Regional wind and solar resources clustering methods considering spatial-temporal characteristics to accelerate production cost simulation

Zixin Huang [1, a], Huawei Li [1, b], Pei Zhang [1, 2, c] , Long Zhao [3, d], Xiangfei Meng [3, e]

[1] School of Electrical Engineering, Beijing Jiaotong University, Beijing 100044, P. R. China;

[2] School of Electrical Automation and Information Engineering, Tianjin University, Tianjin 300072, P. R. China;

[3] Economic & Technology Research Institute of State Grid Shandong Electric Power Company, Jinan 250021, P. R. China;

[a] h13682866933@163.com, [b] hwli@bjtu.edu.cn, [c] peizhang166@qq.com, [d] 63650828@qq.com, [e] 316327192@qq.com

Abstract. Production cost simulation plays a crucial role in system planning, particularly for power systems with high renewable energy penetration. Balancing computational efficiency and accuracy remains a challenge. This paper introduces a regional division approach for wind and solar resources, utilizing spatiotemporal characteristics to significantly reduce model variables and constraints. Resources are first grouped into geographic grids based on latitude and longitude, with daily wind speed and solar irradiation averages capturing key renewable features. Annual feature vectors and regional correlations serve as clustering indices. The Davies-Bouldin criterion determines the optimal number of clusters, while K-means is employed for classification. A case study in Northwest China demonstrates that this method achieves substantial computational savings without compromising accuracy.

Keywords: Clustering classification; Spatial correlation; K-means; Production cost simulation.

1. Introduction

Power system optimization increasingly relies on production cost simulations for energy equilibrium analysis, economic planning, and renewable integration [1-4]. However, renewable intermittency poses operational challenges, making annual simulations essential for long-term planning. These models involve complex mixed-integer programming over 8760 hours, where system expansion exponentially increases computational demands, necessitating efficient simplification techniques.

Existing studies address this through variable and scenario reduction. Variable reduction clusters similar units and converts binary to integer variables to streamline optimization [5]. Scenario reduction employs temporal compression, such as typical-day selection and output clustering, to capture wind/solar variations while reducing computational costs [6,7]. Advanced clustering methods, including enhanced density peak and hybrid hierarchical-k-means, further improve scenario selection [8,9]. Despite these efforts, large-scale renewable integration remains computationally intensive due to geographic dispersion. Spatial correlations, such as wind farm coherence under shared weather systems, enhance clustering, while principal component analysis refines plant grouping via meteorological parameters [10,11].

To address these challenges, this paper proposes a regional wind and solar clustering method incorporating spatiotemporal characteristics to accelerate production cost simulation. The approach utilizes annual feature vectors and regional correlations as clustering indicators, comparing their effectiveness. By integrating spatial and temporal characteristics, the proposed method significantly reduces variables in annual production cost models.

2. Regional wind and solar resources clustering method

As different stations with similar geographical locations will be affected by the same weather system, the output power of power plants has spatial correlation. Therefore, the related power plants can be divided spatially. This paper partitions regions into grids based on the arrangement of longitude and latitude.

2.1 Clustering indices based on annual feature vectors

The clustering indices for wind resources are determined using the annual wind speed feature vector. The dimension of the annual wind speed feature vector in each grid is 1×365. The feature vector representing the annual wind speed at coordinate point x can be represented as:

$$T_{lat_x,lon_x}=\left[wind_{x,1},\cdots, wind_{x,t},\cdots, wind_{x,365}\right]_{1\times 365} \tag{1}$$

where lat_x denotes the latitude index of coordinate point x; lon_x denotes the longitude index of coordinate point x; $wind_{x,t}$ represents the wind speed corresponding to coordinate point x on day t.

Similarly, the annual irradiance feature vector for coordinate point x can be represented as:

$$T_{lat_x,lon_x}=\left[dswrf_{x,1},\cdots,dswrf_{x,t},\cdots,dswrf_{x,365}\right]_{1\times 365} \tag{2}$$

where $dswrf_{x,t}$ represents the downward shortwave radiation flux corresponding to coordinate point x on day t. The clustering indices for each coordinate point are determined by using the annual feature vector.

