_{b,T}$, diffuse $G_{D,T}$, and ground-reflected $G_{r,T}$ components:

$$G_{POA}=G_{b,T}+G_{D,T}+G_{r,T}, \tag{7}$$

$$G_{\mathrm{b,T}}=G_{\mathrm{b}}\cos\theta, \tag{8}$$

$$G_{\mathrm{r,T}}=\frac{1}{2}\rho G_{\mathrm{H}}(1-\cos\beta), \tag{9}$$

where G_b is the beam normal irradiance, ρ is the ground albedo (typically 0.2), and G_H is the global horizontal irradiance.

In this study, we use the Hay model (Hay, 1979), an anisotropic diffuse sky model, for superior accuracy:

$$G_{\mathrm{d,T}}=G_{\mathrm{D}}\left[\frac{G_{\mathrm{b}}}{G_{\mathrm{ext}}}R_{\mathrm{b}}+\left(1-\frac{G_{\mathrm{b}}}{G_{\mathrm{ext}}}\right)\frac{1+\cos\beta}{2}\right], \tag{10}$$

where G_D is the horizontal diffuse irradiance, G_{ext} is the extraterrestrial irradiance, $R_b=\frac{\cos\theta}{\cos\theta_z}$ is the geometric factor, and θ_z is the zenith angle.

2.1.3 PV module temperature model

The operating temperature of PV modules significantly impacts their conversion efficiency. In this study, we adopt the Mattei model (Mattei et al., 2006), recognized for its accuracy across diverse installations. It models the module temperature T_C as a function of irradiance, ambient temperature, and wind speed:

$$T_{\mathrm{C}}=\frac{HT_{\mathrm{a}}+G_{\mathrm{POA}}\left(\alpha\tau-\eta\left(1+\mu_{\mathrm{P}}T_{\mathrm{a,STC}}\right)\right)}{H-\left(\mu_{\mathrm{P}}\eta G_{\mathrm{POA}}\right)}, \tag{11}$$

where T_a is the ambient temperature (℃), $\alpha\tau$ is the absorptance-transmittance product, η is the module efficiency, and μ_P is the power temperature coefficient of the module. H is a heat transfer coefficient, with W_s representing the wind speed (m/s), defined as H=26.6+2.3W_s. The subscript STC represents the standard test condition of 25 ℃.

2.2 Data-driven model component

The data-driven component uses a CNN−LSTM hybrid architecture. CNN layers extract spatial features and correlations among multi-dimensional meteorological variables, while the LSTM network models the long-term temporal dependencies within the sequential data.

2.3 Hybrid integration strategy

PIDL-HM adopts a sequential integration strategy. The physics-based front-end first generates G_{POA} and T_C, which are concatenated with numerical weather prediction (NWP) variables and historical power data to form an enriched feature set. This multi-variable time-series matrix is processed by one-dimensional convolutional layers for spatial feature extraction. The resulting feature maps are reshaped and fed into LSTM layers to model temporal evolution. Finally, fully connected layers generate the power forecast. This architecture ensures that the learning process is anchored by physical principles while effectively capturing complex residual patterns.

2.4 Model evaluation metrics

1. Root mean square error (RMSE) E_{RMS}: a standard measure of the difference between values predicted by a model and the observed values.

$$E_{\mathrm{RMS}}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(P_i^{\mathrm{pred}}-P_i^{\mathrm{actual}}\right)^2}, \tag{12}$$

where n is the number of samples, P_i^{pred} is the predicted power, and P_i^{actual} is the actual power.

2. Forecasting accuracy F_A: as defined by the SAMR (2021), this metric provides a normalized accuracy score:

$$F_{\mathrm{A}}=\left(1-\frac{E_{\mathrm{RMS}}}{C}\right)\times 100\%, \tag{13}$$

where C is the installed capacity of the PV plant. This metric offers an intuitive percentage representation of forecast performance. Since this indicator F_A is derived from E_{RMS}, only E_{RMS} values are discussed during the case analysis.

3 Case analysis

3.1 Data description

The datasets are sourced from public databases and industry-academia collaborations, encompassing multiple PV power stations in China (Guangzhou and Sanming) and Australia (Alice Springs and Canberra) with installed capacities ranging from 1.00 to 8.67 MW. The data resolution varies from 5 min to 1 h, and the dataset lengths span from 3 to 27 months. All data are split in chronological order (70%/15%/15% for training/validation/testing). Detailed information is summarized in Table 1.

