



ARTICLE

An Equivalent Modeling Method for Distributed Photovoltaic Systems Based on Voltage Sensitivity and Spectral Clustering

Yizhao Cheng*, Li Zhang and Bo Zhao

Automatization Engineering College, Beijing Information Science & Technology University, Beijing, China

*Corresponding Author: Yizhao Cheng. Email: cyz020522@163.com

Received: 02 June 2026; Accepted: 21 July 2026; Published: 18 October 2026

ABSTRACT: To address the increasing complexity of simulation and computation caused by the integration of a high proportion of distributed photovoltaics into the grid, this paper proposes an equivalent cluster modeling method for distributed photovoltaic systems based on electrical distance and Spectral Clustering (SC). By inverting the system Jacobian matrix, the independent response relationships of voltage to active power and reactive power disturbances are derived, and a global voltage increment matrix that comprehensively accounts for the bidirectional influence of power is further constructed. On this basis, the SC algorithm is employed to achieve a rational partitioning of photovoltaic nodes in a low-dimensional space, and the capacity-weighted method is adopted to determine the equivalent parameters. The IEEE 33-bus system is built in PSASP for simulation-based validation. The results show that under voltage fluctuation and short-circuit fault conditions, the equivalent model established by the proposed method can accurately track the power response of the detailed model, with an error of less than 2%.

KEYWORDS: Distributed photovoltaics; electrical distance; spectral clustering; voltage increment matrix; capacity-weighted method

1 Introduction

With the rapid development of new energy technologies, the penetration of distributed photovoltaic generation in distribution networks has continued to increase [1,2]. A large number of photovoltaic power stations are connected to distribution systems through inverters, driving traditional distribution networks to gradually evolve into power-electronics-dominated grids with a high proportion of converter-interfaced devices. Compared with conventional synchronous generators, photovoltaic generation systems mainly rely on power electronic equipment and control strategies to achieve grid-connected operation, and their dynamic characteristics differ significantly from those of traditional power sources [3]. As the scale of distributed photovoltaics continues to expand, the coupling interactions among numerous grid-connected inverters exert an important influence on system voltage stability and transient response characteristics [4]. Therefore, electromechanical transient simulation of distribution networks with a high penetration of distributed photovoltaics is of great significance for power grid planning and stability analysis.

However, in practical engineering simulations, establishing a detailed model for each distributed photovoltaic power station would lead to a rapid increase in system size, significantly raising the model dimension and computational complexity. This, in turn, would prolong the simulation time and reduce computational efficiency. Especially in distribution networks containing a large number of distributed photovoltaic power stations, directly applying detailed models for dynamic simulation often fails to meet

the engineering requirements for computational efficiency [5,6]. Therefore, under the premise of ensuring simulation accuracy, the aggregated equivalent modeling of multiple distributed photovoltaic power stations through reasonable equivalencing methods has become an important research direction for improving simulation efficiency [7].

Research on the equivalent modeling of new energy power stations can be broadly classified into voltage–power characteristic-based methods, topology-oriented aggregation methods, and data-driven clustering-based equivalent modeling methods [8–10]. Voltage–power characteristic-based methods, such as [11], construct equivalent models according to the distribution characteristics of voltage and power to address the possible entry of generating units into the low-voltage ride-through state under fault conditions. Although these methods can reflect the operating characteristics of new energy stations under disturbances, they usually show strong dependence on operating conditions and involve relatively high topological modeling complexity. Topology-oriented aggregation methods, represented by [12], consider the structural characteristics of the distribution network when grouping distributed photovoltaic units. However, such methods mainly rely on topological information and therefore cannot fully capture the dynamic coupling characteristics among photovoltaic units, which limits equivalencing accuracy. Data-driven clustering-based methods have also attracted considerable attention. For example, Ref. [13] combines photovoltaic output characteristics with node sensitivity indices and applies the K-means algorithm to aggregated modeling, while Ref. [14] establishes an equivalent model for photovoltaic power stations using the AP clustering algorithm. These methods improve the identification of similarity among photovoltaic nodes, but they still have limitations. K-means, based on Euclidean distance, is more suitable for convex data distributions and has difficulty characterizing the complex topology of distribution networks, whereas AP becomes computationally expensive when the number of photovoltaic nodes is large. In addition, the method in [15] increases model complexity through multi-source and multi-scale data fusion, while the method in [16] requires multiple parameter settings, which increases the difficulty of model tuning. In summary, although existing studies have made important progress, challenges remain in balancing electrical interpretability, modeling simplicity, computational efficiency, and equivalencing accuracy. Compared with conventional clustering methods, SC, based on graph theory and matrix spectral analysis, can characterize the intrinsic relationships among samples through a similarity matrix and achieve effective partitioning in a low-dimensional feature space [17]. Meanwhile, voltage sensitivity can reflect the influence of nodal power disturbances on system voltage distribution and serves as an important indicator for describing electrical coupling among distribution network nodes [18]. Therefore, combining electrical distance constructed from voltage sensitivity with SC enables more accurate identification of distributed photovoltaic nodes with similar electrical characteristics, providing a reliable basis for equivalent cluster modeling.

Based on the above analysis, this paper develops an equivalent modeling method for distributed photovoltaic systems based on joint active–reactive global voltage–response characteristics and spectral graph clustering. The main contributions are as follows. First, a global voltage–response feature jointly considering active- and reactive-power variations is constructed for each photovoltaic node, and an electrically interpretable distance is defined according to the difference between the response vectors. Second, an adaptive cluster-number determination strategy combining the eigengap and weighted modularity is introduced to reduce the dependence on a manually specified cluster number. Third, the proposed clustering method is integrated with capacity-weighted parameter equivalencing and is quantitatively compared with K-means- and FCM-based equivalent models under typical disturbances.

2 Development of the Electromechanical Transient Model of Distributed Photovoltaics

2.1 Photovoltaic Inverter Model

The essence of a photovoltaic inverter is to utilize the switching characteristics of power semiconductor devices to precisely convert the randomly fluctuating direct current generated by the photovoltaic array into sinusoidal alternating current whose frequency, phase, and amplitude strictly satisfy grid-connection standards [19]. Unlike conventional synchronous power sources, the output characteristics of a photovoltaic generation system are mainly determined by its controller and corresponding parameters. Its topological structure is shown in Fig. 1.

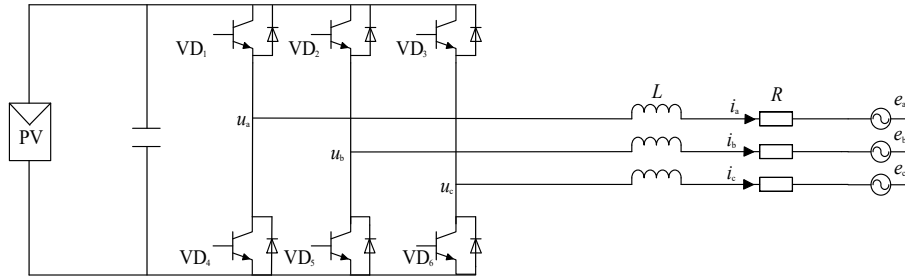


Figure 1: Main circuit topology of the photovoltaic inverter.