2.2 Clustering indices based on regional correlations

For wind resources, the daily wind speed data of each coordinate point are processed into a matrix with dimensions of the number of coordinate points×365. The annual wind speed matrix for an area containing N coordinate points can be expressed as follows:

$$W_N=\begin{bmatrix} wind_{1,1} & wind_{2,1} & \cdots & wind_{N,1} \\ \vdots & \vdots & \ddots & \vdots \\ wind_{1,t} & wind_{2,t} & \cdots & wind_{N,t} \\ \vdots & \vdots & \cdots & \vdots \\ wind_{1,365} & wind_{2,365} & \cdots & wind_{365,1} \end{bmatrix}_{N\times 365} \tag{3}$$

where $wind_{N,t}$ denotes the wind speed corresponding to coordinate point N on day t.

The column N in the annual wind speed matrix represents the annual wind speed feature vector of coordinate point N. A correlation coefficient matrix with dimensions of N×N is computed by calculating the correlation coefficient between each pair of coordinate points. The correlation coefficient is calculated using Pearson correlation coefficient:

$$\rho_{X,Y}=\frac{\mathrm{cov(X,Y)}}{\sigma_X\sigma_Y}=\frac{E[\mathrm{(X-EX)(Y-EY)}]}{\sigma_X\sigma_Y} \tag{4}$$

where X and Y denote two coordinate points; $\rho_{X,Y}$ denotes the Pearson correlation coefficient between the two coordinate points; $\mathrm{cov(X,Y)}$ denotes the covariance of the wind speed data between the two coordinate points; σ_X and σ_Y denote the standard deviation of the wind speed data at the two coordinate points; $E()$denote the mathematical expectation of the data.

Similarly, the annual irradiance matrix for an area containing N coordinate points can be expressed as follows:

$$D_N=\begin{bmatrix} dswrf_{1,1} & dswrf_{2,1} & \cdots & dswrf_{N,1} \\ \vdots & \vdots & \ddots & \vdots \\ dswrf_{1,t} & dswrf_{2,t} & \cdots & dswrf_{N,t} \\ \vdots & \vdots & \cdots & \vdots \\ dswrf_{1,365} & dswrf_{2,365} & \cdots & dswrf_{365,1} \end{bmatrix}_{N\times 365} \tag{5}$$

where $dswrf_{N,t}$ represents the downward shortwave radiation flux corresponding to coordinate point N on day t.

Column N in the annual irradiance matrix is the annual irradiance feature vector of coordinate point N. The Pearson correlation coefficient between each coordinate point is calculated to form a correlation coefficient matrix. The correlation coefficient between each coordinate point serves as the clustering indices.

2.3 K-means clustering process

The K-means algorithm is an iterative optimization method for partitioning data into distinct groups. Its workflow consists of four key stages:

(1)Initialization: Randomly select k data points as initial cluster centroids.

(2)Cluster Assignment: Compute the distance between each data point and the centroids, assigning each point to the nearest cluster to form k preliminary groups.

(3)Centroid Update: Recalculate the centroids by averaging all data points within each cluster, then reassign points based on the updated centroids.

(4)Termination: Repeat step 3 until centroids stabilize (no further changes) or a predefined iteration limit is reached.

The algorithm's performance heavily depends on the choice of k, which influences both clustering quality and computational efficiency. To determine the optimal k, this study employs the Davies-Bouldin (DB) Index, a metric that quantifies cluster validity by balancing intra-cluster cohesion and inter-cluster separation. The DB Index is defined as follows:

$$DB = \frac{1}{k}\sum_{i=1}^{k} R_i \tag{6}$$

where R_i denotes the worst intra-cluster to inter-cluster ratio of cluster i. The objective of the optimal clustering solution is to minimize the value of DB criterion.

3. Case studies

3.1 Dataset description

The datasets used in this study came from a provincial system of China in 2016.There are 348 generators in the datasets, including 53 thermal power generators, 2 hydropower generators,140 wind power generators and 153 solar power generators. These wind and solar power generators are distributed in 81 wind power plants and 127 solar power plants. The upper and lower limits of output, ramp rate and generation cost coefficient of each generator are obtained. Besides, the datasets also consist of historical wind and solar power data, load demand data and network data for 2016.Build production simulation problems using this system.

The latitude of coordinate points ranges from 59.875 degrees north latitude to 89.875 degrees south latitude. The longitude of the coordinated points ranges from 0.125 east longitude to 0.125 west longitude. The latitude and longitude of coordinate points are represented by indexes with an interval of 0.25 degrees. In this study, the latitude index ranges from 81 to 99, and the longitude index ranges from 416 to 431.

