International Journal of Gaming and Computer-Mediated Simulations
Volume 16 • Issue 1 • January-December 2024

Simulation Framework for Substation Siting Integrating Load, Land Use, Neighborhood, and Cost Analysis

Jing Xiong
Shaanxi University of Science and Technology, China

Jinming Yang
Shaanxi University of Science and Technology, China

Yueling Deng
Shaanxi University of Science and Technology, China

Zixuan Chen
Shaanxi University of Science and Technology, China

Jihaoyu Yang
Shaanxi University of Science and Technology, China

Chen Xu
Shaanxi University of Science and Technology, China

Yong Qi
Shaanxi University of Science and Technology, China

ABSTRACT

This study presents a comprehensive framework for substation siting to address power demand and urbanization challenges. It integrates economic, social, and environmental dimensions to ensure reliable power supply systems. Using simulation techniques, the paper evaluates substation placement by analyzing load distribution, land use, neighborhood satisfaction, and construction costs. The Analytical Hierarchy Process (AHP) assesses load distribution, while a hierarchical clustering algorithm maximizes land use efficiency. Neighborhood satisfaction is measured using hierarchical analysis and fuzzy logic. Construction costs are optimized via a genetic algorithm. These simulations formulate a multi-criteria decision-making tool, proposing an optimized siting strategy balancing technical, socio-economic, and environmental considerations. The framework enhances the accuracy of substation siting studies and provides practical implementation guidance, offering new insights and refined methodologies for modern power system planning and development.

KEYWORDS

Substation Siting, Simulation, Hierarchical Analysis, Hierarchical Clustering Algorithm, Fuzzy Comprehensive Evaluation, Genetic Algorithm

INTRODUCTION

With the continuous growth of power demand and the acceleration of urbanization, the siting and construction of substations have become crucial to ensuring the stable operation of power systems and meeting economic and social development needs. Despite numerous studies in this area, substation siting remains a complex challenge. The problem we address in this study is the lack of a comprehensive framework that integrates economic, social, and environmental factors in substation siting. Current methods often focus on specific technical aspects without fully considering the multidimensional nature of the problem.

Substation siting must account for economic benefits, social impacts, and environmental considerations to ensure reliability and stability. This makes it a multicriteria, multiobjective

DOI: 10.4018/IJGCMS.356272

decision-making problem. The development of new technologies and evolving social needs necessitate more scientific and efficient site selection methods, which have significant theoretical and practical implications.

Previous research has primarily utilized optimization algorithms, geographic information systems (GISs), and multicriteria decision-making methods to address substation siting. For example, heuristic algorithms like genetic algorithms (GAs) and simulated annealing, combined with GIS spatial analysis, help identify optimal siting plans based on geographical, environmental, and economic factors. Multicriteria decision support systems balance various decision criteria under complex conditions (Zeng, 2022; Sun et al., 2020).

Recent advancements have introduced new methodologies to enhance the siting process. Improved firefly algorithms optimize substation siting and capacity determination (Li, 2022). Gas-insulated switchgear technology benefits high pollution environments (Wu, 2019). Hierarchical analysis methods systematically evaluate multiple criteria for high voltage substation siting (Luo et al., 2021). Integrating GISs and variable weight models offers a comprehensive framework for evaluating multiple factors (L. K. Xu et al., 2022). Environmental impacts like noise pollution must be evaluated to ensure compliance with standards (Li, 2022).

However, these studies often focus on specific techniques or methods and lack the comprehensive consideration of the multidimensional influencing factors of the substation siting problem. This study aims to construct a more comprehensive analytical framework for substation siting by integrating key factors such as load distribution, land use, residential satisfaction, and construction cost. Through the in-depth analysis and modeling of different factors, this paper aims to propose a comprehensive assessment method to support more scientific and rational siting decisions.

RELATED WORK

Recent studies have explored various methodologies and tools to improve substation siting decisions. For instance, Zeng (2022) proposed a method for 110kV substation siting based on current distribution network and load conditions, emphasizing the importance of considering both technical and economic factors in the decision-making process. Similarly, Țboacă-Ciupăgeanu and Țboacă-Ciupăgeanu (2024) explored optimal substation placement through machine learning, presenting it as a sustainable solution for electrical grids.

Mirshekali et al. (2023) developed a deep learning-based framework for fault location in power distribution grids, emphasizing its potential in improving substation fault management. Yao et al. (2023) proposed a microscale substation siting framework utilizing spatial optimization and geospatial big data, underlining the significance of integrating machine learning in spatial decision making.

Li (2022) introduced an improved firefly algorithm for substation siting and capacity determination, highlighting its advantages in handling multiobjective optimization problems. Building on similar optimization techniques, Hrgović and Pavić (2024) developed a substation reconfiguration selection algorithm based on power transfer distribution factors and reinforcement learning, demonstrating its effectiveness in managing congestion. Garcia et al. (2022) employed machine learning to estimate circuit topology at the substation level, indicating its benefits in communication-limited environments.

Azhar et al. (2022) used machine learning to assist in selecting communication technologies for distribution substations, showing how decision makers can optimize technology choices. This was complemented by the work of Dai et al. (2023), who developed an electric-vehicle charging station siting model using machine learning, emphasizing its applicability in optimizing energy infrastructure siting decisions.

