Location and Size Planning of Distributed Photovoltaic Generation in Distribution network System Based on K-means Clustering Analysis

Siqi Lu 1, Xiaorong Wang 2 and Junyong Wu 1,*

1School of Electrical Engineering, Beijing Jiaotong University, Beijing, China
2China Electric Power Research Institute, Beijing, China

*Corresponding author e-mail: wujy@bjtu.edu.cn

Abstract. The paper presents a method to generate the planning scenarios, which is based on K-means clustering analysis algorithm driven by data, for the location and size planning of distributed photovoltaic (PV) units in the network. Taken the power losses of the network, the installation and maintenance costs of distributed PV, the profit of distributed PV and the voltage offset as objectives and the locations and sizes of distributed PV as decision variables, Pareto optimal front is obtained through the self-adaptive genetic algorithm (GA) and solutions are ranked by a method called technique for order preference by similarity to an ideal solution (TOPSIS). Finally, select the planning schemes at the top of the ranking list based on different planning emphasis after the analysis in detail. The proposed method is applied to a 10-kV distribution network in Gansu Province, China and the results are discussed.

1. Introduction

With the integration of renewable energy, distributed energy storage systems and electric charging and changing facilities for electric vehicles into power system, distribution network, as a link between traditional power system and new things such as distributed energy and interactive users, is the most important part of power system reform. Study on distribution network planning is mainly divided into two types: the distributed energy supply capacities and locations planning; the distribution network expansion planning considering the whole system with distributed power feeder or distribution substation equipment [1]. In this paper, the research on distributed PV planning is presented.

There have been a lot of existing studies dealing with the improvement of every aspect of distribution network planning with different views, optimization algorithms, constraints and objectives. In terms of optimization algorithms, many algorithms such as genetic algorithm (GA)[2-4], tabu search (TS)[5, 6], particle swarm optimization (PSO)[7, 8], and evolutionary algorithm (EA)[9] are used to solve the nonlinear programming problem. In [10], the most famous optimization algorithms for solving the distribution network planning problem are reviewed and compared, and some points are proposed to improve the performance of the algorithms. Besides, some new optimization algorithms are developed to conquer the shortcomings of the existing algorithms. In [11], the multi-objective hybrid big band-big crunch algorithm is proposed for simultaneous network reconfiguration and power allocation of DGs in distribution networks. In [12], an evolutionary algorithm-based solution method called binary chaotic shark smell optimization algorithm is used to solve a multiyear...
expansion planning of distribution networks including the reinforcement pattern of primary feeders and location and size of DGs. And meanwhile, the objective functions of distribution network planning vary with the progress of the distribution network research. In [13], the objective function is to manage the output of battery energy storage system (BESS) so that PV station output swings are reduced significantly. In [14], the objective function of the optimization problem includes only the battery installation costs so that a direct cost comparison with other PV integration measures can be performed. In [15], the place and capacity of the energy storage systems (ESSs) and DGs in the coordinated ESS and DG planning model are determined to maximize the profit of the distribution company subject to the secure operation of the network. Multi-objective planning techniques should be exploited to sort out viable trade-off solutions among conflicting objectives that satisfy multiple stakeholders. A multi-objective method for optimal network reconfiguration as well as reactive power dispatches of distributed generations (DGs) has been proposed to minimize system power loss, voltage deviation and energy wastage from solar and wind generation system [16]. In [8], the planning aims at minimizing the cost of energy not supplied as well as ESSs cost at the same time by an optimal approach to denote the location and size of energy storage systems (ESS). The distribution network reconfiguration problem in a multi-objective scope is addressed in [17], aiming to determine the optimal radial configuration by means of minimizing the active power losses and a set of commonly used reliability indices formulated with reference to the number of customers. But DGs are not considered in the paper.

