Journal of Zhejiang University-SCIENCE A in press
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 PV 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 510630, 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 of up to 8.93% in root mean square error (*RMSE*) 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 has provided a reliable and data-efficient forecasting solution, establishing the PIDL-HM as a foundational element for next-generation forecasting systems.

Key words: Photovoltaic 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) solar becoming a key renewable source. PV generation accounted for 12% of global electricity in 2024 (Iea, 2024), but its intermittency and volatility challenge grid stability. High-accuracy forecasting is thus critical for enhancing renewable integration and reducing operational costs, supporting a sustainable energy future.

1.2 Current Status of Photovoltaic Power Forecasting Research

PV forecasting techniques can be categorized based on methodology 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 generalizability and robustness in practical applications due to sensitivity to external disturbances. For instance, Mayer and Gróf (2021) constructed a model chain based on the seven key processes of PV power generation, systematically comparing 32,400 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, which excel at capturing nonlinear relationships and long-term temporal dependencies, have been widely adopted. For instance, Rao et al. (2024) integrated a Gated Recurrent Unit (GRU) with a Temporal

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

Received Nov. 10, 2025; Revision accepted Feb. 9, 2026; Crosschecked

Convolutional Network (TCN) and attention mechanisms to model complex input-output mappings, significantly improving forecasting accuracy over traditional methods. AI techniques can accurately characterize complex mapping relationships between inputs and outputs under large-data conditions. AI methods widely used in PV forecasting include Support Vector Regression (SVR) (Das et al., 2017), Extreme Gradient Boosting (XGBoost) (Zhu et al., 2025), Backpropagation Neural Networks (Liu et al., 2017), Convolutional Neural Networks (CNN) (Agga et al., 2021), and Long Short-Term Memory (LSTM) networks (Sun et al., 2025). Beyond model construction, scholars have also explored forecasting across different time scales. For example, Zhu et al. (2023) proposed a multi-scale fluctuation characteristics 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, some scholars 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; and 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 adopt mainly 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 aimed to address.

1.3 Research Gap and Contribution

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:

(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) Various Forecasting Horizons: 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) Demonstrating that PIDL-HM establishes a physically-constrained baseline that synergistically enhances performance when fused with models like XGBoost.

2 Physics-informed deep learning hybrid model (PIDL-HM)

The 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 (G_{POA} , T_C). 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.

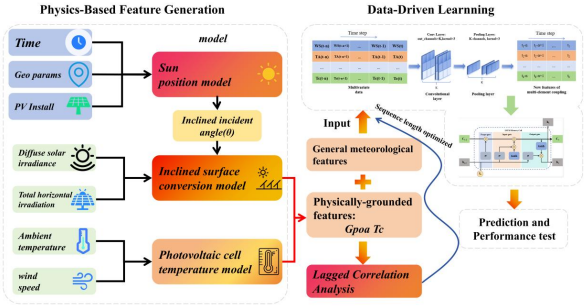


Fig. 1 Workflow of the PIDL-HM

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 sun's position is defined by several angles, calculated based on the location's latitude, longitude, and time.

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

$$T_{sun} = T_{local} + \frac{LC + EOT}{60} - D \tag{1}$$

where T_{local} is the local time [hours], EOT is the equation of time [minutes], LC is the longitude correction [hours], and D is the daylight saving time indicator (1 if active, 0 otherwise). The solar hour angle (ω) is derived from T_{sun}.

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

$$EOT = 9.87\sin(2B) - 7.53\cos(B) - 1.5\sin(B) \tag{2}$$

where B=360(N_{day}−81)/365 and N_{day} is the day of the year.

The Declination Angle (δ) is the angle between the rays of the sun and the plane of the Earth's equator:

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

The Solar Altitude Angle (α) is the angle between the sun's rays and the horizontal plane:

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

The Solar Azimuth Angle (γ_s) is the compass direction from which the sunlight is coming, 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. For a fixed panel with tilt angle β and azimuth angle γ_p.

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

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.

