doi: 10.20532/cit.2024.1005870

Improved Multiobjective Genetic Algorithm for Partitioning Distributed Photovoltaic Clusters: Balancing Spatial Distance and Power Similarity

Yansen Chen[1], Kai Cheng[1], Zhuohuan Li[1], Shixian Pan[1, 2] and Xudong Hu[1, 2]

[1]China Southern Grid Digital Grid Research Institute Co. Ltd., Guangzhou, China
[2]China Southern Power Grid Artificial Intelligence Technology Co. Ltd., Guangzhou,China

The prediction of power output from photovoltaic generation clusters is crucial for optimizing the dispatch of regional photovoltaic generation. Enhancing the accuracy of power prediction for photovoltaic power plant clusters requires the segmentation of distributed photovoltaic systems into clusters. This paper proposes a method for partitioning distributed photovoltaic clusters using a multiobjective genetic algorithm NSGA-II, with spatial distance modularity and electricity similarity as optimization objectives to determine the optimal cluster partitioning scheme. The numerical examples and experimental results of the case analysis demonstrate a significant improvement in the convergence speed of the prediction system when employing the clustering partitioning method. This cluster segmentation algorithm significantly reduces the complexity and investment cost of the prediction system.

ACM CCS (2012) Classification: Computing methodologies → Artificial Intelligence → Planning and Scheduling

Keywords: multiobjective genetic algorithm, distributed photovoltaic, cluster partitioning

1. Introduction

The output of photovoltaic generation is influenced by meteorological factors such as actual irradiance, temperature, wind speed, and humidity, resulting in intermittent, random, and fluctuating behavior [1]. With the promotion of photovoltaics throughout the county, a substantial number of photovoltaic systems on the user side of the power grid are being connected to the distribution network. This trend leads to two distinct modes of photovoltaic power generation: (1) large-scale decentralized development, low-voltage connection, and local consumption, and (2) large-scale centralized development, medium and high voltage connection, and high-voltage long-distance transmission and reception [2]. Since these developments will profoundly impact the power system, accurate prediction of PV power output is essential for power system administrators to adjust the generation plan of PV power plants or develop on-site PV consumption strategies in advance, thereby achieving balanced grid dispatch.

Based on historical output data, numerical weather prediction (NWP) data, and actual meteorological data, a photovoltaic power prediction model is established to forecast future photovoltaic output power. Currently, photovoltaic power forecasting can be categorized into centralized photovoltaic forecasting and distributed photovoltaic forecasting [3].

In the early stages of photovoltaic (PV) technology development, large-scale centralized PV power plants were the primary focus. These centralized PV farms are typically located in remote areas such as deserts or open farmland, where abundant sunlight resources are available. Centralized PV plants require substantial land areas and a unified grid connection, relying on centralized generation and transmission systems to deliver electricity. While centralized

PV systems can provide large-scale clean energy, they depend on long-distance transmission, leading to power loss and higher grid stability requirements. This is particularly challenging in regions with significant fluctuations in power demand, where centralized systems may struggle to respond flexibly.

In contrast, distributed PV systems enable users to generate and consume their own electricity, and any surplus power can be fed back into the grid through small-scale grid connections, enhancing both system flexibility and cost-efficiency. This model effectively reduces reliance on long-distance transmission, minimizing transmission losses and costs.

For centralized PV forecasting, the primary methods include direct methods and data-driven methods. In recent years, big data and artificial intelligence techniques such as component decomposition, clustering, and neural networks [4-6] have gained significant attention. These methods leverage additional information like satellite and ground-based cloud images to enhance the accuracy of ultrashort-term forecasts.

In distributed photovoltaic forecasting, the centralized photovoltaic forecasting approach can serve as a reference. However, there are notable disparities between distributed photovoltaic stations and centralized photovoltaic stations concerning data availability and predictability. Existing centralized photovoltaic forecasting models rely on local data with consistent spatiotemporal resolution, including actual measurements, forecasts, weather, and electrical data. Yet, a considerable portion of distributed photovoltaic systems lack such comprehensive data conditions. For instance, over 80% of household photovoltaic installations connected to low-voltage distribution networks possess only measured irradiance and NWP data, along with daily electricity consumption data from the electricity distribution marketing system, making it challenging to directly replicate mature technological approaches [7]. Due to data limitations, distributed photovoltaic systems are unable to conduct high-level short-term forecasting akin to centralized photovoltaic installations. An effective approach involves aggregating all photovoltaic stations with similar characteristics into a single large photovoltaic station, where the equivalent photovoltaic station can perform power prediction.

Currently, the utilization of clustering algorithms to partition distributed photovoltaic station clusters [8] represents one of the conventional methods, which primarily encompasses the following approaches:

a) The K-means clustering algorithm [9] iteratively identifies clustering centers, assigning data points to their nearest centroids. To address the challenge of rapid photovoltaic power fluctuations, Tai *et al.* [10] proposed an index-weighted K-means++ algorithm. This method focuses on optimizing economic utilization and coordinated control of distributed resources. By computing the weighted Euclidean distance matrix, it effectively groups resources and minimizes regulation costs, thereby mitigating the impact of power fluctuations through resource aggregation. Wu *et al.* [11] utilized an improved K-means++ algorithm to aggregate a significant portion of distributed photovoltaic resources into distinct clusters. They constructed an aggregation-peak regulation-decomposition model to tackle optimization challenges in power system peak regulation. Additionally, studies [12] have combined the K-means clustering algorithm with the gravity model to propose an enhanced clustering model for load clustering, optimizing demand-side resources efficiently. In other research [13], the improved K-means clustering method was employed to determine power generation states and establish uncertainty models for wind and photovoltaic power generation, selecting an appropriate probability density function for fitting. Furthermore, considering electrical distance, cluster power balance, and cluster size, a comprehensive cluster partition index system was proposed, and the K-means algorithm was enhanced by integrating the gray wolf optimization algorithm with the Levy flight strategy for cluster partitioning [14]. Chen *et al.* [15] improved the K-means clustering algorithm using the gray wolf optimization algorithm to analyze datasets containing output values of photovoltaic power stations and environmental parameters, aiming to achieve the best dynamic supply-demand balance in distributed photovoltaic station clusters. Liu *et al.* [16] discussed cluster division and reactive power optimization for large-scale distributed generation. They utilized the K-means clustering algorithm improved by IGA to identify better initial clustering centers and divide the distribution network based on electrical distances between nodes for different clusters. However, the K-means algorithm still faces challenges in determining the optimal number of clusters and is highly sensitive to the selection of initial cluster centers, potentially leading to instability in photovoltaic cluster division results. Zhang *et al.* [17] using a modified AP-TD-K-medoids trilevel clustering algorithm that was designed to cluster and partition the distribution network.

