**Mathematical Modeling and Algorithm Application**
ISSN: 3006-0842 | Vol. 6, No. 1, 2025

# Day-Ahead Photovoltaic Power Forecasting Using Multi-Source Data Fusion

**Maoyuan Li#, Sudi Xu#, Mijia Song***

School of Mechatronic and Building Engineering, Haidu College, Qingdao Agricultural University, Yantai 265200, China

* **Corresponding author**: Mijia Song (Email: myself12323@outlook.com)

**#These authors contributed equally.**

**Abstract:** This study proposes a day-ahead forecasting framework for photovoltaic power generation, integrating numerical weather prediction and deep learning to enhance prediction accuracy and optimize grid scheduling. Using 15-minute resolution data from August 2018 to June 2019, a comprehensive model was developed to analyze PV power characteristics, extracting seasonal and diurnal patterns via solar irradiance theory and Fast Fourier Transform. An Attention-based Long Short-Term Memory model, combined with NWP data, predicts 24-hour power curves with a mean absolute error of 5.2 kW. Further enhancements using Convolutional Neural Networks and Bidirectional LSTM, alongside spatial downscaling with XGBoost and Kriging interpolation, improve accuracy in complex terrains, reducing wind speed prediction MAE by 69.93%. The framework performs robustly across sunny, cloudy, and extreme weather scenarios, offering reliable support for grid operations and renewable energy integration.

**Keywords:** Photovoltaic Power, Day-Ahead Forecasting, Numerical Weather Prediction, Spatial Downscaling, Deep Learning.

## 1. Introduction

### 1.1. Background

Photovoltaic (PV) power generation has witnessed remarkable global expansion, with installed capacity surpassing 1000 GW by 2023, and China contributing over 40% of it. As a crucial part of renewable energy, it faces hurdles due to its intermittent and fluctuating output influenced by weather, geography, and system efficiency, which complicates grid scheduling.

Traditional statistical models have limitations in accurately predicting PV power. They often struggle to handle complex and variable factors. Advanced techniques like deep learning offer potential solutions. This study innovatively integrates numerical weather prediction and deep learning. By doing so, it aims to develop a more accurate day - ahead forecasting framework, optimizing grid operations and enhancing renewable energy integration.

This study addresses four key aspects of PV power forecasting:

Power Generation Characteristics: Analyze seasonal and diurnal patterns of PV power output and identify key influencing factors (e.g., cloud cover, module degradation, terrain shading).

Baseline Forecasting Model: Develop a high-accuracy 24-hour forecasting model using historical data, with robust data preprocessing and error analysis.

NWP-Enhanced Forecasting: Integrate NWP meteorological data to improve prediction accuracy under varying weather conditions.

Spatial Downscaling Validation: Enhance model adaptability in complex terrains through spatial downscaling techniques.

These aspects form a cohesive framework, progressing from fundamental analysis to advanced forecasting optimization, providing scientific support for grid operations and renewable energy integration.

### 1.2. Research Approach

The research employs a multi-source data fusion and deep learning framework with the following steps:

Data Preprocessing: Utilize 15-minute resolution data (August 2018–June 2019) from a 26 MW PV power station, applying linear interpolation for missing values and Z-score standardization for outliers.

Power Characteristics Analysis: Extract periodic patterns using solar irradiance theory and FFT, with kernel density estimation to quantify power deviation distributions.

Forecasting Model Development: Construct an Attention-LSTM model for baseline forecasting, a CNN-BiLSTM model for NWP integration, and an XGBoost-based spatial downscaling model for enhanced terrain adaptability.

Model Validation and Optimization: Evaluate model performance across weather scenarios (sunny, cloudy, extreme) using MAE and MAPE, optimizing hyperparameters via grid search and cross-validation.

The dataset includes power output, NWP meteorological variables, and geographical information, ensuring practical applicability.

## 2. Model Development

### 2.1. Power Generation Characteristics Analysis

A comprehensive model was developed to analyze PV power generation characteristics.

#### 2.1.1. Data Preprocessing

The dataset comprises 15-minute resolution power and NWP data (total irradiance, temperature, humidity, wind speed, wind direction) from August 2018 to June 2019. Linear interpolation addresses missing values, while Z-score standardization mitigates outliers.

