Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence (IJCAI-25)
Special Track on AI and Social Good

MCloudNet: An Ultra-Short-Term Photovoltaic Power Forecasting Framework With Multi-Layer Cloud Coverage

Meng Wan[1,2†], **Tiantian Liu**[1,3†], **Yuxuan Bi**[4], **Jue Wang**[1,3*], **Hui Cui**[5], **Rongqiang Cao**[1,3], **Jiaxiang Wang**[2], **Peng Shi**[2], **Ningming Nie**[1,3], **Yangang Wang**[1,3]
[1]Computer Network Information Center, Chinese Academy of Sciences, Beijing, China
[2]University of Science and Technology Beijing, Beijing, China
[3]University of Chinese Academy of Sciences, Beijing, China
[4]North China Electric Power University,Baoding, China
[5]China Unicom Software Research Institute, Beijing, China
wanmengdamon@cnic.cn, liutiantian21a@mails.ucas.ac.cn, {biyuxuann, wjx3023}@gmail.com, {wangjue, caorq, nienm, wangyg}@sccas.cn, pshi@ustb.edu.cn

Abstract

Over 4.15 million low-income households across nearly 60,000 villages in China benefit from photovoltaic (PV) poverty alleviation power stations. However, weak infrastructure and limited capabilities make these systems vulnerable to fluctuations. One of the United Nations' Sustainable Development Goals (SDG 7) seeks to ensure access to affordable and reliable energy for all, especially in underdeveloped regions. This paper proposes MCloudNet, a multi-modal framework designed to improve ultra-short-term PV prediction in data-scarce, cloud-dynamic environments. MCloudNet explicitly models multi-layer cloud structures from satellite imagery and fuses them with time-series meteorological data to enhance prediction accuracy and interpretability. A province-level dispatch system with MCloudNet has been deployed in Hebei, supporting scheduling across rural PV stations. Experiments conducted in counties such as Shexian and Luxi highlight the framework's effectiveness for use in underdeveloped micro-grids. Operational results show that the system has reduced over 60 million kWh of solar curtailment and generated 24 million CNY in economic value, benefiting approximately 50,000 rural households. By minimizing power fluctuations and improving rural energy scheduling, MCloudNet supports essential services such as lighting, medical facilities, and communications. The source code is available at: https://github.com/AI4SClab/MCloudNet.

1 Introduction

Solar energy is a crucial renewable resource that plays a key role in the global energy transition and carbon neutrality goals [Saeed and Siraj, 2024]. The efficient utilization of solar power depends not only on the deployment of photovoltaic (PV) systems but also on accurate PV power forecasting [Iheanetu, 2022]. Reliable forecasting helps optimize energy dispatch, balance electricity supply-demand, and enhance grid stability [Zhang *et al.*, 2024]. In many underdeveloped regions, PV systems often serve as the primary energy source for small, independent power networks, known as micro-grids [Yan *et al.*, 2017]. According to the white paper *Energy in China's New Era* released by the State Council Information Office, **China has constructed photovoltaic (PV) poverty alleviation power stations as rural micro-grids with a total installed capacity of 26.36 GW, benefiting nearly 60,000 impoverished villages and 4.15 million low-income households** [State Council Office of the People's Republic of China, 2020]. However, weak infrastructure and limited grid regulation capacity make these micro-grids vulnerable to fluctuations (e.g., dynamic cloud shifts, solar intensity drops, and sudden weather changes), which can lead to instability in electricity supply and disrupt basic services such as lighting in schools, cold storage in clinics, or communications in emergency scenarios [van de Beek, 2024]. Additionally, the lack of real-time monitoring equipment constrains energy management, making satellite cloud imagery a cost-effective alternative for PV power forecasting.

Since 2014, the Chinese government has launched a Photovoltaic Poverty Alleviation Program (PPAP), driving continuous advancements in PV power forecasting technology [Wang *et al.*, 2020; Bai *et al.*, 2021]. Early approaches rely on simple statistical models and empirical formulas, which are easy to implement but lack accuracy in complex weather and geographical conditions, such as autoregression (AR) [Hannan and Kavalieris, 1986], autoregressive moving average (ARMA) [Swami *et al.*, 1994] and autoregressive integrated moving average (ARIMA) [Ho and Xie, 1998]. The introducing of physical models improve prediction accuracy by incorporating meteorological data and PV system characteristics [Dolara *et al.*, 2015]. In recent years, deep learning models such as MLP[Wang *et al.*, 2016], RNN[Rajagukguk *et al.*, 2020], CNN[Qin *et al.*, 2022], have shown capibility to effectively capture non-linear relationships between power

*Corresponding author

generation and multiple influencing factors, significantly improving prediction accuracy. Furthermore, multi-modal approaches have integrated satellite cloud imagery with time-series datasets, further enhancing predictive performance by capturing spatial-temporal cloud variations that directly impact solar radiation [Yao *et al.*, 2021; Hajjaj *et al.*, 2023; Xie *et al.*, 2024].

Despite significant advancements in PV power forecasting, challenges remain in achieving accurate predictions under complex weather conditions and rural deployment scenarios. Studies have shown that China's average curtailment rates rise to 2.8% in 2024, resulting in significant renewable energy waste due to insufficient forecasting precision and dispatch support [Ratings, 2024]. Specifically, the first major challenge is the inadequate complexity of single-layer cloud cover, especially in regions with highly dynamic weather patterns. Existing methods oversimplify irregular, complicated cloud cover as a single-layer uniform structure [Son *et al.*, 2023; Wang *et al.*, 2024b], which fails to capture the fine-grained temporal and spatial changes in cloud movements. As a result, these models struggle with providing accurate real-time forecasts, particularly during fast-changing weather events such as cloud formations and dissipations that significantly impact solar radiation. Another challenge is effectively fusing heterogeneous data features, such as multi-layer satellite cloud imagery and time-series meteorological data, into a unified framework for micro-grids forecasting. The varying resolutions and formats of these data limit predictive accuracy and real-time applicability in micro-grid environments.

Figure 1: Left: Rural PV station in Yunnan Province. Right: Local grid-side distribution infrastructure supporting dispatch operations.

In this paper, we propose a novel multi-modal framework MCloudNet that improves PV power forecasting by introducing a composite multi-layer cloud coverage approach and a time-series prediction model. To the best of our knowledge, this is the first work that explicitly models multi-layer cloud structures based on satellite imagery for PV forecasting. Figure 1 shows a rural PV station in Yunnan Province, where MCloudNet has been integrated into local dispatch operations. Our main contributions are as follows:

- **Multi-layer cloud feature capture modeling**: We propose a composite multi-layer cloud modeling approach that extracts high, middle, and low-altitude cloud layers from satellite imagery. By incorporating cloud motion vectors and occlusion coefficients, our method effectively captures cloud dynamics and inter-layer interactions, enhancing both accuracy and interpretability.
- **Multi-modal PV prediction framework**: We design a multi-modal learning framework that fuses multi-layer clouds and ground-based data, leveraging multi-scale information to improve predictive accuracy under diverse weather conditions.
- **State-of-the-art (SOTA) performance**: Our model achieves SOTA performance on multiple real-world datasets, consistently outperforming existing methods in both accuracy and generalization.
- **Social and economic benefits**: The MCloudNet framework has been successfully applied in underdeveloped villages and micro-grids in Hebei, Yunnan, and Shandong, across over 50 photovoltaic stations. By providing accurate PV forecasting, it reduces 60 million kWh of curtailment power, generates significant economic benefits, and ensures the security of micro-grid dispatch.

