Day-ahead Forecasting of PV Power Based on TCN and Temporal Query Network Variant

Hongyi Peng
State Key Laboratory of Advanced Electromagnetic Technology
Huazhong University of Science and Technology
Wuhan, China
penghy0403@hust.edu.cn

Mingyu Yan*
State Key Laboratory of Advanced Electromagnetic Technology
Huazhong University of Science and Technology
Wuhan, China
mingyuyan@hust.edu.cn

Yize Liu
State Key Laboratory of Advanced Electromagnetic Technology
Huazhong University of Science and Technology
Wuhan, China
yizeliu@hust.edu.cn

Yijia Zhou
State Key Laboratory of Advanced Electromagnetic Technology
Huazhong University of Science and Technology
Wuhan, China
zhouyijia@hust.edu.cn

***Abstract*—In this paper, a method for day-ahead photovoltaic (PV) power forecasting is proposed. The method integrates new feature construction, a neural network architecture, and output data constraints. The neural network includes a Temporal Convolutional (TCN), Temporal Query (TQ) module, Multi-head Attention, Multi-Layer Perceptron (MLP), Reversible Instance Normalization (ReVIN), and projection layers. Experiments on the PVOD dataset demonstrate that introducing a physically meaningful feature dimension effectively constrains nighttime outputs and enhances forecasting accuracy. The proposed method outperforms 5 baseline models (BiLSTM, SegRNN, PatchTST, DLinear, and Transformer) in the PVOD dataset with RMSE = 0.3652, MAE = 0.1736, and R^2 = 0.9024. The proposed method uses the inherent periodic characteristics of PV power, employs a Multi-Head Attention mechanism to enhance the learning of important periodic information, and applies physical constraints to correct nighttime forecasting results, demonstrating excellent performance in day-ahead PV power forecasting.**

Keywords*—*PV Power Forecasting, Day-ahead Forecasting, TCN, TQ Network, Multi-Head Attention

I. Introduction

The global climate crisis is worsening. Replacing fossil fuels with renewable energy to reduce carbon emissions has become a matter of urgency. In recent years, photovoltaic (PV) power has developed rapidly and has emerged as a highly competitive approach to electricity generation [1]. However, PV generation is highly sensitive to meteorological variability, posing challenges for its effective utilization. Accurate day-ahead PV forecasts enhance the benefits of day-ahead grid dispatch [2].

Common PV power forecasting approaches fall into two types: physical models and data-driven models [3]. Physical modeling requires extensive design data from PV plants, involves a complex modeling process, and typically yields lower accuracy [4]. By contrast, data-driven models require less data and can learn latent patterns of PV power. Therefore, data-driven methods are currently widely used for PV forecasting [3].

Numerous time-series forecasting models have been applied to PV forecasting [5]. Recurrent Neural Networks (RNNs) and their variants, such as Long Short-Term Memory (LSTM) and Bidirectional LSTM (BiLSTM), have been widely used to capture temporal dependencies [6]. However, these recurrent architectures often struggle with long-term dependency modeling due to gradient vanishing problems. To address this, Transformer-based models (e.g., standard Transformer [7] and PatchTST [8]) utilizing self-attention mechanisms have emerged, offering superior capabilities in capturing global correlations. Additionally, recent linear models like DLinear and segmentation-based models like SegRNN have shown promising results in general time-series forecasting [9], [10]. However, these models do not explicitly represent periodic patterns as learnable variables, thereby limiting their capacity to learn periodicity. In addition, model learning is not perfectly accurate, which leads to nighttime PV power forecasts often deviating from reality [11]. To address such physical inconsistencies, Physics-Informed Neural Networks (PINNs) have been introduced; however, they typically impose physical laws as soft constraints via penalty terms, which cannot strictly guarantee the satisfaction of physical limits [12]. Reference [13] proposed a time-series forecasting method based on temporal periodic queries. However, it does not further explore model improvements tailored to the PV forecasting task.

