Received March 21, 2019, accepted April 18, 2019, date of publication April 30, 2019, date of current version May 13, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2913946

Equivalent Modeling of Photovoltaic Power Plant Based on Factor Analysis and Correlation Clustering

PINGPING HAN, ZIHAO LIN, JINGJING ZHANG, YU XIA, AND LEI WANG
Anhui Province Laboratory of New Energy Utilization and Energy Conservation, Hefei University of Technology, Hefei 230009, China

Corresponding author: Zihao Lin (lzh36566@163.com)

This work was supported in part by the National Key Research and Development Program of China under Grant 2016YFB0900601, and in part by the Technology Projects of State Grid Corporation of China under Grant 52094017000W.

ABSTRACT The equivalent model of photovoltaic (PV) power plant can greatly improve the model simulation speed on the basis of satisfying the simulation accuracy. However, in equivalent modeling of PV power plant, there are strong correlations and subjectivity between the selected initial variables, which will cause inaccurate equivalent results. Therefore, this paper proposes an equivalent modeling method of PV power plant based on factor analysis and correlation clustering. First, the factor analysis is used to reduce the dimensions of initial variables to eliminate the correlations between clustering indexes. The factor rotation and factor scores are adapted to mine the actual physical meaning of data, extracting the operation states of PV units, and eliminating the subjectivity of clustering indexes. Second, the correlation analysis and significance test are applied to analyze the extracted operation state indexes to obtain the correlations between PV units for clustering. Third, by the contour visualization, the correlations between PV units are qualitatively analyzed and the rationality of clustering results is judged. Finally, the equivalent parameters are calculated based on clustering results to establish the equivalent model of the PV power plant. The simulation results show that the proposed equivalent modeling method not only improves the simulation speed but also fit the simulation accuracy as well. The adaptability of the equivalent modeling method under different working conditions is also verified. The established equivalent model can be used to analyze the interaction between the PV power plant and power grid as well as the power system analysis with large-scale new energy, which is practical and popularized.

INDEX TERMS Correlation clustering, contour visualization, equivalent modeling, factor analysis, PV power plant.

I. INTRODUCTION

The maturity of PV power generation technology promotes the large-scale development of PV power plant. By 2017, the global accumulative PV installed capacity has reached 405 GW. By September 2018, China's total PV installed capacity has reached 165 GW, ranking first in the world, in which large-scale PV power plants account for more than 80%. Due to the large installed capacity, high voltage level and large occupied area, the large-scale PV power plant connecting to power grid has a great impact on the dynamic characteristics of the power grid, which will cause an important impact on the safe and stable operation of the power grid [1]. Therefore, the modeling and simulation of PV power plant is indispensable [2]–[4]. Large-scale PV power plant usually consists of dozens or even hundreds of PV units. Due to the huge model data, complex structure, slow simulation speed and large amount of calculation, the scheme of detailed modeling is not desirable. Therefore, it is of great practical significance to study and establish the equivalent model of PV power plant instead of the detailed model for the simulation and analysis of practical engineering [5]–[7].

Equivalent modeling of PV power plant is to simplify the detailed model of PV power plant with multiple PV units into the equivalent model of PV power plant with only one or a few PV units [8]. The dynamic response characteristics of point

The associate editor coordinating the review of this manuscript and approving it for publication was B. Chitti Babu.

of common coupling (PCC) are consistent before and after equivalent modeling. The traditional single-machine equivalence equates all PV units to one PV unit, which neglects the difference of operation characteristics of each PV unit in PV power plant and the equivalent error is large [9]. Therefore, it is necessary to cluster all PV units according to the internal differences of PV power plant and establish the multi-machine equivalent model for meeting the requirements of simulation speed and accuracy. The core of multi-machine equivalence in PV power plant is the selection of clustering indexes and the use of clustering methods.

Clustering indexes need to be able to characterize the operation conditions and fault information of PV units. The weather parameters such as solar irradiance and temperature are direct factors affecting the fluctuation and difference of PV output [10], [11]. However, due to the complex and varied meteorological factors, there are uncertainties on using environmental parameters as clustering indexes and the operation characteristics of PV units are not considered. The internal control parameters of PV inverters have a significant impact on the dynamic response characteristics of PV power plant, which essentially reflects the differences between PV units. Taking control parameters as clustering indexes has high equivalent accuracy [12], [13]. However, the control parameters are not easy to obtain in practical engineering. In the process of equivalence, parameter identification is used to identify the control parameters, which increases the difficulty of equivalence. Operation states of PV units include voltage drops, the action of protection circuit, low voltage ride through (LVRT) performance, transient and steady-state operation conditions and so on. Clustering with such indexes has clear physical meaning and good identification [14], [15], but these operation states are difficult to quantify and cannot be directly obtained as clustering indexes. Output characteristics of PV units are easy to obtain in practical engineering and can well characterize the operation states of PV units, which is more practical to be used as clustering indexes [16], [17]. However, output characteristics are mostly multi-dimensional time series data. The selected index variables have strong correlations and subjectivity, resulting in inaccurate clustering results.

Clustering methods can be summarized into two types: One is clustering with algorithms, such as k-means algorithm based on partition [18], DBSCAN algorithm based on density [19], fuzzy clustering algorithm [20] and support vector machine algorithm [21]. These algorithms usually measure the feature distance between data points in multi-dimensional space to cluster [22], which has fast processing speed. However, the unsupervised classification makes the clustering results vulnerable to sensitive data such as data types, isolated data points and edge data points and cannot obtain the global optimal solution, which affects the accuracy of clustering results. The other is clustering with feature information of PV units. This kind of feature information includes the action state of protection circuit [23], the landmark demarcation point [24], [25] obtained by analyzing the clustering effect of response curve, the correlations [26] between PV units and so on. These methods do not depend on complex algorithms and PV units are directly clustered according to their significant feature differences. The principle is simple and the calculation is convenient, but the feature information used for directly clustering is difficult to obtain or quantify effectively, which needs further study.

Based on the above problems, this paper analyses the change of operation states of PV power plant before and after three-phase short-circuit fault, selecting the transient and steady-state data of output response characteristics as initial variables. Factor analysis is used to reduce the dimensions of the initial variables and mine the intrinsic information of the data and by which the correlations and subjectivity of the initial variables are eliminated. Taking the objective data after factor analysis as clustering indexes, correlation analysis and significance test are applied to measure the correlations between PV units for clustering analysis. By the contour visualization method, the correlations between PV units are qualitatively analyzed and the rationality of clustering results is judged. The equivalent model of PV power plant is established in DIgSILENT/PowerFactory software, in which the numerical simulation is carried out to verify the effectiveness and the adaptability to different working conditions of the equivalent modeling method proposed in this paper.

II. SELECTION AND ANALYSIS OF INITIAL VARIABLES IN PV POWER PLANT

A. SELECTION OF INITIAL VARIABLES

In order to meet the requirement of the model for electromechanical transient stability analysis, the equivalent model of PV power plant is established under the three-phase short-circuit fault condition at PCC. Because the time scale of short-circuit fault (10^{-3}-10^{0}s) is very small, there are differences only between the initial steady-state operation points of PV units in the PV power plant before the fault. Therefore, the solar irradiance variation in the long-time scale (10^{0}-10^{3}s) is not considered. In the equivalent modeling of PV power plant, the selected initial variables should be able to fully characterize the operation states of PV units under the whole operating conditions, so as to make the clustering results more accurate. In this section, the output response characteristics of PV units which are easy to obtain in practical projects are selected to analyze the operation states of PV power plant under three-phase short-circuit fault conditions and appropriate sampling time points are selected for subsequent analysis.

