Distributed photovoltaic siting optimisation considering load centre similarity

He Cheng [1, a, *], Sipeng Hao [1, b], Jitao Zhang [2, c], Kai Xu [2, d], Yuankai Ma [1, e]

[1] School of Electric Power Engineering, Nanjing Institute of Technology, Nanjing 211176,China;

[2] State Grid Xiangshui Power Supply Company, Yancheng 224600, Jiangsu, China.

[a, *] 2513497528@qq.com, [b] bhspnj@sohu.com, [c] 870669001@qq.com,

[d] 2523927221@qq.com, [e] 1561242066@qq.com

Abstract. The load fluctuation law remains challenging for traditional distribution line users. Furthermore, as the number of distributed photovoltaic installations connected to the distribution network rises, the issue of voltage overload on the distribution lines becomes more prevalent. This, in turn, adds complexity to power system operation. To configure the PV access location, whilst ensuring the line voltage is qualified, this paper applies the DBSCAN density clustering and K-means optimization algorithm to classify the distribution network line load. Different characteristics of the load collection are then constructed, and the concept of the load centre distance is proposed. PV access points are selected based on the load center characteristic curve and PV curve similarity. Model simulation is used to verify the conclusions, which demonstrate the rationality and effectiveness of the proposed method. The model simulation is conducted to confirm the soundness and credibility of the conclusions drawn.

Keywords: DBSCAN; K-means; electricity distribution network; load center; similarity.

1. Introduction

With the rapid development of the social economy, the entire power grid is increasingly prioritising the power quality of the distribution network. During the 75th session of the United Nations General Assembly, General Secretary Xi proposed the "carbon peak, carbon neutral" strategic objective, which will be integrated into China's overall ecological civilization construction plan [1]. The state is providing significant support for the development of new energy generation, including large-scale distributed photovoltaic systems that are connected to the electricity grid. According to statistics from the National Energy Administration, as of the end of June 2023, the total capacity of domestic photovoltaic power generation connected to the grid reached 47,000,000,000 kW. This comprises 27,200,000,000 kW of centralised power generation and 198,000,000 kW of distributed power generation.

With the increasing proportion of distributed photovoltaic access to the distribution network, leading to significant changes in the traditional distribution network, the traditional line in the current from the original single flow direction into the current backward, resulting in the PV system point of common coupling and even the surrounding nodes of voltage rise or overvoltage [2]. Literature [3] analysed the causes of PCC (point of common coupling) voltage rise caused by grid-connected PV power generation system through power system power transfer theory, analysed the voltage enhancement strategies based on active and reactive power for PV power generation system respectively, and proposed dynamic voltage adjustment based on instantaneous voltage-current control strategy. Literature [4] gives the relationship between the two types of criteria based on the existence of the current solution and the load-voltage characteristics, and expands and improves the voltage stability performance indexes of distribution networks. Literature [5] analyses and calculates the voltage stability of distribution networks in different locations and weather conditions through the distribution network voltage stability L index. Literature [6] proposed a new voltage stability H index by deriving the L index and criterion and verified it by arithmetic simulation. Literature [7] theoretically analyses the network loss of distribution networks containing distributed photovoltaic and establishes a two-layer optimization loss reduction model based on structural change and operation regulation to maximize the loss reduction benefit of distribution networks. Literature [8] studies the reactive power impact of distributed PV in active distribution networks and the reasons for it, and proposes the concept of trend reversal point (stage), and gives the reactive power fine automatic compensation scheme. In the face of large-scale distributed PV grid-connected, many scholars choose to study the capacity as well as location of PV grid-connected. Literature [9] proposes a distributed PV access planning method considering power quality constraints, and verifies the effectiveness of its planning through the improved particle swarm optimisation algorithm solution and data envelopment analysis comprehensive assessment method. Literature [10] considers the distributed power supply as well as the timing characteristics of the load, builds a two-layer optimisation model with the objective of the comprehensive economic cost of the system throughout the year and the optimal operating voltage level of the system in each time period, combines it with the adaptive genetic algorithm solution, and finally verifies the economy and reasonableness of the layout by simulation.