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

The beam and ground-reflected components are calculated geometrically:

$$G_{b,T} = G_b\frac{\cos\theta}{\cos\theta_z} \tag{8}$$

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

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

In this study we used the HAY model (E.Hay, 1979), an anisotropic diffuse sky model, for superior accuracy:

$$G_{d,T} = G_D[\frac{G_b}{G_{ext}}R_b + (1 - \frac{G_b}{G_{ext}})\frac{(1+\cos\beta)}{2}(1 + \sin^3\frac{\beta}{2})] \tag{10}$$

where G_D is the horizontal diffuse irradiance, G_{ext} is the extraterrestrial irradiance, and R_b=cosθ/cosθ_z is the geometric factor.

2.1.3 PV Module Temperature Model

The operating temperature of PV modules significantly impacts their conversion efficiency. In this study we adopted the MATTEI model (Keddouda et al., 2024), recognized for its accuracy across diverse installations. It models the module temperature (T_c) as a function of irradiance, ambient temperature, and wind speed:

$$T_C = \frac{HT_a + G_{POA}(\alpha\tau - \eta(1 + \beta T_{a,STC}))}{(H - (\beta\eta G_{POA}))} \tag{11}$$

where T_a is the ambient temperature [°C]; W_s is the wind speed [m/s]; $\alpha\tau$ is the absorptance-transmittance

product; η is the module efficiency; β is the power temperature coefficient of the module, H is a heat transfer coefficient where $H = 26.6 + 2.3W_s$ (Mattei et al., 2006); and subscript STC represents the Standard Test Conditions of 25 °C.

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 1D 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 the learning process is anchored by physical principles while effectively capturing complex residual patterns.

2.4 Model Evaluation Metrics

To ensure a comprehensive and standardized evaluation, two key metrics were used:

Root Mean Square Error (*RMSE*): A standard measure of the difference between values predicted by a model and the observed values.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(P_i^{pred} - P_i^{actual}\right)^2} \quad (12)$$

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

Forecasting Accuracy (F_A): As defined by (Press, 2021), this metric provides a normalized accuracy score:

$$F_A = (1 - \frac{RMSE}{C}) \times 100\% \quad (13)$$

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 the *RMSE*, only *RMSE* values are discussed during the case analysis.

3 Case Analysis

3.1 Data Description

The datasets were 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 MW to 8.67 MW. The data resolution varied from 5 min to 1 h, and the dataset lengths spanned from 3 to 27 months. All data were split in chronological order (70%/15%/15% for training/validation/testing). Detailed information is summarized in Table 1.

3.2 Correlation Analysis

Correlation analysis (Fig. 2) guided our feature and sequence length selection. In the Alice Springs dataset, G_H and G_{POA} showed strong correlations with power (P), unlike the weak correlation with Rain. This pattern generally held for the Sanming and Guangzhou datasets, though humidity (Hum) exerted a more significant influence at these sites.

Furthermore, lagged correlation analysis was 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) revealed 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 was selected (e.g., 4, 10, 15, ..., 106 data points) for subsequent experimental optimization. A similar strategy was applied to other regions and the 4-h forecasting task. These points included lags of 12, 24, 36, and 48 h (Fig. 3b). These four time points covered different temporal ranges from short-term intra-day patterns to long-term cross-day periodic information.

Table 1 Detailed information for each station

	Location and size	Category		Dataset Source	Data length	Resolution	Feature Parameters
1	Alice Springs 4.95 MW	M	√	(Dkasc, 2025)	12 Months	5 min	Wind Speed (W_s)
		F	×			-	Ambient Temperature

2	Sanming 2.77 MW	M	√	(Committee, 2024)	5.5 Months	15 min
		F	√	(Platform, 2025)		15 min
3	Guangzhou 8.67 MW	M	√	Skyworth Photovoltaic Power Station of Guangzhou Bureau	6.5 Months	1 h
		F	√			15 min
4	Canberra 1.00 MW	M	×	(Hong et al., 2016)	27 Months	-
		F	√			1 h

Note: 'M' and 'F' represent the Measured Data and Forecast Data, respectively, '√' and '×' indicate that the corresponding data/variable is available or unavailable.

