

## Research Article

# PV Power Forecasting in the Hexi Region of Gansu Province Based on AP Clustering and LSTNet

Xujiong Li <sup>1</sup>, Guoming Yang,<sup>2</sup> and Jun Gou<sup>2</sup>

<sup>1</sup>Lanzhou Resources & Environment Voc-Tech University, Yellow River Basin Ecotope Integration of Industry and Education Research Institute, Key Laboratory of Climate Resources Utilization and Disaster Prevention and Mitigation of Gansu Province, Lanzhou 730021, China

<sup>2</sup>Lanzhou Power Supply Company, State Grid Gansu Electric Power Company, Lanzhou 730000, China

Correspondence should be addressed to Xujiong Li; [lxj8836@163.com](mailto:lxj8836@163.com)

Received 12 June 2023; Revised 4 March 2024; Accepted 8 April 2024; Published 26 April 2024

Academic Editor: Dhanamjayulu C.

Copyright © 2024 Xujiong Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Accurate PV power forecasting is becoming a mandatory task to integrate the PV system into the power grid, schedule it, and ensure the safety of the power grid. In this paper, a novel model for PV power prediction using AP-LSTNet has been proposed. It consists of a combination of affinity propagation clustering and long-term and short-term time series network models. First, the affinity propagation algorithm is used to divide the regionally distributed photovoltaic station clusters into different seasons. The Pearson correlation coefficient is used to determine the strong correlation between meteorological factors of photovoltaic power, and the bilinear interpolation method is used to encrypt the meteorological data of the corresponding photovoltaic station cluster. Furthermore, LSTNet is used to mine the long-term and short-term temporal and spatial dependence of photovoltaic power, and meteorological factor series and linear components of auto-regression are superimposed to realize the simultaneous prediction of multiple photovoltaic stations in the group. Finally, PV power plants in five cities, Wuwei, Jinchang, Zhangye, Jiuquan, and Jiayuguan in the Hexi region of Gansu Province, China, will be selected to test the proposed model. The experimental comparison shows that the prediction model achieves high prediction accuracy and robustness.

## 1. Introduction

In recent years, PV power generation has developed rapidly, relying on the advantages of clean energy. In China, PV power plants include centralized and distributed two categories; centralized power plants are large-scale PV power plant clusters constructed in desert areas, making full use of abundant and relatively stable solar energy resources and accessing high-voltage transmission systems to supply long-distance loads. Distributed power stations are mainly built on the surface of scattered buildings, solving the power consumption problems of users in the vicinity, realizing the compensation of power supply difference, and sending out through grid connection. Self-generation and self-consumption surplus power on-grid is the main mode of distributed photovoltaic power generation, users of photovoltaic power stations issued by the power to meet their load first and the excess can be sold to the power

company; if the power generated by the energy does not meet their load, the grid power supply to supplement it. The power company only collects daily power generation data but lacks monitoring of the operating conditions of the power generation equipment and is unable to carry out routine maintenance and repair. With the continuous improvement of the penetration rate of distributed PV, its volatility and randomness have become uncontrollable factors in grid-coordinated dispatch. Therefore, accurate short-term power forecasting of distributed PV is of great importance to optimize power system dispatching and ensure safe operation of the power system [1].

Short-term PV power forecasting methods can be divided into statistical methods and machine learning methods. Statistical methods mainly include grey theory [2], regression analysis [3], time series [4], and so on. These methods have simple models but relatively poor prediction accuracy and stability. Machine learning methods mainly include support

vector machine [5], random forest [6], extreme learning machine, neural network [7], and so on. In recent years, deep neural networks such as convolutional neural networks (CNN) [8, 9] and long short-term memory (LSTM) networks [10, 11] have also been introduced to improve the fitting ability of models. Combining the advantages of different neural networks, hybrid network prediction models such as CNN-LSTM [12, 13] and recurrent neural network (RNN)-LSTM [14, 15] further improve the prediction accuracy of the model by capturing the time series and spatial correlation between PV power generation power sequence and related influencing factors. Since the attention mechanism can measure the importance of input features, it is introduced into various neural network units to improve the generalization ability of the model [16, 17].

The above power prediction method is mainly applicable to PV power stations with meteorological acquisition systems and complete power information, while the distributed PV power stations with low-voltage access are small and scattered, and the investment cost of separately configuring meteorological measurement equipment is too high [18]. In [19, 20], by analyzing the spatial correlation of PV power between the neighboring power station and the power station to be predicted, the decision tree and neural network model are, respectively, established to construct the nonlinear mapping relationship between them. In [21, 22], the BP neural network model is established to improve the PV prediction accuracy of the slave station by studying the correlation between the “PV master station—PV slave station” clustered in the same category. In [23, 24], spatial correlation cluster analysis is performed on historical sample groups of weather types, the deployment of weather stations is optimized, and a multi-PV user power prediction model based on “space-time correlation” is proposed. In [25, 26], missing power data are reconstructed based on space-time correlation, and an irradiance encryption model is established by using a 3D convolutional neural network to achieve full grid coverage of power and meteorological data. In [27], an improved version of the PV2-state model is introduced for intra-hour PV power prediction. Reference [28] proposes a new model for predicting photovoltaic power generation using LSTM-TCN. In [29], LSTM-based predictive models are used to control solar PV systems and effectively prepare for future battery system consumption. Reference [30] proposes a new two-stage deep learning method for photovoltaic power generation prediction, which has significant improvement and robustness in point prediction and probabilistic prediction tasks. In [31], a graph-based multisite daytime PV generation prediction model is presented. It is possible to interpret which PV stations and time steps influence the prediction. Reference [32] analyses the performance of 12 different models that predict day-ahead generation in line with market conditions.

The current short-term photovoltaic power prediction technology has achieved certain results, but it still needs to solve the following aspects: (1) all of the above solutions effectively improve the power prediction accuracy of distributed PV, but the high economic cost makes it impossible to carry out large-scale engineering practice. (2) The power output of distributed PV has strong randomness and volatility, and has obvious daily periodicity [33]. The power prediction of a single distributed PV power plant not only

has low accuracy but also has little impact on power system planning and scheduling. The model adaptability between different power plants is poor and cannot be shared.

The traditional single model cannot effectively solve the above problems. Regarding the photovoltaic power prediction technology for the multivariable time series prediction scenario, this paper proposes a short-term power forecasting model for regional distributed PV power plants based on affinity propagation (AP) clustering and a long short-term time series network (LSTNet). The regional forecasting of multiple distributed PV power plants can effectively improve the accuracy and applicability. According to the local climate characteristics of the PV power plant, the PV output data are identified in seasons, and the distributed PV power plant groups in different seasons in the region are divided by AP clustering so that the weather in each power plant group is consistent. Due to the geographic dispersion of distributed PV power stations, it is impossible to set up a meteorological station for each station and only the meteorological data of centralized PV power stations that are far away can be shared. Therefore, the Pearson correlation coefficient is first used to determine the strong correlation between meteorological factors of PV output power and then the corresponding meteorological data are encrypted by bilinear interpolation so that each station group has its meteorological data, thus realizing the full-area coverage of key meteorological data. Finally, the LSTNet is used to forecast multiple PV power plants in the cluster simultaneously to improve the short-term power forecast accuracy of the large-scale distributed PV power plant cluster. Combined with the measured data, the results show that the power prediction model has excellent generalization, high accuracy, and strong robustness.

## 2. Distributed PV Power Plant Cluster Partition Based on AP Clustering

*2.1. Correlation between PV Power and Meteorological Factors.* PV performance is related to many factors, such as external meteorological factors like light intensity, temperature, humidity, cloud movement, wind speed and direction, etc. [34], and internal technical parameters such as installed location and capacity, PV conversion efficiency, and PV panels. However, large-scale distributed PV users not only seriously lack technical parameters such as installation location but also do not provide public weather stations information, such as illumination and other meteorological data, and cannot obtain strict geographical zoning information in accordance with meteorological associations. Therefore, it is necessary to find the space-time distribution rule equivalent irradiance, which describes the nature of PV power from the existing PV data.

As shown in Figure 1, since there is an approximately linear relationship between the power plant output power and the solar irradiance, the PV power curve of the large-scale distributed PV power plant can be used as the equivalent irradiance to reflect the change process of meteorological information such as light intensity and temperature of the location. According to the spatial correlation of the historical data of PV power, the PV power plant cluster is divided so that the data characteristics of each

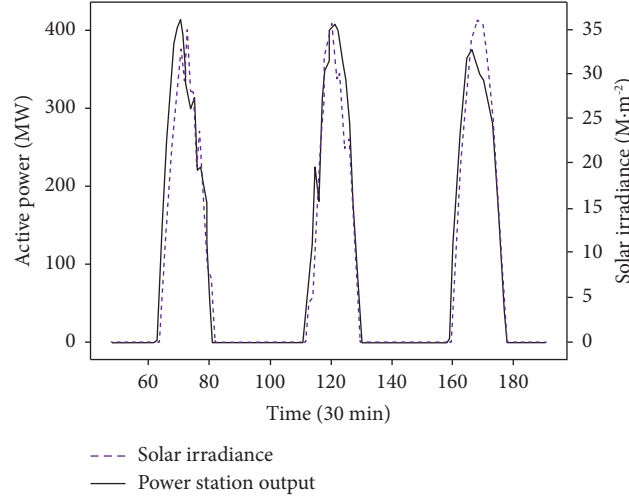


FIGURE 1: Relationship between output power and irradiance.

power plant in the cluster are similar and have meteorological consistency. On the one hand, this provides a basis for the subsequent reconstruction of regional meteorological data as the meteorological data of each distributed power plant. On the other hand, regional forecasting of distributed PV power plants can simplify the tedious modelling of each individual station and reduce the difficulty of forecasting.

**2.2. Classification of PV Power Station Group by AP Clustering.** AP clustering is an unsupervised clustering algorithm for clustering based on information transfer [35, 36]. Compared with traditional clustering algorithms such as  $K$ -means clustering, this algorithm is insensitive to outliers, and the clustering results are more stable [37, 38]. In this paper, AP clustering is used to cluster the PV power of PV power plants in the region, and power plants with the same fluctuation trend are divided into the same group of power plants. The algorithm first takes all distributed PV power stations as potential cluster centers, and then iteratively competes for cluster centers based on “mutual information transmission” among power stations. As shown in Figure 2, there are two information exchange mechanisms between the power plant data points: attraction information  $r(i, k)$  and assignment information  $a(i, k)$ , where the attraction information  $r(i, k)$  represents the degree to which plant  $k$  is suitable as the cluster center of the candidate central plant  $i$ . The assignment information  $a(i, k)$  represents the suitability of plant  $i$  to select plant  $k$  as the cluster center.

