IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 20, 2019, accepted July 22, 2019, date of publication July 26, 2019, date of current version August 12, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2931444

Equivalent Modeling of Photovoltaic Power Station Based on Canopy-FCM Clustering Algorithm

HONGBIN WU[1], (Member, IEEE), JIAJIA ZHANG[1], CHEN LUO[1], AND BIN XU[2]
[1]School of Electrical Engineering and Automation, Hefei University of Technology, Hefei 230009, China
[2]State Grid Anhui Electric Power Research Institute, Hefei 230601, China

Corresponding author: Hongbin Wu (hfwuhongbin@163.com)

This work was supported by the National Key Research and Development Program of China under Grant 2016YFB0900400.

ABSTRACT Grid-connected photovoltaic power generation has become an effective way to utilize solar energy. In order to accurately analyze the dynamic characteristics of a grid-connected photovoltaic power station, an equivalent modeling method based on the Canopy-FCM clustering algorithm is proposed. Considering the lack of theoretical basis for the selection of the traditional clustering index, this paper deduces the transfer function of the grid-connected inverter control system based on the detailed model and proposes a new clustering index, which takes the zero-pole expression of the closed-loop transfer function as the prototype. In order to improve the fuzzy C-means clustering method, which is sensitive to the initial clustering centers and the outliers, a Canopy algorithm with lower computation cost is introduced. Before fuzzy C-means clustering analysis, the Canopy algorithm is used to determine the initial clustering centers and the number of clusters to improve the accuracy and efficiency of the clustering algorithm. The effectiveness of the proposed method is verified by simulation examples.

INDEX TERMS Grid-connected photovoltaic station, improved fuzzy C-means, clustering and equivalence, canopy algorithm.

I. INTRODUCTION

With the gradual depletion of global fossil energy, the use of renewable energy has become an inevitable choice for people to survive and develop. In 1993, Germany built the world's first MW-scale grid-connected photovoltaic (PV) power station. Since then, the construction of large-scale grid-connected PV power stations has become a trend in the development of PV industry in various countries [1]. However, unlike traditional power generation methods, PV system achieves power output through power electronic devices with fast response characteristics. Its dynamic characteristics depend entirely on the performance of the inverter and its control system. With the increase of its permeability in power grid, the dynamic characteristics of power grid will change [2]–[4], and then it will affect the safe and stable operation of power grid and the configuration of protection devices after faults. Therefore, it is of great significance to accurately model the grid-connected PV power station and simulate its dynamic performance.

The dynamic characteristics of a PV power station can be accurately simulated by establishing a detailed device-level model [5]–[7]. However, the station contains a large number of power electronic devices. Building detailed models for them will not only increase the workload but also increase simulation time and reduce simulation efficiency. To this end, scholars have carried out a lot of research on the equivalent modeling and simulation of grid-connected PV power stations. Clustering provides a reference and basis for equivalent modeling. Reference [8] obtained a two-machine equivalent model of the PV power plant under the same conditions of all PV cells working mode and load distribution. Assuming that the types of inverters in the PV power station are the same, the equivalent model is established in references [9]–[10]. The above assumptions are too ideal. In view of the fact that there are many kinds of inverters and the load distribution is uneven in PV power plants, reference [11] takes the control parameters and filtering parameters of grid-connected

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

inverters as clustering indicators. By calculating the distance between them, PV units with similar distances are merged into an equivalent machine. In references [12], the clustering index considering sensitivity coefficient is used to cluster PV power plants based on K-means clustering algorithm, which improves the clustering accuracy. In addition, the dynamic performance of PV power plants can also be affected by special devices. In reference [13], considering the dynamic characteristics of a synchronous power control device (SPC) in inverters, a method to obtain an equivalent model of a large-scale PV power plant is presented. Reference [14] estimates the equivalent resistance and reactance between the inverters and the grid based on the model reference adaptive control (MRAC) method of active and reactive power, which can be used as a reference for the equivalent modeling of PV power plants after clustering.

For most PV power plants, the dynamic characteristics are mainly determined by the inverters and their control links [15]. As a matter of fact, the control parameters and filter parameters of the inverters play a major role in the dynamic characteristics of PV power plants. By adjusting and adapting parameters, excellent external characteristics of PV power plants can be realized. When clustering PV power plants, PV units with smaller Euclidean distances of clustering parameters are often grouped into one category. This clustering method is too rough, which will lead to inaccurate clustering results. Some scholars use more complex K-means and FCM algorithms for clustering to improve the accuracy of clustering. However, the random selection of the initial clustering centers and the number of clusters may lead to the algorithm falling into local optimum, which requires repeated validation of the clustering results and increases the computational complexity. In order to reduce the simulation scale without losing accuracy, it is urgent to explore clustering parameters that can more accurately characterize the dynamic performance of PV power plants. At the same time, in order to improve the shortcomings of traditional clustering algorithm, it is necessary to find a more scientific clustering method.

In this paper, the modeling method and control strategy principle of a grid-connected PV system are deeply analyzed. Through deriving the transfer function of control system, the clustering parameters of the grid-connected PV system are determined. Before clustering analysis, the Canopy algorithm is used to preprocess the data to get the initial clustering centers and the number of clusters. On this basis, the clustering results can be obtained by combining the traditional FCM clustering algorithm. After that, the ''coherent equivalence method'' is used to model the PV units clustered into one category [16], [17]. Finally, through simulation examples, the dynamic characteristics of equivalent models and the detailed model under different fault conditions are compared, which verifies the correctness of the equivalent modeling method adopted in this paper for large-scale PV power stations.