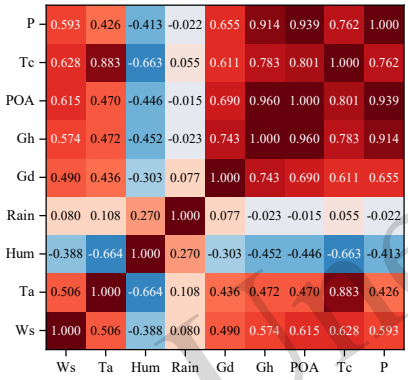
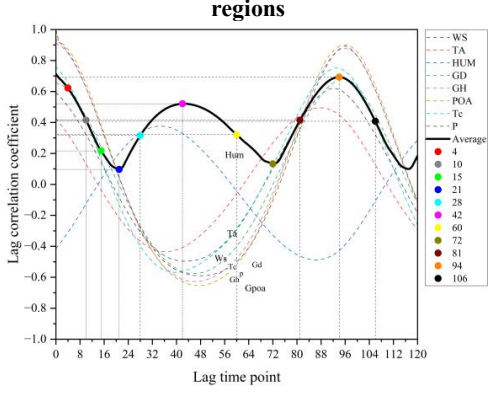
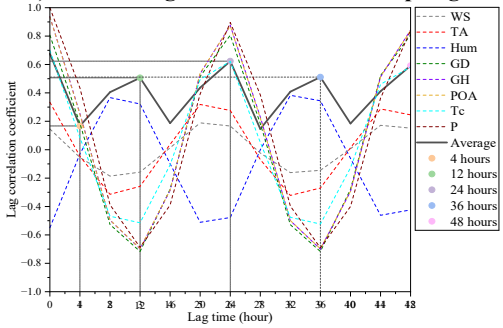


Fig. 2 Real-time correlation heatmap for the different regions



a). 15-min Lag Correlation of Alice Springs



b). 4-hour Interval Lag Correlation of Alice Spring

Fig. 3 15-minute interval lag correlation of the Alice Springs dataset

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 was adaptively scaled. Specifically, the number of LSTM hidden units was increased from 64 to 144 to capture broader temporal dependencies, while the Dropout ratio was raised from 0 to 0.4 to prevent overfitting. Concurrently, CNN output channels were optimized to maintain computational efficiency while reducing feature redundancy.

Table 2 CNN-LSTM neural network simulation parameters

Parameter Name	15-minute	4-hour	24-hour
Convolutional Layer and Kernel	1, 3	1, 3	1, 3
Pooling Layer and Kernel	1, 3	1, 3	1, 3
Output Unit	32	16	8
Number of LSTM Layers	1	1	1
LSTM Hidden Unit	64	128	144
Dropout	0	0.2	0.4
Rate		0.001	
Iterations		100	
Optimizer		Adam	

3.4 Various Forecasting Horizons Forecasting Experiment

A controlled experiment was designed to evaluate the PIDL-HM's effectiveness and optimize feature engineering. Two case groups were 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 physical-ground 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 were applied consistently across all forecasting horizons and regions.

3.4.1 15-minute forecasting experiment

Table 4 presents the *RMSE* metrics across various input sequence lengths and feature combinations for the 15-min task. Guided by lagged

correlation analysis, optimal sequence lengths (lag 94 for Alice Springs; lag 95 for Sanming) were identified to balance historical context with noise, effectively capturing diurnal patterns.

Table 3 Case settings for Alice Springs

Cases	Description	W_s	T_a	Hum	Rain	G_D	G_H	G_{POA}	T_C
Case1-1	All available meteorological features	√	√	√	√	√	√		
Case1-2	Moderately correlated features	√	√	√		√	√		
Case1-3	Highly correlated features	√				√	√		
Case2-1	Full meteorological & physically-grounded features	√	√	√	√	√	√	√	√
Case2-2	M-correlated & physically-grounded features	√	√	√		√	√	√	√
Case2-3	H-correlated & physically-grounded features	√				√	√	√	√
Case2-4	Pure physically-grounded features							√	√
Case2-5	Single-information features			√				√	√
Case2-6	Overlapping-information features		√				√	√	√

