A Review of Multitemporal and Multispatial Scales Photovoltaic Forecasting Methods

CAICHENG LIU¹, MING LI¹, YUNJUN YU¹,², (Member, IEEE), ZIYANG WU³, HAI GONG³, AND FEIER CHENG¹

¹College of Information Engineering, Nanchang University, Nanchang 330027, China
²Institute of AI, Nanchang University, Nanchang 330027, China
³Samueli School of Engineering, Irvine, CA 92697, USA

Corresponding author: Yunjun Yu (yuyunjun@ncu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61563034, and in part by the International S&T Cooperation Program of China under Grant 2014DFG72240.

ABSTRACT Reliable photovoltaic (PV) forecasting can provide important data support for power system operation, which is the key to realize the large-scale consumption of solar energy resources. PV forecasting task becomes crucial to ensure power system stability and economic operation. This paper reviews the existing research of PV forecasting methods from the perspective of multi-temporal scale and multi-spatial scale. Firstly, according to the forecasting process, demand, temporal and spatial scale, the forecasting methods are classified and the evaluation indicators involved in the research are listed. Secondly, based on the temporal scale of PV power generation, the results are combed through the three kind of scale of ultra-short-term, short-term and medium and long-term prediction. Thirdly, on each kind of temporal scale, the results are subdivided into single-site prediction and regional prediction to sort out in detail. Finally, the results are analyzed on the basis of the predicted temporal scale, spatial scale and input data. It has been observed that most recent papers highlight the importance of short-term predictions. The machine learning method shows excellent nonlinear description ability in short-term prediction, the prediction results are satisfactory. The spatial average effect of regional prediction reduces the variability of solar energy, the prediction results are reliable.

INDEX TERMS Solar energy, photovoltaic forecasting, machine learning, multi-temporal scale, multi-spatial scale.

I. INTRODUCTION
Solar energy engineering is one of the energy alternatives for global sustainable development goals, photovoltaic panels are the biggest driving force for the rapid growth of solar power generation [1], [2]. According to the “Renewable Energy Installed Capacity Statistics 2020” report released by the International Energy Agency, the total global installed photovoltaic capacity reached 578GW in 2019, and the new global installed capacity of photovoltaic is expected to have a year-on-year growth of 14% in 2020 [3]. However, photovoltaic systems are easily affected by environmental factors such as sunlight, seasons, time, geographic location, etc., resulting in the characteristics of stochastic, intermittent and variable of photovoltaic [4]. These characteristics will reduce the efficiency of photovoltaic output and the power quality of grid connected photovoltaic. The reliability and stability of the power supply connected to the main network will also be affected [5], [6]. With the penetration of large-scale solar energy resources into the power grid, how to ensure the effective consumption of variable and stochastic solar energy resources under the normal operation of the power system is an urgent problem. Reliable prediction of photovoltaic can provide important data support for power system operation, which is the basis and key to realize the large-scale consumption of solar energy resources [7], [8]. Photovoltaic forecasting can help to adjust the scheduling plan timely and improve the efficiency of solar energy utilization. Improving the accuracy of photovoltaic prediction is of great significance to the safe and stable operation of the system and the maintenance of power quality [9], [10].

The existing research has analyzed the prediction methods of photovoltaic and solar radiation from different angles. From the perspective of forecasting scales, some studies...
consider the temporal and spatial scales. On the temporal scale, Wang et al. [11] reviewed the photovoltaic prediction methods based on deep learning, and they compared and analyzed the prediction effects of different deep learning models under the time scale. Barbieri et al. [12] summarized the specific research and application of large-scale photovoltaic grid connected power generation field. According to the grid operation considered, the authors defined the prediction time range and spatial-temporal resolution. Ellahi et al. [13] investigated the photovoltaic prediction technology for the availability of resource supply in different periods. They considered the supply scheduling mechanism extracted from these resources. At the spatial scale, Antonanzas et al. [14] summarized the solar power generation prediction methods from the perspective of time prediction range from hours to days or weeks and spatial scale from single site to regional power group prediction. Furthermore, the prediction accuracy was included in the economic evaluation of the impact on the power grid. Pazikadin et al. [15] summarized the application of artificial neural network in the prediction of solar power generation, and they analyzed and discussed the instruments for measuring solar irradiance. Based on the temporal and spatial scales, the authors evaluated the prediction performance of different types of neural networks for photovoltaic. In view of the vulnerability problem caused by intermittent solar and wind energy to power grid, Ssekulima et al. [16] introduced the current main prediction methods of physics, statistics and artificial intelligence in time and space, and then proposed key prediction problems and future development trend.

From the perspective of forecasting methods, conventional methods and ensemble methods are usually considered. For a single method, Ssekulima et al. [16] and Sobri et al. [17] reviewed photovoltaic prediction technology from the perspective of current main physical models, time series statistics and artificial intelligence. Compared with different prediction methods, artificial intelligence method that can solve the nonlinear and complex structure of data has been widely used. Digne et al. [18] introduced the prediction technology based on cloud image or numerical weather forecast model when they reviewed the current solar radiation intensity prediction methods. In addition, Dliento et al. [19] reviewed the research progress of photovoltaic system monitoring, diagnosis and artificial neural network prediction methods in recent years. Akhter et al. [20] reviewed and compared machine learning and meta-heuristic methods based on historical data, prediction horizon and input parameters. Moreover, they summarized the advantages and disadvantages of various methods. Das et al. [21] and Mellit et al. [22] comprehensively and systematically reviewed the direct prediction of photovoltaic, and they discussed the importance of input-output data correlation and model input data preprocessing. The authors considered the advantages and disadvantages of different prediction models, which include the hybrid model and the performance matrix of the evaluation model. Law et al. [23] summarized the prediction accuracy of numerical weather prediction model, time series analysis method, cloud motion vector and hybrid method.

From the perspective of other reviews, Viscondi et al. [24] conducted a literature review on the big data model of solar photovoltaic prediction by using the method of system literature review (SLR), mainly introducing machine learning and data mining technology. The authors aim to evaluate the most applicable and accurate technology for photovoltaic prediction. Yang et al. [25] introduced how to use text mining technology for studying solar radiation and photovoltaic power prediction. The proposed text mining technology can be transferred to other solar engineering topics. Van der Meer et al. [26] mainly introduced the latest progress in solar energy probabilistic forecasting and load forecasting. Furthermore, the research focused on uncertainty prediction and probabilistic prediction. Lai et al. [27] analyzed three types of photovoltaic forecasting demand (point forecast, interval forecast, probability forecast), and they respectively sorted out, categorized, summarized and reviewed the current forecasting methods and measures of forecasting effects. The factors affecting the prediction accuracy of photovoltaic power generation and the typical technical means to improve the prediction accuracy are summarized.

Distributed grid related operations which include unit commitment, planning, scheduling, load balancing and load following are performed on different spatial and temporal scales. Photovoltaic forecasting related to these operations also needs to be considered about different spatial and temporal scales. Existing photovoltaic forecasting review papers analyze and study photovoltaic forecasting methods from the perspective of time scale or space scale, but they do not combine time scale and space scale to discuss photovoltaic forecasting methods. In this paper, the photovoltaic prediction methods will be taxonomically and comparatively studied from the view of combining time scale and spatial scale, meticulous to the comparison of methods at the same spatial scale on different time scales and the comparison of methods at different spatial scales on the same time scale. The differences and correlations of PV prediction methods between different time scales and different spatial scales are discussed.

The structure of this paper is as follows: Section II classifies the photovoltaic prediction method according to the prediction process, forecast demand, forecast spatial scale and temporal scale, and evaluation metrics involved in this study are listed. According to the time scale, photovoltaic forecasting methods are divided into three categories: ultra-short-term(Section III), short-term(Section IV) and medium and long-term(Section V). In each section, the prediction method is subdivided in terms of the spatial scale (single-site, regional). Section VI discussion the correlations and differences of prediction methods on the same spatial scale at different temporal scales and different spatial scales at the same temporal scale. Section VII summarizes and concludes the paper.
II. CLASSIFICATION AND EVALUATION METRICS OF PHOTOVOLTAIC FORECASTING METHODS

A. CLASSIFICATION OF PHOTOVOLTAIC FORECASTING METHODS

With a large number of photovoltaics connected to the grid, accurate photovoltaic forecasting method can provide information support for grid dispatching plans and reduce the phenomenon of “abandonment of solar energy”. According to different forecasting standards, forecasting methods can be divided into many categories. The classification of photovoltaic forecasting methods is shown in Figure 1. Based on the different forecasting process, it can be divided into direct forecasting and indirect forecasting [28]. For different forecasting needs, it can be divided into point forecasting, interval forecasting and probability forecasting [29]. According to different forecasting time scales, it can be divided into ultra-short-term forecasting (0-4h), short-term forecasting (4h-72h) and medium and long-term forecasting (1m-1y) [30]. In the light of different forecasting spatial scales, it can be divided into single site prediction and regional prediction [31]. For different forecasting processes and different forecasting demands, photovoltaic can be predicted from the perspective of temporal-spatial scale. This article will focus on comprehensively combing, summarizing and reviewing photovoltaic forecasting methods from the perspective of different temporal scales and spatial scales.

B. EVALUATION METRICS OF PHOTOVOLTAIC FORECAST METHODS

The performance and accuracy of the photovoltaic prediction model can be evaluated through several metrics [32]. According to different forecast demands, there are different evaluation metrics. The evaluation metrics involved in this research are shown in Table 1.