In another significant contribution, Kang and Kim (2022) applied machine learning for fault location estimation in power distribution networks, illustrating its adaptability under constrained conditions. Similarly, Livani (2024) utilized supervised learning for fault location in power grids, underscoring the method's precision in detecting and locating faults, which is critical for enhancing grid resilience. Trivedi et al. (2024) extended these efforts by integrating analytical hierarchy process-technique for order of preference by similarity to ideal solution (TOPSIS) methodologies in substation technology selection, particularly in environmentally sensitive areas.

Rizeakos et al. (2023) developed a deep learning application for fault location and classification in distribution grids, demonstrating its utility in active grid management. This is echoed by the work of Mirshekali et al. (2022), who introduced a machine learning-based fault location method tailored for smart distribution networks equipped with micro-phasor measurement units (PMUs) highlighting the advantages of leveraging high-resolution data.

In terms of environmental considerations, Azhar et al. (2023) proposed criteria for selecting communication technologies for substations, incorporating machine learning to enhance decision-making processes. Meanwhile, Elomiya et al. (2024) presented a hybrid suitability mapping model integrating GIS, machine learning, and multicriteria decision analytics, aimed at optimizing service quality of electric-vehicle charging stations, further demonstrating the versatility of machine learning in energy infrastructure planning.

Nhung-Nguyen et al. (2024) investigated anomaly detection using machine learning in digital substations, showcasing its effectiveness in enhancing operational reliability. In a related effort, Poudel et al. (2022) introduced a zonal machine learning-based protection strategy for distribution systems, demonstrating its capability in optimizing protection schemes.

Townsend and Reid (2022) presented a three-dimensional fault location program using machine learning for high-voltage substations, which integrates multilateration and geolocation techniques. Tavoosi et al. (2022) combined impedance-based techniques with machine learning for fault location in power networks, highlighting the hybrid approach's effectiveness in enhancing accuracy.

Joga et al. (2023) utilized a novel graph search and machine learning method to detect and locate high impedance fault zones in distribution systems, showcasing its potential to improve reliability and efficiency in power networks. Noebels et al. (2022) further expanded on machine learning applications by proposing a real-time selection of preventive actions to improve power network resilience, reinforcing the growing importance of real-time data in decision-making processes.

These studies collectively contributed to the advancement of substation siting methodologies, integrating advanced technologies and comprehensive evaluation frameworks to address the multifaceted challenges of modern power distribution networks.

MATERIALS AND METHODS

Model Preparation

Data Sources and Interpretation

This paper assesses the impact of substation siting from several perspectives, including load distribution, land use, neighborhood satisfaction, and construction costs. Each perspective can be evaluated by specific indicators.

Load distribution: The impact of load distribution on substation siting, reflected in the optimization of economic costs and the improvement of grid stability, which determines its efficiency and sustainability.

Land use: Evaluating the impact of substation construction on land use, including indicators such as land use area, land utilization rate, land use efficiency, and environmental protection index.

Neighborhood satisfaction: Assessing the impact of substation construction on neighborhood satisfaction by evaluating the quality of the substation, environmental impacts, traffic impacts, and safety indices.

Construction cost: Evaluating the construction cost of the substation, considering factors such as the generation of demand points, the determination of the number of people, the amount of material demand, the difficulty level of material transportation, and the cost of transportation.

Our data set comprises data from three substations in western China, referred to as G, Q, and S. To ensure the robustness and reliability of our findings, we collected comprehensive data from multiple sources. The data used in this study were obtained from four sources:

Load distribution data: The load distribution data were sourced from publicly available reports and data sets provided by the State Grid Corporation of China for the G, Q, and S substations. These data sets include hourly load demands for various regions over a five-year period (2018-2022), offering a robust temporal and spatial representation of load distribution patterns.

Land use data: Land use data were acquired from satellite imagery and GIS databases maintained by the Ministry of Natural Resources of China. These data included detailed land use classifications, land utilization rates, and environmental protection indices for urban and rural areas.

Neighborhood satisfaction data: Neighborhood satisfaction data were collected through surveys conducted in regions near the G, Q, and S substations. The surveys included questions on environmental impact, noise levels, air quality, transportation accessibility, electromagnetic field exposure, and safety concerns. Feedback was gathered from over 1,000 residents to ensure a diverse and representative sample.

Construction cost data: Construction cost data were compiled from project reports and financial records of the G, Q, and S substation construction projects. These data included material requirements, transportation costs, construction equipment costs, and project management expenses. Table 1 shows the data set elements.

Data Preprocessing

Standardized Processing

Key performance indicators have different ranges for different applications and systems. Normalization removes the range differences and helps to calculate the similarity between key performance indicators. For these data, we use the normalization formula, as shown in Equation 1.

$$\hat{x}_t = \frac{x_t - \mu_x}{\sigma_x} \tag{1}$$

PauTa Criterion

PauTa criterion: Let the measured variable be measured with the same precision, obtain, calculate its arithmetic mean and residual error, and calculate the standard error according to Bessel's formula; if the residual error of the measured value meets the requirements, the value is considered to be a bad value with a coarse error value, and it should be eliminated. Bessel's formula is as shown in Equation 2.

$$\sigma = \left[\frac{1}{n-1}\sum_{i=1}^{n} v_i^2\right]^{1/2} = \left\{\left[\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2 / n\right] / (n-1)\right\}^{1/2} \tag{2}$$

In Equation 2, to quantify the identification effectiveness of each method, a robust distance value method for outlier data is used. The method for detecting outliers is performed by calculating the sum of the distances between each observation and the mean value. MAD (mean absolute deviation) is used to further quantify the distance value of each group of data with respect to the concentration value, and the concentration value of the distance value is used to calculate the degree of deviation of the distance value of each group of data. This method is mainly applicable to normally distributed data sets, and lognormally distributed data must be logarithmized before it can be used.