3.2 Clustering indices based on annual feature vectors

Firstly, the wind resources are clustered and partitioned. According to the latitude and longitude index of coordinate points, the annual wind speed of each coordinate point in the data is transformed into a 1 × 365-dimensional feature vector. The annual wind speed feature vector of each coordinate point is used as the clustering indicator. The DB criterion is used to determine the value of clustering K. The number of clusters is determined to be 3. The result of wind resource clustering division is shown in Fig. 1. In the figure, the horizontal and vertical coordinates represent the longitude and latitude indexes respectively. Points with different colors represent different categories of clustering division. The 140 wind generators in these 81 wind power plants can be aggregated to 3 wind generators equivalently.

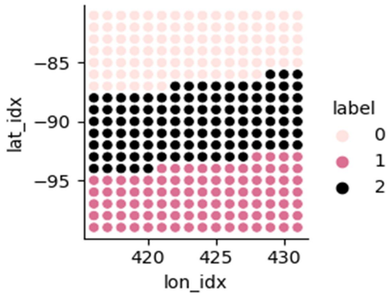

Fig. 1 The result of wind resource zoning with wind speed feature vector as clustering indicator

For solar resources, the annual irradiance feature vector is used as the clustering indicator.The DB criterion is used to determine the optimal number of clusters, which is set to 3.The K-means clustering method is applied to perform cluster partitioning. The result of solar resource clustering division is shown in Fig. 2. The 151 solar power generators in these 127 solar power plants can be aggregated to 3 solar power generators equivalently.

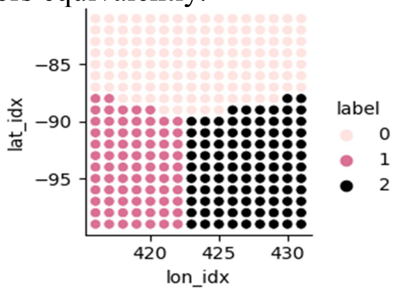

Fig. 2 The result of solar resource zoning with wind speed feature vector as clustering indicator

3.3 Clustering indices based on regional correlations

Firstly, the wind resources are clustered and partitioned. According to the latitude and longitude index of coordinate points, the annual wind speed of all coordinated points in the data is processed into a 304 × 365-dimensional matrix. According to the feature vectors, Pearson correlation coefficient formula is used to calculate the correlation coefficient matrix between coordinate points. The heat map of wind speed correlation matrix is shown in Fig. 3. The light-colored area along the diagonal is the area with high correlation.

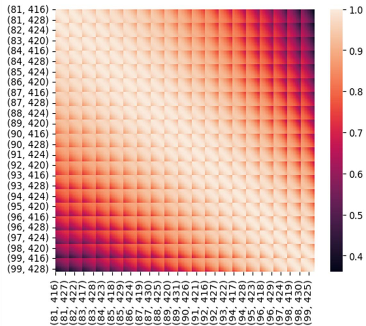

Fig. 3 Wind speed correlation matrix heat map

The result of wind resource clustering division is shown in Fig. 4.

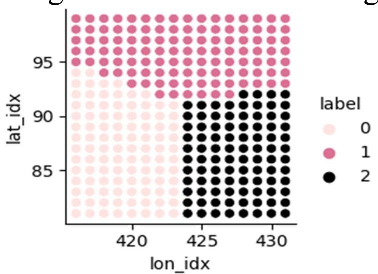

Fig. 4 Wind resource zoning with regional relevancy as clustering indices

For solar resources, Pearson correlation coefficient formula is used to calculate the correlation coefficient matrix between coordinate points. The heat map of irradiance correlation matrix is shown in Fig. 5.

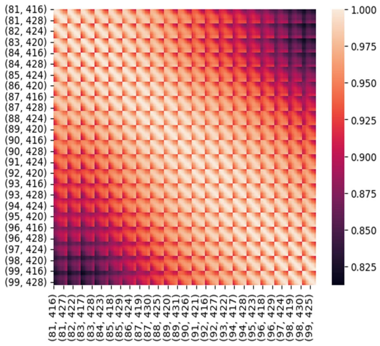

Fig. 5 Irradiance correlation matrix heat map

The result of solar resource clustering division is shown in Fig. 6.