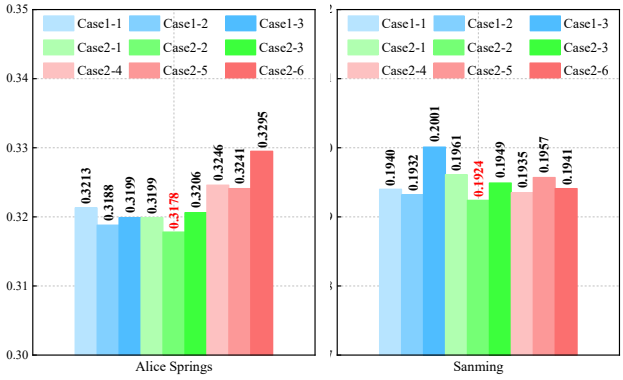
Note: ' √ ' indicates that the feature is used

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

Input Sequence Length (data points)	Case1-1	Case1-2	Case1-3	Case2-1	Case2-2	Case2-3	Case2-4	Case2-5	Case2-6
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

Note: *RMSE*s are in [MW]. Best and second-best results are bold-underlined and bold, respectively.

As detailed in Table 4, integrating physical features (G_{POA} , T_C) consistently enhanced prediction precision. The hybrid Case 2-2, which synergizes these physical priors with meteorological variables (e.g., Ws, G_H), achieved the optimal performance, outperforming both purely data-driven and standalone physical benchmarks. Notably, this configuration reduced *RMSE* by 1.09% compared to the all-feature baseline (Case 1-1). This superiority persisted across both the Alice Springs and Sanming datasets, confirming the cross-regional reliability of the PIDL-HM for ultra-short-term forecasting (Fig. 4).



a). Alice Springs **b). Sanming**

Fig.4 15-minute *RMSE* results of Alice Springs (at lag 94) and Sanming (at lag 95)

3.4.2 4-hour forecasting Experiment

For the 4-h forecast, *RMSE* values naturally increased across all three regions (Alice Springs, Sanming, Guangzhou) due to greater inherent

uncertainty (Tables 5-7). The optimal input sequence length consistently proved to be 24 h, a full diurnal cycle, as suggested by the ALC analysis. At this optimal length, the PIDL-HM's advantage remained significant.

Case 2-6 emerged as the optimal configuration for the Alice Springs and Guangzhou datasets, reducing the *RMSE* 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 proved region-dependent. In Sanming, high meteorological volatility rendered standalone physical features (Case 2-4) suboptimal due to error accumulation from noisy inputs. Conversely, the stable climates of Alice Springs and Guangzhou yielded smoother feature sequences, enabling the hybrid Case 2-6 to effectively capture system dynamics.

3.4.3 24-hour forecasting experiment

Note that for the 24-h forecast task, lagged correlation analysis was omitted as it transitioned 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 confirmed that integrating physical features consistently enhances accuracy, with the optimal hybrid configuration (Case 2-6) in Guangzhou achieving a reduction in *RMSE* 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.

3.5 Data Efficiency Experiment

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 conducted experiments 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 demonstrated a decisive advantage. While the optimal feature combination varied, the best-performing model in every scenario was invariably a PIDL-HM variant (Tables 9-11). This advantage magnified with system scale and forecast horizon: for Guangzhou's 24-h forecast, the optimal PIDL-HM reduced the *RMSE* by 17.8% compared to the pure data-driven baseline.

Table 5 Evaluation metrics for the 4-hour forecasting task in Alice Springs

Input Sequence Length (hours)	Case1-1	Case1-2	Case1-3	Case2-1	Case2-2	Case2-3	Case2-4	Case2-5	Case2-6
4	0.5116	0.5161	**0.5056**	**0.4952**	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	**0.4507**
24	0.4788	0.4536	0.4607	0.4547	0.4632	0.4486	**0.4476**	0.4521	**0.4382**
36	0.4903	0.4745	0.4798	0.4600	0.5187	**0.4576**	0.4761	0.4583	**0.4573**
48	0.4864	**0.4508**	0.4925	0.4860	0.4704	**0.4569**	0.4889	0.4810	0.4723