This paper categorizes line loads in the distribution network using the DBSCAN density clustering and K-means optimization algorithm. It creates a load collection with various features and introduces the concept of load centre distance. Through studying PV field operations, load timing characteristics, and load centre characteristics, the study analyses the impact of accessing PV at different load centres on the distribution line node voltage. The simulation results are presented. In the wider context of photovoltaic (PV) across the whole county, this paper suggests characterizing load centres of a specific type of load, focusing solely on the load centre and PV access points. Specific analysis of all loads on the line will not be undertaken, despite the fact that complex line PV access planning is of significant practical value.

2. Load classification and load centre calculation in distribution lines

Traditional distribution line users, load change law is difficult to figure out, random transient events and many other factors affect the line voltage, now a large number of distributed photovoltaic access to the distribution network distribution line, making the traditional distribution network single trend mode change, line voltage stability, line voltage difference between the first and the end of the line fluctuation is large and with the time, the weather changes. Therefore, it is necessary to classify the load of distribution lines, to achieve based on the load centre to plan the location and capacity of photovoltaic access.

2.1 Load clustering based on DBSCAN with K-means algorithm

2.1.1 The DBSCAN density clustering algorithm

DBSCAN（Density-Based Spatial Clustering of Applications with Noise） is a density-based clustering algorithm [11] for discovering clusters with different shapes and sizes and is able to identify noisy points. Compared to traditional distance-based clustering algorithms, DBSCAN focuses more on density connections of points rather than Euclidean distances, and is robust to noise in the dataset.The key parameters of DBSCAN include radius (ε , epsilon) and minimum number of neighbours (MinPts). The radius is the neighbourhood size, and MinPts is the threshold for determining whether a data point is a core point within the radius.

D is the line load power usage dataset, and Eps(p) is defined as:

$$Eps\ (p)\ =\{\, q \in D \mid distance\ (p,\ q) \le Eps\}$$

Based on the set a1, a2, ..., ak of class K generated by epsilon and MinPts, define the points in D that do not belong to one of the above sets as noise points (noises):

$$noise = \{ p \in D \mid \forall i\colon p \notin C_i \},\ i = 1,2,\ldots,\ k$$

2.1.2 K-means optimisation algorithm considering clustered evaluation metrics

Considering that the classical K-means algorithm needs to be set manually for the number of clusters and randomly selects the initial clustering centre [12], in the face of the complex and variable electricity consumption dataset, the algorithm produces a different set of categories when clustering and may select the noisy points in the dataset, which makes clustering fall into the local optimum. In this paper, based on the classical K-means algorithm, we combine the contour coefficient and CH index as the clustering index:

a) Contour coefficient

This index measures the closeness of data points to the cluster to which they belong and their separation from other clusters. Equation (1) is given below:

$$S(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{1}$$

Where a(i) is the intra-cluster dissimilarity, which represents the average distance from sample i to other samples in the same cluster; b(i) is the inter-cluster dissimilarity, which represents the average distance from sample i to samples in other clusters.

b) CH index

This index is the ratio of inter-cluster distance to intra-cluster distance. Equation (2) is given below:

$$\mathrm{CH}(k) = \frac{trB(k)/(k-1)}{trW(k)/(n-k)} \tag{2}$$

Where trB(k) denotes the sum of squared inter-cluster errors; trW(k) denotes the sum of squared intra-cluster errors.

2.1.3 Load classification based on DBSCAN and K-means optimisation algorithm

Combining the DBSCAN density clustering algorithm with the K-means optimisation algorithm can effectively eliminate the noise of the dataset and avoid falling into local optimal solutions when classifying loads, as well as combining the contour coefficients and the validity of the quantitative K-value of the CH index. The rough steps are as follows:

(1) Read the load data D and normalise it.