Table 1 Detailed information for each station

Location and size	Category		Dataset source	Data length	Resolution	Feature parameter
Alice Springs, 4.95 MW	M	√	Desert Knowledge Australia Solar Centre	Jan. 1, 2015–Dec. 31, 2015	5 min	Wind speed W_s, ambient temperature T_a, rainfall R, humidity H_u, global horizontal irradiance G_H, diffuse horizontal irradiance G_D, and photovoltaic power P
	F	×			–	
Sanming, 2.77 MW	M	√	Digital China Innovation Contest 2024	Nov. 17, 2022–Apr. 30, 2023	15 min	
	F	√	Xihe-Energy		15 min	
Guangzhou, 8.67 MW	M	√	Skyworth Photovoltaic Power Station of Guangzhou Bureau	Jan. 1, 2025–June 19, 2025	1 h	
	F	√			15 min	
Canberra, 1.00 MW	M	×	Global Energy Forecasting Competition 2014	Jan. 1, 2012–Mar. 31, 2014	–	
	F	√			1 h	

M: measured data; F: forecast data. "√" and "×" indicate whether the corresponding data/variable is available or unavailable. "–" means no corresponding content

3.2 Correlation analysis

Correlation analysis (Fig. 2) guides our feature and sequence length selection. In the Alice Springs dataset, G_H and G_{POA} show strong correlations with power P, unlike the weak correlation with rainfall R. This pattern generally holds for the Sanming and Guangzhou datasets, although humidity H_u exerts a more significant influence at these sites.

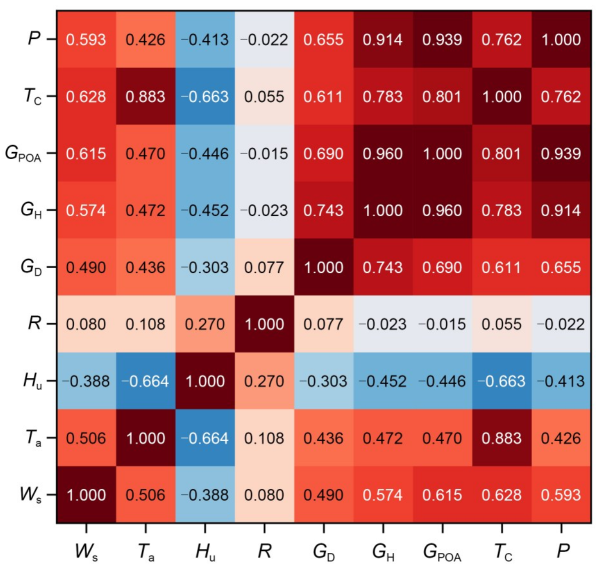

Fig. 2 Real-time correlation heatmap of Alice Springs

Furthermore, lagged correlation analysis is performed to identify the optimal input sequence length for the time-series model. The average lagged correlation (ALC) for Alice Springs at 15-min resolution (Fig. 3a) reveals a significant periodic pattern of about 24 h. Based on the ALC curve's key inflection points (e.g., peaks and troughs), a set of candidate input lengths is selected (e.g., 4, 10, ⋯, 106 data points) for subsequent experimental optimization. A similar strategy is applied to other regions and the 4-h forecasting task. These points include lags of 12, 24, 36, and 48 h (Fig. 3b). These four time points cover different temporal ranges, from short-term intra-day patterns to long-term cross-day periodic information.

3.3 Model parameter settings

Table 2 summarizes the hyperparameter configuration for the CNN–LSTM model. To accommodate the increasing complexity of input patterns as the forecasting horizon extended from 15 min to 24 h, the model capacity is adaptively scaled. Specifically, the number of LSTM hidden units is increased from 64 to 144 to capture broader temporal dependencies, while the dropout ratio is raised from 0.0 to 0.4 to prevent overfitting. Concurrently, CNN output channels are optimized to maintain computational efficiency while reducing feature redundancy.

3.4 Forecasting experiments across various forecasting horizons

A controlled experiment is designed to evaluate the PIDL-HM's effectiveness and optimize feature engineering. Two case groups are established: (1) pure data-driven baselines (Cases 1-1 to 1-3) with various feature correlation strengths; (2) PIDL-HM variants (Cases 2-1 to 2-6) incorporating physically grounded features (G_{POA}, T_C) in different combinations (taking the Alice Springs dataset as an example, all case details are summarized in Table 3). These case settings are applied consistently across all forecasting horizons and regions.

3.4.1 15-min forecasting experiment

Table 4 presents the RMSE metrics across various input sequence lengths and feature combinations for the 15-min task, using the Alice Springs dataset as a representative case. Guided by lagged correlation analysis, optimal sequence lengths (94 for Alice Springs; 95 for Sanming, whose ablation table is not included here for brevity) are identified to balance historical context with noise, effectively capturing diurnal patterns.

As detailed in Tables 3 and 4, integrating physical features (G_{POA}, T_{C}) consistently enhanced the prediction precision. The hybrid Case 2-2, which synergizes these physical priors with meteorological variables (e.g., W_{s} and G_{H}), achieved the optimal performance, outperforming both purely data-driven and standalone physical benchmarks. Notably, this configuration reduced RMSE by 1.09% (when the input sequence length is 94) compared to the all-feature baseline (Case 1-1). This superiority persisted across both the