In addition, the renewable DG technologies like PV have special characteristics due to their dependence on climate condition, therefore the uncertainty and variability in the PV output power should be taken into account. At present, researchers have proposed some methods to address the problem, especially probabilistic methods. Probability density functions are widely used to model the uncertainties related to solar irradiance, load demand and future load growth in [18], and in general, Monte Carlo simulation is employed to the generated probability density functions. Latin Hypercube Sampling with Cholesky Decomposition and point estimate method [5, 6] are used to take place of Monte Carlo simulation to reduce the computational burden [19-21]. In the probabilistic method above, DGs output power and load power are considered as independent events without correlation. The recent technological developments monitoring the electricity use of small customers provides with a whole new view to build DG output power and load models. In [22], hourly input data (electricity demand, onshore wind, offshore wind, solar availability) is clustered via a K-means algorithm to reduce the computational effort and increase solution speed. Homes with similar hourly electricity use patterns are clustered into groups using the K-means clustering algorithm to determine the shape of seasonally-resolved residential demand profiles within each season [23]. Though the data-driven method above takes the correlation into account, historical data are separated into four groups of seasonal data by experience, which leads to the inaccuracy of the weight of the generated scenarios.

To address the problem above, this paper proposed a method based on the data-driven cluster analysis technology to determine the number of clusters (planning scenarios), the weight of each cluster and the duration curves of PV and load in each cluster. The analysis of PV output power and load power is implemented to extract the appropriate characteristic variables for K-means clustering algorithm, which has positive effects on clustering result and keeping the correlation of PV output power and load power. The objective functions are to minimize the power loss of the network, the installation and maintenance costs of distributed PV and the voltage offset, and to maximize the generation energy of distributed PV. Because of the conflict objective functions, a self-adaptive genetic algorithm is used to obtain Pareto optimal front of decision variables (the location and size of PV units). Then the solutions are ranked by TOPSIS. Finally, determine the recommended planning scheme from the optimal result, which can be referenced to actual demands. The proposed method is applied to a 10-kV distribution network in Gansu Province, China and the results are discussed.
2. Planning scenarios based on K-means clustering algorithm

2.1. Analysis of photovoltaic generation output power and load power and determination of their characteristic variables

To generate the planning scenarios, the methods proposed in previous research highly depend on planners’ experience. Typically, a single year is divided into four seasons and each season lasts for three months [6, 15, 17]. Previous studies indicate that the distributed PV output power and load power have shown obvious daily periodicity. The daily fluctuation of distributed PV output power and load are similar while the power fluctuation amplitude varies with the seasons. Four PV output power curves of a city in different seasons and the corresponding residential load duration curves are shown in Figure 1. It can be obtained that the daily fluctuations of all PV output power curves are similar, but the maximum amplitude varies with the season of the daily PV output power curve. The conclusion also applies to the load power curves.

However, the scenarios in the whole year cannot be fully represented by four seasons of the same length. This experience-driven method to generate scenarios cannot be appropriately adapted to different planning areas under different climate conditions.

Figure 1. Comparison of daily PV (load) power curves.

To address the above problem, this paper proposed a method for clustering analysis of distributed PV generation output power curves and historical load power data to obtain the planning scenarios and the corresponding duration curves of PV and load power in each planning scenario. It can make the planning scenarios more accurate.

The clustering result is based on the extraction of the characteristic variables. According to the above analysis, the output power of distributed PV has obvious daily periodicity, so the daily average PV output power and the daily maximum PV output power are selected as the characteristic variables of the distributed PV generation. PV output can be calculated using historical weather data though Formula (1).

\[
P_{PV} = P_{STC} \frac{I_{ING}}{I_{STC}} \left[ 1 + K_f (T - T_r) \right]
\]  

Where \( P_{PV} \) is the output power of PV. \( P_{STC} \) is the output power of PV under the standard test condition (the PV battery temperature: 25°C). \( I_{ING} \) is the actual light intensity can be obtained by solar
irradiance; \( I_{STC} \) is standard test light intensity; \( K_T \) is the temperature factor; \( T \) is PV module temperature (generally approximately equal to ambient temperature); \( T_r \) is referenced temperature.

Similarly, the distribution network load duration curves have obvious daily periodicity. The daily average load power and the daily maximum load power of each load node are taken as the characteristic variables of the load duration curves.

2.2. K-means clustering analysis of distributed PV output power and load duration curves

The clustering algorithm is mainly divided into two types: the division method and the hierarchical method. The division method divides the data set into \( K \) parts by optimizing the evaluation function, and the value of \( K \) should be given before clustering. Hierarchical clustering algorithm consists of different levels of division clustering operation, and the division between levels has nested relations. It does not require input parameters, which is an obvious advantage over the division clustering algorithm, but the termination condition of the clustering algorithm should be specified in advance. Because the huge amount of data used in the planning process, it is necessary to consider whether the clustering algorithm is suitable for clustering large data sets. In addition, the value of \( K \) has a certain reference value based on the traditional method of dividing the whole year into four seasons in planning process. Therefore, K-means clustering algorithm is selected for the following research.