The main contributions of this paper are as follows:

1) A day-ahead PV power forecasting method based on the inherent periodic characteristics is proposed. A Temporal Query (TQ) module is introduced as a query, using learnable cyclic parameters to explicitly model the periodic patterns of the PV data.

2) A Multi-Head Attention mechanism is introduced to the TQ module to model global and robust correlations in multivariate data. This mechanism prevents the model from focusing solely on correlations within a local sample window.

3) Physical constraints are proposed to correct nighttime forecasting results. It avoids the unphysical nighttime forecasting in direct forecasting and achieves better forecasting performance.

The remainder of this paper is organized as follows. Section II presents the methodology of the proposed method. Section III introduces the experimental settings. Section IV presents the results and discussion of the experiments. Section V concludes the paper.

II. Methodology

In this section, we introduce a new feature and present the network components of the proposed method. Finally, the physical constraints on the forecast outputs are discussed.

This work was supported by the National Key Research and Development Program of China (2023YFB4005601).

979-8-3315-8138-1

A. New Feature Construction

The solar zenith angle has a strong influence on irradiance, particularly in clear-sky conditions. At night, when the angle is $\geqslant 90^{\circ}$, irradiance and PV generation are both zero. During the day, the angle is $<90^{\circ}$ and decreases toward noon. Under ideal clear-sky conditions, higher irradiance associated with smaller zenith angles leads to greater PV output, as shown in Fig. 1. Therefore, we added the solar zenith angle to the features of the training data. The solar zenith angle is determined using the solar position algorithm, based on the latitude and longitude of the PV station:

$$Z = SPA\left(t, loc\left(lat, lon\right)\right) \tag{1}$$

where Z represents the solar zenith angle at the latitude and longitude coordinates *(lat,lon)* at time *t*.

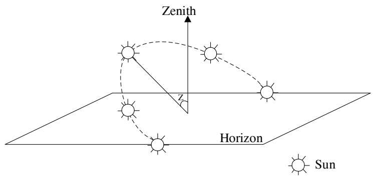

Fig. 1 Diagram of the Solar Zenith Angle

B. Components of Network

Temporal Convolutional Module

Temporal Convolutional Networks (TCN) are used as a Temporal Convolutional Module. TCN has multiple advantages, such as long-term dependency modeling and multi-scale information extraction. The architecture integrates causal convolution, dilated convolution, and residual connections.

TCN employs causal convolution to ensure the model's compliance with the temporal order. In causal convolution, only the present and preceding inputs contribute to the computation, preserving the temporal direction of information propagation and avoiding any influence from future data. This is commonly implemented by padding zeros on the right side of the kernel (referred to as causal padding), so that the output depends solely on information available up to the current time step. The mathematical representation of causal convolution is:

$$y\left(t\right) = \sum_{i=0}^{k-1} f\left(i\right) \cdot x\left(t-i\right) \tag{2}$$

where f represents the convolution kernel, k represents the size of the convolution kernel, x represents the input sequence.

To expand the receptive field without introducing additional parameters, TCN employs a dilated convolution mechanism. In this mechanism, gaps are inserted between the convolution kernels (i.e., skips certain input units). It enables the model to capture a broader receptive field without introducing additional parameters, thereby improving its capacity to model the context information in the input sequence. The mathematical representation of dilated convolution is:

$$y\left(t\right) = \sum_{i=0}^{k-1} f\left(i\right) \cdot x\left(t-d\cdot i\right) \tag{3}$$

where d represents the expansion rate.

TCN integrates residual connections to address the vanishing gradient problem and support the training of deeper network structures. In this framework, the output of an intermediate layer is directly added to that of a subsequent layer, forming a skip pathway. This design ensures a smoother flow of gradients during backpropagation, thereby reducing gradient attenuation and enhancing the network's ability to efficiently learn from deeper hierarchies. The output of a residual block can be expressed as:

$$Y_{tcn} = activation\left(X_t + F\left(X_t\right)\right) \tag{4}$$

where F represents a combination of a convolutional layer and an activation function.