As an unsupervised algorithm, AP clustering cannot directly evaluate the clustering effect. In order to obtain accurate and stable clustering results, this paper chooses the silhouette coefficient [39, 40] to evaluate the clustering results, and its formula is as follows:

$$s(i) = \frac{n(i) - m(i)}{\max\{m(i), n(i)\}}, \quad (1)$$

where  $s(i)$  is the profile coefficient of power plant  $i$ ;  $m(i)$  is the average distance between sample  $x_i$  of power plant  $i$  and other samples in the same cluster, which is called cohesion;  $n(i)$  is the average distance between  $x_i$  and all samples in other

clusters, which is called separation degree; the average contour coefficient is the average value of all sample contour coefficients and the value range is  $[-1, 1]$ . The larger the value, the smaller the intracluster distance, and the larger the intercluster distance, the better the clustering effect.

The steps for dividing PV power plant groups by AP clustering are as follows:

Step 1: Divide the PV power data into four seasons and standardize each season.

Step 2: Calculate the similarity  $s(i, k)$  between the power samples of the distributed PV power plants as shown in (2) to obtain the similarity matrix  $S$ , and its diagonal element  $s(k, k)$  is the evaluation standard for whether the power plant  $k$  sample  $x_k$  can become the cluster center of the power plant cluster, which is called the reference value, and its size will affect the number of clusters.

$$s(i, k) = -\|x_i - x_k\|^2. \quad (2)$$

Where  $\|\cdot\|^2$  means to find the Euclidean distance.

Step 3: Calculate the element  $r(i, k)$  of the attractiveness matrix  $r$  and the element  $a(i, k)$  of the attribution matrix  $a$ .

$$r(i, k) \leftarrow s(i, k) - \max_{k' \neq k} \{a(i, k') + s(i, k')\}, \quad (3)$$

$$a(i, k) \leftarrow \min \left\{ 0, r(k, k) + \sum_{i' \neq (i, k)} \max \{0, r'(i', k)\} \right\}, \quad (4)$$

$$a(k, k) \leftarrow \max_{i' \neq k} \{0, r(i', k)\}. \quad (5)$$

Step 4: Update  $r(i, k)$  and  $a(i, k)$ , and introduce a damping coefficient  $\lambda$  to adjust the convergence speed and the iterative stability.

$$\begin{cases} r^{(d+1)}(i, k) = \lambda r^{(d)}(i, k) + (1 - \lambda) r^{(d+1)}(i, k), \\ a^{(d+1)}(i, k) = \lambda a^{(d)}(i, k) + (1 - \lambda) a^{(d+1)}(i, k). \end{cases} \quad (6)$$

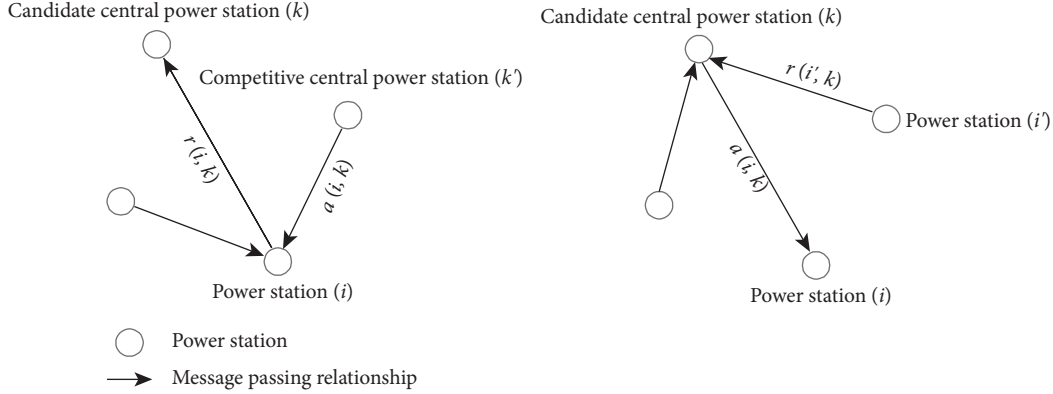


FIGURE 2: AP messaging mechanism.

Where  $d$  represents the number of iterations.

Step 5: If the number of iterations exceeds the preset number or the clustering distance is not changed, go to step 6. Otherwise, repeat steps 3 and 4 to continue the calculation.

Step 6: Calculate the profile coefficient according to the clustering result under the current reference value, determine the clustering center and different PV power plant clusters, go to step 2 and change the reference value.

Step 7: Analyze the contour coefficients under different clustering numbers, select the best clustering result, and complete the division of the PV power station group.

### 3. Selection and Encryption of Meteorological Factors

Considering that after the PV power plant group is divided, the meteorological consistency exists in the region of the same PV power plant group, the most relevant cluster center power plant of each type of PV power plant group can be selected as the representative power plant, and the meteorological data of the representative power plant is used as the overall meteorological data of the region.

Since most distributed PV power plants only have meteorological data from central PV power plants located far away and coarse-grained weather forecast, the bilinear interpolation method can be used to encrypt meteorological data of representative power plants of different station groups when there is no corresponding meteorological station for representative power plants, so that each station group has its meteorological data. In this way, the coverage of key meteorological data for the whole region by station groups can be achieved. Therefore, this paper first uses the Pearson correlation coefficient to determine the strong correlation between meteorological factors of PV output power, and then uses the bilinear interpolation method to encrypt the corresponding meteorological data for the subsequent prediction of neural network models.

**3.1. Selection of Meteorological Factors.** Assuming that the internal technical parameters of the PV power plant remain basically unchanged, it is very important to isolate the main

factors influencing the PV output power from the external meteorological factors (irradiance, temperature, humidity, etc.). The Pearson correlation coefficient  $R(X, Y)$  is used to measure the degree of correlation between the PV output power and each influencing factor, namely.

$$R(X, Y) = \frac{\sum_{c=1}^C (X_c - \bar{X})(Y_c - \bar{Y})}{\sqrt{\sum_{c=1}^C (X_c - \bar{X})^2 \sum_{c=1}^C (Y_c - \bar{Y})^2}}, \quad (7)$$

where  $X_c$  and  $Y_c$  are the output power and meteorological factor sample points of the standardized PV plant sample  $c$ , respectively.  $X$  and  $Y$  are matrices consisting of  $X_c$  and  $Y_c$ , respectively.  $\bar{X}$  and  $\bar{Y}$  are the mean values of  $X_c$  and  $Y_c$ , respectively.  $C$  is the number of samples. The value range of the correlation coefficient is  $[-1, 1]$ , and the larger the absolute value, the stronger the correlation.

**3.2. Densification of Meteorological Factors Based on Bilinear Interpolation.** Bilinear interpolation, as a classical statistical encryption method, is widely used in signal processing, digital image processing, and so on [41]. The core idea is to perform a linear interpolation in two directions, respectively. The basic principle is shown in Figure 3. In weather data encryption, the two directions of the interpolation function are the east-west and north-south distances between weather stations, as shown on the  $x$  and  $y$  axes.  $P$  is the meteorological station to be sought, and  $Q_{11}$ ,  $Q_{12}$ ,  $Q_{21}$ , and  $Q_{22}$  are the four known meteorological stations closest to them.  $Q_{11}$  and  $Q_{21}$  obtain the meteorological data  $f(R_1)$  of the point  $R_1$  by linear interpolation in the  $x$ -axis direction, as shown in (8), and similarly obtain the meteorological data  $f(R_2)$  of the point  $R_2$ . By interpolation in the  $y$ -axis direction from points  $R_1$  and  $R_2$ , the meteorological data  $f(P)$  of point  $P$  can be obtained, as shown in equation (9):

$$\begin{cases} f(R_1) \approx \frac{x_2 - x_0}{x_2 - x_1} f(Q_{11}) + \frac{x_0 - x_1}{x_2 - x_1} f(Q_{21}), \\ f(R_2) \approx \frac{x_2 - x_0}{x_2 - x_1} f(Q_{12}) + \frac{x_0 - x_1}{x_2 - x_1} f(Q_{22}). \end{cases} \quad (8)$$

$$f(P) \approx \frac{y_2 - y_0}{y_2 - y_1} f(R_1) + \frac{y_0 - y_1}{y_2 - y_1} f(R_2), \quad (9)$$

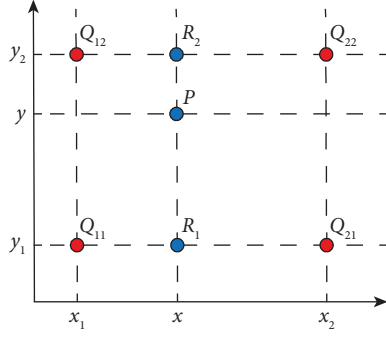


FIGURE 3: Principle of bilinear interpolation.

where  $x_1$  and  $y_1$  are the abscissa and ordinate of weather station  $Q_{11}$ , respectively.  $x_2$  and  $y_2$  are the abscissa and ordinate of weather station  $Q_{22}$ , respectively.  $x_0$  and  $y_0$  are the abscissa and ordinate of point  $P$ , respectively.

#### 4. Prediction Model Based on LSTNet

The LSTNet model [42, 43] needs to keep the input and output dimensions consistent, and it is impossible to input only meteorological data as the characteristic variable of multistation PV power. In this paper, the original LSTNet model is modified. The PV power and meteorological factors are taken as the input of the model. The output dimensions of the fully connected layer of the nonlinear branch and the autoregressive layer of the linear branch are modified to be the same as the quantity dimension of the PV power plant. The final prediction result is obtained by superimposing the prediction results of the two branches. In this way, the model can extract the meteorological characteristics without increasing the output dimension, so as to better extract the long-term trend and short-term fluctuation characteristics of the time series of multistation PV output and related meteorological factors, and combine the robustness of linear and nonlinear branch addition models to effectively predict the PV power generation output in the future. The structure of LSTNet is shown in Figure 4, which consists of a nonlinear branch composed of convolutional layers, loops, loop-skip layers, and fully connected layers, and a linear branch composed of autoregressive layers.

**4.1. Convolution Layer.** The first layer of LSTNet is a convolutional layer, which extracts short-term patterns and local dependencies between variables from time series of PV power and meteorological data, and mines correlations between multiple features.

The convolutional layer uses the convolutional kernel to traverse the input layer and complete the data transformation to the input layer. The output obtained by extracting the time series features by one-dimensional convolution is

$$\mathbf{h}_v = \mathbf{f}(\mathbf{W}_v * \mathbf{X}_t + \mathbf{b}_v), \quad (10)$$

where  $X_t$  is the input time series.  $*$  denotes convolution operation.  $\mathbf{W}_v$  and  $\mathbf{b}_v$  are the weight matrix and bias vector of the  $v$ th filter, respectively.  $\mathbf{h}_v$  is the output feature.  $\mathbf{f}(\cdot)$  denotes the ReLU activation function.

