

Research Article

Short-Term Master-Slave Forecast Method for Distributed Photovoltaic Plants Based on the Spatial Correlation

Jia Ning ^{1,2}, Guanghao Lu ^{1,2}, Sipeng Hao ^{1,2}, Aidong Zeng ^{1,2} and Hualei Wang ³

¹School of Electric Power Engineering, Nanjing Institute of Technology, Nanjing 211167, China

²Jiangsu Collaborative Innovation Center for Smart Distribution Network, Nanjing 211100, China

³State Grid Lianyungang Power Supply Company, Lianyungang 222000, China

Correspondence should be addressed to Jia Ning; ningjia@njit.edu.cn

Received 19 March 2021; Revised 17 April 2021; Accepted 10 May 2021; Published 20 May 2021

Academic Editor: Jun Shen

Copyright © 2021 Jia Ning et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the large-scale integration of distributed photovoltaic (DPV) power plants, the uncertainty of photovoltaic generation is intensively influencing the secure operation of power systems. Improving the forecast capability of DPV plants has become an urgent problem to solve. However, most of the DPV plants are not able to make generation forecast on their own due to the constraints of the investment cost, data storage condition, and the influence of microscope environment. Therefore, this paper proposes a master-slave forecast method to predict the power of target plants without forecast ability based on the power of DPV plants with comprehensive forecast system and the spatial correlation between these two kinds of plants. First, a characteristics pattern library of DPV plants is established with K-means clustering algorithm considering the time difference. Next, the pattern most spatially correlated to the target plant is determined through online matching. The corresponding spatial correlation mapping relationship is obtained by numerical fitting using least squares support vector machine (LS-SVM), and the short-term generation forecast for target plants is achieved with the forecast of reference plants and mapping relationship. Simulation results demonstrate that the proposed method could improve the overall forecast accuracy by more than 52% for univariate prediction and by more than 22% for multivariate prediction and obtain short-term generation forecast for DPV or newly built DPV plants with low investment.

1. Introduction

Electricity power consumption increases drastically in recent years, and with the decreasing supply of fossil fuels, the renewable generation, especially photovoltaic (PV) generation, has developed rapidly as well [1]. In the background of green energy strategy, the global PV installed capacity has reached 300 GW. However, the large utility-scale generation is typically deployed in rural areas, which are far from the load centers; thus, the generated power is not efficiently used. Integration of distributed PV generation with the distribution network could contribute to solving the unmatched location of generation and consumption [2]. However, distribution network is the terminal end of power system, with weak infrastructure and low reserve capacity. The increasing amount of highly intermittent and variant

DPV generation will greatly affect the stability of power systems [3, 4]. Therefore, the accurate generation forecast of DPV is significant for the scheduling and stable operation of power systems. Generation forecasts could be categorized into short-term forecasts (0–72 hours ahead of the next day) and ultra-short-term forecasts (15 minutes–4 hours ahead) [5, 6]. Short-term generation forecast provides supportive data for decision-making of power system scheduling and helps improve operational reliability.

1.1. Literature Review. There is a considerable amount of scientific literature on renewable energy forecasting, and current research [7–10] on generation forecast of intermittent renewable energy has made great achievements, but the forecast methods are mostly focused on large capacity

wind and solar power plants, in which a single generating unit has an installed capacity at MW level. The renewable energy forecast methods could be classified into two major categories: time series forecast method and spatial distribution forecast method.

Time series forecast methods analyze the trends of the past to predict future events, with the assumption that future trends will hold similar to historical trends. Two numerical weather prediction models are utilized to forecast the weather variables used by the third module to predict the hourly energy production in the PV plant in [11]. Weather status pattern recognition model for short-term PV forecasting is presented using a solar irradiance feature extraction and support vector machine [12]. These references perform time series forecasting to predict weather and then obtain the short-term PV forecast power. Some other references on time series forecasting focus on different algorithms, e.g., traditional physical model prediction [13], BP-artificial neural network (ANN) prediction with accurate numeric weather forecast [14], extreme learning machine (ELM) [15], and support vector machine (SVM) [16, 17]. In [18], a new model combines two well-known methods: the seasonal auto-regressive integrated moving average method and support vector machines method are proposed for short-term power forecasting of a grid-connected photovoltaic plant. A short-term forecasting method is presented for large-scale grid-connected PV plants using ANN in [19]. A genetic algorithm-based SVM model for short-term power forecasting of residential scale PV system is proposed in [20]. Reference [21] provides a review about the methods used to predict PV power, with the main focus being on the metaheuristic and machine learning methods. In general, these time series methods rely on a large amount of historical generation data and numerical weather forecasts and could obtain high forecast accuracy. However, the spatial characteristics of distributed PV systems are not considered. This paper focuses on distributed PV generations, which have the problem of deficient historical data, and considers their spatial distribution characteristics to realize their short-term power prediction.

Spatial distribution forecast method considers the geographic information and the spatial distribution characteristics of PV systems. The effect of spatial and spectral nonuniform irradiance distribution on multijunction solar cell performance is analyzed using an integrated approach [22], and the spatial dependence of variations for small residential PV system power output is investigated, indicating that the fluctuations are correlated up to a certain decorrelation length [23]. In [24], Karakaya applies the finite element method to forecast the diffusion of solar PV systems in time and space, in which the time-varying parameters are arduous to determine. Spatial clustering of PV systems and quantitative analysis of PV adoption drivers in the time dimension are investigated to propose a data-driven forecasting approach of PV diffusion in [25]. These references are studied to verify the spatial distribution characteristics or forecast the diffusion of PV systems. Our research is to utilize the spatial distribution characteristics for DPV power prediction.

These methods are not suitable to be applied to DPV prediction due to the data constraints and distributed characteristics of DPV [26]. In terms of data constraints, in actual DPV projects, most of the DPVs are not equipped with their own forecasting module and are not capable of storing a large amount of historical data or obtaining weather forecast data because of the limited investment.

1.2. Explanation of Spatial Correlation. In terms of distributed characteristics, the affecting factors of generation include not only natural factors such as radiation and temperature, but also the installed tilt angle, construction layout, vegetation, and microscope weather, which could vary widely even in a small range [27].

Figure 1 illustrates the spatiotemporal distribution characteristics of DPV. The DPVs are distributed in 6 areas across 3 time zones. The microscope environments in each area are different from each other, and the generation of DPV may be more closely related to its surrounding environment than the area it is in. For example, the generation pattern of the DPV in area A may be similar to that in area F, even if it is located far away and in a different time zone, because the microscope environments (shadow of obstacles, moisture, and building height) are similar. The installation details also vary, such as the tilt angle and direction. This similarity, regardless of time-space continuity, is revealed in data correlation, instead of physical connections [25]. We define this correlation as a spatial correlation as follows:

Spatial correlation refers to the numerical correlation of DPV generation at different locations. When analyzing the spatial correlation, eliminate the time difference of generation curve with data processing.

1.3. Contribution. The current technique bottleneck of DPV generation forecast is caused by data deficiency and complex influencing factors, making the traditional method of mathematically modelling infeasible in DPV forecast. A new method considering the data deficiency and spatiotemporal distribution characteristics is required to meet the need of DPV forecast. In the current installation, there are a few DPV plants with functional forecast system, which are used as reference plants in the following paper. Meanwhile, most of the DPV plants are not able to make generation forecasts on their own due to the economic and technological constraints. These plants are later referred to as target plants. According to this reality, this paper takes advantage of big data methodology and proposes a master-slave forecast technique based on spatial correlation between reference plants and target plants considering multiple affecting factors including radiation, temperature, time zone, etc., which were not studied before. The technique utilizes a master-slave forecast framework, matching the generation characteristics of target plants to reference plants using data correlation, forecasting the generation of slave target plants with the forecast data of spatially correlated master reference plants, and realizing DPV generation forecasting with data correlation relationship.