The simulation model of single grid-connected PV unit is established in DIgSILENT/PowerFactory software, as shown in Figure 1. The PV unit is composed of PV array, DC bus capacitor, DC-side chopper protection circuit [27] and PV inverter. The PV inverter has the capability of LVRT control. Three-phase short-circuit fault simulation for the model is carried out to obtain four output external characteristic waveforms at PCC: active power P, reactive power

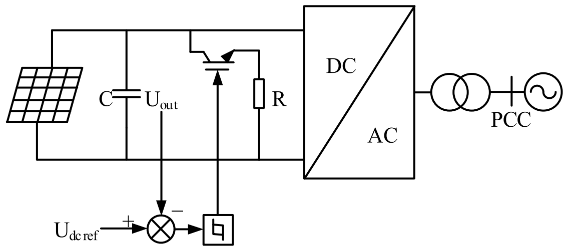


FIGURE 1. Single grid-connected PV unit.

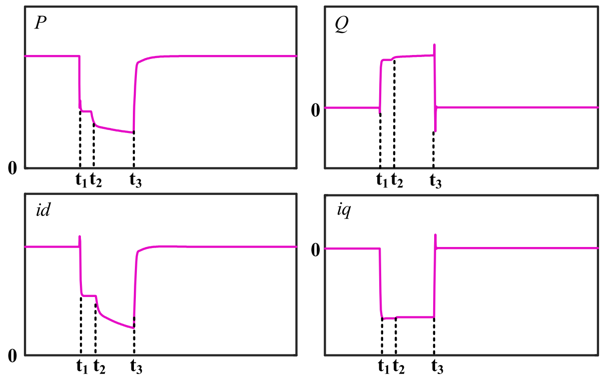


FIGURE 2. Output external characteristic of single grid-connected PV unit.

TABLE 1. Analysis of operation states of PV unit.

Time	Operation track	Operation states
[0, t_1]	Operate steadily	Initial steady-state operation
[t_1, t_2]	Be out of control for a short time	The start of short-circuit fault; The action of protection circuit
[t_2, t_3]	P and Q drop; id and iq uplift	Transient stability of voltage drop; The realization of LVRT capability
[t_3, ~)	Restore stable operation	Fault clearance; Restoring steady-state operation

Q, active current id and reactive current iq, as shown in Figure 2.

The operation states characterized by the four output variables of the PV unit are analyzed and summarized as Table 1. According to the operation states described in Table 1, the transient and steady-state data of P, Q, id and iq are selected as the initial variables. The selection basis of the transient and steady-state data is as follows: Select one time point for the initial steady-state period in [0, t_1] with subscript 1; Select two time points during the transition period of protection circuit acting in [t_1, t_2] with subscript 2,3; Select two time points during the transient stability period in [t_2, t_3] with subscript 4,5; Select one time point during the steady-state period after fault clearance in [t_3, $\sim$) with subscript 6. Finally, the 24-dimensional initial variables of $P_1 \sim P_6$, $Q_1 \sim Q_6$, $id_1 \sim id_6$ and $iq_1 \sim iq_6$ are obtained.

B. CORRELATION ANALYSIS OF INITIAL VARIABLES

Four output external characteristics indexes can fully characterize the operation states of PV power plant. However, the strong correlations between variables will affect the accuracy of clustering results. In the selected 24-dimensional initial variables, four groups of invalid zero variables Q_1, Q_6, iq_1 and iq_6 are eliminated. The correlation analysis (shown in Section IV) of 20-dimensional initial variables is carried out and the obtained correlation coefficients between variables are shown as Figure 3. In clustering analysis, the weight of strongly correlated index data will be enlarged, resulting in the tendency of clustering results. Therefore, before clustering the initial variables, it is necessary to eliminate the correlations between the indexes.

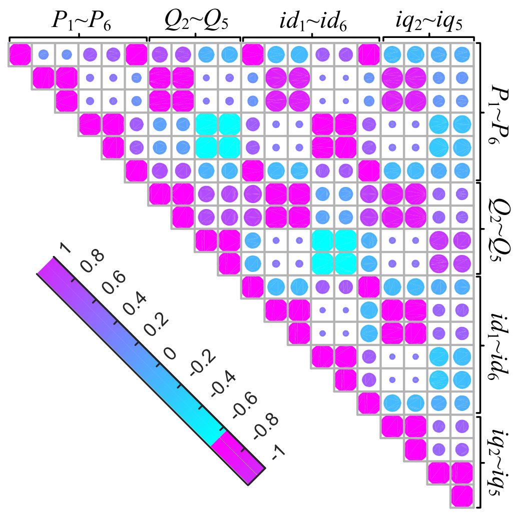


FIGURE 3. Correlation analysis of initial variables.

C. SUBJECTIVITY ANALYSIS OF INITIAL VARIABLES

If the initial variables are not processed by data mining, the time seriality of artificial selection and the uncertainty of self-weight make the clustering indexes too abstract to observe the internal logic, so that the degree of impact of clustering indexes on clustering results cannot be evaluated, which weakens the objectivity and authenticity of the clustering results. In view of the existing subjectivity of the initial variables, it is necessary to adopt certain methods to fully mine the essential information of the index variables and enhance the guidance and objectivity of clustering indexes, so that the clustering indexes can reflect the importance of a state variable in PV power plant and realize the objective information expression of the subjective data.

In this paper, factor analysis is used to reduce the dimensions of the initial variables to eliminate the correlations between the initial variables. Furthermore, the intrinsic information of the data is fully excavated to extract the operation state indexes of PV units, which eliminates the subjectivity of the initial variables.

III. FACTOR ANALYSIS MODEL OF CLUSTERING INDEXES

The idea of factor analysis is to group the initial variables according to the correlations, so that the correlations between variables in the same group are higher, while the correlations between variables in different groups are lower. Each group of variables is represented by an unobservable synthetic variable, which converts the initial variables with intricate relationships into a few synthetic factor variables [28].

The normalized 20-dimensional initial variable matrix is denoted as $X = (X_1, X_2, \cdots, X_i, \cdots X_{20})^{\mathrm{T}}$. Factor analysis model is established as Formula (1), which is used to reduce the dimensions of the initial variable matrix:

$$X = AF + \varepsilon, \tag{1}$$

where $F = (F_1, F_2, \cdots, F_k, \cdots, F_m)^{\mathrm{T}}$ is the vector of common factors and the number of common factors is m; $A = (a_{ik})_{20\times m}$ is the factor load matrix and a_{ik} represents the load of the ith dimensional variable X_i on the kth common factor F_k. The bigger the absolute value of a_{ik}, the greater the dependence of X_i on F_k; $\varepsilon = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_{20})^{\mathrm{T}}$ is the vector of special factors which is the part of the initial variables that cannot be explained by common factors.

Establishing factor analysis model includes three steps: calculating factor load, implementing factor rotation and calculating factor scores.

A. CALCULATING FACTOR LOAD

The factor load matrix is calculated according to the principal component method as Formula (2):

$$A = \left(\sqrt{\lambda_1}u_1, \sqrt{\lambda_2}u_2, \cdots \sqrt{\lambda_m}u_m\right), \tag{2}$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_m > 0$ are the first m largest eigenvalues of the initial variable correlation matrix R; $u_1, u_2, \cdots, u_m$ are eigenvectors corresponding to eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_m$. The number m of common factors is determined by the cumulative variance contribution rate $\sum_{k=1}^{m} \lambda_k / \sum_{k=1}^{20} \lambda_k$. When the cumulative variance contribution rate reaches 90%, the information amount of the selected common factors can effectively represent the information of initial variables.