By improving forecast accuracy and energy management efficiency, our research directly contributes to the Sustainable Development Goals of the United Nations (SDG 7: Affordable and Clean Energy, SDG 9: Industry, Innovation and Infrastructure, SDG 13: Climate Action). Furthermore, it aligns with the principle "Leave No One Behind" (LNOB) by promoting equitable access to renewable energy and ensuring that underserved communities benefit from sustainable energy solutions [Carlsen and Bruggemann, 2022].

2 Related Work

In recent years, photovoltaic power prediction models have undergone continuous development and evolution [Al-Dahidi *et al.*, 2024]. With the rapid advancement of deep learning technology, researchers have gradually introduced various deep learning architectures into photovoltaic power prediction [Aslam *et al.*, 2021; Wazirali *et al.*, 2023]. For example, [Agga *et al.*, 2022] proposed a hybrid architecture combining a convolutional neural network (CNN) and a long short-term memory network (LSTM), which fully utilized the spatial characteristics and temporal dependencies of photovoltaic data, thereby improving the prediction accuracy. At the same time, [Chen *et al.*, 2024] proposed a CGAformer model that integrates the global additive attention mechanism, which effectively solves the problem of multi-scale feature fusion in short-term photovoltaic prediction. However, most of these methods rely on a single type of data (such as historical power data or meteorological data), and fail to fully utilize the complementarity of multi-source data, resulting in certain limitations under complex meteorological conditions [Tan *et al.*, 2024]. To further improve the prediction accuracy, researchers began to explore deep learning methods based on multi-source heterogeneous data fusion. These methods can capture meteorological and spatial information more comprehensively by integrating local measurement data, numerical weather forecast (NWP) data, and satellite images. For example, [Yao *et al.*, 2021] used U-Net and an encoder-decoder architecture to jointly process multi-source data, thereby improving the accuracy of photovoltaic forecasts. [Yang *et al.*, 2023] fused historical power data, NWP data, and irradiance

data of adjacent sites to effectively reduce the impact of meteorological factors on forecast errors. However, the current multi-source data fusion method still does not fully consider the impact of cloud movement and distribution on ultra-short-term photovoltaic forecasts, which limits the adaptability and prediction accuracy of the model under complex meteorological conditions.

3 The Overall Framework

The overall framework of the MCloudNet we proposed is shown in Figure 2, which mainly consists of four stages: data collection, cloud image reconstruction and layering, cloud forecasting and GHI (Global Horizontal Irradiance) correction, and the time-series PV prediction [Stein *et al.*, 2012]. First, in the data collection stage, the input data includes NWP data, local meteorological data, and satellite cloud images, providing comprehensive and multi-dimensional environmental and meteorological information support for the model. Subsequently, in the cloud image reconstruction and layering stage, channel extraction and temperature stratification operations are performed to generate layered historical cloud maps, which can provide richer and more accurate meteorological data compared to traditional single cloud layer research. In the cloud forecasting and GHI Correction stage, multi-layer satellite cloud images and the optical flow method are used to predict the movement trajectory of clouds. The NWP data is corrected by convolutional neural networks to generate corrected NWP-GHI data, improving the accuracy of the prediction data. Finally, in the time-series PV prediction stage, in order to combine multi-layer high-resolution satellite data with time-series data, the model optimizes the input data through the time alignment and noise reduction module, and extracts deep features through the hierarchical feature extraction module and cross-scale attention module . At the same time, combined with the corrected NWP-GHI data, the final prediction result is generated. This framework can efficiently integrate satellite images and ground-based meteorological data and can be applied to various scenarios, such as energy prediction, carbon emission analysis, and power grid dispatch.

4 Methods

This section presents the MCloudNet framework for ultra-short-term PV power forecasting, consisting of three main modules. Specifically, referring to predictions of photovoltaic generation within the next 15 minutes to 4 hours, with a time resolution of 15 minutes.

4.1 Cloud Image Reconstruction and Layering

The Himawari-8 satellite, operated by the Japan Meteorological Agency, is a geostationary meteorological satellite covering 16 bands. Among these, the infrared Top of Atmosphere Brightness Temperature (TBB) channel records the black body temperature derived from the radiative brightness observed by the satellite, which can be used to generate cloud images. The TBB_{13} channel is selected for cloud image reconstruction because it represents the cloud top temperature, which can indirectly reflect the cloud height. Specifically, we use the temperature values from the TBB_{13} channel as the RGB values without additional mapping. The image is then normalized, adjusting the pixel values to the range of[0,255], resulting in a three-channel cloud image based on cloud top temperature. According to the World Meteorological Organization [Organization, 2017; Molero *et al.*, 2022], cloud layers are classified as shown in Table1. The classification follows the temperature thresholds defined for low, mid, and high clouds, which help in determining the distinct characteristics and weather patterns.

Cloud Type	Temperature Range (°C)	Height (m)
Low Clouds	$> 0°$C	$0 - 2000$
Mid Clouds	$0°$C to $-20°$C	$2000 - 6000$
High Clouds	$< -20°$C	> 6000

Table 1: Classification of layers based on cloud top temperatures and heights.

In each cloud layer image, we preserve the pixels corresponding to the specific cloud layer with their original temperature values, while pixels from the other layers are masked. Mathematically, the processes can be represented as follows:

$$I_{\mathrm{t}}(x,y) = \begin{cases} T_{\mathrm{t}}(x,y) & \text{if Layer}(x,y) = \text{target layer} \\ \text{Mask} & \text{if Layer}(x,y) \neq \text{target layer} \end{cases} \quad (1)$$

where $I_{\mathrm{t}}(x,y)$ denotes the pixel value at coordinates (x,y) in the cloud image for the target cloud layer at time t, and $T_{\mathrm{t}}(x,y)$ represents the temperature value at the same coordinates for the target layer.

4.2 Cloud Forecasting and GHI Correction

We employ the optical flow method to obtain the predicted cloud map based on historical cloud imagery. Subsequently, the predicted cloud map is processed through a feature extraction network to achieve a more accurate estimation of the GHI value at the forecasted time.

[Optical Flow Cloud Prediction] The position and extent of cloud cover directly affect PV power generation efficiency, and the cloud's shape and movement may change within a short period (the 10-minute interval). To accurately predict PV power generation, we obtain cloud maps at the predicted time. Considering the time delay of satellite cloud images, we employed the Dual TV-L1 optical flow method [Zach *et al.*, 2007] to predict the cloud maps at the predicted time from two consecutive cloud maps. We estimate motion between two image frames by minimizing both image intensity gradients and flow field variation.Specifically, to obtain the predicted cloud map I_t from times t, we use the cloud maps I_{t-1} and I_{t-2} from times $t-1$ and $t-2$, respectively, to calculate the optical flow field u between them:

$$E(u,v) = \iint \left[(I_{t-1}(x,y) - I_{t-2}(x + u(x,y), y + v(x,y)))^2 + \lambda(|\nabla u(x,y)| + |\nabla v(x,y)|) \right] dx\, dy \quad (2)$$

where $u(x,y)$ and $v(x,y)$ are the components of the optical flow field in the x and y directions, respectively. λ is the regularization parameter controlling the smoothness of the optical flow field.