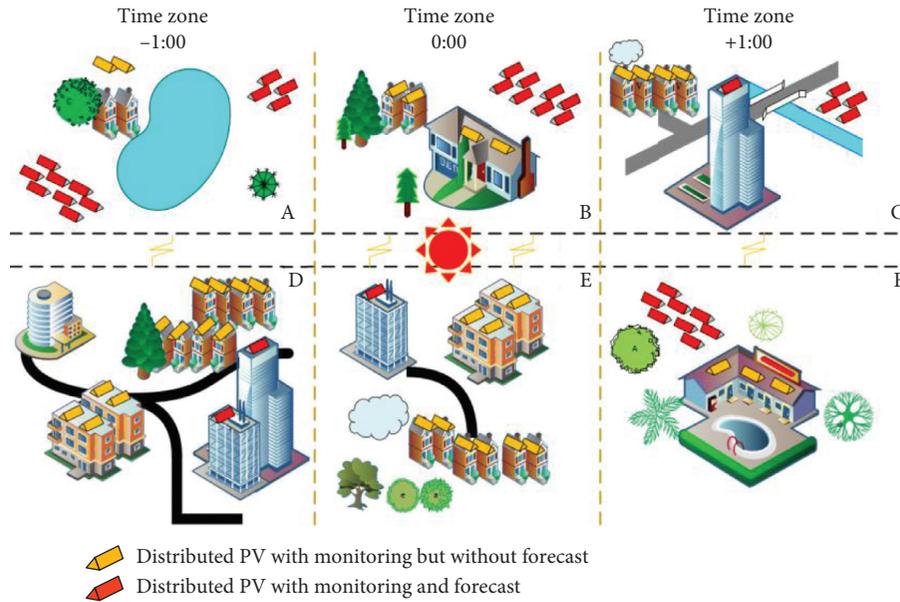


FIGURE 1: Spatiotemporal distribution of DPV plants considering the microscope environment.

Based on the bottleneck analysis of DPV generation forecast and the characteristics of DPV, the main contributions of this paper are listed as follows:

- (1) A spatial correlation matching method is proposed to obtain the data correlation relationship across time and space between target plants and reference plants, in which the K-means clustering algorithm is utilized to cluster reference plants into groups with individual patterns on the basis of their generation characteristics. The clustering method could reduce the computation time of online matching and improve the matching accuracy.
- (2) A master-slave forecast method is presented to make the generation forecast for a large number of target plants in short-term time scale, in which the LS-SVM algorithm is utilized to obtain the spatial correlation mapping relationship. Therefore, the power of target plants as slave could be predicted based on the power of reference plants as master and the spatial correlation between these two kinds of plants.

1.4. Article Organization. The following paper is composed as follows. Section 2 gives the introduction of the master-slave forecast framework. Section 3 describes the matching method for spatial correlation relationship and studies the time difference characteristics of DPV generation curves. Section 4 conducts a case study, validating the advantage of the proposed technique. Finally, Section 5 concludes the paper.

2. Framework of Short-Term Master-Slave Forecast Technique Based on Spatial Correlation

In this section, the framework of the proposed master-slave forecast technique is illustrated and explained. The forecast method is based on the spatial correlation between the generation characteristics of different DPV plants.

Data mining shows that the generation trajectories of different DPV plants in the same time dimension have a certain numerical correlation; that is, two or more numerical trajectories approximately fit in some correlation relationship. For example, Figure 2 shows the generation curves of some randomly chosen DPV plants in 3 different areas and the comparison of selected curves from all areas. It is seen that the generation curves in the same area have different shapes, while a curve might share more similarity with curves from other areas than the curves within the same area, although the DPV plants are geologically closer in one area. Therefore, the spatial correlation is defined as a numerical correlation between the generation data of different DPV plants, and the geological relationship is ignored.

The master-slave spatial correlation based forecast technique is to utilize short-term forecast of reference DPV plants (master plant) and spatial correlation relationship to forecast short-term generation of target DPV plants (slave plant) indirectly. The master-slave DPV generation forecast framework based on spatial correlation is shown in Figure 3.

As shown in Figure 3, the framework of master-slave prediction method consists of three parts, namely, left part, middle part, and right part. The left part is the forecast results

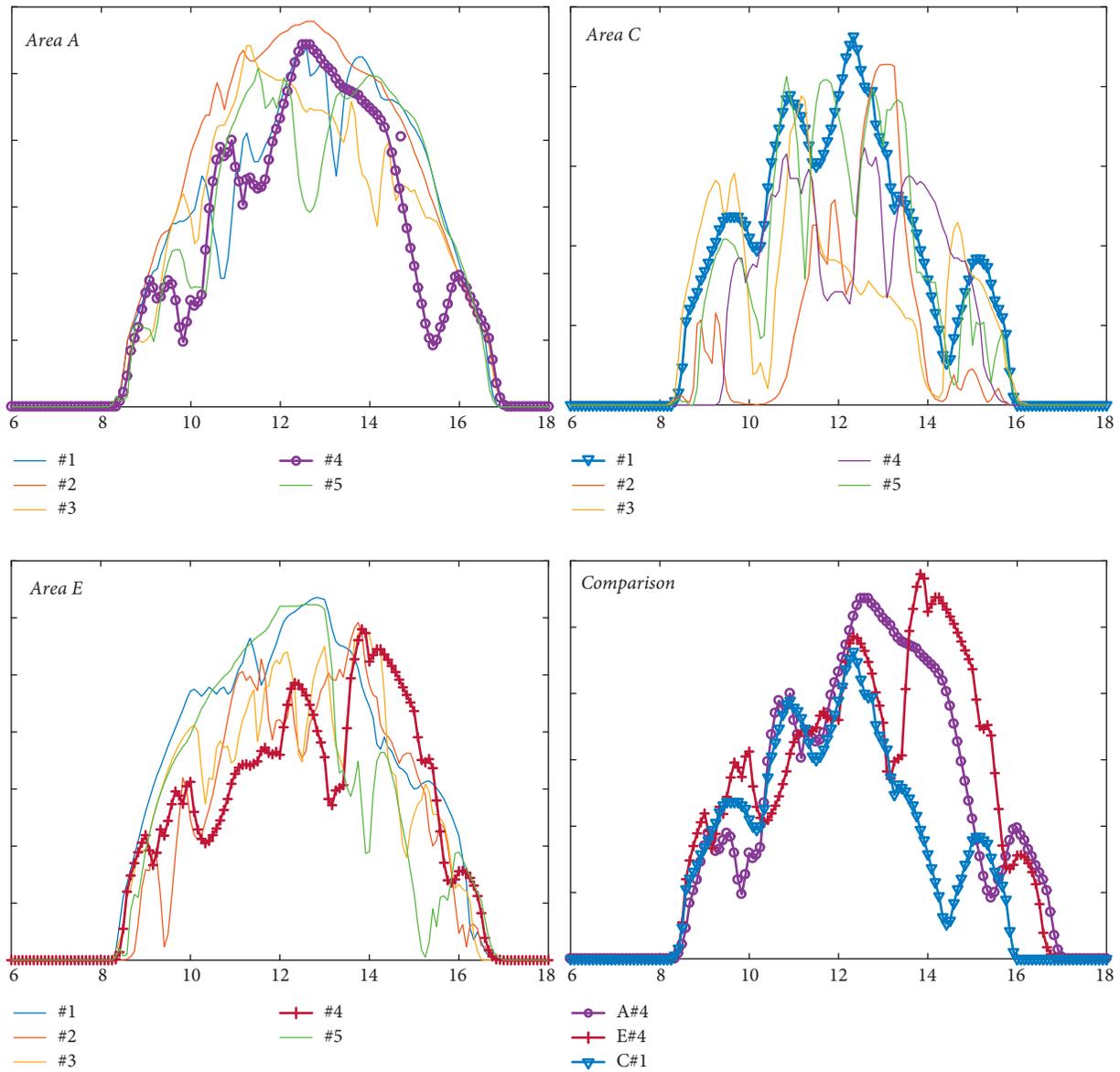


FIGURE 2: The principle of spatial correlation prediction.