Table 6 Evaluation metrics for the 4-hour forecasting task in Sanming

Input Sequence Length (hours)	Case1-1	Case1-2	Case1-3	Case2-1	Case2-2	Case2-3	Case2-4	Case2-5	Case2-6
4	0.2698	0.2632	0.2629	**0.2564**	**0.2572**	0.2580	0.2860	0.2999	0.2649

12	0.2825	0.2616	0.2661	**0.2574**	**0.2602**	0.2613	0.2988	0.2914	0.2680
24	0.2614	0.2565	0.2551	**0.2483**	0.2594	**0.2419**	0.2962	0.2760	0.2624
36	0.2655	0.2902	0.2752	0.2606	0.2601	**0.2540**	0.2759	0.2854	**0.2594**
48	0.2870	0.2658	0.2760	0.2706	0.2697	**0.2618**	0.3006	0.2895	**0.2639**

Table 7 Evaluation metrics for the 4-hour forecasting task in Guangzhou

Input Sequence Length (hours)	Case1-1	Case1-2	Case1-3	Case2-1	Case2-2	Case2-3	Case2-4	Case2-5	Case2-6
4	0.4917	0.4851	0.4987	0.4854	**0.4771**	**0.4782**	0.5075	0.5065	0.5144
12	0.4949	**0.4878**	0.5273	0.4977	**0.4812**	0.4944	0.5039	0.5017	0.4886
24	0.4668	0.4586	0.4668	0.4619	0.4406	0.4554	**0.437**	0.4373	**0.4344**
36	0.4808	0.4527	0.4846	0.4754	0.4554	0.4724	0.4603	**0.4508**	**0.4478**
48	0.4849	0.4577	0.4814	0.4725	0.4582	0.4716	**0.4557**	0.4642	**0.4518**

Table 8 Day-ahead (24-hour) forecasting evaluation metrics for Guangzhou, Canberra, and Sanming

Regions	Case1-1	Case1-2	Case1-3	Case2-1	Case2-2	Case2-3	Case2-4	Case2-5	Case2-6
Guangzhou	0.6851	**0.6347**	0.6639	0.6586	0.6447	0.6561	0.6435	0.6546	**0.6239**
Canberra	**0.1163**	0.1170	0.1172	**0.1162**	0.1168	0.1165	0.1211	0.1169	0.1218
Sanming	0.3353	0.3254	0.3429	0.3289	0.3304	**0.3079**	**0.2983**	0.3160	0.3083

Table 9 15-minute small-batch data efficiency experiment

Regions	Case1-1	Case1-2	Case1-3	Case2-1	Case2-2	Case2-3	Case2-4	Case2-5	Case2-6
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

Table 10 4-hour small-batch data efficiency experiment

Regions	Case1-1	Case1-2	Case1-3	Case2-1	Case2-2	Case2-3	Case2-4	Case2-5	Case2-6
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**

Table 11 24-hour small-batch data efficiency experiment

Regions	Case1-1	Case1-2	Case1-3	Case2-1	Case2-2	Case2-3	Case2-4	Case2-5	Case2-6
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**

Note: *RMSE* are in [MW].

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

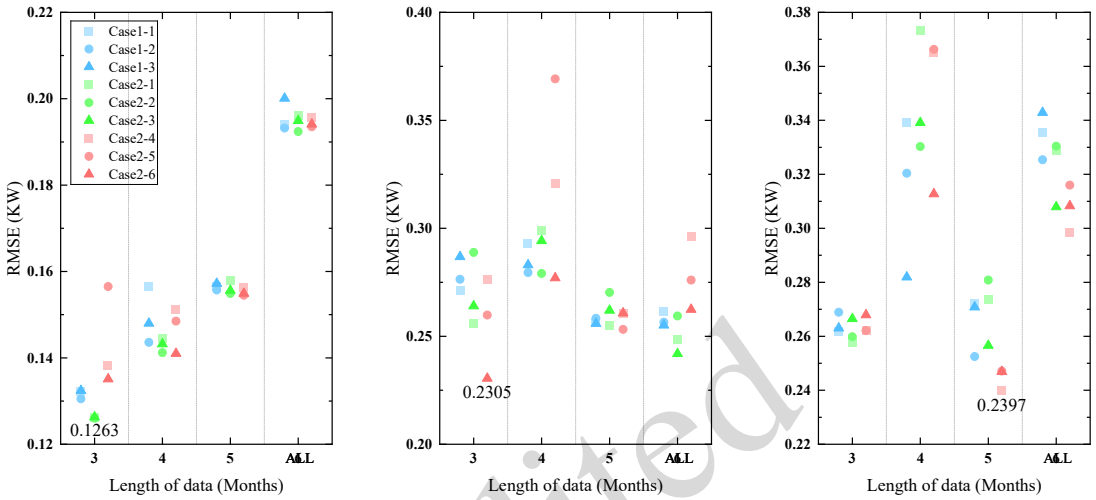
To evaluate robustness under seasonal distribution shifts, we used two experimental designs:

Ex1 (Chronological Test): The training set's start date was fixed, and its length was varied. The test set immediately followed the training period.

Ex2 (Fixed Test): The test set was fixed on a specific period (e.g., summer). The training set's length was varied by extending its start date further into the past, thus incorporating different seasons.

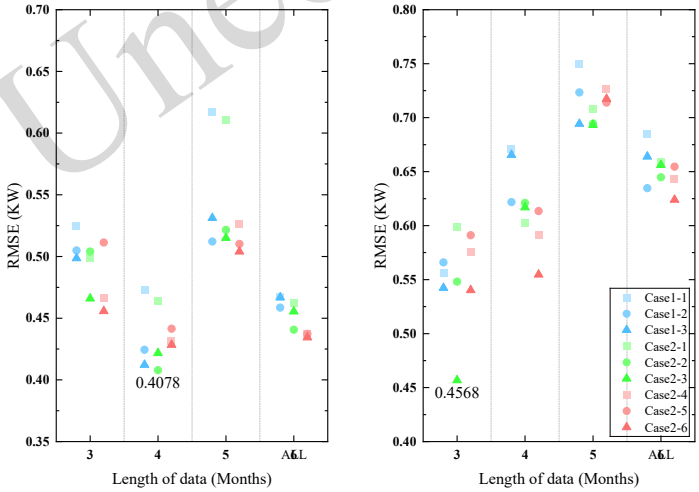
For stations with shorter datasets (Sanming,

Guangzhou), performance in Ex1 fluctuated significantly due to the training and test sets covering different, biased seasonal segments (Figs. 5, 6). Despite this inherent volatility, the PIDL-HM achieved 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.



a) 15 min **b) 4 hour** **c) 24 hour**

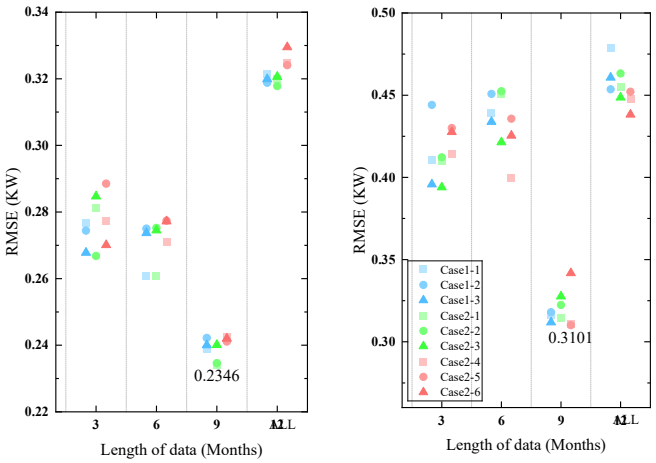
Fig. 5 Data efficiency experiment: Sanming (started 17 Nov 2022)



a) 4hour **b) 24hour**

Fig. 6 Data efficiency experiment: Guangzhou (started 1 Jan 2025)

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



a) 15 min **b) 4 hour**

Fig. 7 Data efficiency experiment: Alice Springs (started at 1 Jan 2015)