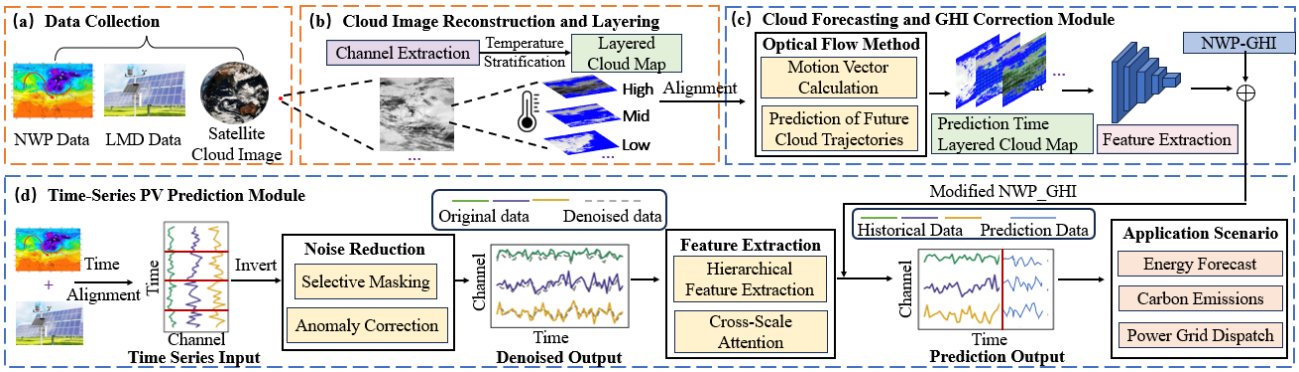


Figure 2: The overall framework of MCloudNet.

By minimizing this energy function, we obtain the optical flow field:

$$(u^*, v^*) = \arg\min_{u,v} E(u, v) \tag{3}$$

Then, we obtain the cloud image at the predicted time:

$$I_t(x, y) = I_{t-1}(x + u(x, y), y + v(x, y)) \tag{4}$$

[Feature Extraction and Representation]

One of the most critical factors influencing photovoltaic (PV) power generation is the global horizontal irradiance (GHI) received by the PV panels [Kallio-Myers *et al.*, 2020; Wang *et al.*, 2024a]. While numerical weather prediction (NWP) models estimate GHI using physical principles and empirical formulas, these methods primarily capture diurnal and seasonal patterns, and often fall short in accurately representing short-term fluctuations due to limited responsiveness to rapid weather changes.

To address this limitation and enhance the precision of radiation forecasting, we propose a heterogeneous data fusion network that integrates satellite cloud imagery with NWP data to predict the GHI value G_t at a target time t. The proposed network comprises two branches that independently process cloud images and NWP features, enabling complementary information extraction.

1. **Cloud Imagery branch:** For a given time t, the cloud imagery input is a set of three-channel cloud images (high-, middle-, and low-level clouds), denoted as $I_{\text{t}} = [I_t^{\text{high}}, I_t^{\text{mid}}, I_t^{\text{low}}]$, where $I_t^{\text{high}}, I_t^{\text{mid}}, I_t^{\text{low}} \in \mathbb{R}^{C\times H\times W}$. These images are processed using three separate convolutional residual networks [He *et al.*, 2016] to extract feature vectors $F_{\text{high}}, F_{\text{mid}}, F_{\text{low}} \in \mathbb{R}^{d_{\text{model1}}}$, respectively.

$$F_{\text{high}}, F_{\text{mid}}, F_{\text{low}} = \text{FE}(I_t^{\text{high}}), \text{FE}(I_t^{\text{mid}}), \text{FE}(I_t^{\text{low}}) \tag{5}$$

To better explore the contribution of different cloud layers to the results, cloud images features are multiplied by three learnable parameters weights w_{high}, w_{mid}, and w_{low} respectively, before being fed into the fusion branch.

$$F_{\text{t}} = [w_{\text{high}} \times F_{\text{t}}, w_{\text{mid}} \times F_{\text{mid}}, w_{\text{low}} \times F_{\text{low}}] \tag{6}$$

2. **NWP branch:** The NWP-GHI data x_t is processed through an embedding layer to convert it into a token $T_{\text{NWP}} \in \mathbb{R}^{d_{\text{model2}}}$:

$$T_{\text{NWP}} = \text{Embedding}(x_t) \tag{7}$$

3. **Fusion Network:** Subsequently, a fusion network integrates the cloud image feature vectors and the NWP-GHI token to model their interaction, producing a more accurate GHI prediction G_t. The cloud feature vector F_{t} and NWP-GHI token T_{NWP} are fused using an MLP to produce the final predicted GHI value G_t:

$$G_t = \text{MLP}(F_{\text{t}}, T_{\text{NWP}}) \tag{8}$$

4.3 Time-Series PV Prediction

To enhance the time series analysis capability of our multimodal model, we design an advanced cross-scale fusion transformer. Our model consists of three key components: Noise Reduction Module, Hierarchical Feature Extraction Module, and Cross-Scale Attention Module. Given an input time series $\mathbf{X} \in \mathbb{R}^{C\times L}$, where C represents the number of channels and L represents the length of the time steps.

[Noise Reduction Module] The noise reduction module aims to enhance the quality of input data by dynamically adjusting the contribution of points within each channel. Specifically, it refines the input data through selective masking and anomaly correction. Given an inverted tensor $\mathbf{X}_{\text{inv}} \in \mathbb{R}^{L\times C}$, obtained by independently flipping $\mathbf{X}$ along each channel, the noise reduction operation can be expressed as:

$$\mathbf{M} = f(\mathbf{X}_{\text{inv}}; \theta) \tag{9}$$

$$\mathbf{X}_{\text{denoised}} = \mathbf{X}_{\text{inv}} \odot \mathbf{M} \tag{10}$$

where $\mathbf{M}$ is a learnable mask matrix, f is a parameterized function, and θ represents the parameters of this function. Through training, the values in the mask matrix $\mathbf{M}$ gradually approach either 0 or 1, effectively selecting features and suppressing noise.

[Hierarchical Feature Extraction Module] The hierarchical feature extraction module constructs multi-level representations of the denoised data to capture features at different scales. The core idea is to extract fluctuations and trends across various temporal scales using a multi-level structure. The output of this module can be represented as:

$$\mathbf{H} = \text{HierarchicalFeatureExtraction}(\mathbf{X}_{\text{denoised}}) \tag{11}$$

where $\mathbf{H}$ contains feature representations at different scales. Specifically, hierarchical feature extraction can be achieved through down-sampling and up-sampling operations:

$$\mathbf{H}_i = \text{Downsample}(\mathbf{X}_{\text{denoised}}, i) \tag{12}$$

$$\mathbf{H}'_i = \text{Upsample}(\mathbf{H}_i, i) \tag{13}$$

where i is the layer of feature representations.