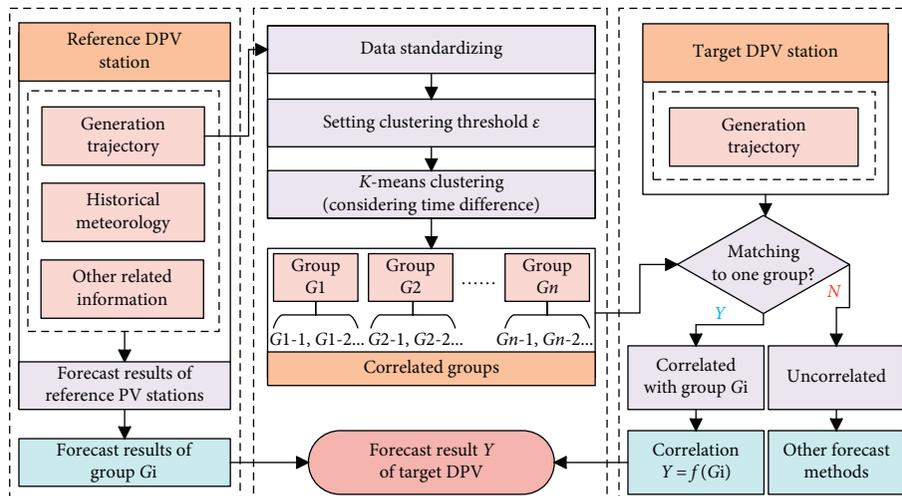


FIGURE 3: Framework of master-slave prediction based on spatial correlation.

of reference DPV stations. Based on the generation trajectory, historical meteorology, and other related information, the power of reference DPV plants is predicted to benefit the prediction of target DPV stations. The middle part is offline clustering of reference plants on the basis of their generation trajectory in history. Because there are a large number of reference plants, and many reference plants are spatial-correlated, a pattern library is established using offline K-means clustering to reduce the searching time for online matching [28]. The right part is an online matching process of spatial correlation between target plants and patterns in library. If the matching is successful, the forecast results of target plants are obtained based on the forecast results of the correlated pattern and correlation relationship. If the matching fails, other forecast methods should be adopted.

3. Spatial Correlation Matching with K-Means Clustering

To utilize the spatial correlation between reference plants and target plants, the pattern matching method to find the correlated reference master plants with target slave plants is given in this section. K-means clustering algorithm is used to cluster master plants into groups with individual patterns according to their generation characteristics, thus constructing the standard pattern library. Next, the clustering significance index (CSI) is defined to set the cluster number, and standardized Euclidean distance (SED) is used to match the standardized data of DPV generation to the data patterns in the pattern library to establish the spatial correlation mapping. The spatial correlation matching process using K-means clustering is shown in Figure 4. The detailed corresponding algorithms are described in Sections 3.1–3.4.

3.1. Data Standardization. Data standardization is the process of data scaling and nondimensionalization so that the data could be compared. In this paper, the raw data of DPV generation A are processed row by row using normalization, with the following equation:

$$B_{ij} = \frac{(A_{ij} - \text{mean}(A_j))}{\text{std}(A_j)}, \quad (1)$$

where A_j is the j th row of the matrix A , A_{ij} is the i th element in A_j , $\text{mean}(A_j)$ denotes the mean value of vector A_j , $\text{std}(A_j)$ denotes the standard deviation of vector A_j , and B_{ij} is the i th element of the j th row of matrix B . After standardization, A_j is converted to B_j . The mean of vector B_j is 0, and the variance of vector B_j is 1. Vector B_j is called a standard vector, and matrix B is the standardized matrix of A .

The data standardization could reduce the influence of DPV installed capacity difference on spatial correlation and preserve the characteristics of the trend of historical DPV generation data.

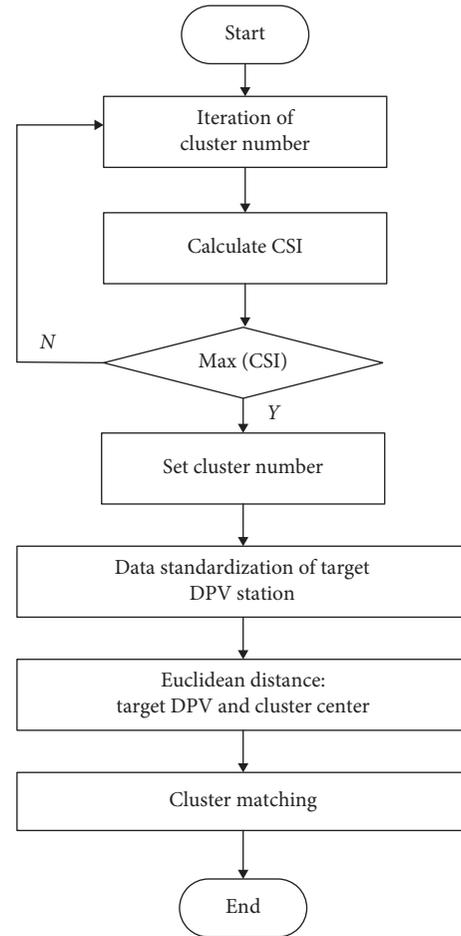


FIGURE 4: Flow chart of spatial correlation matching.

3.2. K-Means Clustering of Reference Plants. PV generation shows the characteristics of uncertainty and fluctuation, and the generation curve of a random day could not represent the general generation pattern of the plant. Therefore, the average of several generation days' data is used to establish the pattern library of reference plants.

In this paper, K-means algorithm is adopted for clustering, and several groups are formed of plants with similar generation pattern inside each group. The cluster number needs to be inputted when using K-means algorithm. The cluster significance index (CSI) is given in equation (2) to determine the group number; that is,

$$\text{CSI} = \frac{\sum_{j=1}^N \sum_{i=1}^{n_j} |X_{cj} - X_{ci}|}{\sum_{j=1}^N \sum_{i=1}^{n_j} |X_{ji} - X_{cj}|}, \quad (2)$$

where N is the number of clustering groups, n_j is the number of plants in the j th group, X_{cj} is the eigenvector of the j th group, and X_{ji} is the vector of the i th plant in the j th group.

The number of groups is determined through iteration. Different values of N are selected, and CSI is calculated for each N . The value N with the largest CSI is chosen as the input of the group number for K-means algorithm, and the PV generation patterns are obtained subsequently.

3.3. Online Matching of Spatial Correlation. The online pattern matching process is described as follows. Extract n monitoring points from recent historical data backward from the forecast point of the target DPV plant as a prediction window vector. Standardize the prediction window vector and add it to the pattern library as the $(T+1)$ cluster and perform a clustering process. If the current window vector could be put into the same cluster with the i th vector pattern, the target DPV plant is determined to have spatial correlation with the i th type of reference DPV plants.