**4.2. Loop and Loop-Skip Layer.** The output of the convolutional layer is input to both the recurrent layer and the loop-skip layer. The recurrent layer uses LSTM network units to selectively remember or forget sequence information, which can capture relatively long-term dependencies in historical information while reducing the risk of gradient explosion during model training. The state of the cycle unit at time  $t$  is calculated according to the following formula:

$$\begin{cases} \mathbf{r}_t = \sigma(\mathbf{x}_t \mathbf{W}_{xr} + \mathbf{h}_{t-p} \mathbf{W}_{hr} + \mathbf{b}_r), \\ \mathbf{u}_t = \sigma(\mathbf{x}_t \mathbf{W}_{xu} + \mathbf{h}_{t-p} \mathbf{W}_{hu} + \mathbf{b}_u), \\ \mathbf{o}_t = \sigma(\mathbf{x}_t \mathbf{W}_{xo} + \mathbf{h}_{t-p} \mathbf{W}_{ho} + \mathbf{b}_o), \\ \tilde{\mathbf{c}}_t = \mathbf{f}(\mathbf{x}_t \mathbf{W}_{xc} + \mathbf{h}_{t-p} \mathbf{W}_{hc} + \mathbf{b}_c), \\ \mathbf{c}_t = \mathbf{u}_t \cdot \mathbf{c}_{t-1} + \mathbf{r}_t \cdot \tilde{\mathbf{c}}_t, \\ \mathbf{h}_t = \mathbf{o}_t \cdot \mathbf{f}(\mathbf{c}_t), \end{cases} \quad (11)$$

where  $\mathbf{r}_t$ ,  $\mathbf{u}_t$ ,  $\mathbf{o}_t$ ,  $\tilde{\mathbf{c}}_t$ ,  $\mathbf{c}_t$ , and  $\mathbf{h}_t$  are input gate, forget gate, output gate, input node, memory unit, and hidden layer output, respectively.  $p$  is the number of hidden units that are skipped, where  $p = 1$ .  $\mathbf{x}_t$  is the input at time  $t$ .  $\mathbf{W}_{xr}$ ,  $\mathbf{W}_{hr}$ ,  $\mathbf{W}_{xu}$ ,  $\mathbf{W}_{hu}$ ,  $\mathbf{W}_{xo}$ ,  $\mathbf{W}_{ho}$ ,  $\mathbf{W}_{xc}$ , and  $\mathbf{W}_{hc}$  are weight matrices.  $\mathbf{b}_r$ ,  $\mathbf{b}_u$ ,  $\mathbf{b}_o$ , and  $\mathbf{b}_c$  are the corresponding bias vectors.  $\sigma(\cdot)$  is the Sigmoid function.

The PV power sequence has an obvious periodic pattern, while the LSTM network cannot capture the long-term repetitive pattern in the sequence, which can be solved by adding the loop-skip layer. In this paper, a recurrent RNN with cycle skipping is used to extend the time span of the information flow and simplify the optimization process of the model. In particular, the currently hidden unit is concatenated with hidden units of the same historical period in the adjacent time period, thus exploiting a periodic pattern that combines short- and long-term repetitive patterns of the partial sequence data. The cell state update procedure for the loop-skip layer is the same as (11), and the value of  $p$  ( $24N$ ,  $N$  is a positive integer) can be easily determined for a PV power dataset with a clear periodic pattern.

**4.3. Fully Connected Layer.** The cycle unit state of the cycle layer at time  $t$  is connected to the fully connected layer, the cycle skip layer has multiple connections to the fully connected layer, and the number of connections is related to the number of data points with skip length  $p$  as the period in the time window. The fully connected layer integrates the outputs of the loop layer and the loop-skip layer to obtain the prediction result of the nonlinear branch.

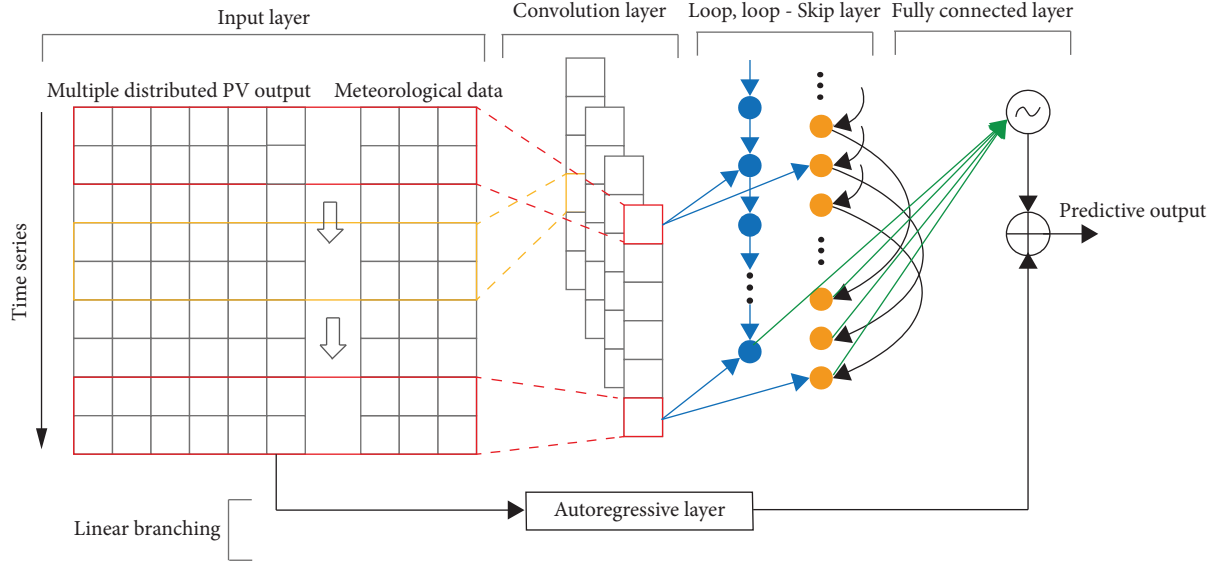


FIGURE 4: Structure of the LSTNet network model.

$$h_t^D = W_R^R h_t^R + \sum_{r=0}^{p-1} W_r^S h_{t-r}^S + b, \quad (12)$$

where  $W_R$  and  $W_r^S$  are weight matrices.  $b$  is the bias vector.  $h_t^R$  is the output of the loop layer at time  $t$ .  $h_{t-p+1}^S, h_{t-p+2}^S, \dots, h_t^S$  are the outputs of the cyclic skip layers at time instants  $t-p+1$  to  $t$ .

**4.4. Autoregressive Layer.** LSTNet adds a linear branch prediction module consisting of an autoregressive model that uses the values of previous time points of the same variable to predict the value of the current time point. The linear branch is predicted as follows:

$$h_{t,l}^L = \sum_{\varphi=0}^{q^{\text{ar}}-1} W_{\varphi}^{\text{ar}} y_{t-k,l} + b^{\text{ar}}, \quad (13)$$

where  $h_{t,l}^L$  is the prediction result of prediction autoregressive component  $l$  at time  $t$ .  $y_{t-k,l}$  is the value matrix at each previous time point.  $q^{\text{ar}}$  is the size of the input window.  $W_{\varphi}^{\text{ar}}$  and  $b^{\text{ar}}$  are autoregressive model coefficients.

Finally, the prediction results of the linear branch and the nonlinear branch are superimposed to obtain the final prediction result  $\hat{Y}_t$ :

$$\hat{Y}_t = h_t^D + h_t^L, \quad (14)$$

where  $h_t^D$  is the prediction result of the autoregressive model at time  $t$ .

For the PV power sequence, the value at a given time is highly dependent on the power value of previous time steps, and there is a nonperiodic change in a short time range with autocorrelation, as shown in Figure 5. The use of an autoregressive model can well capture the linear characteristics in the PV sequence and improve the prediction accuracy.

To sum up, the overall framework of the AP-LSTNet model proposed in this paper is shown in Figure 6.

## 5. Example Analysis

In this paper, measured power data and meteorological data of 48 distributed PV power plants in five cities, namely, Wuwei, Jinchang, Zhangye, Jiuquan, and Jiayuguan, in the Hexi region of Gansu Province, China, from January to December 2021 are selected for simulation experiments. The installed capacity of each PV system ranges from 30 MW to 200 MW.

**5.1. Division of PV Power Plant Group.** Due to the rotation of the earth, the illumination time and the angle of the sun will change with the seasons, and the power generation of the PV power station will change seasonally. Therefore, it is necessary to divide the group of stations into different seasons and establish the prediction model. Annual PV power data are used, first divided into four seasonal prediction units, the different clusters of distributed PV plants in each season are divided by the step of dividing the clusters of PV plants using AP clustering in Section 2.2, where the damping coefficient  $\lambda=0.5$ , the maximum number of iterations is 200, and the initial value is the median of the nondiagonal elements in the matrix  $S$ . Figure 7 shows the four seasons of the standing group of classification results.

Obviously, the station group divided according to PV power characteristics basically belongs to the same geographical area, that is, the power station in this area can be equivalent to a small-scale PV power station group with meteorological consistency. There are some differences in the division results of the same station group in different seasons, so it is necessary to divide the station group into four seasons before modelling. The division results are shown in Table 1.

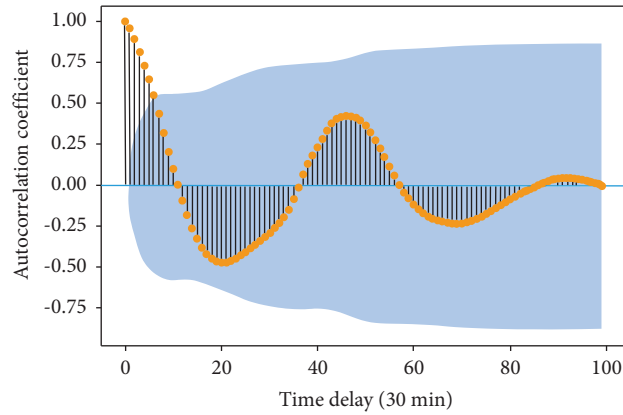


FIGURE 5: Autocorrelation coefficient of the photovoltaic output of a power station.