In the proposed model, the input and output sequence lengths of the TCN module are consistent. The structure of the TCN module is shown in Fig. 2. More detailed information can be found in [14].

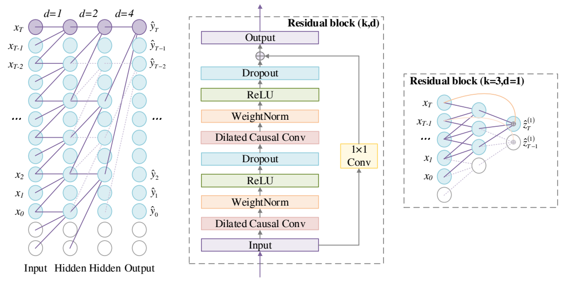

Fig. 2 The Structure of the TCN Module

Temporal Query Module

In the task of PV power output day-ahead forecasting, the sequentiality and periodicity of time are extremely crucial. However, the attention mechanism usually only implicitly incorporates temporal information through position encoding, focusing more on the numerical features of time steps and having limited ability to model periodicity. Therefore, a TQ module is introduced as the Query before applying the attention mechanism.

A period length W is defined to specify both the length of the learnable vector and the interval of the periodic offset. The learnable parameter matrix $\theta_{TQ} \in \mathbb{R}^{C\times W}$ is initialized to 0. During the training process, this matrix is dynamically optimized to adaptively model the stable dependency relationships among variables.

The sequence processed by the TCN module is taken as the input sample, and a vector of length L is periodically extracted from θ_{TQ} based on the sample position index t mod W to serve as the Query $Q = \theta_{TQ}^{t,L} \in \mathbb{R}^{C\times L}$ in the attention mechanism, where W is the period length, as shown in Fig. 3.

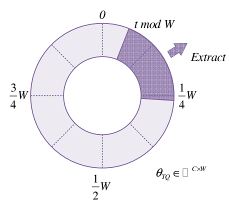

Fig. 3 TQ Module

Multi-Head Attention Module

As mentioned above, in the multi-head attention mechanism, the Query Q_h adopts the periodic query vector constructed by TQ. The Key K_h and Value V_h are input with the original sequences processed by the TCN module. These components interact via standard multi-head attention to capture periodic dependencies. This periodic mechanism enables samples separated by W steps to have the same query vector, thereby achieving parameter sharing and stable modeling. The multi-head attention mechanism is calculated as follows:

$$Head_h = Softmax\left(\frac{Q_h K_h^T}{\sqrt{L}}\right)V_h \tag{5}$$

The output of multi-head attention is concatenated and mapped to a unified representation:

$$MHA(Q,K,V) = Concat(head_1,\ldots,head_H)W^O \tag{6}$$

where W^O represents the output projection matrix of the multi-head attention module. The structure of the multi-head attention module is shown in Fig. 4.

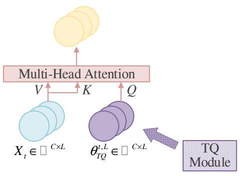

Fig. 4 Multi-Head Attention Module

Multi-Layer Perceptron Module

After the attention module, a lightweight Multi-Layer Perceptron (MLP) Module is used to further model the temporal dependencies. The MLP consists of two fully connected layers, each followed by a GeLU activation function. Its structure is as follows:

$$Y_{mlp} = GeLU\left(Linear\left(GeLU\left(Linear\left(Y_{attn}\right)\right)\right)\right) \tag{7}$$

where Y_{attn} represents the output of the multi-head attention module.

Projection and Normalization

In addition to the above-mentioned main modules, the proposed model also involves dimension projection and normalization processing.

The output of the multi-head attention is added to the original sequence via a residual connection. The resulting representation is then linearly mapped to a unified hidden space through a linear layer with nonlinear activation, serving as input to the MLP module:

$$Y_{attn} = GeLU\left(Linear\left(Y_{tcn} + Y'_{attn}\right)\right) \tag{8}$$

where Y'_{attn} represents the output of the multi-head attention.