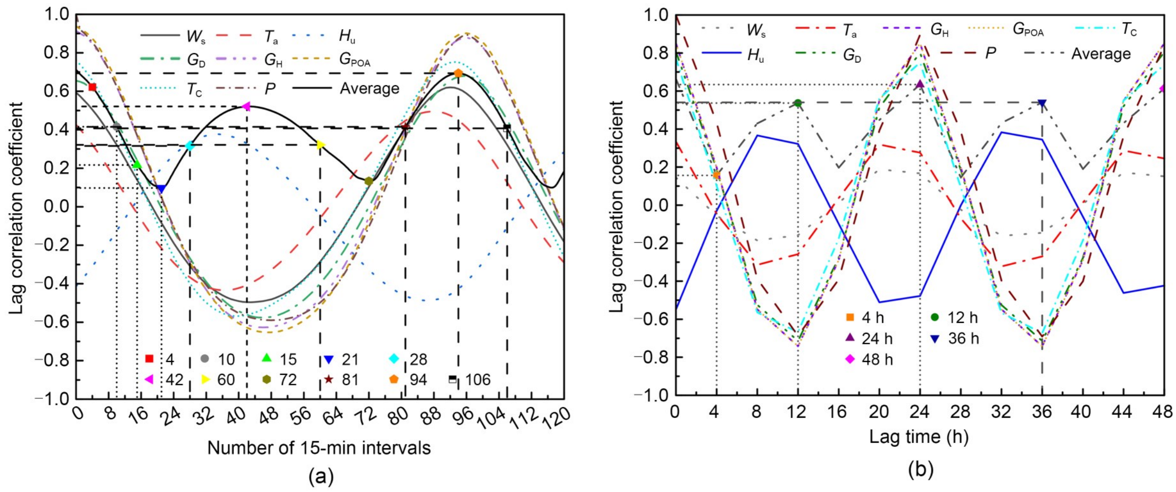

Fig. 3 15-min interval lag correlation of the Alice Springs dataset: (a) 15-min lag correlation of Alice Springs; (b) 4-h interval lag correlation of Alice Springs

Table 2 CNN–LSTM neural network simulation parameters

Parameter	Detail			Parameter	Detail		
	15-min	4-h	24-h		15-min	4-h	24-h
Convolutional layer and kernel	1, 3	1, 3	1, 3	Dropout	0.0	0.2	0.4
Pooling layer and kernel	1, 3	1, 3	1, 3	Rate	0.001	0.001	0.001
Output unit	32	16	8	Number of iterations	100	100	100
Number of LSTM layers	1	1	1	Optimizer	Adam	Adam	Adam
LSTM hidden unit	64	128	144				

Table 3 Case settings for Alice Springs

Case	Description	Combination							
		W_{s}	T_{a}	H_{u}	R	G_{D}	G_{H}	G_{POA}	T_{C}
Case 1-1	All available meteorological features	√	√	√	√	√	√	–	–
Case 1-2	Moderately correlated features	√	√	√	–	√	√	–	–
Case 1-3	Highly correlated features	√	–	–	–	√	√	–	–
Case 2-1	Full meteorological and physically grounded features	√	√	√	√	√	√	√	√
Case 2-2	Moderately correlated and physically grounded features	√	√	√	–	√	√	√	√
Case 2-3	Highly correlated and physically grounded features	√	–	–	–	√	√	√	√
Case 2-4	Pure physically grounded features	–	–	–	–	–	–	√	√
Case 2-5	Single-information features	–	–	√	–	–	–	√	√
Case 2-6	Overlapping-information features	–	√	–	–	–	√	√	√

"√" indicates that the feature is used. "–" indicates that the feature is not used

Table 4 RMSE metrics for the 15-min forecasting task in Alice Springs

Input sequence length (number of 15-min intervals)	E_{RMS} (MW)								
	Case 1-1	Case 1-2	Case 1-3	Case 2-1	Case 2-2	Case 2-3	Case 2-4	Case 2-5	Case 2-6
4	0.3389	0.3391	0.3387	0.3366	**0.3360**	**0.3330**	0.3387	0.3367	0.3398
10	0.3360	0.3379	0.3347	**0.3288**	0.3321	**0.3307**	0.3386	0.3337	0.3402
15	0.3267	0.3276	0.3281	**0.3241**	**0.3221**	0.3264	0.3340	0.3324	0.3407
21	0.3224	**0.3201**	0.3279	0.3251	0.3216	**0.3211**	0.3296	0.3282	0.3383
28	0.3217	0.3231	0.3244	0.3222	**0.3207**	**0.3214**	0.3257	0.3237	0.3304
42	0.3223	0.3257	**0.3211**	0.3254	**0.3212**	0.3246	0.3246	0.3274	0.3274
60	0.3225	**0.3203**	**0.3215**	0.3232	0.3232	0.3240	0.3251	0.3251	0.3310
72	0.3229	0.3269	0.3288	**0.3226**	**0.3227**	0.3239	0.3261	0.3269	0.3300
81	0.3246	**0.3215**	0.3219	0.3220	**0.3208**	0.3237	0.3255	0.3256	0.3287
94	0.3213	**0.3188**	0.3199	0.3199	**0.3178**	0.3206	0.3246	0.3241	0.3295
106	0.3231	0.3225	**0.3224**	0.3228	**0.3216**	0.3230	0.3253	0.3261	0.3333

Best and second-best results are bold-underlined and bold, respectively

Alice Springs and Sanming datasets, confirming the cross-regional reliability of the PIDL-HM for ultra-short-term forecasting (Fig. 4).