Table 1. Data set elements

Module	Column Name	Description
Load Distribution	Hour	Hour of data recording
Load Distribution	Substation G Load Demand (MW)	Hourly average load demand of Substation G (MW)
Load Distribution	Substation Q Load Demand (MW)	Hourly average load demand of Substation Q (MW)
Load Distribution	Substation S Load Demand (MW)	Hourly average load demand of Substation S (MW)
Land Use	Substation	Identifier for the substation (G, Q, S)
Land Use	Land Use Area (km^2)	Land area used by the substation (in km^2)
Land Use	Land Utilization Rate (%)	Percentage of land utilization
Land Use	Environmental Protection Index	Environmental impact score (0 to 1)
Neighborhood Satisfaction	Substation	Identifier for the substation (G, Q, S)
Neighborhood Satisfaction	Environmental Impact	Satisfaction score for environmental impact (0 to 1)
Neighborhood Satisfaction	Noise Level	Satisfaction score for noise level (0 to 1)
Neighborhood Satisfaction	Air Quality	Satisfaction score for air quality (0 to 1)
Neighborhood Satisfaction	Transportation Accessibility	Satisfaction score for transportation accessibility (0 to 1)
Neighborhood Satisfaction	EMF Exposure	Satisfaction score for electromagnetic field exposure (0 to 1)
Neighborhood Satisfaction	Safety Concerns	Satisfaction score for safety concerns (0 to 1)
Neighborhood Satisfaction	Overall Satisfaction	Overall satisfaction score (0 to 1)
Construction Costs	Substation	Identifier for the substation (G, Q, S)
Construction Costs	Material Cost (CNY Million)	Cost of materials for construction (in million CNY)
Construction Costs	Transportation Cost (CNY Million)	Cost of transporting materials (in million CNY)
Construction Costs	Equipment Cost (CNY Million)	Cost of equipment for construction (in million CNY)
Construction Costs	Management Cost (CNY Million)	Cost of management during construction (in million CNY)
Construction Costs	Total Construction Cost (CNY Million)	Total cost of construction (in million CNY)

In view of the actual situation of the problem, we have improved the above calculation method, and the improvement steps are:

1. Calculate the median of each index before and after anomaly identification.
2. Calculate the absolute deviation from the median of each index before and after the anomaly identification.
3. Calculate the median of absolute deviation before and after anomaly identification.
4. Calculate the ratio of the absolute deviation values before and after anomaly identification to the median absolute deviation value, with MAD as the center distance value, as shown in Equation 3.

$$dv_i = \frac{\left|X_i - (X_i)_{med}\right|}{\left[\left|X_i - (X_i)_{med}\right|\right]_{med}} \tag{3}$$

5. The ratio of the maximum distance value to the mean value of each metric before and after anomaly identification was taken as the deviation of the distance value, as shown in Equation 4.

$$D_i = \frac{dv_{\max} - \overline{dv_i}}{dv'_{\max} - \overline{dv_i}} \tag{4}$$

Where, in Equations 3 and 4, the distance from the center of the indicator is the measured value of the indicator, the median of the indicator, the maximum distance after the anomalies are identified and eliminated, the maximum distance before the anomalies are identified, the degree of deviation from the value of the distance, and the mean value of the center of the distance.

Load Distribution Impact Modeling

In the current field of power system planning and design, load distribution has a direct impact on the economic efficiency and grid stability of substation siting, which requires site optimization to reduce transmission and distribution losses and investment costs. In this study, a comprehensive load distribution impact model is proposed, which aims to comprehensively evaluate the factors of load distribution affecting substation siting. These factors include the direct influence of load distribution, the stability and anti-interference ability of the power grid, the scientific and rationality of site selection, the relative distance to the load center, and the accuracy of load forecasting. By mathematically describing these factors, the model not only ensures a high degree of accuracy, but also guarantees its scientificity and wide applicability.

First of all, the investment cost required for substation design and sizing (C), as a core indicator for economic analysis of power systems, is closely related to the amount of load (L) and the distance from the load center (D). This relationship reveals the combined effect of load and distance on investment cost through explicit expression. This formula is based on practical experience and theoretical analysis, aiming at optimizing resource allocation and maximizing economic benefits.

Next, the stability (S) and immunity (V) of the grid are the key to ensure the reliability and security of power supply. These two parameters are evaluated by the equalization of the load distribution and quantified by $S = g(L, V)$. This function reflects how to improve the overall stability of the grid and its ability to withstand external disturbances, by optimizing the load distribution, and is an indispensable consideration in power system design and planning.

In addition, the scientific validity and reasonableness (R) of siting is assessed by the accuracy of load forecasting, as indicated by $R = h(P, L)$. This aspect highlights the important role of high-quality data and advanced forecasting techniques in ensuring the rationality of siting decisions. The model evaluates the accuracy of the forecasting tool and the effectiveness of the siting decision by comparing the predicted load (P) with the actual load (L).