b) The Fast Unfolding Clustering Algorithm [18] is a modularity optimization-based community detection algorithm designed for efficiently clustering large-scale networks. In [19], the fast unfolding clustering algorithm was applied to partition photovoltaic power clusters. The analysis considered the correlation characteristics between photovoltaic power supply and load at the access node. Division indicators included the net load of the node and the active and reactive power regulation capacity of the photovoltaic power supply, taking into account electrical distance characteristics. However, despite its effectiveness, there remains an efficiency issue when processing large-scale data.

c) The Density Peak-based Clustering Algorithm (DPC) [20] identifies cluster centers by locating high-density areas' center points, making it suitable for detecting clusters of arbitrary shapes. In a study by Zhang and Shi [21], the electrical distance between nodes is computed based on comprehensive voltage sensitivity, and a node similarity matrix is constructed accordingly. The DPC algorithm then rapidly partitions the distribution network using a fast search and discovery method. Given the lack of local photovoltaic independent voltage regulation ability, a coordinated control strategy for voltage partitioning, starting with reactive power and then active power, is proposed. Recognizing the complexity of the centralized voltage control method and the time-consuming nature of traditional partitioning methods [22], the authors first analyze the impact of active and reactive power on distribution network voltage post-installation of distributed photovoltaic power. Then, the authors of the study employ combined sensitivity as electrical distance to determine the proximity of electrical connections between nodes. Subsequently, the calculated electrical distance serves as input for the DPC clustering algorithm, facilitating rapid zoning of the distribution network. However, since the DPC algorithm's efficacy can be significantly influenced by the subjective definition of cutoff distance during application, authors in [23] propose a K-nearest neighbor optimization DPC algorithm (KNN-DPC). This approach employs a comprehensive sensitivity matrix to represent electrical distance, thereby accelerating the search for density peaks, mitigating the impact of cutoff distance, and enhancing clustering accuracy. Nevertheless, the DPC algorithm may encounter challenges when processing data with non-uniform density distributions, especially in scenarios with non-uniform distribution of photovoltaic generation.

d) The Fuzzy C-means clustering algorithm (FCM) [24] is a flexible soft clustering method that permits a single data point to belong to multiple clusters simultaneously, assigning a certain degree of membership to each cluster center. In the context of dynamic grouping modeling for regional centralized photovoltaic generation systems, Sheng *et al.* [25] employ an improved fuzzy C-means clustering algorithm to obtain clustering results for photovoltaic generation units under varying operating conditions, enabling dynamic unit clustering. Chen *et al.* [26] utilize the traditional fuzzy C-means clustering algorithm for photovoltaic subclustering to address fault diagnosis issues in large photovoltaic arrays. Considering the spatial distribution and source-load coupling of distributed photovoltaics, Hu *et al.* [27] propose a multivariate performance index system for group classification. They enhance the traditional FCM algorithm by utilizing the point density function to initialize clustering centers and constructing clusters based on the cohesive coupling degree of information granularity. A clustering validity function is formulated to evaluate cohesion and coupling degrees, and the improved FCM algorithm is employed to achieve reasonable grouping results. In developing cluster partition strategies, various factors including electrical distance, regulation difficulty, and regulation cost are comprehensively considered, and indicators of different dimensions are unified. A partition method based on fuzzy clustering is proposed in [28]. However, the FCM algorithm may lack robustness when dealing with noise and outliers, and its computational cost is relatively high. This poses challenges in real-time or photovoltaic cluster partition applications that necessitate rapid response times (Table 1).

Additionally, employing intelligent optimization algorithms for equivalent modeling of distributed photovoltaic clusters is an important research direction. Lv *et al.* [29] utilized Particle Swarm Optimization (PSO), Genetic Algorithm (GA), and the K-means clustering algorithm for cluster division. Meng *et al.* [30] introduced a genetic algorithm to partition distributed photovoltaic clusters in a distribution network, using the electrical distance-based modularity index and active power balance index as comprehensive classification indicators. Chen *et al.* [31] proposed a framework for distributed photovoltaic cluster delineation based on federated learning synthetic clustering. Wu *et al.* [32] improved modular increment by selecting an appropriate clustering partition index, employed the Fast Newman algorithm to partition the distribution network into multiple clusters and verified the clustering effectiveness on the IEEE 33-bus standard system. Wang and Gao [33] used Smart Local Moving (SLM) to overcome the problem of challenging regulation of distributed power supply in modern distribution networks. Liu *et al.* [34] proposed an optimal cluster partitioning method based on a graph-based genetic algorithm (GA). In this method, a novel graph-based structure is proposed for representing chromosomes, and improvements are introduced to the evolutionary selection, crossover, and mutation processes. These enhancements facilitate the generation of a search population, which is employed to partition distributed photovoltaic (PV) power grids into distinct clusters. Huang *et al.* [35] applied an improved genetic algorithm to search for the optimal partition scheme. Li *et al.* [36] introduced a method for partitioning and dynamically adjusting distributed photovoltaic clusters, based on an improved adaptive genetic algorithm (AGA).