[Cross-Scale Attention Module] The cross-scale attention module utilizes attention mechanisms to fuse semantic information from different scales, deeply analyzing interactions between different scales and capturing a broader spectrum of complex temporal patterns. Assuming that the feature representations at different scales are $\mathbf{H}'_i$, the output of this module can be represented as:

$$\mathbf{Z} = \text{Cross-Scale Attention}([\mathbf{H}'_1, \mathbf{H}'_2, ..., \mathbf{H}'_N]) \tag{14}$$

where $\mathbf{Z}$ is the final aggregated feature representation. Specifically, cross-scale attention is applied to the concatenated multi-scale feature set:

$$\mathbf{Q}_i = \mathbf{W}_Q \mathbf{H}'_i, \quad \mathbf{K}_i = \mathbf{W}_K \mathbf{H}'_i, \quad \mathbf{V}_i = \mathbf{W}_V \mathbf{H}'_i \tag{15}$$

$$\text{Attention}(\mathbf{Q}_i, \mathbf{K}_j, \mathbf{V}_j) = \text{softmax}\left(\frac{\mathbf{Q}_i \mathbf{K}_j^T}{\sqrt{d_k}}\right)\mathbf{V}_j \tag{16}$$

The final cross-scale attention output aggregates information from all scales:

$$\mathbf{Z} = \sum_{i=1}^{N}\sum_{j=1}^{N} \text{Attention}(\mathbf{Q}_i, \mathbf{K}_j, \mathbf{V}_j) \tag{17}$$

The output processor aggregates the processed features and generates the prediction output. Assuming that the final feature representation is $\mathbf{Z}$, the predicted output $\hat{\mathbf{Y}}$ can be represented as:

$$\hat{\mathbf{Y}} = \mathbf{W}_O \mathbf{Z} + \mathbf{b}_O \tag{18}$$

where $\mathbf{W}_O$ and $\mathbf{b}_O$ are the weights and biases of the output layer, respectively.

After obtaining the PV power predictions C_{PV} from CSFformer, an MLP is employed to learn the relationship between C_{PV} and the current GHI G_t. This ensures that the final PV power not only follows the historical trend but also accurately accounts for the impact of the current GHI.

$$\mathbf{Y}_{PV} = \text{MLP}(\hat{\mathbf{Y}}, G_t) \tag{19}$$

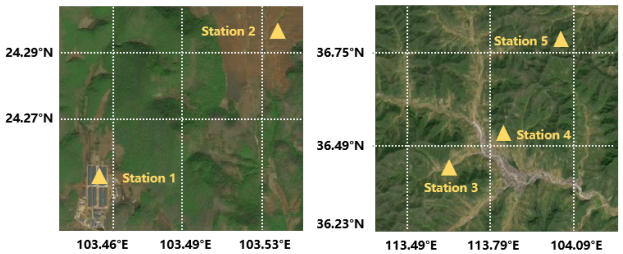
The model is trained using a combination of cloud image data, NWP data, and real-world PV power generation data. The training process involves optimizing the weights and parameters of the MCloudNet modules, ensuring that the model can accurately predict solar power output for different regions and weather conditions.

5 Experiments

In this section, we evaluate the proposed model through experiments and ablation studies, demonstrating its effectiveness in PV power forecasting and analyzing the contribution of each component.

5.1 Experimental Settings

Datasets. As shown in Figure 3, we construct a PV power generation dataset based on two stations in Luxi County, Yunnan, and three in Shexian County, Hebei. The dataset integrates multi-source inputs, including Local Meteorological Data (LMD), Numerical Weather Prediction (NWP), and stratified satellite cloud images. LMD and NWP provide 15-minute resolution weather indicators (e.g., irradiance, temperature, wind speed), while satellite imagery from Himawari-8 offers 10-minute interval cloud-top temperature data. All modalities are temporally aligned at 15-minute intervals. An overview of spatial, temporal, and real-time characteristics is summarized in Table 3.



(a) Luxi County, Yunnan (b) Shexian County, Hebei

Figure 3: Location of photovoltaic stations.

Baselines. We select eight competitive models as baselines, including A-UNET[Yao *et al.*, 2021], DLinear[Zeng *et al.*, 2023], PatchTST[Nie *et al.*, 2022], TimesNet[Wu *et al.*, 2022], STLAN[Yang *et al.*, 2023], iTransformer[Liu *et al.*, 2023b], SDformer[Li *et al.*, 2024].

Parameter Details. We divide the dataset into a training set and a test set in an 8:2 ratio. In the GHI correction module, the training uses 15 epochs, with a sincos learning rate schedule and an initial learning rate of 0.001, and a batch size of 24. In the image branch feature dimension of each layer of cloud images d_{model1} is 256, while the NWP-GHI feature dimension d_{model2}in the NWP branch is 512. The fusion network contains a hidden layer, with dropout set to 0.5. In the MCloudNet model, the training uses 10 epochs, with the learning rate following exponential decay, starting at 0.0001. The input sequence length is configurable with four preset options: $\{24\ (6\text{h}), 48\ (12\text{h}), 96\ (24\text{h}), 192\ (48\text{h})\}$. In the GHI-based correction of PV, the number of training epochs is 200, using the Adam optimizer with an initial learning rate of 0.01.All experiments were conducted on an Intel(R) Xeon(R) CPU E5-2650 (16 cores) and GeForce-RTX-4090-24GB GPUs.

5.2 Main Experiments Results

In the Table 2, we present a comparison of our model's results with baseline models. Our model achieves SOTA performance in most cases. It is observed that the average prediction error in Luxi County (Stations 1 and 2) is higher than in Shexian County (Stations 3, 4, and 5). We speculate that this is due to the complex mountainous terrain in Luxi County, which leads to more dynamic cloud movements and localized microclimate effects, such as valley winds and slope radiation variations. These factors cause frequent and unpredictable changes in cloud coverage, making cloud motion estimation via optical flow more challenging. Despite these difficulties, our model demonstrates a more significant improvement over baseline methods in Luxi County, suggesting that the multi-