III. ULTRA-SHORT-TERM PHOTOVOLTAIC FORECAST METHODS

Ultra-short-term forecast refers to the forecast time range between 0 and 4 hours, and its time resolution is small. When sudden large-scale fluctuations in photovoltaic occur, accurate ultra-short-term photovoltaic prediction can obtain power transient information, which can provide powerful
TABLE 1. The evaluation metrics involved in this paper.

| Metrics | Equation | Description |
|---------|----------|-------------|
| RMSE    | \[ \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left( \frac{F_i - P_i}{P_i} \right)^2} \] | evaluate the overall accuracy of the prediction, and punish large prediction errors with the square level. |
| MSE     | \[ \sqrt{\frac{1}{N} \sum_{i=1}^{N} (P_i - F_i)^2} \] | refers to the expected value of the square of the difference between the predicted value and the real value, and evaluates the change degree of the predicted value. |
| MBE     | \( \frac{1}{N} \sum_{i=1}^{N} (P_i - F_i) \) | evaluates the average deviation between predicted and true values. |
| MAPE    | \( \frac{1}{N} \sum_{i=1}^{N} \left| \frac{P_i - F_i}{P_i} \right| \) | suitable for evaluating unified prediction error. |
| MAE     | \( \frac{1}{N} \sum_{i=1}^{N} |P_i - F_i| \) | suitable for evaluating unified prediction error. |
| R       | \[ \frac{\text{Cov}(P, F)}{\text{Var}(P)\text{Var}(F)} \] | represents the linear correlation between predicted value and actual value. |
| R2      | \[ 1 - \frac{\text{Var}(P - F)}{\text{Var}(P)} \] | represents the correlation between predicted value and actual value. |
| CRPS    | \( \frac{1}{N} \int_{-\infty}^{\infty} (F(x) - H(y - x))^2 dy \) | measures the difference between the predicted value distribution and the true value distribution in the probability prediction case. |
| PICP    | \[ \frac{1}{N} \sum_{i=1}^{N} \lambda_i \] | refers to the probability that the actual value of the predicted target falls within the prediction area. |

The main challenges of ultra-short-term photovoltaic power prediction are as follows: 1) the time scale of ultra-short-term prediction is small, the generation, dissipation and movement of clouds are the main reasons for the fluctuation of photovoltaic output. How to mine the relationship between cloud cover and ground irradiance is one of the means to improve the prediction accuracy. 2) Due to the characteristics of fluctuation and stochasticity of photovoltaic power generation and the lack of appropriate descriptive models, it is difficult to provide a reference prediction model for researchers. Therefore, researchers need to keep trying to find out the factors that affect photovoltaic in accordance with different prediction situations, and then formulate appropriate ultra-short-term prediction model for photovoltaic. The following analyzes the ultra-short-term photovoltaic prediction methods from the perspective of spatial scale (single-site, regional).

A. SINGLE-SITE PHOTOVOLTAIC FORECAST METHODS

Photovoltaic utilizes the photovoltaic effect, and ground irradiance is the decisive factor that directly affects the photovoltaic output. Single-site ultra-short-term photovoltaic forecasting methods are mainly divided into two categories: cloud image-based and data-driven.

1) SINGLE-SITE PHOTOVOLTAIC FORECASTING METHODS BASED ON CLOUD-IMAGES

Photovoltaic is easily affected by meteorological factors. When the forecast time range is less than 4 hours, the generation, dissipation and movement of clouds are the fundamental factors to generate the fluctuation of photovoltaic output. At the same time, the solar radiation received by the photovoltaic power station will also be influenced by the geographical environment. In order to study the relationship between cloud cover and ground irradiance, researchers often adopt cloud image-based prediction methods, which mainly combine ground-based cloud images, satellite observation cloud images or radar observation cloud images. Firstly, the cloud above the photovoltaic power station is observed by image acquisition instruments (all sky imager, satellite, radar, etc.). The cloud images are collected according to the fixed sampling time. Then the cloud image is processed and classified by pattern recognition. Finally, the movement trend of cloud clusters is predicted to provide data support for ground irradiance prediction. The flow chart of prediction algorithm based on cloud chart is shown in Fig.2.

Hu et al. [33] proposed an ultra-short-term prediction model that based on the dynamic characteristics of overcast clouds. The motion vector was used to predict the motion trajectory of the cloud, and the cloud blocking the sun would be selected. Digital image processing technology was
adopted to extract the dynamic characteristics of the target cloud that has a greater impact on photovoltaic power generation. Meteorological data and historical photovoltaic data were chosen as radial basis function (RBF) neural network model inputs to predict photovoltaic. Venugopal et al. [34] established a model based on convolutional neural network (CNN) to predict the power output of photovoltaic panels in the next 15 minutes. In addition, the input data were photovoltaic power output data and ground Sky images in the past 15 minutes. Ai et al. [35] emplaced a fisheye camera to collect sky images. The authors implemented optical flow and an adaptive threshold scheme for cloudy days to predict cloud movement. They build a support vector machine (SVM) prediction model to predict solar irradiance. Chu et al. [36] proposed an automatic Smart Adaptive Cloud Identification (SACI) system for sky image and solar irradiance prediction. The system applied off-the-shelf fisheye cameras for deployment, and it adopted Smart Image Classification (SIC) algorithms to combine sky images and solar irradiance measurements. The cloud layer retrieved by SACI is used as the input of the artificial neural network (ANN) model to predict the global average irradiance.

2) SINGLE-SITE PHOTOVOLTAIC FORECASTING METHODS BASED ON DATA-DRIVEN

The stochasticity and variability of photovoltaic mainly occur in the period with short prediction time scale. The output power of photovoltaic will change due to the location, environment, forecast time and season of photovoltaic power station, and it will be difficult to detect because of the impact of noise. Therefore, the ultra-short-term prediction mainly considers how to mine these characteristics from the data and deal with the current sequence to match the features for getting the predicted value. The mining features can also be used to summarize the error change patterns between the predicted value and the real value so as to help correct the predicted value. The framework is shown in Fig.3.

Data-driven ultra-short-term prediction models include: decision tree (DT) models, regression models, markov chain (MC) models, neural network models (NN) and hybrid models. These models can achieve multi-step predictions. For decision trees, Ibrahim et al. [37] adopted Random Forest (RF) technology to predict the hourly solar radiation, and then they used the firefly algorithm to optimize the number of trees and leaves in each tree. Compared with the prediction data of ANN and SVM, the experimental results of RF model are closer to the real data. In terms of regression model, Yang et al. [38] used the lasso model to automatically reduce and select the appropriate lagging solar radiation time series data for regression. This model predicted the solar radiation for 5 minutes at a time interval of 1s. The experimental results show that the proposed model is significantly better than univariate time series models and least squares (LS) regression, especially in the case of less training data and more predictors. Regarding the MC model, Hocaoglu et al. [39] assumed that solar radiation data appeared repeatedly in historical data, and then they adopted Mycielski algorithm to check historical solar radiation data for predicting the data in time. In order to ensure that all sub-patterns obtained from Mycielski most likely history of the sub-mode, the authors took advantage of the MC to transform the measurement data into state. About the neural network model, Qing et al. [40] proposed a new solution for day-ahead solar irradiance forecasting using weather forecast data. The forecasting problem was described as a structured output forecasting problem that predicts multiple outputs simultaneously. This scheme used the data collected by the numerical weather prediction (NWP) as the input of the long short-term memory (LSTM) prediction model. In order to improve the prediction accuracy, the correlation between several consecutive hours in the same day was considered. For the hybrid model, Chen et al. [41] proposed the K-nearest neighbor(K-NN) method to predict the meteorological data of 8 stations near the target station. The authors applied the ANN model to model the available relationship between the K-NN prediction and the solar radiation value of the target site to predict the solar radiation value of the next hour.

B. FORECASTING METHODS OF REGIONAL PHOTOVOLTAIC POWER STATIONS

Regional photovoltaic power station group usually refers to the collection of a certain scale of photovoltaic power station in geographical location. These photovoltaic power stations are often located in similar or related geographical locations, and the power generated is usually packaged and connected to the grid. The power generation of each photovoltaic power station in the photovoltaic power station group is related due to the connection of meteorological change process.

Based on the regional prediction process, the prediction methods can be divided into three categories: 1) superposition method: the output of all photovoltaic power stations in the region is predicted, and then the regional output value is obtained by simply adding the predicted results of single
Researchers choose clustering algorithm to divide the whole region into several sub regions. In order to obtain the spatial mapping relationship between the target photovoltaic power station and the reference photovoltaic power station, Zhang et al. [43] adopted hierarchical clustering algorithm to judge and match the spatial correlation between photovoltaic power stations. The back propagation neural network algorithm was applied to predict the ultra-short-term power of the reference photovoltaic power station, and the predicted results were used as the input of the spatial mapping relationship to obtain the output power of the target power station. Jiao et al. [44] took cluster analysis method to divide photovoltaic power stations of each type of weather type according to the region, and then they formed a group of time-space photovoltaic prediction through distributed power stations with meteorological consistency. Saint et al. [45] proposed a probability area photovoltaic model, which uses the average photovoltaic model with very limited input number (two module orientation angles) to calculate the generation of the most frequent module orientation angles. The probability of their occurrence was finally weighted according to the generated power value to estimate the actual photovoltaic output. Saint et al. [46] took k-means clustering method to divide the region into multiple sub-regions and used bayesian method to parameterize the regional photovoltaic model. The deviation of the initial guess and variance matrix is restricted to standardize the linear system. Generally speaking, cluster analysis consists of the following four basic steps, which are equally important and closely related, and a feedback pathway that indicates that cluster analysis consists of a series of trials and repetitions rather than a one-shot process (see Fig.4).

C. SUMMARY OF ULTRA-SHORT-TERM PHOTOVOLTAIC FORECASTING METHODS

This section mainly analyzes the ultra-short-term photovoltaic forecasting methods, which are divided into single-site prediction and regional power station group prediction according to spatial scale. For single site prediction, the photovoltaic forecasting methods based on cloud images and data-driven are summarized. About regional power station group prediction, the prediction method has spatial correlation obviously. The relationship between single photovoltaic station and the whole power station group is analyzed, and the regional ultra-short-term prediction method is described.

As the prediction time scale is 0-4h, the fluctuation of single-site prediction is caused by the movement and dissipation of clouds. The cloud images-based prediction method uses cloud amount and historical data as the input of the model so as to improve the prediction accuracy. The prediction method based on data-driven improves the prediction accuracy from the perspective of data-driven. It selects the model with strong ability to deal with nonlinear predict problem. Due to the wide coverage of the regional photovoltaic power station group and the inconsistent weather changes in the area, the idea of clustering algorithm can be fully utilized to divide the entire area into multiple sub-areas. In addition, the sub regions can be predicted to fit the output power curve with higher accuracy.

Table 2 compares the prediction methods, input data and prediction errors in different spatial scales for ultra-short-term photovoltaic forecasting. In terms of prediction error, there is no uniform evaluation standard for the prediction error of solar radiation and photovoltaic in the world. In addition, Sources of used data are not the same. Therefore, there is no comparability between the prediction errors.