B. IMPLEMENTING FACTOR ROTATION

The purpose of establishing factor analysis model is not only to find common factors and to group variables, but also to make a scientific analysis of practical problems with an explanation of the physical significance of each common factor. According to the non-uniqueness of the factor load matrix, the maximum variance orthogonal rotation of the factor load matrix is implemented.

The maximum variance orthogonal rotation is to multiply the factor load matrix by the orthogonal matrix to maximize the square variance of each column of factor loads. After maximum variance orthogonal rotation, each variable has a larger load on only one common factor, of which the absolute value tends to be 1, while the loads on the other common factors are smaller, of which the absolute values tend to be 0. Factor load a_{ik} reflects the correlation between variable X_i and common factor F_k. The greater the absolute value, the stronger the correlation.

In the light of initial variables selected from PV power plant in this paper, the correlations between variables and common factors can be judged according to the rotated factor load matrix. The common factors are named according to the operation states of PV power plant characterized by variables and endowed with practical physical significance.

C. CALCULATING FACTOR SCORES

Factor scores are the estimates of common factors corresponding to each sample, which can be calculated by the idea of least square method as follows:

$$\hat{F} = A'^{\mathrm{T}} R^{-1} X, \tag{3}$$

where $\hat{F}$ is the factor score matrix of common factors on each sample; A' is the rotated factor load matrix; R is the correlation matrix of the initial variable matrix X.

In the equivalent modeling of PV power plant, potential common factors are obtained by factor analysis on initial variables. Using factor scores of each PV unit as clustering indexes to cluster can reduce the complexity of problem analysis and improve the accuracy of clustering results.

IV. CORRELATION CLUSTERING OF PV UNITS

A. CORRELATION ANALYSIS AND SIGNIFICANCE TEST

The idea of clustering analysis is to use similarity to measure the degree of affinity and sparsity between samples to achieve classification. In this paper, correlation analysis and significance test are used to measure the correlations between PV units to accomplish clustering.

Correlation analysis is to analyze two or more related variables or samples to measure the degree of correlation between variables or samples. In the correlation analysis on clustering indexes of PV units, the factor scores are regarded as samples and PV units are regarded as population. The correlation coefficients between samples are calculated as the estimates of the correlation coefficients between population. The formula for calculating the correlation coefficients is as follows:

$$c_{hj} = \frac{Cov(\hat{F}_h, \hat{F}_j)}{\sqrt{Var(\hat{F}_h)}\sqrt{Var(\hat{F}_j)}}, \tag{4}$$

where c_{hj} is the correlation coefficient between sample $\hat{F}_h$ and sample $\hat{F}_j$; $Cov(\hat{F}_h, \hat{F}_j)$ is the covariance between sample $\hat{F}_h$ and sample $\hat{F}_j$; $Var(\hat{F}_h)$ and $Var(\hat{F}_j)$ are variances of sample $\hat{F}_h$ and sample $\hat{F}_j$ respectively.

In correlation analysis, the selected samples are only a part of population. The correlation coefficients between samples may not reflect the correlations between population. In order to judge the representativeness of samples to population, it is necessary to carry out significance test on the correlation coefficients. Significance test is to make a hypothesis about the distribution of population at first and then use the sample information to accept or reject the hypothesis according to the principle of "the actual impossibility of small probability events".

In the process of correlation clustering of PV units, the correlation coefficients between factor scores are obtained by correlation analysis. Then the correlations between PV units are obtained by significance test on correlation coefficients. Finally, the strongly correlated PV units are clustered into the same group according to the correlations between PV units. The flow chart is shown in Figure 4.

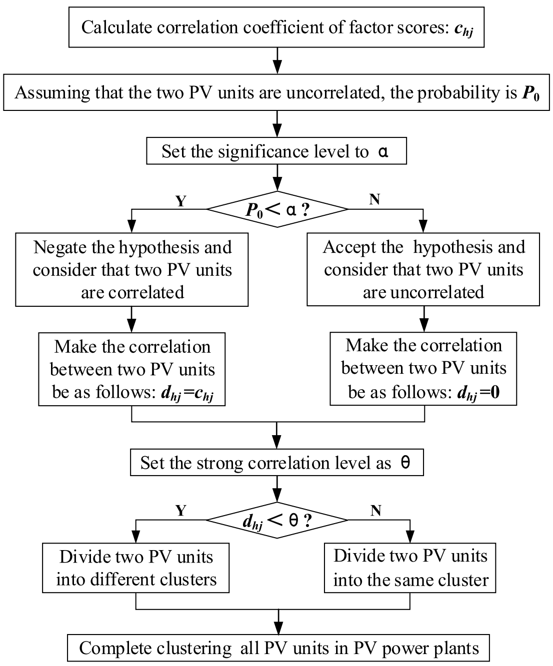


FIGURE 4. The flow chart of correlation clustering of PV units.

B. THE ANALYSIS OF CLUSTERING RESULT BASED ON CONTOUR VISUALIZATION

In this paper, clustering is carried out by measuring the correlations between PV units. The quantitative analysis and unsupervised clustering process make the correlations abstract and the clustering results are not intuitive enough. There is no specific criterion to prove the rationality of clustering results. Based on this, a line-based multi-dimensional data visualization method is proposed. The contour diagram is used to visualize the correlations as well as clustering results of PV units, which visually display the correlations between PV units and qualitatively analyze the rationality of clustering results. Combining factor analysis and contour visualization, this paper uses dimension-reduced data to construct contour diagram, which enhances the visualization effect. This method can not only analyze the correlations between PV units in the overall trend, but also show the difference of PV units of each operation state in the corresponding common factor dimension. The specific steps are as follows:

Step 1: Make a plane coordinate system. Take m points with equal spacing on abscissa to represent m common factors;

Step 2: Taking the factor scores as ordinates, make m points of the PV units in m common factors;

Step 3: Connect m points to get the contour of each PV unit which is a polygonal line and n PV units can make n polygonal lines to form a contour diagram.

Step 4: Display n PV units in the form of clustering results in the contour diagram for visual analysis.

V. EQUIVALENT PARAMETERS

After clustering the PV units, it is necessary to equate all PV units in each cluster into one PV unit and to equate power collection lines as well as unit transformers to establish the equivalent model of PV power plant. Figure 5 is a schematic diagram of the equivalent model of PV power plant, which consists of s equivalent PV units, s equivalent unit transformers and power collection lines. The equivalence of power collection lines is regarded as the calculation of equivalent impedance. The main transformer does not need to be established its equivalent model, which can directly adopt a detailed transformer model.

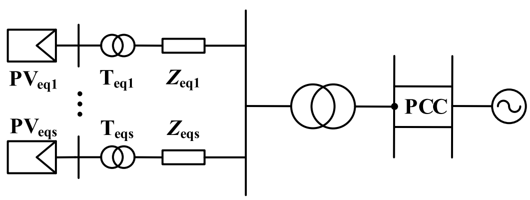


FIGURE 5. Equivalent Model of PV power plant.

The basic principle of equivalence is to ensure that the injection power at PCC is the same and the losses of power collection lines are equal before and after equivalence. Suppose that there are δ PV units in a cluster, the equivalent parameters are calculated as follows:

A. EQUIVALENT PV UNITS AND EQUIVALENT UNIT TRANSFORMERS

For the parameters of equivalent PV units and equivalent unit transformers, the multiplier method is adopted according to the series-parallel relationship [29]. The calculation formulas are as follows:

$$\begin{cases} S_{\mathrm{eq}} = \sum_{j=1}^{\delta} S_j, & S_{\mathrm{Teq}} = \sum_{j=1}^{\delta} S_{\mathrm{T}j} \\ C_{\mathrm{eq}} = \sum_{j=1}^{\delta} C_j, & L_{\mathrm{eq}} = \dfrac{L_j}{\delta} \\ Z_{\mathrm{ceq}} = \dfrac{Z_{\mathrm{c}j}}{\delta}, & Z_{\mathrm{Teq}} = \dfrac{Z_{\mathrm{T}j}}{\delta}, \end{cases} \tag{5}$$

where S_j, S_{T} are respectively the rated capacities of the jth PV unit and the corresponding unit transformer in this cluster; C_j,

L_j are respectively the capacitance parameter and inductance parameter of the jth PV unit in this cluster; Z_{cj}, Z_{Tj} are respectively the impedances of the jth PV unit inverter and the corresponding unit transformer in this cluster; Subscript eq stands for equivalent parameters.