The value of \( K \) should be specified in advance. Calculate the distance between clustering centers (labelled as betweenSS) and the total distance (labelled as totSS). The total distance includes the distance between different clustering centers and the distances between the clustering objects with the clustering center among the same cluster. Calculate the ratio of betweenSS and totSS as a measurement of the clustering result. The larger the ratio is, the more effective the clustering result is.

It should be noticed that orders of magnitude of PV output power and load power are different. The comparison between two kinds of data will influence the contribution weights of PV output power data and load power data to the clustering results. Therefore, it is necessary to normalize the PV power data and load power data before the clustering analysis.

3. Mathematical formulation

3.1. Objective functions

In this paper, the multi-objective planning should consider economic indicators and technical indicators. Select the power loss of the network, the installation and maintenance costs of distributed PV, the profit of generation power energy of distributed PV and the node voltage offset as objective functions.

Planning objective 1: minimize the annual power loss of the network with the distributed PV, \( P^{loss} \)

\[
P^{loss} = \sum_{d=1}^{K} \left( \sum_{i=1}^{N_{node}} \sum_{j=1}^{N_{node}} G_{ij,d} \right) \left( U_{i,t,d}^2 + U_{j,t,d}^2 - 2U_{i,t,d} U_{j,t,d} \cos \delta_{ij,t,d} \right) \omega_d \alpha \tag{2}
\]

Where \( K \) is the number of clusters, \( N_{node} \) is the total number of nodes, \( G_{i,j,d} \) is the conductance of the bus from node \( i \) to node \( j \), at time \( t \), in cluster \( d \), \( U_{i,t,d} \) is the node voltage amplitude at node \( i \), time \( t \), in cluster \( d \), \( \delta_{i,j,t,d} \) is the voltage phase difference between node \( i \) and node \( j \), at time \( t \), in cluster \( d \), \( \omega_d \) is the number of days of cluster \( d \) and \( \alpha \) is the price of electricity.

Planning objective 2: maximize the profit of distributed PV generation, \( C_{pv} \)
The profit of distributed PV generation includes the energy cost saving by PV and the state subsidy of PV.

\[ C_{pv} = \sum_{d=1}^{24} \sum_{t=1}^{N_{node}} N_{PV}^{d} P_{i,t,d}^{PV} \cdot (a+b) \]  

(3)

Where \( N_{PV}^{d} \) is the number of the PV generation units, \( P_{i,t,d}^{PV} \) is output power of PV generation of node \( i \) at time \( t \) in cluster \( d \), and \( b \) is the state subsidy.

Planning objective 3: minimize installation and maintenance costs of distributed PV generation, \( C \)

\[ C = N_{PV}^{d} P_{0} \left( \frac{r(1+r)^{n}}{(1+r)^{n}-1} \right) \left( C_{1} + C_{2} \right) \]  

(4)

Where \( P_{0} \) is the installed capacity of the PV unit, \( r \) is the annual interest rate, \( n \) is the number of planning horizon, \( C_{1} \) is the installation cost of distributed PV unit, \( C_{2} \) is the operation and maintenance costs of distributed PV unit.

Planning objective 4: minimize the average voltage offset, \( \Delta U \)

\[ \Delta U = \frac{\sum_{d=1}^{K} \sum_{t=1}^{24} \sum_{i=1}^{N_{node}} \left( \frac{U_{i,t,d}^{d} - U_{N}}{U_{N}} \right)^{2}}{8760 \times N_{node}} \]  

(5)

Where \( U_{N} \) is the rated voltage.

3.2. Constraints

In the process of distribution network planning, the constraints include network constraints and distributed PV generation constraints. Network constraints include the power flow constraints and the safe operation constraints of power quality indexes. Distributed PV generation constraints include capacity constraint for distributed PV accessed to each load node and the safe penetration constraint of distributed PV to prevent the influence of PV beyond control.

