EAI Endorsed Transactions
on Energy Web

Research Article EAI.EU

Distributed photovoltaic power prediction considering spatiotemporal correlation and dual Attention-LSTM

Yueyuan Zhang[1], Yifan Zhang[2], Dazhi Pan[3], Mingyu Sun[4,*], Huawei Mei[4], and Wangbin Cao[4]

[1]Inner Mongolia Power (Group) Limited Liability Company Marketing Services, Ordos 017010, China
[2]Inner Mongolia Electric Power Research Institute, Inner Mongolia Key Laboratory of Smart Grid of New-type Power System, Ordos 017010, China
[3]Inner Mongolia Power (Group) Co., Ltd., Inner Mongolia Enterprise Key laboratory of Smart Grid Simulation of Electrical Power System, Ordos 017010, China
[4]North China Electric Power University (Baoding), Baoding 071000, China

Abstract

Predicting the power output of photovoltaic clusters is crucial for optimizing regional solar power scheduling. To enhance the accuracy of distributed photovoltaic station power forecasts, a method incorporating spatiotemporal correlation and dual Attention-LSTM is introduced. First, the K-means algorithm is used to classify the distributed photovoltaic power plant clusters in the region based on the photovoltaic power curve. The reference station for the target photovoltaic station is determined by calculating the Euclidean distance between the target station and the typical daily power profiles of other stations in the cluster. Simultaneously, pivotal weather features that influence photovoltaic output are ascertained through computation of the Pearson correlation coefficient. Subsequently, an Attention-LSTM-based power prediction and error correction model is constructed, utilizing both meteorological and power traits as input variables to finalize the photovoltaic power generation forecast. To validate the approach, a simulation analysis is conducted using empirical data from Arizona, USA. The experimental results indicate that the suggested method greatly improves the accuracy of predictions for distributed photovoltaic power.

Keywords: Distributed Photovoltaic, Power Prediction, Feature Fusion, K-means Algorithm, Attention-LSTM Model.

Received on 11 October 2024, accepted on 27 February 2025, published on 06 October 2025

doi: 10.4108/ew.7530

1. Introduction

In alignment with China's objectives for "carbon peak and carbon neutrality," and the ongoing advancement of energy transition strategies[1], renewable energy sources, predominantly solar (photovoltaic) and wind, are witnessing sustained growth with rising installation capacities annually. Moreover, distributed photovoltaic (PV) power generation has emerged as a significant direction in the swift evolution of solar energy production in recent times. A distributed photovoltaic system involves installing solar power generation equipment on rooftops of buildings, industrial zones, residential areas, and other decentralized locations. Initially, we apply the K-means algorithm to classify clusters of distributed photovoltaic power stations in the region. We identify the reference station for the target photovoltaic station by calculating the Euclidean distance to the typical daily power profiles of other cluster stations. The experimental results show that this method greatly enhances the accuracy of photovoltaic power predictions. The system

*Corresponding author. Email: 220222221110@ncepu.edu.cn

generates electricity for self-consumption, feeding any excess back into the grid. Currently, distributed photovoltaic power generation is extensively implemented and promoted worldwide. Numerous countries and regions have introduced various policies and initiatives to foster the growth of distributed PV, aiming to reduce reliance on traditional energy sources and enhance the use of sustainable energy. In the context of a modern power system, the large-scale integration of distributed PV into the distribution network can change network currents, creating challenges for grid management. If regional distribution networks cannot handle the randomness and variability of PV generation, it may result in voltage fluctuations and reduce the capacity for PV consumption. Thus, precise short-term power forecasting for distributed PV is essential for optimizing power system operations and ensuring the grid operates safely[2].

Extensive research on centralized photovoltaic (PV) forecasting has been conducted both domestically and internationally. For example, literature[3] utilizes Principal Component Analysis (PCA) to reduce the dimensionality of multivariate data series that are highly correlated with PV power. The resulting principal component data series are then used to develop a short-term PV power prediction model based on the Long-Short Term Memory (LSTM) neural network. Literature[4] in the Gate Recurrent Unit (GRU) network algorithm on the basis of a time-shared GRU ultra-short-term power prediction method, the model for each moment to establish a GRU model, a model predicts the power of a moment, to achieve time-shared photovoltaic power prediction. The above prediction methods all use a single prediction model, which has the problems of low prediction accuracy and poor stability, for this reason, a combination model is considered for prediction, and the combination prediction integrates each single prediction method, which can combine the advantages of each type of prediction method[5-6], and usually obtains a higher prediction accuracy. Extensive research into centralized PV power forecasting has been explored globally. Literature[7] utilized a Convolutional Neural Network (CNN) to extract spatial features from the data, followed by the application of Long-Short Term Memory (LSTM) for temporal feature extraction. This approach combines both methods with the Extreme Gradient Boosting (XGBoost) model through an error backpropagation technique, facilitating comprehensive forecasts of photovoltaic power generation. Another study[8] created a dataset of historical power output and correlated characteristics after applying wavelet noise reduction. It proposed a short-term PV power forecasting model that combines bidirectional Long-Short Term Memory (Bi-LSTM) and Random Forest (RF), resulting in high-precision, ultra-short-term power predictions. Furthermore, the Attention mechanism, known for highlighting the significance of input features, was incorporated into core neural network units to enhance model's generalizability[9-10].