The relative distance of the substation from the load center (D) directly affects the length of the transmission and distribution lines and the power loss (E), which is calculated by the formula $E = k \cdot D^2 \cdot L$. This relationship illustrates the importance of considering distance in the siting process to minimize energy losses and costs.

The accuracy of load forecasting is critical to planning the size and configuration of substations, and the forecast error is denoted by $E_p = |P - L|$. Accurate forecasting ensures the effectiveness of power system planning and reduces the risk of over-investment and resource wastage.

Finally, the comprehensive assessment through the analytical hierarchy process integrates the above factors and provides a multidimensional and all-encompassing assessment of the siting options through a composite score of $S_c = \sum_{i=1}^{n} w_i \cdot F_i$ calculated from the weights of the factors w_i and the score F_i. This comprehensive assessment method considers the multiple influencing factors of substation siting, which makes the decision-making process more scientific, rational, and adaptable to the ever-changing power market and technological environment.

A comprehensive load distribution impact model is constructed to support the decision-making process of substation siting. The mathematical method of the model not only enhances the accuracy and depth of the theoretical analysis but also provides reliable guidance for practical application, promoting the scientific and efficient power system planning and design.

Land Use Model

The land use optimization model can take the land use area, land use rate, land use efficiency, environmental protection index, and so on as constraints in the indicators of the substation and take the maximization of land use as the objective function. Through the optimization algorithm the optimal land use scheme can be obtained, to realize the maximization of land use efficiency of the substation.

At present, most of the similarity measures of hierarchical clustering algorithms use single connection, full connection, or average connection. Single connection means that given two classes, in which the samples are denoted as $C_i C_j p_i p_j$, the Euclidean distance between the two samples closest to each other between the subclasses is used as the similarity between the subclasses. The single connection is shown in Equation 5.

$$dist\left(C_i, C_j\right) = \min\left\{ dist(p_i, p_j) \middle| p_i \in C_i, p_j \in C_j \right\} \tag{5}$$

The fully connected approach is to use the Euclidean distance between the two most distant samples as the similarity between the subclasses as shown in Equation 6.

$$dist\left(C_i, C_j\right) = \max\left\{ dist(p_i, p_j) \middle| p_i \in C_i, p_j \in C_j \right\} \tag{6}$$

Mean connectivity is based on the average distance between two samples within a subclass as the similarity between subclasses, as shown in Equation 7.

$$dist\left(C_i, C_j\right) = average\left\{ dist(p_i, p_j) \middle| p_i \in C_i, p_j \in C_j \right\} \tag{7}$$

The principle of hierarchical clustering is shown in Figure 1.

Neighborhood Satisfaction Model

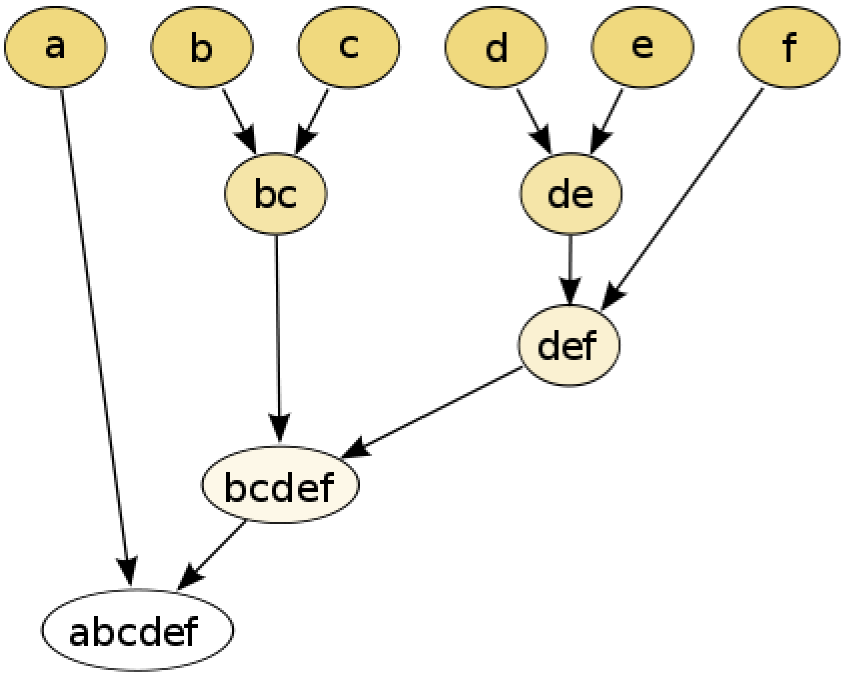
Hierarchical analysis can be used to establish a system of satisfaction indicators for residents near the selected site, with environmental impact, noise level, air quality, transportation accessibility, and other indicators as the levels of the hierarchy, and then the weights of the indicators are determined through the feedback from residents' surveys or expert evaluations. Finally, the residents' satisfaction index is calculated by weighted average.

Basic Steps in Analyzing Hierarchy

Establishment of a hierarchical chart of analysis generally includes an objective level, criterion level, and indicator level.