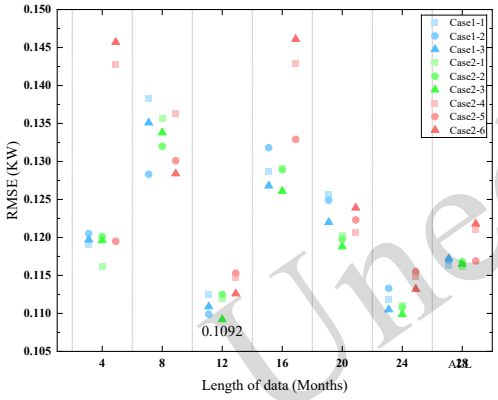


Fig. 8 *RMSE* result of 24-hour data efficiency experiment: Canberra (started 1 Jan 2014)

For the Guangzhou region's 24-h forecasting task, Ex2 was implemented, fixing the test set (Aug 18 to Sep 18) and setting up five cases where the training data length gradually increased, with start dates from May 1st to January 1st. The performance improvement of the PIDL-HM over the pure data-driven model for different cases is shown in Table 12.

Table 12 provides compelling quantitative evidence of the PIDL-HM's value under distribution shift. The performance gain of the PIDL-HM over the best pure data-driven model monotonically increased from 1.05% to a peak of 6.87% as the seasonal mismatch between training and test data increased (Ex2-4, spanning winter to summer). This trajectory confirms that the more pronounced the distribution shift, 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 integrated it into an ensemble framework. For key scenarios, we fused predictions from the baseline pure data-driven model and PIDL-HM models with those from an XGBoost model via simple averaging (Table 13).

The fusion strategy showed significant efficacy; for instance, in Sanming's 4-h task, integrating the data-driven baseline with XGBoost reduced *RMSE* by 7.2%. Notably, replacing the baseline with PIDL-HM as the base component yielded a further 3.8% *RMSE* 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:

(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.

Table 12 Performance improvement of hybrid-driven model under different cases

	Case1-1	Case1-2	Case1-3	Case2-1	Case2-2	Case2-3	Case2-4	Case2-5	Case2-6	Gain
Ex2-1	0.5047	0.5026	0.5023	0.5069	0.514	0.504	**0.5023**	**0.4994**	0.5158	1.05%
Ex2-2	0.519	0.5003	0.5185	0.5193	0.5049	0.5336	**0.4952**	**0.4972**	0.5252	4.58%
Ex2-3	0.5274	0.5029	0.5226	0.527	**0.4987**	0.5094	**0.4989**	0.502	0.5266	5.44%
Ex2-4	0.5314	0.5155	0.5143	0.5317	0.5208	0.5207	**0.4949**	**0.4966**	0.5004	6.87%

Ex2-5	0.5263	0.5146	0.5369	0.5377	0.5146	**0.4924**	**0.5093**	0.5112	0.5178	6.44%

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

Case Description	Pure data-driven version		PIDL-HM	
	Original	+ XGBoost	Original	+ 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**

(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.

4.1 Limitations and Future Work

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 was supported by the National Natural Science Foundation of China (52306013), Guangdong Basic and Applied Basic Research Foundation (2025A1515012288), and the Fundamental Research Funds for the Central Universities (21624212).

Author contributions

Chang Huang designed the research, wrote the first draft of the manuscript and revised and edited the final version. Xuanbin Huang, Jinmin Guo and Zhuo Cao processed the corresponding data. Ting He and Wentao Shang helped to organize the manuscript.

Conflict of interest

No potential conflict of interest was reported by the authors.

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

Committee D. 2024. Distributed pv power forecasting—track 3 dataset. Available from https://www.dcic-china.com/competition/2024/track3

Das U, Tey K, Seyedmahmoudian M, et al., 2017. Svr-based model to forecast pv power generation under different weather conditions. *Energies*, 10(7) https://doi.org/10.3390/en10070876

Dkasc. 2025. Notes on the data. Available from https://dkasolarcentre.com.au/download/notes-on-the-data

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

Hong T, Pinson P, Fan S, et al., 2016. Probabilistic energy forecasting: Global energy forecasting competition 2014

and beyond. *International Journal of Forecasting*, 32(3):896-913. https://doi.org/10.1016/j.ijforecast.2016.02.001