Constraint 1: Power flow equation

\[
\begin{align*}
P_{i,t,d}^{PV} &= P_{i,t,d}^{L} + U_{i,t,d} \sum_{j \in N_{i}} U_{j,t,d} \left( G_{ij,t,d} \cos \delta_{ij,t,d} + B_{ij,t,d} \sin \delta_{ij,t,d} \right) \\
Q_{i,t,d}^{PV} &= Q_{i,t,d}^{L} + U_{i,t,d} \sum_{j \in N_{i}} U_{j,t,d} \left( G_{ij,t,d} \sin \delta_{ij,t,d} + B_{ij,t,d} \cos \delta_{ij,t,d} \right)
\end{align*}
\]  

(6)

Where \( P_{i,t,d}^{PV} \) and \( Q_{i,t,d}^{PV} \) are the active and reactive power of PV at node \( i \), time \( t \), in cluster \( d \).

Constraint 2: Node voltage constraint

\[ U_{\text{min}} \leq U_{i,t,d} \leq U_{\text{max}} \]  

(7)

Where \( U_{\text{max}} \) and \( U_{\text{min}} \) are the upper and lower limitation, respectively.
Constraint 3: Upper limit of distributed PV penetration

\[ N^{PV} P_0 \leq \eta P_{load} \] (8)

Where \( P_{load} \) is the sum power of peak load. \( \eta \) is the maximum penetration of DG.

Constraint 4: Upper limit of distributed PV accessed to each load node

\[ P_i^{DG} \leq P_i^{DG_{max}} \] (9)

Where \( P_i^{DG_{max}} \) the maximum output power of distributed PV is accessed to the node \( i \).

4. Multi-objective optimization method

GA has advantages of high parallelism, strong randomness and great performance on global searching. It has been widely applied to solve combinatorial optimization problem and nonlinear optimization problem. GA also shows obvious disadvantages. The computation speed of GA is highly affected by the objective functions. It could be easy to fall into the local optimum during the iterative process. A large number of infeasible solutions would be generated during the crossover and mutation. So, some specific improvements are made to address these problems in traditional GA.

4.1. Chromosome encoding and population initialization

Chromosomes serve as the basis for GA, determining the performance of individuals. In this paper, integer encoding is used. The number of the genes is equal to the number of nodes. The value of the gene represents the number of accessed distributed PV units. The proposed encoding strategy is shown below:

\[ X = [x_1, x_2, \ldots, x_{N_{bus}}] \] (10)

4.2. Fast Non-dominated sorting

There are two main methods to solve the multi-objective optimization problem. The first method is to incorporate the objective functions with different weight, which is the most commonly used method [3]. This method is easy to be implemented, but the weight is highly dependent on the experience of planners. The second method is to obtain Pareto optimal front through optimization algorithms [11, 15, 19]. In this paper, the second method is used to solve the multi-objective optimization problem.

During the process of distribution network planning, the value of each objective function is obtained by power flow calculation. Forward and backward substitution method is used to solve the power flow of distribution network. The genetic process of GA needs to repeatedly call the fitness function. Take the distance between the individual and the nearest Pareto optimal solution in the current population as its fitness.

\[ f_i = \frac{1}{1 + \| X_i - X_p \|_b} \] (11)
Where $f_i$ is the fitness of the $i$th individual in the current population, $\|X_i - X_p\|_2$ is the Euclidean distance between the individual $X_i$ and the nearest Pareto optimal solution $X_p$.

4.3. Genetic Operation

Genetic operation includes selection, crossover, mutation, and retention. Crossover rate and mutation rate have great influence on the performance of GA.

Selection

To ensure the evolution of the population toward the Pareto optimal front, the individuals in the Pareto optimal front are directly involved in crossover and mutation, and the rest of individuals participate in the selection. Two integers that are less than or equal to the number of individuals of a population are randomly generated, and the non-dominated level and the crowding distance of these two individuals are compared. Select the individual in low non-dominated level (the Pareto optimal front is regarded as level 1). If two individuals are in the same level, compare the crowding distance between the two individuals and select the one with larger crowding distance.