(2) Find the noise points by DBSCAN algorithm and remove the noise points.

(3) Average the load for the same user to get the data set D'.

(4) From the dataset D, select a data as the initial clustering centre a1 and calculate the distance D(i) between each data in the dataset and the nearest clustering centre, and calculate the probability of each data to be the next clustering centre, Equation (3) as follows:

$$P_i = \frac{D(i)^2}{\sum_{i \in x} D(i)^2} \tag{3}$$

Select the next clustering centre by roulette and repeat this step until K clustering centres are selected.

(5) For each sample in the dataset D' calculate its distance from the K clustering centres and classify it according to the clustering centre with the shortest distance;

(6) For the clustering centre aj(j $\in$ (1, k)),it needs to be recalculated each time new data is filled in, and Equation (4) is as follows:

$$A_{jnew} = \frac{1}{|A_j|} \sum_{D'(i) \in a_j} D'(i) \tag{4}$$

Repeat steps (5-6) until all clustering positions no longer appear to change, forming a classification result.

2.2 Load centre distance calculations

According to the above steps to obtain the classification results, solve for different load types of centres. In the case of distribution line current distribution is known, simplify the line model, to the equivalent power source and a certain type of load centre to the equivalent power source of the virtual equivalent resistance as the basis for the construction of the power transfer model, as shown in Fig. 1:

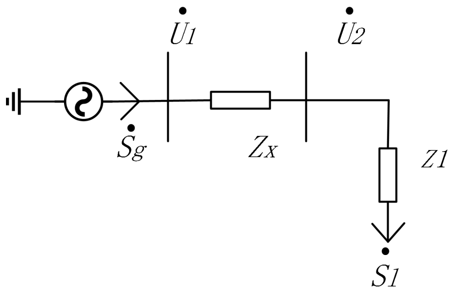


Fig.1 Equivalent line model

Where Sg is the branch injection power, Sl is the load power, Zl is the equivalent load corresponding impedance, Zx is the equivalent load distance from the power supply of the virtual line impedance, U1 is the power supply at the voltage, U2 is the equivalent impedance access to the voltage, where the branch injection power Sg and are Sl known quantity. The loss power Δ Sx of the virtual impedance Zx can be found by power balancing,Equation (5) is as follows:

$$\dot{S}_g = \Delta \dot{S}_x + \dot{S}_l \tag{5}$$

The voltage U2 at the equivalent impedance access can be found according to Eq. (6) as follows:

$$\dot{U}_2 = \frac{\dot{S}_1}{\dot{S}_g} \dot{U}_1 \tag{6}$$

The load equivalent impedance Zl, equation (7) is obtained as follows:

$$Z_l = \frac{\dot{U}_2 \dot{U}_2^*}{\dot{S}_l^*} \tag{7}$$

Finally, Zx is obtained based on the ratio of branch input power to output power [13], and Eq. (8) is given below:

$$Z_x = (\frac{\dot{S}_g}{\dot{S}_l} - 1)\ Z_l \tag{8}$$

Where Zx is the virtual line impedance of the equivalent load distance from the power source , i.e. the electrical distance between the centre of that type of load and the power source.

3. Mechanism of photovoltaic timing variations on distribution lines

In this paper, we study distributed photovoltaic integration into low-voltage distribution lines, so the theoretical analysis of radial low-voltage distribution network feeder, whose typical structure is shown in Figure 2.