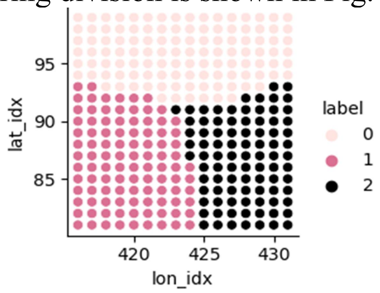

Fig. 6 Solar resource zoning with regional relevancy as clustering indices

3.4 Clustering indices based on regional correlations

The annual power generation of each station in a given region is divided by the installed capacity, that is, the number of annual utilization hours, serving as an indicator to validate the accuracy of the partitioning. Within the same region, wind farms and solar power plants share similar resources so their annual utilization hours should be approximately equivalent. After clustering, the standard deviation of wind farms in each region is shown in Table 1. The annual utilization hours curve is shown in Fig. 7 .

Table 1. The standard deviation of annual utilization hours in each region

Standard deviation	Region 0	Region 1	Region 2
Clustering indices based on regional correlations	763	701	1287
Clustering indices based on annual characteristic vectors	2032	2010	3481

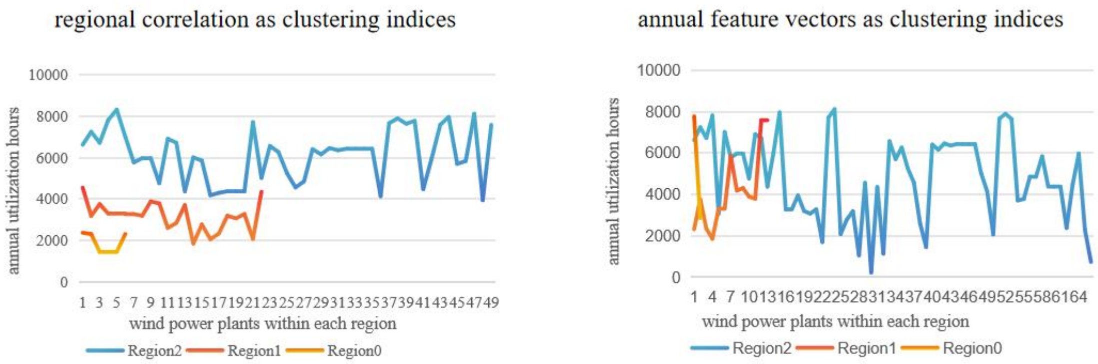

Fig. 7 Comparison of annual utilization hours of wind farms using two clustering indicators

Using regional correlations as cluster indices, the standard deviation of annual utilization hours for wind farms in each region is small. At the same time, there is a significant disparity in annual utilization hours across different regions. However, when annual feature vectors are utilized as the clustering index, the annual utilization hours of wind farms in each region show a relatively discrete distribution pattern. Therefore, indicating that utilizing regional correlations as clustering indices leads to more accurate results.

Similarly, cluster the data of solar power plants. After clustering, the standard deviation of solar power plant in each region is shown in Table 2, and the annual utilization hours curve is shown in Fig. 8.

Table 2 The standard deviation of annual utilization hours in each region

Standard deviation	Region 0	Region 1	Region 2
Clustering indices based on regional correlations	424	712	1116
Clustering indices based on annual characteristic vectors	1848	1445	1632

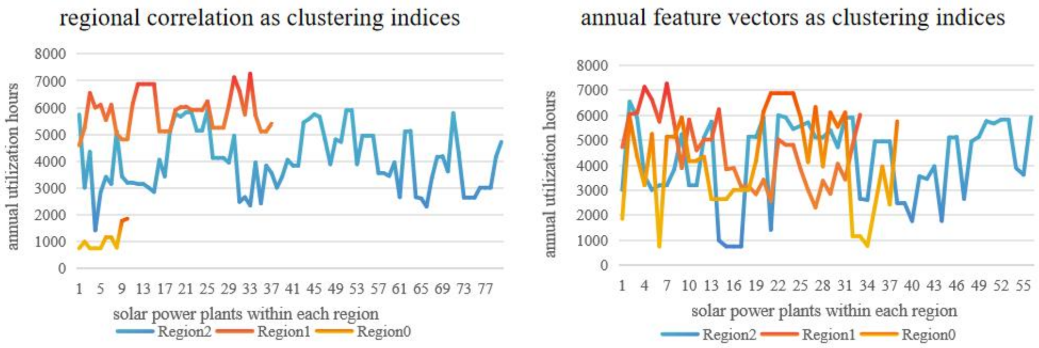

Fig. 8 Comparison of annual utilization hours of solar power plants using two clustering indicators

According to the comparisons, the spatial-temporal characteristic region partition method using regional correlation as the clustering indices is superior to the spatial-temporal characteristic region partition method using the annual feature vectors as the clustering indices.