Adopt the mature "1~9 scale method" for evaluation indicators. Compare and analyze the indicators at the same level, determine the relative importance, and construct the judgment matrix. Calculate the maximum eigenvalue and eigenvector of the judgment matrix, and normalize the

Figure 1. Structure diagram of hierarchical clustering principle



eigenvector to the weight of the corresponding indicator of the same level to the relative importance of the indicator of the previous level. Then verify the consistency.

The total ranking weight of a factor at a level is equal to the sum of the product of the single-level ranking weight of the relevant factor at the previous level and the total ranking weight of the relevant factor at the previous level. The total CI and total RI of the total ranking of a level are calculated from the CI and RI of the single-level ranking of the relevant factor at the previous level.

Determination of Indicator Weights

Based on the public data released by the state grid, the hierarchical analysis method discussed in the previous section is used to calculate the index weights. The judgment matrix of correlation coefficients between standard and target layers is constructed, and the maximum eigenvalue and eigenvector of the matrix are calculated by using MATLAB software. Subsequently, the consistency is verified. If the consistency test is met, the eigenvectors associated with the largest eigenvalue are normalized to derive the weights of the corresponding indicators. If the consistency test is not passed, the judgment matrix must be reconstructed until the consistency test is passed. Figure 2 shows the judgment matrices, consistency test results, and index weights of the target and standard layers.

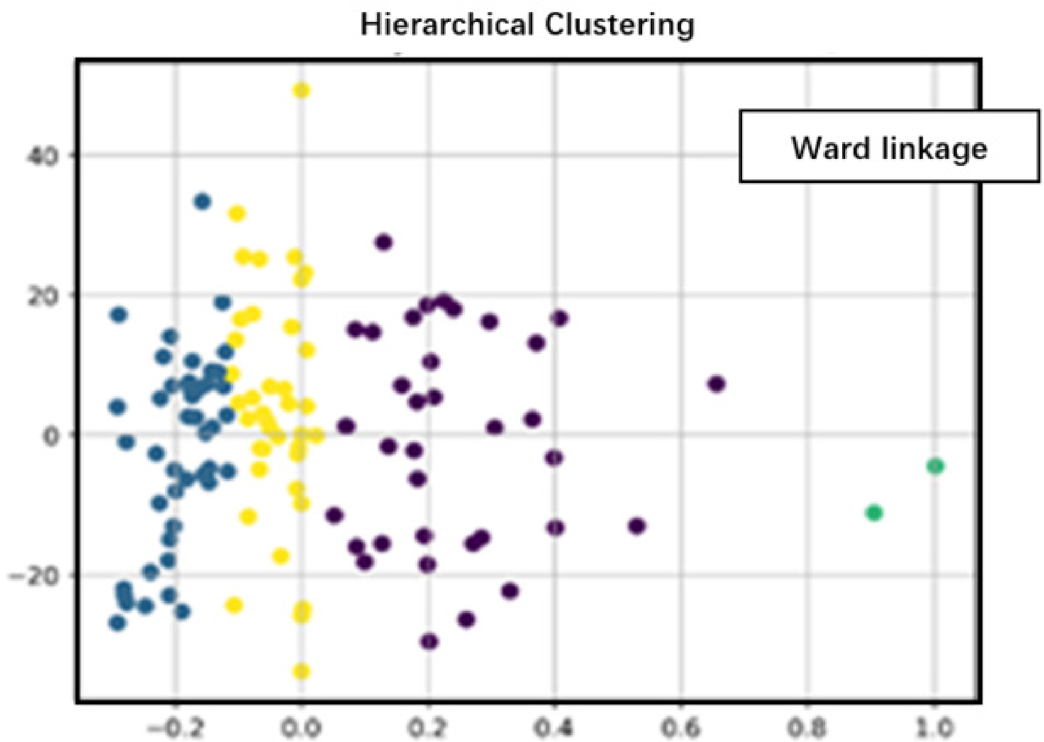
Compute the weight vector and consistency test for hierarchical unit ranking:

The maximum eigenvalue of the pairwise comparison matrix A1 is λ=11.2200, and the corresponding normalized eigenvector is ω =(0.1311, 0.0380, 0.0511, 0.2107, 0.1235, 0.1227, 0.0240, 0.1347, 0.0416, 0.0927, 0.0178, 0.0122).

$$RuleCI = \frac{\lambda\text{-N}}{\text{N-1}} = 0.1536.$$

Therefore, it passes the consistency test. The consistency of this matrix is consistent with$CR = \frac{CI}{RI} = 0.0910 < 0.1$.

Figure 2. Judgment matrix for hierarchical analysis



Fuzzy Integrated Evaluation

1. Determination of the matrix of eigenvalues of the indicators.

There is a set of samples to identify the risk of the resource system, each sample is represented by m indicators, and the $m \times n$ matrix of eigenvalues of the sample set of indicators is as shown in Equation 8.

$$X = (x_{ij}) = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \tag{8}$$

Where, in Equation 8, x_{ij} is the eigenvalue of sample j indicator i ($i = 1,2,3,\ldots,m$; $j = 1,2,3,\ldots n.$).

2. Matrix of standardized eigenvalues of indicators.

Determine the sample set according to the standardized eigenvalues of m indicators and c rank indicators, so as to get a matrix of standardized eigenvalues of indicators of size$m \times c$. Where $h = 1,2,3,\ldots,c$ is the standardized eigenvalue of the rank indicators.