As can be seen from Fig. 1, in the three-phase stationary coordinate system, the voltage–current relationship on the grid side of the inverter can be expressed as:

$$\begin{bmatrix} e_a \\ e_b \\ e_c \end{bmatrix} = \begin{bmatrix} u_a \\ u_b \\ u_c \end{bmatrix} + L \frac{d}{dt} \begin{bmatrix} i_a \\ i_b \\ i_c \end{bmatrix} + R \begin{bmatrix} i_a \\ i_b \\ i_c \end{bmatrix} \quad (1)$$

where e_a , e_b and e_c are the instantaneous values of the three-phase grid-side voltages; u_a , u_b and u_c are the instantaneous values of the three-phase output voltages of the inverter; i_a , i_b and i_c are the instantaneous values of the three-phase currents; L and R denote the filter inductance and resistance.

In the control strategy of three-phase grid-connected inverters, controller design in the dq rotating coordinate system is more commonly adopted than that in the stationary coordinate system [20]. Its core advantages are as follows. First, through coordinate transformation, AC time-varying signals are converted into DC steady-state quantities, so that zero steady-state error tracking can be achieved by using a standard PI controller. Second, this scheme effectively realizes the decoupling of active and reactive power, thereby enabling the system to flexibly regulate the power factor and provide reactive power support. This decoupling characteristic significantly improves the grid-connected performance of the inverter. The voltage equations of the inverter in the dq coordinate system can be expressed as:

$$\begin{cases} u_d = Ri_d + L \frac{di_d}{dt} - \omega_0 Li_q + e_d \\ u_q = Ri_q + L \frac{di_q}{dt} + \omega_0 Li_d + e_q \end{cases} \quad (2)$$

where e_d and e_q represent the grid electromotive forces on the d- and q-axes; u_d , u_q , i_d , and i_q denote the output voltages and currents of the inverter on the d- and q-axes.

The three-phase grid-connected inverter adopts a double closed-loop control structure, which enables independent control of active and reactive power while ensuring both efficiency and stability in grid-connected operation. Fig. 2 illustrates its control process under normal operating conditions in detail.

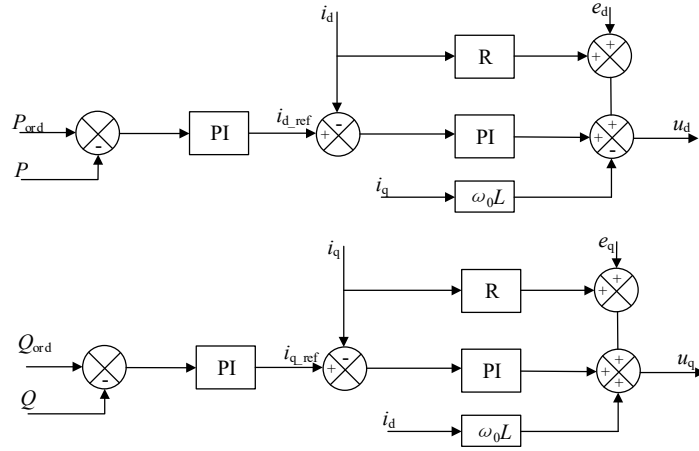


Figure 2: Control block diagram of the inverter under normal operating conditions.

As can be seen from Fig. 2, the vector control equations of the inverter current inner loop are given as follows:

$$\begin{cases} u_d = Ri_d + (i_{d_ref} - i_d)(K_{ip1} + \frac{1}{s}K_{ii1}) - \omega_0 Li_q + e_d \\ u_q = Ri_q + (i_{q_ref} - i_q)(K_{ip2} + \frac{1}{s}K_{ii2}) + \omega_0 Li_d + e_q \end{cases} \quad (3)$$

2.2 Electromechanical Transient Model of Photovoltaic Systems

To balance simulation accuracy and computational efficiency in electromechanical transient analysis of large-scale power systems, PSASP applies a simplified model to photovoltaic generation systems. At the physical architecture level, based on the assumption of three-phase symmetry, only the fundamental-frequency component is retained while higher-order harmonics are neglected, and inductors as well as power electronic devices are idealized so as to eliminate nonlinear effects caused by magnetic saturation and switching losses. At the control dynamic level, according to the principle of time-scale separation, the dynamic delays of PWM modulation and the current inner loop are neglected because the simulation time step is much larger than the microsecond-level switching period. Meanwhile, the MPPT tracking process is omitted because environmental conditions remain relatively stable within the second-level simulation time scale. In this way, the complex electromagnetic transient details are equivalently transformed into a quasi-steady-state power control model with fast response and reduced computational burden [21]. The overall structure of the simplified electromechanical transient model of distributed photovoltaics is shown in Fig. 3.

Among them, the control block diagram of the inverter electromechanical transient model is shown in Fig. 4.

The adopted photovoltaic model is intended for electromechanical transient analysis rather than electromagnetic transient simulation. The model retains the outer-loop active- and reactive-power control characteristics, voltage-dependent control response, and fault ride-through behavior that dominate the second-level dynamic response at the PCC. Fast switching processes, PWM delays, semiconductor-level dynamics, and high-frequency harmonics are neglected according to the time-scale separation principle.

Therefore, the proposed equivalent model is suitable for system-level power and voltage response studies, but it is not intended for harmonic, switching transient, or device-level stability analysis.

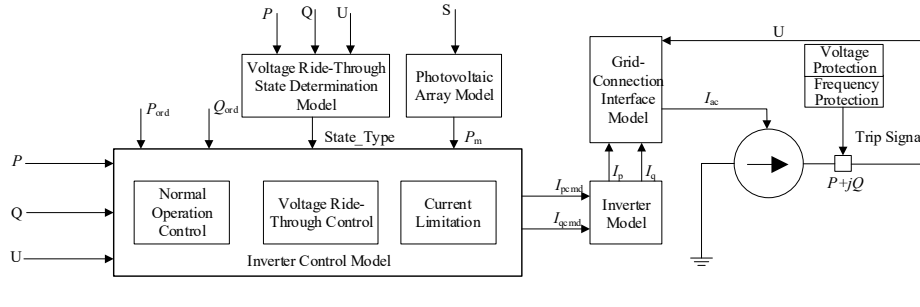


Figure 3: Overall structure of the electromechanical transient model of distributed photovoltaics.

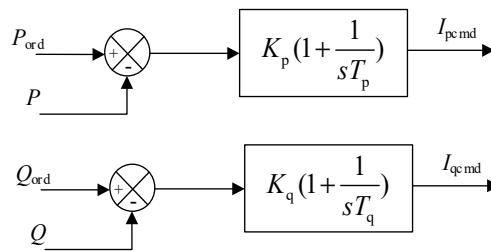


Figure 4: Control block diagram of the inverter electromechanical transient model.

3 Distributed Photovoltaic Clustering Based on Spectral Clustering

3.1 Electrical Distance Based on Joint Active–Reactive Voltage Responses

The dynamic similarity among distributed photovoltaic nodes is determined not only by their geographical or topological proximity, but also by the system-wide voltage responses caused by their active- and reactive-power variations [22]. Conventional clustering indices based on geographical distance, line impedance, photovoltaic output, or a single voltage-sensitivity component may fail to fully describe the coupled influence of active and reactive power on the voltage distribution of an active distribution network. Therefore, this paper constructs a joint active–reactive voltage-response feature for each photovoltaic node. Specifically, the voltage–active-power and voltage–reactive-power sensitivity matrices are first derived from the reduced Jacobian matrix. The two sensitivity components are then combined with the corresponding active- and reactive-power increments to obtain the global voltage-response vector caused by each photovoltaic node. The electrical distance between two photovoltaic nodes is defined according to the difference between their global voltage-response vectors. In this way, the clustering index represents not merely the physical distance between nodes, but the similarity of their system-wide electrical influence under power disturbances.