B. THE EQUIVALENCE OF POWER COLLECTION LINES

In practical projects, PV power plants are mostly trunk topology. The equivalence of power collection lines is usually based on the principle that total loss equals the sum of branch losses to obtain the equivalent impedance [30]. However, this method is not suitable for cross equivalence between multi-trunk lines. Therefore, supposing that the voltage amplitude of each PV unit is equal, this paper proposes an improvement on the equivalence of power collection lines.

Step 1: Equivalent impedance of the power collection line of each PV unit on trunk lines is calculated by equal power loss method, by which the trunk topology is transformed into radial topology. As shown in Figure 6, there are n PV units connected by trunk topology to PCC and the equivalent impedance of the power collection line of each PV unit is calculated according to the following formula after being converted into the radial topology:

$$Z_{\mathrm{eq}w} = \sum_{j=w}^{n} \frac{(P_1 + P_2 + \cdots + P_j)^2}{P_1^2 + P_2^2 + \cdots + P_j^2} Z_j, \tag{6}$$

where w is one of the PV units, $1 \leq w \leq n$; Z_j is the line impedance of the jth PV unit; P_j is the active power flowing through Z_j; $Z_{\mathrm{eq}w}$ is the equivalent line impedance of PV unit w in radial topology.

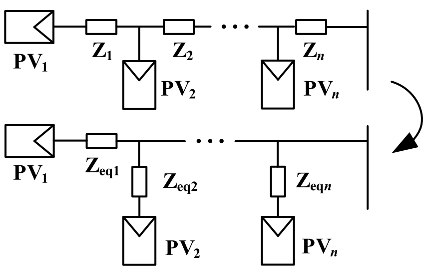


FIGURE 6. Trunk topology to radial topology.

Step 2: According to the clustering results, the total equivalent impedance of all PV units is calculated by equal power loss method as the equivalent impedance of power collection lines. The calculation formula is as follows:

$$Z_{eq} = \sum_{h=1}^{\delta} (P_h^2 Z_{\mathrm{eq}h}) \Big/ (\sum_{h=1}^{\delta} P_h)^2, \tag{7}$$

where $Z_{\mathrm{eq}h}$ is the equivalent line impedance of PV unit h in radial topology; P_h is the active power flowing through $Z_{\mathrm{eq}h}$; Z_{eq} is the equivalent impedance of power collection lines in the clustering.

VI. SIMULATION AND VERIFICATION

A. THE SETTING OF EXAMPLES

Take the PV power plant with 20MW capacity shown in Figure 7 as an example system. The PV power plant consists of 40 PV units and power collection system. Each PV unit has 400V outlet voltage, which is connected to 35KV bus by unit transformer and power collection line. All PV units are connected to the bus bar, boosted by main transformer, and connected to 220KV power grid through double-circuit line.

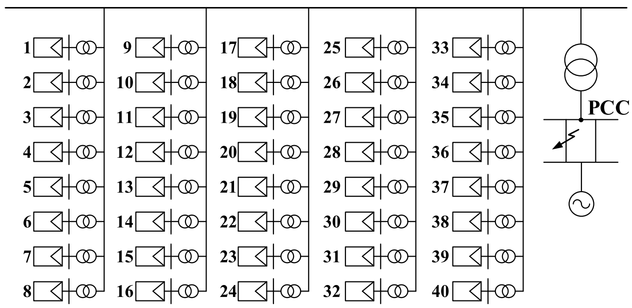


FIGURE 7. The topology of PV power plant.

The detailed model of PV power plant was established in DIgSILENT/PowerFactory software. Suppose that the three-phase short-circuit fault occurred at the midpoint of the double-circuit line of the PV power plant. The fault started at 0.1s and was cleared at 0.2s, during which the voltage at PCC dropped to 0.3p.u.. The transient and steady-state data of P, Q, id, iq were selected as the 20-dimensional initial variables. Factor analysis of the initial variables showed that when the number of common factors was 4, the cumulative variance contribution rate reached 98%. Finally, the rotating factor load matrix and the factor scores of each PV unit were obtained. The rotated factor load matrix was counted by bar chart, as shown in Figure 8.

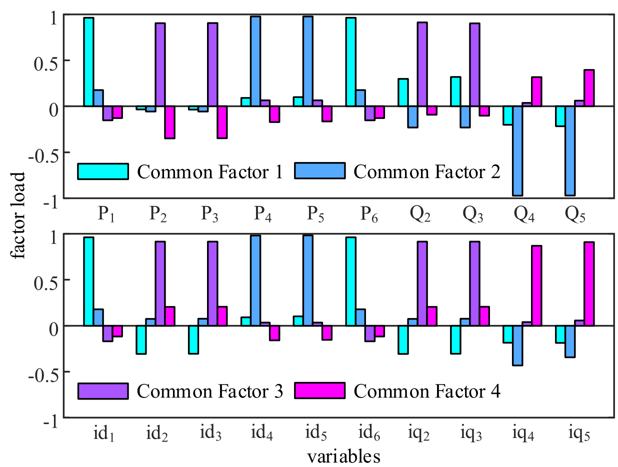


FIGURE 8. The rotated factor load matrix.

From Figure 8, it can be seen that the Common Factor 1 is strongly positively correlated with the active power and active current at steady state; Common Factor 2 is strongly positively correlated with the active power and active current at transient time, and strongly negatively correlated with the reactive power at transient time; Common Factor 3 is strongly positively correlated with the active power, reactive power, active current and reactive current when protection circuit acts; Common Factor 4 is strongly positively correlated with the reactive current at transient time. Combined with operation states analysis of PV power plant in Section II, four common factors were given practical physical significance, named as steady-state factor, drop factor, protection factor and LVRT factor. Factor scores of these four factors on each PV unit were used as clustering indexes for clustering analysis.

B. CLUSTERING ANALYSIS

The correlation analysis was carried out on the factor scores of 40 PV units obtained by factor analysis to obtain the correlation coefficient matrix of the factor scores. The significance test was carried out on the correlation coefficient matrix to obtain the correlations between the PV units. According to the correlations, the clustering results of PV units were obtained in Table 2.

TABLE 2. Clustering results of PV units.

Equivalent model	The serial number of PV units
Cluster 1	1, 2, 6, 7, 9, 14, 15, 17, 25, 28
Cluster 2	3, 4, 11, 12, 13, 18, 19, 20
Cluster 3	5, 8, 24, 31, 32, 39, 40
Cluster 4	10, 16, 21, 22, 23, 26, 27, 29, 30, 33, 34, 35, 36, 37, 38

In Table 2, 40 PV units were divided into four categories. The chopper protection circuits of PV units in Cluster 1, 2, 4 were triggered due to the increase of DC bus voltages, while the chopper protection circuits of PV units in Cluster 3 were not. In order to qualitatively analyze the correlations and clustering results of PV units, the contour diagram of factor scores was constructed for visualization analysis, as shown in Figure 9.