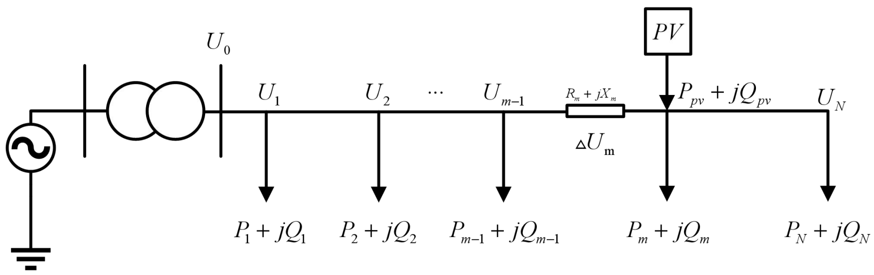


Fig. 2 Typical structure of a low-voltage feeder for a radial distribution network

The feeder is distributed along the line N node users, U0 is the first end voltage, Um is the voltage of the mth node, Pm + jQm is the m-node apparent power, Rm + jXm is the impedance between the m-1 node and the m-node, Ppv + jQpv is the PV merge point k merge power, assuming that the positive direction of the feeder is the positive direction of the power flow, ignoring the line loss and the line impedance is uniformly distributed, then the m-1 and m node voltage difference on the line is:

$$\Delta U_m = \frac{\sum_{i=m}^{N} P_i R_m + \sum_{i=m}^{N} Q_i X_m}{U_{m-1}} \tag{9}$$

Then the m-node voltage is:

$$U_m = U_0 - \sum_{i=1}^{m} \Delta U_i = U_0 - \sum_{i=1}^{m} \frac{\sum_{j=i}^{N} P_j R_i + \sum_{j=i}^{N} Q_j X_i}{U_{i-1}} \tag{10}$$

It can be seen that in the case of distribution lines without PV access, the Um line voltage decreases as the number of nodes increases. Assuming that the PV system only outputs active power, the PV access case is divided into access before node m and after node m:

(1) PV access point k is after node m. The voltage at point m is:

$$U_m^{'} = U_0 - \sum_{i=1}^{m} \frac{rl_i(\sum_{j=i}^{N} P_j - P_{pv}) + \sum_{j=i}^{N} Q_j xl_i}{U_{i-1}} \tag{11}$$

Compared to the case where PV is not connected, the value of U'm is greater than Um, which shows that the access of PV provides some support to the voltage. At this time, the voltage difference between nodes m-1 and m is:

$$U'_{\mathrm{m}} - U'_{\mathrm{m\text{-}1}} = -\frac{rl_m(\sum_{j=m}^{N} P_j - P_{pv}) + \sum_{j=m}^{N} Q_j x l_m}{U_{m-1}} \tag{12}$$

Considering that the voltage phase difference between two nodes in the actual line is not large, the transverse component of the voltage landing in the line can be ignored. That is, the magnitude of both Pj and Ppv determines the voltage difference between the two points, i.e. the direction of current flow between the two points. According to the light intensity and temperature in a day, it can be seen that the photovoltaic output power in a day with the sun began to increase, until the midday photovoltaic output power to reach the peak, when the photovoltaic output power is greater than the m node, and after the sum of all the loads, then the m node voltage is greater than the m-1 node voltage, resulting in a reverse tidal wave problem, which seriously affects the safety of the distribution line. Subsequently, the light intensity and temperature gradually decrease, when the PV output power is less than the m node and after the sum of all loads, the m node voltage is less than the m-1 node voltage, but the m node voltage has a certain degree of uplift, and finally the sun goes down to stop the PV output, that is, at this time, the line first and last end of the voltage landing is the largest, and even the end of the voltage over the lower limit of the situation.

(2) The PV access point k is before m, and the voltage at point m is:

$$U_{\mathrm{m}} = U_0 - \sum_{i=1}^{k} \frac{rl_i(\sum_{j=i}^{N} P_j - P_{pv}) + \sum_{j=i}^{N} Q_j x l_i}{U_{i-1}} - \sum_{i=k+1}^{m} \frac{\sum_{j=i}^{N} P_j r l_i + \sum_{j=i}^{N} Q_j x l_i}{U_{i-1}} \tag{13}$$

The voltage difference between the m-1 and m nodes is:

$$U'_{\mathrm{m}} - U'_{\mathrm{m\text{-}1}} = -\frac{\sum_{j=m}^{N} P_j r l_m + \sum_{j=m}^{N} Q_j x l_m}{U_{m-1}} \tag{14}$$