	Methods	Ours (2025)		SDformer (2024)		iTransformer (2024)		STLAN (2023)		TimesNet (2023)		PatchTST (2023)		DLinear (2023)		A-UNET (2022)	
	Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Station 1	24	**1.998**	**0.954**	2.017	0.975	2.027	0.994	2.021	0.963	2.227	0.995	2.232	1.228	2.436	1.454	2.369	1.372
	48	2.009	0.971	**2.002**	**0.952**	2.005	0.969	1.997	0.965	2.193	0.969	2.384	1.051	2.526	1.303	2.478	1.262
	96	**1.874**	0.952	1.976	**0.917**	1.987	0.953	1.979	0.899	2.136	1.021	2.135	1.169	2.357	1.326	2.243	1.237
	192	**2.007**	**0.812**	2.011	0.980	2.010	0.921	2.101	0.938	2.190	0.993	2.273	1.235	2.402	1.457	2.352	1.304
	avg	**1.972**	**0.872**	2.012	0.941	2.007	0.959	2.025	0.942	2.187	0.995	2.256	1.171	2.430	1.385	2.361	1.294
Station 2	24	**2.012**	**0.779**	2.251	0.927	2.313	0.980	2.268	0.932	2.212	1.012	2.169	1.156	2.365	1.314	2.254	1.276
	48	**2.006**	0.984	2.041	0.943	2.011	**0.932**	2.024	0.941	1.141	0.998	2.254	1.233	2.415	1.390	2.348	1.353
	96	**1.989**	**0.892**	2.033	0.939	1.996	0.897	2.021	0.954	1.118	1.217	2.212	1.223	2.432	1.455	2.340	1.370
	192	**1.999**	**0.877**	2.023	0.935	2.004	0.896	2.013	0.917	1.188	1.126	2.199	1.179	2.377	1.397	2.274	1.265
	avg	**2.001**	**0.883**	2.087	0.936	2.079	0.916	2.082	0.936	1.415	1.088	2.209	1.198	2.397	1.389	2.304	1.316
Station 3	24	**1.283**	0.741	1.371	0.732	1.368	**0.724**	1.355	0.725	1.487	0.895	1.545	0.998	1.698	1.218	1.588	1.072
	48	**1.198**	**0.702**	1.351	0.729	1.317	0.726	1.313	0.745	1.443	0.855	1.473	0.926	1.631	1.156	1.524	1.040
	96	1.233	0.725	1.330	0.717	1.244	**0.706**	**1.229**	0.739	1.356	0.950	1.503	0.964	1.705	1.226	1.605	1.037
	192	**1.262**	0.745	1.339	**0.728**	1.281	0.730	1.292	0.748	1.490	0.903	1.556	0.985	1.692	1.157	1.603	1.034
	avg	**1.244**	0.728	1.348	0.727	1.302	**0.722**	1.297	0.739	1.444	0.901	1.519	0.968	1.682	1.189	1.580	1.046
Station 4	24	**1.471**	0.766	1.589	0.774	1.563	**0.720**	1.579	0.764	1.588	0.968	1.698	1.142	1.852	1.266	1.715	1.194
	48	**1.451**	**0.702**	1.525	0.763	1.498	0.723	1.512	0.733	1.660	0.930	1.673	1.131	1.814	1.192	1.785	1.061
	96	**1.444**	**0.695**	1.508	0.747	1.452	0.697	1.491	0.723	1.521	0.849	1.596	0.885	1.797	1.142	1.692	1.033
	192	**1.899**	**0.802**	1.915	0.808	1.996	0.805	1.925	0.892	1.824	1.228	2.162	1.201	2.330	1.342	2.241	1.295
	avg	**1.566**	0.741	1.634	0.773	1.627	**0.736**	1.628	0.779	1.648	0.994	1.782	1.017	1.948	1.236	1.858	1.146
Station 5	24	**1.332**	**0.549**	1.489	0.674	1.459	0.665	1.479	0.664	1.488	0.753	1.545	0.748	1.687	0.966	1.595	0.894
	48	**1.323**	0.678	1.455	0.683	1.425	**0.675**	1.451	0.682	1.460	0.810	1.533	0.812	1.714	1.092	1.665	0.961
	96	**1.288**	0.671	1.418	**0.653**	1.404	0.673	1.411	0.676	1.521	0.849	1.597	0.845	1.591	1.043	1.492	0.933
	192	**1.371**	0.754	1.445	0.668	1.470	0.704	1.415	**0.643**	1.524	1.048	1.638	1.024	1.730	1.298	1.541	1.115
	avg	**1.328**	**0.663**	1.452	0.669	1.440	0.684	1.439	0.666	1.498	0.834	1.546	0.857	1.682	1.099	1.573	0.976
	Count	36		5		7		2		0		0		0		0	

Table 2: Results on Luxi and Shexian County, where the best and second-best results are shown in **bold** and underlined, respectively. 'Avg' represents the average across four look-back window lengths.

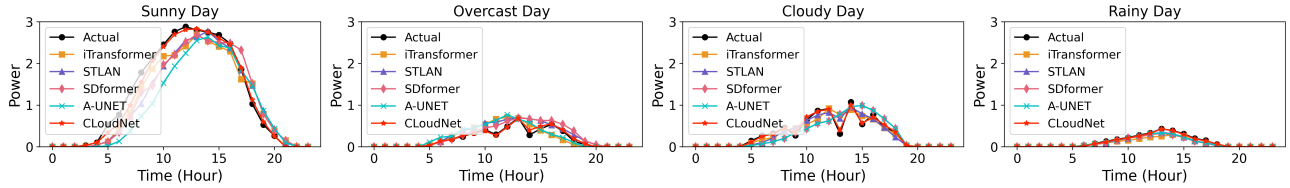


Figure 4: Photovoltaic forecast results under different weather conditions.

Data Type	Spatial Resolution	Temporal Resolution	Real-time Availability
Satellite	5 km × 5 km	10 min	20 min delay
LMD	Station-based	15 min	Real-time
NWP	9 km × 9 km	15 min	Twice per day

Table 3: Overview of Multi-Source Data for PV Power Forecasting.

layer cloud modeling effectively captures cloud dynamics and enhances prediction accuracy in complex environments. Furthermore, we observe that increasing the historical window initially reduces prediction error, but excessive data introduces noise, leading to degraded performance. Notably, in all tested historical windows, our model consistently outperforms baseline methods, demonstrating its ability to effectively capture the ultra-short-term forecasting characteristics while mitigating the impact of excessive temporal dependencies. Moreover, the consistent performance gain across both regions indicates the strong robustness of MCloudNet.

5.3 Ablation Study

To investigate the effectiveness of different components in MCloudNet, we conduct an ablation study by removing key modules and evaluating their impact on forecasting performance. The compared variants are defined as follows:

- **W/O-Correction**: Remove the GHI correction module, relying solely on time-series module with LMD and NWP data.
- **W/O-Former**: Remove the time-series prediction module, directly mapping the GHI correction module to PV values with MLP.

- **W/O-MCloud**: Exclude the multi-cloud imagery component from GHI correction module.
- **Single-Cloud**: Replace multi-layer cloud as a single-layer cloud component.

Models	Metrics	Station1	Station2	Station3	Station4	Station5
W/O Correction	MSE	1.987	2.058	1.322	1.554	1.273
	MAE	0.927	0.985	0.747	0.679	0.744
W/O Former	MSE	8.121	8.459	5.187	5.574	5.184
	MAE	2.827	2.753	1.751	1.796	1.737
W/O MCloud	MSE	2.001	2.021	1.262	1.528	1.423
	MAE	0.987	0.984	0.751	0.791	0.712
Single Cloud	MSE	1.931	1.997	1.250	1.458	1.391
	MAE	0.960	0.925	0.740	0.728	0.701
MCloudNet	MSE	**1.874**	**1.989**	**1.233**	**1.444**	**1.288**
	MAE	**0.952**	**0.892**	**0.725**	**0.695**	**0.671**

Table 4: Ablation study on different modules and components on look-back window length=96.