Similarly, after the MLP module, its output is fused with the original sequence. The fused representation is then projected to the target forecast length space through a linear layer with dropout:

$$Y_t' = Linear\left(Dropout\left(Y_{attn} + Y_{mlp}\right)\right) \tag{9}$$

Each input sequence is normalized using its own mean and variance instead of the information from the global dataset. This normalization step alleviates the distribution shift across samples. The output is then denormalized to restore to the original scale, which constitutes the Reversible Instance Normalization (ReVIN) process:

$$X_t = \frac{X_t - \mu}{\sqrt{\sigma^2 + \varepsilon}} \tag{10}$$

$$Y_t = Y_t' \times \sqrt{\sigma^2 + \varepsilon} + \mu \tag{11}$$

where μ and σ^2 represent the mean and variance of the input sequence data, respectively. ε is a small constant.

C. *Output Data Constraints*

Previously, we have calculated the zenith angle and input it as training data into the model. However, the output obtained through training still fails to accurately reflect the actual characteristics of PV power generation. At night, the predicted output shows oscillations and drifts, whereas theoretically, the PV output must be zero.

Statistics show that when the zenith angle $Z \geq 90^\circ$, 99.63% of samples have a PV output < 10^{-4} (and 99.63% are < 1). Similarly, for $Z \geq 86^\circ$, 96.16% of samples are < 10^{-4}, while 99.63% remain < 1. Considering that oscillations near zero in prediction results initiate while $Z < 90^\circ$ rather than strictly after 90°. Therefore, we impose physical constraints on the model output using the zenith angle:

$$output = \begin{cases} Y_t, Z < 86^\circ \\ 0, Z \geq 86^\circ \end{cases} \tag{12}$$

where Y_t is the output of the aforementioned proposed model.

The process of the proposed method is shown in Fig. 5.

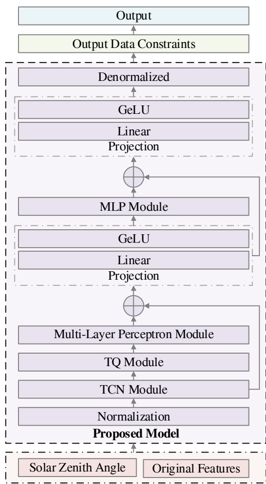

Fig. 5 The Process of the Proposed Method

III. EXPERIMENTS

A. *Experimental Settings*

A photovoltaic power output dataset (PVOD) [15] is applied to test the proposed model. The data in PVOD is at

15-minute intervals, including Numerical Weather Prediction and Local Measurements Data. We selected the data from Station 01, whose dataset ranges from June 30, 2018 to June 13, 2019, and it is the station with the longest data collection period in the PVOD dataset. We use 12 features from the original data and 1 newly constructed feature for training, as shown in TABLE I. The forecasting experiment is conducted as a day-ahead forecast, forecasting the PV power output for the next 24 hours, i.e., 96 steps. The input look-back window size was set to 384 time steps, corresponding to historical data from the past 4 days. In the training and validation sets, we use a sliding window to select the forecast time. In the test set, we issue day-ahead PV power forecasts at 00:00 each day. The forecast experiment was conducted on an RTX 4090 GPU. Mean squared error (MSE) is used as the loss function with the Adam optimizer with an initial learning rate of 2.75×10^{-4}. The training process spanned 500 epochs with a batch size of 64, employing an early stopping strategy with a patience of 5 epochs to prevent overfitting.