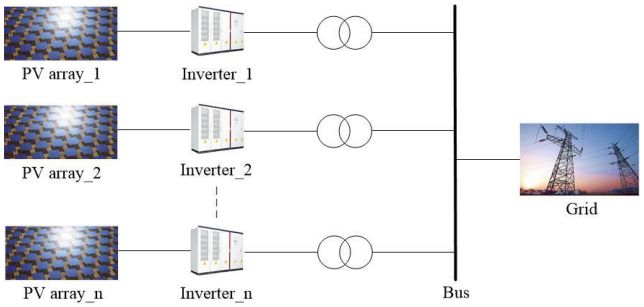


FIGURE 1. Structure of a grid-connected PV power station.

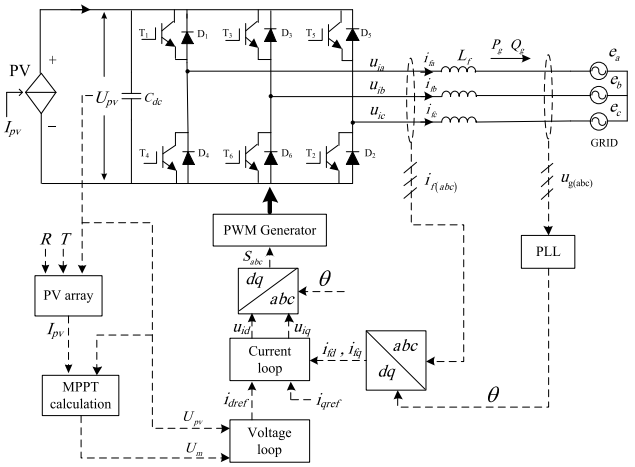


FIGURE 2. Topology of inverter control strategy.

II. CLUSTERING INDEX OF GRID-CONNECTED PV

A. MATHEMATICAL MODEL OF GRID-CONNECTED PV

The construction of a large-scale grid-connected PV power station is an effective method to use solar energy for power generation. There are many devices in a high-capacity PV power station, and the performance of the components is different. The structure of a typical grid-connected PV power station is shown in Fig. 1.

In order to realize the effective utilization of solar energy, the grid-connected inverters often use a MPPT-unit power factor control strategy to achieve the maximum output of active power. To facilitate analysis, the mathematical model in a three-phase static *abc* coordinate system is transformed into the model on a synchronously rotating *dq*-rame [18]–[21]. The typical control strategy is shown in Fig. 2.

In Fig. 2, R and T are illumination intensity and temperature, respectively. The output filter adopts single inductance filter [22], L_f is the filter inductance, i_f is the current flowing through the filter inductor, u_g is the grid side voltage, u_i is the voltage on the AC side of the inverter, and r is the equivalent resistance representing the internal resistance of the filter inductor and the influence of the inverter switch.

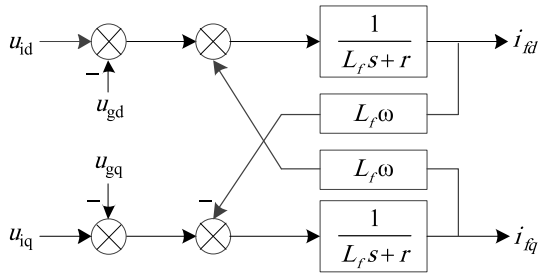


FIGURE 3. Grid-connected PV system model.

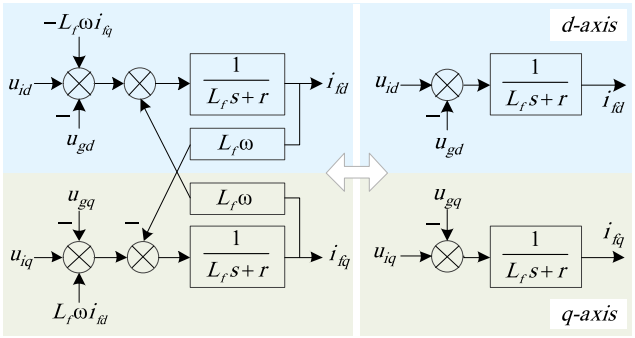


FIGURE 4. Decoupled system model.

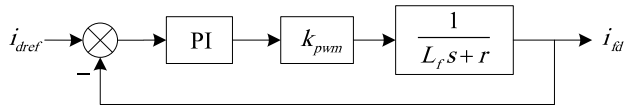


FIGURE 5. Control structure of current loop.

The output voltage equation of the inverter in the frequency domain can be written as follows:

$$\begin{cases} u_{\mathrm{id}} = (L_f s + r) i_{fd} - L_f \omega i_{fq} + u_{\mathrm{gd}} \\ u_{\mathrm{iq}} = (L_f s + r) i_{fq} + L_f \omega i_{fd} + u_{\mathrm{gq}} \end{cases} \tag{1}$$

The model structure corresponding to (1) is shown in Fig. 3.

In Fig. 3, the system model is coupled on the d-axis and the q-axis. In order to facilitate the design of the control system, the feed-forward decoupling strategy is adopted to weaken the coupling. In the AC side of the inverter,$-L_f \omega i_{fq}$ and $L_f \omega i_{fd}$ are added into the d-axis and the q-axis, respectively, to offset the coupling term. The decoupled system model is shown in Fig. 4, which is transformed into two independent and equivalent parts.

Since the inverter switching frequency is much larger than the grid frequency, the amplification characteristic of the inverter can be expressed by the proportional gain k_{pwm} [23].

The feed-forward compensation of the grid voltage is adopted to eliminate the adverse effect of the grid voltage on the grid-connected current control. The d-axis control structure of the current inner loop after decoupling is shown in Fig. 5.

The transfer function of the PI regulator is as follows:

$$G_I(s) = k_p + \frac{k_i}{s} \tag{2}$$

Therefore, the closed-loop transfer function of the inner current loop is as follows:

$$G_{cl}(s) = \frac{k_{pwm}\left(k_i + k_p s\right)}{L_f s^2 + \left(r + k_{pwm} k_p\right) s + k_{pwm} k_i} \tag{3}$$

The value of k_{pwm} can be determined according to the actual situation. Since the closed-loop transfer function of the current loop is consistent on the d-axis and the q-axis, this paper only analyses the dynamic characteristics of the d-axis current. The PI parameters of the q-axis are the same as those of the d-axis. The equivalent resistance r is relatively small and neglected.