Crossover and Mutation

To avoid the local optimum, the crossover operator and mutation operator are adjusted linearly according to the fitness of the population, as follows:

$$p_c = \begin{cases} p_{c1} - \frac{(p_{c1} - p_{c2})(f' - f_a)}{f_m - f_a}, & f' \geq f_a \\ p_{c1}, & f' < f_a \end{cases}$$

(12)

$$p_m = \begin{cases} p_{m1} - \frac{(p_{m1} - p_{m2})(f - f_m)}{f_m - f_a}, & f \geq f_a \\ p_{m1}, & f < f_a \end{cases}$$

(13)

Where $p_{c1}, p_{c2}$ are the upper and lower bounds of the crossover rate. $p_{m1}, p_{m2}$ are the upper and lower bounds of the mutation rate. $f_m$ is the max fitness in the current generation of population. $f_a$ is the average fitness of the current generation of population. $f'$ is the greater fitness of the two crossover individuals. $f$ is the fitness of the individual.

Retention

Merge Pareto optimal front of the parent population and the child population into a new population. Use the fast non-dominated sorting method for new population, and calculate the crowding distance of each individual. The top ranked individuals are selected to be part of the next generation population and the following genetic operations continue to be performed.

4.4. Infeasible Solutions Processing

In genetic operations, infeasible individuals are generated during the process of crossover and mutation. The method of dealing with infeasible solutions consists of two parts:

For distributed PV constraints, the infeasible solutions can be caused by distributed PV being accessed to non-load node or the size of PV exceeding the limitation. The maximum and minimum accessed distributed PV capacity of each node could be respectively determined by two matrixes. For a new generation of population, check the gene of each individual, and the accessed PV size exceeding the limitation is placed by given upper and lower limitations.

For power flow constraints, the infeasible individuals are caused by node overvoltage related to the distributed PV. It could be solved by introducing the penalty function to give the infeasible individual a small fitness value, which leads to the elimination of natural genetic operation.
4.5. The Decision Making Strategy
After obtaining the Pareto optimal front, TOPSIS has been implemented. Introduce the conception of the ideal and the negative-ideal solution. The ideal solution is the objective vector which all the objective functions reach the optimum and the negative-ideal solution is the objective vector which all objective functions reach the worst.

\[
val_+ = [f_1, f_2, \ldots, f_P]
\]

(14)

\[
val_- = [f_1, f_2, \ldots, f_P]
\]

(15)

Where \( P \) is the number of objective functions.

Calculate the relative closeness to the negative-ideal solution, \( M_i \). According to the ascending order of relative closeness, the ranking of the solutions is performed with respect to the relative closeness calculated by using (16):

\[
M_i = \frac{\|f - val_+\|}{\|val_- - f\| + \|f - val_+\|}
\]

(16)

Sort the solutions in the Pareto optimal front according to the relative closeness.

4.6. Algorithm Flow Chart

5. Simulation results and discussions
The proposed method above is applied to determine the location and size of distributed PV in a 10kV distribution network in Gansu Province, in China as shown in Figure 3. Peak load of the network is occurred at 3088.28 kW and 1831.83 kVar. The historical solar irradiance data and load power data from Gansu Province in China have been used in the process of planning scenario generation.

The betweens/totss-K curve obtained by clustering the PV power data and load power data of a city for a whole year is shown in Figure 4.
From Figure 4:

When $K \leq 8$, the ratio of betweenss and totss increases as the K value increases and the change of its growth is obvious, which indicates that the growth of K value can improve the clustering result effectively in this case.

When $K > 8$, the ratio of betweenss and totss is approximately 0.85, close to 1. And the ratio changes little with the growth of K. It indicates that the clustering result has been effective enough and the growth of K has little contribution to the improvement of the clustering result.

Therefore, it is appropriate to set the number of clusters K to 8. In this case, the ratio of betweenss and totss equals to 88.4%, which means that the clustering result is satisfying. K-means clustering result of the distributed PV output power and load power is shown in Figure 5.