Standard Euclidean Distance (SED) is more commonly adopted in actual application as the criterion of correlation. Therefore, in this paper, SED is used to quantify the correlation, and the optimal delay value Δt is determined by searching for the minimum SED. (3) gives the equation to calculate SED:

$$\rho(A, B) = \sqrt{\sum (a[i] - b[i])^2}, \quad (i = 1, 2, 3, \dots, n), \quad (3)$$

where $a[i]$ and $b[i]$ are the i th element of vector A and vector B , respectively.

In theory, the probability of successful matching of spatial correlation is higher if the reference PV power plants are distributed more evenly and with larger number. For the target DPV plants that fail to match reference DPV plants, temporal correlation based forecast or other forecast methods are recommended.

3.4. Numerical Fitting Using LS-SVM. After spatial correlation matching, a single one or multiple reference DPV plants are chosen from the spatially correlated reference plant groups. The spatial correlation model is obtained by numerical fitting of the prediction window historical data of reference plants and target plant. Next, the short-term generation could be calculated with short-term forecast of reference DPV plants and the spatial correlation relationship.

least squares support vector machine (LS-SVM) regression is applied to perform numerical fitting, which could achieve better results of multivariate regression. The equation is shown as

$$P_{f2} = f(X) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(X_i, X_j) + b, \quad (4)$$

where α_i and b are the coefficients to be determined, and $K(X_i, X_j)$ is the kernel function. Radial basis function (RBF) is often used as the kernel function to solve a regression problem, which is given in

$$K(X_i, X_j) = \exp\left(\frac{-1}{2\sigma^2} \|X_i - X_j\|^2\right), \quad (5)$$

where σ is called the extension constant of RBF, which reflects the width of the function image. The smaller the width σ is, the more selective the function is.

3.5. Forecast Performance Evaluation. Although the prediction graph could show the results of all forecasting

methods intuitively, it is arduous to quantitatively judge the pros and cons of each prediction method objectively. Therefore, this paper applies the root mean square error (RMSE) and mean absolute error (MAE) to compensate the shortcomings of the prediction graph. The two error formulas are as shown in equations (6) and (7):

$$\text{RMSE} = \sqrt{\frac{1}{m} \sum_{i=1}^m (P_p - P_r)^2}, \quad (6)$$

$$\text{MAE} = \frac{1}{m} \sum_{i=1}^m |P_p - P_r|, \quad (7)$$

where P_p is the prediction value of PV power, P_r is the actual power, and m is the total number of prediction points.

3.6. Influence of Time Difference Characteristics on Spatial Correlation Matching. Considering the widespread distribution characteristics of DPV, the correlation relationship between the reference plant X and target plant Y may show some time and space difference characteristics; that is, $Y(t)$ is more correlated to $X(t + \Delta t)$. This characteristic is referred to as time difference characteristics in the following paper. As shown in Figure 5, curves A and B have similar changing trend, but the starting and ending points are different. By moving the curve B to the right with a period of Δt , the distance between the curves is reduced, and the similarity of the trend is highlighted.

In references [10, 11], the Pearson product-moment correlation coefficient (PPMCC) is used to describe the correlation between vectors. The optimal value of Δt is determined by finding the value of PPMCC. Equation (7) gives the equation to calculate PPMCC:

$$r_p(X, Y) = \frac{\sum_{i=1}^n (x_i - x_{av})(y_i - y_{av})}{\sqrt{\sum_{i=1}^n (x_i - x_{av})^2 \sum_{i=1}^n (y_i - y_{av})^2}}, \quad (8)$$

where x_{av} and y_{av} represent the arithmetic mean of vector X and vector Y , respectively. PPMCC value close to 1 denotes strong correlation, while a value close to 0 denotes weak correlation.

Considering the fact that the DPV plants may be distributed in different time zones, a time shift method to improve pattern matching effects is given as follows. Set a unified reference time as 0 points, search from 0 points backward and forward with the time of Δt , and obtain the monitoring points within the range of $[-p, p]$.

For every iteration of spatial correlation matching, move the target plant vector backward or forward for one monitoring point and keep the other vectors in the pattern matrix unchanged. Calculate the correlation of the target plant and all other patterns and find the pattern with minimum SED, and the most spatially correlated DPV plant is found globally with consideration of time difference characteristics.

In summary, the advantages of considering Δt include the following:

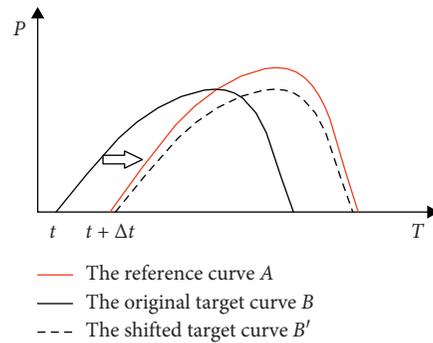


FIGURE 5: Time difference characteristics and time shift method for matching.

- (1) Increasing the probability of successfully matching a target plant to the reference DPV power plants.
- (2) Searching for the most correlated reference PV power plant globally; i.e., the global minimum is achieved rather than the local minimum, which could improve the forecast accuracy.

Therefore, the reference PV power plants matched with target PV plants in this paper are the most correlated plants globally considering the time difference characteristics.

4. Case Studies

The case used in this paper to demonstrate the forecast method is the actual historical generation data of 5166 DPV in the USA [29]. The DPV plants are located from 73° to 125°W, 25°–49°N, as shown in Figure 6. The monitored time is from 0:00 on 1 January 2006 to 23:45 on 31 December 2006, with a sampling interval of 15 minutes. The total number of sampling points is 30540. 1000 of the DPV plants (19.3%) are chosen randomly as reference DPV plants, and the remaining 4166 (81.7%) are regarded as target plants. The forecast target is the generation curve of day 307 in the year. The prediction window is set to be 3 days before the target forecast day, and the number of monitoring points is 288.

The preparation for online forecast is offline clustering. The 3616 reference plants are clustered into 50 spatial correlated groups; i.e., 50 patterns are generated. The calculated largest CSI is obtained to be 1.15485 when the N equals 50 based on equation (2). The clustering results are shown in Section 4.1.

Next, the online forecast process of master-slave short-term DPV generation forecast method is presented. In Section 4.2 and Section 4.3, two target plants T1 (3903#) and T2 (1346#) are chosen to show the forecast process. In Section 4.4, the forecasts of 1550 target plants are obtained, and the statistic error is compared. Section 4.5 discusses the situation in which multiple reference plants are used to make forecasts. Section 4.6 discusses the choice of prediction window size and its influence on forecast accuracy.

4.1. Clustering of Reference DPV Plants. 1000 DPV plants with forecast ability are chosen as reference plants to generate a pattern library. Using the clustering method in

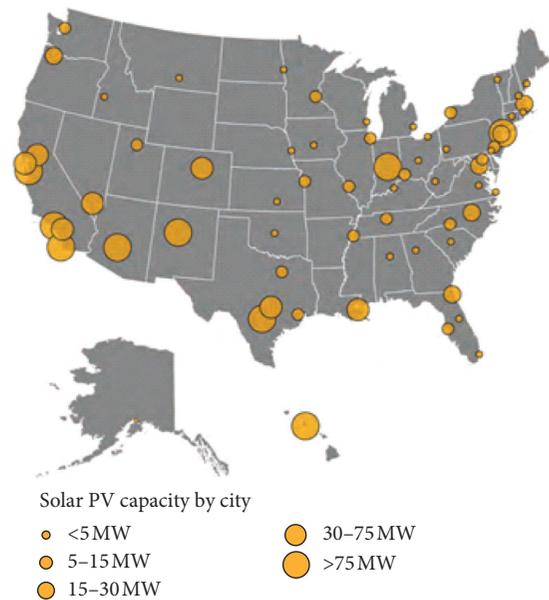


FIGURE 6: PV distribution in the USA [30].