3. Calculate the relative membership matrix.

Based on the definition of the relative membership function, the relative membership matrix of the indicator eigenvalues and the fuzzy concept $R = (r_{ij})$ is determined.

The formula for the relative membership function of the forward indicator pair is as shown in Equation 9.

$$r_{ij} = \begin{cases} 0 & x_{ij} < y_{i\min} \\ \dfrac{x_{ij} - y_{i\min}}{y_{i\max} - y_{i\min}} & y_{i\min} < x_{ij} < y_{i\max} \\ 1 & x_{ij} > y_{i\max} \end{cases} \quad (9)$$

The formula for the relative membership function of the inverse pair is as shown in Equation 10.

$$\mathrm{r}_{ij} = \begin{cases} 0 & \mathrm{x}_{ij} > y_{i\max} \\ \dfrac{y_{i\max} - x_{ij}}{y_{i\max} - y_{i\min}} & y_{i\min} < x_{ij} < y_{i\max} \\ 1 & x_{ij} < y_{i\min} \end{cases} \quad (10)$$

Where, in Equation 10, $\mathrm{y}_{i\min}, y_{i\max}$ are the minimum and maximum values of the indicator criteria, respectively.

Similarly, the relative membership matrix between the standardized eigenvalues of the indicator and the fuzzy concept $S = (s_{ih})$ is calculated.

4. Determine the fuzzy weight vector of the indicators.

Since the indicators in the sample set have different degrees of influence on fuzzy identification, the weights of the indicators are different. Let the indicator weight vector be $W = (w_1, w_2, w_3, \ldots, w_m)$, and satisfy the normalization condition $\sum_{i=1}^{m} w_i = 1$.

5. Determine the fuzzy pattern recognition model.

Take the relative membership matrix $U = (u_{hj})$ of the sample set for each class and satisfy the constraint $\sum_{h=1}^{c} u_{hj} = 1$.

Weight the degree of membership of the samples to each risk class, and set the difference between the samples and the risk class indicator. Let the sum of squares of the weighted generalized distances be minimized, and construct the Lagrangian function according to the constraints, and finally obtain the pattern recognition model as shown in Equation 11.

$$u_{hj} = \frac{1}{\sum_{i=1}^{m} \left[w_i(r_{ij} - s_{ih})\right]^2 \sum_{k=1}^{c} \frac{1}{\sum_{i=1}^{m} \left[w_i(r_{ij} - s_{ik})\right]^2}} \quad (11)$$

Construction Cost Modeling

The construction cost model can be constructed, taking the demand point generation, the number of people, the material demand, the difficulty level of material transportation, and the transportation cost as independent variables, and taking the total cost and cost of constructing the substation as dependent variables. Then, we can simulate the natural genetic process, set specific objectives and constraints to optimize the calculation of each index, and obtain the degree of contribution of each index to reduce the construction cost.

Algorithmic Step

A GA is an evolutionary algorithm that encodes the problem parameters into chromosomes and then uses an iterative method to perform selection, crossover, and mutation operations to exchange information in the population and ultimately produce chromosomes that meet the optimization objective.

The basic steps of the GA are:

1. Encoding. The GA represents the solution space data as genotype string structure data in genetic space and combines these data into different points before performing the search.
2. Generating initial groups. Randomly generate N initial string structure data, each data becomes an individual, and the individuals form a group among themselves. GA takes this group as the starting point of evolution.
3. Robustness assessment. Fitness indicates the strengths and weaknesses of an individual or solution.
4. Selection. The purpose of selection is to pick out the best individuals from the current population and give them a chance to become parents.
5. Crossover. The crossover algorithm is the most important genetic operation in GAs, and through crossover new individuals can be obtained, which inherit the characteristics of the parents.
6. Mutation. Mutation first selects an individual randomly in the population and randomly changes a value in the string structure data with random probability.

Algorithm for Building Substations at Least Cost

The basic idea of GAs is to encode a random selection of individuals into a primitive solution, to cultivate a new generation by simulating the processes of selection, hybridization, and mutation, and to determine the formation of individuals in a new population through an adaptive function. By doing so, the overall characteristics of the population can be gradually improved and approach optimal understanding.

This model optimizes the cost of substation construction through GAs. By simulating the natural genetic process (including selection, hybridization, and mutation) and setting specific objectives and constraint functions, the model achieves efficient planning of substation construction with minimal human and material costs.

Single Parent Genetic Operator

In this paper, we use PGA to solve the HV-DRP problem separately, which is essentially a two-layer uniparental GA.

In this paper, we chose a single-point shift operator with shift probability PS to select the genes between any two positions to be shifted to generate a new chromosome. We randomly selected a set of Positions 3 and 8, moved the corresponding alleles backward by one position, and moved the gene at Position 8 to Position 3 to generate a new set of chromosomes.

Gene shift is shown in Figure 3.

In this paper, we chose a single-point inversion operator with inversion probability Pi to invert the order of gene segments between any two positions to generate new chromosomes. A new set of chromosomes wss generated by randomly selecting a set of Positions 3 and 8 and reversing the order of the corresponding allele segments between them.

Gene reversal is shown in Figure 4.