B. SELECTION METHOD OF CLUSTERING INDEX

Completing clustering of different PV units is the premise of equivalent modeling of PV power plants. The control parameters of grid-connected inverters can reflect the dynamic characteristics of the PV system to a certain extent. However, the exact influence degree of each control parameter and filter parameter of the inverters is unknown. The clustering method which only takes the control parameters as the clustering indexes is not rigorous. There are also simulation error and artificial judgment error when the sensitivity coefficient and weight coefficient of each parameter are given to distinguish the influence degree of different parameters on the dynamic performance by the simulation method.The transfer function of the system is the most intuitive mathematical expression of its dynamic characteristics. This paper establishes the equivalent clustering index by deriving the transfer function of a grid-connected PV system, and the validity of the equivalent clustering index is verified by simulation comparison.

The expressions of zero and pole points of the d-axis closed-loop transfer function obtained by (3) are as follows:

$$\begin{cases} z = -k_i / k_p \\ p_{1,2} = -k_p/(2L_f) \pm \sqrt{(k_p^2/(4L_f^2)) - (k_i/L_f)} \end{cases} \tag{4}$$

After randomly changing the values of k_p, k_i and L_f, the distribution of zero and pole points is shown in Table 1.

In Table 1, groups 2, 3 and 4 differ only from group 1 in terms of k_p, k_i and L_f, respectively, and the other parameters are identical, but the distribution of zero and pole points in the system differs greatly. The values of k_p, k_i and L_f in group 5 differ greatly from those in group 1, but their zero and pole distributions are almost the same. According to the transfer function represented by five sets of data, the unit step response curve of the corresponding system is drawn, as shown in Fig. 6. Corresponding to the zero-pole distribution of the system, the system with similar zero and pole points has similar dynamic response curves.

The above analysis shows that the zero and pole points of the closed-loop transfer function of the system are closely

TABLE 1. Distribution of zero and pole points.

N	k_p	k_i	L_f	P_1	P_2	z
1	0.5	250	0.01	-25+156i	-25-156i	-500
2	1.1	250	0.01	-55+148i	-55-148i	-227
3	0.5	300	0.01	-25+171i	-25-171i	-600
4	0.5	250	0.02	-13+111i	-13-111i	-500
5	1.1	500	0.02	-28+156i	-28-156i	-455

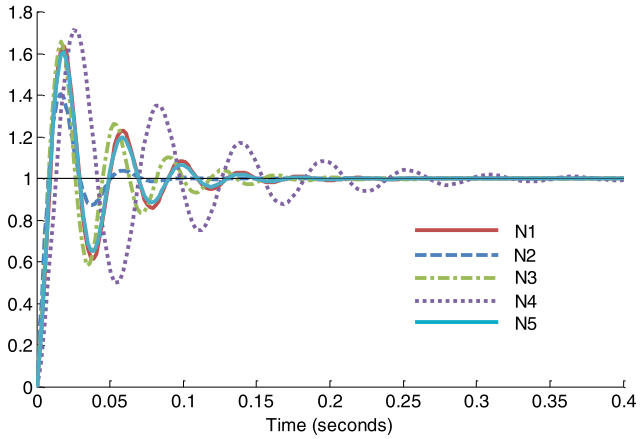


FIGURE 6. System dynamic response curve.

related to its dynamic characteristics. In this paper, the zero-pole expression of closed-loop transfer function is used as the clustering index to cluster the grid-connected PV system. For the above description system, the clustering index vector is a three-dimensional vector, that is, $X = [x_1, x_2, x_3]$, where

$$\begin{cases} x_1 = z = -k_i/k_p \\ x_2 = p_1 = -k_p/(2L_f) + \sqrt{(k_p^2/(4L_f^2)) - (k_i/L_f)} \\ x_3 = p_2 = -k_p/(2L_f) - \sqrt{(k_p^2/(4L_f^2)) - (k_i/L_f)} \end{cases} \quad (5)$$

It should be pointed out that for the system with dipoles in the zero and pole points of the closed-loop transfer function, only the influence of the dominant poles on the dynamic characteristics of the system is considered. For simplified analysis, dipoles are not involved in this paper.

III. CLUSTERING AND EQUIVALENT METHOD

A. TRADITIONAL FCM ALGORITHM

The fuzzy C-means algorithm is a flexible clustering method. It takes the sum of Euclidean distances between sample points and clustering centers as the objective function and establishes an iterative equation that satisfies the constraint condition. Finally, the sample attribution is determined according to the membership degree matrix.

Suppose there are n samples $[x_1, x_2, \ldots, x_n]$, and the dimension of each sample space is s. In the traditional FCM algorithm, we first determine c original clustering centers $[v_1, v_2, \ldots, v_n], (c < n)$, divide n sample vectors into c groups randomly, and set the convergence precision ε, constructing the objective function shown as follows:

$$J(U, V) = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^m d_{ij}^2$$
$$\text{s.t.} \sum_{j=1}^{c} u_{ij} = 1, 0 < u_{ij} < 1 \quad (6)$$

where u_{ij} represents the membership value of the jth sample relative to the ith clustering center. m is the fuzzy coefficient, which usually takes 2. $d_{ij} = ||x_j - v_i||$ denotes the Euclidean distance from the jth sample to the ith clustering center.