As shown in Table 4, W/O-former is unable to capture historical weather patterns, and W/O-Corrction fails to capture cloud fluctuations, both leading to a decline in performance, highlighting the importance of considering both historical trends and ultra-short-term fluctuations in photovoltaic power forecasting. In the W/O-Former experiment, the model error significantly increases, indicating that photovoltaic (PV) power prediction requires not only consideration of GHI fluctuations but also the influence of covariates such as temperature and wind speed.

Besides, the increases in MSE rates by 1.3% and 2.3% for the W/O-MCloud and Single-Cloud respectively, demonstrate that employing multi-layer cloud images yield superior results compared to single cloud images, which in turn outperform the absence of cloud images altogether. This underscores the efficacy of the GHI correction module, highlighting that multi-layer cloud images facilitate the attainment of a more precise real-time GHI, which enhances the accuracy of PV prediction.

5.4 Weather Conditions

As shown in Figure 4, this study compares the predictive performance of MCloudNet and baseline models under different weather conditions. Under sunny conditions, baseline models and MCloudNet can effectively capture long-term trends, resulting in predictions that are relatively close to the actual data. However, MCloudNet further improves the prediction accuracy by capturing features of different scales through the noise reduction module and the hierarchical feature extraction module. In overcast and cloudy conditions, baseline models struggle to accurately capture fine-grained spatiotemporal variations in cloud movements, which limits their predictive performance. In contrast, MCloudNet utilizes multi-layer cloud and ground data, enabling it to effectively capture the dynamic characteristics and irregular changes of clouds.

5.5 Visual Analysis

[Optical Flow Prediction Visualization] To evaluate the feasibility of short-term cloud motion forecasting, we visualize the optical flow derived from sequential cloud images. As shown in Figure 5, the optical flow is computed from the original cloud image at 13:50, capturing the estimated motion of cloud patterns over time. The visualization includes actual and predicted cloud images across multiple time steps, along with a delta image highlighting prediction discrepancies. This setup facilitates a direct visual assessment of prediction performance and confirms that optical flow can effectively capture coherent cloud movement patterns, supporting its utility in spatiotemporal modeling for short-term GHI value and PV power prediction.

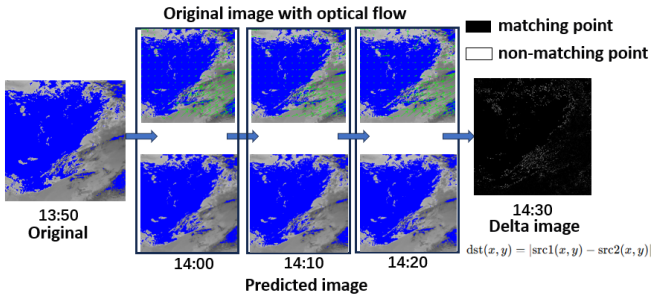


Figure 5: Cloud motion forecasting using optical flow. The large image on the far left shows the original cloud image at 13:50. Green arrows overlayed on the actual cloud images indicate the estimated optical flow toward the next time step; arrow direction and length denote motion direction and magnitude, respectively. Predicted cloud images are shown below the observations for comparison. The delta image on the far right shows pixel-wise differences between the predicted and actual cloud image at 14:30.

[Contribution of Multi-layer Clouds to the GHI] In the GHI Correction module, three learnable parameters w_{high}, w_{mid}, and w_{low} are introduced to quantify the contributions of high, mid, and low cloud layers to the final results. These weights, as learnable parameters, are adjusted during the network training process. The trend of these weight parameters changing with training epochs is depicted in Figure 6, showing that initially all weights are equal, but eventually the weight for low clouds is the highest, followed by mid clouds, and high clouds is the lowest. This outcome aligns with meteorological conclusions, indicating that

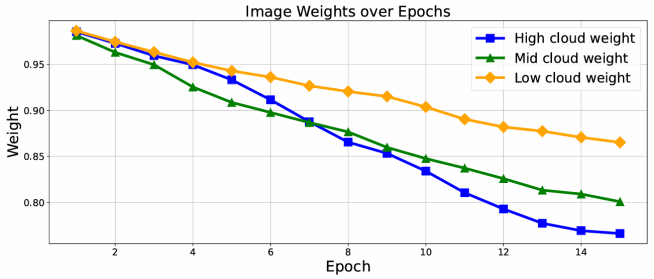


Figure 6: The contribution of each cloud layer to the predicted GHI.

the GHI correction module captures the impact of different cloud layers on the GHI value [Miyamoto *et al.*, 2021; Luo *et al.*, 2023].

5.6 Scales Evaluation of Satellite Image

To explore the impact of spatial resolution on the accuracy of PV power prediction, four image sizes were selected in the experiment: 4×4, 8×8, 16×16, and 32×32. Considering that the spatial resolution of Himawari-8 satellite cloud images is 5 kilometers per pixel, these configurations correspond to ground coverage areas ranging from approximately 20 kilometers to 160 kilometers. The evaluation criterion for determining the optimal input resolution was the final PV power prediction error (MSE and MAE). As shown in Table 5, smaller images like 4×4 showed significantly higher errors, likely due to excessive loss of spatial information. Moreover, increasing image size beyond 16×16 did not yield further improvements and may introduce computational burdens and potential overfitting risks. Based on the optimal results, we chose the 16×16 satellite cloud images for the main experiment.

Image Size	Coverage Area	MSE	MAE
4×4	20 km × 20 km	1.230	0.709
8×8	40 km × 40 km	1.206	0.713
16×16	80 km × 80 km	1.198	0.702
32×32	160 km × 160 km	1.202	0.708

Table 5: Performance metrics for different image sizes with look-back window length = 48

5.7 Transferability and Data Scarcity Experiment

Newly established PV stations in rural areas often lack historical LMD data, making classical forecasting methods ineffective. This experiment evaluates MCloudNet's ability to adapt to new stations with limited data by assessing its transferability. We pretrain MCloudNet on Hebei PV stations and test its performance on newly established stations with Cambricon MLU270 GPU [Liu *et al.*, 2023a] in Yunnan, using only 6 months time-series data. As shown in Table 6, MCloudNet significantly outperforms baseline models, demonstrating strong transfer capability. The effectiveness is attributed to the abundant availability of satellite cloud images, which provide critical GHI estimates that drive PV power fluctuations. By leveraging multi-layer cloud information, MCloudNet effectively compensates for the absence of local meteorological data, ensuring practical applicability in data-scarce rural PV projects.

Station	Ours		SDformer		iTransformer		A-UNET	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Station 1	**2.438**	**1.212**	4.615	1.998	4.443	2.012	3.140	1.625
Station 2	**2.343**	**1.179**	4.396	1.984	6.127	2.025	3.355	1.840

Table 6: Performance comparison on look-back window length=96.

6 Deployment and Social Impact

6.1 PV Prediction System Interface

To support real-time PV dispatch, we developed a visualized forecasting platform that has been deployed across Hebei Province, China. As shown in Figure 7, the system integrates multi-scale prediction models (15-minute, 4-hour, and 24-hour horizons) and currently connects to 128 PV stations. As a province-level deployment, the system demonstrates how AI-driven forecasting can empower large-scale renewable integration and inform data-driven energy governance.