Section 3, a prediction window composed of the average of 10 days’ generation data before the forecast target day is chosen as the pattern mining and clustering data, and the group number is set to be 50.

Figure 7 shows the curves of 4 typical generation patterns in the pattern library. The plants in the same pattern group share similar generation characteristics, and the generation patterns between groups are extremely different. Therefore, the K-means clustering method could put the reference plants with similar generation patterns into the same groups and form a pattern library.

The choice of clustering group number should not only consider the CSI, which affects the clustering performance, but also the time consumption for matching target plants to reference plants. In short-term generation forecast, the forecast interval is 15 minutes. If the number of groups is too large, the matching process will be extremely time-consuming, and the forecast timeliness could not be guaranteed.

4.2. Searching for Most Correlated Plants considering Time Difference. This example shows the effect of the time shift

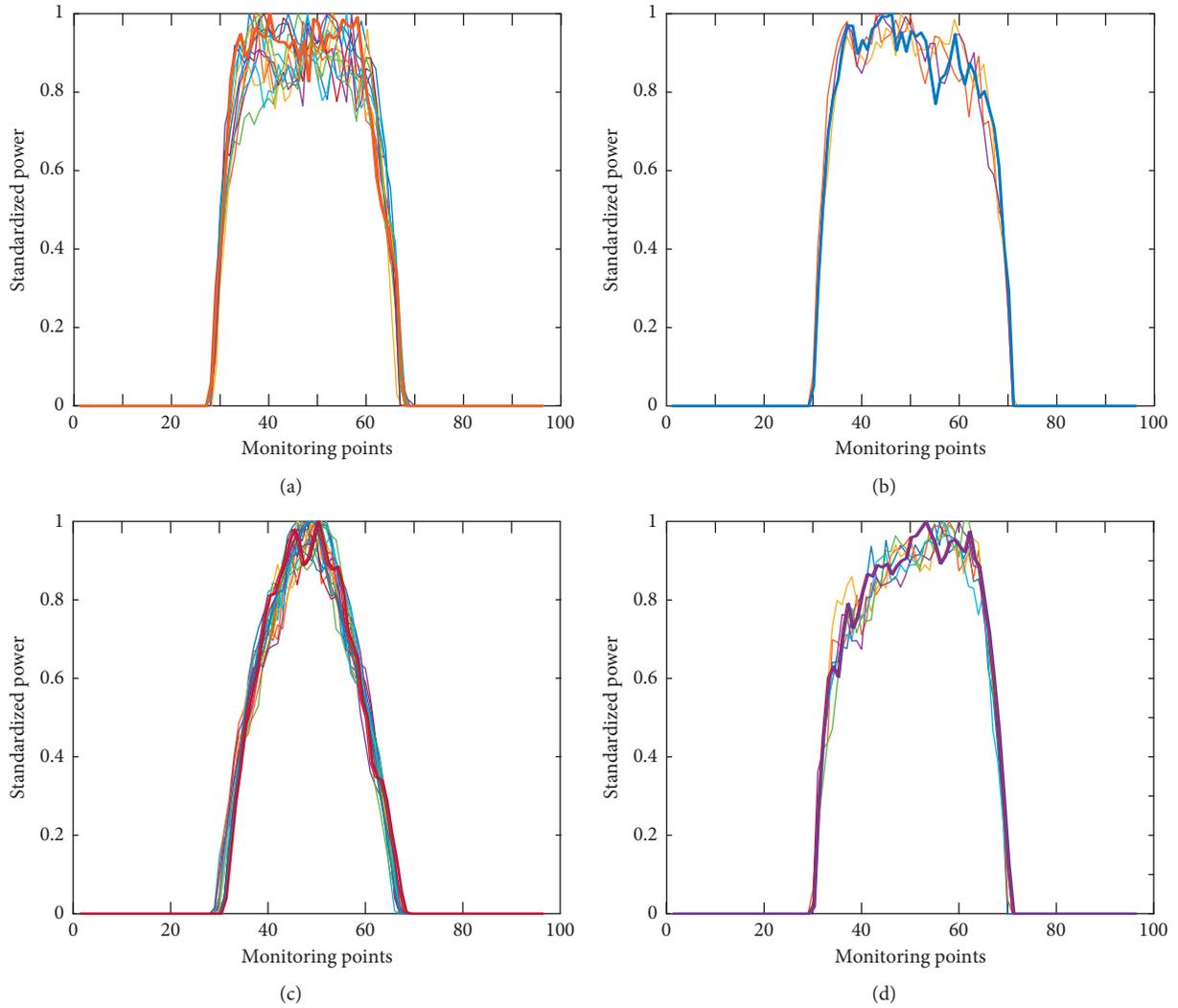


FIGURE 7: Examples of generation patterns obtained with K-means clustering. The numbers of plants in the pattern groups in subplots (a), b, (c), and d are, respectively, 24, 8, 35, and 7.

method given in Section 3. Following the proposed time shift method, the globally most spatially correlated reference plants to the target plant T1 are obtained. The time shift, minimum SED, PPMCC values, and chosen reference stations are listed in Table 1, and the search process is illustrated in Figure 8.

Several randomly chosen target plants are simulated, and the results show saddle-shaped curves similar to those in Figure 8, and the spatially correlated reference plants are usually located in time zones close to the target plants. There exists a minimum among the SED values achieved with different time shifts, and the most spatially correlated reference plant may not be synchronous with the target plants. Therefore, the most spatially correlated reference plants could be found globally with the time shift method, considering the time difference. In addition, the results in Table 1 show that the matched reference plants are different when different time shifts are applied, which means that the consideration of time difference could affect the matching results and further affect the forecast performance.

4.3. Spatial Correlation Matching considering Time Difference. Target plants T1 and T2 are added as two new patterns (patterns 51 and 52) into the pattern library. The clustering threshold is set as 1.40. K-means clustering is performed on the new library, and the results show that T1 is most strongly correlated with pattern 48. The reference plant R1 (785#), which has the highest correlation in that pattern group, is chosen as the master station, and the SED between the standard vectors of T1 and R1 is 1.1221. The spatial correlation results are shown in Figure 9.

In Figure 9, curves *a* and *b* are the real power of the reference PV plant and target PV plant, respectively, and *c* and *d* are the standard vectors of *a* and *b*, respectively. We compare the trajectory curves given that the nominal values of generation output of the two plants are quite different, which is the result of differences in installed capacity, converting efficiency, etc., but the overall changing trends are similar. Therefore, it is verified that the standardized trajectory curve could preserve the similarity of changing trend and could present the significant numerical correlation.

TABLE 1: Reference plant matching results for target plant T1 with different time shifts.

Time shift	Min.(SED)	rp	Ref_plant
-5	3.2552	0.97226	2057#
-4	2.3362	0.98571	2057#
-3	1.6388	0.99297	2057#
-2	1.2218	0.99609	2225#
-1	0.81687	0.99825	2239#
0	0.95176	0.99763	5141#
1	0.85250	0.9981	756#
2	0.88669	0.99794	2176#
3	0.84598	0.99813	4558#
4	1.3721	0.99507	452#
5	2.1725	0.98764	547#

However, the SED between T2's standard vector and the closed pattern's vector is 1.6794, which is higher than the clustering threshold. Thus, T2 will be regarded as a new pattern, and no match is found in the reference plant groups. The forecasting for unmatched DPV plants should adopt other forecast methods.