However, traditional centralized PV power plants are usually established in specific geographical locations, and their power prediction is usually modeled based on time series. In contrast, distributed PV systems comprise numerous plants spread across various geographic locations, exhibiting spatial correlations between them. Considering both temporal and spatial correlations can enhance the accuracy of power predictions for distributed PV systems[11]. Literature[12] combines data from neighboring multi-user sources to enhance the sample size and introduces a similar day search method that considers power correlation and correlation weights. This approach is used to implement day-ahead forecasting with LSTM. On the other hand, Literature[13] employs meteorological data from nearby public weather stations and corrects for any discrepancies in this information to achieve accurate power predictions for the target site. Literature[14] performs gridded interpolation of meteorological resource data across a wide area and groups photovoltaic power stations with similar meteorological characteristics based on the interpolation results. This method is then integrated with LSTM to develop a dual migration model that shifts from the source domain to the target domain. The aforementioned methods tackle the issue of limited distributed PV data by enhancing it through geographic interpolation of meteorological resources or by fusing data from neighbouring PV stations. While these approaches improve the accuracy of distributed PV power predictions, they do not fully exploit the potential connections between the power sequences of nearby PV stations. Literature[15] utilizes the K-means clustering method, grouping stations based on the Euclidean distance of historical power generation data. It applies the Least Squares Support Vector Machine (LSSVM) model to analyse single-station predictions for spatially correlated power stations. Literature[16] develops a PV clustering method aimed at predicting power for large-scale distributed PV users. It selects PV power stations with spatial correlation to the target stations within the clusters and establishes an Autoregressive Moving Average (ARMA) model for power prediction of the identified PV stations. Literature[17] first employs an integrated XGBoost-LSTM model to forecast based on historical time series data of PV installations. It then utilizes the Least Squares Support Vector Machine (LSSVM) model to leverage spatial correlations between PV plants for predictive purposes. Ultimately, by evaluating the prediction errors from both models, it calculates the weights for each model. These weights are then combined to derive the forecasted power of the time-combined model. The method described above forecasts the power output of target PV plants by establishing a connection between the power series of target and reference PV plants. However, it relies solely on a single power characteristic as the model's input, overlooking the influence of weather features and additional power traits. Consequently, there remains potential for further enhancement of the model's predictive accuracy. Furthermore, the aforementioned methods either do not incorporate error correction models or utilize overly simplistic ones, which fail to adequately mitigate the impact of random errors and noise on model predictions. This results in a prediction model that is highly sensitive to data noise, lacking robustness and stability, and exhibiting significant errors under complex and fluctuating weather conditions.

Considering these factors, this study presents a short-term power forecasting methodology for distributed photovoltaic (PV) systems that utilizes spatiotemporal feature fusion and model error correction to improve the accuracy of distributed PV power predictions. Section 2 divides distributed photovoltaic clusters based on power curves. After division, each photovoltaic power station group has meteorological consistency. Section 3 uses the Pearson correlation coefficient to determine the key meteorological characteristics of the target photovoltaic power station, and selects the reference photovoltaic power station by calculating the Euclidean distance between the power sequence of the target power station and other photovoltaic power stations in the cluster, thereby completing the model input characteristics Build. Section 4 builds a power prediction model and error correction model based on Attention-LSTM. The weather characteristics of the target photovoltaic power station and the power of the reference photovoltaic power station are jointly input into the power prediction model for prediction, and the error correction model is used to reduce the error of the power prediction model forecast error. Experimental simulations are carried out in Section 5. The results indicate that the method presented in this paper significantly improves the accuracy of distributed PV power predictions.

2. Distributed PV cluster segmentation

2.1. Correlation between photovoltaic power and meteorological factors

The power production from photovoltaic (PV) systems is affected by various meteorological elements such as irradiance, temperature, humidity, wind velocity, and wind bearing. Reflecting the significant connection between PV power output and solar irradiance, the power curve of a PV plant can serve as an equivalent indicator of irradiance[18]. This represents the local light intensity, temperature, and additional meteorological details. By dividing the PV power station clusters based on the PV power history data, the data characteristics of each station within the cluster can be converged and have similar meteorological conditions. This categorization considers the sensitivity of the photovoltaic (PV) power system to solar irradiance and how PV modules respond under varying meteorological conditions. As a result, it allows for a better understanding and analysis of PV power plant operations. By grouping stations with similar meteorological characteristics, the power output of a PV cluster can be assessed and predicted more accurately, offering a more reliable reference for the operation and management of PV power systems.

2.2. Distributed PV cluster segmentation based on K-means algorithm

There are two main types of cluster partitioning algorithms commonly used today, one is cluster analysis (cluster analysis) algorithm and the other is association structure discovery (community detection) algorithm[19].

The K-means algorithm is a widely used technique for cluster analysis. In a scenario with n data points to be grouped into k clusters, the procedure starts by randomly choosing k points as the initial centroids. Subsequently, the remaining n-k data points are allocated to the closest cluster by minimizing their distance to the centroids. Following this allocation, the mean of all data points within each cluster is computed to update and establish the new centroids. The steps for dividing the clusters of distributed PV plants based on K-means algorithm are as follows:

Step 1: For each distributed PV power plant, the annual PV power data are standardized.

Step 2: Select n typical days and calculate the mean, standard deviation, coefficient of variation, kurtosis and skewness of their power, respectively, where the expressions for the mean, standard deviation, and maximum value are not repeated, and the coefficient of variation, kurtosis, and skewness are defined as:

$$C=\frac{\sigma}{\bar{X}} \tag{1}$$

$$K=\frac{\sum_{i=1}^{n}\left(Xi+\bar{X}\right)^{4}}{\left(n-1\right)\sigma} \tag{2}$$

$$S=\frac{\sum_{i=1}^{n}\left(Xi-\bar{X}\right)^{3}}{\left(n-1\right)\left(n-2\right)\sigma} \tag{3}$$

In the formula, Xi denotes the value of the i-th data point in the sample for the variable X; $\bar{X}$ represents the mean of the variable X; X indicates the sample size; and σ signifies the standard deviation of the variable.