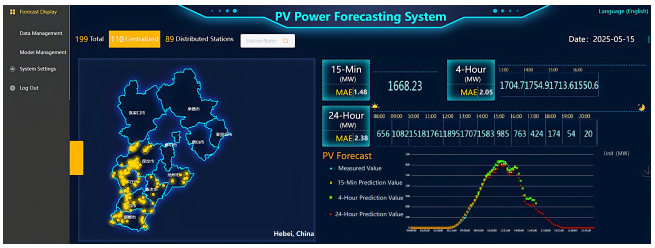


Figure 7: Deployed PV prediction and dispatch system interface across Hebei Province.

6.2 Real-World Achievements and Impact

In collaboration with the Hebei Electric Power Company and the Weiqiao Institute of Science and Technology, the proposed MCloudNet framework has been successfully deployed in over 50 photovoltaic stations across underdeveloped villages in Hebei, Yunnan, and Shandong, China, contributing to improved energy access and stability. Over the past six months, it has reduced 60 million kWh (kilowatt-hour) of curtailment power from renewable energy, generating an economic benefit of 24 million CNY (Chinese Yuan) and benefiting approximately 50,000 low-income rural households. Beyond economic benefits, MCloudNet plays a crucial role in supporting micro-grids where a stable electricity supply remains a challenge. By providing accurate ultra-short-term PV power predictions, MCloudNet enables proactive energy scheduling, reducing dependency on expensive backup power sources, such as diesel generators. Furthermore, enhanced forecasting accuracy minimizes power fluctuations, reducing the risk of voltage instability and supply disruptions in isolated micro-grids. This improvement is particularly crucial for supporting essential services such as lighting, medical facilities, and communication infrastructure in these regions. These results contribute to the achievement of the SDGs and support the LNOB by promoting equitable access to renewable energy.

7 Conclusion

This paper introduces MCloudNet, a multi-modal framework that improves ultra-short-term photovoltaic power forecasting by multi-layer cloud modeling and time-series analysis. The approach shows promising results for accurate predictions in regions with limited monitoring infrastructure. Future work will focus on expanding the application across more regions to optimize energy dispatch and reduce carbon emissions.

Ethical Statement

There are no ethical issues.

Acknowledgements

This work is supported by the National Science and Technology Major Project of China (2023ZD0120503). We thank VenusAI Platform (http://data.aicnic.cn) for kindly providing the GPU clusters for model training. We also express our sincere gratitude to Hebei Electric Power Company and Weiqiao National Science and Technology Research Institute for providing access to real-world data and deployment environments, which enabled practical validation of the proposed system.

Contribution Statement

Meng Wan and Tiantian Liu have made equal contributions.

References

[Agga *et al.*, 2022] Ali Agga, Ahmed Abbou, Moussa Labbadi, Yassine El Houm, and Imane Hammou Ou Ali. Cnn-lstm: An efficient hybrid deep learning architecture for predicting short-term photovoltaic power production. *Electric Power Systems Research*, 208:107908, 2022.

[Al-Dahidi *et al.*, 2024] Sameer Al-Dahidi, Manoharan Madhiarasan, Loiy Al-Ghussain, Ahmad M Abubaker, Adnan Darwish Ahmad, Mohammad Alrbai, Mohammadreza Aghaei, Hussein Alahmer, Ali Alahmer, Piero Baraldi, et al. Forecasting solar photovoltaic power production: A comprehensive review and innovative data-driven modeling framework. *Energies*, 17(16):4145, 2024.

[Aslam *et al.*, 2021] Sheraz Aslam, Herodotos Herodotou, Syed Muhammad Mohsin, Nadeem Javaid, Nouman Ashraf, and Shahzad Aslam. A survey on deep learning methods for power load and renewable energy forecasting in smart microgrids. *Renewable and Sustainable Energy Reviews*, 144:110992, 2021.

[Bai *et al.*, 2021] Bo Bai, Siqin Xiong, Xiaoming Ma, and Yushen Tian. Effectiveness evaluation of photovoltaic poverty alleviation project in china: From a capital perspective. *Journal of Cleaner Production*, 284:124684, 2021.

[Carlsen and Bruggemann, 2022] Lars Carlsen and Rainer Bruggemann. The 17 united nations' sustainable development goals: A status by 2020. *International Journal of Sustainable Development & World Ecology*, 29(3):219–229, 2022.

[Chen *et al.*, 2024] Rujian Chen, Gang Liu, Yisheng Cao, Gang Xiao, and Jianchao Tang. Cgaformer: Multi-scale feature transformer with mlp architecture for short-term photovoltaic power forecasting. *Energy*, 312:133495, 2024.

[Dolara *et al.*, 2015] Alberto Dolara, Sonia Leva, and Giampaolo Manzolini. Comparison of different physical models for pv power output prediction. *Solar energy*, 119:83–99, 2015.

[Hajjaj *et al.*, 2023] Charaf Hajjaj, Massaab El Ydrissi, Alae Azouzoute, Ayoub Oufadel, Omaima El Alani, Mohamed Boujoudar, Mounir Abraim, and Abdellatif Ghennioui. Comparing photovoltaic power prediction: Ground-based measurements vs. satellite data using an ann model. *IEEE Journal of Photovoltaics*, 2023.

[Hannan and Kavalieris, 1986] Edward James Hannan and Laimonis Kavalieris. Regression, autoregression models. *Journal of Time Series Analysis*, 7(1):27–49, 1986.

[He *et al.*, 2016] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[Ho and Xie, 1998] Siu Lau Ho and Min Xie. The use of arima models for reliability forecasting and analysis. *Computers & industrial engineering*, 35(1-2):213–216, 1998.

[Iheanetu, 2022] Kelachukwu J Iheanetu. Solar photovoltaic power forecasting: A review. *Sustainability*, 14(24):17005, 2022.

[Kallio-Myers *et al.*, 2020] Viivi Kallio-Myers, Aku Riihelä, Panu Lahtinen, and Anders Lindfors. Global horizontal irradiance forecast for finland based on geostationary weather satellite data. *Solar Energy*, 198:68–80, 2020.

[Li *et al.*, 2024] Duantengchuan Li, Tao Xia, Jing Wang, Fobo Shi, Qi Zhang, Bing Li, and Yu Xiong. Sdformer: A shallow-to-deep feature interaction for knowledge graph embedding. *Knowledge-Based Systems*, 284:111253, 2024.

[Liu *et al.*, 2023a] Weihong Liu, Jiawei Geng, Zongwei Zhu, Yang Zhao, Cheng Ji, Changlong Li, Zirui Lian, and Xuehai Zhou. Ace-sniper: cloud–edge collaborative scheduling framework with dnn inference latency modeling on heterogeneous devices. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 43(2):534–547, 2023.

[Liu *et al.*, 2023b] Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. itransformer: Inverted transformers are effective for time series forecasting. *arXiv preprint arXiv:2310.06625*, 2023.