The Lagrange multiplier method is used to obtain the extreme value of (6), and the iteration formulas of membership degree and clustering centers are shown in (7) and (8), respectively.

$$v_{i(k+1)} = \sum_{j=1}^{n} u_{ij(k)}^m x_j / \sum_{j=1}^{n} u_{ij(k)}^m, i = 1, 2, \cdots c \quad (7)$$

$$u_{ij(k+1)} = \left(\sum_{h=1}^{c} (d_{ij(k)}/d_{hj(k)})^{2/(m-1)} \right)^{-1} \quad (8)$$

Iterating over the value of (7) and (8), until the iteration error satisfy $\left\| v_{(k+1)} - v_{(k)} \right\| \leq \varepsilon$ (k is the number of iterations), stops iteration and outputs the final clustering centers and membership matrix. The clustering center corresponding to the maximum membership value of each sample is taken as the final attribution category of the sample, and then the clustering results are obtained.

B. CANOPY-FCM ALGORITHM

From the above introduction, it can be seen that the initial clustering centers of the traditional FCM algorithm are generated randomly. If the initial clustering centers are not selected reasonably, the algorithm can easily fall into the local optimal solution. In addition, the clustering number of FCM is determined artificially.

In Fig. 7, the number of clusters is artificially determined to be 2, and all data are classified into two categories. The distance between point K and two clustering centers is d_1 and d_2, respectively, and $d_1 < d_2$. Point K is classified as category A data, but there is a big gap between them (point K is called

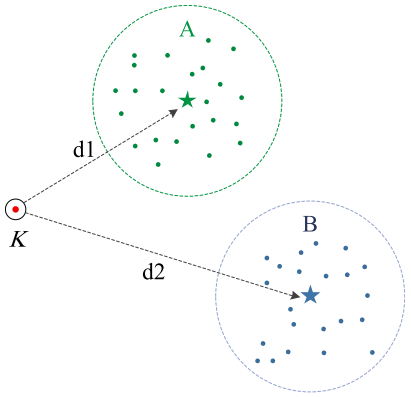


FIGURE 7. FCM clustering diagram.

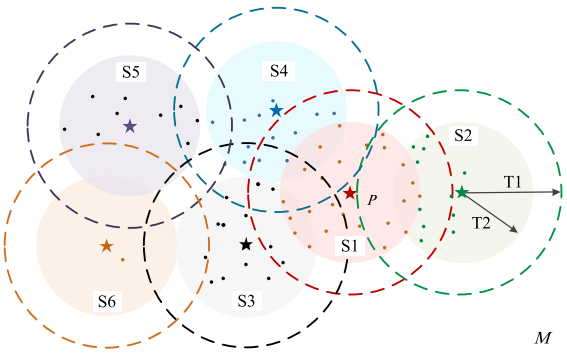


FIGURE 8. Initial clustering schematic.

outlier point). It is obviously unreasonable to determine the number of clusters artificially.

In order to improve the clustering accuracy in traditional FCM algorithm, it is usually necessary to evaluate the clustering effect under different clustering numbers in order to obtain the optimal number of clusters. This undoubtedly increases the amount of calculation.

To avoid the defects of the traditional FCM algorithm, this paper obtains the initial clustering centers and the number of clusters by the Canopy algorithm.

In the first stage, Canopy clustering chooses a simple and inexpensive method to calculate object similarity and puts similar objects in a subset, as shown in Fig. 8:

The concrete steps of the Canopy algorithm are as follows:

Step 1: Obtain the original cluster data set M. Determine two distance thresholds T_1 and T_2 such that $T_1 > T_2$. The values of T_1 and T_2 are determined by a cross-check method;

Step 2: Randomly select an unclassified data point P from M as a clustering center. Calculate the Euclidean distance d from other data points in M to point P;

Step 3: If $d \leq T_2$, classify the data point and point P into one class and delete the data point from M. If $T_1 \leq d \leq T_2$, classify the data point and point P into one class and keep this data point;

Step 4: Repeat step 2 and step 3 until set M is empty.

The data are preprocessed by the Canopy algorithm and then clustered by FCM. This avoids blind selection of initial clustering centers and number of clusters and greatly reduces the computational complexity.

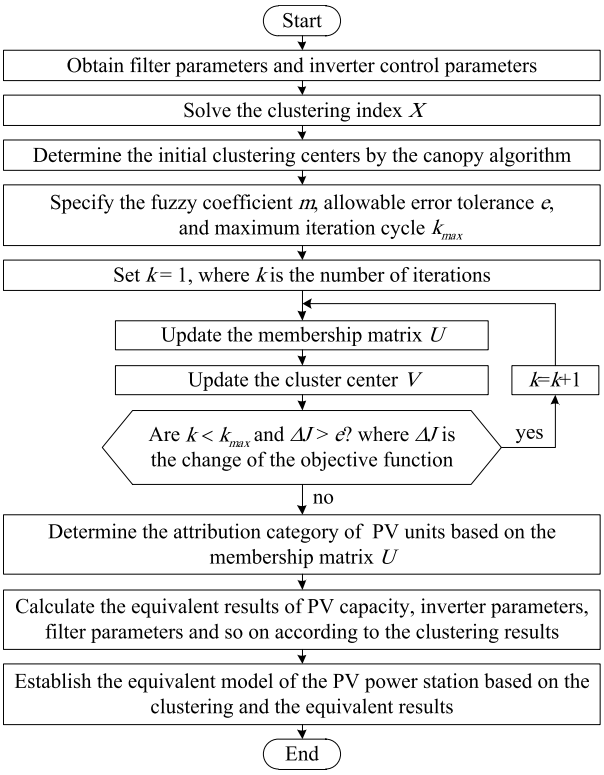


FIGURE 9. Flow chart of clustering and equivalence.

C. CLUSTERING AND EQUIVALENT PROCESS OF PV SYSTEM

After the clustering of the PV system is completed, the PV units clustered into one group need to be equivalent. The method of calculating the parameters of the system after the equivalence is described in reference [24].

The process of clustering and equivalence is shown in Fig. 9.