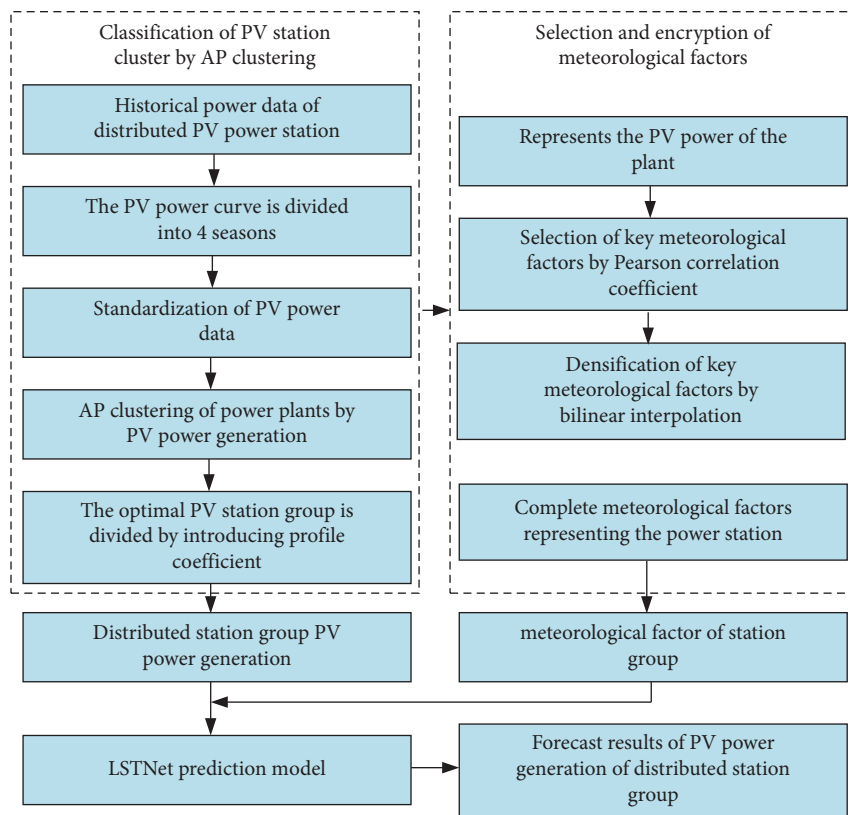


FIGURE 6: Framework of AP-LSTNet model.

To verify the superiority of AP clustering over traditional clustering algorithms, taking spring as an example, different clustering algorithms are used to partition the PV power of 56 power plants in this season. The corresponding contour coefficients under different classifications  $K$  are shown in Table 2. Obviously, when the number of classes is the same, AP clustering has a better clustering effect and a larger contour coefficient.

*5.2. Selection of Meteorological Factors and Encryption Results.* Considering the distribution of meteorological stations and the results of station clustering in different seasons, the PV power station cluster located at  $38.9\sim 39.5^{\circ}\text{N}$  latitude and  $96.8\sim 97.4^{\circ}\text{E}$  longitude is selected as the typical station cluster in each season. The cluster results of nine power plants in the power plant group in four seasons all belong to the same power plant group, which has a strong correlation.

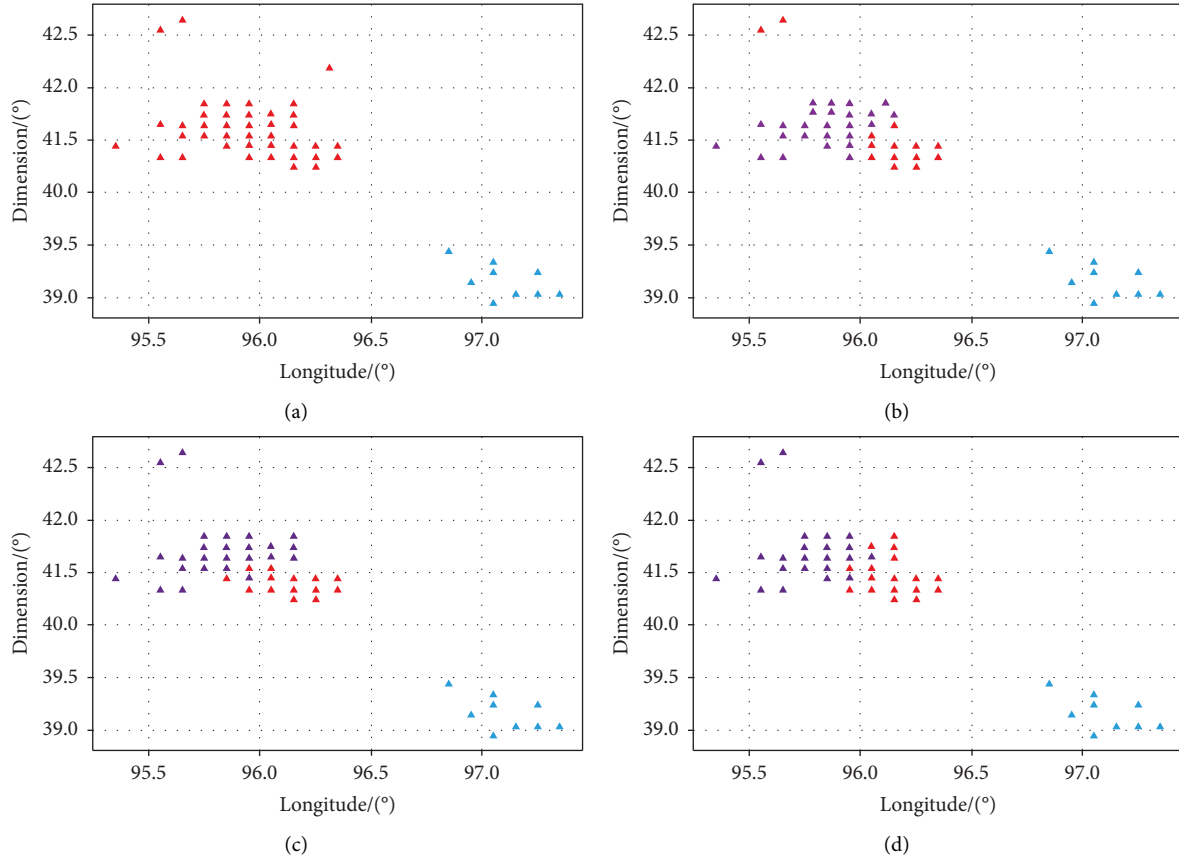


FIGURE 7: Division results of distributed PV clusters. (a) Spring; (b) summer; (c) autumn; (d) winter.

TABLE 1: Forecast seasonal unit data.

Season	Spring	Summer	Autumn	Winter
Time span	2021.4.1–2021.6.30	2021.7.1–2021.9.30	2021.10.1–2021.12.31	2021.1.1–2021.3.31
Optimal classification $K$	2	4	3	3
Contour coefficient	0.443	0.240	0.222	0.195

TABLE 2: Contour coefficients of different clustering algorithms.

Classification $K$	2	3	4	5
AP	0.443	0.435	0.178	0.155
$K$ -means	0.443	0.150	0.153	0.129
AHC	0.443	0.410	0.162	0.149

The installed capacity of the stations ranges from 43 MW to 104 MW. A suitable weather station was selected as the representative weather station for the cluster center site. It can be seen from Figure 8 that there is no meteorological station that completely matches the cluster center station. First, the nearest weather station is selected to determine the strong correlation between meteorological factors of PV power and Pearson correlation coefficient. Then, the meteorological data of the four nearby weather stations are encoded by bilinear interpolation, and the encoded

meteorological data of the representative power plant are used as the meteorological data of the whole area for the subsequent model prediction.

The Pearson correlation coefficient between the output power of a representative power plant of a typical distributed PV power plant cluster and meteorological factors in different seasons is shown in Table 3. It can be seen that although the Pearson correlation coefficients between meteorological factors and PV power in different seasons are different, ultraviolet index (UVI), global horizontal



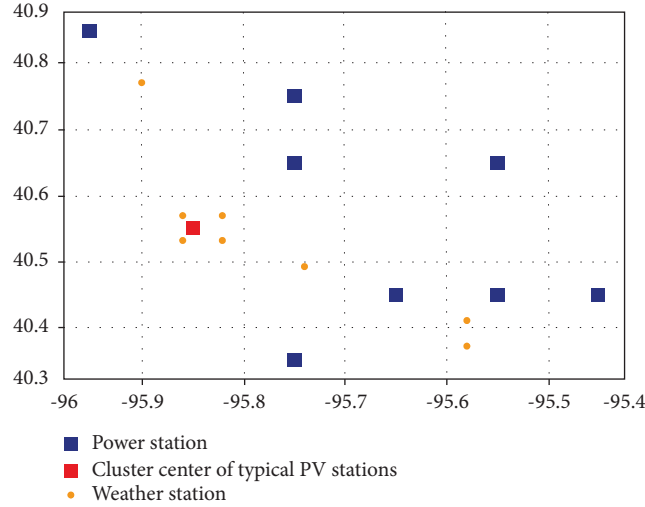


FIGURE 8: Distribution of distributed PV clusters and meteorological stations.

TABLE 3: Pearson correlation coefficient of PV and weather factors.

Meteorological factor	Spring	Summer	Autumn	Winter
UV irradiance	0.96	0.95	0.9	0.89
GHI	0.9	0.86	0.79	0.78
DNI	0.82	0.78	0.79	0.77
Relative humidity	-0.81	-0.87	-0.64	-0.71
Solar zenith angle	-0.8	-0.75	-0.6	-0.63
DHI	0.59	0.64	0.59	0.68
Ambient temperature	0.74	0.6	0.36	0.35
Wind speed	-0.44	-0.36	-0.31	-0.22
Dew point	0.003	0.083	0	-0.1
Intensity of pressure	-0.003	0.12	0.12	0.091
Wind direction	0.067	0.003	0.001	-0.11

irradiance (GHI), direct normal irradiance (DNI), and PV power have strong correlations. Secondly, relative humidity, solar zenith angle, diffuse horizontal irradiance (DHI), and ambient temperature have some correlation, while dew point and pressure have little correlation. In this paper, four meteorological characteristics with high correlation, namely, UVI, GHI, relative humidity, and DNI, are selected as the main meteorological input parameters affecting PV power generation.

Encryption based on bilinear interpolation of meteorological factors of meteorological data operator site analysis of the meteorological condition validation is shown in Figure 9.

### 5.3. Short-Term Power Prediction Results

**5.3.1. Evaluation Index.** The robustness of the model was verified using three indicators: mean absolute error (MAE)  $e_{MAE}$ , mean absolute percentage error (MAPE)  $e_{MAPE}$ , and degree of fit (RS)  $e_{RS}$ . The specific expression of evaluation indicators is as follows:

$$e_{MAE} = \frac{1}{M} |y_m - y'_m|, \quad (15)$$

$$e_{MAPE} = \frac{1}{M} \sum_{m=1}^M \frac{|y_m - y'_m|}{y_m} \times 100\%, \quad (16)$$

$$e_{RS} = \left[ 1 - \frac{\sum_{m=1}^M (y_m - y'_m)^2}{\sum_{m=1}^M (y_m - \bar{y})^2} \right] \times 100\%, \quad (17)$$

where  $y_m$  is the actual value of sample  $m$  (only the part greater than 0 is calculated in  $e_{MAPE}$ ).  $y'_m$  is the predicted value of sample  $m$ .  $\bar{y}$  is the average of the actual values.  $M$  is the number of samples.