In summary, currently, there is limited research on the use of multiobjective genetic algorithms (NSGA-II) for distributed PV system clustering. To address this gap and enhance the efficiency and accuracy of PV cluster partitioning, which in turn optimizes power generation planning and grid balancing, this paper proposes a distributed cluster partitioning method based on the multiobjective genetic algorithm (NSGA-II). This method provides a solid foundation to address the broad application needs of distributed photovoltaic generation. Acknowledging the shortcomings in the convergence speed and parameter selection of the standard NSGA-II algorithm, this paper introduces

Table 1. Deficencies of traditional methods.

Method	**Deficiencies**
K-means	The results of the partition are unstable
Fast Unfolding	Less efficient
DPC	Dealing with data with uneven density distribution is challenge
FCM	Lack of robustness when dealing with noise and outliers

adaptive mutation and crossover operators to enhance its performance. The improved algorithm exhibits higher computational efficiency compared to the standard NSGA-II algorithm while also improving solution accuracy and convergence through the incorporation of adaptive crossover and mutation probabilities. Additionally, by employing a strategy that extracts historical optimal solutions, the algorithm effectively avoids the pitfall of converging to local optima.

2. Objective Function

To facilitate cluster power prediction structurally, synoptic characteristics within each cluster should be tightly coupled, while inter-cluster weather characteristics should be loosely coupled. Spatial distance modularity serves as the structural indicator. Functionally, to ensure consistent output characteristics and enhance prediction accuracy, the similarity of active power is utilized as the evaluation index. Initially, assume the presence of N nodes (comprising distributed photovoltaic power stations), ultimately segmented into N_c groups (for cluster prediction).

2.1. Spatial Distance Modularity ϕ

Distributed photovoltaic clusters and community structures share similar properties in terms of resource sharing and modularity. Consequently, the division index of complex networks can be applied. Spatial distance modularity is employed to assess the reasonableness of cluster division results. A higher value indicates a more reasonable division; when the entire network is grouped into one cluster, modularity is 0. Conversely, if each node is allocated to different clusters, modularity is negative, indicating the absence of a community structure in such a scenario.

Modularity ϕ is defined by Equation 1 as follows:

$$\begin{cases} \phi = 1 - \dfrac{1}{m}\sum\limits_{i}\sum\limits_{j} {A_{ij}}^2 \delta(i,j) \\ m = \sum\limits_{i}\sum\limits_{j} {A_{ij}}^2 \end{cases}, \qquad (1)$$

where A_{ij} represents the distance connecting nodes i and j, and δ is a 0-1 matrix, The variable m represents the sum of squared distances between all node pairs. If node i and node j are located in the same cluster, then $\delta(i, j) = 1$; if they are not in the same cluster, then $\delta(i, j) = 0$. If all the distributed photovoltaics are in one cluster, then the modularity is 0, which is the worst case. If each cluster contains only one node, then the modularity is 1, but that is not what we need at this time. Therefore, our algorithm specifies the maximum number of clusters.

2.2. Electricity Similarity φ

In terms of functionality, to enhance the accuracy of distributed power forecasting, power similarity is utilized to measure the generation similarity within the region, serving as a key metric for cluster partitioning. Initially, two characteristic quantities are defined for each photovoltaic power plant.

Maximum power W_{max}: The peak power output of the generating station during the training period.

Average power percentage W_{ave}: The ratio of the average power output to the maximum power output of the generating station during the training period.

Electricity median percentage R_{50}: The ratio of the median power output to the maximum power output of the generating station during the training period.

The power factor r is defined by Equation 2:

$$r = \sqrt{\frac{W_{50}^2 + W_{ave}^2}{2}} \qquad (2)$$

The definition of electric quantity similarity is given by Equation 3:

$$\begin{cases} \varphi_c = \dfrac{1}{k}\sum_{c=1}^{k} \dfrac{r}{\max r} \\ \varphi = \dfrac{100}{N_c}\sum_{cc \in M} \varphi_c \end{cases} \qquad (3)$$

where: N_c represents the number of clusters to divide; M represents the set of all clusters; φ_c represents a cluster c with an active power similarity degree; k is the number of distributed photovoltaics in cluster; r represents a cluster with a similarity coefficient of all power stations; and φ represents the active similarity of the overall network.

The values of spatial distance modularity and active power similarity are bounded between 0 and 1. For a cluster, higher internal cohesion, indicated by closer geographic locations, leads to a spatial distance modularity closer to 1. Similarly, greater similarity among quantities within the cluster results in a higher active power similarity degree index approaching 1. Both indicators are expected to be of the same order of magnitude.

2.3. Objective Function Formula

The expression for the comprehensive performance index is provided by Equation 4:

$$max\rho = w_1\phi + w_2\phi \tag{4}$$

where w_1 and w_2 represent the weights assigned to the modularity index and the active similarity index, respectively. In this context, the similarity in capacity is selected as the primary optimization objective.

3. Solution Method

3.1. Multiobjective Genetic Algorithm

The steps of the proposed method for partitioning distributed photovoltaic clusters using a multiobjective genetic algorithm NSGA-II is shown in Figure 1.

1. At the outset of the calculation and solving process, the information regarding the PV nodes needs to be obtained, along with the relevant parameters set, including:
   a) Pc.agv is the average crossover probability, with Pc.max and Pc.min denoting the maximum and minimum crossover probabilities, respectively.
   b) Pm.agv is the average mutation probability, with Pm.max and Pm.min representing the maximum and minimum mutation probabilities, respectively.
   c) NG is the maximum number of iterations.
   d) Group size Q (inclusive of the number of individuals).