IV. SHORT-TERM PHOTOVOLTAIC FORECAST METHODS

Short-term photovoltaic prediction refers to the prediction time range of 4h 72h. The prediction of short-term photovoltaic can provide help for power system dispatching and management departments to formulate daily generation plan of conventional power supply and adjust maintenance plan. It is helpful to optimize frequency modulation and spinning reserve capacity. Moreover, accurate prediction can provide effective data support for micro grid energy management, power grid dispatching management, pricing, load management and other fields.
TABLE 2. Comparison of ultra-short-term photovoltaic forecasting methods.

| reference | temporal scale | spatial scale | method                  | input                                                                 | error          |
|-----------|----------------|---------------|-------------------------|-----------------------------------------------------------------------|----------------|
| [33]      | 1h             | Single-site   | RBF                     | PV sequence, solar irradiance, cloud image, temperature.              | RMSE: 7.4%     |
| [34]      | 15min          | Single-site   | CNN                     | PV sequence, solar irradiance, cloud image.                          | RMSE: 14.99%   |
| [35]      | 1min           | Single-site   | SVM                     | solar irradiance, cloud image.                                       | RMSE: 11%      |
| [36]      | 5min           | Single-site   | ANN                     | solar irradiance, cloud image.                                       | RMSE: 14%      |
| [37]      | 1h             | Single-site   | RF + Firefly Algorithm  | sunshine rate, humidity, temperature.                                | RMSE: 18.98%   |
| [38]      | 5min           | Single-site   | Lasso Algorithm         | solar irradiance.                                                     | MAPE: 6.38%    |
| [39]      | 1h             | Single-site   | Mycielski Algorithm + Markov Chain | solar irradiance.                                        | MBE: 2.86%     |
| [40]      | 1h             | Single-site   | LSTM                    | solar irradiance, temperature, humidity, cloud cover.                | RMSE: 9.68%    |
| [41]      | 1h             | Single-site   | KNN + ANN               | solar irradiance, temperature, humidity, wind speed, wind direction. | MAE: 10.96%    |
| [43]      | 4h             | Regional      | BPNN                    | PV sequence, NWP.                                                    | MAPE: 9.23%-10.77% |
| [44]      | 1h             | Regional      | Probability Model       | PV sequence, solar irradiance, temperature, tilt angle, azimuth angle. | RMSE: less than 5% |
| [45]      | 15min          | Regional      | Pair-Copula Theory + ANN| PV sequence, solar irradiance, temperature.                         | Kendall rank correlation coefficient: 0.8128 |
| [46]      | 4h             | Regional      | K-Means + BN            | PV sequence, solar irradiance, temperature, PV geographic location.  | RMSE: 3.90%    |

For the short-term photovoltaic power forecasting, there are two main difficulties: 1) Owing to the influence of meteorology and environment, short-term forecast presents randomness and uncertainty. How to explore the relationship between photovoltaic, meteorological factors and environmental factors is one of the main difficulties. 2) At present, researchers mostly adopt shallow learning method to predict. But the learning ability of this kind of algorithm is limited. It cannot show excellent prediction level in the deep feature training process, and the prediction accuracy is difficult to improve. In order to improve the prediction accuracy of the model, researchers need to make use of the existing knowledge to solve the problems of data processing, feature extraction and parameter optimization. In this section, the research results of short-term photovoltaic forecasting methods are introduced from the perspective of spatial scale (single-site, regional).

A. SINGLE-SITE PHOTOVOLTAIC FORECAST METHODS

Due to the short prediction time scale, the factors affecting the short-term prediction accuracy of single photovoltaic station generally include the fluctuation and non-stationarity of photovoltaic power generation process and the selection of data interval and reconstruction step size in the prediction process. Finding the variation characteristics in the process of unstable solar radiation or photovoltaic is the key to prediction. It can eliminate the influence of non-stationarity on prediction. The research results of single-site short-term photovoltaic forecasting is analyzed from two aspects: historical data, comprehensive historical and other meteorological data.

1) SINGLE-SITE PHOTOVOLTAIC FORECASTING METHODS BASED ON HISTORICAL DATA

For the prediction method based on historical data, the selection of data interval and reconstruction step size directly affects the prediction results of photovoltaic. Reference [47]–[50] study the influence of the length of historical data on the prediction accuracy. In [51]–[55], some researchers hope to decompose the historical time series by signal decomposition technology. The block diagram of the method that predict photovoltaic/solar radiation by using signal decomposition is shown in Fig.5. First, the non-stationary sequence is decomposed into a stationary component and a random component. The unpredictable random component accounts for a smaller proportion of the overall amplitude, while the stationary component accounts for a larger proportion of the overall amplitude. Then, for each group of components, the predictor of current component is trained through phase space reconstruction analysis or regression analysis. Finally, the predicted value of short-term photoelectric or solar radiation is obtained by using the parameters of signal decomposition to reconstruct the signal. Behera et al. [56] adopted empirical model decomposition (EMD) method to decompose and filter the photovoltaic series at each time interval. To obtain the denoising sequence, they took the local extreme values far away from the entire sequence as outliers and filters them out. Li et al. [57] used wavelet packet decomposition (WPD) technology to decompose and filter the photovoltaic series at each time interval. The authors trained the LSTM prediction model for these four subsequences. The prediction results of each LSTM network are reconstructed, and the final prediction results are obtained.
by linear weighting method. Based on the Variational Mode Decomposition (VMD), Zang et al. [58] decomposed different frequency components from the historical time series of photovoltaic. They constructed frequency components into a two-dimensional data form with correlations in both daily and hourly timescales that can be extracted by convolution kernels. The VMD residual time series were refined into high-level features by a CNN, which could reduce the data size.

Compared with ultra-short-term prediction, the time scale of short-term prediction is extended, and the uncertainty of photovoltaic power generation process will also increase. At this time, the prediction method based on state transition and linear regression may not be applicable. Instead, prediction methods based on deep search for high-dimensional features hidden in data, including NN, SVM, DT and other soft computing methods are used [59]–[65]. These methods take solar radiation and photovoltaic sequences of the past period as input vectors, and the predicted single-step or multi-step solar radiation or photovoltaic as output variables. A low-cost and more robust black box model is used to simulate this dynamic relationship. Some researchers predicted photovoltaic through simulated ANN, which overcomes the limitation of traditional network that is difficult to obtain optimal structure and fall into local optimum [66]. In [67], the authors proposed a probabilistic bayesian learning technique to estimate the parameters of a multilayer perceptron NN. The bayesian framework allows to obtain confidence intervals and estimate the error line predicted by the NN, which helps to simplify the network structure and learning process. In [68], researchers adopted the periodic graph method to extract the random component of photovoltaic power and the least square (LS) method to determine the dynamic weight value of each NN. The random component of photovoltaic is predicted through the combined model, and then superimposed it with the periodic component to obtain the final photovoltaic prediction result.

2) SINGLE-SITE PHOTOVOLTAIC FORECASTING METHODS BASED ON HISTORICAL AND OTHER METEOROLOGICAL DATA

Other meteorological data include measured information of the meteorological environment such as atmospheric temperature, atmospheric pressure, relative humidity and geographic information around photovoltaic power stations as well as include NWP information based on these measured information and dynamic equations. Meteorological information will be packaged with solar radiation information as the input of the photovoltaic prediction methods mentioned in this subsection.

With the addition of meteorological environmental information such as temperature, humidity, air pressure, geographic location and other meteorological environmental information, it is more likely to find factors more relevant to the change process of photovoltaic power generation. For the data-driven prediction method, improving the dimension of input data is more conducive to finding the change characteristic of photovoltaic system. Reference [69] analyzed the relationship between each weather variables and photovoltaic output. It proposed a data-driven recursive arithmetic average integration model, which can iteratively improve the weight of sub independent models according to the recursive process. And this model is suitable for processing large amounts of data. Reference [70] proposed a multi time scale data-driven model to predict single-step photovoltaic based on the spatial and temporal correlation between adjacent solar stations. NWP has been used for a long time. Since 1904, NWP model has been developed. By gradually improving the calculation accuracy, more accurate weather forecast results are obtained [71]. According to the results of NWP in reference [72]–[75], the short-term predicted power can be obtained by using photoelectric conversion model, in which solar irradiance is the most important meteorological factor affecting photovoltaic power generation.

There are studies using feature analysis methods such as clustering or component analysis to find similar photoelectric change processes from historical data and meteorological data. According to different processes, the prediction models are designed to realize the prediction of photoelectric or solar radiation. In [76], the kernel principal component analysis (PCA) was adopted to preprocess the sample data of weather type, temperature, humidity, atmospheric pressure, wind speed, etc. And the obtained characteristic sequence was used as the input of Elman network to predict photovoltaic. Reference [77] discussed the influence of missing data in photovoltaic prediction. For the training and testing data of SVM prediction model, K-NN method was applied to fill the missing data. The prediction effect is good. In [78], Fuzzy C-Means clustering algorithm (FCM) was used to cluster the
the synchronization of change trend between sequences is. The stronger the correlation is, the more obvious is obvious spatial correlation between photovoltaic power stations have certain similarity and delay. Therefore, there and weather change between similar photovoltaic power stations not only has a certain degree of autocorrelation, but it has high similarity in photovoltaic power station group is weaker than that of the single photovoltaic power station.

As a result, the stochasticity and non-stationarity of the generation power of the regional photovoltaic power station group is weaker than that of the single photovoltaic power station.

Photovoltaic not only has a certain degree of autocorrelation in time, but also it has high similarity in photovoltaic sequence of similar geographical locations. The degree of similarity can be described by the spatial correlation of photovoltaic. The uncertainty factors of cloud disturbance and weather change between similar photovoltaic power stations have certain similarity and delay. Therefore, there is obvious spatial correlation between photovoltaic power station. The stronger the correlation is, the more obvious the synchronization of change trend between sequences is.

original meteorological data according to the membership degree. Fig.6 summarizes the methodology framework for short-term photoelectric / solar radiation prediction using abundant meteorological data. Firstly, different types of data are preprocessed with the same resampling interval to form a high-dimensional input vector based on historical data. Then, in order to determine the parameters of the predictor, the high-dimensional regression characteristics of the vector and its delay matrix are analyzed by using the method of correlation analysis, principal component analysis, empirical mode analysis or spectral analysis. Finally, the predicted results of photovoltaic generation/solar radiation can be obtained by input the feature sequences into the predictor.

Short term prediction models such as neural network, support vector machine, time series method, etc. [79]–[92] are widely used in photovoltaic forecasting. By using meteorological data to analyze the randomness and uncertainty of photovoltaic, the generalization ability of the prediction model is improved and the power curve of actual photovoltaic is approximately fitted.

**B. FORECASTING METHODS OF REGIONAL PHOTOVOLTAIC POWER STATIONS**

The forecast of photovoltaic Station Group is the overall estimation of photovoltaic. Assuming that the generation power of N photovoltaic power stations in a group of photovoltaic power stations is the same as opposite distribution, then according to the central limit theorem, the variance of average power generation is 1 / N of the power generation method of a single power station. As a result, the stochasticity and non-stationarity of the generation power of the regional photovoltaic power station group is weaker than that of the single photovoltaic power station.