4.4. Univariate Prediction Based on Spatial Correlation.

LS-SVM regression method is utilized to perform the numerical fitting of the prediction window generation data of R1 and T1, and the correlation relationship model is obtained. Considering that the actual generation in night time is 0, the following modification of the correlation relationship model is made to avoid human introduced error: if the reference plant generation is 0, the target plant generation should also be 0.

As the main purpose of the case study is to examine the forecast performance of the spatial correlation based method, the actual generation data of reference plants is utilized as the short-term forecast results to avoid the forecast errors of the reference plants. The short-term forecast generation is utilized as input of the correlation relationship model, and the entire day-ahead generation trajectory of target plant T1 with rolling calculation is obtained. Figure 10 shows the day-ahead forecast generation curve (green dotted line) and the actual generation curve (black line), with the comparison of forecast results using the temporal correlation method (blue broken line).

As shown in Figure 10, the predictive power of target PV plants with spatial correlation (green dotted line) is basically consistent with the predictive power of reference plants (red dotted line) and is closer to the real power of target plants (black line) compared with the predictive power of target PV considering timing correlation (blue broken line). It is obvious to know that the proposed spatial correlation method is effective and has high precision.

The forecast performance is evaluated with the forecast errors given in Section 3. The forecast errors are given in Table 2. It can be seen from Table 2 that both RMSE and MAE are smaller for the spatial correlation forecast method compared with the temporal correlation forecast method, which signifies that the proposed spatial correlation method achieves higher forecast accuracy.

4.5. Multivariate Prediction Based on Spatial Correlation.

The spatial correlation matching is performed for 1550 target plants randomly chosen from all target plants, and 493 of the target plants fail to find a matching correlated pattern group, taking up 31.8% of all target plants. The short-term generation forecast for these plants should consider using temporal correlation forecast or other forecast methods. Among the rest 1057 target plants, which are matched to reference plant groups, 583 of them have 4 or more reference plants. Using the multivariate prediction function of LS-SVM, the generation forecasts for these 583 plants are obtained. The forecast statistic mean errors are shown in Table 3.

The longitudinal comparison of Table 3 shows that the more reference plants are matched, the more reference information is given, the less the forecast error is. Therefore, when there is more than one match of reference plants, the result of multivariate prediction is better than that of univariate prediction.

The horizontal comparison of Table 3 shows that the forecast based on spatial correlation is more accurate than the forecast based on temporal correlation. The reason is that the temporal forecast method only utilizes the historical generation data, and no information of future change is involved. The spatial correlated forecast method, on the other hand, uses the numeric weather forecast data (in the generation forecast of reference plants) and historical generation data, hence achieving higher forecast accuracy.

4.6. Influence of Prediction Window Size on Spatial Correlation Forecast.

This part discusses the choice of prediction window size and its influence on forecast performance. Considering the limited data storage capability of target plants, we assume that only ten days of historical generation data is available. Use the first nine days' data to generate prediction window data and make a forecast, and the tenth day's data to examine the forecast performance. Figure 11 shows the forecast error of a randomly chosen target plant (plant 1000#) with different prediction window sizes, from 1 day to 9 days. It can be seen that, for example, the forecast error of spatial correlation method is larger than that of temporal correlation using MAE as a criterion if the prediction window size is 3 days or 4 days. The

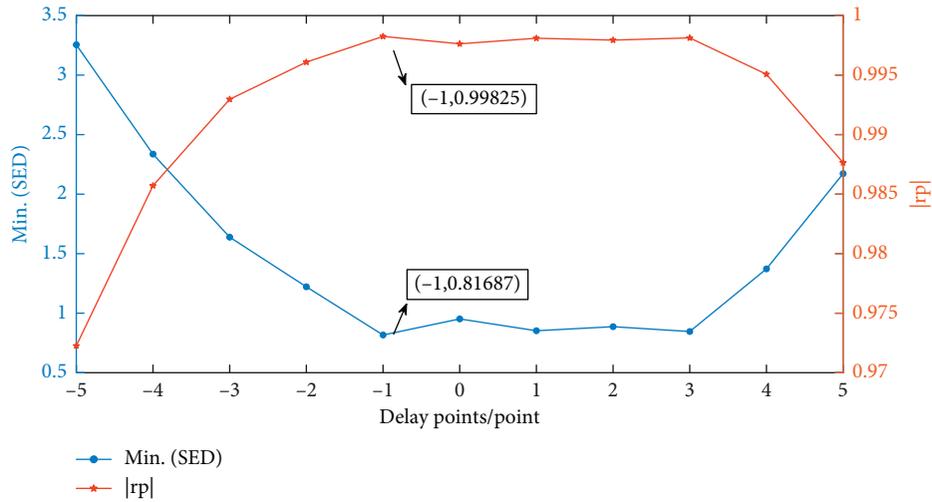


FIGURE 8: Min. (SED) and $|rp|$ values with different time shifts.

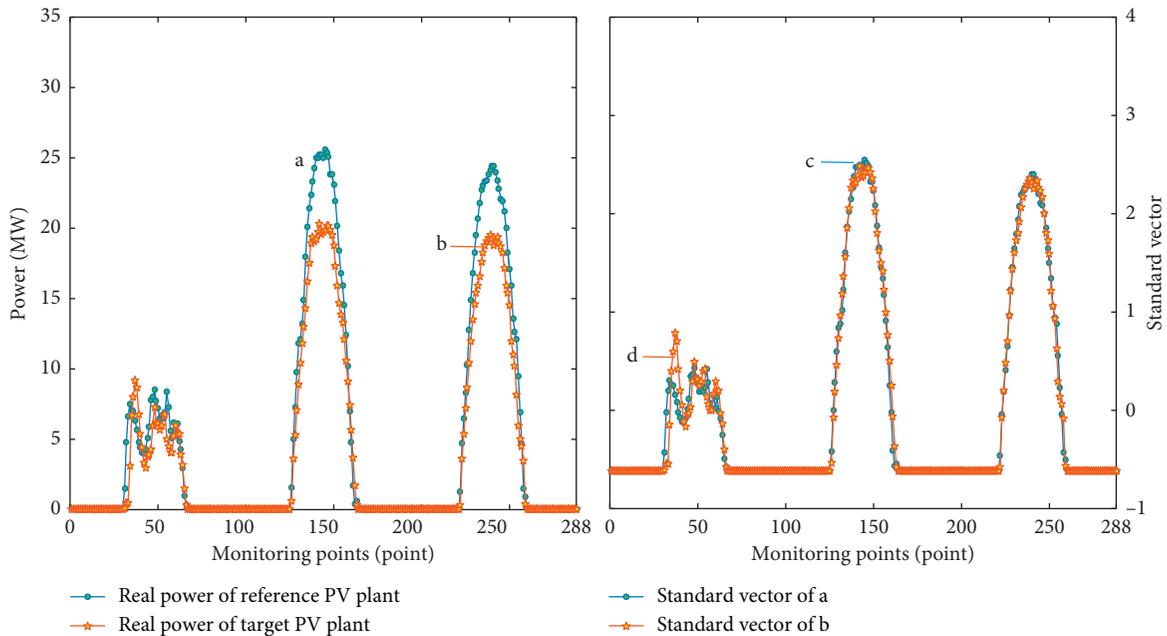


FIGURE 9: The spatial correlation of DPV plant T1 and R1.

forecast results of spatial correlation method with other prediction window sizes are better than those of the temporal correlation method. The optimal prediction window size is 2 days.

Next, 641 target plants are randomly chosen, and the optimal prediction window sizes are counted. As shown in Figure 12, it is noted that the optimal prediction window size is different for each plant, which is influenced by the characteristics of the plant and the surrounding environment. The majority of the plants could achieve good

forecast performance with a prediction window of 3–7 days. Therefore, in practical application, the forecast scheme should be customized for each target plant according to its historical data, the prediction window size should be appropriately selected, and the value of prediction window size should be updated as time goes by. To reproduce the cases, there are four limitations including the data source, the number of reference/target DPV plants, the prediction window size, and offline clustering threshold value setting.