**5.3.2. Experimental Setup.** Regarding the original LSTM model, the search range of neurons and parameters of each layer are set, the set of values of the number of hidden neurons of the convolutional layer and the recurrent layer is {48, 64, 96, 128}, the set of values of the number of hidden neurons of the recurrent skip layer is {10, 20, 30,

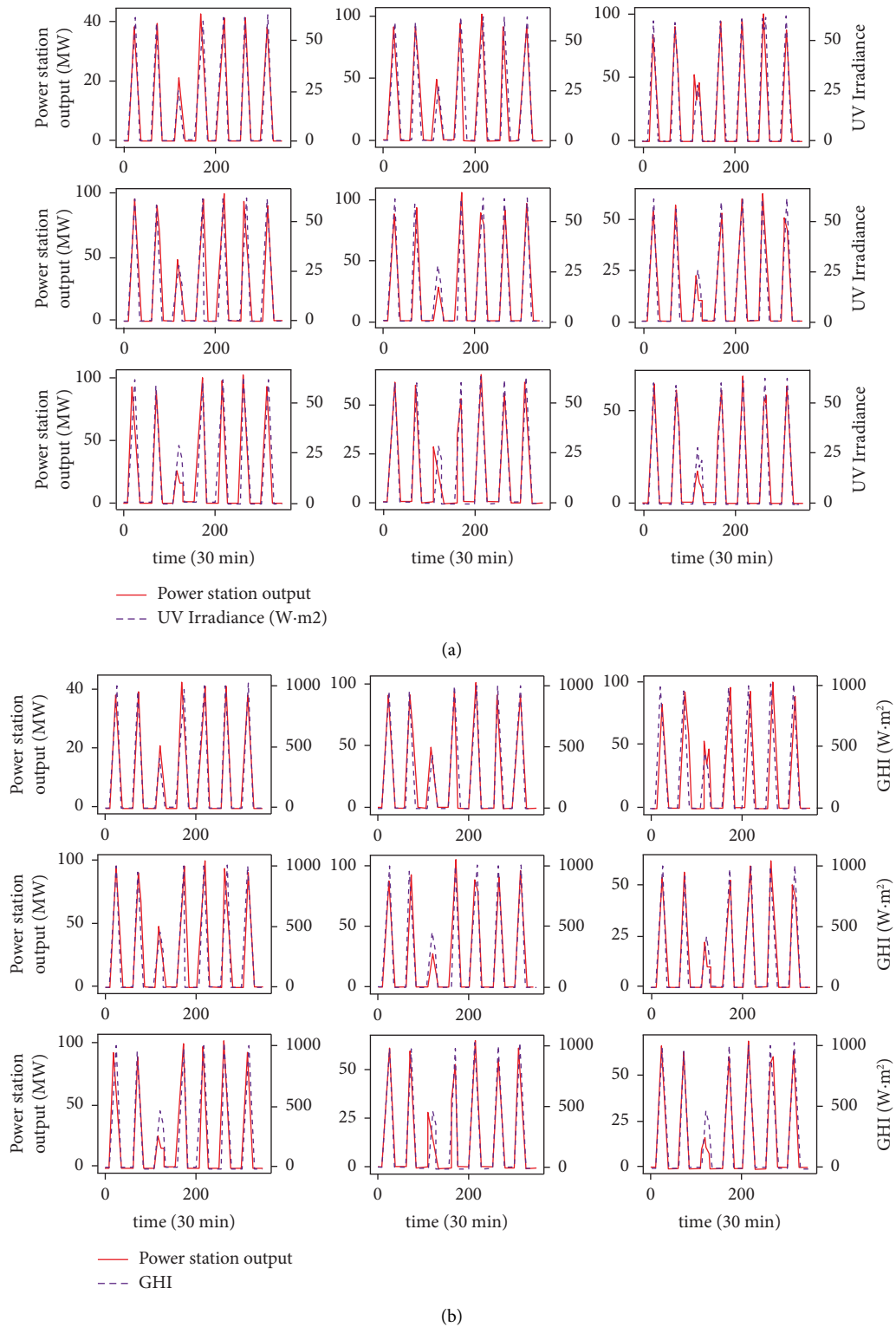
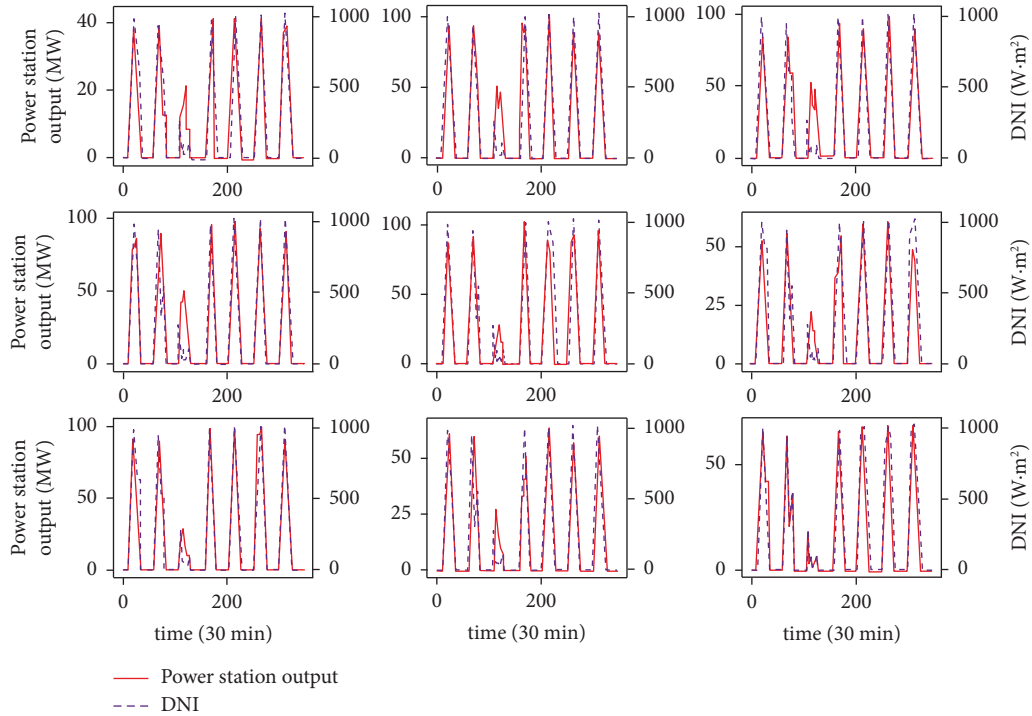
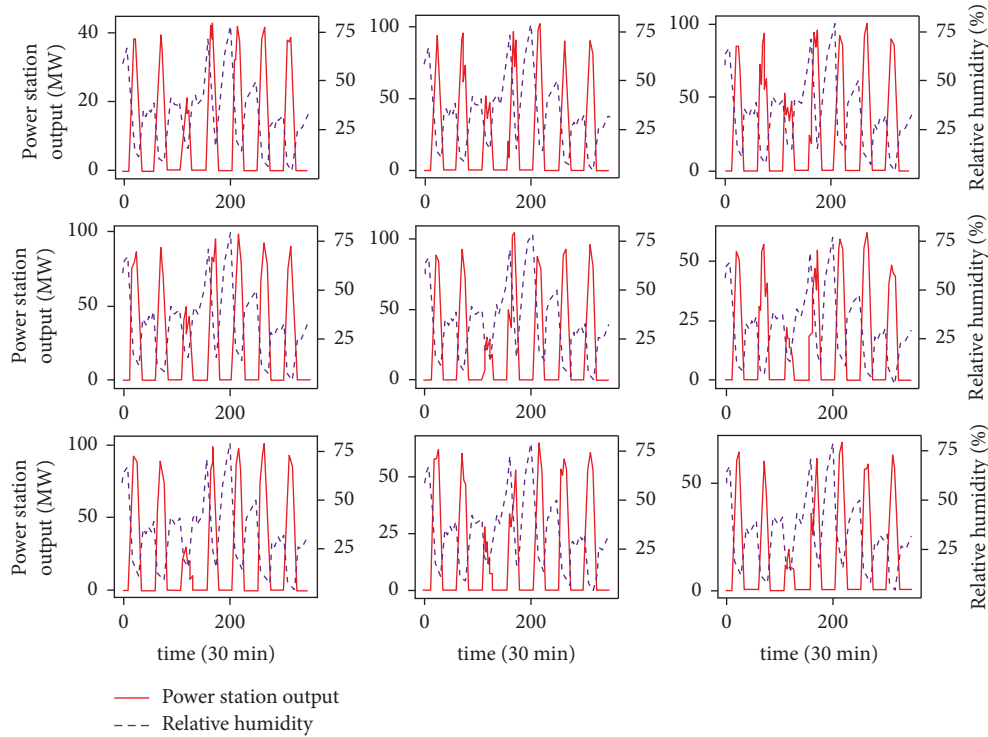


FIGURE 9: Continued.