Step 3: The mean, standard deviation, maximum, coefficient of variation, kurtosis and skewness of the n typical daily powers are used as clustering features and clustered using K-means clustering algorithm.

Step 4: Analyse the contour coefficients under different number of clusters, select the best clustering results, and complete the distributed PV power plant cluster division.

One of the contour coefficients is a metric used to evaluate the clustering results[20] and is used to measure the tightness and separation of the clusters. Where the definition of the profile coefficient $S\left(i\right)$ is shown below, the closer the value of the profile coefficient is to 1 the better the clustering effect of the model, and the closer it is to -1 the worse the clustering effect of the model.

$$S\left(i\right)=\frac{b\left(i\right)-a\left(i\right)}{max\left\{a\left(i\right),b\left(i\right)\right\}} \tag{4}$$

In the formula, $a(i)$ denotes the average distance from sample point i to other sample points in the same cluster;

$b(i)$ denotes the average distance from sample point i to all other sample points in some cluster.

3. Input feature construction based on spatio-temporal correlation models

3.1. Analysis of photovoltaic power influencing factors

PV power generation is affected by various meteorologic factors, such as irradiance, temperature, humidity, wind speed, and wind direction. Therefore, it is essential to analyse the correlations among these variables to identify the primary factors impacting PV power output[21]. Pearson coefficients can better reflect the degree of linear correlation between 2 random variables[22], and for the 2-length n of the data series *X* and *Y*, the Pearson correlation coefficient between series *X* and *Y* is expressed using $\rho_{X,Y}$, which is calculated as:

$$\rho_{X,Y} = \frac{\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\overline{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i-\overline{Y})^2}} \tag{5}$$

In the formula: X_i and Y_i represent the value of the *i*-th data point in the sample on variable *X* and variable *Y*, respectively; $\overline{X}$ and $\overline{Y}$ represent the mean values of variable *X* and variable *Y*; *n* represents the sample size.

3.2. Proximity to photovoltaic power plant selection

The Euclidean distance is a commonly used metric for calculating the straight-line distance between two points in a multidimensional space. Most traditional similar day theories identify a similar day to the predicted day by calculating the Euclidean distance between the attribute values[23]. Applying it to the comparison of power profiles among different PV plants, their differences can be assessed by calculating the Euclidean distances between typical daily power profiles of different PV plants. To identify the PV plant with the most similar power characteristics to the target plant, the reference PV plant is chosen based on the smallest Euclidean distance. This is done by calculating the Euclidean distances between the target PV plant and the *n* typical daily power curves of each plant within the cluster after it has been segmented. The formula for calculating the Euclidean distance is shown below:

$$dist(X,Y) = \sqrt{\sum_{i=1}^{n}(X_i - Y_i)^2} \tag{6}$$

In the formula: X_i and Y_i represent the values of the *i*-th data point in the sample on variable *X* and variable *Y* respectively

3.3. Model of input feature construction

Currently on the construction of power prediction model input features are only based on meteorological features or a single power feature, so in this paper, we construct an extended matrix that integrates meteorological features and power features:

$$F = \{W_1, W_2 \ldots Wn, Ppv\} \tag{7}$$

In the formula: $W_1, W_2 \ldots Wn$ is a key meteorological feature that affects PV output; Ppv predicted power for reference photovoltaic plants.

The construction method of specific model input features is as follows:

Step 1: Weather features with high impact on PV power are selected from all-weather features based on the method in Section 2.1.

Step 2: A reference PV plant for the target PV plant is selected based on the method in Section 2.2, and the predicted power of the reference PV plant is calculated based on the model proposed in Section 3.

Step 3: Combine the weather features of the target time period with the predicted power sequence of the reference PV power plant at the target time period to complete the construction of the model input features.

4. Attention-LSTM-based power prediction model and error correction model construction

4.1. LSTM model

LSTM is a modified Recurrent Neural Network (RNN)[24] that removes or adds information through forgetting gates, input gates, and output gates so that only the important information is retained, thus avoiding the phenomenon of gradient explosion that exists in RNNs.

4.2. Attention mechanism

The Attention mechanism is a widely used technique in machine learning[24] that mimics human attention. It allows models to selectively focus on the most relevant aspects of the input by dynamically assigning varying weights to different positions within the input sequence. In models using the Attention mechanism, the correlations between each location in the input sequence and the current output location are computed, and then these correlations are converted into weights that are used to weight the average input sequence representation. In this way, the model is able to adaptively focus more attention on the information that is more important to the current task, thus improving model performance.

4.3. Attention-LSTM-based power prediction model and error correction model construction

The structure of the Attention-LSTM model is shown in Figure 1. The LSTM network depicted in the figure consists of two LSTM layers, with a Dropout layer added after each to mitigate overfitting during training. The Attention mechanism enhances the LSTM model by dynamically learning and assigning weights, enabling the model to better capture relevant information from the input sequences that pertains to the prediction outcomes. Ultimately, the predictions are produced through a fully connected layer (Dense).