In this paper, a new chromosome was generated by choosing a single-point translocation operator with a translocation probability of PV and exchanging the genes at that position. A new set of chromosomes was generated by randomly selecting a set of Positions 3 and 8 and exchanging the corresponding alleles between them. This is shown in Figure 5.

Algorithm Termination Condition

In general, the termination condition is divided into the fitness value judgment method and the iteration number judgment method; this paper adopts the latter method as the termination condition, according to the complexity and scale of the problem to set a reasonable number of iterations; when the algorithm runs the number of times to reach the set number of iterations to terminate, no more cycle.

In the establishment of an improved GA to build substations with minimum cost, the objective function is to optimize the cost of building the substation with minimum number of people and money.

Figure 3. Gene shift operation

$$A = \begin{bmatrix} 1 & 3 & \frac{1}{5} & \frac{1}{5} & 5 & \frac{1}{6} & \frac{1}{6} & \frac{1}{7} & \frac{1}{8} & 7 & \frac{1}{2} & 3 \\ \frac{1}{3} & 1 & \frac{1}{7} & \frac{1}{7} & 3 & \frac{1}{7} & \frac{1}{7} & \frac{1}{8} & \frac{1}{9} & 5 & \frac{1}{5} & 2 \\ 5 & 7 & 1 & 1 & 7 & \frac{1}{2} & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & 9 & 4 & \frac{1}{3} \\ 5 & 7 & 1 & 1 & 7 & \frac{1}{2} & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & 9 & 4 & \frac{1}{5} \\ \frac{1}{5} & \frac{1}{3} & \frac{1}{7} & \frac{1}{7} & 1 & \frac{1}{7} & \frac{1}{7} & \frac{1}{8} & \frac{1}{9} & 4 & \frac{1}{8} & 3 \\ 6 & 7 & 2 & 2 & 7 & 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{2} & 8 & \frac{1}{2} & 5 \\ 6 & 7 & 2 & 2 & 7 & 2 & 1 & \frac{1}{3} & \frac{1}{3} & 7 & 3 & \frac{1}{2} \\ 7 & 8 & 3 & 3 & 8 & 3 & 3 & 1 & \frac{1}{2} & 8 & \frac{1}{4} & 4 \\ 8 & 9 & 4 & 4 & 9 & 2 & 3 & 2 & 1 & 9 & \frac{1}{3} & \frac{1}{2} \\ \frac{1}{7} & \frac{1}{5} & \frac{1}{9} & \frac{1}{9} & \frac{1}{4} & \frac{1}{8} & \frac{1}{7} & \frac{1}{8} & \frac{1}{9} & 1 & \frac{1}{2} & \frac{1}{3} \\ 2 & \frac{1}{2} & 4 & 4 & 4 & \frac{1}{6} & 5 & \frac{1}{3} & 3 & \frac{1}{3} & 1 & 5 \\ 3 & 5 & \frac{1}{3} & 2 & \frac{1}{3} & 2 & 3 & \frac{1}{5} & 5 & \frac{1}{2} & 5 & 1 \end{bmatrix}$$

1. Coverage objective function:

In the coverage objective function, U represents a randomly generated demand point, and all demand fields are denoted by US, where T is used to determine the number of people, and all numbers are denoted by TS; after that, any number of people T will be a set of satisfactions, and the time between these demand points and the number of people T will be less than the time between any other number of people $C_U(t)$ represented by Equation 5.1

Figure 4. Gene reversal operation

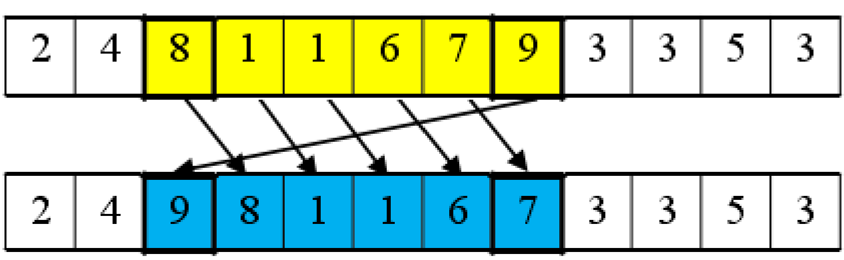
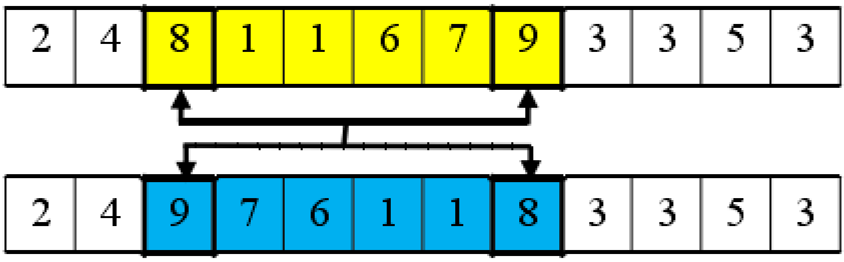


Figure 5. A translocation operation



$$C_U(t) = \left\{U \in US \middle| d_{T,U} \leq r_T\right\} \tag{12}$$

$$US = \underset{t \in sdf}{U} C_U(T), s.t..Min|TS| \tag{13}$$

Here, in Equations 12 and 13, the BPR road resistance function is used to determine the cost; in this case, the smallest cost is the optimal solution we seek. The optimal objective function is expressed in Equation 14.

$$Minf_{CT} = \frac{\sum_{t \in df} C_U(T)}{|US|} \tag{14}$$

2. Binding objectives:

Through the target expenditure, the utility coefficient is introduced on the basis of considering the time cost and the difficulty level of material transportation.