In an n-bus distribution network with an arbitrary topology, the voltage–power sensitivity matrix, which describes the mapping relationship between nodal voltage variations and power variations, can be obtained by inverting the Jacobian matrix under the equality constraints of the power flow equations.

$$\begin{bmatrix} \Delta P \\ \Delta Q \end{bmatrix} = \begin{bmatrix} J_{P\theta} & J_{PU} \\ J_{Q\theta} & J_{QU} \end{bmatrix} \begin{bmatrix} \Delta \theta \\ \Delta U \end{bmatrix} \quad (4)$$

where ΔP and ΔQ denote the active- and reactive-power injection increments of the non-slack buses, respectively; $\Delta \theta$ and ΔU denote the increments of nodal voltage phase angles and voltage magnitudes, respectively. The four Jacobian submatrices are defined as $J_{P\theta} = \partial P / \partial \theta$, $J_{PU} = \partial P / \partial U$, $J_{Q\theta} = \partial Q / \partial \theta$, and $J_{QU} = \partial Q / \partial U$. The rows and columns associated with the slack bus are excluded from the reduced Jacobian matrix.

According to the mapping relationship shown in (4), the voltage fluctuation of system nodes is jointly affected by active and reactive power disturbances. To derive the voltage–active-power sensitivity matrix, the reactive-power injection increment is set to zero, i.e., $\Delta Q = 0$. From the second row of Eq. (4),

$$J_{Q\theta}\Delta\theta + J_{QU}\Delta U = 0 \quad (5)$$

Therefore,

$$\Delta\theta = -J_{Q\theta}^{-1}J_{QU}\Delta U \quad (6)$$

Substituting this expression into the first row of Eq. (4) gives,

$$\Delta P = (J_{PU} - J_{P\theta}J_{Q\theta}^{-1}J_{QU})\Delta U \quad (7)$$

Hence, the voltage–active-power sensitivity matrix is expressed as

$$S_{UP} = \left[\frac{\partial U}{\partial P} \right] = (J_{PU} - J_{P\theta}J_{Q\theta}^{-1}J_{QU})^{-1} \quad (8)$$

Similarly, by setting the active power variation to $\Delta P = 0$, the voltage–reactive power sensitivity matrix characterizing the response relationship between system voltage and reactive power can be obtained as:

$$S_{UQ} = \left[\frac{\partial U}{\partial Q} \right] = (J_{QU} - J_{Q\theta}J_{P\theta}^{-1}J_{PU})^{-1} \quad (9)$$

The matrices $S_{UP} = \partial U / \partial P$ and $S_{UQ} = \partial U / \partial Q$ characterize two complementary voltage-response mechanisms. The former describes the redistribution of nodal voltages caused by active-power variations, whereas the latter reflects the voltage response associated with reactive-power variations. Using either sensitivity matrix alone may omit part of the electrical coupling characteristics of inverter-interfaced photovoltaic systems. Accordingly, the two sensitivity components are combined in this paper to construct a unified global voltage-response feature.

Based on the voltage–power mapping relationships established by (8) and (9), the response characteristics of nodal voltages to power disturbances can be further derived. For a distribution network model containing n nodes, the point of common coupling (PCC) at the sending end is set as the slack bus. For each photovoltaic node j , its electrical influence on the distribution network is represented by a global voltage-response vector. When the active- and reactive-power injections at node j vary by ΔP_j and ΔQ_j , respectively, the voltage response of all non-slack buses is expressed as:

$$\Delta U^{(j)} = S_{UP}^{(:,j)}\Delta P_j + S_{UQ}^{(:,j)}\Delta Q_j \quad (10)$$

where $\Delta U^{(j)} \in \mathbb{R}^{n-1}$ is the global voltage-response vector associated with the power variation at photovoltaic node j ; $S_{UP}^{(:,j)}$ and $S_{UQ}^{(:,j)}$ are the j -th column vectors of the voltage–active-power and voltage–reactive-power sensitivity matrices, respectively; and ΔP_j and ΔQ_j are the active- and reactive-power increments at node j .

Each element of $\Delta U^{(j)}$ represents the voltage variation of one non-slack bus caused by the joint active- and reactive-power variation at node j .

By applying the same construction to all photovoltaic nodes, a global voltage-response matrix is obtained as:

$$U_{\text{res}} = [\Delta U^{(1)}, \Delta U^{(2)}, \dots, \Delta U^{(m)}] \quad (11)$$

The j -th column of U_{res} describes the system-wide voltage-response pattern caused by the active- and reactive-power variations at photovoltaic node j . Therefore, photovoltaic nodes with similar column vectors have similar electrical influence on the distribution network and are suitable for aggregation into the same cluster.

Based on the global voltage-response features, the electrical distance between photovoltaic nodes i and j is defined as the Euclidean distance between their response vectors:

$$d_{ij} = \|\Delta U^{(i)} - \Delta U^{(j)}\|_2 = \sqrt{\sum_{r=1}^{n-1} (\Delta U_r^{(i)} - \Delta U_r^{(j)})^2} \quad (12)$$

A smaller d_{ij} indicates that power variations at photovoltaic nodes i and j produce more similar voltage-response patterns throughout the distribution network. Therefore, these two nodes have stronger electrical similarity and are more suitable for aggregation into the same cluster. Unlike conventional distance indices based only on geographical location, line impedance, or a single local operating variable, the proposed distance simultaneously incorporates active-power sensitivity, reactive-power sensitivity, disturbance magnitude, and system-wide voltage propagation characteristics.

3.2 Cluster Validity Evaluation Index

Given the significant dispersion of distributed photovoltaic integration locations, the modularity concept ρ is introduced in this paper to quantitatively characterize the electrical coupling strength among nodes. By extending this concept to weighted network partitioning, a modularity evaluation function suitable for distribution network characteristics is constructed to accurately measure the internal cohesion of network cluster partitioning and the rationality of the partition structure.

Let A denote the weighted similarity matrix after partitioning, where A_{ij} represents the similarity between nodes i and j and is constructed based on the reciprocal of the electrical distance. The modularity $\rho(k)$ is defined as follows:

$$\rho(k) = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(\text{id}x_i, \text{id}x_j) \quad (13)$$

$$k_i = \sum_j A_{ij}, k_j = \sum_i A_{ij} \quad (14)$$

$$m = \frac{1}{2} \sum_{i,j} A_{ij} \quad (15)$$

where k_i and k_j are the sums of the weights of all edges connected to nodes i and j , respectively; m is the sum of the weights of all edges in the network; and $\delta(\text{id}x_i, \text{id}x_j)$ is the Kronecker delta function, which equals 1 if and only if nodes i and j belong to the same partition, and 0 otherwise.