There are 8 planning scenarios in total with PV output power curves and load duration curves in each planning scenario. The clustering result and the number of days of each cluster are presented in Table 1. In other words, the fluctuation of PV output power and load power of 8760 hours in a year can be summarized by the 8 planning scenarios in Figure 5, and the number of days of each scenario varies as shown in Table 1.
Table 1. Three Scheme comparing.

| Num | Daily max. PV power | Daily average PV power | Daily max. load power | Daily average load power | Num. of days |
|-----|---------------------|------------------------|-----------------------|-------------------------|-------------|
| 1   | 0.3567715           | 0.1278125              | 0.7719438             | 0.5319071               | 17          |
| 2   | 0.7722859           | 0.2445904              | 0.8437805             | 0.5730979               | 25          |
| 3   | 0.8972864           | 0.3012879              | 0.3359646             | 0.2222485               | 78          |
| 4   | 0.8739324           | 0.2881783              | 0.538945              | 0.3716012               | 37          |
| 5   | 0.4072870           | 0.1423051              | 0.4297497             | 0.2754488               | 40          |
| 6   | 0.2114304           | 0.1056693              | 0.4202594             | 0.2729958               | 60          |
| 7   | 0.5963232           | 0.1907256              | 0.5309545             | 0.3474800               | 49          |
| 8   | 0.6516436           | 0.2046796              | 0.3768330             | 0.2366256               | 59          |

To examine that the representativeness of PV output power curves and load duration curves obtained by the clustering analysis as well as to compare the K-means clustering method with the traditional four-season method, the standard deviation of two methods are calculated by formula (17):

\[
\sigma_{d,t} = \sqrt{\frac{1}{\omega_d} \sum_{h=1}^{\omega_d} (x_{h,d,t} - \mu_{d,t})^2}
\]  

(17)

Where \(\sigma_{d,t}\) is the standard deviation of PV power (or load) of cluster d, at time t. \(\omega_d\) is the number of days of cluster d. \(x_{h,d,t}\) is the magnitude of PV output (or load power) in cluster d, on day h, at time t. \(\mu_{d,t}\) is the average output of PV at time t in cluster d.

The average standard deviation of PV output power and load power curves in each cluster by K-means clustering method is shown in the left part of Figure 6, while those in each cluster by the traditional four-season method is shown in the right part of Figure 6.

![Figure 6. Standard deviations of PV and load power in two methods](image)

The average and maximum standard deviations of the two methods are presented in Table 2.

Table 2. Comparison between the standard deviations of two methods

|                  | K-means standard deviation | Experience cluster standard deviation |
|------------------|----------------------------|--------------------------------------|
| PV               | 0.059597                   | 0.20556                              |
| LOAD             | 0.068913                   | 0.157545                             |
| Average          | 0.091301                   | 0.295783                             |
| Maximum          | 0.10804                    | 0.234587                             |
From Figure 6 and Table 2, it comes to conclusion that the average and maximum standard deviation of the PV output power curves and load duration curves in planning scenarios obtained by K-means cluster algorithm are smaller than those in the planning scenarios obtained by the traditional experience-driven method. Curves of PV output power and load power data in the whole year in the planning scenarios obtained by K-means algorithm are more representative than those in the planning scenarios obtained by the traditional four-season method, which means that the planning scenarios obtained by K-means algorithm are more suitable.

Assume that the distributed generators can only be connected to the load nodes. In the distribution network, the 3th, 4th, 5th, 6th, 7th, 9th, 10th, 11th and 12th nodes are load nodes. Set the parameters of GA. The number of individuals in each population is 60. The number of iterations is 100. The maximum cross rate is 0.9. The minimum crossover rate is 0.5. The maximum mutation rate is 0.1. The minimum variance rate is 0.05. The annual interest rate is 0.06. The installation cost of distributed PV is ¥8000/kVA. The maintenance cost is ¥0.06/kWh. The electricity price is ¥0.6/kWh. The distributed PV tariff is ¥0.55/kWh. The national subsidy is ¥0.2/kWh. The maximum individual number of Pareto front is 40. The top 10 solutions obtained by adaptive GA are shown in Table III.