After clustering and equivalence are completed, the equivalent model of the PV power station is established. We need to obtain the dynamic response characteristic curve of the equivalent model and compare it with the original model. In order to quantify the fitting error of the dynamic response curve, the error quantization index is established, shown as follows:

$$E = \sqrt{\frac{1}{n}\sum_{i=1}^{n}((x_i - x_i^*)/x_i^*)^2} \times 100\% \tag{9}$$

where n is the number of sample points, x_i and x_i^* are the values at the ith sample point on the equivalent model and the original model dynamic response curve, respectively.

IV. ANALYSIS EXAMPLE

A. INTRODUCTION OF THE EXAMPLE

In order to verify the above clustering and equivalent methods, this paper builds a detailed model of a 2 MW PV power

TABLE 2. Parameters of PV units.

PV number	k_p	k_i	L_f	x_1	x_2	x_3
1	0.3	50	0.01	-167	-15+69i	-15-69i
2	0.64	64	0.005	-100	-64+93i	-64-93i
3	0.43	70	0.015	-163	-14+67i	-14-67i
4	0.9	65	0.005	-72	-90+70i	-90-70i
5	2.3	160	0.002	-69	-74	-1075
6	0.34	30	0.004	-88	-42+75i	-42-75i
7	0.33	60	0.01	-182	-17+76i	-17-76i
8	0.20	42	0.01	-210	-10+64i	-10-64i
9	0.25	65	0.002	-260	-63+169i	-63-169i
10	0.21	55	0.015	-262	-7+60i	-7-60i
11	0.8	61	0.005	-76	-80+76i	-80-76i
12	0.63	38	0.008	-60	-39+57i	-39-57i
13	0.62	54	0.004	-87	-78+87i	-78-87i
14	0.9	50	0.005	-56	-90+44i	-90-44i
15	0.43	63	0.01	-147	-22+76i	-22-76i

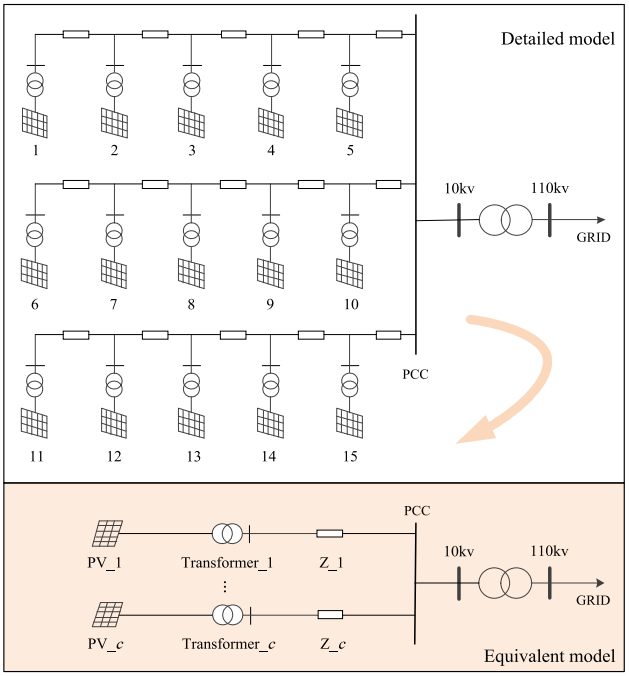


FIGURE 10. Simulation model of a grid-connected PV power station.

station on PSCAD/EMTDC platform. The PV power station has 15 PV arrays with a capacity of 130 kW. The PV arrays are connected to a PCC point by inverters, transformers and cables, sent to the upper 110 kV substation by single-circuit overhead lines, and then connected to the power grid by overhead lines. The photovoltaic grid-connected inverters adopt the MPPT-unit power factor control strategy described in this paper. After clustering, the multi-machine equivalent model of the PV power station is established, as shown in Fig. 10. The number c of equivalents varies according to the clustering index and clustering method.

The basic parameters and equivalent clustering index of each PV unit are shown in Table 2.

B. CLUSTERING RESULT

In order to independently verify the correctness of the clustering index and clustering method proposed in this paper,

TABLE 3. Clustering result.

	Method 1	Method 2	Method 3	Method 4
Category 1	1,3,6,7,8 9,10	1,3,6,7,8 9,10	1,3,6,7,8 10,12,15	1,3,6,7,8 10,12,15
Category 2	2,4,5,11,12 13,14,15	2,4,11,12,13 14,15	2,4,5,9,11 13,14	2,4,9,11,13 14
Category 3	-	5	-	5
Run times	15	1	12	1

the following four methods are used to cluster the above example system.

Method 1: The clustering index is selected as the inverter control parameters k_p, k_i and the filtering parameter L_f. The clustering index needs to be normalized. The traditional FCM algorithm is used in the clustering algorithm. The number of clusters is set artificially at 2, and the initial clustering centers are selected randomly.

Method 2: The clustering index is selected as the inverter control parameters k_p, k_i and the filtering parameter L_f. The clustering index needs to be normalized. The Canopy-FCM algorithm is selected as the clustering algorithm.

Method 3: The clustering index is selected as the equivalent clustering index based on zero and pole points of system closed-loop transfer function, and the clustering algorithm adopts the traditional FCM algorithm. The number of clusters is set artificially at 2.

Method 4: The clustering index is selected as the equivalent clustering index based on zero and pole points of system closed-loop transfer function, and the clustering algorithm adopts the Canopy-FCM algorithm.

In order to obtain relatively reliable results, when using the FCM algorithm for clustering, this paper runs the program multiple times until the results of the two runs are the same. The clustering result is shown in Table 3.