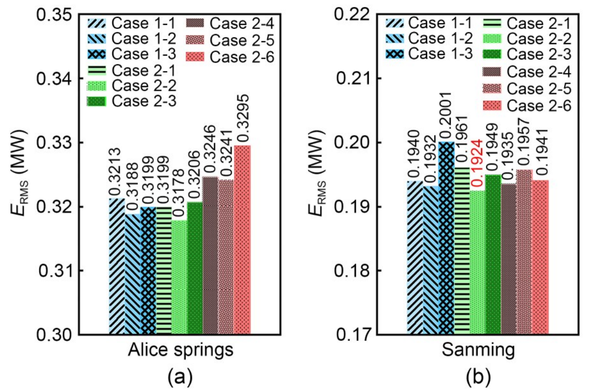

Fig. 4 15-min RMSE results of Alice Springs (at input sequence length of 94) and Sanming (at input sequence length of 95): (a) Alice Springs; (b) Sanming

3.4.2 4-h forecasting experiment

For the 4-h forecast, E_{RMS} values naturally increase across all three regions (Alice Springs, Sanming, and Guangzhou) due to greater inherent uncertainty (Tables 5–7). The optimal input sequence length consistently proves to be 24 h, a full diurnal cycle, as suggested by the ALC analysis. At this optimal length, the PIDL-HM's advantage remains significant.

Case 2-6 emerges as the optimal configuration for the Alice Springs and Guangzhou datasets, reducing the E_{RMS} by 8.48% (to 0.4382 MW) and 6.94% (to 0.4344 MW), respectively, compared to the Case 1-1 baseline. However, the optimal feature strategy proves region-dependent. In Sanming, high meteorological volatility renders standalone physical features (Case 2-4) suboptimal due to error accumulation from noisy inputs. Conversely, the smoother irradiance- and temperature-related feature sequences in Alice Springs and Guangzhou reduce the risk of error accumulation in physical feature generation, enabling the hybrid Case 2-6 to effectively capture system dynamics.

3.4.3 24-h forecasting experiment

Note that for the 24-h forecast task, lagged correlation analysis is omitted as it transitions to an NWP-driven point-to-point mapping paradigm, prioritizing physical causality over historical inertia.

Table 8 summarizes the day-ahead (24-h) forecasting performance across the three climatically diverse regions. The results confirm that integrating physical features consistently enhances accuracy, with the optimal hybrid configuration (Case 2-6) in Guangzhou achieving a reduction in E_{RMS} of 8.93% compared to the Case 1-1 baseline, reaching 92.8% accuracy.

The optimal feature strategy is dictated mainly by regional NWP reliability. In regions characterized by high NWP uncertainty (e.g., Sanming), standalone physical features (Case 2-4) excel by acting as noise filters. Conversely, in environments with stable, high-quality NWP data (e.g., Canberra), a hybrid of physical and meteorological features (Case 2-1) yields the best results. These findings demonstrate that for point-based day-ahead forecasting, the physical robustness of G_{POA} and T_{C} provides a stable causal anchor, effectively mitigating the impact of meteorological forecast errors.

Table 5 RMSE metrics for the 4-h forecasting task in Alice Springs

Lag time (h)	E_{RMS} (MW)								
	Case 1-1	Case 1-2	Case 1-3	Case 2-1	Case 2-2	Case 2-3	Case 2-4	Case 2-5	Case 2-6
4	0.5116	0.5161	**0.5056**	**<u>0.4952</u>**	0.5162	0.5090	0.5161	0.5143	0.5273
12	0.4675	0.4633	0.4783	0.4821	0.4768	0.4742	**0.4584**	0.4759	**<u>0.4507</u>**
24	0.4788	0.4536	0.4607	0.4547	0.4632	0.4486	**0.4476**	0.4521	**<u>0.4382</u>**
36	0.4903	0.4745	0.4798	0.4600	0.5187	**0.4576**	0.4761	0.4583	**<u>0.4573</u>**
48	0.4864	**<u>0.4508</u>**	0.4925	0.4860	0.4704	**0.4569**	0.4889	0.4810	0.4723

Best and second-best results are bold-underlined and bold, respectively

Table 6 RMSE metrics for the 4-h forecasting task in Sanming

Lag time (h)	E_{RMS} (MW)								
	Case 1-1	Case 1-2	Case 1-3	Case 2-1	Case 2-2	Case 2-3	Case 2-4	Case 2-5	Case 2-6
4	0.2698	0.2632	0.2629	**<u>0.2564</u>**	**0.2572**	0.2580	0.2860	0.2999	0.2649
12	0.2825	0.2616	0.2661	**<u>0.2574</u>**	**0.2602**	0.2613	0.2988	0.2914	0.2680
24	0.2614	0.2565	0.2551	**0.2483**	0.2594	**<u>0.2419</u>**	0.2962	0.2760	0.2624
36	0.2655	0.2902	0.2752	0.2606	0.2601	**<u>0.2540</u>**	0.2759	0.2854	**0.2594**
48	0.2870	0.2658	0.2760	0.2706	0.2697	**<u>0.2618</u>**	0.3006	0.2895	**0.2639**