The number of clusters k is a key parameter in spectral clustering. In the original procedure, modularity was used only to evaluate the clustering results after the partitioning process. However, modularity alone is a posterior indicator and cannot fully characterize the intrinsic spectral structure of the similarity graph. Therefore, an adaptive cluster-number determination strategy combining the eigengap and weighted modularity is introduced in this paper.

Let the eigenvalues of the normalized Laplacian matrix be arranged in ascending order as:

$$0 \leq \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n \quad (16)$$

The eigengap corresponding to the candidate cluster number k is defined as:

$$g(k) = \lambda_{k+1} - \lambda_k \quad (17)$$

A relatively large eigengap indicates that the first k eigenvectors form a stable low-dimensional representation and that the graph has a relatively clear k -cluster structure.

To further evaluate the electrical cohesion within clusters, the weighted modularity $\rho(k)$ is calculated for each candidate partition. The modularity measures the difference between the actual intra-cluster similarity and the expected similarity under a random weighted network. A larger value of $\rho(k)$ indicates stronger electrical coupling within the clusters and weaker coupling between different clusters.

Since the eigengap and modularity have different numerical ranges, they are normalized as:

$$\hat{g}(k) = \frac{g(k) - g_{\min}}{g_{\max} - g_{\min}} \quad (18)$$

$$\hat{\rho}(k) = \frac{\rho(k) - \rho_{\min}}{\rho_{\max} - \rho_{\min}} \quad (19)$$

A comprehensive cluster-validity index is then constructed as:

$$F(k) = \alpha \hat{g}(k) + (1 - \alpha) \hat{\rho}(k) \quad (20)$$

where α is the weighting coefficient. In this study, $\alpha = 0.5$ is adopted to give equal importance to the spectral separation and the electrical cohesion of the partition. The optimal cluster number is determined as:

$$k^* = \arg \max_{k \in \kappa} F(k) \quad (21)$$

where κ denotes the predefined candidate set of cluster numbers. This strategy incorporates both the intrinsic spectral structure of the similarity graph and the electrical coupling quality of the obtained partition, thereby reducing the dependence on a manually specified cluster number.

3.3 Spectral Clustering Algorithm

SC is a clustering method based on graph theory. Its core idea is to use the spectrum of the similarity matrix for dimensionality reduction, so as to achieve accurate partitioning of samples in a low-dimensional space. Compared with the traditional K-means algorithm, Spectral Clustering can identify sample spaces of arbitrary shapes and exhibits stronger global optimization capability when dealing with node clustering problems in distribution networks with complex topological structures [23]. Its specific implementation procedure is shown in Fig. 5.

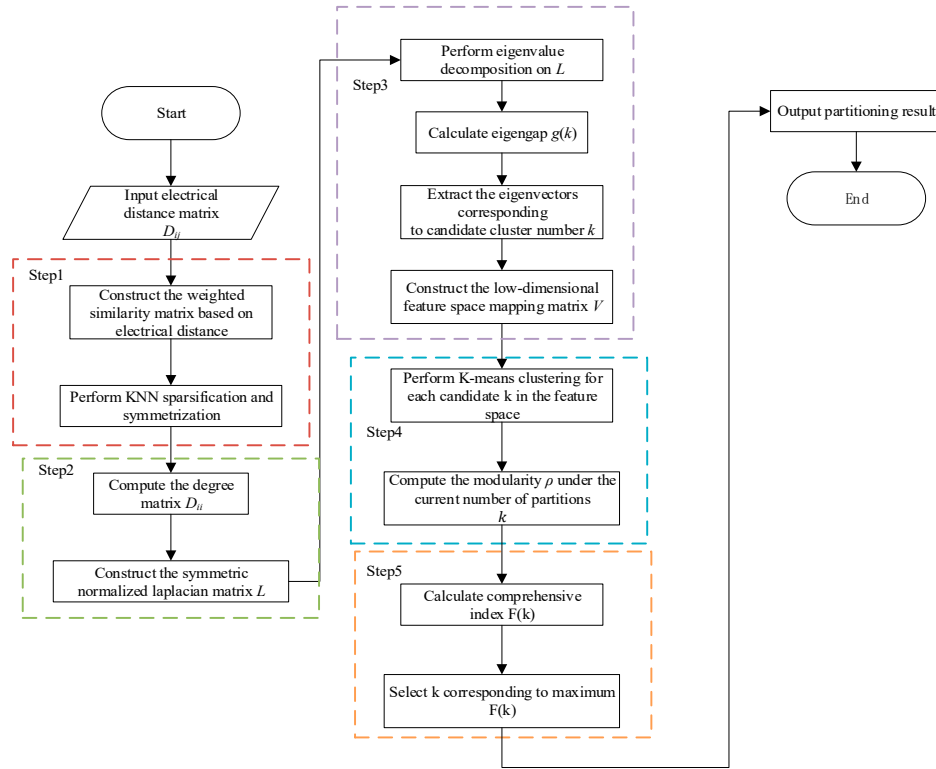


Figure 5: Detailed flowchart of the spectral clustering algorithm.

Step 1: Construct the similarity matrix A . To characterize the similarity between nodes, this paper constructs a weighted similarity matrix based on the reciprocal of electrical distance. A larger A_{ij} indicates that nodes i and j are closer in terms of voltage response characteristics. To enhance local coupling features, the K-nearest neighbor (KNN) algorithm is employed to sparsify the similarity matrix, ensuring that the clustering process focuses on node groups with relatively small electrical distances.

Step 2: Calculate the normalized Laplacian matrix L . First, compute the degree matrix D , which is a diagonal matrix whose diagonal elements are given by:

$$D_{ii} = \sum_{j=1}^n A_{ij} \quad (22)$$

D_{ii} represents the sum of the weights of all edges connected to node i . To eliminate the influence of differences in node degree on the clustering results and improve the robustness of the algorithm to partitions of unbalanced sizes, the symmetric normalized Laplacian matrix L is introduced in this paper as:

$$L = I - D^{-1/2} A D^{-1/2} \quad (23)$$

where I is the identity matrix. The spectral properties of this matrix can effectively characterize the manifold structure of distributed photovoltaic nodes in the electrical topology.

Step 3: Perform spectral analysis and construct the low-dimensional feature space. Eigenvalue decomposition is performed on the normalized Laplacian matrix, and the obtained eigenvalues are arranged in ascending order. The differences between adjacent eigenvalues are calculated for different candidate cluster numbers. A larger eigengap indicates a clearer separation between the leading spectral components and the remaining components, suggesting that the corresponding cluster number may represent a relatively stable

partition structure. For each candidate cluster number, the eigenvectors corresponding to the smallest eigenvalues are extracted and assembled into a feature matrix. The feature matrix is then normalized row by row to construct the low-dimensional mapping matrix V . Through this spectral mapping, the complex electrical coupling relationships in the original space are transformed into a point distribution in a low-dimensional Euclidean space.

Step 4: Perform clustering and evaluate each candidate partition. For each candidate cluster number k , K-means clustering is performed on the row vectors of the low-dimensional feature matrix V . According to the correspondence between the feature vectors and the original photovoltaic nodes, the clustering labels are mapped back to the distribution network. The weighted modularity $\rho(k)$ is then calculated for the partition obtained under the current value of k . The eigengap $g(k)$ reflects the intrinsic spectral separation of the similarity graph, whereas the modularity $\rho(k)$ evaluates the electrical cohesion within clusters and the separation between different clusters.