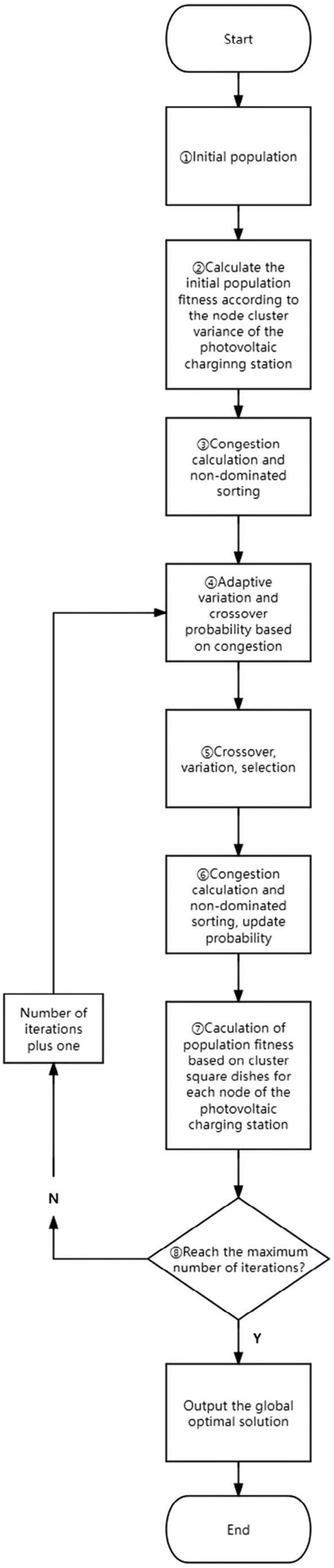
Start
①Initial population
②Calculate the initial population fitness according to the node cluster variance of the photovoltaic charginng station
③Congestion calculation and non-dominated sorting
④Adaptive variation and crossover probability based on congestion
⑤Crossover, variation, selection
⑥Congestion calculation and non-dominated sorting, update probability
⑦Caculation of population fitness based on cluster square dishes for each node of the photovoltaic charging station
Number of iterations plus one
N
⑧Reach the maximum number of iterations?
Y
Output the global optimal solution
End

Figure 1. Algorithm flow of the proposed mehod.

The node information primarily comprises the distance between substations, the average electricity percentage Wave, and the median percentage of electricity for each photovoltaic power station.

2. Initial Population: Following the initial settings, an initialization population is generated by utilizing the state of each node as an object to formulate a cluster scheme of photovoltaic power plants with the specified number of populations. Specifically, N nodes are randomly assigned to Nc clusters, resulting in the generation of Q allocation schemes. Each allocation scheme can be created through randomization. In this step, Q cluster solutions are inputted.
3. Calculate the initial population fitness based on the clustering scheme of each node of the photovoltaic charging station.
   a) Perform congestion degree calculation and non-dominated sorting to obtain the optimal solution set. If the optimal solution set stabilizes, save the extreme points; otherwise, adjust the crossover and mutation probabilities of each individual based on the crowding degree, *i.e.*, apply the adaptive crossover mutation operator. Then, perform crossover, mutation, and selection to reset the optimal solution set and continue optimization.
   b) Upon reaching the maximum number of iterations, output the global optimal solution by comparing it to the historical optimal solutions, thus concluding the optimization process.

The specific algorithm process is as follows:

1. Using the node information obtained earlier, an initial population of clustering solutions is generated. The states of each node are used as objects for forming clustering schemes of the photovoltaic charging stations. N nodes are randomly assigned to Nc clusters, forming Q clustering schemes. These schemes are essentially different ways of grouping the nodes and can be generated randomly. At the end of this step, Q cluster solutions are inputted.
2. Each clustering scheme's fitness is calculated based on the cluster variance of the photovoltaic charging station nodes. This metric helps to evaluate how well the clustering solution works in terms of energy distribution and other factors.
3. Perform congestion degree calculation and non-dominated sorting to identify the best solution set from the current population. Non-dominated sorting is used to categorize solutions based on whether they are dominated by others in terms of multiple objectives.
4. In this step, adaptive mechanisms are applied. Depending on the congestion (how crowded a particular region of solutions is), the algorithm adjusts the crossover and mutation probabilities of each individual in the population. If a region is highly congested, the probabilities are adjusted to introduce more diversity. This step enhances the GA by allowing it to focus more on exploring less congested areas of the solution space.
5. With the newly adjusted crossover and mutation probabilities, the algorithm performs crossover, mutation, and selection operations.Crossover involves combining parts of two solutions to form a new one, while mutation introduces random changes to individual solutions. Selection ensures that only the fittest individuals are carried over to the next generation.
6. After the new population is generated, the congestion calculation and non-dominated sorting are repeated to update the population's fitness and probabilities for the next iteration.
7. For each individual in the updated population, recalculate its fitness based on the clustering variance of the nodes in the photovoltaic charging station. This helps assess whether the clustering solutions are improving with each iteration.
8. The algorithm checks whether the maximum number of iterations NG has been reached. If the number of iterations has not yet been exceeded, the process returns to step 5 for further optimization (incrementing the number of iterations by one).

9. At the end of the iterations, the algorithm compares the solutions generated in all previous generations and selects the global optimal solution from the set of historical best solutions. This final solution represents the best clustering scheme for the photovoltaic charging stations based on the specified objectives.If the maximum number of iterations is reached, the algorithm proceeds to the final step.

3.2. Improved Multiobjective Genetic Algorithm

In the application of the NSGA-II algorithm, the parameter settings of the crossover and mutation probability have a great impact on the optimization result, and different parameter combinations will have different influences on the optimization result. Crossover and mutation are important methods for the NSGA-II algorithm to generate new individuals. Blindly performing crossover and mutation on individuals may destroy the optimal solution. Setting the crossover mutation probability too small will lead to a slow optimization speed, and it is easy to become trapped in a local optimal solution. Therefore, this paper adds adaptive crossover and mutation operators to ensure that there is a higher crossover and mutation probability in the early stage of the iteration to enhance the global search ability; in the later stage, the crossover and mutation probabilities tend to the average value, and the local search is emphasized to obtain the optimal solution. Set. The improved crossover probability P_c and mutation probability P_m are shown in Equation 5:

$$P_c=\begin{cases} P_{c.agv}+\left(\dfrac{NG-k}{NG}\right)^3\times\left(P_{c.max}-P_{c.agv}\right), & d_j(k)<d_{agv}(k) \\ P_{c.agv}, & d_j(k)=d_{agv}(k) \\ P_{c.agv}-\left(\dfrac{k}{NG}\right)^3\times\left(P_{c.agv}-P_{c.min}\right), & d_j(k)>d_{agv}(k) \end{cases}$$

$$P_m=\begin{cases} P_{m.agv}+\left(\dfrac{NG-k}{NG}\right)^3\cdot\left(P_{m.max}-P_{m.agv}\right), & d_j(k)<d_{agv}(k) \\ P_{m.agv}, & d_j(k)=d_{agv}(k)\ , \\ P_{m.agv}-\left(\dfrac{k}{NG}\right)^3\cdot\left(P_{m.agv}-P_{m.min}\right), & d_j(k)>d_{agv}(k), \end{cases} \tag{5}$$

where: $P_{c.agv}$ is the average crossover probability; $P_{c.max}$ and $P_{c.min}$ are the maximum and minimum crossover probabilities, respectively; $P_{m.agv}$ is the average mutation probability; $P_{m.max}$ and $P_{m.mix}$ are the maximum and minimum mutation probabilities, respectively; NG is the maximum number of iterations; k is the current iteration number; $d_j(k)$ is the crowding degree of the j-th individual of the k-th generation population; $d_{agv}(k)$ is the average crowding distance of the population; $P_{c.agv}$ $P_{c.max}$, $P_{c.mix}$ and NG are all preset parameters.

Current methods for assessing the stability of the optimal solution set typically involve monitoring the distance of the Pareto optimal solution set over past iterations. As the distance decreases, the optimal solution set tends to stabilize. However, since this paper prioritizes user's power similarity as the primary optimization goal, a modification is made to the stability assessment criterion. Instead of solely relying on the distance between the Pareto optimal solutions, the stability judgment condition is now based on the smallest power similarity between two optimal solutions separated by several generations. Throughout the optimization process, the distance to the optimal solution gradually diminishes. Once the judgment condition is met and an extreme point emerges, it is preserved as the historical optimal solution. Subsequently, the optimal solution set is reset to continue optimization. Upon reaching the maximum number of iterations, the global optimal solution is output after comparing the historical optimal solutions, thereby concluding the optimization process.

Adaptive mutation adjusts the mutation intensity dynamically, allowing the algorithm to avoid local optima by responding to varying problem complexity or population diversity at different stages. Meanwhile, adaptive crossover adjusts the crossover rate based on the population's evolutionary state. This adaptive adjustment enables the algorithm to balance exploration and exploitation, enhancing the efficiency of both global search and local optimization.

4. Numerical Example Analysis

4.1. Parameter Setting

The data analyzed in this paper are sourced from the distributed photovoltaic generation data of a county in Guizhou. The dataset comprises the annual generation data of 20 photovoltaic power stations in the year 2022. Specifically, each photovoltaic substation contains 365 days of generation data, though a small amount of data may be missing for some stations. Table 2 presents the characteristic data of the 20 photovoltaic power plants, compiled from the original dataset. The horizontal and vertical axes represent the horizontal and vertical distances, respectively, between each photovoltaic installation location and the reference point. Following the selection of a local reference point, the unit of measurement is kilometers (km).

Table 2. Characteristic data of 20 photovoltaic power plants.

Site ID	Mean/maximum value	Median/maximum value	Horizontal axis (km)	Vertical axis (km)
1	0.9724	0.9732	10.562	13.650
2	0.9749	0.9742	1.985	−1.540
3	0.9748	0.9757	4.613	−1.358
4	0.9780	0.9769	13.231	2.931
5	0.9786	0.9773	9.034	−0.433
6	0.9783	0.9781	5.552	3.040
7	0.9792	0.9789	8.252	2.261
8	0.9806	0.9793	3.704	3.051
9	0.9805	0.9799	9.076	4.306
10	0.9811	0.9804	7.812	1.330
11	0.9822	0.9811	6.193	3.139
12	0.9822	0.9813	3.414	4.263
13	0.9836	0.9829	−3.886	8.017
14	0.9838	0.9835	3.208	9.628
15	0.9844	0.9846	4.283	8.404
16	0.9845	0.9850	−1.055	9.595
17	0.9857	0.9851	4.725	7.991
18	0.9907	0.9906	4.409	7.054
19	0.9900	0.9915	7.458	7.826
20	0.9922	0.9920	2.996	4.209

4.2. Evaluation Indicators

To better evaluate the effectiveness of the model and algorithm, two evaluation indicators are employed for effectiveness assessment:

1. The root mean square error (RMSE) is utilized to gauge the deviation between the predicted value and the actual value. The expression for RMSE is depicted in Equation 6:

$$e_{\mathrm{RMSE}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_{fi} - y_{ai}\right)^2} \tag{6}$$

2. Mean absolute error (MAE): The average value of the absolute errors effectively reflects the actual situation of the predicted value errors. The expression for MAE is given by Equation 7:

$$e_{\mathrm{MAE}} = \frac{1}{N}\sum_{i=1}^{N}\left|y_{fi} - y_{ai}\right| \tag{7}$$

where y_{fi} represents the predicted value and y_{ai} represents the true value.

4.3. Experimental Result Analysis

4.3.1. Analysis of Cluster Results

Utilizing the aforementioned algorithm, the 20 stations are divided into 6 clusters. The results are illustrated in Figure 2, where each cluster comprises 2, 2, 2, 4, 5, and 5 photovoltaic power plants, respectively. The partition depicted in the figure aligns well with the actual photovoltaic distribution in the field. As shown in Figure 2, each point represents a power station, with points of the same color indicating stations that have been grouped into the same category. From the clustering results, neighboring stations are well-classified into the same group. Taking geographical factors into account, the two stations marked in red have also been correctly classified into the same category. The experimental outcomes demonstrate that this method effectively addresses the cluster division problem, furnishing a robust foundation for subsequent photovoltaic power prediction development.