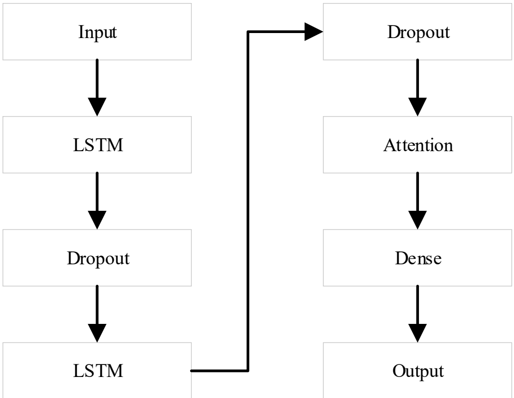


Figure 1. Attention-LSTM model structure

To improve the accuracy of distributed PV power prediction, this paper presents an error correction model based on Attention-LSTM, extending the existing Attention-LSTM power prediction model. The data is first divided into three groups: A, B, and C. The Attention-LSTM power prediction model is first trained using the data from group A. The trained model is then used to predict the power output for the data in group B. The data in group B is predicted using the Attention-LSTM power prediction model. Subsequently, the differences between the predicted values and the actual values for group B are calculated, and the Attention-LSTM error correction model is trained based on this sequence of differences. Finally, both the power prediction model and the error correction model are employed to predict the data in group C. The final prediction result for group C is obtained by summing the outputs from the power prediction model and the error correction model. The flow of power prediction with Attention-LSTM power prediction model and the error correction model is shown in Figure 2.

The main flow of the method proposed in this paper is shown in Figure 3. First, the K-means algorithm is used to divide the distributed photovoltaic power station groups in the region. Then, by calculating the Euclidean distance between the target photovoltaic power station and the typical daily power sequence of other photovoltaic power stations in the cluster, the reference photovoltaic power station of the target photovoltaic power station is selected. At the same time, the key meteorological characteristics that affect photovoltaic output are determined by calculating the Pearson coefficient. Finally, a power prediction model and error correction model based on Attention-LSTM were established, and meteorological characteristics and power characteristics were used as inputs to the model to complete the prediction of photovoltaic power generation.

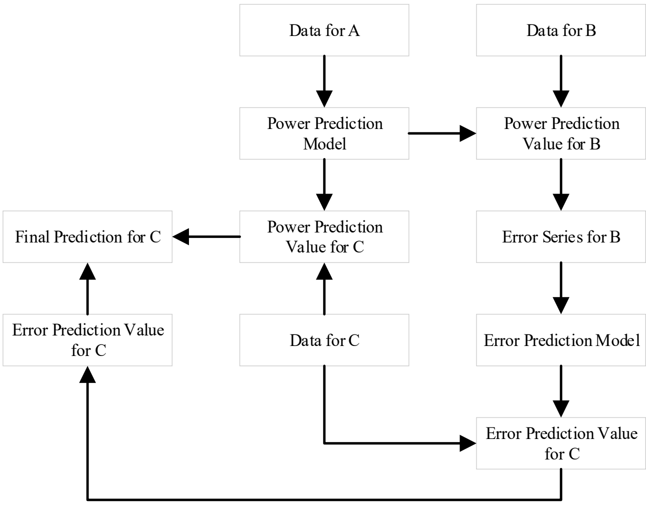


Figure 2. Power prediction process

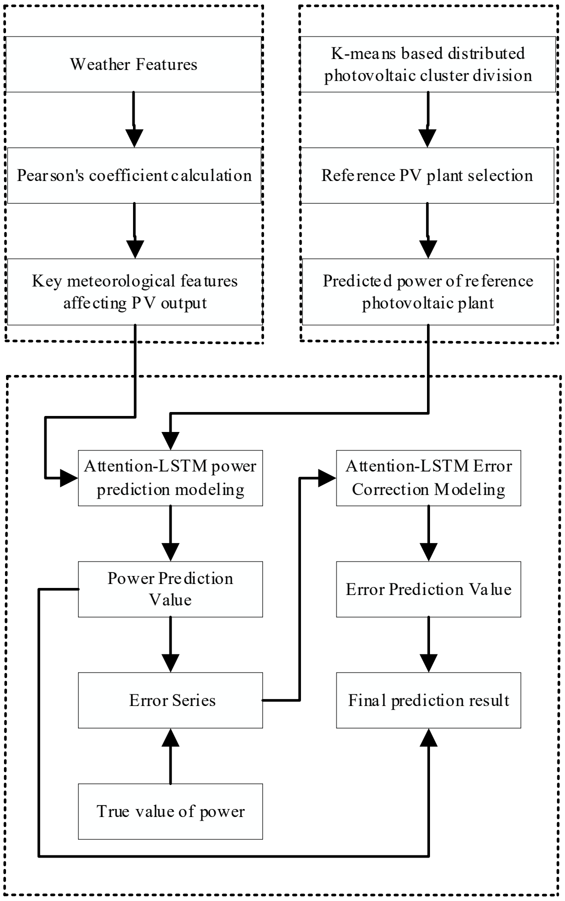


Figure 3. The main method flow of this paper

The main flow of the method proposed in this paper is shown in Figure 3. First, the K-means algorithm is used to divide the distributed photovoltaic power station groups in the region. Then, by calculating the Euclidean distance between the target photovoltaic power station and the typical daily power sequence of other photovoltaic power stations in the cluster, the reference photovoltaic power station of the target photovoltaic power station is selected. At the same time, the key meteorological characteristics that affect photovoltaic output are determined by calculating the Pearson coefficient. Finally, a power prediction model and error correction model based on Attention-LSTM were established, and meteorological characteristics and power characteristics were used as inputs to the model to complete the prediction of photovoltaic power generation.