Step 5: Adaptively determine the number of clusters. For each candidate value of k , the eigengap of the normalized Laplacian matrix and the weighted modularity of the corresponding partition are calculated. After normalization, the two indices are combined to form the comprehensive validity index $F(k)$. The value corresponding to the maximum $F(k)$ is selected as the optimal number of clusters k^* .

4 Calculation of Equivalent Parameters

4.1 Parameter Equivalencing Based on the Capacity-Weighted Method

After clustering the photovoltaic groups, the key step is to perform aggregated calculation of the equivalent unit parameters, with the objective of maintaining a high degree of consistency in the power and voltage response characteristics at the PCC before and after equivalencing. The essence of equivalent modeling for photovoltaic clusters is to use a single equivalent unit to represent the overall operating characteristics of the original group of multiple units. This treatment is applicable when the units within the same cluster adopt identical or dynamically compatible control structures. For heterogeneous controllers, a control-type preclassification step is required before applying the proposed electrical-distance-based clustering method. Therefore, the equivalent model can still follow the original unit topology in structure, while only its physical parameters need to be recalibrated and equivalently calculated. The specific parameter calculation procedure is as follows.

(1) Equivalent parameters of the photovoltaic array

To ensure the equivalence of total installed capacity and output voltage characteristics before and after the aggregation of photovoltaic clusters, the numbers of series- and parallel-connected modules in the equivalent array should be converted according to the consistency principle. Based on the topological equivalence constraint, the module configuration parameters of the equivalent array are calculated as follows:

$$N_{seq} = N_{sc}, N_{peq} = \delta N_{pc} \quad (24)$$

where N denotes the number of series-connected or parallel-connected modules; s , p , eq , and c represent the series value, parallel value, equivalent value, and cluster-center value, respectively; and δ is the ratio of the total capacity S of the photovoltaic power stations in the group to the capacity S_c of the cluster-center power station.

(2) Equivalent parameters of the transformer

Each photovoltaic power station is connected to the grid through a transformer, and the aggregated parameters of transformer capacity and impedance are calculated as follows:

$$S_{teq} = \delta S_{tc}, Z_{teq} = Z_{tc} / \delta \quad (25)$$

where S_t and Z_t are the rated capacity and impedance of the transformer.

(3) Equivalent parameters of the controller

During the equivalencing process of photovoltaic clusters, the chopper and inverter control parameters of the equivalent unit in each subgroup are uniformly assigned to the corresponding control parameters of the photovoltaic power station to which the cluster center belongs. This treatment is intended to use the representative characteristics of the cluster center to characterize the dynamic control response of the cluster.

4.2 Equivalent Parameters of Loads and Collector Lines

Fig. 6 shows the wiring structure of a typical photovoltaic cluster before and after equivalencing, where the PCC denotes the point of common coupling of the photovoltaic cluster. As can be seen from Fig. 6b, the unknown network equivalent parameters include the equivalent line impedance Z_{eq} , the voltage at the access point of the equivalent photovoltaic power station U_{eq} , and the equivalent load S_{eq} .

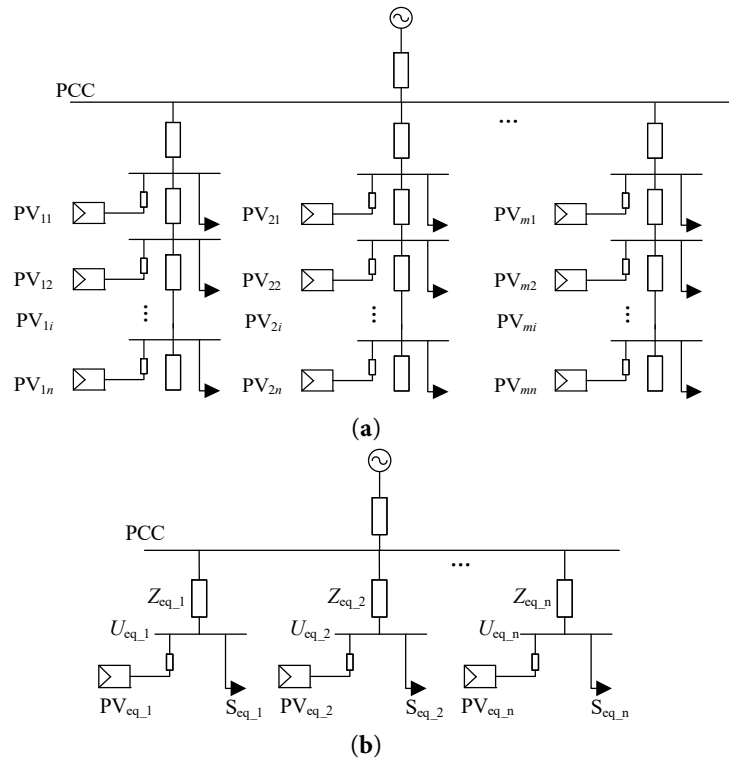


Figure 6: Collector Lines Before and After Equivalence of the Photovoltaic Cluster. (a) Collector lines of the system before equivalencing; (b) Collector lines of the system after equivalencing.

In this paper, the line impedance parameters are equivalenced according to the constant voltage-drop criterion, so as to ensure that the voltage difference between the photovoltaic cluster access point and the PCC remains unchanged before and after equivalencing. First, the voltage deviation ΔU_i of each photovoltaic power station access point relative to the PCC under steady-state conditions is calculated ($i = 1, 2, \dots, n$). On this basis, the weighted average voltage difference between each power station in the k -th cluster and the PCC is obtained by the weighted averaging method, which is expressed as follows:

$$\Delta \bar{U}_k = \frac{\sum_{i \in k} \Delta U_i P_i}{\sum_{i \in k} P_i} \quad (26)$$

After equivalencing, the voltage difference between the access point of the k -th equivalent unit and the PCC is given by:

$$\Delta \bar{U}_{eq_k} = \frac{Z_{eq_i} P_{eq_i}}{\sqrt{3} U_{PCC}} \quad (27)$$

where:

$$P_{eq_i} = \sum_{i \in k} P_i \quad (28)$$

Since the voltage difference remains unchanged, the equivalent line impedance can be calculated as follows:

$$Z_{eq_k} = \frac{\sqrt{3} U_{PCC} \sum_{i \in k} \Delta U_i P_i}{(\sum_{i \in k} P_i)^2} \quad (29)$$

The equivalent load of the k -th cluster is given by:

$$S_{eq_k} = \sum_{i \in k} S_{L_i} \quad (30)$$

5 Simulation Verification

5.1 Simulation Model

The IEEE 33-bus system is selected as the test system because it contains multiple photovoltaic integration points and has been widely adopted as a benchmark for active distribution network studies. Although the topology is radial, the electrical coupling among photovoltaic nodes is diverse due to different locations and feeder impedances. An IEEE 33-bus distribution network model with distributed photovoltaics is established in the PSASP platform. The distribution network contains a total of 20 distributed photovoltaic power stations, and their connection locations are shown in Fig. 7. The installed capacities of the distributed photovoltaic power stations are as follows: PV1–PV8 are 640 kW, PV9–PV14 are 520 kW, PV15–PV17 are 384 kW, and PV18–PV20 are 200 kW. Each photovoltaic generation system is connected to the distribution network through an ideal 0.38 kV/12.66 kV transformer, and an ideal voltage source is adopted to approximately represent the infinite bus.