Iea, 2024. Renewables 2024: Analysis and forecast to 2030. Technical Report,

Keddouda A, Ihaddadene R, Boukhari A, et al., 2024. Photovoltaic module temperature prediction using various machine learning algorithms: Performance evaluation. *Applied Energy*, 363 https://doi.org/10.1016/j.apenergy.2024.123064

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

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

Niccolai A, Dolara A, Ogliari E, 2021. Hybrid pv power forecasting methods: A comparison of different approaches. *Energies*, 14(2) 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

Platform XE. 2025. Meteorological forecasting dataset for fujian, china. Available from https://xihe-energy.com/

Press CS, 2021. Technical requirements for dispatch-side wind power or photovoltaic power prediction systems, GB/T 40607—2021. Beijing.

Rao Z, Yang Z, Yang X, et al., 2024. Tcn-gru based on attention mechanism for solar irradiance prediction. *Energies*, 17(22) https://doi.org/10.3390/en17225767

Sun F, Li L, Bian D, et al., 2025. Photovoltaic power prediction based on multi-scale photovoltaic power fluctuation characteristics and multi-channel lstm prediction models. *Renewable Energy*, 246 https://doi.org/10.1016/j.renene.2025.122866

Zhu C, Tu Y, Wei Q, et al., 2025. Short-term power prediction of photovoltaic power station based on lstm-xgboost model. *Solar Energy*, 300 https://doi.org/10.1016/j.solener.2025.113819

Zhu J, Li M, 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

中文概要

题　目： 基于物理信息的深度学习实现数据高效且稳健的光伏发电功率预测

作　者： 黄畅[1]，黄炫彬[1]，郭锦敏[2]，曹倬[1]，何婷[1,*]，尚文涛[1,*]

机　构：[1]暨南大学,能源电力研究中心,中国珠海,519070；[2]北京清能互联科技有限公司，中国北京，510630；

目　的： 光伏发电功率具有极强的波动性和不确定性。现有的纯数据驱动预测模型在缺乏历史数据（小样本）或面临复杂多变气象条件（分布漂移）时，往往泛化能力较差。本文旨在提出一种融合物理机理与深度学习的混合预测框架（PIDL-HM），通过引入明确的物理约束，解决新建电站数据匮乏及非平稳环境下的预测难题，提高光伏功率预测的准确性、数据效率与鲁棒性。

创新点： 1. 提出了一种物理信息深度学习混合模型（PIDL-HM），通过将“白箱”物理建模与“黑箱”数据驱动挖掘深度融合，建立了一种超越简单误差修正的原则性集成框架。2.构建了基于物理模型链的特征工程方法，将数值天气预报（NWP）数据转化为具有明确物理意义的中间变量（如阵列面辐照度 G_{POA} 和组件温度 T_c），为神经网络提供了稳定的物理链。3.实现了在极小样本数据（仅需 3 个月）下的高精度预测，突破了传统深度学习模型对海量历史数据的依赖，显著降低了新建光伏电站的投运门槛。

方　法： 1. 基于天文学几何关系和热力学传热方程，构光伏物理转换模型链，计算关键物理特征。2. 将生成的物理特征与气象数据共同输入卷积神经网络-长短期记忆网络（CNN-LSTM），利用 CNN 提取局部特征，LSTM 捕捉长时序依赖。3.利用中国（广州、三明）和澳大利亚（爱丽丝泉、堪培拉）等多个不同气候区的光伏电站实测数据，设计 15 分钟、4 小时及 24 小时等多时间尺度的预测实验进行验证。

结　论： PIDL-HM 在多时间尺度预测任务中均展现出优于纯数据驱动模型的综合性能，特别是在日前预测中 *RMSE* 降低了约 8.93%。更关键的是，得益于物理机理的有效引入，该模型展现出卓越的数据效率与鲁棒性，不仅在仅有 3 个月历史数据的极端小样本条件下维持了高精度预测，有效解决了新建电站冷启动难题，而且在面临跨季节数据分布漂移的非平稳环境下仍能保持稳定，相比基准模型性能提升达 6.87%，为复杂气象条件下的光伏功率预测提供了一种可靠且具有物理可解释

性的解决方案。

关键词：光伏发电预测；基于物理的深度学习；多种预测时限；基于物理特征；数据效率