(c)



(d)

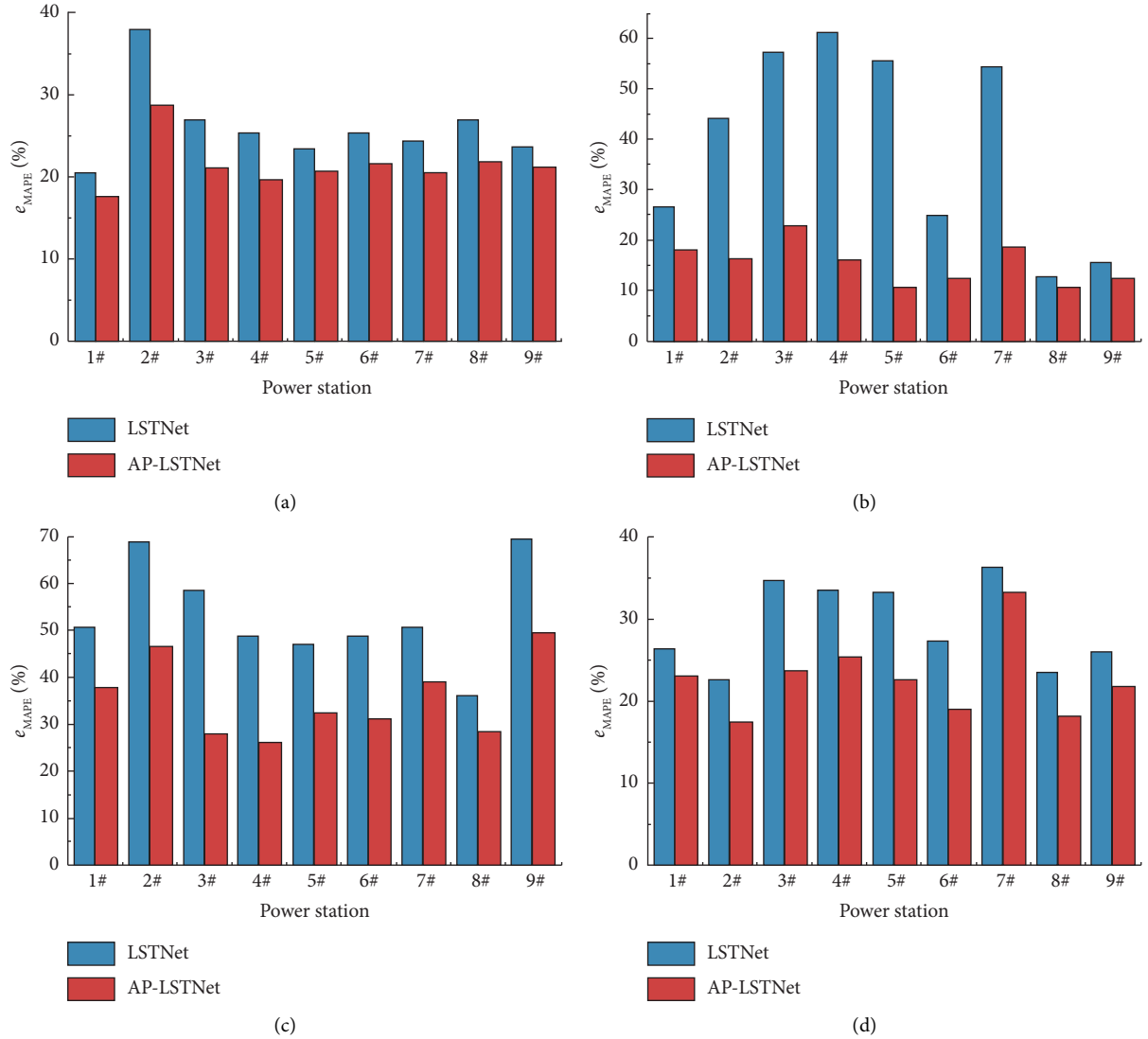
FIGURE 9: Relationship between PV and weather factors at each station. (a) UV Irradiance; (b) GHI; (c) DNI; (d) relative humidity.

40}, and the set of values of the regularization coefficient is {0.1, 1, 10}. The value set of the sliding window size is {48, 48 × 2, 48 × 3, 48 × 4, 48 × 5, 48 × 6, 48 × 7}, and the value set of the dropout layer coefficients is {0.1, 0.2}. The number of neurons in each layer is traversed using a grid

search method within this threshold range; the model prediction loss corresponding to the number of neurons in each group is calculated, and the model parameter with the minimum loss is selected as the final number of neurons; the number of hidden neurons in the

TABLE 4:  $e_{MAE}$  results of single-station forecasting and multistation simultaneous forecasting.

Prediction model	Season	$e_{MAE}$ (MWh)									Training time
		1#	2#	3#	4#	5#	6#	7#	8#	9#	
LSTNet	Spring	1.54	3.43	3.76	3.03	3.27	2.60	3.22	2.20	2.29	1357.97
	Summer	1.28	2.92	2.77	2.60	2.74	1.53	2.67	1.63	1.48	1427.37
	Autumn	1.27	2.39	2.74	1.96	2.62	1.72	2.42	1.54	1.81	1305.85
	Winter	1.42	3.01	3.03	3.51	3.22	2.07	3.50	1.82	2.08	1341.89
AP-LSTNet	Spring	1.24	2.65	2.56	2.43	2.55	1.98	2.44	1.71	1.88	200.81
	Summer	0.81	1.93	1.59	1.74	1.75	1.09	1.86	1.18	1.13	209.39
	Autumn	0.93	1.96	2.24	1.65	2.04	1.30	1.99	1.16	1.38	212.98
	Winter	1.14	2.64	2.43	2.69	2.55	1.50	2.76	1.55	1.61	206.23

FIGURE 10:  $e_{MAPE}$  comparison between single-station prediction and multistation simultaneous prediction in four seasons. (a) Spring; (b) summer; (c) autumn; (d) winter.

convolutional layer, the cyclic layer, and the loop-skip layer is set to 64, 64, and 20, respectively; the regularization coefficient of the autoregressive layer is 1, the sliding window size is  $48 \times 7$ , the skip length  $p$  is 48, and

the coefficient of the dropout layer is 0.2. The experimental batch size is 64, the training rounds are 100, and the Adam algorithm [44, 45] is adopted as the optimization algorithm.

TABLE 5: Forecasting results of different models.

Prediction model	Average value $e_{MAE}$ (MWh)				Average value $e_{MAPE}$ (%)				Average value $e_{RS}$ (%)			
	Spr.	Sum.	Fal.	Win.	Spr.	Sum.	Fal.	Win.	Spr.	Sum.	Fal.	Win.
AP-LSTM	2.26	1.79	1.96	2.33	26.09	21.10	43.98	31.17	95.99	97.86	95.79	97.21
AP-ALSTM	2.27	1.73	1.90	2.33	25.57	19.04	40.78	29.64	95.79	97.89	95.85	97.08
AP-CNN-LSTM	2.43	2.04	1.95	2.63	26.63	21.80	38.79	32.61	95.33	97.36	95.67	96.56
AP-A-CNN-LSTM	2.43	2.00	1.90	2.53	25.49	21.65	36.09	34.43	95.59	97.27	95.82	96.63
AP-LSTNet	2.16	1.45	1.63	2.10	21.50	15.42	35.66	22.79	96.56	98.47	96.84	97.50

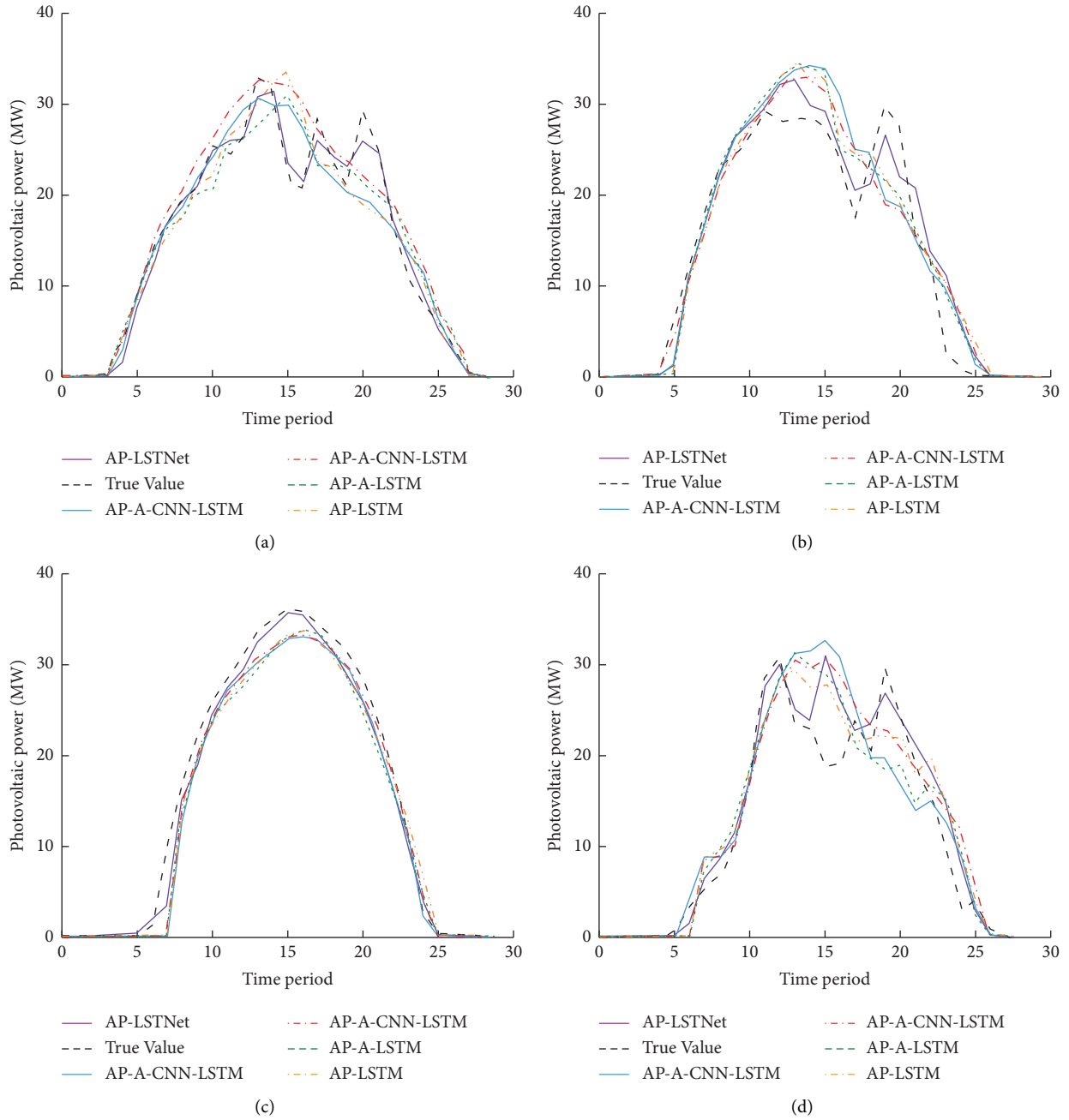


FIGURE 11: Comparison of forecasting results in different seasons. (a) Spring; (b) summer; (c) autumn; (d) winter.

TABLE 6: Prediction results under different weather types.