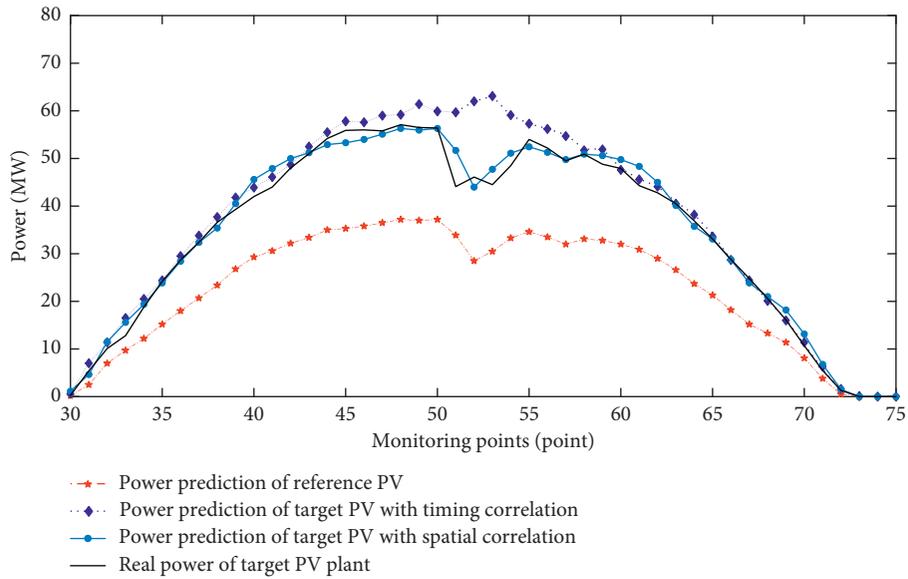


FIGURE 10: Univariate forecast based on spatial correlation and comparison with actual generation and forecast results using temporal correlation.

TABLE 2: Forecast errors of spatial and temporal correlation method (/MW).

Forecast method	RMSE	MAE
Temporal correlation forecast	3.4252	1.2927
Spatial correlation forecast	1.3898	0.6691

TABLE 3: Forecast mean errors of 583 plants using spatial and temporal correlation methods (/MW).

Number (referenced PV plants)	Temporal correlation forecast		Spatial correlation forecast	
	Avg. (RMSE)	Avg. (MAE)	Avg. (RMSE)	Avg. (MAE)
1	2.8990	0.9990	1.8478	0.7819
2	2.4093	0.9538	1.3656	0.7353
3	2.1938	0.9279	1.1381	0.6480
4	2.0283	0.9105	1.0271	0.6161

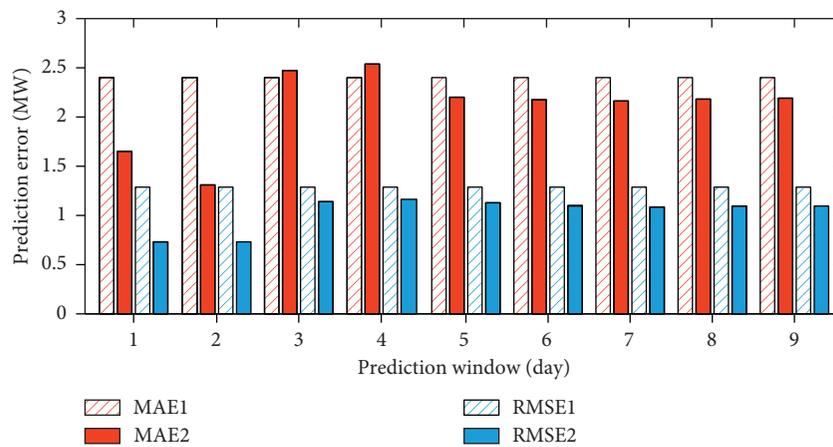


FIGURE 11: Forecast error with different prediction window sizes. The red and blue shady bars are the prediction errors of temporal forecast method, and the red and blue solid bars are the prediction errors of spatial forecast method.

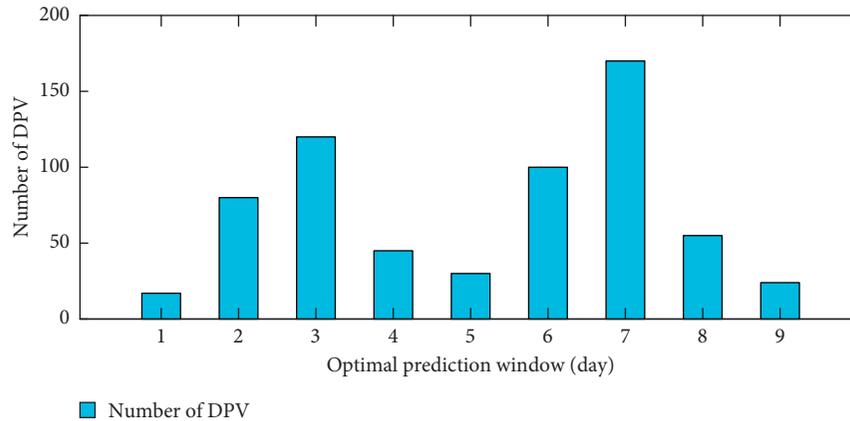


FIGURE 12: Statistical results of different optimal prediction window sizes.

5. Conclusions

Aiming to solve the technique bottleneck of small capacity DPV generation forecast caused by data deficiency and complex influencing factors, this paper proposes an indirect forecast method based on spatial correlation, using a master-slave structure and mapping the target plants incapable of making a forecast on their own to the reference plants, which could make the forecast with sophisticated method. The following conclusions are drawn:

- (1) The historical generation data contain the complete background information such as meteorological data, so that the spatial correlated forecast method for DPV generation could make full use of historical data and achieve accurate short-term forecast.
- (2) Adopting LS-SVM regression for numerical fitting of the spatial correlation relationship could improve the overall forecast accuracy, compared to prediction methods based on temporal correlation and least squares linear regression.
- (3) The proposed spatial correlation forecast method could use the DPV plants that are already equipped with forecast systems and obtain short-term generation forecast for DPV or newly built DPV plants with low investment.

Data Availability

The data used to support the findings of this work are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work is jointly supported by University Natural Science Research General Project of Jiangsu Province (No. 19KJB470004), the High-level Talent Introduction Scientific

Research Foundation of Nanjing Institute of Technology (No. YKJ201820), and Open Research Fund of Jiangsu Collaborative Innovation Centre for Smart Distribution Network, Nanjing Institute of Technology (No. XTCX201906).