[Luo *et al.*, 2023] Hao Luo, Johannes Quaas, and Yong Han. Examining cloud vertical structure and radiative effects from satellite retrievals and evaluation of cmip6 scenarios. *Atmospheric Chemistry and Physics*, 23(14):8169–8186, 2023.

[Miyamoto *et al.*, 2021] Ayumu Miyamoto, Hisashi Nakamura, Takafumi Miyasaka, and Yu Kosaka. Radiative impacts of low-level clouds on the summertime subtropical high in the south indian ocean simulated in a coupled general circulation model. *Journal of Climate*, 34(10):3991–4007, 2021.

[Molero *et al.*, 2022] Francisco Molero, Rubén Barragán, and Begoña Artíñano. Estimation of the atmospheric boundary layer height by means of machine learning techniques using ground-level meteorological data. *Atmospheric Research*, 279:106401, 2022.

[Nie *et al.*, 2022] Yuqi Nie, Nam H Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. *arXiv preprint arXiv:2211.14730*, 2022.

[Organization, 2017] World Meteorological Organization. Classification of clouds. http://www.wmo.int/, 2017.

[Qin *et al.*, 2022] Jun Qin, Hou Jiang, Ning Lu, Ling Yao, and Chenghu Zhou. Enhancing solar pv output forecast by integrating ground and satellite observations with deep learning. *Renewable and Sustainable Energy Reviews*, 167:112680, 2022.

[Rajagukguk *et al.*, 2020] Rial A Rajagukguk, Raden AA Ramadhan, and Hyun-Jin Lee. A review on deep learning models for forecasting time series data of solar irradiance and photovoltaic power. *Energies*, 13(24):6623, 2020.

[Ratings, 2024] SP Global Ratings. Rising curtailment in china: Power producers will push past the pain, 2024. Accessed: 2025-05-20.

[Saeed and Siraj, 2024] Syed Saeed and Tanvir Siraj. Global renewable energy infrastructure:: Pathways to carbon neutrality and sustainability. *Solar Energy and Sustainable Development Journal*, 13(2):183–203, 2024.

[Son *et al.*, 2023] Yongju Son, Xuehan Zhang, Yeunggurl Yoon, Jintae Cho, and Sungyun Choi. Lstm–gan based cloud movement prediction in satellite images for pv forecast. *Journal of Ambient Intelligence and Humanized Computing*, 14(9):12373–12386, 2023.

[State Council Office of the People's Republic of China, 2020] State Council Office of the People's Republic of China. Energy in china's new era, December 2020. Accessed: 2025-05-19.

[Stein *et al.*, 2012] Joshua S Stein, Clifford W Hansen, and Matthew J Reno. Global horizontal irradiance clear sky models: implementation and analysis. Technical report, Sandia National Laboratories (SNL), Albuquerque, NM, and Livermore, CA . . . , 2012.

[Swami *et al.*, 1994] Ananthram Swami, Georgios Giannakis, and Sanyogita Shamsunder. Multichannel arma processes. *IEEE Transactions on Signal Processing*, 42(4):898–913, 1994.

[Tan *et al.*, 2024] Ling Tan, Ruixing Kang, Jingming Xia, and Yue Wang. Application of multi-source data fusion on intelligent prediction of photovoltaic power. *Solar Energy*, 277:112706, 2024.

[van de Beek, 2024] Liselotte van de Beek. Performance analysis of solar micro-grids in rural developing areas. 2024.

[Wang *et al.*, 2016] Lunche Wang, Ozgur Kisi, Mohammad Zounemat-Kermani, Germán Ariel Salazar, Zhongmin Zhu, and Wei Gong. Solar radiation prediction using different techniques: model evaluation and comparison. *Renewable and Sustainable Energy Reviews*, 61:384–397, 2016.

[Wang *et al.*, 2020] Zihan Wang, Jiaxin Li, Jing Liu, and Chuanmin Shuai. Is the photovoltaic poverty alleviation project the best way for the poor to escape poverty?——a dea and gra analysis of different projects in rural china. *Energy Policy*, 137:111105, 2020.

[Wang *et al.*, 2024a] Kai Wang, Shuo Shan, Weijing Dou, Haikun Wei, and Kanjian Zhang. A robust photovoltaic power forecasting method based on multimodal learning using satellite images and time series. *IEEE Transactions on Sustainable Energy*, 2024.

[Wang *et al.*, 2024b] Yufei Wang, Xianzhe Wang, Deyang Hao, Yiyan Sang, Hua Xue, and Yang Mi. Combined ultra-short-term prediction method of pv power considering ground-based cloud images and chaotic characteristics. *Solar Energy*, 274:112597, 2024.

[Wazirali *et al.*, 2023] Raniyah Wazirali, Elnaz Yaghoubi, Mohammed Shadi S Abujazar, Rami Ahmad, and Amir Hossein Vakili. State-of-the-art review on energy and load forecasting in microgrids using artificial neural networks, machine learning, and deep learning techniques. *Electric power systems research*, 225:109792, 2023.

[Wu *et al.*, 2022] Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, and Mingsheng Long. Timesnet: Temporal 2d-variation modeling for general time series analysis. *arXiv preprint arXiv:2210.02186*, 2022.

[Xie *et al.*, 2024] Chen Xie, Yunbing Wei, Sifang Wang, and Jiankang Liu. Research on photovoltaic power prediction based on deep learning. In *2024 The 9th International Conference on Power and Renewable Energy (ICPRE)*, pages 1348–1353. IEEE, 2024.

[Yan *et al.*, 2017] Jinyue Yan, Yongping Zhai, Priyantha Wijayatunga, Abdul Matheen Mohamed, and Pietro Elia Campana. Renewable energy integration with mini/microgrids, 2017.

[Yang *et al.*, 2023] Jing Yang, Hui He, Xuemeng Zhao, Jue Wang, Tiechui Yao, Haizhou Cao, and Meng Wan. Day-ahead pv power forecasting model based on fine-grained temporal attention and cloud-coverage spatial attention. *IEEE Transactions on Sustainable Energy*, 2023.

[Yao *et al.*, 2021] Tiechui Yao, Jue Wang, Haoyan Wu, Pei Zhang, Shigang Li, Ke Xu, Xiaoyan Liu, and Xuebin Chi. Intra-hour photovoltaic generation forecasting based on multi-source data and deep learning methods. *IEEE Transactions on Sustainable Energy*, 13(1):607–618, 2021.

[Zach *et al.*, 2007] Christopher Zach, Thomas Pock, and Horst Bischof. A duality based approach for realtime tv-l 1 optical flow. In *Pattern Recognition: 29th DAGM Symposium, Heidelberg, Germany, September 12-14, 2007. Proceedings 29*, pages 214–223. Springer, 2007.

[Zeng *et al.*, 2023] Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. Are transformers effective for time series forecasting? In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pages 11121–11128, 2023.

[Zhang *et al.*, 2024] Yuanyuan Zhang, Huiru Zhao, Ze Qi, and Bingkang Li. A two-stage low-carbon economic coordinated dispatching model for generation-load-storage resources considering flexible supply-demand balance. *Applied Energy*, 373:123981, 2024.