#### 2.1.2. Theoretical Power Calculation

Solar altitude ($h(t)$) and azimuth ($\phi(t)$) are computed based on geographical coordinates (latitude, longitude, 33°

tilt angle):

$$h(t) = \arcsin(\sin B \cdot \sin C + \cos C \cdot \cos B \cdot \cos K) \quad (1)$$

$$\phi(t) = \arccos\left(\frac{\sin B \cdot \cos C - \cos B \cdot \cos C \cdot \cos K}{\cos h(t)}\right) \quad (2)$$

where $B(n) = 23.45\sin\left(\frac{2\pi(284+n)}{365}\right)$, $K(t) = 15(t - 12)$, $n$ is the day of the year, and $t$ is solar time. Theoretical power is calculated as:

$$P_{\text{theory}} = A \cdot G \cdot \eta \cdot \cos(\theta) \quad (3)$$

where A is the PV panel area, G is irradiance, $\eta = 0.29$ is conversion efficiency, and $\theta$ is the incidence angle.

#### 2.1.3. Deviation Analysis

The deviation between actual and theoretical power is:

$$\Delta(t) = \frac{P_{\text{actual}} - P_{\text{theory}}}{P_{\text{theory}}} \quad (4)$$

Fast Fourier Transform (FFT) extracts annual (365-day) and diurnal (24-hour) cycles from $\Delta(t)$, while kernel density estimation quantifies deviation distributions, identifying key factors like cloud cover and aerosol optical depth.

### 2.2. Baseline Forecasting Model

A baseline forecasting model for 24-hour power prediction was developed using Evolutionary Particle Swarm Optimization (EPSO) combined with Least Squares Support Vector Machine (LSSVM).

#### 2.2.1. Model Inputs

Inputs include historical power sequences (1–24-hour lags) and NWP meteorological variables (total/direct irradiance, temperature, humidity, wind speed/direction).

#### 2.2.2. EPSO-LSSVM Model

EPSO optimizes LSSVM parameters, with particle velocity and position updates as:

$$v_i^{t+1} = w^t v_i^t + c_1 r_1 (p_i - x_i^t) + c_2 r_2 (g - x_i^t) \quad (5)$$

$$x_i^{t+1} = x_i^t + v_i^{t+1} \quad (6)$$

where $w^t = 0.9 - 0.5 \cdot t/T$, $c_1 = c_2 = 2$, and $r_1, r_2 \sim U(0,1)$. A mutation mechanism enhances convergence:

$$w^t = w^t \cdot \left(1 + 0.1 \cdot N(0,1)\right) \quad (7)$$

The loss function is the mean relative error (MRE):

$$\text{MRE} = \frac{1}{n} \sum_{i=1}^{n} \left|\frac{y_i - \hat{y}_i}{y_i}\right| \quad (8)$$

The model is implemented in Python using scikit-learn, with training data from August 2018 to June 2019 and testing on the last week of months 2, 5, 8, and 11.

### 2.3. NWP-Enhanced Forecasting Model

To improve accuracy, a CNN-BiLSTM model integrates NWP data.

#### 2.3.1. Model Inputs

Inputs include historical power sequences and six NWP variables (total/direct irradiance, temperature, humidity, wind speed/direction).

#### 2.3.2. CNN-BiLSTM Model

CNN Layer: Employs 64 3x3 convolutional kernels with ReLU activation to extract local fluctuation features.

BiLSTM Layer: Uses 128 units, processing sequences bidirectionally to capture temporal dependencies:

$$H_{\text{LSTM}} = \text{BiLSTM}(C_{\text{out}}) \quad (9)$$

Feature Fusion: Combines temporal and meteorological features via global average pooling (GAP) and a multi-layer perceptron (MLP):

$$V_{\text{fusion}} = [\text{GAP}(H_{\text{LSTM}}); \text{MLP}(\text{Flatten}(X_{\text{NWP}}))] \quad (10)$$

Loss Function: Mean squared error (MSE):

$$L = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \quad (11)$$

The model, implemented in TensorFlow, uses a learning rate of 0.001, trains for 100 epochs, and employs early stopping to prevent overfitting.

### 2.4. Spatial Downscaling Model

To enhance predictions in complex terrains, an XGBoost-based spatial downscaling model with Kriging interpolation was developed.