In [93], the spatiotemporal correlation of photovoltaic data of distributed photovoltaic system was deeply studied. And a bayesian network prediction model based on spatio-temporal correlation was established. Reference [94] conducts spatial clustering of photovoltaic power stations. The power output of each cluster was estimated and predicted by selecting the representative points in the regional centroid. The model output was averaged to obtain the regional prediction results. In [95], sub regions were divided by EMD and hierarchical clustering. The representative power stations were selected based on the minimum redundancy maximum correlation (MRMC) criterion. The Elman NN was used to predict the power generation of representative power stations in each sub region to realize the photovoltaic prediction of the whole region. Reference [96] adopted the generalized parameter covariance function to simulate the correlation between the irradiances of observation points. And it used the Kriging method of space and time to predict solar radiation. In [97], a chain structure of echo state network (ESN) was proposed. Autocorrelation analysis was used in the time information of each spatial variable to provide appropriate input for each ESN module. Reference [98]–[103] analyzed the spatial correlation of the photovoltaic sequence of each power station in the region. It used the similarity of photovoltaic output changes between photovoltaic power stations to obtain the trend of power generation. In [104], a number of solar radiation measurement stations in the northern states of the United States were modeled as spatial undirected graphs. The convolution graph automatic encoder (CGAE) probability model was used to predict the temporal and spatial probability of solar irradiance. The distribution of future irradiance of each station under the given historical radiation observation values could be estimated.

Fig.7 shows the framework of regional photovoltaic prediction method based on spatial correlation.

In [93], the spatiotemporal correlation of photovoltaic data of distributed photovoltaic system was deeply studied. And a bayesian network prediction model based on spatio-temporal correlation was established. Reference [94] conducts spatial clustering of photovoltaic power stations. The power output of each cluster was estimated and predicted by selecting the representative points in the regional centroid. The model output was averaged to obtain the regional prediction results. In [95], sub regions were divided by EMD and hierarchical clustering. The representative power stations were selected based on the minimum redundancy maximum correlation (MRMC) criterion. The Elman NN was used to predict the power generation of representative power stations in each sub region to realize the photovoltaic prediction of the whole region. Reference [96] adopted the generalized parameter covariance function to simulate the correlation between the irradiances of observation points. And it used the Kriging method of space and time to predict solar radiation. In [97], a chain structure of echo state network (ESN) was proposed. Autocorrelation analysis was used in the time information of each spatial variable to provide appropriate input for each ESN module. Reference [98]–[103] analyzed the spatial correlation of the photovoltaic sequence of each power station in the region. It used the similarity of photovoltaic output changes between photovoltaic power stations to obtain the trend of power generation. In [104], a number of solar radiation measurement stations in the northern states of the United States were modeled as spatial undirected graphs. The convolution graph automatic encoder (CGAE) probability model was used to predict the temporal and spatial probability of solar irradiance. The distribution of future irradiance of each station under the given historical radiation observation values could be estimated.

FIGURE 6. Block diagram of photovoltaic forecasting method using meteorological data.

FIGURE 7. Block diagram of regional photovoltaic forecasting method based on spatial correlation.
C. SUMMARY OF SHORT-TERM PHOTOVOLTAIC FORECASTING METHODS

At present, the vast majority of photovoltaic forecasting work is aimed at short-term forecasting. At the same time, the research on short-term regional forecasting is much less than that of short-term single site forecasting. Compared with ultra-short-term forecasting, the short-term forecasting has a larger time scale, and the stochasticity, variability and non-stationarity of photovoltaic have a obvious influence.

When only solar radiation / photovoltaic output power is used as input, the choice of data and model is very important. Due to the complexity of photovoltaic power generation process, it is difficult to fully reproduce. Generally, only the existing data and methods can be adopted to find the approximate real photovoltaic power generation process. In terms of data, the appropriate data type and scale will reduce the variability of the data series and increase the correlation between the input and the predicted value, which is conducive to the detection of photovoltaic characteristics hidden in the data. In the aspect of model, we mainly consider how to find the features in the data and how to find the corresponding short-term prediction method for the characteristics. In addition to trying one by one, principal component analysis, autoregressive integral moving average and other methods can actively detect the hidden features of photovoltaic / solar radiation sequences. Combined with the selected data and models, it is possible to simulate the dynamic change process of photovoltaic / solar radiation and achieve high-quality photovoltaic prediction. The prediction methods based on historical data are shown in Table 3.

If other meteorological information is used, the abundance of data will be greatly improved and the possibility of finding more relevant data related to the change process of photovoltaic will become greater. Numerical weather prediction (NWP) based on meteorological characteristics can help to provide more accurate and relevant meteorological data (such as solar radiation, temperature, humidity, wind speed, etc.). By using meteorological data to analyze the randomness and uncertainty of photovoltaic power generation process, the generalization ability of the prediction model is improved and the power curve of actual photovoltaic is approximately fitted. The forecast methods of combining historical data with other meteorological data are shown in Table 4.

Compared with a single photovoltaic power station, the power station group has a larger area. The overall power generation change process has a spital average effect on the terrain environment, which reduces the variability of solar energy and reduces the error coefficient. Therefore, regional prediction is more reliable than single-site prediction. The regional photovoltaic prediction method based on spatial correlation describes the similarity of photovoltaic sequences with similar geographic locations. Cluster analysis methods such as k-means and hierarchical clustering are often used to divide the entire area into multiple sub-regions. The representative power stations are used to forecast the photovoltaic power in the sub region, so that the short-term single site forecasting method is applied to the power generation prediction of regional power stations. Table 5 compares the short-term regional forecasting methods.

V. MEDIUM AND LONG-TERM PHOTOVOLTAIC FORECAST METHODS

The medium and long-term prediction time scale of photovoltaic is more than 72h. Through the statistical analysis of long-term meteorological data, geographic location information and historical data of solar radiation in a certain region, photovoltaic situation over 72 hours or within one year can be obtained, which can be applied to the site selection of new photovoltaic power stations and the economic value assessment of new power plants [105].

The following is still from the perspective of spatial scale (single site, regional) to analyze the medium and long-term photovoltaic forecasting methods.

A. SINGLE-SITE PHOTOVOLTAIC FORECAST METHODS

Medium and long-term photovoltaic/solar radiation prediction has a significant influence on the location of photovoltaic facilities. The terrain and meteorological conditions of different locations are different and change over time. It is a challenge to accurately estimate the power generation potential of a new location. For the medium and long-term photovoltaic/solar radiation prediction of a single site, existing researches apply more deterministic models or data-driven models as prediction models. This section summarizes and analyzes the medium and long-term power generation/solar radiation prediction methods based on these two types of prediction models.

1) SINGLE-SITE PHOTOVOLTAIC FORECASTING METHODS BASED ON DETERMINISTIC MODEL

The deterministic model can be further divided into an empirical parameterized model and a physical model in accordance with the mathematical form. At present, the empirical parameterized model is a widely studied method for predicting global solar radiation. Most stations usually measure meteorological data such as sunshine duration, temperature, steam, clouds, pressure and wind speed. So empirical models can be used to estimate solar radiation based on these meteorological variables. Empirical model parameters are often specific to specific locations. Local calibration is always required. It is conducive to solar energy assessment and location of photovoltaic facilities that using empirical models to predict mid and long-term solar radiation at designated locations. In the past few decades, a large number of empirical models have been developed. The model adopts commonly used meteorological data and geographic information to predict and estimate global solar radiation. It is widely used because of its low computational cost and reasonable prediction accuracy. Empirical models can be divided into sunlight-based models, temperature-based models, other weather models and hybrid models.
### TABLE 3. Comparison of single-site short-term photovoltaic forecasting methods based on historical data.