3.5 Simulation result analysis

The output of each generator can be taken as a dynamically controllable variable. The objective is to minimize the total power generation cost of all generations, with 1 hour as the time scale, and the entire year divided into 8760 time periods. Based on the datasets collected for a period of 6 months, the production simulation of the partitioned generators is conducted with 1 hour as the time scale. Taking 24 hours a day as an example, it can be seen that after clustering the number of variables and constraints is greatly reduced. At the same time, the operation time is significantly reduced compared to the unpartitioned case, and the operation speed is greatly improved.

Table 3. The actual model scale before and after clustering

	Number of variables	Number of wind and solar constraints
With clustering partition	144	168
Without clustering partition	7032	7056

Table 4. The actual total solve time before and after clustering

Total Steps	24h	744h	2184h	4368h
With clustering partition	30s	1560s	9250s	25758s
Without clustering partition	123s	13662s	117080s	455161s

Table 5. The objective function before and after clustering

Total Steps	With clustering partition/yuan	Without clustering partition/yuan	Error rate/%
24h	25358296.9326	25357892.3042	0.0016%

744h	681781870.2732	681764061.5978	0.0026%
2184h	1873463060.0298	1873354297.7467	0.0058%
4368h	3871447331.1208	3870972565.3451	0.0123%

After partitioning, the values of the objective function show a certain level of error compared to those before partitioning. However, the error rates are minimal, ranging from 0.0016% to 0.0123%. The disparities in model results before and after partitioning fall within an acceptable range. Considering the solving time, it is evident that the algorithm proposed in this paper significantly enhances solving efficiency while ensuring high-precision calculation results.

4. Conclusion

This paper introduces a novel region partitioning method that leverages the unique spatiotemporal features of wind and solar resources. Its effectiveness is validated in production cost simulation with an actual provincial system data, analyzing the pre and post partitioning scenarios. The subsequent conclusions shed light on the method's capabilities and impact:

The clustering partition method considering temporal and spatial characteristics is based on the fluctuation of regional wind speed and irradiance and the correlation between regions. It can better reflect the real characteristics of the original scene in the space. In practical application, clustering partition can reduce the calculation scale by considering the temporal and spatial characteristics. The method using regional correlation as the clustering indices can further reduce computation time and improve efficiency. It is beneficial for power sector in power grid planning, reducing computational complexity and improving economic feasibility.

References

[1] Hu, QR.; Guo, ZS.; Li, FX. Imitation learning based fast power system production cost minimization simulation. IEEE Trans. Power Syst. 2023, 38, 2951-2954.

[2] Jiang, H.; Zhang, H.; Shi, X. Refined production simulation and operation cost evaluation for power system with high proportion of renewable energy. Energy Rep. 2022, 8, 108-118.

[3] Liang, C.; Meng, J.; Chen, C.; Zhou, Y. A production-cost-simulation-based method for optimal planning of the grid interconnection between countries with rich hydro energy. Global Energy Interconnection 2022, 3, 23-29.

[4] E Kabir, P Kumar, S Kumar, et al. (2018) Solar energy: potential and future prospects. Renewable and Sustainable Energy Reviews, 82:894-900

[5] Dai Jiang, Tian Nianjie, Jiang youquan, et al. Collaborative maintenance scheduling and unit commitment for hydropower and thermal power systems considering cascade hydropower coupling[J]. Electric Power Engineering Technology, 2022,41(03):83-91.

[6] Huang Feng . Generation method of typical renewable energy scenarios for power system analysis [D]. Hefei University of Technology, 2019.

[7] Ding Ming, Xie Xiaolong, Liu Xinyu, et al. The Generation Method and Application of Wind Resources/Load Typical Scenario Set for Evaluation of Wind Power Grid Integration [J]. Proceedings of the CSEE, 2016,36 (15): 4064-4072.

[8] Yuehui, et al. An annual wind power planning method based on time sequential simulations[J]. Automation of Electric Power Systems, 2014, 38(11): 13-19.

[9] Sun Huijuan, Liu Jun, Peng Chunhua. Multi-objective planning of distributed generation based on classified probability comprehensive multi-scenario analysis [J]. Electric Power Automation Equipment, 2018,38 (12): 39-45. DOI: 10.16081/J.ISSN.1006-6047.2018.

[10] Xiao Ziniu, Lu Zongxiang, Ren Zhouyang, et al. Medium and long-term electricity forecasting and optimal dispatching of new energy sources [M]. Beijing: Meteorological Press, 2021: 5-6.