Best and second-best results are bold-underlined and bold, respectively

Table 7 RMSE metrics for the 4-h forecasting task in Guangzhou

Lag time (h)	E_{RMS} (MW)								
	Case 1-1	Case 1-2	Case 1-3	Case 2-1	Case 2-2	Case 2-3	Case 2-4	Case 2-5	Case 2-6
4	0.4917	0.4851	0.4987	0.4854	**<u>0.4771</u>**	**0.4782**	0.5075	0.5065	0.5144
12	0.4949	**0.4878**	0.5273	0.4977	**<u>0.4812</u>**	0.4944	0.5039	0.5017	0.4886
24	0.4668	0.4586	0.4668	0.4619	0.4406	0.4554	**0.4370**	0.4373	**<u>0.4344</u>**
36	0.4808	0.4527	0.4846	0.4754	0.4554	0.4724	0.4603	**0.4508**	**<u>0.4478</u>**
48	0.4849	0.4577	0.4814	0.4725	0.4582	0.4716	**0.4557**	0.4642	**<u>0.4518</u>**

Best and second-best results are bold-underlined and bold, respectively

Table 8 RMSE metrics for the day-ahead (24-h) forecasting in Guangzhou, Canberra, and Sanming

Region	E_{RMS} (MW)								
	Case 1-1	Case 1-2	Case 1-3	Case 2-1	Case 2-2	Case 2-3	Case 2-4	Case 2-5	Case 2-6
Guangzhou	0.6851	**0.6347**	0.6639	0.6586	0.6447	0.6561	0.6435	0.6546	**<u>0.6239</u>**
Canberra	**0.1163**	0.1170	0.1172	**<u>0.1162</u>**	0.1168	0.1165	0.1211	0.1169	0.1218
Sanming	0.3353	0.3254	0.3429	0.3289	0.3304	**0.3079**	**<u>0.2983</u>**	0.3160	0.3083

Best and second-best results are bold-underlined and bold, respectively

3.5 Data efficiency experiments

To further investigate the robustness and generalization capability of the PIDL-HM, in this section we examine the impact of historical data volume on model performance.

3.5.1 Small-sample data experiment

To evaluate the efficacy of PIDL-HM under data scarcity, we conduct the experiment using only three months of training data across all regions and forecasting horizons. Specific results are shown in Tables 9–11.

Under small-sample conditions (three months of training data), the PIDL-HM demonstrates a decisive advantage. While the optimal feature combination varies, the best-performing model in every scenario is invariably a PIDL-HM variant. This advantage magnifies with the system scale and forecast horizon: for Guangzhou's 24-h forecast, the optimal PIDL-HM

Table 9 RMSE metrics for the 15-min small-sample data efficiency experiment

Region	E_{RMS} (MW)								
	Case 1-1	Case 1-2	Case 1-3	Case 2-1	Case 2-2	Case 2-3	Case 2-4	Case 2-5	Case 2-6
Alice Springs	0.2767	0.2744	**0.2678**	0.2812	**0.2668**	0.2847	0.2774	0.2885	0.2701
Sanming	0.1322	0.1305	0.1324	**0.1263**	**0.1260**	0.1263	0.1382	0.1565	0.1351

Best and second-best results are bold-underlined and bold, respectively

Table 10 RMSE metrics for the 4-h small-sample data efficiency experiment

Region	E_{RMS} (MW)								
	Case 1-1	Case 1-2	Case 1-3	Case 2-1	Case 2-2	Case 2-3	Case 2-4	Case 2-5	Case 2-6
Alice Springs	0.4105	0.4441	**0.3958**	0.4099	0.4122	**0.3940**	0.4145	0.4300	0.4276
Sanming	0.2711	0.2764	0.2868	**0.2556**	0.2888	0.2640	0.2764	0.2598	**0.2305**
Guangzhou	0.5246	0.5049	0.4985	0.4986	0.5040	**0.4659**	0.4667	0.5113	**0.4557**

Best and second-best results are bold-underlined and bold, respectively

Table 11 RMSE metrics for the 24-h small-sample data efficiency experiment

Region	E_{RMS} (MW)								
	Case 1-1	Case 1-2	Case 1-3	Case 2-1	Case 2-2	Case 2-3	Case 2-4	Case 2-5	Case 2-6
Canberra	**0.1190**	0.1205	0.1197	**0.1161**	0.1201	0.1196	0.1427	0.1195	0.1457
Sanming	0.2618	0.2689	0.2630	**0.2576**	**0.2598**	0.2665	0.2620	0.2621	0.2679
Guangzhou	0.5559	0.5659	0.5423	0.5987	0.5481	**0.4568**	0.5757	0.5911	**0.5402**

Best and second-best results are bold-underlined and bold, respectively

reduces the E_{RMS} by 17.8% compared to the pure data-driven baseline.