| reference | temporal scale | spatial scale | method | input                  | error                  |
|-----------|----------------|---------------|--------|------------------------|------------------------|
| [47]      | 24h            | Single-site   | Weighted GPR | solar irradiance       | nRMSE: 6.614%          |
| [48]      | 24h            | Single-site   | LS-SVM   | PV sequence            | NMAE: 8.17%            |
| [49]      | 12h            | Single-site   | Sub AT + Multi Task Learning | PV sequence | RMSE: 7.139%          |
| [50]      | 24h            | Single-site   | Dynamic Harmonic Regression | solar irradiance | MAE: 4.0098% |
| [51]      | 12h            | Single-site   | Ensemble EMD + SVM | PV sequence | MBE: 3.82% |
| [52]      | 24h            | Single-site   | WT + SARIMA-RVFL | PV sequence | RMSE: 11.61% |
| [53]      | 24h            | Single-site   | WT + SVM | PV sequence            | RMSE: 12.49%          |
| [54]      | 12h            | Single-site   | EMD + ELM | solar irradiance       | RMSE: 8.54%            |
| [55]      | 6h             | Single-site   | WPD + LSTM | PV sequence | MAPE is 69.66% higher than BPNN |
| [56]      | 24h            | Single-site   | WMD + CNN | PV sequence            | RMSE: 4.08%            |
| [59]      | 24h            | Single-site   | K-NN + SVM | solar irradiance       | MAP: 3.12%             |
| [60]      | 12h            | Single-site   | Nonlinear Autoregressive NN | solar irradiance | RMSE: 0.2375kW/m² |
| [61]      | 6h             | Single-site   | AT + GRU | solar irradiance       | RMSE: 1.96%            |
| [62]      | 24h            | Single-site   | K-Means+GRU | PV sequence | RMSE: 4.476%          |
| [63]      | 24h            | Single-site   | Assemble Model | PV sequence | RMSE: 13%             |
| [64]      | 24h            | Single-site   | (GRU+MLP+XGB) | solar irradiance | RMSE: 7.4%             |
| [65]      | 6h             | Single-site   | Mutual Information Optimization of Packaging Machine + ELM | solar irradiance | MAPE: 17.4% |
| [66]      | 24h            | Single-site   | BN + MLP  | solar irradiance       | RMSE: 3.6%             |
| [67]      | 12h            | Single-site   | LS + Dynamic Weighted Combination | solar irradiance | nRMSE: 14.32%          |
| [68]      | 24h            | Single-site   | WNN       | solar irradiance       | nRMSE: 10.84%          |
|           |                |               |          |                        | nRMSE: 9.1%            |
|           |                |               |          |                        | RMSE: 14.07%           |
| reference | temporal scale | spatial scale | method                          | input                                                                 | error               |
|-----------|----------------|---------------|---------------------------------|----------------------------------------------------------------------|---------------------|
| [69]      | 24h            | Single-site   | Recursive Assemble Model (SVM+MLP+MARS) | wind speed, temperature, humidity, global horizontal radiation, horizontal scattering, wind direction. | RMSE:11.97%        |
| [70]      | 6h             | Single-site   | RF                              | solar irradiance, cloud cover, temperature, wind speed and wind direction. | MAE:8.61%          |
| [72]      | 72h            | Single-site   | Simulation Aggregation + ANN     | PV sequence, NWP.                                                    | RMSE:7.39%          |
| [73]      | 24h            | Single-site   | Gradient Boosting DT            | PV sequence, NWP.                                                    | nRMSE:7.19%         |
| [74]      | 24h            | Single-site   | Deep NN                        | solar irradiance, satellite images, NWP.                              | RMSE:25.07%         |
| [76]      | 24h            | Single-site   | Kernel PCA + ELMAN              | PV sequence, solar radiance, weather type, temperature, humidity, wind. | MBE:-1.04W/m²       |
| [79]      | 24h            | Single-site   | ANN                             | solar irradiance, temperature.                                       | RMSE:9.064%         |
| [80]      | 6h             | Single-site   | Graph NN                       | PV sequence, solar irradiance, weather type, temperature, humidity    | MAE:7247%           |
| [81]      | 12h            | Single-site   | LUBE + ELM                      | PV sequence, solar irradiance, temperature, humidity, wind speed, precipitation. | MAE:2.93%          |
| [82]      | 24h            | Single-site   | ELM                             | PV sequence, solar radiance, wind speed, temperature.                | RMSE:17.89%         |
| [83]      | 6h             | Single-site   | ARIMA                           | PV sequence, solar irradiance, air pressure, temperature.             | nRMSE:9.4%          |
| [84]      | 12h            | Single-site   | ELM                             | PV sequence, solar irradiance, temperature, PV module temperature, wind speed. | PICP:95.14%        |
| [85]      | 12h            | Single-site   | LSTM                            | solar irradiance, NWP.                                               | RMSE:3.46-13.46% higher than ANN |
| [86]      | 72h            | Single-site   | RBF                             | PV sequence, cloud type, air pressure, sunshine time and temperature. | RMSE:7.63%          |
| [87]      | 24h            | Single-site   | SVR                             | temperature, humidity, wind speed.                                   | RMSE:6.25%          |
| [88]      | 24h            | Single-site   | K-Means + LSTM                  | global horizontal irradiance, cloud type, precipitation, humidity, solar zenith angle, wind speed, wind direction. | New York:11.1% Hawaii:8.2% |
| [89]      | 24h            | Single-site   | LSTM                            | global horizontal irradiance, temperature, humidity, total cloud cover, wind speed, wind direction. | New York:0.95 Hawaii:0.91 |
| [90]      | 24h            | Single-site   | SVR                             | solar irradiance, mean Earth Solar distance, zenith angle, cloud image. | MAE:8.65W/m²        |
| [91]      | 24h            | Single-site   | Conditional Mutual Information + GPR | solar irradiance, air pressure, humidity, wind speed, wind direction, temperature, rainfall. | RMSE:65.76W/m²      |
| [92]      | 24h            | Single-site   | MLP                             | solar irradiance, time, month, solar zenith angle, wind speed, wind direction, aerosol optical depth. | MAPE was 4.33% lower than that of non characteristic GPR |
| reference | temporal scale | spatial scale | method | input                                                                 | error                  |
|-----------|----------------|---------------|--------|----------------------------------------------------------------------|------------------------|
| [93]      | 24h            | Regional      | Assemble Model (Mars+SVM+MLP) | PV sequence, temperature, wind speed, wind direction, rainfall, rainfall density, air pressure. | RMSE:5.98%             |
| [94]      | 24h            | Regional      | PCA + ANN | solar azimuth, altitude, clear sky irradiance, temperature, relative humidity, wind speed. | MAE:4%-5%              |
| [95]      | 24h            | Regional      | EOF + MRMC + ELMAN | PV sequence. | MAE:4.69%       |
| [96]      | 6h             | Regional      | Kriging  | solar irradiance, cloud movement direction. | RMSE:7.92%             |
| [97]      | 10h            | Regional      | ESN      | solar irradiance. | RMSE:123.1L_j | MAE:89.8L_j |
| [98]      | 72h            | Regional      | PCA + LSTM | solar irradiance, cloud cover, dew point temperature, wind speed, precipitation. | MAE:1.786W/m² | RMSE:4.242W/m² |
| [99]      | 24h            | Regional      | Mutual Information + MRMC + Cuckoo Search Algorithm | PV sequence. | RMSE:8.35%       |
| [101]     | 24h            | Regional      | Cluster Analysis + ANN | solar irradiance, relative humidity, temperature, wind speed, cloudiness index. | MSE:20.95W/m² | RMSE:29.48W/m² |
| [102]     | 6h             | Regional      | Greedy Adjacency Algorithm + Spatio-temporal CNN | PV sequence, site geographic information. | RMSE: California:2.40% New York:2.70% Alabama:3.39% |
| [103]     | 24h            | Regional      | VAR      | PV sequence. | NA                   |
| [104]     | 6h             | Regional      | CGAE     | GHI time series. | CRPS:4.90%          |
a: SUNLIGHT-BASED EMPirical MODELs

The duration of sunshine is the most commonly used parameter for estimating global solar radiation. Since most weather stations in the world can reliably and accurately measure the duration of sunshine, the sunlight-based model becomes the most extensive empirical model for solar radiation estimation. Angstrom and Prescott proposed the first and most widely used empirical model for estimating monthly average daily global solar radiation. It referred to as AP model [106], which is based on radiation from a horizontal surface. The expression is shown in formula (1).

\[
\text{AP model} : \frac{H}{H_0} = a + b \left( \frac{S}{S_0} \right) \tag{1}
\]

where \( H \) is horizontal surface, \( H_0 \) is extraterrestrial solar radiation, \( \frac{H}{H_0} \) is clearness index, \( S \) is sunshine duration, \( S_0 \) is potential sunshine duration, \( \frac{S}{S_0} \) is sunshine duration ratio, \( a \) and \( b \) are empirical coefficient.

Many researchers have conducted verification studies and improvements to the AP model. Besharat [107], Giva [108], Chukwujindu [109] and Yao [110] verified the reliability of the AP model in Iran, Nigeria, Africa and China respectively. Mecibah [111] relates the monthly average daily solar radiation of six cities in Algeria with the monthly average sunshine record and temperature data. The AP model is modified and the performance of six platforms in Algeria is optimized by the quadratic and cubic regression models. Bayrakci [112] summarized the polynomial (linear, quadratic and cubic), logarithmic, exponential, power, rational and combined function regression models based on the AP model. The summarized model was used to provide an estimate of the monthly mean global solar radiation for the city of Mula in southwest Turkey. Studies [113]–[115] have also shown that modified forms of the AP model (such as two-time, three-time, logarithmic, exponential, power, rational and combined function forms) may improve global solar radiation estimates. In addition to the most widely used AP model, empirical models based on sunlight include Rietveld model [116], Ogelman model [117], Bahel model [118], Newland model [119], Jin model [120], etc. Some sunlight-based models are listed as follows.

\[
\text{Rietveld model} : \frac{H}{H_0} = 0.18 + 0.62 \left( \frac{S}{S_0} \right) \tag{2}
\]

\[
\text{Ogelman model} : \frac{H}{H_0} = 0.195 + 0.676 \left( \frac{S}{S_0} \right) - 0.142 \left( \frac{S}{S_0} \right)^2 \tag{3}
\]

\[
\text{Bahel model} : \frac{H}{H_0} = 0.16 + 0.87 \left( \frac{S}{S_0} \right) - 0.61 \left( \frac{S}{S_0} \right)^2 + 0.34 \left( \frac{S}{S_0} \right)^3 \tag{4}
\]

\[
\text{Newland model} : \frac{H}{H_0} = 0.34 + 0.40 \left( \frac{S}{S_0} \right) + 0.17 \log \left( \frac{S}{S_0} \right) \tag{5}
\]

\[
\text{Jin model} : \frac{H}{H_0} = a + b \cos \varphi + cZ + d \left( \frac{S}{S_0} \right) \tag{6}
\]

where \( \varphi \) is latitude of the site, \( a, b, c \) and \( d \) are empirical coefficient.

b: TEMPERATURE-BASED EMPirical MODELs

Temperature is a commonly used measurement parameter. Due to the common availability of daily maximum and minimum temperatures, researchers have proposed empirical methods to estimate solar radiation based on these variables. The temperature-based model assumes that the difference between the highest and lowest temperature is directly related to the proportion of extraterrestrial radiation received by the ground. Hargreaves and Samani proposed an empirical equation that only uses the highest and lowest temperature to estimate solar radiation, and it referred to as the HS model [121]. The specific expression is shown in equation (7), and the coefficient \( a \) for arid and semi-arid regions is set to 0.17. In addition, Bristow and Campbell developed the Bristow-Campbell (BC) model [122]. The empirical formula is expressed in the form of an exponential function. The specific expression is shown in equation (8). The coefficient \( a \) represents the maximum radiation that can be predicted on a sunny day, and the coefficient \( b \) and \( c \) control the rate of approaching \( a \) as the temperature difference increases. Apart from the HS and BC models, temperature-based models include Allen model [123], Hunt model [124], Goodin model [125], Chen model [126], etc. Many researchers analyze and study temperature-based empirical models. Sharifi [127] adopted five temperature-based empirical models with wavelet regression, artificial neural networks and gene expression programming to estimate solar radiation values and analyze performance. Li et al. [128] divided areas according to solar radiation. Then the solar radiation was estimated by the empirical model of selecting the best temperature in each radiation area. Hassan et al. [129] proposed a new global solar radiation estimation model based on ambient temperature, which was tested in ten different locations around Egypt.

\[
\text{HS model} : \frac{H}{H_0} = a(T_{\text{max}} - T_{\text{min}})^{0.5} \tag{7}
\]

\[
\text{BS model} : \frac{H}{H_0} = a[1 - \exp(-b \Delta T^c)] \tag{8}
\]

\[
\text{Allen model} : \frac{H}{H_0} = K_{ra}\left( \frac{P_s}{P_0} \right) (T_{\text{max}} - T_{\text{min}})^{0.5} \tag{9}
\]

\[
\text{Hunt model} : \frac{H}{H_0} = a(T_{\text{max}} - T_{\text{min}})^{0.5}H_0 + b \tag{10}
\]

\[
\text{Goodin model} : \frac{H}{H_0} = a[1 - \exp(-b \Delta T^c)] \tag{11}
\]

\[
\text{Chen model} : \frac{H}{H_0} = a \ln(T_{\text{max}} - T_{\text{min}}) + b \tag{12}
\]

where \( T_{\text{min}} \) is minimum air temperature, \( T_{\text{max}} \) is maximum air temperature, \( \Delta T \) is difference between maximum and minimum air temperatures, \( a \) and \( b \) are empirical coefficient.
c: OTHER WEATHER MODELS AND MIXED EMPIRICAL MODELS

In addition to models based on sunlight and temperature, many researchers have established corresponding empirical models based on meteorological parameters such as relative humidity [130], precipitation [131], clouds [132], [133] and water vapor [134]. At the same time, a mixed model composed of parameters such as sunlight, temperature, relative humidity, precipitation, altitude, etc. also be extensively studied. Fan et al. [135] introduced precipitation and relative humidity data as input of the temperature model for tropical and subtropical regions in China, which can greatly improve the accuracy of a single temperature model. In order to estimate solar radiation in the Mediterranean, Yildirim et al. [136] added relative humidity to a single sunlight model. The results showed that the newly proposed model improves the predictive accuracy of solar radiation. Jahani et al. [137] introduced a complex model that takes temperature, precipitation, relative humidity and sunshine duration as inputs. And he verified the accuracy of the complex model. Fan et al. [138] combined sunshine time, overall radiation, temperature and relative humidity to form a hybrid model, and then they verified the prediction accuracy of the proposed hybrid model in five different climate regions in China. Pandey et al. [139] built a corresponding empirical model by correlating the diffusion ratio with the relative sunshine duration. They estimated the monthly average daily diffuse solar radiation in the four main locations in India.