Under the trend of large-scale PV power access, the voltage volatility of distribution lines increases. Therefore, it is crucial to analyse the similarity between different load characteristics and PV output. Euclidean distance [14] focuses on the calculation of numerical distance, which to some extent can well identify the single-peak type loads and peak-avoidance type loads. In this paper, Euclidean distance is used as the similarity index, defined as follows:

$$d = \sqrt{\sum_{j=1}^{n} (x_{ij} - x_{i'j})^2}$$

Where d denotes the Euclidean distance between load curve i and PV curve i', and xij and xi'j are the values obtained after normalisation of load curve i and PV curve i' at moment j, respectively.

Based on the above analysis, it can be seen that when the degree of similarity between load and PV timing characteristics is high, the proportion of local consumption of PV power generation is high, and the degree of fluctuation of Pj-Ppv difference in Eq. (12) over time is reduced to improve the real-time voltage stability of the system. When the degree of similarity between load and PV timing characteristics is low, the proportion of local consumption of PV power generation is low, and the degree of fluctuation of the Pj-Ppv difference in Eq. (12) increases with time, which reduces the stability of the system voltage.

4. case analysis

In this paper, a typical day's data of 56 users in a county is collected for cluster analysis, and the data contains one electricity consumption data collection per user every 15 minutes. The clustering evaluation index is used to determine the best clustering value. In order to avoid the impact of holidays leading to the clustering effect, this paper selects a weekday as a typical day.

The contour coefficient interval range is [-1,1], the closer to 1 indicates that the clustering effect is better; CH index is the larger the better, as shown in Figure 3, the comprehensive consideration of the contour coefficient and the CH index changes to determine the optimal number of clusters K = 3.

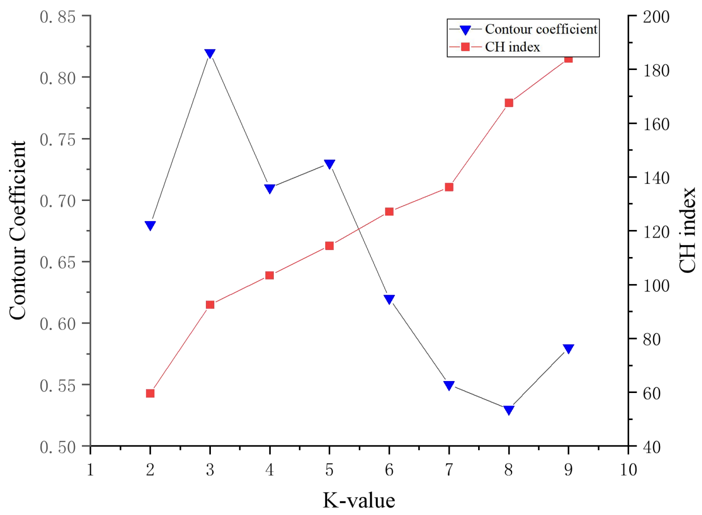


Fig. 3 Evaluation of clustering effects with different K-values

The characteristics of the 3 load classifications are shown in Fig. 4:

a) Load Category I: This type of load is at peak consumption during daytime (9.00-18.00) as compared to other time periods.

b) Load Category II: This load peaks in the early morning (1.00-5.00) and then falls back quickly and is in the trough in the afternoon (13.00-18.00).

c) Load Category III: The peak of this category occurs in the morning (6.00-8.00) and in the evening (18.00-20.00).