In the contour visualization diagram of Figure 9, the abscissa represents the four common factors of the initial variables after factor analysis and the ordinate represents the factor scores of each PV unit. Each polygonal line represents one PV unit. From the figure, we can see that the change trend of polygonal lines is consistent and the correlations are strong for PV units in the same cluster, while the change trend of polygonal lines is different and the correlations are weak for PV units in different clusters. In different operation states dimensions, each cluster has a good discrimination. The above visualization analysis validates the rationality of clustering results effectively. In particular, there are significant differences in the change trend of polygonal lines of cluster 3 in which the chopper protection circuits did not act

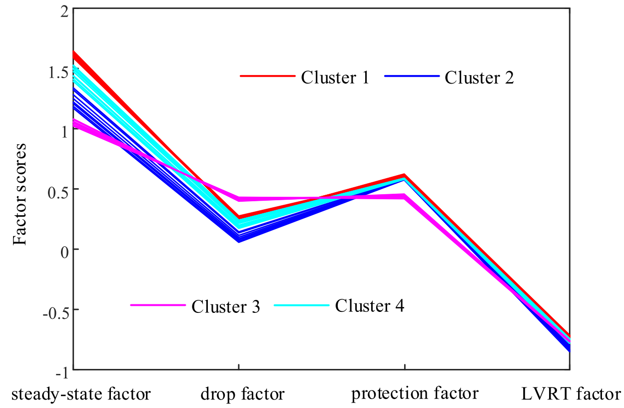


FIGURE 9. Contour visualization analysis of factor scores.

and cluster 1, 2, 4 in which the chopper protection circuits acted, as well as the significant differences in protection factor dimension.

C. VERIFICATION OF EFFECTIVENESS

The equivalent model of PV power plant was established in DIgSILENT/PowerFactory software. According to the above clustering results, 40 PV units were clustered into four clusters. The equivalent parameters were calculated according Section V to establish the equivalent model. In order to verify the effectiveness of the equivalent modeling method used in this paper, the fitting accuracy of three sets of equivalent models for the output characteristics of detailed models under the same short-circuit fault conditions was compared and analyzed. The first group was the single-machine equivalent model of PV power plant. The second group was the equivalent model A of PV power plant which was clustered directly by correlations without factor analysis on the initial variables. The third group was the equivalent model B of PV power plant obtained by the proposed method. The clustering results of PV units of equivalent model A are shown in Table 3.

TABLE 3. Clustering results of equivalent model A.

Equivalent model	The serial number of PV units
Cluster 1	1, 2, 6, 7, 9, 14, 15, 16, 17, 23, 25, 27, 28, 29, 30, 33, 34, 35, 36
Cluster 2	3, 4, 10, 11, 12, 13, 18, 19, 20, 21, 22, 26, 27, 38
Cluster 3	5, 8, 24, 31, 32, 39, 40

The fault started at 0.1s and was cleared at 0.2s, during which the voltage at PCC dropped to 0.3p.u.. Compare the accuracy of three sets of equivalent models fitting the output characteristics of detailed models, as shown in Figure 10. Based on the output characteristics of the detailed model of PV power plant, the error evaluation index of the equivalent model of PV power plant is defined as follows:

$$e = \frac{1}{s}\sum_{a=1}^{s}\frac{T_a^* - T_a}{T_a}, \tag{8}$$

where T_a^* is the output characteristic at PCC of the equivalent model of PV power plant; T_a is the output characteristic at PCC of the detailed model of PV power plant; s is the number of sampling points.

TABLE 4. Error evaluation index of equivalent model.

Error evaluation index	Active power	Reactive power	Output current
Single-machine equivalent model	0.01829	0.22566	0.00994
Equivalent model A	0.00535	0.03540	0.00326
Equivalent model B	0.00062	0.00205	0.00039

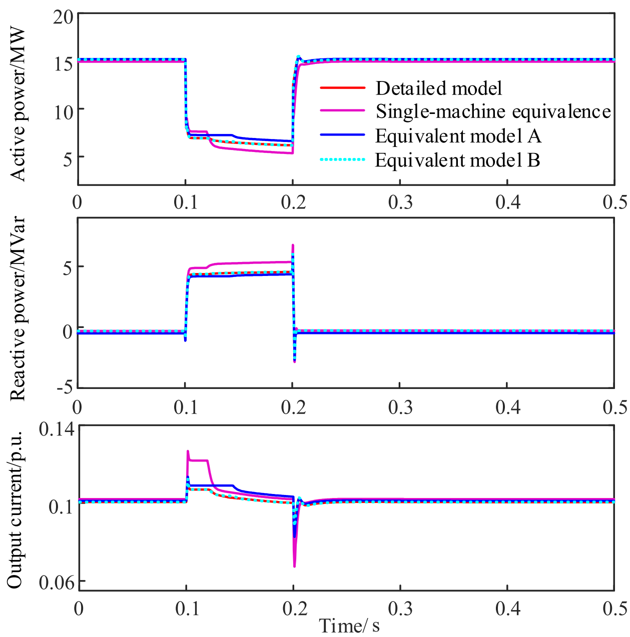


FIGURE 10. The effect comparison of equivalent models.

Calculate the error evaluation index of the equivalent model with the sampling step of 0.001s according Formula (8), as shown in Table 4. Combining with Figure 10 and Table 4, it can be seen that the equivalent model obtained by factor analysis and correlation clustering has higher accuracy in fitting the output characteristics of PV power plant. The fitting accuracy of single-machine equivalence is the lowest, for which the single-machine equivalence regarded the operation states of all PV units as coherence change, using capacity weighting to equate them to one PV unit, which failed to effectively distinguish the action state of chopper protection circuits and the difference of each operation state. Equivalent model A used the initial variables that had not been processed by factor analysis to carry out correlation clustering, of which the principle was clear, the calculation was simple, and the fitting accuracy was higher than that of single-machine equivalence. However, due to the unprocessed initial variables, the correlations and subjectivity between variables would affect the judgment of data changes, causing a certain tendency for clustering results. It was impossible to divide the PV units more accurately and meticulously of which chopper protection circuits acted. Equivalent model B used clustering indexes that had been processed by factor analysis to carry out correlation clustering, which eliminated the correlations between initial variables. At the same time, the rotating technology was used to explain the factors, fully mining the intrinsic information of the data and effectively quantifying the operation states of PV units, which realize the objective information expression of the subjective data. It made the clustering indexes more instructive, so as to display more data information and effectively measure the correlations between PV units. Using factor data to cluster with actual information can objectively guide the clustering process and evaluate the clustering results, eliminating the drawbacks of subjective data affecting the judgment of clustering results.

D. ADAPTABILITY TO DIFFERENT VOLTAGE DROP DEPTHS

In order to verify the adaptability of the equivalent modeling method to different voltage drop depths, the voltages at PCC were dropped to 0.9 p.u., 0.8 p.u., 0.7 p.u., 0.6 p.u., 0.5 p.u., 0.4 p.u., 0.3 p.u., 0.2 p.u., 0.1 p.u. and 0.0 p.u., respectively. The equivalent models of PV power plant with all voltage drop depths were established by the proposed method. Calculate the active power errors according to Formula (8) and compare them with single-machine equivalence and correlation clustering equivalence without factor analysis, as shown in Figure 11.