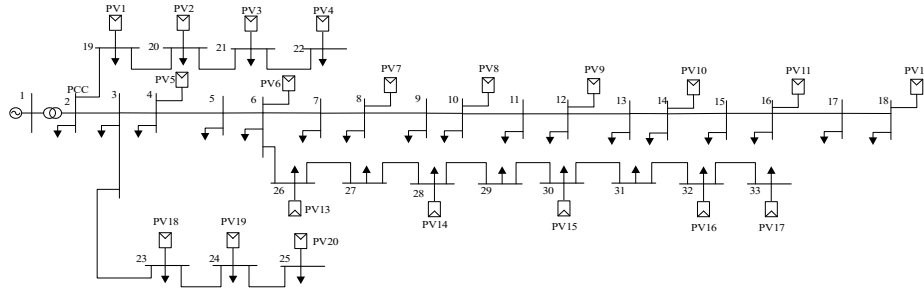


Figure 7: Topology of the IEEE 33-Bus distribution network with integrated distributed photovoltaics.

5.2 Analysis of Clustering Results

To determine the optimal number of clusters for the distributed photovoltaic power stations, the candidate cluster numbers from 2 to 8 are evaluated using both the eigengap of the normalized Laplacian matrix and the weighted modularity of the corresponding partitions. The eigengap reflects the intrinsic spectral separation of the similarity graph, whereas the modularity measures the electrical cohesion within each cluster and the separation between different clusters. After normalization, the two indicators are combined to form a comprehensive validity index. Fig. 8 shows the variations in the eigengap, modularity, and comprehensive validity index under different candidate cluster numbers.

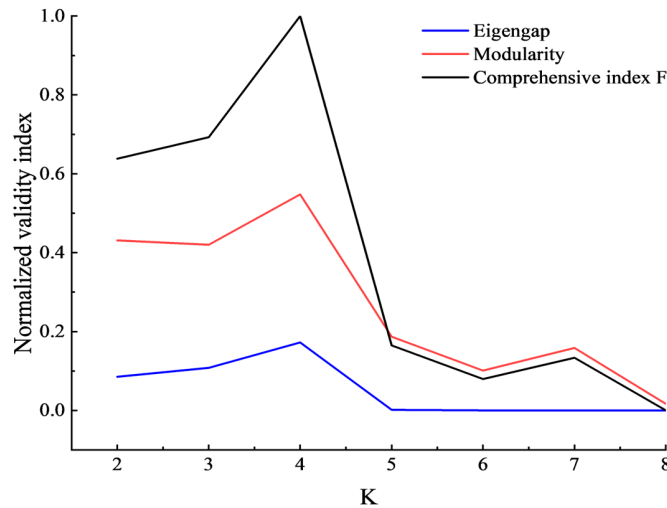


Figure 8: Comparison of cluster-validity indices under different cluster numbers.

As shown in Fig. 8, the eigengap and modularity exhibit different variation trends as the number of clusters changes. When the number of clusters is four, the comprehensive validity index reaches its maximum value. This result indicates that the four-cluster scheme provides the best balance between the spectral separation of the similarity graph and the electrical cohesion of the obtained partitions. Therefore, the optimal number of clusters is adaptively determined as four.

Based on the selected cluster number, the first four eigenvectors corresponding to the smallest eigenvalues of the normalized Laplacian matrix are extracted to construct the low-dimensional feature space. K-means clustering is then performed in this feature space, and the resulting labels are mapped back to the original photovoltaic nodes. The final clustering results are presented in Table 1.

Table 1: Clustering results.

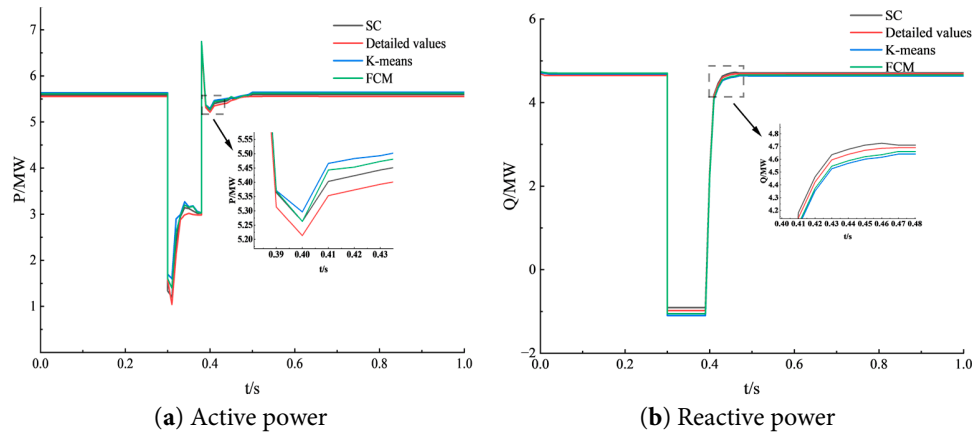
Group Number	Distributed Photovoltaic Unit Number
1	PV1, PV2, PV5, PV6, PV10, PV11, PV12, PV13, PV14, PV15, PV17, PV18, PV19, PV20
2	PV3, PV8
3	PV7, PV4
4	PV9, PV16

As shown in Table 1, the partitioning results obtained by the voltage-sensitivity-based electrical distance and the Spectral Clustering algorithm divide the 20 distributed photovoltaic power stations into four clusters. Subsequently, an equivalent modeling method is applied to aggregate the photovoltaic power stations within the same cluster, thereby reducing the system modeling scale. To verify the effectiveness of the proposed clustering-based equivalencing method, both a detailed model and an equivalent model are further constructed in the PSASP platform, and their dynamic response characteristics at the PCC are compared under multiple typical disturbance conditions so as to evaluate the simulation accuracy and applicability of the equivalent model.

5.3 Analysis of Simulation Results

5.3.1 Comparison of Dynamic Responses under Typical Disturbance Conditions

In this paper, three typical operating conditions are imposed at the PCC, namely voltage sag, voltage swell, and three-phase short-circuit fault, so as to compare and analyze the fitting performance and dynamic simulation accuracy of equivalent models under different clustering schemes. Specifically, the voltage sag is set to 80%, and the voltage swell is set to 20%. For all operating conditions, the total simulation time is 1 s; the fault occurs at 0.3 s and is cleared after 0.08 s. The comparisons of active and reactive power responses at the PCC between different equivalent models and the detailed model are shown in Figs. 9–11.

**Figure 9:** Power response at the PCC under voltage sag.

Overall, the simulation results show that under three typical disturbance conditions, namely voltage sag, voltage swell, and three-phase short-circuit fault, the active and reactive power response curves at the PCC obtained from the distributed photovoltaic equivalent model based on Spectral Clustering can accurately track the dynamic variation trends of the detailed model. This indicates that by reasonably partitioning distributed photovoltaic power stations using the electrical-distance-based Spectral Clustering method and establishing a multi-machine equivalent model combined with the capacity-weighted equivalencing

method, the system modeling complexity can be significantly reduced while maintaining high dynamic response accuracy. In addition, the comparison results under different operating conditions demonstrate that the proposed equivalent model exhibits good adaptability and stability in various scenarios, including voltage disturbances and short-circuit faults.