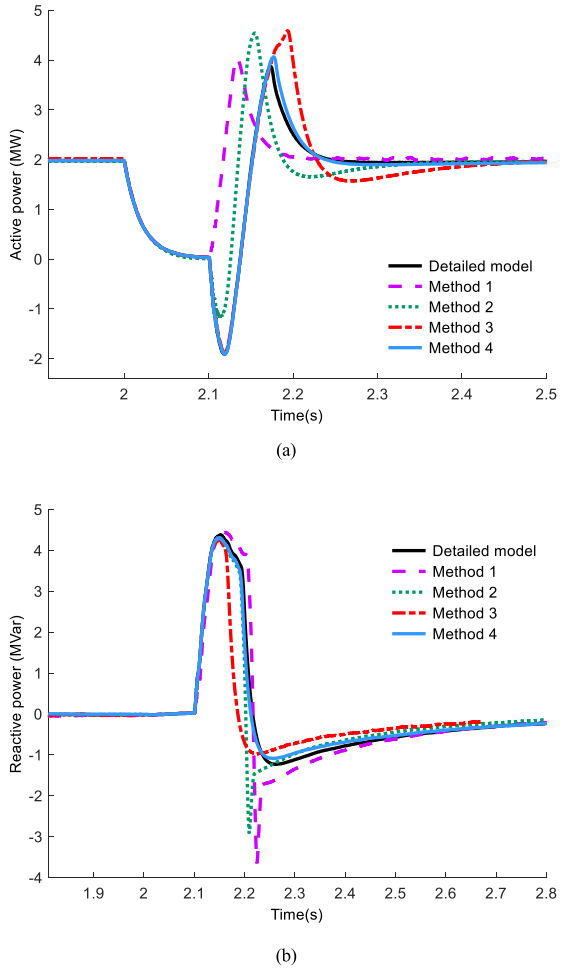


FIGURE 11. Power curves: (a) Active power. (b) Reactive power.

Regardless of which clustering index is used, the Canopy-FCM algorithm judges the fifth group of data as outliers, which are clustered into a single group, and the whole PV system is clustered into three groups. In addition, it can be seen from Table 3 that the random selection of initial clustering centers leads to a large increase in the number of program runs.

C. SIMULATION EXAMPLE

According to the clustering results, the equivalent models of the PV power station corresponding to the four clustering methods are established in PSCAD. For comparison, a detailed model is built. The dynamic characteristics of the equivalent models are compared and analyzed under the following disturbance conditions.

1) THREE-PHASE SHORT CIRCUIT

The system operates normally. At 2 second, a three-phase short circuit occurs at the PCC point, and the short circuit is removed after 0.1 seconds. The change of active and reactive power of the system is shown in Fig. 11.

TABLE 4. Error and running time in three-phase short circuit.

Clustering method	Active power error	Reactive power error	Running time
Method 1	1.018	0.748	42.1s
Method 2	0.714	0.689	29.5s
Method 3	0.109	0.251	38.4s
Method 4	0.022	0.137	28.6s
Detailed model	-	-	954.7s

In Fig. 11, the power change curve of the equivalent model based on the equivalent clustering index and Canopy-FCM clustering algorithm is closer to that of the original model in the case of the three-phase short circuit.

From 2 second to 3 second of simulation operation, 50 points are sampled equidistantly. The active power and reactive power error between the equivalent models and the detailed model are calculated by (9). In order to clearly demonstrate the superiority of the proposed algorithm in efficiency, the recorded system running time includes clustering time and simulation time. For the clustering program using the FCM algorithm, the clustering time is the sum of the multiple running time. The results are shown in Table 4.

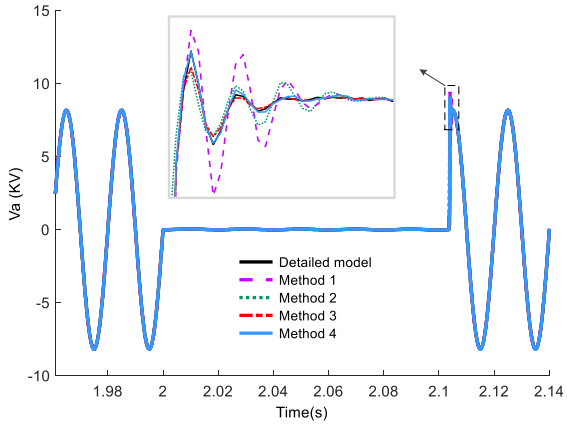
The error of the equivalent model in Method 2 is smaller than that in Method 1, indicating that the Canopy-FCM clustering method has better clustering effect. The error of the equivalent model in Method 3 are smaller than that in Method 1, which shows that the equivalent clustering index is more effective than the original clustering index. Method 4 combines the equivalent clustering index and the Canopy-FCM clustering method, and the error of the established model is relatively smaller. In addition, it can be seen that the model after clustering and equivalence can greatly reduce the simulation time. In the overall running time, the Canopy-FCM algorithm has advantages over the FCM algorithm. It further demonstrates that the clustering index and clustering method proposed in this paper can reduce the clustering error to a certain extent and improve the accuracy of the equivalent model while improving simulation efficiency.

2) SINGLE-PHASE SHORT CIRCUIT

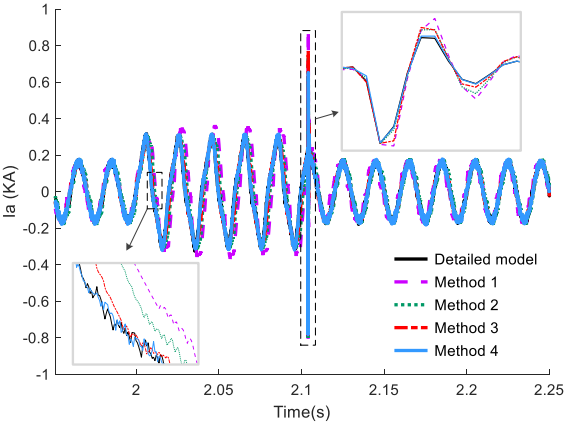
In order to verify the validity of the model under asymmetric faults, a single-phase short-circuit fault disturbance is added to the system. The system operates normally. At 2 second, a single-phase short circuit (phase A) occurs at the PCC point, and the fault is removed after 0.1 seconds. The voltage and current of phase A are shown in Fig. 12.