TABLE I. TRAINING FEATURES OF THE EXPERIMENT

Numerical Weather Prediction	Local Measurements Data	Constructed feature
Global irradiation	Total irradiation	Solar zenith angle
Direct irradiation	Diffuse irradiation	
Temperature	Temperature	
Humidity	Pressure	
Windspeed	Windspeed	
Pressure	PV Power	

B. Data Preprocessing

The data of each feature is standardized as follows:

$$x'_{i,j} = \frac{x_{i,j} - \mu_j}{\sigma_j} \tag{13}$$

where $x_{i,j}$ represents the original value of the i-th sample of the j-th feature. μ_j represents the mean of the j-th feature. σ_j represents the standard deviation of the j-th feature.

C. Evaluation Metrics

Three evaluation metrics are used to measure the performance of the model: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination (R^2). The calculation of these metrics follows the standard approach and thus is not further elaborated in this paper for the sake of brevity.

IV. RESULTS AND DISCUSSION

A. Model Training

We trained the proposed and 5 baseline models, and recorded the MSE losses on the training and validation sets during training, as shown in Fig. 6. During the training of BiLSTM and PatchTST, the MSE loss on the training set decreases with increasing epochs, whereas the MSE loss on the validation set continuously rises. Although Transformer achieves a lower training loss than the proposed method, it performs worse on the validation set. In addition, both DLinear and SegRNN exhibit poorer results on the training and validation sets compared with the proposed model. Overall, the proposed method achieves the best convergence and lowest validation loss, confirming its effectiveness.

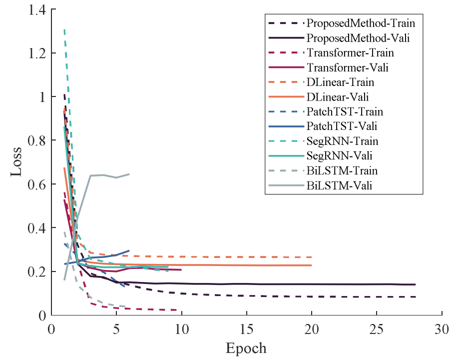

Fig. 6 Train vs Validation Loss Curves of Different Models

B. Model Comparison

We compared the evaluation metrics of the output data before and after applying the constraints described in Section II-C and the proposed model against 5 baseline models, as shown in TABLE III. Without physical constraints, the forecasting of nighttime PV output fluctuates around zero, which is inconsistent with real conditions. The introduction of constraints helps correct nighttime forecasting bias and thereby improves model performance. The inferior performance of the baseline models aligns with the limitations discussed in Section I. The proposed method takes into account the inherent periodic characteristics of the PV output forecasting task and introduces a multi-head attention mechanism. Therefore, the model enables training better using the daily cycle and key time nodes. The proposed model achieves better forecasting performance than BiLSTM and other comparative models (RMSE = 0.3652, MAE = 0.1736, R^2 = 0.9024). As shown in Fig. 7, the forecasting results for 3 selected days are visualized.

C. Ablation Experiments

We conducted ablation experiments on the proposed model, as shown in TABLE II. The TCN module extracts fundamental local trends and short-term meteorological dependencies, serving as the backbone of the feature extractor. Removing the TCN module results in a significant drop in performance. The TQ module and Multi-Head Attention explicitly encode daily periodic patterns, enabling the model to retrieve correlations from the same time of day across historical data, which effectively compensates for the limited receptive field of standard convolutions. The ReVIN module mitigates the impact of non-stationary distribution shifts common in PV data, ensuring training stability. The MLP module further refines the high-dimensional features for the final output. The integration of these components allows the proposed method to capture both local fluctuations and global periodic dependencies. The proposed method, which integrates all these components, achieved the optimal forecasting performance across all metrics (RMSE = 0.3652, MAE = 0.1736, R^2 = 0.9024).