It is easy to see that the characteristics of the three classes of loads are obvious, indicating that the clustering effect of this paper is better and the classification results are stable. At the same time, the typical daily PV output curve of a household PV is normalized, as shown in Fig. 5

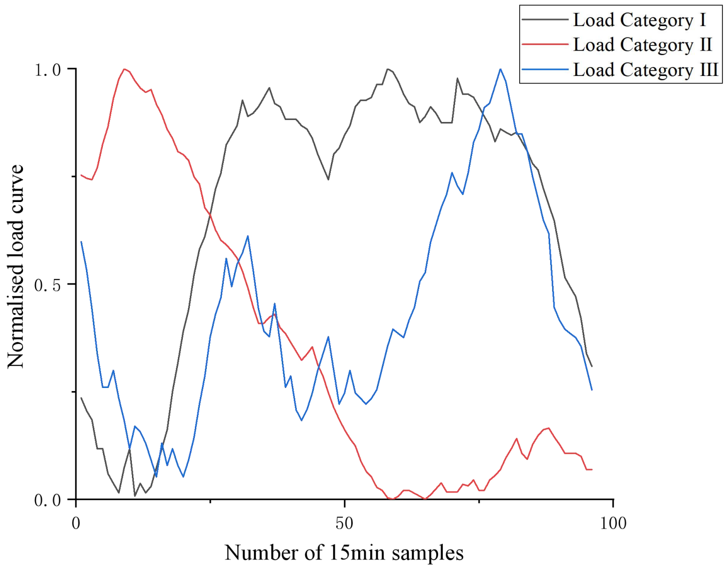


Fig. 4 Load clustering effect

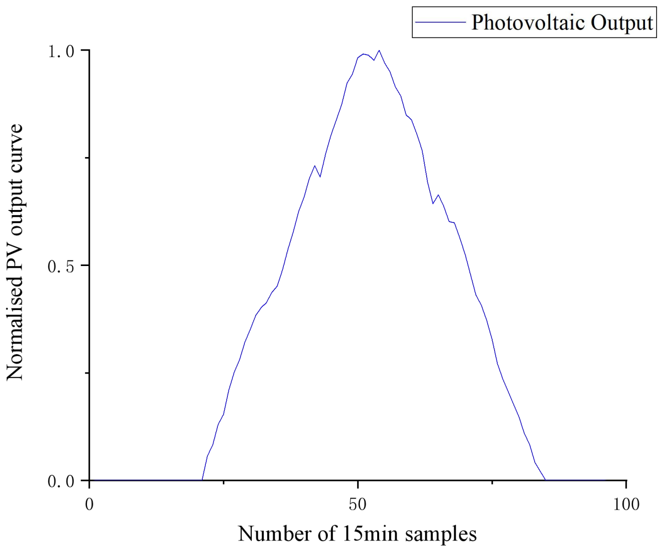


Fig. 5 Normalised PV power curve

Find the degree of similarity between the three and PV outflows, plotting Table 1.

Table 1. Load curve similarity and ranking

	Similarity	Sequence
Load Category I	4.38	2
Load Category II	5.88	3
Load Category III	3.97	1

Based on a section of the county distribution line structure, using the IEEE9 node distribution system for example simulation analysis, grid voltage of 10KV, grid frequency of 50Hz, unlimited high-power power supply and low-voltage radial feeder interconnections, conductor type JKLYJ-240, topology as shown in Figure 6, to load rated value of the three kinds of load centre point location, drawing Table 2.