Forecasting method	$e_{MAE}$ (MWh)			
	Sunny	Cloudy	Clear to overcast	Shower
AP-LSTM	0.91	2.22	2.27	2.49
AP-ALSTM	0.70	1.96	2.14	2.30
AP-CNN-LSTM	1.48	2.03	2.86	2.51
AP-A-CNN-LSTM	0.74	1.88	2.34	2.90
AP-LSTNet	0.62	1.36	1.59	1.40

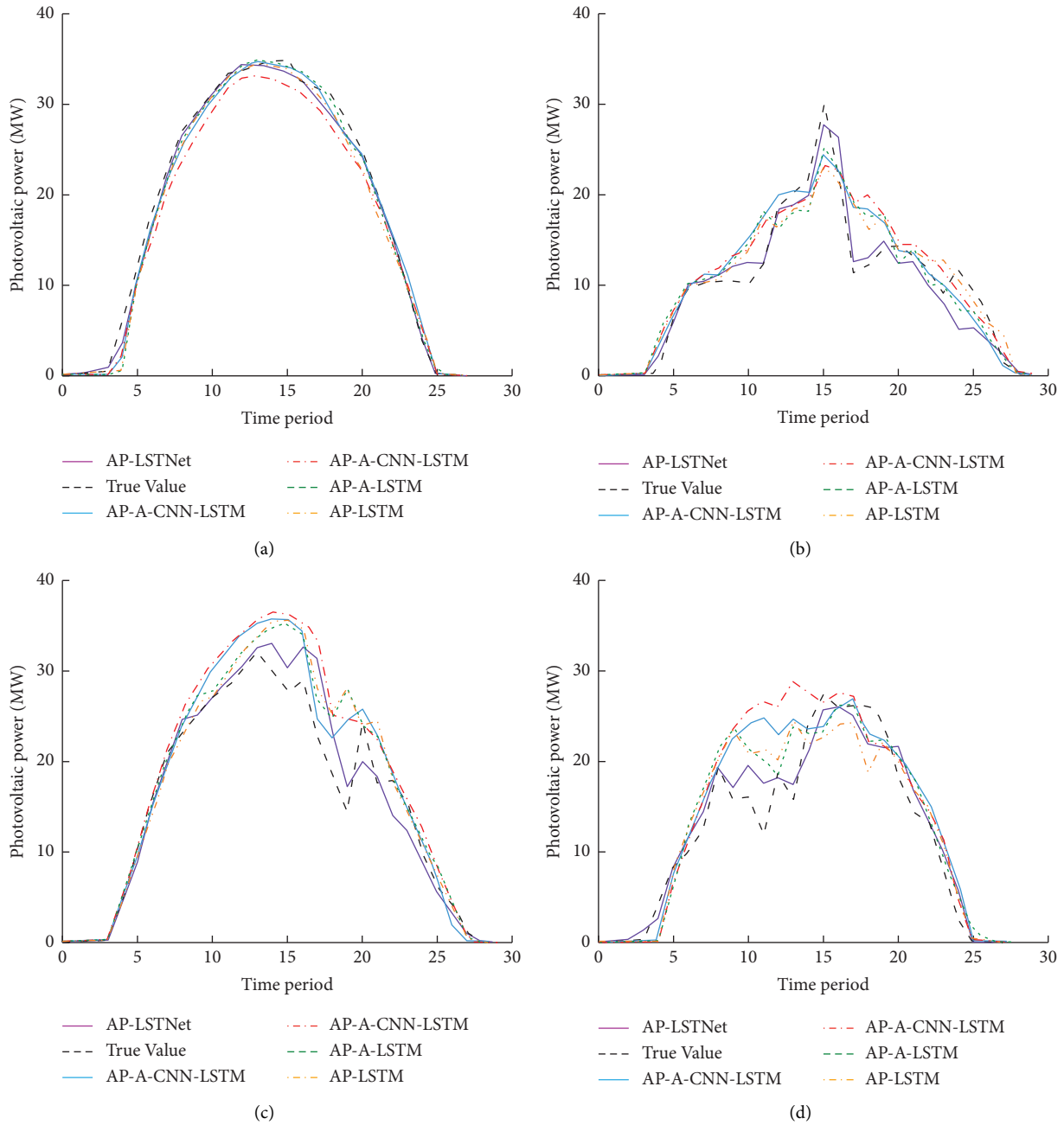


FIGURE 12: Comparison of forecasting results with different weather types. (a) Sunny; (b) cloudy; (c) clear to overcast; (d) shower.

### 5.3.3. Performance Evaluation and Error Analysis of the Prediction Model

- (1) Comparative analysis of single-station prediction and multistation simultaneous prediction

To verify the effectiveness of using AP clustering to predict the output power of multiple PV power plants in a typical cluster, one AP-LSTNet model and nine LSTNet models are built. The AP-LSTNet model simultaneously predicts nine PV power plants in the cluster, and its inputs are the PV power of nine power plants and the meteorological data of typical meteorological stations. The nine LSTNet models each predict the nine PV stations in the cluster, and the input is the PV power of each station and the meteorological data of typical meteorological stations. The prediction results of the two types of models are compared and analyzed, and the prediction results are shown in Table 4 and Figure 10.

It can be seen that the multistation simultaneous prediction in spring, summer, autumn, and winter has lower error indexes  $e_{MAE}$  and  $e_{MAPE}$  than that of single-station prediction, and the average training time is greatly reduced. It can be seen that the prediction accuracy and speed can be effectively improved by using AP clustering to divide the PV power plant group and simultaneously predict the power plants in the power plant group.

- (2) Comparative analysis of different prediction models

To verify the validity and reliability of the LSTNet model, the prediction results of the AP-LSTNet model were compared with those of the AP-LSTM model, the APCNN-LSTM model, the AP-ALSTM model with attention mechanism, and the AP-A-CNN-LSTM model. To ensure the objectivity and fairness of the control experiment, all models use AP clustering to divide the station group and predict the nine stations of the typical station group simultaneously. The parameter selection method of the comparison model is the same as that of the LSTNet model, and the grid search method is used to optimize the parameters of the comparison model.

The experimental results are shown in Table 5 and Figure 11. The PV power prediction curves of power station 1 from 5:00 to 20:00 in spring, summer, autumn, and winter are randomly selected for display. Table 5 shows the average  $e_{MAE}$ ,  $e_{MAPE}$ , and  $e_{RS}$  of nine power stations in each season.

As shown in Table 5, AP-LSTNet has a lower prediction error than each comparison model in different seasons, and the  $e_{RS}$  of the model is above 96%. The accuracy of  $e_{MAE}$  and  $e_{MAPE}$  of AP-A-LSTM and AP-A-CNN-LSTM models with attention mechanism is improved compared with the original model, indicating that the attention mechanism can improve the ability to extract key feature information. Compared with the suboptimal AP-A-LSTM model, the

$e_{MAE}$  and  $e_{MAPE}$  of the AP-LSTNet model in four seasons are further reduced by more than 4%, and the  $e_{RS}$  is the optimal value in four seasons, indicating that the loop-skip layer in this model can extract the time series characteristics of ultra-long time series more effectively.

To further verify the prediction effectiveness of the AP-LSTNet model under different weather types, the prediction results and measured power generation data of six prediction models of typical days under different weather types of Power Plant 1 are shown in Table 6 and Figure 12. When comparing the prediction results under different weather types, AP-A-LSTM and AP-CNN-LSTM, the prediction error  $e_{MAE}$  of the AP-LSTNet model is significantly lower than that of the other models under the four weather types. The prediction error of the model is the lowest in sunny weather with an  $e_{MAE}$  of 0.62 MWh. The prediction error of the model is the highest in sunny to cloudy conditions, and the  $e_{MAE}$  is 1.59 MWh. Compared to the contrast model, the model has a significant improvement in accuracy in cloudy and showery weather.

## 6. Conclusions

In this paper, a short-term PV power plant group forecasting model, the AP-LSTNet model, is proposed. According to the seasonal variation, four forecasting units are constructed to simultaneously predict the output power of multiple power plants in the group. Simulation results show that the prediction results of the AP-LSTNet model are closest to the actual output. The main conclusions are as follows:

- (1) AP clustering is used to divide multiple distributed PV power plants into small power plant clusters so that the meteorological data in each power plant cluster is consistent. The meteorological data are encrypted by the linear interpolation method, and the coverage of key meteorological data in the whole region is realized.
- (2) Meteorological features were added to the LSTNet model, and the short-term dependence and cycle repetition patterns between PV sequence and meteorological factors were considered. The nonlinear branches composed of different neural networks and linear branches composed of autoregressions were integrated to perform feature mining, and more accurate prediction results were obtained.
- (3) In the same PV power station group, the PV power of multiple stations has a strong coupling. Compared with single-station prediction, the LSTNet model can better learn the dependence of multistation PV output power, improve the accuracy of regional PV prediction, and reduce the training cost.

It is necessary to further improve the clustering algorithm to accurately partition the distributed PV by integrating geographical location information, meteorological characteristics, and other factors; improve the accuracy of regional PV power plant group prediction; and explore the

use of high-precision data enhancement technology to solve the problems of distributed PV power data and lack of key weather elements at low cost.

## Data Availability

The data that is used to support the findings of this study are available from the corresponding author upon reasonable request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Authors' Contributions

X.L. and G.Y. conceptualized the study, contributed to visualization, performed funding acquisition, performed supervision, contributed to validation, performed formal analysis, investigated the study, contributed to data curation, and performed project administration. X.L. and J.G. proposed the methodology. X.L. was in charge of software and writing of the original draft. X.L., G.Y., and J.G. wrote, reviewed, and edited the manuscript. All authors have read and agreed to the published version of the manuscript.

## Acknowledgments

This research was funded by the Science and Technology Development Guiding Plan Project Lanzhou under Grant No. 2022-5-36; Lanzhou City Science and Technology Programme Project No. 2022-2-22; 2023 Lanzhou Resources & Environment Voc-Tech University Research Capacity Enhancement Project No. X2023A-13; Lanzhou Resources & Environment Voc-Tech University, Yellow River Basin Ecotope Integration of Industry and Education R&D Fund under Grant No. XHYF2023-02; and 2020 University-Level Scientific Research Project under Grant No. Y2020B-02.