5. Experimental results and analysis

In order to validate the effectiveness of using spatio-temporal fusion features as well as complex error correction models. This paper utilizes power data and meteorological data from 48 distributed PV power plants in Arizona, USA, covering the period from January 1, 2006, to December 31, 2006, for simulation and analysis. Each PV plant has an installed capacity ranging from 30 to 150 MW, with data sampled at 30-minute intervals, resulting in 48 observations collected daily. The dataset is divided into a training set and a test set in an 80% to 20% ratio.

5.1. Data preprocessing

(1) Outliers handling

The initial PV power and meteorological data have some outliers and missing values, which affect the accuracy and reliability of the data, resulting in the inability to use them directly for predictive analysis. Therefore, before prediction, the data need to be processed by data cleaning, so as to remove outliers and fill in missing values.

(2) Data standardization

In order to eliminate the differences in physical dimensions and accelerate the convergence speed of the model degree, The original PV power and meteorological data are normalized using the maximum-minimum method, ensuring that all values fall within the range of [0, 1] after normalization.

5.2. Distributed photovoltaic power plant cluster segmentation results

For each PV plant, May 15 (spring), August 15 (summer), November 15 (fall), and February 15 (winter) are selected as the feature days, and the mean, standard deviation, coefficient of variation, kurtosis, and skewness of the power of each feature day are computed respectively, and the power features of the four feature days are clustered as the cluster features. The contour coefficients corresponding to different classification numbers are shown in the table, and the results of cluster division of distributed PV are shown in Figure 4.

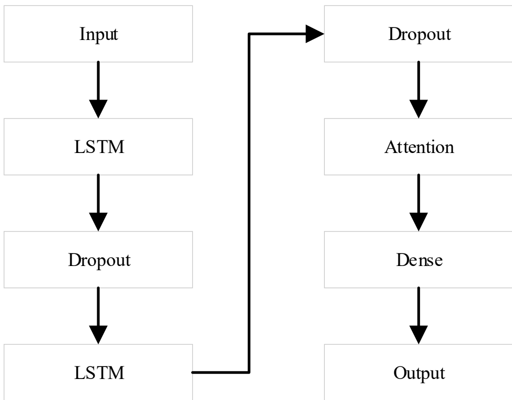


Figure 4. Results of distributed PV cluster division

Among the clusters, PV clusters 1, 2, and 3 consist of 18, 10, and 20 distributed PV power stations, respectively. It is clear that clusters formed based on PV power characteristics generally belong to the same geographic area. This indicates that the stations within these clusters can be viewed as smallscale PV power station groups sharing consistent meteorological conditions.

5.3. Distributed PV power prediction results

5.3.1. Analysis of influencing factors of photovoltaic power generation

Using the No. 1 power station in the No. 1 PV cluster as an example, Table 1 displays the Pearson correlation coefficients between the power station's output and various meteorological factors. It is evident that Global Horizontal Irradiance (GHI) has a very strong correlation with PV power output. Direct Normal Irradiance (DNI) and the solar zenith angle also exhibit strong correlations with PV power. Relative humidity shows a medium correlation, while Direct Horizontal Irradiance (DHI), wind speed, dew point, and temperature have weak correlations. Pressure, wind direction, and surface albedo have almost no correlation with PV power. Four features with high correlation, i.e., GHI, DNI, solar zenith angle, and relative humidity are selected in the text as the main meteorological output parameters affecting PV power.

Table 1. Pearson coefficient of each meteorological factor

Weather Characteristics	Pearson Factor	Weather Characteristics	Pearson Factor
DHI	0.192	Temperature	0.311
DNI	0.683	Pressure	-0.006
Dewpoints	-0.205	GHI	0.851
Surface Albedo	0.128	Wind Direction	0.062
Wind Speed	0.255	Solar Zenith Angle	-0.693
Relative Humidity	-0.476		

5.3.2. Reference PV plant selection results

Using PV cluster No. 1 as an example, for each PV plant in the cluster, the dates May 15 (spring), August 15 (summer), November 15 (fall), and February 15 (winter) are selected as typical days. PV No. 1 is designated as the target PV plant, and the sum of the Euclidean distances between the typical days of the target PV plant and those of the other PV plants in the cluster is presented in Table 2. From the table, it can be seen that the Euclidean distance between PV power plant No. 10 and the typical day of the target PV power plant is the smallest, so PV power plant No. 10 is selected as the reference power plant for PV power plant No. 1.

Table 2. Euclidean distance between meteorological stations

Power Station Number	European Distance /MW	Power Station Number	European Distance /MW
Power Station 2	3.516	Power Station 10	3.132
Power Station 3	6.133	Power Station 11	3.742
Power Station 4	3.467	Power Station 12	3.698
Power Station 5	3.348	Power Station 13	4.532
Power Station 6	3.298	Power Station 14	4.429
Power Station 7	3.459	Power Station 15	4.288
Power Station 8	3.845	Power Station 16	5.187
Power Station 9	3.272	Power Station 17	5.559

5.3.3. Predictive model performance evaluation and error analysis

(1) Evaluation indicators

Mean Absolute Error (MAE) and Mean Squared Error (MSE) are utilized as error evaluation metrics. The formulas for calculating the two indicators are shown below:

$$MAE=\frac{1}{n}\sum_{i=1}^{n}\left|yi-\hat{y}i\right| \tag{8}$$

$$MSE=\frac{1}{n}\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2 \tag{9}$$

In the formula: y_i stands for the value of the i-th data point in the sample at the true value y , respectively; $\hat{y}_i$ stands for the value of the ith data point in the sample at the predicted value $\hat{y}$, respectively; and n stands for the number of samples.