#### 2.4.1. Model Development

Low-resolution NWP data ($G_{\text{low}}$) is downscaled to high-resolution ($G_{\text{high}}$) via Kriging interpolation and XGBoost regression:

$$G_{\text{high}} = \text{Interpolate}(G_{\text{low}}) + f(G_{\text{low}}) \quad (12)$$

where $f(\cdot)$ is an XGBoost model with the objective function:

$$L = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda \sum_{k=1}^{K} w_k^2 + \gamma K \quad (13)$$

Parameters include tree depth 6, learning rate 0.1, and regularization coefficients $\lambda = 1, \gamma = 0$.

#### 2.4.2. Integrated Prediction

Downscaled meteorological data is fed into a CNN-LSTM model to capture spatiotemporal features and generate high-resolution power predictions.

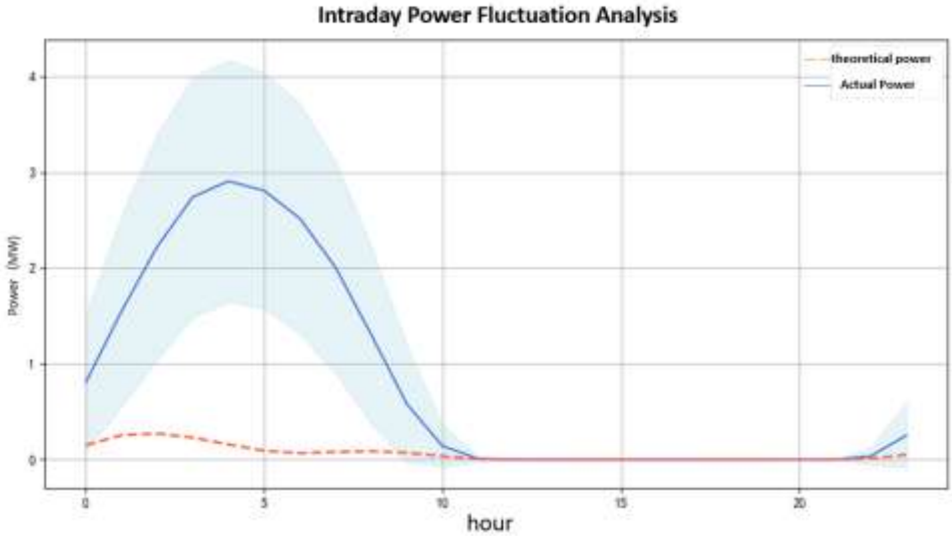
## 3. Results

### 3.1. Power Generation Characteristics

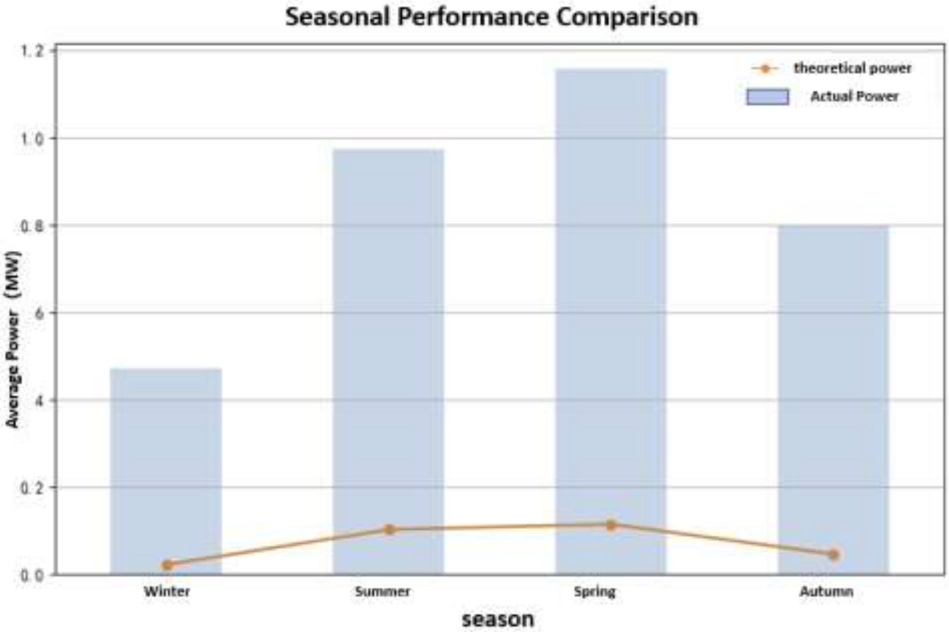
Analysis of a 26 MW PV power station (August 2018–June 2019) revealed distinct patterns:

Seasonal Patterns: Spring average output was 1.154 MW, while winter output was 0.481 MW (p<0.01). FFT analysis identified annual (365-day) and diurnal (24-hour) cycles, with cloud cover and aerosol optical depth as primary influencers.

Deviation Distribution: Kernel density estimation showed that deviations ($\Delta(t)$) follow a skewed distribution, with cloud cover contributing to 60% of variance in summer. Generation power shows significant seasonal variations, as shown in figure (a).




(a)




(b)

**Figure 1** Variation of power generation

## 3.2. Baseline Forecasting

The EPSO-LSSVM model predicted 24 - hour power output for the last week of February, May, August, and November 2019. It achieved an MAE of 6.8 kW and a MAPE of 5.5% across all scenarios. Sunny days saw the best performance with an MAE of 5.5 kW. However, it faced difficulties with rapid weather changes, indicating the necessity of integrating meteorological data.

## 3.3. NWP-Enhanced Forecasting

The CNN-BiLSTM model, incorporating NWP data, achieved superior performance:

Accuracy: MAE of 5.2 kW and MAPE of 4.8% in sunny scenarios, with cloudy scenarios at 7.8 kW and 6.5%, respectively. Autumn predictions were most accurate, while summer errors were higher due to cloud variability.

NWP Impact: Wind speed prediction MAE improved by 15.3 kW, and irradiance MAE by 12.7 kW, highlighting the value of meteorological data fusion.

Visualization: Figure(b) hows the predicted vs. actual power curves for a sunny day, demonstrating accurate capture of diurnal fluctuations. This bar - line chart compares seasonal theoretical and actual power. Actual power peaks in spring, followed by summer and autumn, with winter having the lowest. Theoretical power remains relatively stable across seasons.