Moreover, with the incorporation of adaptive mutation and crossover operators, the convergence speed is enhanced, enabling the attainment of optimized results within 50 iterations.

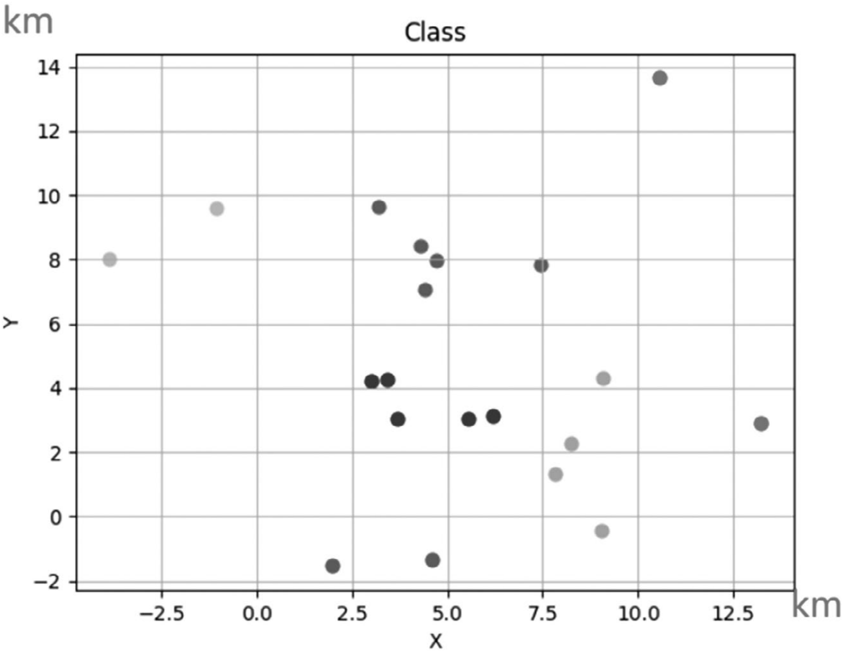

Figure 2. The cluster partition result.

4.3.2. Analysis of Forecast Results

To demonstrate the superiority of clustering, two schemes were employed to validate the prediction effectiveness of the cluster prediction method:

Option 1. No cluster partitioning is performed, and a single station aggregation method is used. The prediction model is BPNN.

Option 2. Cluster partitioning is performed, and the typical stations obtained using the proposed method are aggregated. The prediction model is BPNN.

Option 3. Cluster partitioning is performed, and the typical stations are obtained using the optimal settings of adaptive crossover and mutation operators, followed by aggregation. The prediction model is BPNN.

Option 4. Cluster partitioning is performed, with the typical stations obtained using the GA algorithm, and aggregation is used. The prediction model is BPNN.

The predicted results of the four options are shown in Table 3.

The experimental findings indicate that, under the same forecasting scheme, there is no substantial difference between the errors of cluster forecasting and single-station cumulative forecasting. However, in comparison to the single-site cumulative method, the clustering algorithm can notably reduce the complexity and investment costs of the forecasting system, demonstrating its clear superiority. Similarly, the results obtained using the adaptive operators in this paper are significantly better than those using the predefined optimal operators. Compared to the GA algorithm, the method proposed in this paper demonstrates superior performance.

5. Conclusion

In this paper, the partitioning of distributed photovoltaics into clusters is investigated, and a partitioning method is proposed:

1. Two indicators, spatial distance modularity, and electric quantity similarity, are introduced to balance the cluster division.

2. A multiobjective genetic algorithm is proposed to address the problem. The standard NSGA-II algorithm is known for its slow convergence speed and the challenge of parameter selection. To address this, adaptive mutation and crossover operators are incorporated into the standard NSGA-II algorithm to enhance convergence speed. Compared to the standard NSGA-II algorithm, the improved version exhibits greater computational efficiency. The addition of adaptive crossover and mutation probabilities enhances solution accuracy and convergence of the algorithm. By leveraging historical optimal solutions, the issue of falling into a local optimal solution is mitigated.

Table 3. Cluster power prediction results for different experimental schemes.

Scheme	RMSE	MAE
Option I	0.1283	0.06842
Option II	0.1284	0.06735
Option III	0.1174	0.06624
Option IV	0.1233	0.06691

Simulation results and practical applications demonstrate that the method proposed in this paper effectively divides distributed photovoltaic clusters, yielding division results consistent with real-world scenarios. This research can be applied to distributed photovoltaic power prediction in the future to enhance accuracy, further reduce the gap between predicted and actual power output, and improve the overall safety of grid operations. This study conducted simulation experiments solely on the proposed method. In future work, the method will be applied to distributed photovoltaic power prediction to further verify its effectiveness.

Acknowledgement

This research was funded by China Southern Power Grid Science and Technology Project (Grant NO.: 670000kk52210042).

References

[1] W. Chen *et al.*, "Influence of Grid-connected Photovoltaic System on Power Network", *Electric Power Automation Equipment*, vol. 33, no. 2, pp. 26–32, 2013.

[2] M. Ding *et al.*, "A Review on the Effect of Large-scale PV Generation on Power Systems", *Proceedings of the CSEE*, vol. 34, no. 1, pp. 1–14, 2014. https://doi.org/10.13334/j.0258-8013.pcsee.2014.01.001

[3] X. M. Zhang *et al.*, "Overview for Large-scale Distributed Photovoltaic Power Prediction with Clustered Method", *North China Electric Power*, vol. 6, no. 7, pp. 1–7, 2017. https://doi.org/10.16308/j.cnki.issn1003-9171.2017.06.001

[4] Q. B. Sun *et al.*, "Enhancing Session-based Recommendations with Popularity-aware Graph Neural Networks", *Acadlore Transactions on AI and Machine Learning*, vol. 1, no. 1, pp. 22–29, 2022. https://doi.org/10.56578/ataiml010104

[5] B. Li *et al.*, "Research on Prediction of Regional Distributed Photovoltaic Output Considering Spatial Relevance", *Power Technology*, vol. 45, pp. 1048–1051, 2021.