(2) Comparative analysis of different input features

To validate the effectiveness of using spatio-temporal information fusion features as model inputs, the prediction model incorporating these features is compared with models that utilize only meteorological or power features. To ensure the objectivity of the comparison, all experiments employ the Attention-LSTM-based power prediction model along with the error correction model.

As an example, the experimental results of power station No. 1 in PV cluster No. 1 are shown in Figure 5 and Table 3. Figure 5 demonstrates the comparison of PV power prediction curves randomly selected from November 17 (sunny), November 12 (cloudy), and November 28 (rainy) from 7:00 to 19:00, and Table 3 demonstrates the comparison of the errors under different input characteristics.

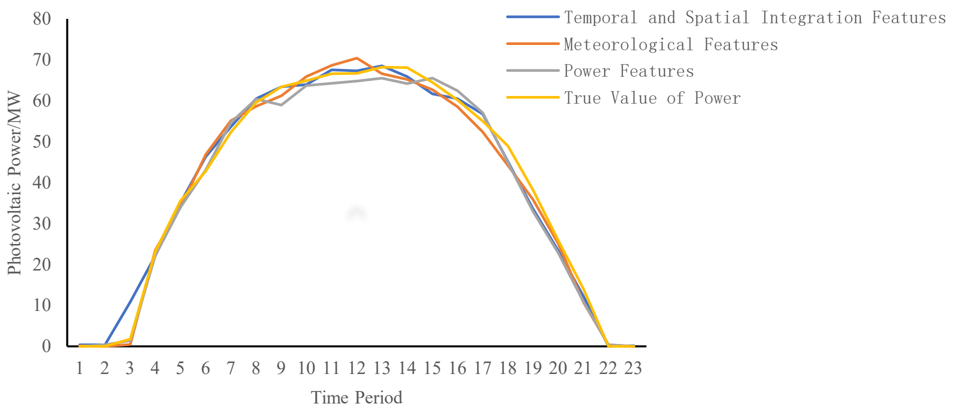
Table 3. Error comparison of different input features

Input Features	MAE/MW	MSE/MW
Temporal and Spatial Integration Features	2.71	16.72
Meteorological Feature	2.93	18.91
Power Characteristics	3.26	21.33
Meteorological Feature	2.93	18.91

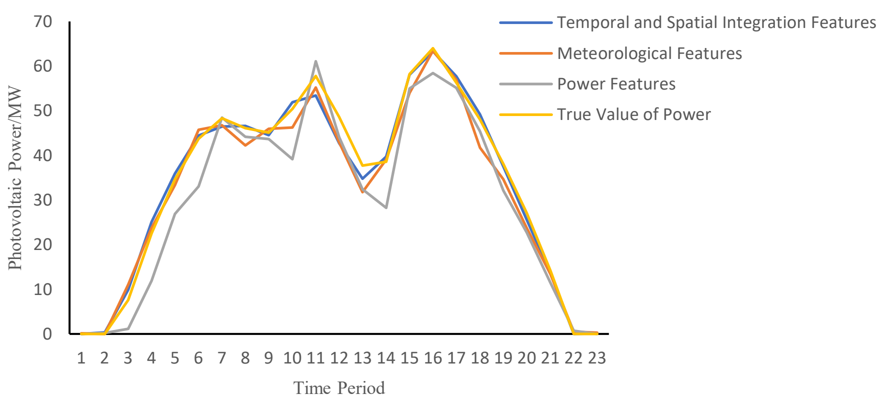
From Table 3, it can be seen that the MAE and MSE of the method proposed in this paper with spatio-temporal information fusion features as model input features are 2.71 MW and 16.72 MW, respectively. In comparison to the meteorological features., the MAE and MSE are reduced by 7.51% and 11.58%, respectively; compared with the power features, the MAE and MSE are reduced by 16.87% and 21.61%, respectively. It is evident that utilizing spatiotemporal information fusion features as model inputs can significantly enhance the accuracy of distributed PV power prediction.

(3) Comparative analysis of the results of different error correction models

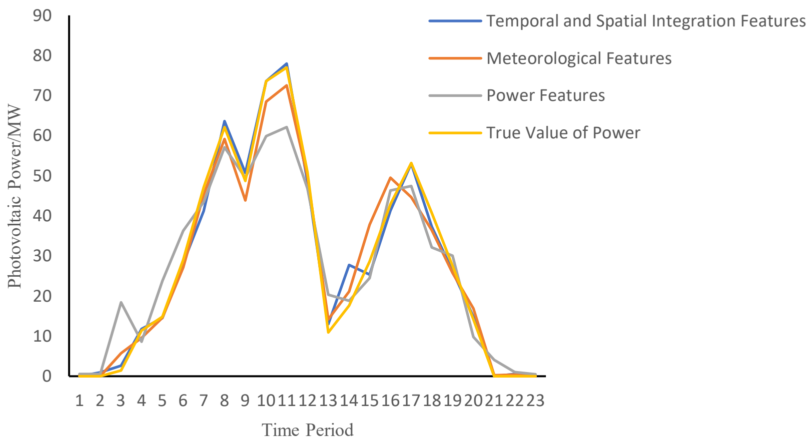
To verify the effectiveness of the Attention-LSTM error correction model, a comparative analysis will be conducted among the Attention-LSTM error correction model, the XGBoost error correction model, and a scenario without any error correction. To ensure objectivity, all experiments will use the Attention-LSTM power prediction model, with spatiotemporal information fusion features as input.