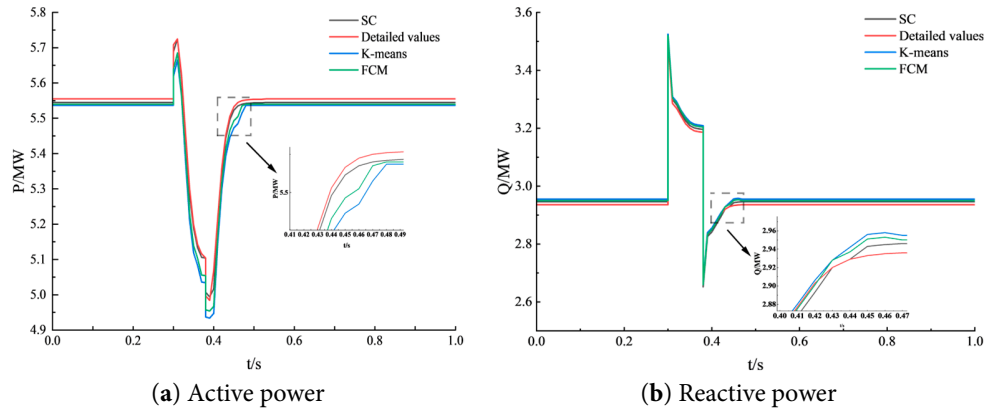


Figure 10: Power response at the PCC under voltage swell.

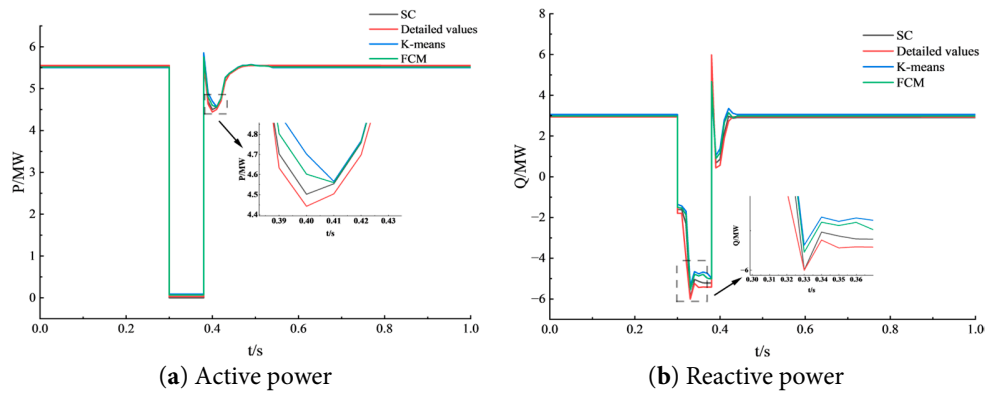


Figure 11: Power response at the PCC under three-phase short-circuit fault.

5.3.2 Quantitative Error Analysis

To accurately quantify the error between different equivalent models and the detailed model, the normalized root mean square error (*NRMSE*) [24] is adopted in this paper to evaluate the performance of different equivalent models. The calculation formula is given as follows:

$$NRMSE_P = \frac{\sqrt{\frac{(P-P_0)^2}{N}}}{P_{0\max} - P_{0\min}} \times 100\% \quad (31)$$

$$NRMSE_Q = \frac{\sqrt{\frac{(Q-Q_0)^2}{N}}}{Q_{0\max} - Q_{0\min}} \times 100\% \quad (32)$$

where P_0 and Q_0 are the active and reactive power of the detailed model; P and Q are the active and reactive power of each equivalent model; and N is the number of sampling points. The error comparison results are shown in Table 2.

Table 2: Error comparison results of different equivalent models.

Fault Type	Equivalent Model	NRMSEP/%	NRMSEQ/%
Voltage Sag	Equivalent Model Based on SC	1.03	0.61
	Equivalent Model Based on K-means	2.33	1.04
	Equivalent Model Based on FCM	1.64	0.80
Voltage Swell	Equivalent Model Based on SC	1.53	1.19
	Equivalent Model Based on K-means	4.67	2.24
	Equivalent Model Based on FCM	3.58	1.67
Three-Phase Short-Circuit	Equivalent Model Based on SC	0.43	1.57
	Equivalent Model Based on K-means	1.15	3.11
	Equivalent Model Based on FCM	0.83	2.51

As can be seen from the error comparison results in Table 2, the equivalent model established based on Spectral Clustering (SC) achieves lower NRMSE values under all three typical operating conditions, indicating better dynamic response consistency with the detailed photovoltaic model. Although the response curves of different equivalent models exhibit similar trends in Figs. 9–11, the quantitative error results demonstrate clear differences among these methods.

Compared with the K-means-based equivalent model, the proposed SC-based model reduces the active-power NRMSE by approximately 55.8%, 67.2%, and 62.6% under voltage sag, voltage swell, and three-phase short-circuit conditions, respectively. Meanwhile, the reactive-power NRMSE is reduced by approximately 41.3%, 46.9%, and 49.5%, respectively. Compared with the FCM-based equivalent model, the proposed method also achieves lower errors under all tested disturbances, demonstrating improved dynamic equivalence accuracy.

These results indicate that the advantage of the proposed method is mainly reflected in the quantitative consistency of dynamic responses rather than obvious visual differences in the response curves. By constructing electrical-distance-based clustering features that characterize the global voltage-response similarity among photovoltaic nodes, the proposed SC method can better preserve the electrical coupling relationships within photovoltaic clusters and improve the accuracy of the equivalent model.

As shown in Table 3, all three equivalent models reduce the number of photovoltaic dynamic models from 20 to 4, resulting in a substantial reduction in simulation time. Since the three equivalent models have the same model scale, their simulation times are relatively close. The proposed method therefore achieves higher dynamic-response accuracy while maintaining a computational cost comparable to those of the K-means- and FCM-based equivalent models.

Table 3: Comparison of model scale and simulation efficiency.

Model	Number of PV Dynamic Models	Simulation Time/s	Time Reduction/%
Detailed model	20	1.229	-
K-means equivalent	4	0.486	60.46
FCM equivalent	4	0.474	61.43
Proposed model	4	0.458	62.73

6 Conclusion

This paper proposes an equivalent cluster modeling method for distributed photovoltaic systems based on joint active–reactive voltage-response characteristics and spectral graph clustering. By deriving the

voltage–active-power and voltage–reactive-power sensitivity matrices from the reduced Jacobian matrix, a global voltage-response feature is constructed for each photovoltaic node. The electrical distance defined by the difference between voltage-response vectors can characterize the electrical coupling relationship among photovoltaic nodes from a system-wide perspective.

The proposed method is validated on the modified IEEE 33-bus active distribution system with multiple photovoltaic integration points. The simulation results under voltage sag, voltage swell, and three-phase short-circuit disturbances demonstrate that the proposed equivalent model can accurately track the active-power, reactive-power, and voltage responses of the detailed photovoltaic model. Compared with K-means- and FCM-based equivalent models, the proposed method achieves lower dynamic response errors and better clustering performance.