[6] N. S. Divya and R. Vatambeti, "Detecting False Data Injection Attacks in Industrial Internet of Things Using an Optimized Bidirectional Gated Recurrent Unit-swarm Optimization Algorithm Model", *Acadlore Transactions on AI and Machine Learning*, vol. 2, no. 2, pp. 75–83, 2023. https://doi.org/10.56578/ataiml020203

[7] Y. Qiao *et al.*, "Distributed Photovoltaic Station Cluster Gridding Short-term Power Forecasting Part I: Methodology and Data Augmentation", *Power System Technology*, vol. 5, pp. 1799–1808, 2021. https://doi.org/10.13335/j.1000-3673.pst.2021.0305

[8] M. Shi *et al.*, "Overview of Flexible Grid-connected Cluster Control Technology for Distributed Photovoltaic", *Electr. Meas. Instrum*, vol. 58, pp. 1–9, 2021.

[9] J. C. Yang and C. Zhao, "Survey on K-Means Clustering Algorithm", *Computer Engineering and Applications*, vol. 55, no. 23, pp. 7–14, 2019.

[10] B. Tai *et al.*, "Smoothing Control Method of Distributed Photovoltaic Power Fluctuation Based on the Index Weighted K-means++ Algorithm", *Engineering Journal of Wuhan University*, vol. 56, no. 11, pp. 1413–1424, 2023. https://doi.org/10.14188/j.1671-8844.2023-11-012

[11] Y. Wu *et al.*, "A Peak Shaving Method of Aggregating the Distributed Photovoltaics and Energy Storages Based on the Improved K-means++ Algorithm", *Power Syst. Technol*, vol. 46, pp. 3923–3931, 2022. https://doi.org/10.13335/j.1000-3673.pst.2021.1555

[12] S. X. Guo *et al.*, "Optimal Construction of Industrial Park Load Polymers Concerning the Distributed Photovoltaic and Electric Vehicles Based on Improved K-means Clustering", *Electric Power Science and Engineering*, vol. 34, no. 3, pp. 14–21, 2018.

[13] X. W. Song *et al.*, "Wind and Photovoltaic Generation Scene Division Based on Improved K-means Clustering", *Power Generation Technology*, vol. 41, no. 6, pp. 625–630, 2020. https://doi.org/10.12096/j.2096-4528.pgt.19090

[14] T. Z. Wei *et al.*, "Distributed Power Cluster Partitioning Method Based on LGWO Improved K-means Clustering Algorithm", *Journal of North China Electric Power University (Social Sciences)*, pp. 1–9, 2023.

[15] S. Chen *et al.*, "Cluster Dynamic Partitioning Strategy Based on Distributed Photovoltaic Output Prediction and Improved Clustering Algorithm", in *2023 IEEE/IAS Industrial and Commercial Power System Asia (I&CPS Asia)*, pp. 1630–1635, IEEE, July 2023. https://doi.org/10.1109/ICPSAsia58343.2023.10294606

[16] Y. Liu *et al.*, "Distributed Photovoltaic Cluster Partition and Reactive Power Optimization Strategy Based on BAS-IGA Algorithm", in *2022 7th Asia Conference on Power and Electrical Engineering (ACPEE)*, Hangzhou, China, pp. 2198–2203, IEEE, April 2022. https://doi.org/10.1109/ACPEE53904.2022.9784084

[17] Y. Zhang *et al.* "Planning Strategies for Distributed PV-Storage Using a Distribution Network Based on Load Time Sequence Characteristics Partitioning", *Processes*, vol. 11, no. 2, p. 540, 2023.
https://doi.org/10.3390/pr11020540

[18] V. D. Blondel *et al.*, "Fast Unfolding of Communities in Large Networks", *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 10, p. 10008, 2008.
https://doi.org/10.1088/1742-5468/2008/10/P10008

[19] L. Wang *et al.*, "Large-scale Distributed PV Cluster Division Based on Fast Unfolding Clustering Algorithm", *Acta Energiae Solaris Sinica*, vol. 42, no. 10, pp. 29–34, 2021.
https://doi.org/10.19912/j.0254-0096.tynxb.2018-0896

[20] X. Xu *et al.*, "Survey on Density Peaks Clustering Algorithm", *Journal of Software*, vol. 33, no. 5, pp. 1800–1816, 2022.
https://doi.org/10.13328/j.cnki.jos.006122

[21] Y. N. Zhang and Z. Shi, "Voltage Partition Coordinated Control of Distribution Network with Distributed Photovoltaic Based on CFSFDP Algorithm", *Modern Electric Power*, vol. 37, no. 1, pp. 35–41, 2020.
https://doi.org/10.19725/j.cnki.1007-2322.2019.0263

[22] H. Chen *et al.*, "Distributed Photovoltaic Power Cluster Partition Based on DPC Clustering Algorithm", in *Proc. of the 2022 4th International Conference on Electrical Engineering and Control Technologies (CEECT)*, 2022, pp. 974–979.
https://doi.org/10.1109/CEECT55960.2022.10030647

[23] Z. Wu *et al.*, "Distributed PV Cluster Partition Based on KNN-DPC Clustering Algorithm", in *Proc. of the 2023 5th International Conference on Electrical Engineering and Control Technologies (CEECT)*, 2023, pp. 1–5.
https://doi.org/10.1109/CEECT59667.2023.10420687

[24] S. Yadav *et al.*, "Task Allocation Model for Optimal System Cost Using Fuzzy C-means Clustering Technique in Distributed System", *Ingénierie des Systèmes d'Information*, vol. 25, no. 1, pp. 59–68, 2022.
https://doi.org/10.18280/isi. 250108