(A) sunny



(B) cloudy



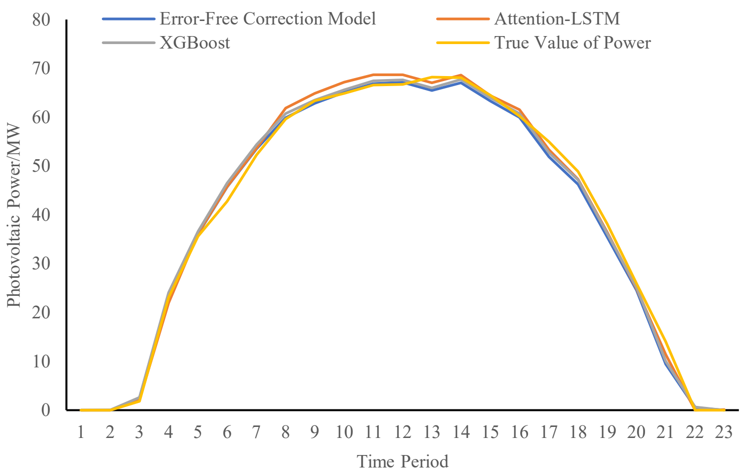
(C) rainy

Figure 5. Comparison of prediction results under different input characteristics

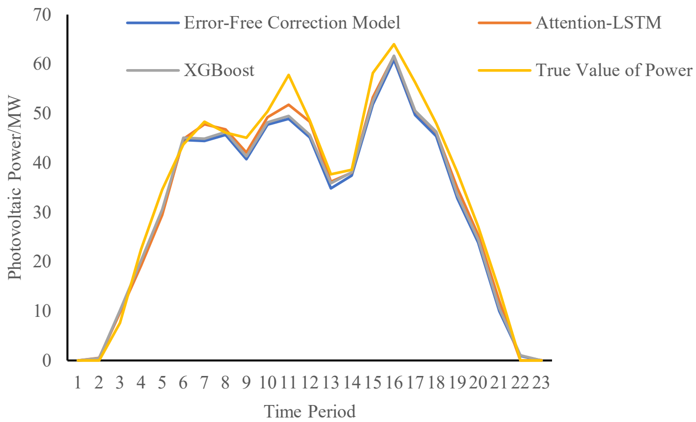
Taking the No. 1 power station in the No. 1 PV cluster as an example, the experimental results are shown in Figure 6 and Table 4. Figure 6 shows the comparison of the PV power prediction curves for November 17 (sunny), November 12 (cloudy), and November 28 (rainy) from 7:00 to 19:00, and Table 4 shows the comparison of the errors under different error correction models.

Table 4. Error comparison of different error models

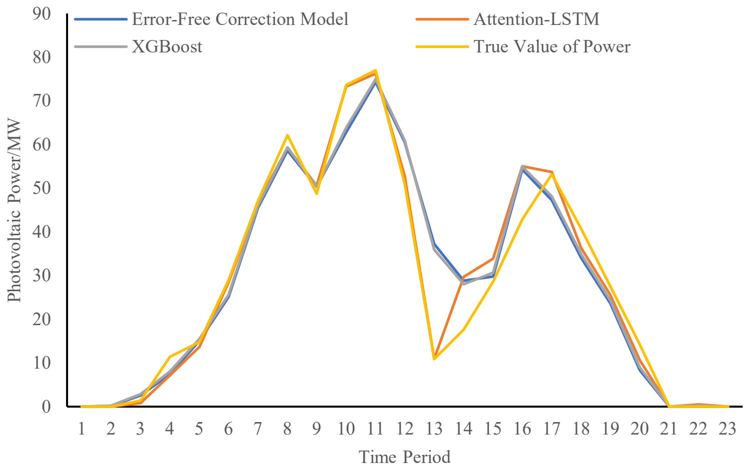
Input Features	MAE/MW	MSE/MW
Error-Free Correction Model	3.24	23.12
XGBoost Model	2.97	20.14
Attention-LSTM Modeol	2.78	17.95



(A) sunny



(B) cloudy



(C) rainy

Figure 6. Comparison of prediction results under different error correction models

As shown in Table 4, the Mean Absolute Error (MAE) and Mean Squared Error (MSE) of the Attention-LSTM error correction model proposed in this paper are 2.78 MW and 17.95 MW, respectively. Compared to the XGBoost error correction model, the MAE and MSE are reduced by 6.39% and 10.87%, respectively. Additionally, when compared to the scenario without any error correction model, the MAE and MSE decrease by 14.19% and 22.36%, respectively. These results indicate that the Attention-LSTM error correction model effectively enhances the accuracy of distributed PV power predictions.

6. Conclusion

This paper presents a method for short-term power prediction of distributed photovoltaic (PV) systems that utilizes a fusion of spatial and temporal features, combined with model error correction, to improve prediction accuracy, and the main conclusions are as follows:

(1) A model input feature construction method that accounts for spatiotemporal correlation is proposed. This approach jointly utilizes the weather features of the target PV plant and the power features of the reference PV plant as input features for the model. By incorporating these fused features, the accuracy of distributed PV power prediction is enhanced compared to models that use only weather features or power features as inputs.

(2) An Attention-LSTM-based error correction model is proposed, building upon the Attention-LSTM power prediction model. Compared to scenarios with no error correction or the use of a simple error correction model, the Attention-LSTM error correction model effectively reduces the prediction error of the power prediction model, yielding superior results.