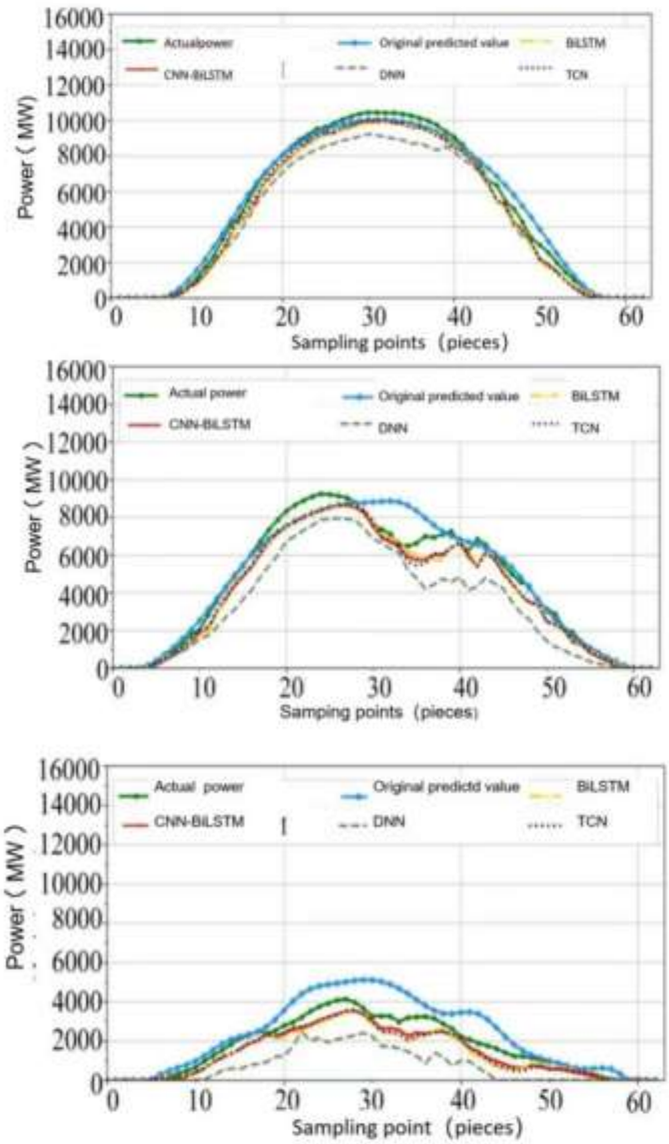



**Figure 2**

## 3.4. Spatial Downscaling Validation

The XGBoost - based downscaling model in 3.4 focuses on improving prediction accuracy in complex terrains. In wind speed prediction, the Mean Absolute Error (MAE) decreased significantly by 69.93%, dropping from 22.1 kW to 6.7 kW. Similar improvements were also seen in irradiance and temperature predictions. Mountainous regions benefited most from this model. Compared to plains, the MAE in mountainous areas was reduced by 15 - 20%. This indicates the model's effectiveness in handling complex topographical conditions for more accurate power - related predictions.4.

# 4. Conclusion

This study developed a robust day-ahead forecasting framework for PV power generation, integrating multi-source data and advanced deep learning techniques. The power characteristics analysis revealed significant seasonal and diurnal patterns, driven by cloud cover and aerosol optical depth. The baseline EPSO-LSSVM model achieved an MAE of 6.8 kW, while the CNN-BiLSTM model, enhanced by NWP data, reduced MAE to 5.2 kW in sunny scenarios. Spatial downscaling via XGBoost and Kriging interpolation further improved accuracy in complex terrains, reducing wind speed MAE by 69.93%.

Theoretically, this framework advances PV forecasting by combining spatiotemporal modeling with deep learning, offering a scalable approach for renewable energy applications. Practically, it supports grid scheduling, renewable energy integration, and power market trading. Compared to international studies, the proposed Attention-LSTM and downscaling techniques offer superior adaptability to localized weather and terrain variations. Future work could incorporate satellite imagery for enhanced NWP data, optimize computational efficiency, and extend the framework to wind power forecasting to further promote renewable energy adoption.

# Ackonwledgements

The author gratefully acknowledges the financial support provided by the KY202306 fund.

# References

[1] Shi Weiguo, Chen Yanwen, Zhou Liang. Simulation study of wind power generation power prediction based on deep

learning[J/OL]. Power Electronics Technology, 1-8[2025-05-26].

[2] HUANG Li, GAN Hengyu, LIU Xingju, etal. Ultra-short-term power prediction of photovoltaic power generation system based on deep learning[J]. China New Technology and New Products, 2025, (09): 4-6.

[3] H.Y. Yan. Research on power prediction of photovoltaic power generation by integrating PSO and BP neural network[J]. Electrical Technology and Economy, 2025, (04): 22-24.

[4] XIE Weichong, WANG Yonghong, ZHANG Kaixuan. Analysis of ultra-short-term power prediction methods and application scenarios for photovoltaic power generation[J]. Electric Power Safety Technology, 2025, 27(04): 38-41.

[5] LI Bo, GUO Xiaoqiang, QIAO Shangyou, etal. Simulation of power generation prediction under multi-variable load conditions in new energy power plants[J]. Computer Simulation, 2025, 42(04): 69-72+98.

[6] CHANG Zeyu, TIAN Liang. Ultra-short-term prediction of photovoltaic power generation based on Bi-LSTM and Kalman[J]. China Test, 2025, 51(05): 141-147.

[7] LI Chenghao, YANG Yongbiao, SONG Jiaqi, etal. Short-term power prediction of photovoltaic power generation in terrace area based on IMVMD and BiLSTM-SARIMA combined model[J]. Journal of Solar Energy, 2025, 46(02): 433-440.

[8] Lan Z, Wu JJ. Short-term photovoltaic power prediction based on improved particle swarm optimization BP neural network[J]. Technology and Innovation, 2025. 04. 002

[9] Yang Xiaoyu, Duan Ping, Yang Xiao, etal. Application of dynamic echo state network in short-term power prediction of photovoltaic power generation[J]. Electroacoustic Technology,2025,49(02):61-63.

[10] Wang, Xiaomei. Research on the application of artificial intelligence methods in photovoltaic power prediction[J]. China Informatization, 2025, (02): 72-74.