### Table 3. The top 10 solutions obtained by adaptive GA

| No. | Accessed Node | PV capacity/kW | $P^{\text{loss}}$ ¥ | $C_m$ ¥ | $C$ ¥ | $\Delta U$ /p.u. |
|-----|---------------|----------------|----------------------|----------|--------|-----------------|
| Orig. | \ / | 329600 \ / | 155753 | 93724 | 259284 | 0.0379 |
| 1 | 3,4,6,7,9,10,12 | 90,150,120,30,90,60,180 | 156720 | 942737 | 507230 | 0.0382 |
| 3 | 3,4,6,7,9,12 | 90,150,120,30,90,150 | 157667 | 942736 | 463123 | 0.0385 |
| 4 | 3,4,6,7,9,12 | 30,150,120,30,90,180 | 157991 | 819770 | 441070 | 0.0387 |
| 5 | 3,4,6,7,12, | 90,150,120,30,180 | 159503 | 778782 | 419016 | 0.0388 |
| 6 | 3,4,7,9,12 | 90,150,30,90,150 | 159208 | 696805 | 374909 | 0.0391 |
| 7 | 3,4,7,9,12 | 30,150,30,90,180 | 159839 | 655816 | 352856 | 0.0394 |
| 8 | 3,4,5,7,12 | 90,150,30,30,150 | 161943 | 614882 | 330802 | 0.0397 |
| 9 | 3,4,7,12 | 90,150,30,150 | 162292 | 573839 | 308749 | 0.0398 |
| 10 | 3,4,7,12 | 30,150,30,180 | 162846 | 532851 | 286695 | 0.0401 |

According to Table III, compared with the original network, all kinds of optimization scheme reduce the network losses and the voltage offset. Meanwhile, power generation from distributed PV can create huge profit without pollution. It is noticeable that the installation and maintenance cost of distributed PV are huge. Comparing the 10 schemes of distributed PV planning, the result of the network power loss cost, the profit of PV, the installation and maintenance costs and the average voltage offset are shown in Figure 7:
According to Figure 7, the objective functions are conflict. The above 10 optimization schemes are sorted by the distance to the virtual worst solution. Scheme 1 shows great performance on reduction of network power loss, distributed PV power generation and the voltage offset, despite a high distributed PV costs. With the adequate budget, scheme 1 would be a good choice. Scheme 10 is more applicable with a limited budget. In the actual planning, the final scheme selection should take more specific requirements into considerations. The proposed method find out a series of optimization schemes as alternatives, providing references and effective assists in decision making.

6. Conclusion

The correlation of PV output and load curve is fully considered during the process of K-means clustering analysis. The PV output power and load changes of the whole year are clustered into 8 planning scenarios, preparing for the following planning. The example is performed to verify the effectiveness of proposed method compared with the conventional seasonal method. Distributed PV generation output power curve and load curve are more accurate comparing to the traditional seasonal method. Then, distribution network losses, distributed PV generation, installation and maintenance cost and average voltage offset are taken into considerations as objective functions. A refined self-adaptive GA is used to solve the location and size planning of distributed PV. The results provide the reference for the planner in the practical problem.

In the future study, more aspects should be considered, including the greater network scale, various kinds of DG, the electricity market operation and other directly or indirectly related factors. Further researches on distribution network planning need to be carried out.

Acknowledgments

The work was supported by Research Program of State Grid Cooperation (DZ71-15-004) of China. The data used in the paper are from China Electric Power Research Institute.

References

[1] Kazmi, A.A., R.S. Dong and M.K. Shahzad, "Multi-objective planning techniques in distribution networks: a composite review". Energies, 2017. 10: p. 44.

[2] Heidari, S., M. Fotuhi-Firuzabad and S. Kazemi, "Power distribution network expansion planning considering distribution automation". IEEE Transactions on Power Systems, 2015. 30(3): p. 1261-1269.
[3] Sardi Junainah, Mithulananthan N., Gallagher M., Hung Duong Quoc, Yan J., "Multiple community energy storage planning in distribution networks using a cost-benefit analysis". Applied Energy, 2017. 190: p. 453-463.

[4] Heidari, S., M. Fotuhi-Firuzabad and M. Lehtonen, "Planning to equip the power distribution networks with automation system". IEEE Transactions on Power Systems, 2017. PP(99): p. 1-1.

[5] Ahmadian, Ali, Sedghi, Mahdi, Aliakbar-Golkar, Masoud, Fowler, Michael, Elkamel, Ali, "Two-layer optimization methodology for wind distributed generation planning considering plug-in electric vehicles uncertainty: A flexible active-reactive power approach". Energy Conversion & Management, 2016. 124: p. 231-246.

[6] Ahmadian, Ali, Sedghi, Mahdi, Aliakbar-Golkar, Masoud, Elkamel, Ali, Fowler, Michael, "Optimal probabilistic based storage planning in tap-changer equipped distribution network including PEVs, capacitor banks and WDGs: A case study for Iran". Energy, 2016. 112: p. 984-997.