However, the problem of gradient disappearance or explosion of the Attention-LSTM model used in this article still exists. Therefore, more advanced models such as LSTNet can be considered in future work to further improve the accuracy of photovoltaic power prediction.

Acknowledgements.

The work is supported by Inner Mongolia Electric Power (Group) Co., Ltd Technology Project [Grant Nos. 2023-3-2, 2023KYPT0020].

References

[1] XIN Baoan, SHAN Baoguo, LI Qionghui, et al. Rethinking of the “Three Elements of Energy” Toward Carbon Peak and Carbon Neutrality[J]. Proceedings of the CSEE, 2022,42(09):3117-3126.

[2] ARGUELLO A, LARA J D, ROJAS J D, et al. Impact of rooftop PV integration in distribution systems considering socioeconomic factors[J]. IEEE Systems Journal, 2017, 12(4): 3531-3542.

[3] SONG Shaojian, LI Bohan. Short-term forecasting method of photovoltaic power based on LSTM[J]. Renewable Energy,2021,39(05):594-602.

[4] TIAN Jiangang. Ultra short-term forecast of photovoltaic generation based on gru of deep learning algorithm[D]. East China Jiao Tong University,2021.

[5] WEN S, ZHANG C, LAN H, et al. A hybrid ensemble model for interval prediction of solar power output in ship onboard power systems[J]. IEEE Transactions on Sustainable Energy, 2019, 12(1): 14-24.

[6] ABDEL-NASSER M, MAHMOUD K, LEHTONEN M. Reliable solar irradiance forecasting approach based on choquet integral and deep LSTMs[J]. IEEE Transactions on Industrial Informatics, 2020, 17(3): 1873-1881.

[7] TANG Deqing,ZHU Wu,HOU Linchao. Ultra-short-term photovoltaic power prediction based on cnn-lstm-xgboost model[J]. Chinese Journal of Power Sources,2022,46(09):1048-1052.

[8] LIU Zhichao,YUAN,SannanTANG Wancheng. Output power prediction of short-term photovoltaic power generation based on BLSTM-random forest[J]. Chinese Journal of Power Sources, 2021,45(04):495-498.

[9] YU Guangzheng, LU Liu, TANG Bo, et al. An Improved Hybrid Neural Network Ultra-Short-Term Photovoltaic Power Forecasting Method Based On Cloud Image Feature Extraction[J]. Proceedings of the CSEE, 2021,41(20):6989-7003.

[10] ZHOU H, ZHANG Y, YANG L, et al. Short-term photovoltaic power forecasting based on long short term memory neural network and attention mechanism[J]. Ieee Access, 2019, 7: 78063-78074.

[11] QI Xiaoxia. Research on photovoltaic/wind power forecasting based on deep learning[D]. Harbin Engineering University,2020.

[12] ZHENG Ruonan, LI Guojie, HAN Bei, et al. ay-ahead power forecasting of distributed photovoltaic generation based on weighted expanded daily feature matrix[J]. Electric Power Automation Equipment,2022,42(02):99-105

[13] WANG Biao, LYU Yang, CHEN Zhong, et al. Hybrid mechanism-data-driven short-term power forecasting of distributed photovoltaic considering information time shift [J]. Automation of Electric Power Systems,2022,46(11):67-74.

[14] ZHANG Tongyan, LIAO Qingfen, TANG Fei, et al. Wide-area distributed photovoltaic power forecast method based on meteorological resource interpolation and transfer learning[J/OL]. IEEE Systems Journal:1-12[2023-07-22]

[15] MIRJALILI S, MIRJALILI S M, HATAMLOUq A. Multi-verse optimizer: a nature-inspired algorithm for global optimization[J]. Neural Computing and Applications, 2016, 27: 495-513.

[16] JIAO Tianli. Study on key models of wide area distributed photovoltaic power generation output prediction based on spatio-temporal relationship[D]. Hangzhou Dianzi University,2019.

[17] YANG Xiyun, ZHAO Zeyu, YANG Yan. Research on distributed photovoltaic power prediction method based on combination of spatiotemporal information[J]. Thermal Power Generation,2022,51(08):64-72.

[18] WANG Hao. Reliability evaluation of power generation systems with high-proportion of renewable energy[D]. North China Electric Power University,2020.

[19] FANG Hui. Optimization planning and application research of high penetrated distributed photovoltaic generation cluster[D]. Hefei University of Technology,2019.

[20] WANG Xiaoxia, YU Min, HUO Zejian, et al. Short-term power forecasting of distributed photovoltaic station clusters based on affinity propagation clustering and long short-term time-series network[J]. Automation of Electric Power Systems, 2023,47(06):133-141.

[21] WANG Kaiyan, DU Haodong, JIA Rong, et al. Short-term interval probability prediction of photovoltaic power based on similar daily clustering and QR-CNN-BiLSTM model[J]. High Voltage Technology,2022,48(11):4372-4388.

[22] YANG Guohua,, ZHANG Honghao, ZHENG Haofeng. Short-term photovoltaic power forecasting based on similar weather clustering and IHGWO-WNN-AdaBoost model[J]. High Voltage Technology, 2021,47(04):1185-1194.

[23] LIU Guohai, SUN Wenqing, WU Zhenfei, et al. Short-term photovoltaic power forecasting based on Attention-GRU model[J]. Journal of Solar Energy,2022,43(02):226-232.

[24] BAN Lan. Research on short-term photovoltaic power prediciton based on long short memory neural network[D]. Guangxi Universiy,2021.