[25] W. Sheng *et al.*, "Dynamic Clustering Modeling of Regional Centralized Photovoltaic Power Plant Based on Improved Fuzzy C-Means Clustering Algorithm", *Power System Technology*, vol. 41, no. 10, pp. 3284–3291, 2017.
https://doi.org/10.13335/j.1000-3673.pst.2017.1861

[26] Y. Chen *et al.*, "An Unsupervised Deep Learning Approach for Scenario Forecasts", in *Proc. of the 2018 Power Systems Computation Conference (PSCC)*, pp. 1–7, 2018.
https://doi.org/10.23919/PSCC.2018.8442500

[27] Y. Hu *et al.*, "A Group Division Method for Voltage Control Based on Distributed Photovoltaic Scaled Access to Rural Power Grid", in *Proc. of the 2020 5th Asia Conference on Power and Electrical Engineering (ACPEE)*, 2020, pp. 580–586.
https://doi.org/10.1109/ACPEE48638.2020.9136231

[28] M. Zhu and D. Z. Zhang, "Distributed Generation Cluster Division Strategy Based on Fuzzy Clustering Method", *Electric Engineering*, vol. 2023, no. 18, pp. 26–28, 2023.
https://doi.org/10.19768/j.cnki.dgjs.2023.18.008

[29] N. Lv *et al.*, "Cluster Partition Method Development for High Penetration of Distributed Photovoltaics Using Intelligent Algorithm and K-Means", in *Proc. of the 2022 IEEE 5th International Conference on Electronics Technology (ICET)*, Chengdu, China, 2022, pp. 1255–1260.
https://doi.org/10.1109/ICET55676.2022.9824625

[30] T. Meng *et al.*, "Research on Partition Method of Distributed Photovoltaic Cluster Based on Modularity Index and Absorption Ability", in *Proc. of the 2023 2nd International Conference on Smart Grids and Energy Systems (SGES)*, Guangzhou, China, 2023, pp. 323–327.
https://doi.org/10.1109/SGES59720.2023.10366932

[31] L. Chen *et al.*, "Distributed PV Cluster Partitioning Strategy Based on GAN Data Synthesis Federation Clustering", in *Proc. of the 2023 8th International Conference on Power and Renewable Energy (ICPRE)*, Shanghai, China, 2023, pp. 1730–1735.
https://doi.org/10.1109/ICPRE59655.2023.10353797

[32] N. Wu *et al.*, "Voltage Control Strategy of High-Penetration Photovoltaic Distribution Network Based on Cluster Division", in *Proc. of the 2022 China Automation Congress (CAC)*, Xiamen, China, 2022, pp. 5701–5706.
https://doi.org/10.1109/CAC57257.2022.10055653

[33] Z. Y. Wang and Q. Z. Gao, "Large-Scale Distributed Photovoltaic Cluster Partition Method Based on SLM Algorithm", *Journal of Physics: Conference Series*, vol. 2703, p. 012025, 2024.
https://doi.org/10.1088/1742-6596/2703/1/012025

[34] Z. Liu *et al.*, "A Graph-Based Genetic Algorithm for Distributed Photovoltaic Cluster Partitioning", *Energies*, vol. 17, no. 12, p. 2893, 2024.
https://doi.org/10.3390/en17122893

[35] M. Huang *et al.*, "Cluster Partition Method of High-Penetration Distributed Photovoltaics for Autonomous Operation and Control of Distribution Network", in *Proc. of the 5th International Conference on Information Science, Electrical, and Automation Engineering (ISEAE 2023)*, 2023, vol. 12748, pp. 1134–1139.
https://doi.org/10.1117/12.2689759

[36] J. L. Li *et al.*, "Research on Distributed Photovoltaic Cluster Partition and Dynamic Adjustment Strategy Based on AGA", *Electrical Engineering*, pp. 1–13, 2024.
https://doi.org/10.1007/s00202-024-02549-8

Received: July 2024
Revised: September 2024
Accepted: September 2024

Contact addresses:
Yansen Chen*
China Southern Grid Digital Grid Research Institute Co. Ltd
Guangzhou
China
e-mail: chenyansen1985@126.com
*Corresponding author

Kai Cheng
China Southern Grid Digital Grid Research Institute Co. Ltd
Guangzhou
China
e-mail: chengkai@csg.cn

Zhuohuan Li
China Southern Grid Digital Grid Research Institute Co. Ltd
Guangzhou
China
e-mail: 292990775@qq.com

Shixian Pan
China Southern Grid Digital Grid Research Institute Co. Ltd
Guangzhou
China
China Southern Power Grid Artificial Intelligence
Technology Co. Ltd.
Guangzhou
China
e-mail: pansx@csg.cn

Xudong Hu
China Southern Grid Digital Grid Research Institute Co. Ltd
Guangzhou
China
China Southern Power Grid Artificial Intelligence
Technology Co. Ltd.
Guangzhou
China
e-mail: huxd1@csg.cn

Yansen Chen holds a bachelor's degree in engineering. He is currently a senior researcher (software architecture) at the Southern Power Grid Digital Grid Research Institute, China, mainly engaged in the research and application of new energy prediction and control, big data processing, high availability, and high concurrency information systems.

Kai Cheng holds a master's degree in engineering. He is currently an assistant researcher at the Southern Power Grid Digital Grid Research Institute, China, mainly engaged in new energy prediction and control, product research and development, and engineering applications.

Zhuohuan Li holds a master's degree in engineering. He is currently a researcher at the Southern Power Grid Digital Grid Research Institute, China, mainly engaged in the analysis and operation control research of high proportion new energy power systems.

Shixian Pan holds a master's degree in engineering. He is currently a researcher at the Southern Power Grid Digital Grid Research Institute, China, mainly engaged in research on new energy power forecasting technology.

Xudong Hu holds a master's degree in applied statistics. He is currently a researcher at the Southern Power Grid Digital Grid Research Institute, China, mainly engaged in research on new energy power forecasting.