This superior data efficiency stems from the physical prior knowledge embedded in the model, which acts as a powerful inductive bias. While data-driven models overfit to sparse correlations, the physical front-end constrains the solution space to physically plausible outcomes, guiding the learning process effectively, even with limited data.

3.5.2 Data distribution shift experiment

To evaluate robustness under seasonal distribution shifts, we use two experimental designs:

Ex 1: chronological test. The training set's start date is fixed, and its length is varied. The test set immediately follows the training period.

Ex 2: fixed test. The test set is fixed on a specific period (e.g., summer). The training set's length is varied by extending its start date further into the past, thus incorporating different seasons.

For stations with shorter datasets (Sanming, Guangzhou), the performance in Ex 1 fluctuates significantly because the training and test sets cover different, biased seasonal segments (Figs. 5 and 6). Despite this inherent volatility, the PIDL-HM achieves the best performance in most cases. The key conclusion is that the optimal feature combination is context-dependent and should be selected based on the specific training data distribution.

The longer datasets (Alice Springs: 12 months; Canberra: 27 months) reveal distinct distribution shift patterns. In the Alice Springs dataset (Fig. 7), performance dips because the model, trained on mixed seasons (first 8.5 months), is tested solely during the high-radiation southern hemisphere summer (November–December). Similarly, the Canberra dataset (Fig. 8) exhibits clear annual periodicity, with accuracy peaking at full yearly cycles (12/24 months) and degrading at partial intervals (16/27 months). Despite these shifts significantly impacting the baseline models, PIDL-HM consistently maintains superior generalization. However, the variable optimal feature combinations across lengths suggest that strategy selection must align with specific training data distributions.

To evaluate the model's performance in the 24-h forecast task in the Guangzhou region under different training data lengths, Ex 2 is implemented. The test set has a strictly fixed time frame from Aug. 18 to Sept. 18. To construct the training set, the end date is set for July 18, while the start date is progressively extended from May 1 to Jan. 1. This configuration results in five different experimental cases, labeled Ex 2-1 to

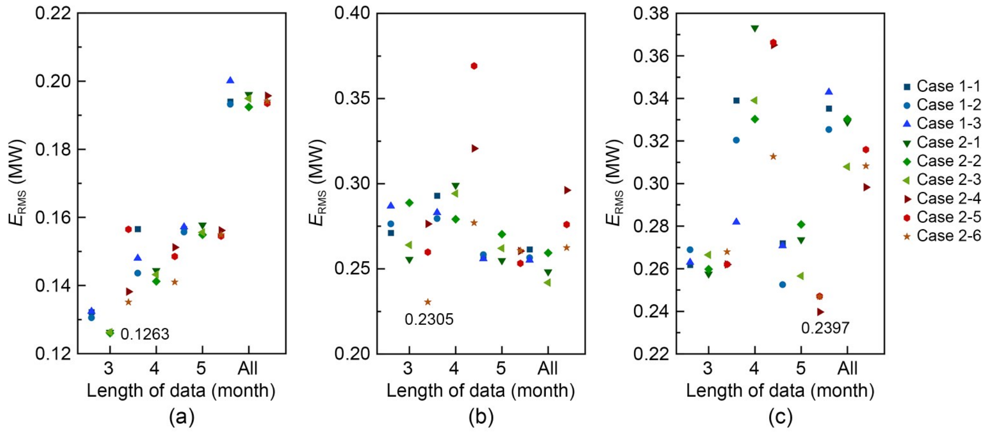

Fig. 5 Data efficiency experiment for Sanming (started on Nov. 17, 2022): (a) 15-min; (b) 4-h; (c) 24-h. Label "All" represents 5.5 months

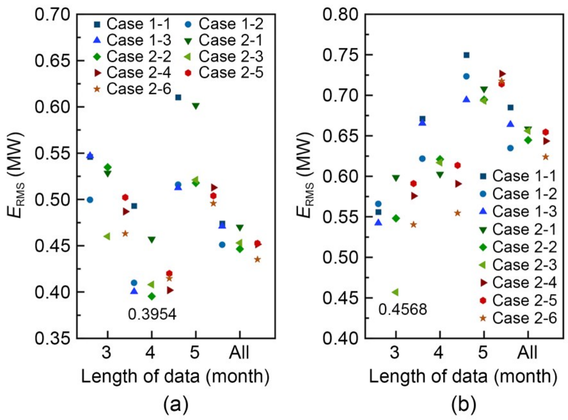

Fig. 6 Data efficiency experiment for Guangzhou (started on Jan. 1, 2025): (a) 4-h; (b) 24-h. Label "All" represents 6.5 months

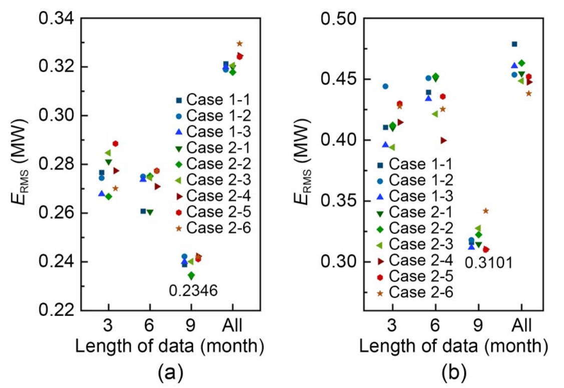

Fig. 7 Data efficiency experiment for Alice Springs (started on Jan. 1, 2015): (a) 15-min; (b) 4-h. Label "All" represents 12 months

Ex 2-5. The relative performance improvement of PIDL-HM compared to the purely data-driven baseline in these five cases is detailed in Table 12.