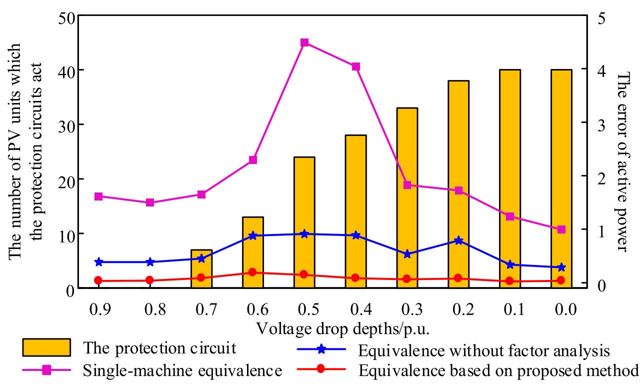


FIGURE 11. Comparison of equivalent models with different voltage drop depths.

Figure 11 shows that for three-phase short-circuit faults with different voltage drop depths, the errors of single-machine equivalence are always the largest, while the errors of the equivalent model obtained by factor analysis and correlation clustering in fitting detailed model keep at a low level. With the increase of voltage drop depth, the error of equivalent model, especially that of single-machine model, increases first and then decreases. This is because the number of PV units which the chopper protection circuits acted increased with the increase of voltage drop depth at PCC. When the voltage drop depth was small, the chopper protection circuits of all PV units did not act. When the voltage drop depth was large, the chopper protection circuits of all PV units acted because of the dramatic increase of the DC bus voltage. In both cases the difference of the operation states of PV units in PV power plant is small with relatively consistent coherence and the equivalent accuracy is relatively high. For the voltage drop depth of only part of the chopper protection circuits acting in PV power plant, if it is not distinguished and clustered according to the operation states of PV power unit, it will cause large errors. By factor analysis to quantify the operation states of PV units and correlations to cluster the factor scores, the obtained equivalent models have good adaptability to different voltage drop depths.

In order to further verify the fitting effect of the equivalent model to the detailed model, choose the voltage at PCC dropping to 0.7p.u. for further simulation. In the case of the number of PV units which the chopper protection circuits acted is small, the simulation results of the equivalent model D established by the proposed method, single-machine equivalent model and the equivalent model C obtained directly by correlation clustering without factor analysis in fitting the output characteristics of detailed model were compared, as shown in Figure 12. The simulation results in Figure 12 show the high accuracy of the equivalent model established by proposed method in fitting the output characteristics of the detailed model, which further verifies that the proposed method is suitable for the equivalent modeling of different voltage drop depths.

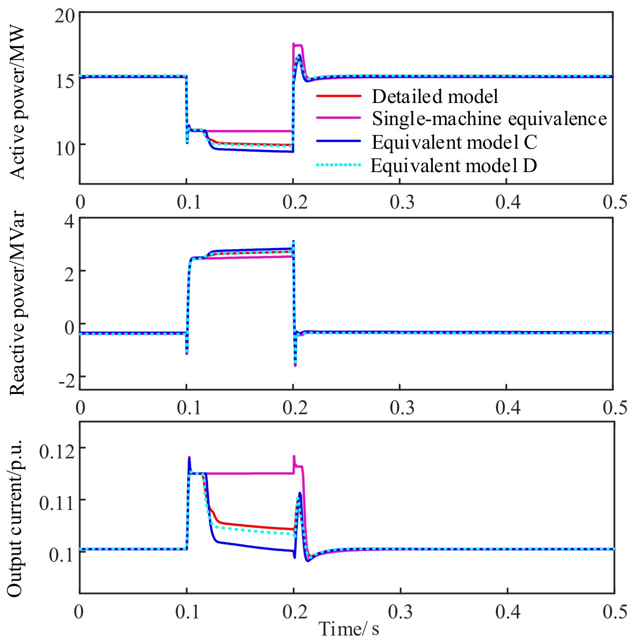


FIGURE 12. The effect comparison of equivalent models (0.7p.u.).

E. ADAPTABILITY TO THE SOLAR IRRADIANCE VARIATION

In a short-time scale, by three-phase short-circuit fault simulation of PV power plant, the rationality of the equivalence modeling method based on factor analysis and correlation clustering of the PV power plant has been verified. However, the short-circuit fault on the power grid side belongs to abnormal operation states. In the actual environment, it is more common for PV power plant to occur dynamic change process such as cloud generation and disappearance in a long-time scale at the plant side. In order to verify the adaptability of the proposed equivalent modeling method to the solar irradiance variation, this section set the conditions of solar irradiance variation due to cloud occlusion and simulated the output characteristics of equivalent model and detailed model of the PV power plant. In the case of initial operating conditions of PV power plant being unchanged, the same change rate of solar irradiance was set to make the solar irradiance of each PV unit change according to the curve shown in Figure 13 within 20 seconds.

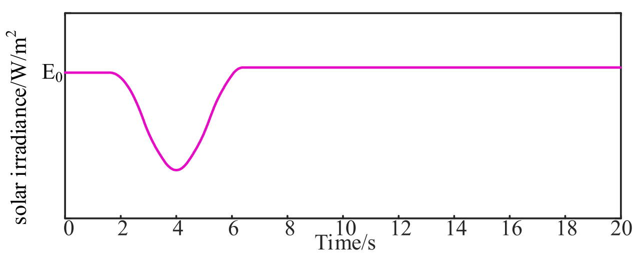


FIGURE 13. Solar irradiance curve.

Under the long-time scale condition of solar irradiance variation, the action state of chopper protection circuits and LVRT operation state were not involved, the PV inverters not outputting reactive current. Therefore, this section selected the time series representative values of solar irradiance, active power and output current of PV units as initial variables, using proposed method to carry out equivalent modeling. The obtained clustering results of PV units are as shown in Table 5.

TABLE 5. Clustering results of equivalent model with solar irradiance variation.

Equivalent model	The serial number of PV units
Cluster 1	1, 2, 7, 9, 14, 15, 25, 28
Cluster 2	3, 4, 11, 12, 13, 16, 19, 20, 21, 22, 23, 37, 38
Cluster 3	5, 10, 17, 18, 26, 27, 29, 30, 33, 34, 36
Cluster 4	6, 8, 24, 31, 32, 35, 39, 40

According to the clustering results in Table 5, the Equivalent Model E under the condition of solar irradiance variation was established. Compare the accuracy of Equivalent Model E and single- machine equivalent model in fitting the output characteristics of the detailed model under the same working conditions, as shown in Figure 14. The error evaluation index of the equivalent model was calculated with the sampling step of 0.001 s, as shown in Table 6.

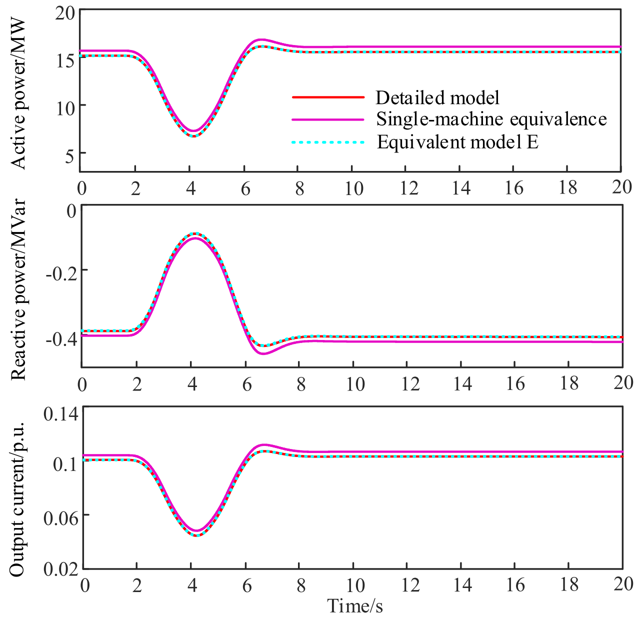


FIGURE 14. The effect comparison of equivalent models with solar irradiance variation.

TABLE 6. Error evaluation index of equivalent model with solar irradiance variation.