The proposed method effectively reduces the number of detailed photovoltaic models and decreases the computational burden of dynamic simulation while maintaining satisfactory accuracy. It should be noted that this study mainly focuses on photovoltaic aggregation in a typical active distribution network. Further studies considering comprehensive irradiance variations, meshed distribution networks, and multi-type distributed energy resources will be carried out in future work.

Acknowledgement: Not applicable.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: Conceptualization, Yizhao Cheng, Li Zhang, Bo Zhao; methodology, Yizhao Cheng, Li Zhang, Bo Zhao; software, Yizhao Cheng, Li Zhang, Bo Zhao; validation, Yizhao Cheng, Li Zhang, Bo Zhao; formal analysis, Yizhao Cheng, Li Zhang, Bo Zhao; investigation, Yizhao Cheng, Li Zhang, Bo Zhao; writing—original draft preparation, Yizhao Cheng, Li Zhang, Bo Zhao. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The authors confirm that the data supporting the findings of this study are available within the article.

Ethics Approval: This study did not involve human participants, animal subjects, or any clinical data. Therefore, ethical review and approval were not required for this research.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Rong S, Wang C, Dong Y, Shi B, Ma H. Adaptive modeling and analysis of distributed photovoltaic absorptive capacity based on improved simulated annealing algorithm. *Int J Low Carbon Technol.* 2024;19:2032–9. [[CrossRef](#)].
2. Niu X, Luo X. Policies and economic efficiency for distributed photovoltaic and energy storage industry. *Distrib Gener Altern Energy J.* 2023;38(4):1197–222. [[CrossRef](#)].
3. Zhang Z, Liao S, Sun Y, Xu J, Ke D, Wang B, et al. Unified dynamic equivalent model for distributed photovoltaic generation systems with different fault-ride-through strategies. *IEEE Trans Sustain Energy.* 2023;14(4):2062–78. [[CrossRef](#)].
4. Ndirangu K, Tafti HD, Fletcher JE, Konstantinou G. Validation of solar photovoltaic inverter behavior to grid disturbances through power-hardware-in-the-loop testing. In: *Proceedings of the 2021 IEEE 12th Energy Conversion Congress & Exposition-Asia (ECCE-Asia)*; 2021 May 24–27; Singapore. p. 250–5.
5. Qi J, Guo Q, Zhu Y, Yu Y, Tu L, Luo C, et al. Dynamic equivalent modeling of distributed photovoltaic generation systems in microgrid considering LVRT active power response difference. *Electronics.* 2025;14(22):4355. [[CrossRef](#)].
6. Li K, Huang Q, Fan R, Gao S, Wang A, Huang T, et al. Neural network-based aggregated equivalent modeling of distributed photovoltaic external characteristics of faults. *Electronics.* 2024;13(16):3232. [[CrossRef](#)].

7. Fang H, Zhang X, Ling X, Lu C, Ye Z. A linearization based model equivalent method for distributed photovoltaic generation clusters. In: Proceedings of the 2022 IEEE 6th Conference on Energy Internet and Energy System Integration (EI2); 2022 Nov 11–13; Chengdu, China. p. 1–5.
8. Xing F, Li X, Fan H, Zhao K, Li S, Li C. Static equivalent of distribution network with distributed PV considering correlation between fluctuation of PV and load. *Front Energy Res.* 2023;10:1119174. [[CrossRef](#)].
9. Liu W, Gu W, Liu W, Pan J, Du X, Xu B. Equivalent modeling of distributed photovoltaic power station clusters based on deep belief network. In: Proceedings of the 2018 International Conference on Power System Technology (POWERCON); 2018 Nov 6–8; Guangzhou, China. p. 4333–9.
10. Tan H, Miao S, Wang T, Wang B, Wang J, Yao F, et al. Research on transient equivalent modeling method for distribution networks with high-penetration distributed photovoltaics. *Sol Energy.* 2025;299:113725. [[CrossRef](#)].
11. Zheng T, Deng J, Wang X, Xia N, Wang T. Equivalent modeling method based on voltage and power distribution characteristics of a renewable energy station. *Power Syst Prot Control.* 2024;52(22):47–58. [[CrossRef](#)].
12. Tang Y, Sun S, Zhao B, Ni C, Che L, Li J. Distributed photovoltaic generation aggregation approach considering distribution network topology. *Energies.* 2024;17(12):2990. [[CrossRef](#)].
13. Yang S, Hu X, Tu Z, Yin R, Zhang Y, Li J, et al. A distributed photovoltaic aggregation method considering photovoltaic output characteristics and active voltage sensitivity. *Electronics.* 2025;14(5):917. [[CrossRef](#)].
14. Jin B, Xing J, San Y, Hou M, Wan J. Equivalent model research of photovoltaic powerstation based on AP clustering algorithm. *Acta Energiæ Solaris Sin.* 2025;46(12):186–94. (In Chinese). [[CrossRef](#)].
15. Yang S, Hu X, Wang L, Sun G, Shi M, Meng Z, et al. Distributed photovoltaic-storage hierarchical aggregation method based on multi-source multi-scale data fusion. *Electronics.* 2026;15(2):464. [[CrossRef](#)].
16. Zhang T. Dynamic equivalent model of two-staged distributed photovoltaic cluster based on multiple clustering algorithms. *IET Circuits Devices Syst.* 2026;2026(1):8895067. [[CrossRef](#)].
17. Qu D, Yao Y, Xiao Y, Ye W, Ding G, Wang Q. Short term output prediction of distributed photovoltaic grid connected generation in low-voltage distribution networks based on spectral clustering. *J Phys Conf Ser.* 2025;2979(1):012013. [[CrossRef](#)].
18. Wu F, Sun W, He G, Xiong J. Voltage control method of distributed PV based on voltage sensitivity. *Acta Energiæ Solaris Sin.* 2024;45(11):313–9. (In Chinese). [[CrossRef](#)].
19. Liu DS, Xiong SH, Song ZW, Qiu HJ. Review of harmonic suppression technology based on photovoltaic grid-connected inverter. *J Electr Eng.* 2025;20(6):223–38. (In Chinese). [[CrossRef](#)].
20. Shi J, Hu S, Fu R, Zhang Q. Convex optimization and PV inverter control strategy-based research on active distribution networks. *Energies.* 2025;18(7):1793. [[CrossRef](#)].
21. Chen J, Zhao B, Zhang F, Hu J, Zhang L. Identification of high and low voltage ride-through control parameters for electromechanical transient modeling of photovoltaic inverter. *Distrib Gener Altern Energy J.* 2024:195–220. [[CrossRef](#)].
22. Chen X, Li S, Wang F, Li J, Tang C. Power estimation method of low-voltage distributed photovoltaic generation based on similarity aggregation. *Energy Rep.* 2021;7:1344–51. [[CrossRef](#)].
23. Itote FM, Shigenobu R, Takahashi A, Ito M, Faggianelli GA. Cluster-driven multi-objective PV curtailment for fair and efficient voltage regulation. *Int J Sustain Energy.* 2025;44(1):2586928. [[CrossRef](#)].
24. Li W, Yang X, Zhao X, Xiong X. Photovoltaic equivalent modeling considering inverter parameters and distribution network characteristics. *Power Electron.* 2025;59(7):53–61.