TABLE II. THE RESULTS OF THE ABLATION EXPERIMENTS

	RMSE	MAE	R^2
Without TQ	0.3721	0.1760	0.8986
Without Attention	0.3834	0.1859	0.8924
Without TCN	0.4637	0.2373	0.8426
Without MLP	0.3681	0.1857	0.9008
Without ReVIN	0.3773	0.1930	0.8958
TCN only	0.4050	0.2103	0.8799
Proposed method	**0.3652**	**0.1736**	**0.9024**

TABLE III. COMPARISON OF THE EVALUATION METRICS OF DIFFERENT MODELS

		BiLSTM	SegRNN	PatchTST	DLinear	Transformer	Proposed Method
Before Constraints	**RMSE**	0.4603	0.5651	0.5171	0.5244	0.4193	0.3693
	MAE	0.3336	0.3433	0.3155	0.3207	0.2354	0.2027
	R^2	0.8449	0.7662	0.8042	0.7986	0.8713	0.9001
After Constraints	**RMSE**	0.4361	0.5624	0.5100	0.5161	0.4152	**0.3652**
	MAE	0.2398	0.3134	0.2678	0.2675	0.2151	**0.1736**
	R^2	0.8608	0.7684	0.8096	0.8050	0.8738	**0.9024**

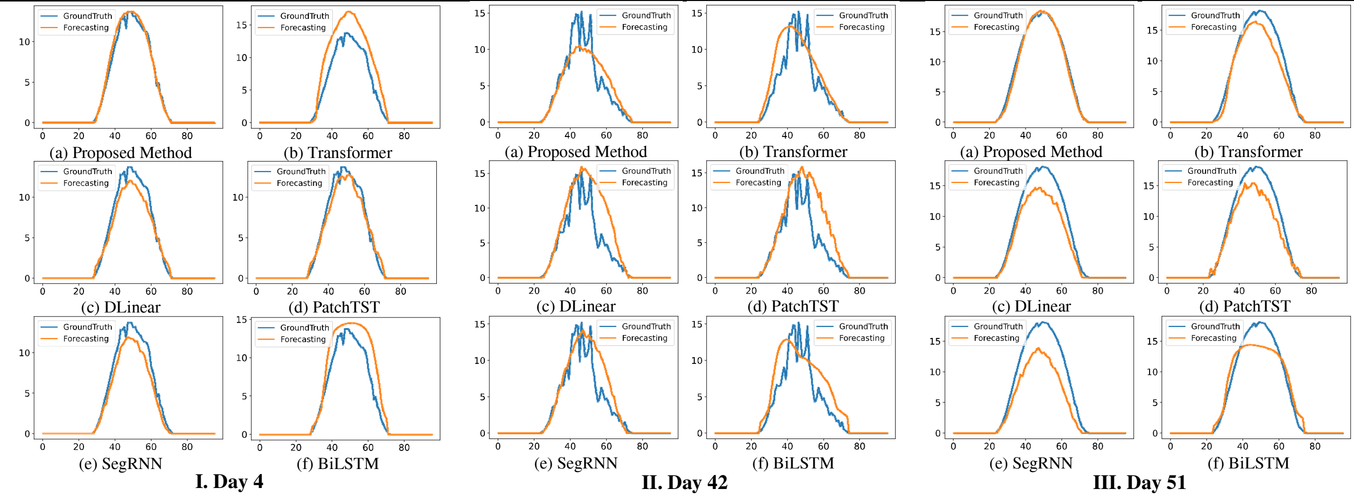

Fig. 7 Comparison of Forecasting Curves and Groundtruth of Different Models

V. CONCLUSION

The paper proposes a method for day-ahead PV power forecasting. The proposed method consists of new feature construction, a neural network, and output data constraints. The neural network component further includes a TCN module, a TQ module, a multi-head attention module, an MLP module, and projection and normalization layers. Compared with traditional general time series forecasting models, it has a clearer objective and is better suited to the specific characteristics of PV forecasting. The proposed method is applicable to various photovoltaic stations when the geographical coordinates are known.

ACKNOWLEDGMENT

This work was supported by the National Key Research and Development Program of China (2023YFB4005601).

REFERENCES

[1] K. Tao, J. Zhao, Y. Tao, Q. Qi, and Y. Tian, "Operational day-ahead photovoltaic power forecasting based on transformer variant," Appl. Energy, vol. 373, p. 123825, Nov. 2024, doi: 10.1016/j.apenergy.2024.123825.