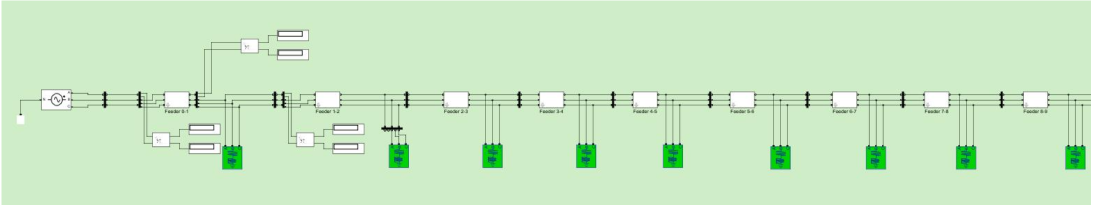


Fig. 6 Distribution line topology simulation

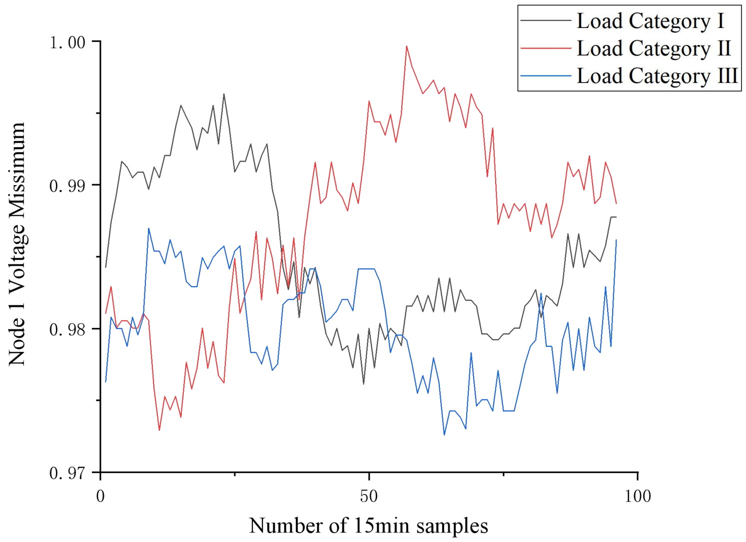
	Centre distance /km
Load Category I	2.16
Load Category II	5.31
Load Category III	7.63

Table 2. Load centre distance

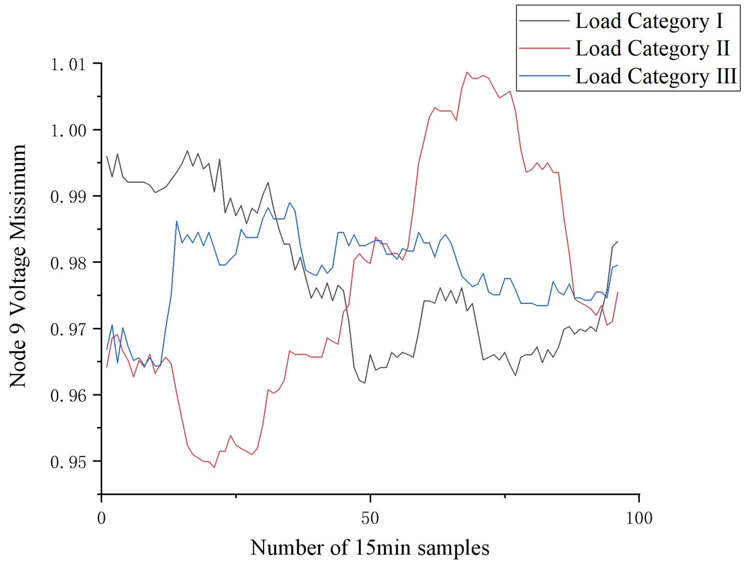
In the three types of load centre from the nearest node access to the PV power supply, the peak output of 100kW, considering the voltage overrun problem, selected node 1 and node 9 for stability analysis, the simulation results are shown in Figure 7.

Table 3 shows the standard deviation and relative change of the standard deviation of the voltage at the first and last node of the distribution line in the case of different PV installation locations, and it can be seen that as the similarity of the load centre of the PV installation point is continuously improved, the fluctuation of the voltage amplitude at the first and last node of the distribution line is gradually reduced. At the same time, regardless of the PV installation location, the distribution line first-end voltage fluctuations have been less than the distribution line first-end voltage fluctuations, and the first-end relative fluctuation difference is positively correlated with the load centre similarity of the PV installation point.

Based on the above analysis, it can be seen that the voltage stability of the distribution line has a greater relationship with the load centre similarity of the PV installation point; the higher the similarity, the better the line voltage stability. For the distribution line PV access planning, can give priority to choose the same type of load centre similarity is higher as the PV access point.