Similar to the three-phase faults, the voltage and current curves of phase A of the equivalent model established by Method 4 are more consistent with the original model in the case of single-phase fault. The simulation error and running time are shown in Table 5.

The results show again the superiority of the new clustering index and clustering method in simulation accuracy and running time.



(a)



(b)

FIGURE 12. Voltage and current curves: (a) Voltage of phase A. (b) Current of phase A.

TABLE 5. Error and running time in single-phase short circuit.

Clustering method	Voltage error	Current error	Running time
Method 1	5.771	7.452	40.3s
Method 2	2.421	3.234	20.4s
Method 3	0.763	0.943	38.1s
Method 4	0.454	0.365	21.1s
Detailed model	-	-	897.2s

3) VOLTAGE DISTURBANCE

In order to analyze the validity of the clustering index and clustering method proposed in this paper under small disturbance, voltage disturbance is applied to the stable operation system. At 2 second of normal system operation, the voltage of the PCC point drops from 10 kV to 9 kV and returns to normal after 0.3 seconds. The change of active power output of the PV system is shown in Fig. 13.

From 1.9 second to 2.5 second of simulation operation, 50 points are sampled equidistantly. The calculated error and running time are shown in Table 6.

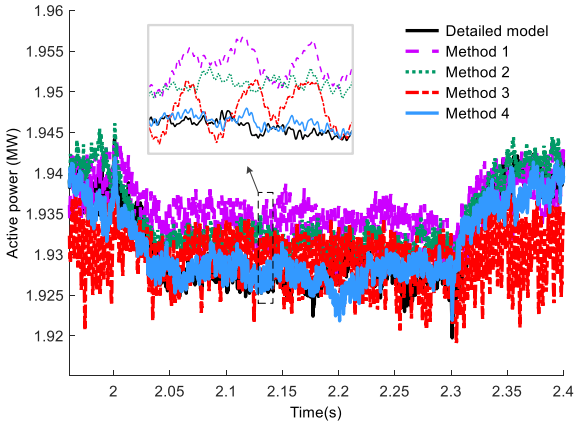


FIGURE 13. Active power curve.

TABLE 6. Error and running time in voltage disturbance.

Clustering method	Voltage error	Running time
Method 1	5.771	30.6s
Method 2	2.421	14.6s
Method 3	0.763	31.5s
Method 4	0.454	14.3s
Detailed model	-	680.6s

The comparison of the error conditions of the active power curves and the running time confirms the validity of the clustering index and clustering method proposed in this paper in the case of small disturbance.

In summary, the simulation results show that, compared with the traditional clustering index and the traditional FCM clustering algorithm, the method of selecting parameters to characterize the dynamic characteristics of the PV system is more scientific in this paper. The Canopy-FCM algorithm avoids blind selection of the initial clustering center and the number of clusters. The equivalent model of the PV power plant based on the equivalent clustering index and the Canopy-FCM algorithm can more accurately characterize the dynamic characteristics of the original PV power plant under different disturbance types.

V. CONCLUSION

This paper describes the mathematical modeling process of a grid-connected PV system in detail. According to the mathematical model of the grid-connected inverter and its control system, an equivalent clustering index based on the zero-pole expression of the closed-loop transfer function is established. The rationality of index selection is verified by simulation. In order to improve the shortcomings of the traditional FCM clustering algorithm, the low-complexity Canopy algorithm is used to generate the initial clustering centers and the number of clusters for the FCM algorithm, which can increase the clustering accuracy and reduce the amount of calculation.

The results of the example show that the equivalent clustering index established in this paper can characterize the dynamic performance of the PV power station more accurately. Compared with the traditional FCM clustering algorithm, the Canopy-FCM algorithm proposed in this paper avoids the defects caused by the random selection of clustering centers and clustering numbers.

In order to simplify the analysis, the equivalent clustering index established in this paper is based on the MPPT-unit power factor control strategy of a photovoltaic grid-connected inverter. For other control strategies, a similar method can be used for analysis. It should be pointed out that when discussing the clustering index of PV power plants, some omissions and simplifications are made, and the accuracy of clustering is thus reduced.

REFERENCES

[1] T. Tomita, "Toward giga-watt production of silicon photovoltaic cells, modules and systems," in *Proc. 31st IEEE Photovolt. Spec. Conf.*, Lake Buena Vista, FL, USA, Jan. 2005, pp. 7–11.

[2] Y. T. Tan and D. S. Kirschen, "Impact on the power system of a large penetration of photovoltaic generation," in *Proc. IEEE Power Eng. Soc. Gen. Meeting*, Tampa, FL, USA, Jun. 2007, pp. 1–8.

[3] J. V. Paatero and P. D. Lund, "Effects of large-scale photovoltaic power integration on electricity distribution networks," *Renew. Energy*, vol. 32, no. 2, pp. 216–234, Feb. 2007.

[4] C. Rodriguez and G. A. J. Amaratunga, "Dynamic stability of grid-connected photovoltaic systems," in *Proc. IEEE Power Eng. Soc. Gen. Meeting*, Denver, CO, USA, Jun. 2004, pp. 2193–2199.

[5] R. Naik, N. Mohan, M. Rogers, and A. Bulawka, "A novel grid interface, optimized for utility-scale applications of photovoltaic, wind-electric, and fuel-cell systems," *IEEE Trans. Power Del.*, vol. 10, no. 4, pp. 1920–1926, Oct. 1995.

[6] T.-F. Wu, C.-L. Shen, H.-S. Nein, and G.-F. Li, "A 1ϕ3W inverter with grid connection and active power filtering based on nonlinear programming and fast-zero-phase detection algorithm," *IEEE Trans. Power Electron.*, vol. 20, no. 1, pp. 218–226, Jan. 2005.