[7] Bezerra, José Roberto, Barroso, Giovanni Cordeiro, Leão, Ruth Pastôra Saraiva, Sampaio, Raimundo Furtado, "Multiobjective optimization algorithm for switch placement in radial power distribution networks". IEEE Transactions on Power Delivery, 2015. 30(2): p. 545-552.

[8] Saboori, Hedayat, Hemmati, Reza, Jirdehi, Mehdi Ahmadi, Lund, Henrik, Kaiser, Mark J, "Reliability improvement in radial electrical distribution network by optimal planning of energy storage systems". Energy, 2015. 93(2): p. 2299-2312.

[9] Qiu, J., et al., "Multi-objective transmission expansion planning in a smart grid using a decomposition-based evolutionary algorithm". Iet Generation Transmission & Distribution, 2016. 10(16): p. 4024-4031.

[10] Sedghi, M., A. Ahmadian and M. Aliakbar-Golkar, "Assessment of optimization algorithms capability in distribution network planning: Review, comparison and modification techniques". Renewable & Sustainable Energy Reviews, 2016. 66: p. 415-434.

[11] Esmaeili, M., M. Sedighizadeh and M. Esmaeili, "Multi-objective optimal reconfiguration and DG (Distributed Generation) power allocation in distribution networks using Big Bang-Big Crunch algorithm considering load uncertainty". Energy, 2016. 103: p. 86-99.

[12] AhmadiGorji, M. and N. Amjady, "A multiyear DG-incorporated framework for expansion planning of distribution networks using binary chaotic shark hunt optimization algorithm". Energy, 2016. 102: p. 199-215.

[13] Abdelrazek, S. and S. Kamalasadan, "A weather-based optimal storage management algorithm for PV capacity firming". IEEE Transactions on Industry Applications, 2016. 52(6): p. 5175-5184.

[14] Poulis, V., et al. "Optimal placement and sizing of battery storage to increase the PV hosting capacity of low voltage grids", in International ETG Congress 2015; Die Energiewende - Blueprints for the new energy age; Proceedings of. 2016.

[15] Saboori, H. and R. Hemmati, "Maximizing DISCO profit in active distribution networks by optimal planning of energy storage systems and distributed generators". Renewable & Sustainable Energy Reviews, 2017. 71: p. 365-372.

[16] Tuladhar S.R., J.G. Singh and W. Ongsakul, "Multi-objective approach for distribution network reconfiguration with optimal DG power factor using NSPSO". Iet Generation Transmission & Distribution, 2016. 10(12): p. 2842-2851.

[17] Paterakis, N.G., et al., "Multi-Objective reconfiguration of radial distribution systems using reliability indices". IEEE Transactions on Power Systems, 2016. 31(2): p. 1048-1062.

[18] Mokryani G., "Active distribution networks planning with integration of demand response". Solar Energy, 2015. 122(3): p. 1362-1370.

[19] Cao Yongji, Zhang Yi, Zhang Hengxu, Shi Xiaohan, and Terzija Vladimir, "Probabilistic optimal PV capacity planning for wind farm expansion based on NASA data". IEEE
Transactions on Sustainable Energy, 2017. PP(99): p. 1291-1300.

[20] Kabir, M.N., Y. Mishra and R.C. Bansal, "Probabilistic load flow for distribution systems with uncertain PV generation". Applied Energy, 2016. 163: p. 343-351.

[21] Ravadanegh S. Najafi, Jahanyari N., Amini A., and Taghizadeh N., "Smart distribution grid multistage expansion planning under load forecasting uncertainty". Iet Generation Transmission & Distribution, 2016. 10(5): p. 1136-1144.

[22] Heuberger Clara F, Staffell Iain, Shah Nilay, and Dowell Niall Mac, "A systems approach to quantifying the value of power generation and energy storage technologies in future electricity networks". Computers & Chemical Engineering, 2017.

[23] Rhodes Joshua D, Cole Wesley J, Upshaw Charles R, Edgar Thomas F, and Webber Michael E, "Clustering analysis of residential electricity demand profiles". Applied Energy, 2014. 135: p. 461-471.