## References

- [1] A. Arguello, J. D. Lara, J. D. Rojas, and G. Valverde, "Impact of rooftop PV integration in distribution systems considering socioeconomic factors," *IEEE Systems Journal*, vol. 99, pp. 1–12, 2017.
- [2] L. Zhang, M. Zhang, W. Ji, L. Fang, Y. Qin, and L. Ge, "Virtual acquisition method for operation data of distributed PV applying the mixture of grey relational theory and bp neural work," *Electr. Power Constr.*, vol. 42, pp. 125–131, 2021.
- [3] T. AlSkaif, S. Dev, L. Visser, M. Hossari, and W. van Sark, "A systematic analysis of meteorological variables for PV output power estimation," *Renewable Energy*, vol. 153, pp. 12–22, 2020.
- [4] Y. Jin and R. Zhang, "Short term photovoltaic output prediction based on singular spectrum analysis," in *Proceedings of the 2021 3rd Asia Energy and Electrical Engineering Symposium (AEEES)*, pp. 903–910, IEEE, Chengdu, China, March 2021.
- [5] L. Shan, X. Peizhe, and Z. Guohui, "Short-term forecasting of PV power based on the fuzzy clustering algorithm and support vector machine in smart distribution planning," in *Proceedings of the 2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, pp. 643–647, Chongqing, China, October 2018.
- [6] G. Wang, M. Yang, and Y. Yu, "A short-term forecasting method for photovoltaic power based on ensemble adaptive boosting random forests," in *Proceedings of the 2020 IEEE/IAS 56th Industrial and Commercial Power Systems Technical Conference (I&CPS)*, pp. 1–8, Las Vegas, NV, USA, June 2020.
- [7] X. J. Dong, J. N. Shen, G. X. He, Z. F. Ma, and Y. J. He, "A general radial basis function neural network assisted hybrid modeling method for photovoltaic cell operating temperature prediction," *Energy*, vol. 234, 2021.
- [8] Y. Nie, Y. Sun, Y. Chen, R. Orsini, and A. Brandt, "PV power output prediction from sky images using convolutional neural network: the comparison of sky-condition-specific sub-models and an end-to-end model," *Journal of Renewable and Sustainable Energy*, vol. 12, no. 4, 2020.
- [9] H. Zhang, Y. Zhao, H. Kang, E. Mei, and H. Han, "Multi-input deep convolutional neural network model for short-term power prediction of photovoltaics," *Computational Intelligence and Neuroscience*, vol. 2022, 2022.
- [10] A. Agga, A. Abbou, M. Labbadi, and Y. El Houm, "Short-term self-consumption PV plant power production forecasts based on hybrid CNN-LSTM, ConvLSTM models," *Renewable Energy*, vol. 177, pp. 101–112, 2021.
- [11] A. Mellit, A. M. Pavan, and V. Lughi, "Deep learning neural networks for short-term photovoltaic power forecasting," *Renewable Energy*, vol. 172, pp. 276–288, 2021.
- [12] K. Wang, X. Qi, and H. Liu, "Photovoltaic power forecasting based LSTM-Convolutional Network," *Energy*, vol. 189, 2019.
- [13] S. C. Lim, J. H. Huh, S. H. Hong, C. Y. Park, and J. C. Kim, "Solar power forecasting using CNN-LSTM hybrid model," *Energies*, vol. 15, no. 21, p. 8233, 2022.
- [14] F. Wang, Z. Xuan, Z. Zhen, K. Li, T. Wang, and M. Shi, "A day-ahead PV power forecasting method based on LSTM-RNN model and time correlation modification under partial daily pattern prediction framework," *Energy Conversion and Management*, vol. 212, 2020.
- [15] M. N. Akhter, S. Mekhilef, H. Mokhlis et al., "An hour-ahead PV power forecasting method based on an RNN-LSTM model for three different PV plants," *Energies*, vol. 15, no. 6, p. 2243, 2022.
- [16] H. Zhou, Y. Zhang, L. Yang, Q. Liu, K. Yan, and Y. Du, "Short-term photovoltaic power forecasting based on long short term memory neural network and attention mechanism," *IEEE Access*, vol. 7, pp. 78063–78074, 2019.
- [17] G. Yu, L. Lu, B. Tang, S. Wang, X. Yang, and R. S. Chen, "An improved hybrid neural network ultra-short-term photovoltaic power forecasting method based on cloud image feature extraction," in *Proceedings of the Chinese Society for Electrical Engineering*, pp. 1–14, Canada, March, 2021.
- [18] N. Chen, L. Qu, Y. Jia, H. Zhang, and L. Zhang, "Analysis and discussion on technical requirements for operation of photovoltaic power generation," in *Proceedings of the 2021 3rd Asia Energy and Electrical Engineering Symposium (AEEES)*, pp. 922–927, IEEE, Chengdu, China, March, 2021.
- [19] X. G. Agoua, R. Girard, and G. Kariniotakis, "Short-term spatiotemporal forecasting of photovoltaic power production," *IEEE Transactions on Sustainable Energy*, vol. 9, no. 2, pp. 538–546, 2018.
- [20] B. Liu, D. Ke, P. Li, J. Xu, H. Bai, and L. Yu, "Identification method of distributed photovoltaic power in low-voltage distribution networks," *Automation of Electric Power Systems*, vol. 43, no. 19, pp. 111–116, 2019.



- [21] K. Hu, L. Wang, W. Li, S. Cao, and Y. Shen, "Forecasting of solar radiation in photovoltaic power station based on ground-based cloud images and BP neural network," *IET Generation, Transmission and Distribution*, vol. 16, no. 2, pp. 333–350, 2022.
- [22] Q. Li, X. Zhang, T. Ma, D. Liu, H. Wang, and W. Hu, "A Multi-step ahead photovoltaic power forecasting model based on TimeGAN, Soft DTW-based K-medoids clustering, and a CNN-GRU hybrid neural network," *Energy Reports*, vol. 8, pp. 10346–10362, 2022.
- [23] D. Nasri, D. Mokeddem, B. Bourouba, and J. Bosche, "A novel levy flight trajectory-based salp swarm algorithm for photovoltaic parameters estimation," *Journal of Information and Optimization Sciences*, vol. 42, no. 8, pp. 1841–1867, 2021.
- [24] T. L. Jiao, J. M. Zhang, and X. Li, "Spatial clustering method for large-scale distribution user photovoltaic based on spatial correlation," *Automation of Electric Power Systems*, vol. 43, pp. 97–102, 2019.
- [25] G. Li, Y. Cheng, C. Wang, J. Shi, and L. Li, "Ultra short-term output forecasting of distributed photovoltaic power station based on feature extraction," in *Proceedings of the 2020 IEEE/IAS Industrial and Commercial Power System Asia (I&CPS Asia)*, pp. 1114–1119, IEEE, Weihai, China, July, 2020.
- [26] Y. Qiao, R. Sun, R. Ding, S. Li, and Z. Lu, "Distributed photovoltaic station cluster gridding short-term power forecasting Part I: methodology and data augmentation," *Power System Technology*, vol. 5, pp. 1799–1808, 2021.
- [27] M. Paulescu, R. Blaga, C. Dughir et al., "Intra-hour PV power forecasting based on sky imagery," *Energy*, vol. 279, 2023.
- [28] T. Limouni, R. Yaagoubi, K. Bouziane, K. Guissi, and E. H. Baali, "Accurate one step and multistep forecasting of very short-term PV power using LSTM-TCN model," *Renewable Energy*, vol. 205, pp. 1010–1024, 2023.
- [29] S. U. Sabareesh, K. S. N. Aravind, K. B. Chowdary, S. Syama, and K. Devi V S, "LSTM based 24 hours ahead forecasting of solar PV system for standalone household system," *Procedia Computer Science*, vol. 218, pp. 1304–1313, 2023.
- [30] Y. Liu, Y. Liu, H. Cai, and J. Zhang, "An innovative short-term multihorizon photovoltaic power output forecasting method based on variational mode decomposition and a capsule convolutional neural network," *Applied Energy*, vol. 343, 2023.
- [31] J. Simeunović, B. Schubnel, P. J. Alet, R. E. Carrillo, and P. Frossard, "Interpretable temporal-spatial graph attention network for multi-site PV power forecasting," *Applied Energy*, vol. 327, 2022.
- [32] L. Visser, T. AlSkaif, and W. van Sark, "Operational day-ahead solar power forecasting for aggregated PV systems with a varying spatial distribution," *Renewable Energy*, vol. 183, pp. 267–282, 2022.
- [33] S. Li, H. Zhang, L. Ding, Y. Chen, and A. Li, "Research of distributed photovoltaic output fluctuation suppression method based on improved FPA," *Journal of Physics: Conference Series*, vol. 2418, no. 1, 2023.
- [34] C. Zhang, Y. Zhang, J. Zhao, and J. Luo, "Sensitivity analysis of weather factors affecting PV module output power based on artificial neural network and sobol algorithm," in *Proceedings of the 2020 IEEE/IAS Industrial and Commercial Power System Asia (I&CPS Asia)*, pp. 246–250, IEEE, Weihai, China, July, 2020.
- [35] H. K. Yadav, Y. Pal, and M. M. Tripathi, "Short-term PV power forecasting using empirical mode decomposition in integration with back-propagation neural network," *Journal of Information and Optimization Sciences*, vol. 41, no. 1, pp. 25–37, 2020.
- [36] Q. Lei and T. Li, "Semi-supervised selective affinity propagation ensemble clustering with active constraints," *IEEE Access*, vol. 8, pp. 46255–46266, 2020.
- [37] L. Sun, C. Guo, C. Liu, and H. Xiong, "Fast affinity propagation clustering based on incomplete similarity matrix," *Knowledge and Information Systems*, vol. 51, no. 3, pp. 941–963, 2017.
- [38] Y. Lu and Y. Li, "A novel fault diagnosis method for circuit breakers based on optimized affinity propagation clustering," *International Journal of Electrical Power and Energy Systems*, vol. 118, 2020.
- [39] G. Wang, C. Xing, J. S. Wang, H. Y. Wang, and J. X. Liu, "Clustering validity evaluation functions of fuzzy C-means clustering algorithm," *IAENG International Journal of Computer Science*, vol. 49, no. 2, 2022.
- [40] A. Koirala, M. U. Hashmi, D. Van Hertem, and R. D'hulst, "Representative feeders for spatial scaling of stochastic PV hosting capacity," in *Proceedings of the 2022 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, pp. 1–6, IEEE, Novi Sad, Serbia, October, 2022.
- [41] S. Ransome, "Benchmarking PV performance models with high quality IEC 61853 Matrix measurements (Bilinear interpolation, SAPM, PVGIS, MLM and 1-diode)," in *Proceedings of the 2022 IEEE 49th Photovoltaics Specialists Conference (PVSC)*, pp. 0375–0380, IEEE, Philadelphia, PA, USA, June, 2022.
- [42] G. Lai, W. C. Chang, Y. Yang, and H. Liu, "Modeling long-and short-term temporal patterns with deep neural networks," in *Proceedings of the 41st international ACM SIGIR conference on research and development in information retrieval*, pp. 95–104, Shenzhen, China, June, 2018.
- [43] Y. Q. Zhu, Y. M. Cai, and F. Zhang, "Motion capture data denoising based on LSTNet autoencoder," *Journal of Internet Technology*, vol. 23, no. 1, pp. 11–20, 2022.
- [44] M. Gao, J. Li, F. Hong, and D. Long, "Short-term forecasting of power production in a large-scale photovoltaic plant based on LSTM," *Applied Sciences*, vol. 9, no. 15, p. 3192, 2019.
- [45] J. Vicente-Gabriel, A. B. Gil-González, A. Luis-Reboredo, P. Chamoso, and J. M. Corchado, "LSTM networks for overcoming the challenges associated with photovoltaic module maintenance in smart cities," *Electronics*, vol. 10, no. 1, p. 78, 2021.