2) SINGLE-SITE PHOTOVOLTAIC FORECASTING METHODS BASED ON DATA-DRIVEN MODEL

In addition to the determination model used in the previous section for medium and long-term prediction, the prediction method based on data-driven model is also one of the medium and long-term prediction methods. And the prediction method is similar to the prediction method in the previous two chapters. Aslam et al. [140] adopted gated recurrent unit (GRU), LSTM, Feedforward NN, SVM and RF to predict the annual solar radiation. The results show that the prediction effect of GRU is better than other models. To predict the photovoltaic capacity in one month, Jung et al. [141] trained the LSTM from the data sets (power plant capacity, power transaction data, weather conditions and estimated solar radiation) collected from 164 photovoltaic stations. Citakoglu [142] predicted the monthly solar radiation value through ANN, adaptive network fuzzy inference system (ANFIS), multiple linear regression (MLR) and empirical Model. When estimating the solar radiation in Turkey, the performance of ANN model was better than other models. Ozoegwu [143] proposed a hybrid prediction model combining nonlinear autoregression and ANN. And they verified the prediction accuracy of the hybrid method based on various statistical criteria. Lin et al. [144] developed the evolutionary seasonal decomposition LS-SVM to predict the monthly photovoltaic. Seasonal decomposition can be used to reduce the complexity of seasonal time series data. Compared with ARIMA, SARIMA, RNN and LS-SVR models, it shows the effectiveness of long-term prediction.

B. FORECASTING METHODS OF REGIONAL PHOTOVOLTAIC POWER STATIONS

When building a new photovoltaic power plant, it is necessary to evaluate the long-term solar energy resources of the region through long-term prediction. This assessment needs to predict more than a few years in the future, so it needs historical data with roughly the same time range as reference. The available data include the observation data of meteorological observation points (including weather stations, satellites, etc.) that have been built around. Heo et al. [145] proposed a CNN model based on digital elevation map, which is used to predict the annual solar radiation under clear sky conditions. The map data and spatial information set are used as model input. The model can recognize and learn complex terrain features, which provides data support for determining the location of solar panels to be installed. Eom et al. [146] provided a framework for feature selection through correlation analysis and reverse elimination. A feature selection ensemble prediction method based on CNN are adopted, which can use different prediction sources for medium and long-term prediction. Li et al. [147] developed a rolling prediction model based on EMD and ANN to predict the long-term solar radiation in Gonghe County, Qinghai Province. Ghimire et al. [148] constructed the extreme learning machine (ELM) model to predict the monthly average daily solar radiation by combining the medium resolution imaging spectrometer satellite and numerical weather prediction data.

C. SUMMARY OF MEDIUM AND LONG-TERM PHOTOVOLTAIC FORECASTING METHODS

The medium and long-term forecast time scale is above 72 hours. The evaluation of solar energy resources in the designated area will help to the site selection of new photovoltaic facilities. For single-site medium and long-term forecasting, the empirical model has the advantages of simple process and reasonable accuracy for the resource evaluation of the designated location. It only needs to combine meteorological data such as sunshine duration, temperature, relative humidity and precipitation as the input of the empirical model. The results of solar radiation estimation can be obtained. Due to the difficulty of modeling and the relatively lack of data, there are few studies on regional medium and long-term forecasts, but the empirical model has great significance to the rational use of solar energy resources and the safe operation of the power grid. The regional power station group has an average effect on a single photovoltaic. Its variability is smaller than single photovoltaic and its stability is better. After selecting the appropriate scale data, the prediction method for a single site can be directly used on power generation forecast for the regional power station group. The real measurement data can
### TABLE 6. Comparison of medium and long-term photovoltaic forecasting methods.

| reference | temporal scale | spatial scale | method | input | error |
|-----------|----------------|---------------|--------|-------|-------|
| [107]     | 1m             | Single-site   | AP Model | sunshine duration. | R: 0.9969 MABE: 0.4124 M J/m² day MAPE: 0.8485% |
|           |                 |               |         |       | RMSE: 0.5385 M J/m² day MBE: 0.04371 M J/m² day |
| [110]     | 1m             | Single-site   | Bahel Model | sunshine duration. | MAPE: 0.257 M J/m² day MBE: -0.005 M J/m² day |
|           |                 |               |         |       | RMSE: 0.239 M J/m² day |
| [111]     | 1m             | Single-site   | Improved AP Model (cubic regression) | sunshine duration | R2: 0.963 MPE: 0.0244 M J/m² day MAPE: 0.867 M J/m² day |
|           |                 |               | Veeran-Kumar Model | sunshine duration | MBE: 0.0001 M J/m² day RMSE: 0.006 M J/m² day |
| [112]     | 1m             | Single-site   | Sunshine Duration Score Model | sunshine duration | MAPE: 4.6% MABE: 0.19 kW/m² |
|           |                 |               |         |       | RMSE: 0.22 kW/m² |
| [113]     | 1m             | Single-site   | Chen Model | sunshine duration, temperature. | R2 = 0.93 |
| [128]     | 1m             | Single-site   | Temperature Model | temperature. | MPE: 0.18 M J/m² day RMSE: 4.24 M J/m² day |
| [129]     | 1m             | Single-site   | Temperature Model | temperature. | R2: 0.956 MAPE: 4.5302 M J/m² day |
| [135]     | 1m             | Single-site   | Mixed Model | temperature, precipitation, relative humidity. | MBE: 0.9557 M J/m² day |
| [136]     | 1m             | Single-site   | Mixed Model | sunshine duration, relative humidity. | R2: 0.84 RMSE: 0.90 M J/m² day |
| [138]     | 1m             | Single-site   | Mixed Model | sunshine duration, overall radiation, temperature, relative humidity. | RMSE: 0.82 M J/m² day |
| [139]     | 1m             | Single-site   | Mixed Model | sunshine duration, diffusion ratio. | R2: 0.960 MBE: 0.203 M J/m² day |
|           |                 |               |         |       | RMSE: 1.549 M J/m² day |
| [140]     | 1y             | Single-site   | GRU | clear sky solar irradiance, historical hours. | R2: 0.904 RMSE: 0.904 M J/m² day |
|           |                 |               |         | solar irradiance, sunshine duration, month, temperature, relative humidity, wind speed, precipitation, cloud cover. | MABE: 0.664 M J/m² day |
|           |                 |               |         |       | RMSE: Seoul: 4443.65 M J/m² day |
| [141]     | 1y             | Single-site   | LSTM | number of months, extraterrestrial radiation, average temperature, relative humidity. | Busan: 5426.70 M J/m² day |
| [142]     | 1y             | Single-site   | ANN | Non-linear Autoregressive ANN + solar radiation. | R2: 0.723 nRMSE: 7.705% MAPE: 11.535% |
| [143]     | 2y             | Single-site   | LS-SVM | PV sequence. | R2: 0.930 RMSE: 1.650 M J/m² day |
| [144]     | 16m            | Single-site   | Digital Elevation Map + CNN | solar irradiance, spatial information. | R2: 0.96 RMSE: 0.33 M J/m² day |
| [145]     | 1y             | Regional      | Assembly Model | PV equipment installation price, consumption level, gross domestic product, power consumption, temperature. | RMSE: 0.1618 M J/m² day MAPE: 7.8434% |
| [146]     | 1y             | Regional      | EMD + ANN | solar irradiance. | R2: 0.986 MAPE: 0.470% |
| [147]     | 1y             | Regional      | Satellite Image + ELM | solar irradiance, NWP. | MAE: 16.70 MWh/m² MBE: 416.98 MWh/m² |
| [148]     | 40m            | Regional      | Satellite Image + ELM | solar irradiance, NWP. | RMSE: 20.42 MWh/m² |
|           |                 |               |         |       | R: 0.930 RMSE: 0.5849 MWh/m² MAPE: 12.66% |
|           |                 |               |         |       | R2: 0.988 RMSE: 0.503-0.866 M J/m² day |
assist to modify the prediction methods based on the stacking of single photovoltaic sites, model extrapolation and scale upscaling so as to apply the single-site generation prediction method to the prediction of regional power stations. Table 6 compare the medium and long-term photovoltaic forecasting methods.

VI. DISCUSSION OF MULTI-TEMPORAL AND MULTI-SPATIAL SCALES FORECASTING METHODS

In essence, photovoltaic prediction is to find the law of photovoltaic power generation process. It is approximately described by mathematical method to deduce the change process of deterministic photovoltaic output. It is necessary to select appropriate data and methods to analyze the characteristics of the change process of photovoltaic output according to different spatial and time scales. Then select the appropriate mathematical model based on these characteristics.

A. DISCUSSION OF FORECASTING METHODS ON DIFFERENT TEMPORAL SCALES

Data of different time scales have completely different statistical characteristics. The data analysis shows that the variability and stochasticity of photovoltaic usually occur on ultra-short and short-term time scales. The consistent change of photovoltaic occurs on a longer time scale, and the sudden change of photovoltaic occurs on a shorter time scale.

The ultra-short-term prediction time scale is less than 4h. The generation, disappearance and movement of clouds are the fundamental factors that produce the variability of photovoltaic. The cloud image-based method can predict the movement trend of clouds and provide data support for surface photovoltaic. The variability of photovoltaic can be found by methods such as Markov chain, random forest, Bayesian inference and artificial neural network. When the time scale increases, the variability and non-stationarity of PV / solar radiation are weaker than that of ultra-short term. Due to the increase of data complexity, it is necessary to use the prediction model with strong ability to describe the nonlinearity to predict photovoltaic power generation.