References

- [1] A. Sangwongwanich, Y. Yang, F. Blaabjerg, and H. Wang, "Benchmarking of constant power generation strategies for single-phase grid-connected photovoltaic systems," *Institute of Electrical and Electronics Engineers Transactions on Industry Applications*, vol. 54, no. 1, pp. 447–457, 2018.
- [2] F. Ding and B. Mather, "On distributed pv hosting capacity estimation, sensitivity study, and improvement," *Institute of Electrical and Electronics Engineers Transactions on Sustainable Energy*, vol. 8, no. 3, pp. 1010–1020, 2017.
- [3] U. K. Das, K. S. Tey, M. Seyedmahmoudian et al., "Forecasting of photovoltaic power generation and model optimization: a review," *Renewable & Sustainable Energy Reviews*, vol. 81, pp. 912–928, 2017.
- [4] N. Haghgadi, A. Bruce, I. MacGill, and R. Passey, "Impact of distributed photovoltaic systems on zone substation peak demand," *Institute of Electrical and Electronics Engineers Transactions on Sustainable Energy*, vol. 9, no. 2, pp. 621–629, 2018.
- [5] B. Jing, Z. Qian, Y. Pei, and J. Wang, "Ultra short-term PV power forecasting based on ELM segmentation model," *The Journal of Engineering*, vol. 2017, no. 13, pp. 2564–2568, 2017.
- [6] F. Liu, R. Li, Y. Li, R. Yan, and T. Saha, "Takagi-Sugeno fuzzy model-based approach considering multiple weather factors for the photovoltaic power short-term forecasting," *IET Renewable Power Generation*, vol. 11, no. 10, pp. 1281–1287, 2017.
- [7] L. Gigoni, A. Betti, E. Crisostomi et al., "Day-ahead hourly forecasting of power generation from photovoltaic plants," *Institute of Electrical and Electronics Engineers Transactions on Sustainable Energy*, vol. 9, no. 2, pp. 831–842, 2018.
- [8] A. Bracale, G. Carpinelli, and P. De Falco, "A probabilistic competitive ensemble method for short-term photovoltaic power forecasting," *Institute of Electrical and Electronics Engineers Transactions on Sustainable Energy*, vol. 8, no. 2, pp. 551–560, 2017.
- [9] C. Wan, J. Lin, Y. Song, Z. Xu, and G. Yang, "Probabilistic forecasting of photovoltaic generation: an efficient statistical

- approach,” *Institute of Electrical and Electronics Engineers Transactions on Power Systems*, vol. 32, no. 3, pp. 2471–2472, 2017.
- [10] A. Tascikaraoglu, B. M. Sanandaji, G. Chicco et al., “Compressive spatio-temporal forecasting of meteorological quantities and photovoltaic power,” *Institute of Electrical and Electronics Engineers Transactions on Sustainable Energy*, vol. 7, no. 3, pp. 1295–1305, 2016.
- [11] L. A. Fernandez-Jimenez, A. Muñoz-Jimenez, A. Falces et al., “Short-term power forecasting system for photovoltaic plants,” *Renewable Energy*, vol. 44, no. 4, pp. 311–317, 2012.
- [12] F. Wang, Z. Zhen, Z. Mi, H. Sun, S. Su, and G. Yang, “Solar irradiance feature extraction and support vector machines based weather status pattern recognition model for short-term photovoltaic power forecasting,” *Energy and Buildings*, vol. 86, pp. 427–438, 2015.
- [13] C. Monteiro, T. Santos, L. Fernandez-Jimenez, I. Ramirez-Rosado, and M. Terreros-Olarte, “Short-term power forecasting model for photovoltaic plants based on historical similarity,” *Energies*, vol. 6, no. 5, pp. 2624–2643, 2013.
- [14] H. M. El-Helw, A. Magdy, and M. I. Marei, “A hybrid maximum power point tracking technique for partially shaded photovoltaic arrays,” *Institute of Electrical and Electronics Engineers Access*, vol. 5, pp. 11900–11908, 2017.
- [15] J. Liu, W. Fang, X. Zhang, and C. Yang, “An improved photovoltaic power forecasting model with the assistance of aerosol index data,” *Institute of Electrical and Electronics Engineers Transactions on Sustainable Energy*, vol. 6, no. 2, pp. 434–442, 2015.
- [16] K. Y. Bae, H. S. Jang, and D. K. Sung, “Hourly solar irradiance prediction based on support vector machine and its error analysis,” *Institute of Electrical and Electronics Engineers Transactions on Power Systems*, vol. 32, no. 2, pp. 935–945, 2017.
- [17] T. V. Da Silva, R. V. A. Monteiro, G. C. Guimaraes, F. A. M. Moura, M. A. Tamashiro, and M. R. M. C. Albertini, “Performance analysis of neural network training algorithms and support vector machine for power generation forecast of photovoltaic panel,” *Institute of Electrical and Electronics Engineers Latin America Transactions*, vol. 15, no. 6, pp. 1091–1100, 2017.
- [18] M. Bouzardoum, A. Mellit, and A. Massi Pavan, “A hybrid model (SARIMA-SVM) for short-term power forecasting of a small-scale grid-connected photovoltaic plant,” *Solar Energy*, vol. 98, pp. 226–235, 2013.
- [19] A. Mellit, A. Massi Pavan, and V. Lughi, “Short-term forecasting of power production in a large-scale photovoltaic plant,” *Solar Energy*, vol. 105, pp. 401–413, 2014.
- [20] W. VanDeventer, E. Jamei, G. S. Thirunavukkarasu et al., “Short-term PV power forecasting using hybrid GASVM technique,” *Renewable Energy*, vol. 140, pp. 367–379, 2019.
- [21] M. N. Akhter, S. Mekhilef, H. Mokhlis, and N. Mohamed Shah, “Review on forecasting of photovoltaic power generation based on machine learning and metaheuristic techniques,” *IET Renewable Power Generation*, vol. 13, no. 7, pp. 1009–1023, 2019.
- [22] M. Victoria, R. Herrero, C. Domínguez, I. Antón, S. Askins, and G. Sala, “Characterization of the spatial distribution of irradiance and spectrum in concentrating photovoltaic systems and their effect on multi-junction solar cells,” *Progress in Photovoltaics: Research and Applications*, vol. 21, no. 3, pp. 308–318, 2013.
- [23] B. Elsinga and W. Van Sark, “Spatial power fluctuation correlations in urban rooftop photovoltaic systems,” *Progress in Photovoltaics: Research and Applications*, vol. 23, no. 10, pp. 1390–1397, 2015.
- [24] E. Karakaya, “Finite Element Method for forecasting the diffusion of photovoltaic systems: why and how?,” *Applied Energy*, vol. 163, pp. 464–475, 2016.
- [25] T. Zhao, Z. Zhou, Y. Zhang, P. Ling, and Y. Tian, “Spatio-temporal analysis and forecasting of distributed PV systems diffusion: a case study of shanghai using a data-driven approach,” *Institute of Electrical and Electronics Engineers Access*, vol. 5, no. 99, pp. 5135–5148, 2017.
- [26] Y. Liu, Z. Li, K. Bai, Z. Zhang, X. Lu, and X. Zhang, “Short-term power-forecasting method of distributed PV power system for consideration of its effects on load forecasting,” *The Journal of Engineering*, vol. 2017, no. 13, pp. 865–869, 2017.
- [27] M. Lave and J. Kleissl, “Cloud speed impact on solar variability scaling - application to the wavelet variability model,” *Solar Energy*, vol. 91, no. 3, pp. 11–21, 2013.
- [28] R. Li, W. Wang, and M. Xia, “Cooperative planning of active distribution system with renewable energy sources and energy storage systems,” *Institute of Electrical and Electronics Engineers Access*, vol. 6, pp. 5916–5926, 2017.
- [29] National Renewable Energy Laboratory, *Solar Power Data for Integration Studies*, National Renewable Energy Laboratory, Golden, CO, USA, 2006, <https://www.nrel.gov/grid/solar-power-data.html>.
- [30] A. Bradford, G. Weissman, R. Sargent, and B. Fanshaw, *Shining Cities 2017*, <https://environmentamerica.org/sites/environment/files/cpn/AMN-033117-REPORT/shining-cities-2017.html>, Environment America Research & Policy Center, Denver, CO, USA, 2017, <https://environmentamerica.org/sites/environment/files/cpn/AMN-033117-REPORT/shining-cities-2017.html>.