[7] Y. T. Tan, D. S. Kirschen, and N. Jenkins, "A model of PV generation suitable for stability analysis," *IEEE Trans. Energy Convers.*, vol. 19, no. 4, pp. 748–755, Dec. 2004.

[8] V. V. Pavlovskyi, L. N. Lukianenko, and A. M. Zakharov, "Software polymodels of solar photovoltaic plants for different types of system studies," in *Proc. IEEE Int. Conf. Intell. Energy Power Syst. (IEPS)*, Kiev, Ukraine, Jun. 2014, pp. 163–167.

[9] S. Soni, G. G. Karady, M. Morjaria, and V. Chadliev, "Comparison of full and reduced scale solar PV plant models in multi-machine power systems," in *Proc. IEEE PES T D Conf. Expo.*, Chicago, IL, USA, Apr. 2014, pp. 1–5.

[10] D. Remon, A. M. Cantarellas, M. A. A. Elsaharty, C. Koch-Ciobotaru, and P. Rodriguez, "Equivalent model of a synchronous PV power plant," in *Proc. IEEE Energy Convers. Congr. Expo. (ECCE)*, Montreal, QC, Canada, Sep. 2015, pp. 47–53.

[11] Y. Chai, J. H. Zheng, L. Shu, and S. Z. Zhu, "Equivalent modeling of large-scale photovoltaic power plant," *Appl. Mech. Mater.*, vols. 448–453, pp. 1419–1422, Oct. 2013.

[12] Z. Ma, J. Zheng, S. Zhu, X. Shen, L. Wei, X. Wang, and K. Men, "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] D. Rodriguez, 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.

[14] S. Mukherjee, V. R. Chowdhury, P. Shamsi, and M. Ferdowsi, "Model reference adaptive control based estimation of equivalent resistance and reactance in grid-connected inverters," *IEEE Trans. Energy Convers.*, vol. 32, no. 4, pp. 1407–1417, Dec. 2017.

[15] P. P. Dash and M. Kazerani, "Dynamic modeling and performance analysis of a grid-connected current-source inverter-based photovoltaic system," *IEEE Trans. Sustain. Energy*, vol. 2, no. 4, pp. 443–450, Oct. 2011.

[16] R. Podmore, "Identification of coherent generators for dynamic equivalents," *IEEE Trans. Power App. Syst.*, vol. PAS-97, no. 4, pp. 1344–1354, Jul. 1978.

[17] J. Conroy and R. Watson, "Aggregate modelling of wind farms containing full-converter wind turbine generators with permanent magnet synchronous machines: Transient stability studies," *IET Renew. Power Gener.*, vol. 3, no. 1, pp. 39–52, Mar. 2009.

[18] Q. Zeng and L. Chang, "Study of advanced current control strategies for three-phase grid-connected pwm inverters for distributed generation," in *Proc. IEEE Conf. Control Appl. (CCA)*, Toronto, ON, Canada, Aug. 2005, pp. 1311–1316.

[19] Q. Zeng, L. Chang, and P. Song, "SVPWM-based current controller with grid harmonic compensation for three-phase grid-connected VSI," in *Proc. IEEE 35th Annu. Power Electron. Spec. Conf. (PESC)*, Aachen, Germany, Jun. 2004, pp. 2494–2500.

[20] M. Prodanovic and T. C. Green, "Control and filter design of three-phase inverters for high power quality grid connection," *IEEE Trans. Power Electron.*, vol. 18, no. 1, pp. 373–380, Jan. 2003.

[21] J. Svensson and M. Lindgren, "Vector current controlled grid connected voltage source converter-influence of nonlinearities on the performance," in *Proc. IEEE 29th Annu. Power Electron. Spec. Conf. (PESC)*, Fukuoka, Japan, May 1998, pp. 531–537.

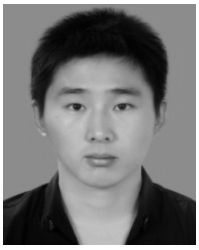
[22] C. N.-M. Ho, V. S. P. Cheung, and H. S.-H. Chung, "Constant-frequency hysteresis current control of grid-connected VSI without bandwidth control," *IEEE Trans. Power Electron.*, vol. 24, no. 11, pp. 2484–2495, Nov. 2009.

[23] A. O. Zue and A. Chandra, "Simulation and stability analysis of a 100 kW grid connected LCL photovoltaic inverter for industry," in *Proc. IEEE Power Eng. Soc. Gen. Meeting*, Montreal, QC, Canada, Jun. 2006, pp. 1–6.

[24] H. Wu, Z. Liu, Y. Chen, B. Xu, and X. Qi, "Equivalent modeling method for regional decentralized photovoltaic clusters based on cluster analysis," *CPSS Trans. Power Electron. Appl.*, vol. 3, no. 2, pp. 146–153, Jun. 2018.

HONGBIN WU was born in 1972. He received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the Hefei University of Technology, Hefei, China, in 1994, 1998, and 2005, respectively. He is currently a Professor with the Hefei University of Technology. His research interests include distributed generation technology and distribution network modeling and simulation.

JIAJIA ZHANG was born in 1995. He received the B.S. degree in electrical engineering and automation from the Hefei University of Technology, in 2017, where he is currently pursuing the M.S. degree in electrical engineering. His research interest includes distributed power generation technology.



CHEN LUO was born in 1990. He received the M.S. degree in electrical engineering and automation from North China Electric Power University, in 2016. He is currently pursuing the Ph.D. degree in electrical engineering with the Hefei University of Technology. His research interest includes distributed power generation technology.

BIN XU was born in 1970. He received the M.S. degree in electrical engineering and automation from Hohai University, in 2000. He is currently a Senior Engineer with the State Grid Anhui Electric Power Research Institute. His research areas include distribution network system planning and design techniques.