Some studies hope to find relevant variables or predictable sequences through principal component analysis, wavelet analysis, empirical mode decomposition and other methods to improve the predictability of input data under the selected model. In the long-term prediction, if the data time scale is large and the volatility of photovoltaic/solar radiation decreases, an empirical model for a specific location can be selected to predict the solar radiation amount of the photovoltaic power station. The calculation process is simple and the prediction accuracy is reasonable.

At present, most researches focus on short-term prediction. The reason for this behavior is that most of the energy is traded in day-ahead markets, when planning and unit commitment take place. With the development of energy market, trading within a day is becoming more important. In this time scale, using feature extraction method and machine learning prediction model (such as ANN, SVM, RF, etc.) with strong nonlinear description ability can obtain excellent prediction accuracy.

B. DISCUSSION OF FORECASTING METHODS IN DIFFERENT SPATIAL SCALES

Different spatial scales have different spatial characteristics including geographical and climatic characteristics. For a single photovoltaic power station, the variation characteristics of power generation and solar radiation of photovoltaic power station are greatly affected by its specific geographical location or geographical environment, which makes the variation rules of solar radiation details of different photovoltaic power stations in the same region are not completely the same or even quite different. How to find the variation law of photovoltaic power generation process is the key to improve the accuracy of single site prediction. Numerical Weather Prediction (NWP) provides more accurate and relevant meteorological data, and it is more likely to find more relevant data related to the change process of single-site photovoltaic. The machine learning model with strong ability to deal with nonlinear problems is selected to predict photovoltaic, and the curve fitting the actual photovoltaic power generation approximately can be obtained.

Stochasticity and variability are the key factors affecting the establishment of photovoltaic power generation prediction model. Regional power stations have spatial average effect, which reduces the variability of solar energy. As a result, the fluctuation of its photovoltaic power generation process is smaller than that of a single photovoltaic station, and its stability is better. After selecting the appropriate scale data, the prediction model for single station can be directly used for the prediction of regional power stations, and the results of regional prediction will be more reliable. Regional prediction can be approached by several ways (superposition method, extrapolation method, statistical scaling method), depending on the information available. In spite of the bottom-up strategy (Superposition methods) outperforming the others, satisfactory results can be obtained with other less computational demanding techniques.

VII. CONCLUSION

According to the temporal scale, spatial scale and input data, this paper classifies the recent research results of photovoltaic prediction and solar radiation prediction. And it summarizes the scientific issues and technical means that the prediction methods focus on under different spatial and temporal scales and input data. For the temporal scale, short-term prediction is a hot topic. With the development of energy market, the status of day-ahead prediction is becoming more important. Machine learning method shows excellent nonlinear description ability in day-ahead prediction, and the prediction results are satisfactory. On the spatial scale, the spatial average effect of the regional reduces the variability of solar energy. The power generation process of the whole region has less fluctuation and better stability. Therefore,
the results of regional prediction are more reliable. The purpose of this research is to provide assistance for absorbing solar resources and provide forecasting methods available for reference when researchers are faced with different photovoltaic forecasting tasks.

REFERENCES

[1] J. D. Sachs, G. Schmidt-Traub, M. Mazzucato, D. Messner, N. Nakicenovic, and J. Rockström, “Six transformations to achieve the sustainable development goals,” Nature Sustainability, vol. 2, no. 9, pp. 805–814, Sep. 2019.

[2] S. Dunnett, A. Sorichetta, G. Taylor, and F. Eigenbrod, “Harmonised global datasets of wind and solar farm locations and power,” Sci. Data, vol. 7, no. 1, Dec. 2020, Art. no. 130.

[3] R. Li, Y. Shi, M. Wu, S. Hong, and P. Wang, “Photovoltaic panel cooling by atmospheric water sorption–evaporation cycle,” Nature Sustainability, vol. 3, no. 8, pp. 636–643, Aug. 2020.

[4] 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 relation modification under partial daily pattern prediction framework,” Energy Convers. Manage., vol. 212, May 2020, Art. no. 112766.

[5] A. Colmenar-Santos, A.-R. Linares-Mena, E.-L. Molina-Ibáñez, E. Rosales-Asensio, and D. Borge-Diez, “Technical challenges for the optimum penetration of grid-connected photovoltaic systems: Spain as a case study,” Renew. Energy, vol. 145, pp. 2296–2305, Jun. 2020.

[6] L. Hernández-Callesio, S. Gallardo-Saavedra, and V. Alonso-Gómez, “A review of photovoltaic systems: Design, operation, and maintenance,” Sol. Energy, vol. 192, pp. 192–203, May 2016.

[7] H. Sharadga, S. Hajimirza, and R. S. Balog, “Time series forecasting of solar power generation for large-scale photovoltaic plants,” Renew. Energy, vol. 150, pp. 797–807, May 2020.

[8] D. Korkmaz, H. Acikgoz, and C. Yildiz, “A novel short-term photovoltaic power forecasting approach based on deep convolutional neural network,” Int. J. Green Energy, vol. 18, no. 3, pp. 525–539, Apr. 2021.

[9] C. Franchucci-Martinez-Anido, B. Botor, A. R. Florita, C. Draxl, S. Lu, H. F. Hamann, and B.-M. Hodge, “The value of day-ahead solar power forecasting improvement,” Sol. Energy, vol. 129, pp. 192–203, May 2016.

[10] S. Theocharides, G. Makrides, A. Livera, M. Theristis, P. Kaimakis, and G. E. Georgiou, “Day-ahead photovoltaic power production forecasting methodology based on machine learning and statistical post-processing,” Appl. Energy, vol. 188, pp. 426–440, Aug. 2019.

[11] K. Wang, X. Qi, and H. Liu, “A comparison of day-ahead photovoltaic power forecasting models based on deep learning neural network,” Appl. Energy, vol. 251, Oct. 2019, Art. no. 113315.

[12] F. Barbieri, S. Rajakaruna, and A. Ghosh, “Very short-term photovoltaic power forecasting with cloud modeling: A review,” Renew. Sustain. Energy Rev., vol. 75, pp. 242–263, Aug. 2017.

[13] M. Ellahi, G. Abbas, I. Khan, P. M. Koola, M. Nasir, A. Raza, and U. Farooq, “Recent approaches of forecasting and optimal economic dispatch to overcome intermittency of wind and photovoltaic (PV) systems: A review,” Energies, vol. 12, no. 22, p. 4392, Nov. 2019.

[14] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F. J. Martinez-de-Pison, and F. Antonanzas-Torres, “Review of photovoltaic power forecasting,” Sol. Energy, vol. 136, pp. 78–111, Oct. 2016.

[15] A. R. Pazikadin, D. Rizaf, K. Ali, M. Z. Malik, A. N. Abdalla, and M. A. Faraj, “Solar irradiance measurement instrumentation and power solar generation forecasting based on artificial neural networks (ANN): A review of five years research trend,” Sci. Total Environ., vol. 715, May 2020, Art. no. 136848.

[16] E. B. Ssekulima, M. B. Anwar, A. Al Hinai, and M. S. El Moursi, “Wind speed and solar irradiance forecasting techniques for enhanced renewable energy integration with the grid: A review,” IET Renew. Power Gener., vol. 10, no. 7, pp. 885–898, Aug. 2016.

[17] S. Sobri, S. Kooi-Kamali, and N. A. Rahim, “Solar photovoltaic generation forecasting methods: A review,” Energy Convers. Manage., vol. 156, pp. 459–497, Jan. 2018.

[18] M. Diagne, M. David, P. Lauret, J. Boland, and N. Schmutz, “Review of solar irradiance forecasting methods and a proposition for small-scale solar grids,” Renew. Sustain. Energy Rev., vol. 27, pp. 65–76, Nov. 2013.
Y. Ju, J. Li, and G. Sun, “Ultra-Short-Term photovoltaic power prediction model—Application to forecasting,” J. Sol. Energy, vol. 188, pp. 760–774, Apr. 2019.

H. Sheng, J. Xiao, Y. Cheng, Q. Ni, and S. Wang, “Short-term solar power forecasting based on weighted Gaussian process regression,” IEEE Trans. Ind. Electron., vol. 65, no. 1, pp. 300–308, Jan. 2018.

F. Almonacid, P. J. Pérez-Higueras, E. F. Fernández, and L. Hontoria, “Short-term photovoltaic power forecasting based on deep convolutional neural network,” IEEE Access, vol. 8, pp. 48421–48428, 2020.

C.-R. Chen and U. Kartini, “K-nearest neighbor neural network machine algorithm,” Energies, vol. 13, no. 1, p. 87, Dec. 2019.

C. Li, Z. J. Liu, Y. Zhao, and Y. Zou, “Short-term photovoltaic power forecasting using a hybrid Wavelet-PSO-SVM model,” J. Cleaner Prod., vol. 195, pp. 1072–1083, Oct. 2018.

Z. Xie, C. Wang, X. H. Xie, Z. J. Liu, and J. F. Bai, “Prediction of solar PV output based on fine-grained characteristics,” J. Power Syst. Technol., vol. 44, pp. 689–696, May 2020.
[83] A. M. Nobre, C. A. Severiano, S. Karthik, M. Kubis, L. Zhao, F. R. Martins, E. B. Pereira, R. Rüther, and T. Reindl, “PV power conversion and short-term forecasting in a tropical, densely-built environment in Singapore,” *Renew. Energy*, vol. 94, pp. 496–509, Aug. 2016.

[84] Q. Ni, S. Zhuang, H. Sheng, S. Wang, and J. Xiao, “An optimized prediction intervals approach for short term PV power forecasting,” *Energies*, vol. 10, no. 10, p. 1669, Oct. 2017.

[85] L. Gao and L. Hong, “Short-term forecasting of power production in a large-scale photovoltaic plant based on LSTM,” *Appl. Sci.*, vol. 9, no. 15, p. 3192, Aug. 2019.

[86] A. Bugada, M. Zaborowicz, P. Boniecki, D. Janczk, K. Koszela, W. Czekaj, and A. Lewicki, “Short-term forecast of generation of electric energy in photovoltaic systems,” *Renew. Sustain. Energy Rev.*, vol. 111, pp. 306–315, Jan. 2018.

[87] K. A. Baharin, H. Abdul Rahaman, M. Y. Hassan, and C. K. Gan, “Short-term forecasting of solar photovoltaic output power for tropical climate using ground-based measurement data,” *J. Renew. Sustain. Energy*, vol. 8, no. 5, Sep. 2016, Art. no. 053701.