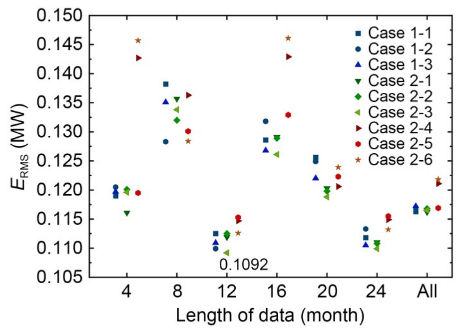

Fig. 8 E_{RMS} results of the 24-h data efficiency experiment for Canberra (started on Jan. 1, 2014). Label "All" represents 27 months

Table 12 provides compelling quantitative evidence of the PIDL-HM's value under distribution shift. The performance gain of the PIDL-HM over the Case 1-1 baseline monotonically increases from 1.05% to a peak of 6.87% as the seasonal mismatch between training and test data increases (Ex 2-4, spanning winter to summer), and it remains high at 6.44% in Ex 2-5. This trajectory confirms that the more pronounced the distribution shift is, the greater the benefit of the embedded physical priors, highlighting the model's particular value for cross-seasonal forecasting and newly built plants.

3.6 Fusion prediction with tree models

To explore PIDL-HM's potential as a foundational component, we integrate it into an ensemble framework. For key scenarios, we fuse predictions from the baseline pure data-driven model and PIDL-HM models with those from an XGBoost model via simple averaging (Table 13).

Table 12 Performance improvement of PIDL-HM under different cases

Scenario	E_{RMS} (MW)									Gain
	Case 1-1	Case 1-2	Case 1-3	Case 2-1	Case 2-2	Case 2-3	Case 2-4	Case 2-5	Case 2-6	
Ex 2-1	0.5047	0.5026	0.5023	0.5069	0.5140	0.5040	**0.5023**	**0.4994**	0.5158	1.05%
Ex 2-2	0.5190	0.5003	0.5185	0.5193	0.5049	0.5336	**0.4952**	**0.4972**	0.5252	4.58%
Ex 2-3	0.5274	0.5029	0.5226	0.5270	**0.4987**	0.5094	**0.4989**	0.5020	0.5266	5.44%
Ex 2-4	0.5314	0.5155	0.5143	0.5317	0.5208	0.5207	**0.4949**	**0.4966**	0.5004	6.87%
Ex 2-5	0.5263	0.5146	0.5369	0.5377	0.5146	**0.4924**	**0.5093**	0.5112	0.5178	6.44%

Best and second-best results are bold-underlined and bold, respectively

Table 13 Performance comparison before and after XGBoost fusion for selected cases

Case	E_{RMS} (MW)			
	Original	XGBoost	PIDL-HM	PIDL-HM+XGBoost
15-min (Alice Springs)	0.3213	0.3138	0.3178	0.3083
4-h (Sanming)	0.2614	0.2426	0.2419	0.2333
24-h (Guangzhou)	0.6851	0.6059	0.6239	0.5789

The fusion strategy shows significant efficacy. For instance, in Sanming's 4-h task, integrating the data-driven baseline with XGBoost reduces E_{RMS} by 7.2%. Notably, replacing the baseline with PIDL-HM as the base component yielded a further 3.8% E_{RMS} reduction. This highlights that PIDL-HM establishes a more accurate and physically constrained foundation for ensemble learning. Consequently, the "PIDL-HM+XGBoost" ensemble consistently outperformed its purely data-driven counterpart, proving that physics-informed outputs provide a superior, high-fidelity basis for broader forecasting pipelines.

4 Conclusions

This study validated a physics-informed deep learning hybrid model (PIDL-HM) for robust and data-efficient PV power forecasting. The main findings are as follows:

1. PIDL-HM achieves deep synergy between physical principles and data-driven learning. This explainable framework yielded significant RMSE reductions across various forecasting horizons, notably up to 8.93% for day-ahead tasks compared to purely data-driven methods.

2. The physics-informed architecture enables high accuracy with only three months of training data. In small-sample experiments, PIDL-HM outperformed the data-driven baseline by up to 17.8% in 24-h forecasting, demonstrating its suitability for newly commissioned power plants.

3. The framework effectively mitigates performance degradation under data distribution shifts. In cross-seasonal tests, PIDL-HM achieved a reduction of up to 6.87% in RMSE over the best data-driven models, confirming its stability in non-stationary environments.