[2] P. D. Leo, A. Ciocia, G. Malgaroli, and F. Spertino, "Advancements and Challenges in Photovoltaic Power Forecasting: A Comprehensive Review," Energies, vol. 18, no. 8, Apr. 2025, doi: 10.3390/en18082108.

[3] P. Gupta and R. Singh, "PV power forecasting based on data-driven models: a review," Int. J. Sustain. Eng., vol. 14, no. 6, pp. 1733–1755, Nov. 2021, doi: 10.1080/19397038.2021.1986590.

[4] M. J. Mayer, "Influence of design data availability on the accuracy of physical photovoltaic power forecasts," Sol. Energy, vol. 227, pp. 532–540, Oct. 2021, doi: 10.1016/j.solener.2021.09.044.

[5] W. Liu and Z. Mao, "Short-term photovoltaic power forecasting with feature extraction and attention mechanisms," Renew. Energy, vol. 226, p. 120437, May 2024, doi: 10.1016/j.renene.2024.120437.

[6] Y.-J. Liu, Y.-D. Lee, C.-Y. Lee, C.-C. Cheng, P.-Y. Hou, and Y.-F. Chen, "A Comparative Analysis of LSTM and BILSTM Network-Based Methods in PV Power Prediction," in 2022 IET International Conference on Engineering Technologies and Applications (IET-ICETA), Oct. 2022, pp. 1–2. doi: 10.1109/IET-ICETA56553.2022.9971566.

[7] N. Kim, H. Lee, J. Lee, and B. Lee, "Transformer based prediction method for solar power generation data," in 2021 International Conference on Information and Communication Technology Convergence (ICTC), Oct. 2021, pp. 7–9. doi: 10.1109/ICTC52510.2021.9620897.

[8] V. Suresh, "Benchmarking Transformer Variants for Hour-Ahead PV Forecasting: PatchTST with Adaptive Conformal Inference," Energies, vol. 18, no. 18, Sep. 2025, doi: 10.3390/en18185000.

[9] G. Wang, Y. Liao, L. Guo, J. Geng, and X. Ma, "DLinear photovoltaic power generation forecasting based on reversible instance normalization," in 2023 IEEE 12th Data Driven Control and Learning Systems Conference (DDCLS), May 2023, pp. 990–995. doi: 10.1109/DDCLS58216.2023.10166973.

[10] S.-G. Jeong, Q. V. Do, and W.-J. Hwang, "Short-term photovoltaic power forecasting based on hybrid quantum gated recurrent unit," ICT Express, vol. 10, no. 3, pp. 608–613, Jun. 2024, doi: 10.1016/j.icte.2023.12.005.

[11] L. Ge, Y. Xian, J. Yan, B. Wang, and Z. Wang, "A Hybrid Model for Short-term PV Output Forecasting Based on PCA-GWO-GRNN," J. Mod. Power Syst. Clean Energy, vol. 8, no. 6, pp. 1268–1275, Nov. 2020, doi: 10.35833/MPCE.2020.000004.

[12] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," J. Comput. Phys., vol. 378, pp. 686–707, Feb. 2019, doi: 10.1016/j.jcp.2018.10.045.

[13] S. Lin, H. Chen, H. Wu, C. Qiu, and W. Lin, "Temporal Query Network for Efficient Multivariate Time Series Forecasting," Sep. 11, 2025, arXiv: arXiv:2505.12917. doi: 10.48550/arXiv.2505.12917.

[14] S. Bai, J. Z. Kolter, and V. Koltun, "An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling," Apr. 19, 2018, arXiv: arXiv:1803.01271. doi: 10.48550/arXiv.1803.01271.

[15] "PVOD v1.0 : A photovoltaic power output dataset." Accessed: Jan. 23, 2026. [Online]. Available: https://www.scidb.cn/en/detail?dataSetId=f8f3d7af144f441795c5781497e56b62