[88] Y. Yu, J. Cao, and J. Zhu, “An LSTM short-term solar irradiance forecasting under complicated weather conditions,” *IEEE Access*, vol. 7, pp. 145651–145666, 2019.

[89] M. Hasein and Y.-Y. Chung, “Day-ahead solar irradiance forecasting for microgrids using a long short-term memory recurrent neural network: A deep learning approach,” *Energies*, vol. 12, no. 10, p. 1856, May 2019.

[90] H.-Y. Cheng, C.-C. Yu, and S.-J. Lin, “Bi-model short-term solar irradiance prediction using support vector regressors,” *Energy*, vol. 70, pp. 121–127, Jun. 2014.

[91] N. Huang, R. Li, L. Lin, Z. Yu, and G. Cai, “Low redundancy feature selection of short term solar irradiance prediction using conditional mutual information and Gauss process regression,” *Sustainability*, vol. 10, no. 8, p. 2889, Aug. 2018.

[92] A. Alfadda, S. Rahman, and M. Pipattanasomporn, “Solar irradiance forecast using aerosols measurements: A data driven approach,” *Sol. Energy*, vol. 170, pp. 924–939, Aug. 2018.

[93] R. Zhang, H. Ma, W. Hua, T. K. Saha, and X. Zhou, “Data-driven photovoltaic generation forecasting based on a Bayesian network with spatial–temporal correlation analysis,” *IEEE Trans. Ind. Informat.*, vol. 16, no. 3, pp. 1635–1644, Mar. 2020.

[94] M. Pierro, M. De Felice, E. Maggioni, D. Moser, A. Perotto, F. Spada, and M. Paul, “Spatio-temporal interpolation and forecast criterion,” *Sol. Energy*, vol. 16, no. 3, pp. 163–173, Mar. 2013.

[95] Q. Li, Z. Wu, and H. Zhang, “Spatio-temporal modeling with enhanced flexibility and robustness of solar irradiance prediction: A chain-structure echo state network approach,” *IEEE Access*, vol. 8, pp. 427–436, Jan. 2020.

[96] J. Prescott, “Evaporation from water surface in relation to solar air temperature,” *Agric. Meteorol.*, vol. 11, nos. 11–12, pp. 159–166, Nov. 1976.

[97] M. R. Rietveld, “A new method for estimating the regression coefficients in the models used to estimate solar radiation,” *Energy Convers. Manage.*, vol. 45, nos. 11–12, pp. 1659–1664, Jun. 2004.

[98] A. Alfadda, S. Rahman, M. Pipattanasomporn, “Solar irradiance forecast using aerosols measurements: A data driven approach,” *Sol. Energy*, vol. 170, pp. 924–939, Aug. 2018.

[99] R. Zhang, H. Ma, W. Hua, T. K. Saha, and X. Zhou, “Data-driven photovoltaic generation forecasting based on a Bayesian network with spatial–temporal correlation analysis,” *IEEE Trans. Ind. Informat.*, vol. 16, no. 3, pp. 1635–1644, Mar. 2020.

[100] M. Pierro, M. De Felice, E. Maggioni, D. Moser, A. Perotto, F. Spada, and M. Paul, “Spatio-temporal interpolation and forecast criterion,” *Sol. Energy*, vol. 16, no. 3, pp. 163–173, Mar. 2013.

[101] Q. Li, Z. Wu, and H. Zhang, “Spatio-temporal modeling with enhanced flexibility and robustness of solar irradiance prediction: A chain-structure echo state network approach,” *IEEE Access*, vol. 8, pp. 427–436, Jan. 2020.

[102] J. Prescott, “Evaporation from water surface in relation to solar air temperature,” *Agric. Meteorol.*, vol. 11, nos. 11–12, pp. 159–166, Nov. 1976.

[103] M. R. Rietveld, “A new method for estimating the regression coefficients in the models used to estimate solar radiation,” *Energy Convers. Manage.*, vol. 45, nos. 11–12, pp. 1659–1664, Jun. 2004.
[127] S. S. Sharifi, V. Rezaevidjinad, and V. Nourani, “Estimation of daily global solar radiation using wavelet regression, ANN, GEP and empirical models: A comparative study of selected temperature-based approaches,” J. Atmos. Solar-Terrestrial Phys., vol. 149, pp. 131–145, Nov. 2016.

[128] M.-F. Li, X.-P. Tang, W. Wu, and H.-B. Liu, “General models for estimating daily global solar radiation for different solar radiation zones in mainland China,” Energy Convers. Manage., vol. 70, pp. 139–148, Jun. 2013.

[129] G. E. Hassan, M. E. Youssef, Z. E. Mohamed, M. A. Ali, and A. A. Hanafy, “New temperature-based models for predicting global solar radiation,” Appl. Energy, vol. 179, pp. 437–450, Oct. 2016.

[130] R. K. Swartman and O. Ogunlade, “Solar radiation estimates from common parameters,” Sol. Energy, vol. 11, nos. 3–4, pp. 170–172, Jul. 1967.

[131] R. D. Jong and D. W. Stewart, “Estimating global solar radiation from common meteorological observations in Western Canada,” Can. J. Plant Sci., vol. 73, no. 2, pp. 509–518, Apr. 1993.

[132] G. W. Patridge and D. Proctor, “Monthly mean solar radiation statistics for Australia,” Sol. Energy, vol. 18, no. 3, pp. 235–243, 1976.

[133] M. S. Ahamed, H. Guo, and K. Tanino, “Cloud cover-based models for estimation of global solar radiation: A review and case study,” Int. J. Green Energy, vol. 19, no. 2, pp. 175–189, Jan. 2022.

[134] A. H. Maghrabi, “Parameterization of a simple model to estimate monthly global solar radiation based on meteorological variables, and evaluation of existing solar radiation models for Tabouk, Saudi Arabia,” Energy Convers. Manage., vol. 50, no. 11, pp. 2754–2760, Nov. 2009.

[135] J. Fan, B. Chen, L. Wu, F. Zhang, X. Lu, and Y. Xiang, “Evaluation and development of temperature-based empirical models for estimating daily global solar radiation in humid regions,” Energy, vol. 144, pp. 903–914, Feb. 2018.

[136] A. Teke and F. Antonanzas-Torres, “Evaluation of classical parametric models for estimating solar radiation in the eastern Mediterranean region of Turkey,” Renew. Sustain. Energy Rev., vol. 82, pp. 2053–2065, Feb. 2018.

[137] B. Jahani, Y. Dinipashoh, and A. R. Nafchi, “Evaluation and development of empirical models for estimating daily solar radiation,” Renew. Sustain. Energy Rev., vol. 73, pp. 878–891, Jun. 2017.

[138] J. Fan, L. Wu, F. Zhang, H. Cai, X. Ma, and H. Bai, “Evaluation and development of empirical models for estimating daily and monthly mean daily diffuse horizontal solar radiation for different climatic regions of China,” Renew. Sustain. Energy Rev., vol. 105, pp. 168–186, May 2019.

[139] C. K. Pandey and A. K. Katiyar, “A comparative study to estimate daily diffuse solar radiation over India,” Energy, vol. 34, no. 11, pp. 1792–1796, Nov. 2009.

[140] M. Aslam, J.-M. Lee, H.-S. Kim, S.-J. Lee, and S. Hong, “Deep learning models for long-term solar radiation forecasting considering microgrid installation: A comparative study,” Energies, vol. 13, no. 1, p. 147, Dec. 2019.

[141] Y. Jung, J. Jung, B. Kim, and S. Han, “Long short-term memory recurrent neural network for modeling temporal patterns in long-term power forecasting for solar PV facilities: Case study of South Korea,” J. Cleaner Prod., vol. 250, Mar. 2020, Art. no. 119476.

[142] H. Citakoglu, “Comparison of artificial intelligence techniques via empirical equations for prediction of solar radiation,” Comput. Electron. Agric., vol. 118, pp. 28–37, Oct. 2015.

[143] C. G. Ozogwu, “Artificial neural network forecast of monthly mean daily global solar radiation of selected locations based on time series and month number,” J. Cleaner Prod., vol. 216, pp. 1–13, Apr. 2019.

[144] K.-P. Lin and P.-F. Pai, “Solar power output forecasting using evolutionary seasonal decomposition least-square support vector regression,” J. Cleaner Prod., vol. 134, pp. 456–462, Oct. 2016.

[145] J. Heo, J. Jung, B. Kim, and S. Han, “Digital elevation model-based convolutional neural network modeling for searching of high solar energy regions,” Appl. Energy, vol. 262, Mar. 2020, Art. no. 114588.

[146] H. Kom, Y. Son, and S. Choi, “Feature-selective ensemble learning-based long-term regional PV generation forecasting,” IEEE Access, vol. 8, pp. 54620–54630, 2020.

[147] F.-F. Li, S.-Y. Wang, and J.-H. Wei, “Long term rolling prediction model for solar radiation combining empirical mode decomposition (EMD) and artificial neural network (ANN) techniques,” J. Renew. Sustain. Energy, vol. 10, no. 1, Jan. 2018, Art. no. 013704.

[148] S. Ghimire, R. C. Deo, N. J. Downs, and N. Raj, “Self-adaptive differential evolutionary extreme learning machines for long-term solar radiation prediction with remotely-sensed MODIS satellite and reanalysis atmospheric products in solar-rich cities,” Remote Sens. Environ., vol. 212, pp. 176–198, Jun. 2018.

CAICHENG LIU received the B.S. degree from the Haibin College, Beijing Jiaotong University, Heibei, China, in 2019. He is currently pursuing the M.S. degree in control engineering with Nanchang University, Jiangxi, China. His research interests include photovoltaic forecasting and acceleration of neural networks.

MING LI received the Ph.D. degree from Nanchang University, in 2005. He is currently a Professor with the School of Information Engineering, Nanchang University. His research interests include measurement technology and intelligent instrument.

YUNJUN YU (Member, IEEE) was born in Shangrao, China. He received the B.Sc. degree in automation and the M.Sc. degree in control theory and control engineering from Nanchang University, China, in 2000 and 2007, respectively, and the Ph.D. degree from the Chinese Academy of Sciences, in 2013. He is currently an Associate Professor with the Department of Automation, Information Engineering School, Nanchang University. His research interests include fault diagnosis, active disturbance rejection control (ADRC), data-driven optimal control and its applications in microgrids, and low-carbon electricity technology.

ZIYANG WU, photograph and biography not available at the time of publication.

HAI GONG, photograph and biography not available at the time of publication.

FEIER CHENG, photograph and biography not available at the time of publication.