4. Explainable optimization: lagged correlation analysis and context-dependent feature engineering provide a principle-guided approach to model construction, offering practical insights for real-world deployment.

5. PIDL-HM serves as a superior base component for ensemble frameworks. The "PIDL-HM+XGBoost" ensemble consistently outperformed pure data-driven counterparts, proving that physical constraints enhance the entire forecasting pipeline.

In summary, the proposed approach provides a comprehensive solution to the critical challenges of data scarcity and distribution shifts in practical PV forecasting.

Current limitations include the model's dependence on the quality of meteorological forecasts and the potential computational overhead for ultra-large-scale applications. Future research will focus on developing explicit NWP bias-correction techniques and exploring more efficient physics-informed architectures to enhance the framework's scalability and practical scope.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No. 52306013), the Guangdong Basic and Applied Basic Research Foundation (No. 2025A1515012288), and the Fundamental Research Funds for the Central Universities, China (No. 21624212).

Author contributions

Chang HUANG designed the research and drafted, revised, and finalized the paper. Xuanbin HUANG, Jinmin GUO, and Zhuo CAO processed the data. Ting HE and Wentao SHANG helped organize the paper.

Conflict of interest

Chang HUANG, Xuanbin HUANG, Jinmin GUO, Zhuo CAO, Ting HE, and Wentao SHANG declare that they have no conflict of interest.

Declaration on the use of generative AI tools

During the preparation of this work, the authors used ChatGPT and Gemini to improve language and readability. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Data availability

The data that support the findings of this study are available from the corresponding authors upon reasonable request.

References

Agga A, Abbou A, Labbadi M, et al., 2021. Short-term self consumption PV plant power production forecasts based on hybrid CNN–LSTM, ConvLSTM models. *Renewable Energy*, 177:101-112.
https://doi.org/10.1016/j.renene.2021.05.095

Das UK, Tey KS, Seyedmahmoudian M, et al., 2017. SVR-based model to forecast PV power generation under different weather conditions. *Energies*, 10(7):876.
https://doi.org/10.3390/en10070876

Hay JE, 1979. Calculation of monthly mean solar radiation for horizontal and inclined surfaces. *Solar Energy*, 23(4): 301-307.
https://doi.org/10.1016/0038-092X(79)90123-3

IEA (International Energy Agency), 2024. Renewables 2024: Analysis and Forecasts to 2030.
https://www.iea.org/reports/renewables-2024

Liu LY, Liu DR, Sun Q, et al., 2017. Forecasting power output of photovoltaic system using a BP network method. *Energy Procedia*, 142:780-786.
https://doi.org/10.1016/j.egypro.2017.12.126

Mattei M, Notton G, Cristofari C, et al., 2006. Calculation of the polycrystalline PV module temperature using a simple method of energy balance. *Renewable Energy*, 31(4): 553-567.
https://doi.org/10.1016/j.renene.2005.03.010

Mayer MJ, 2022. Benefits of physical and machine learning hybridization for photovoltaic power forecasting. *Renewable and Sustainable Energy Reviews*, 168:112772.
https://doi.org/10.1016/j.rser.2022.112772

Mayer MJ, Gróf G, 2021. Extensive comparison of physical models for photovoltaic power forecasting. *Applied Energy*, 283:116239.
https://doi.org/10.1016/j.apenergy.2020.116239

Niccolai A, Dolara A, Ogliari E, 2021. Hybrid PV power forecasting methods: a comparison of different approaches. *Energies*, 14(2):451.
https://doi.org/10.3390/en14020451

Ogliari E, Dolara A, Manzolini G, et al., 2017. Physical and hybrid methods comparison for the day ahead PV output power forecast. *Renewable Energy*, 113:11-21.
https://doi.org/10.1016/j.renene.2017.05.063

Rao Z, Yang ZM, Yang XP, et al., 2024. TCN-GRU based on attention mechanism for solar irradiance prediction. *Energies*, 17(22):5767.
https://doi.org/10.3390/en17225767

SAMR (State Administration for Market Regulation), 2021. Technical Requirements for Dispatching Side Forecasting System of Wind or Photovoltaic Power, GB/T 40607-2021. National Standards of the People's Republic of the China (in Chinese).

Sun FP, Li LH, Bian DX, et al., 2025. Photovoltaic power prediction based on multi-scale photovoltaic power fluctuation characteristics and multi-channel LSTM prediction models. *Renewable Energy*, 246:122866.
https://doi.org/10.1016/j.renene.2025.122866

Zhu CY, Tu YB, Wei QY, et al., 2025. Short-term power prediction of photovoltaic power station based on LSTM-XGBoost model. *Solar Energy*, 300:113819.
https://doi.org/10.1016/j.solener.2025.113819

Zhu JB, Li MR, Luo L, et al., 2023. Short-term PV power forecast methodology based on multi-scale fluctuation characteristics extraction. *Renewable Energy*, 208:141-151.
https://doi.org/10.1016/j.renene.2023.03.029