Error evaluation index	Active power	Reactive power	Output current
Single-machine equivalent model	0.01461	0.09483	0.01455
Equivalent model E	0.00060	0.00400	0.00059

According to Figure 14 and Table 6, it can be seen that the Equivalent Model E based on factor analysis and correlation clustering has better fitting effect and higher fitting accuracy than the single-machine equivalent model in describing the external characteristics of active power, reactive power and output current of the detailed model. Therefore, the proposed method in this paper also has good adaptability to solar irradiance variation.

F. THE COMPARISON OF SIMULATION SPEED

The above simulation verified the high accuracy of the equivalent model in fitting the output characteristics of the detailed model. The purpose of analyzing the interaction between PV power plant and system by using equivalent model instead of detailed model is to reduce the simulation speed and simulation cost and avoid dimension disaster. In order to verify the significant advantage of equivalent model in simulation speed, the actual simulation time of single-machine equivalent model, equivalent model based on proposed method and detailed model under three-phase short-circuit fault condition

TABLE 7. Comparison of simulation time between equivalent model and detailed model.

Simulation time of the model	three-phase short-circuit fault	solar irradiance variation
Detailed model	40.88s	131.56s
Single-machine equivalent model	7.02s	27.55s
Equivalent model based on proposed method	8.06s	31.42s

and solar irradiance variation condition was compared. 5 seconds simulation time was set for three-phase short-circuit fault condition to count the average actual simulation time for each voltage drop depth. 20 seconds simulation time was set for solar irradiance variation condition to count the actual simulation time. The comparative data are shown in Table 7.

From the data of Table 4, Table 6 and Table 7, it can be seen that compared with the detailed model, the equivalent model established by proposed method and single-machine equivalent model can greatly improve the simulation speed. Moreover, the equivalent model established by proposed method has higher fitting accuracy than single-machine equivalent model. With the increase of power system scale and observation scale, the advantages of simulation speed and accuracy will be expanded by replacing the detailed model with the equivalent model based on proposed method in this paper.

VII. CONCLUSION

In this paper, factor analysis was used to preprocess the data indexes and correlations were used to cluster the obtained indexes. The clustering results were analyzed by contour visualization method. The established equivalent model of PV power plant had high accuracy in fitting the output characteristics of the detailed model and the effectiveness and adaptability of the proposed method were verified by different working conditions. The conclusions are as follows:

1) Factor analysis can be used to reduce the dimension of data, which can eliminate the correlations between initial variables. At the same time, the rotating technology can be used to explain the factors, fully mining the intrinsic information of the data and effectively quantifying the operation states of PV units, which can eliminate the subjectivity of the initial variables. The obtained clustering indexes have clearer meanings and stronger guidance.

2) In this paper, the correlations obtained by correlation analysis and significance test can be used to cluster PV units, which has simple calculation process. Moreover, it can effectively distinguish the action state of chopper protection circuits, needless to monitor in real time, which solves the problem that the action signal of protection circuit is not easy to obtain in practical engineering.

3) Compared with the detailed model, the simulation speed of the equivalent model established by proposed method and the traditional single-machine equivalent model is both greatly improved. Moreover, the equivalent model estab-

lished by proposed method has higher fitting accuracy than the single-machine equivalent model. It will be more practical and popularized to use the established equivalent model for the analysis of interaction between PV power plant and system and the analysis of power system with large-scale new energy.

4) When the power system is confronted with uncertainties such as intermittent output of PV power plant and stochastic load demand response, etc., the data for establishing equivalent model need to be changed accordingly. In this case, the off-line database of detailed model simulation conditions can be established and the off-line data can be extracted for online equivalence when monitoring the corresponding actual conditions.

Based on the proposed method in this paper, establishing the equivalent model of PV power plant, researching typical scenarios such as large-scale PV units centralized access and high voltage AC/DC transmission, analyzing the impact of the PV power plant access on the security and stability of large power grid from power angle, frequency, voltage and other aspects of stability and proposing measures to improve the stability of grid-connected PV system will be the following research contents.

REFERENCES

[1] M. Ding, Z. Xu, W. Wang, X. Wang, Y. Song, and D. Chen, "A review on China's large-scale PV integration: Progress, challenges and recommendations," *Renew. Sustain. Energy Rev.*, vol. 53, pp. 639–652, Jan. 2016. doi: 10.1016/j.rser.2015.09.009.

[2] K. Jia, C. Gu, Z. Xuan, L. Li, and Y. Lin, "Fault characteristics analysis and line protection design within a large-scale photovoltaic power plant," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 4099–4108, Sep. 2018. doi: 10.1109/TSG.2017.2648879.

[3] Y. Yu and X. Hu, "Active disturbance rejection control strategy for grid-connected photovoltaic inverter based on virtual synchronous generator," *IEEE Access*, vol. 7, pp. 17328–17336, 2019. doi: 10.1109/ACCESS.2019.2894786.

[4] L. Wang *et al.*, "Electromagnetic transient modeling and simulation of power converters based on a piecewise generalized state space averaging method," *IEEE Access*, vol. 7, pp. 12241–12251, 2019. doi: 10.1109/CCESS.2019.2891122.

[5] X. Pan, Y. Zhang, P. Ju, Y. Jin, T. Liu, and P. Zeng, "Equivalent modeling for photovoltaic power station," *Power Syst. Technol.*, vol. 39, pp. 1173–1178, May 2015. doi: 10.13335/j.1000-3673.pst.2015.05.001.

[6] D. Remon, A. M. Cantarellas, and P. Rodriguez, "Equivalent model of large-scale synchronous photovoltaic power plants," *IEEE Trans. Ind. Appl.*, vol. 52, no. 6, pp. 5029–5040, Nov./Dec. 2016. doi: 10.1109/TIA.2016.2598718.

[7] A. Samadi, L. Söder, E. Shayesteh, and R. Eriksson, "Static equivalent of distribution grids with high penetration of PV systems," *IEEE Trans. Smart Grid*, vol. 6, no. 4, pp. 1763–1774, Jul. 2015. doi: 10.1109/TSG.2015.2399333.

[8] P. Han, Z. Lin, L, Wang, G. Fan, and X. Zhang, "A survey on equivalence modeling for large-scale photovoltaic power plants," *Energies*, vol. 11, p. 1463, Jun. 2018. doi: 10.3390/en11061463.

[9] H. Li, S. L. Pang, and J. D. Huang, "Equivalent modeling and simulation of large photovoltaic station," *Appl. Mech. Mater.*, vol. 615, pp. 27–30, Aug. 2014. doi: 10.4028/www.scientific.net/AMM.615.27.

[10] Y. Guo, H. Gao, and J. Tian, "Photovoltaic output modeling by introducing clustering analysis and its application in reliability evaluation," *Autom. Electr. Power Syst.*, vol. 40, pp. 93–100, Dec. 2016. doi: 10.7500/AEPS20160301008.

[11] E. F. Fernández and F. Almonacid, "A new procedure for estimating the cell temperature of a high concentrator photovoltaic grid connected system based on atmospheric parameters," *Energy Convers. Manage.*, vol. 103, pp. 1031–1039, Oct. 2015. doi: 10.1016/j.enconman.2015.07.034.

[12] Z. Ma, J. Zheng, S. Zhu, X. Shen, L. Wei, and X. Wang, "Online clustering modeling of large-scale photovoltaic power plants," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, Denver, CO, USA, Jul. 2015, pp. 1–5.

[13] W. Sheng, Y. Ji, M. Wu, H. Liu, L. Kou, and L. Sun, "Dynamic clustering modeling of regional centralized photovoltaic power plant based on improved fuzzy C-means clustering algorithm," *Power Syst. Technol.*, vol. 41, pp. 3284–3291, Oct. 2017. doi: 10.13335/j.1000-3673.pst.2017.1861.