(a) Node 1



(b) Node 9

Fig.7 Voltage Missimum Curve

	Type I Load Centre		Type II Load Centre		Type III Load Centre	
	Node 1	Node 9	Node 1	Node 9	Node 1	Node 9
Standard deviation	0.005397	0.011073	0.006665	0.018174	0.003674	0.006006
Relative change	2.07%	3.60%	2.75%	6.29%	1.48%	2.56%

Table 3. Standard deviation and relative change of nodal voltage normals

5. Summary

This paper employs a DBSCAN density clustering and K-means optimization algorithm to classify line loads in the distribution network. A diverse collection of loads is then constructed, and the concept of load centre distance is introduced. The simulation and analysis of various load centres' access to PV on the node voltage of the distribution line is carried out based on field PV operation collection, load timing characteristics, and load centre measurements, among other data. The simulation examines how accessing PV at various load centres affects the voltage at distribution line nodes. The investigation indicates that the output size for PV scale determination is primarily linked to light intensity and temperature, both of which vary over time and influence the access point's flow direction. The simulation verification indicates a positive correlation between the voltage stability of distribution lines and the similarity of the load centre at the PV installation point.

However, the data collection time interval is considerable and the initial data may impact the clustering centre analysis. Furthermore, only the location of PV access is taken into account, and the influence of PV access capacity is not considered. In the forthcoming investigation, we shall further scrutinise the PV access capability whilst incorporating the PV access site for permutation analysis. This will enhance the practicality of the plan.

References

[1] WU Qiang, TU Kun, ZENG Yifan. Reflections on China's energy strategy under the vision of "dual-carbon" target[J]. Science Bulletin,2023,68(15):1884-1898.

[2] ELTAWIL M A,ZHAO zhengming . Grid-connected embedded generate on to long 11KV overhead lines[J].Power Engineering Journal,2002,16(1):5-12.

[3] HUANG Xinke,WANG Huan,WANG Yibo et al. Principle and strategy of voltage rise adjustment at grid-connected point of photovoltaic power generation system[J]. Power System Automation,2014,38(03):112-117.

[4] WU Xiaohui, LIU Jian, BI Pengxiang. Research on voltage stability of distribution network[J]. Grid Technology,2006(24):31-35.

[5] X. Zhang, L. Wei, S. Zhang. Study on the impact of photovoltaic distributed access on distribution network voltage stability[J]. Jilin Electric Power,2021,49(02):12-15.

[6] Zhang Shi, Li Weiguo, Tian Lei et al. Impact of distributed power supply access on distribution network voltage stability[J]. Distributed Energy,2017,2(04):36-39.

[7] WU Wenqing. Research on the impact of distributed photovoltaic grid integration on distribution network network loss[D]. Hebei University of Science and Technology,2022.

[8] YUAN Lintao, SUN Yuanyuan, LIU Zhen et al. Impact of reactive power and optimisation of low-voltage distributed photovoltaic in distribution network based on current transformation point[J]. Shandong Power Technology,2023,50(08):11-19+41.

[9] Zhuo may. Distributed photovoltaic access planning method considering power quality problems[D]. Wuhan University,2017.

[10] Sun Jianhao. Optimal allocation of distributed power supply taking into account demand response and timing characteristics[D]. Northeast Electric Power University,2023.

[11] Zhang Jiali. Research on clustering method based on improved DBSCAN[D]. Jilin University,2023.

[12] LI Dong,LIU Mengmeng,GUO Sha. Application of improved K_means based algorithm in image segmentation[J]. Computer Knowledge and Technology,2016,12(08):166-168.

[13] Bao H,Fang GJ. A new electrical distance calculation method using equivalent transmission impedance method and its application[J]. Grid Technology,2019,43(01):244-250.

[14] B. Zhang, J. Zhuang, J. Hu, et al. An integrated clustering algorithm for power load profiles combining dimensionality reduction techniques[J]. Chinese Journal of Electrical Engineering,2015,35(15):3741-3749.