[14] W. Li, P. Chao, X. Liang, D. Xu, and X. Jin, "An improved single-machine equivalent method of wind power plants by calibrating power recovery behaviors," *IEEE Trans. Power Syst.*, vol. 33, no. 4, pp. 4371–4381, Jul. 2018. doi: 10.1109/TPWRS.2017.2771323.

[15] J. Zheng, Z. Ma, Z. Wang, X. Wang, S. Zhu, and L. Wei, "Feature distance based online cluster modeling of LVRT controlled PV power plants," *Electr. Power Syst. Res.*, vol. 154, pp. 223–233, Jan. 2018. doi: 10.1016/j.epsr.2017.08.028.

[16] P. Wang, Z. Zhang, Q. Huang, N. Wang, X. Zhang, and W.-J. Lee, "Improved wind farm aggregated modeling method for large-scale power system stability studies," *IEEE Trans. Power Syst.*, vol. 33, no. 6, pp. 6332–6342, Nov. 2018. doi: 10.1109/TPWRS.2018.2828411.

[17] P. Li, W. Gu, L. Wang, B. Xu, M. Wu, and W. Shen, "Dynamic equivalent modeling of two-staged photovoltaic power station clusters based on dynamic affinity propagation clustering algorithm," *Int. J. Elect. Power Energy Syst.*, vol. 95, pp. 463–475, Feb. 2018. doi: 10.1016/j.ijepes.2017.08.038.

[18] R. Fang, R. Shang, M. Wu, C. Peng, and X. Guo, "Application of gray relational analysis to k-means clustering for dynamic equivalent modeling of wind farm," *Int. J. Hydrogen Energy*, vol. 42, pp. 20154–20163, Aug. 2017. doi: 10.1016/j.ijhydene.2017.06.023.

[19] H. Pingping, X. Yu, Z. Yan, and L. Kui, "Equivalent model of wind farm based on DBSCAN," in *Proc. IEEE Innov. Smart Grid Technol.-Asia (ISGT)*, Auckland, New Zealand, Dec. 2017, pp. 1–6.

[20] J. Zou, C. Peng, H. Xu, and Y. Yan, "A fuzzy clustering algorithm-based dynamic equivalent modeling method for wind farm with DFIG," *IEEE Trans. Energy Convers.*, vol. 30, no. 4, pp. 1329–1337, Dec. 2015. doi: 10.1109/TEC.2015.2431258.

[21] M. Ali, I.-S. Ilie, J. V. Milanovic, and G. Chicco, "Wind farm model aggregation using probabilistic clustering," *IEEE Trans. Power Syst.*, vol. 28, no. 1, pp. 309–316, Jul. 2012. doi: 10.1109/TPWRS.2012.2204282.

[22] J. Ding, Q. Zhang, S. Hu, Q. Wang, and Q. Ye, "Clusters partition and zonal voltage regulation for distribution networks with high penetration of PVs," *IET Gener., Transmiss. Distrib.*, vol. 12, pp. 6041–6051, Dec. 2018. doi: 10.1049/iet-gtd.2018.6255.

[23] Q. Zhu, M. Ding, and P. Han, "Equivalent modeling of DFIG-based wind power plant considering crowbar protection," *Math. Problems Eng.*, vol. 2016, Jun. 2016, Art. no. 8426492. doi: 10.1155/2016/8426492.

[24] W. Li, P. Chao, X. Liang, J. Ma, D. Xu, and X. Jin, "A practical equivalent method for DFIG wind farms," *IEEE Trans. Sustain. Energy*, vol. 9, no. 2, pp. 610–620, Apr. 2017. doi: 10.1109/TSTE.2017.2749761.

[25] D.-E. Kim and M. A. El-Sharkawi, "Dynamic equivalent model of wind power plant using an aggregation technique," *IEEE Trans. Energy Convers.*, vol. 30, no. 4, pp. 1639–1649, Dec. 2015. doi: 10.1109/TEC.2015.2470531.

[26] P. Chen, S. Tao, X. Xiao, L. Li, and J. Zhang, "Network model for correlation analysis of short-term electricity consumption behavior," *Automat. Electr. Power Syst.*, vol. 41, no. 3, pp. 61–69. doi: 10.7500/AEPS20160127014.

[27] O. Noureldeen and I. Hamdan, "Design and analysis of combined chopper-crowbar protection scheme for wind power system based on artificial intelligence," in *Proc. 20th Int. Middle East Power Syst. Conf. (MEPCON)*, Cairo, Egypt, Dec. 2018, pp. 809–814.

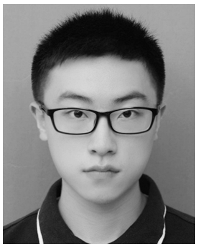
[28] D. Li, Z. Ren, W. Yan, J. Zhu, X. Zhao, and J. Yu, "Month-ahead wind power curve probabilistic prediction based on factor analysis and quantile regression neural network," *Proc. Chin. Soc. Elect. Eng.*, vol. 37, no. 18, pp. 5238–5247, Sep. 2017. doi: 10.13334/j.0258-8013.pcsee.161368.

[29] J. Zou, C. Peng, Y. Yan, H. Zheng, and Y. Li, "A survey of dynamic equivalent modeling for wind farm," *Renew. Sustain. Energy Rev.*, vol. 40, pp. 956–963, Dec. 2014. doi: 10.1016/j.rser.2014.07.157.

[30] H. Wang, Y. Tang, J. Hou, J. Zou, S. Liang, and F. Su, "Composition modeling and equivalence of an integrated power generation system of wind, photovoltaic and energy storage unit," *Proc. CSEE*, vol. 31, pp. 1–9, Dec. 2011.

PINGPING HAN was born in 1981. She received the Ph.D. degree in power system and its automation from the Hefei University of Technology, Hefei, China, in 2008, where she is currently an Associate Professor.

Her research interests are power system analysis, technologies for renewable energy and distributed power generation systems, and applications of power electronics in power systems.



ZIHAO LIN received the B.E. degree in electrical engineering from the Hefei University of Technology, Hefei, China, in 2017, where he is currently pursuing the M.E. degree, and also the master's degree with the Anhui Province Laboratory of New Energy Utilization and Energy Conservation.

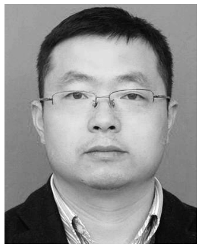
His research interests are modeling and simulation of renewable energy connecting to the power grid.

JINGJING ZHANG was born in 1977. She received the Ph.D. degree in power system and its automation from the Hefei University of Technology, in 2012, where she is currently an Associate Professor with the School of Electrical Engineering and Automation.

Her research interests include power system reliability, protection relaying, and renewable energy and its applications.

YU XIA received the B.E. degree in power system and its automation from Three Gorges University, Yichang, China, in 2016. She is currently pursuing the master's degree with the Anhui Province Laboratory of New Energy Utilization and Energy Conservation, Hefei University of Technology.

Her research interests are control strategies of wind power grid-connected systems and simulation of renewable energy in power systems.



LEI WANG received the B.E. and M.E. degrees in electrical engineering from the Hefei University of Technology, Hefei, China, in 2000 and 2003, respectively, and the Ph.D. degree from the Hefei Institutes of Physical Science, CAS, Hefei, China, in 2010.

He is currently an Associate Professor with the School of Electrical Engineering and Automation, Hefei University of Technology. His research interests relate to renewable energy and its application, and simulation and control of power transmission.