E_T is defined by Equation 15.

$$E_T = \frac{5}{(r_T + 1)^2} \tag{15}$$

Then, q_y^r is the demand for material r at the construction point; q_{ij}^r is the total demand for material r at each construction point; $q_{ij}^r$is the quantity of material r transported from the material dispatch center to the construction point j; L_j is the difficulty level of material transportation J; T_{ij} indicates the optimal route cost of transporting the material vehicle from the material dispatch center i to the construction point J. The objective function and constraints are as shown in Equation 16.

$$Minf_{ST} = \frac{\sum_{t \in df} \left(E_{T,U} / \left(E_{T,U} + \varphi(U)\right)\right)}{|US|} \tag{16}$$

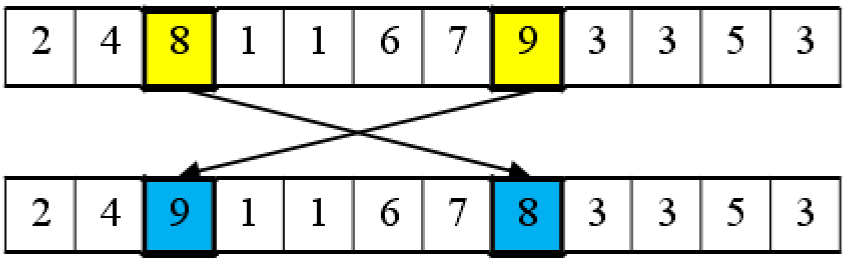
3. Final function objective:

In calculating the GA, the optimization objective is "to minimize the cost and expense of constructing the substation."

The final function has an objective of form, shown in Equation 17.

Objective function:

Figure 6. Load demand and potential site score by proximity to nearest substation



$$\min(TR) = \min \sum_{j \in E} c_{ij} \cdot q_{ij}^{r} \tag{17}$$

Restrictive condition:

$$c_{ij} = \frac{1}{L_j T_{ij}} \tag{18}$$

$$U^{r} \geq Q^{r} \tag{19}$$

$$\sum_{i \in S} q_{ij}^{r} = q_{j}^{r}, \forall j \in E \tag{20}$$

$$\sum_{j \in E} q_{ij}^{r} \leq q_{i}^{r}, \forall i \in S \tag{21}$$

Equation 17 is the objective function, which requires that the substation be constructed at the lowest possible cost and expense; Equation 18 is the utility factor, which is determined on the basis of the cost of constructing the substation; Equation 19 is the maximum amount of money that can be spent on the construction of the substation; Equation 20 requires that the amount of resources to be delivered must satisfy the resource requirements at each point; and Equation 21 requires that the workers work uninterruptedly for the duration of the construction period.

RESULTS

Analysis of Load Distribution Modeling Results

Through the load distribution model, we processed the data on the direct impact of load distribution, the stability and anti-interference ability of the power grid, the scientific and rationality of the site selection, the relative distance to the load center, and the accuracy of the load prediction, and the obtained results are shown in Figure 6.

First, the images show that there is a positive correlation between load demand and potential site scores, that is, an increase in load demand is accompanied by an increase in potential site scores, especially in areas where the load demand is higher than the average value (about 489.62 MW), which is more suitable for constructing a new substation to satisfy the demand for electricity. Second, the range of potential site scores indicates that the suitability of site selection varies significantly from region to region, emphasizing the importance of considering the load distribution in the site selection process to achieve efficient resource allocation and improve the economic efficiency and stability of the power grid. Third, the image indicates that the locations that are farther away from the existing substations receive higher scores, implying that the establishment of new substations in these areas is essential for optimizing the grid layout and improving the reliability of power supply. The analysis of multiple factors such as load demand, geographic location, and distance from existing substations not only helps to identify potential sites in areas of rapid growth in electricity demand but also promotes the sustainable development of the grid by improving the efficiency of power transmission and reducing losses. The load distribution model emphasizes the central role of load distribution in substation siting decisions, pointing out that, through scientific data analysis and rational planning, it can effectively improve the performance and economic efficiency of the power grid to ensure that it can meet the future growth of power demand and promote the sustainable development of the power system.

Analysis of Land Use Model Results

The land use optimization model considers factors such as land use area, land use rate, land use efficiency, and environmental protection index. Through the optimization algorithm, the optimal land use scheme can be determined, and the land use efficiency can be maximized.

The output of the land use optimization model is presented through four different hierarchical clustering methods: average linkage, complete linkage, single linkage and ward linkage. These methods are used to determine the optimal land use scheme by considering factors such as land use area, land utilization rate, land use efficiency, and environmental protection index, to maximize the land use efficiency.

The average linkage clustering method uses the average distance of all pairs of points between clusters as a measure of the distance between clusters. The method tends to produce clusters of relatively balanced sizes, which is conducive to ensuring balanced development of land use, while considering the needs of land use efficiency and environmental protection. The results obtained through average link clustering can help decision makers to identify the balance of land use distribution and thus optimize the allocation of land resources. Average linkage is shown in Figure 7.

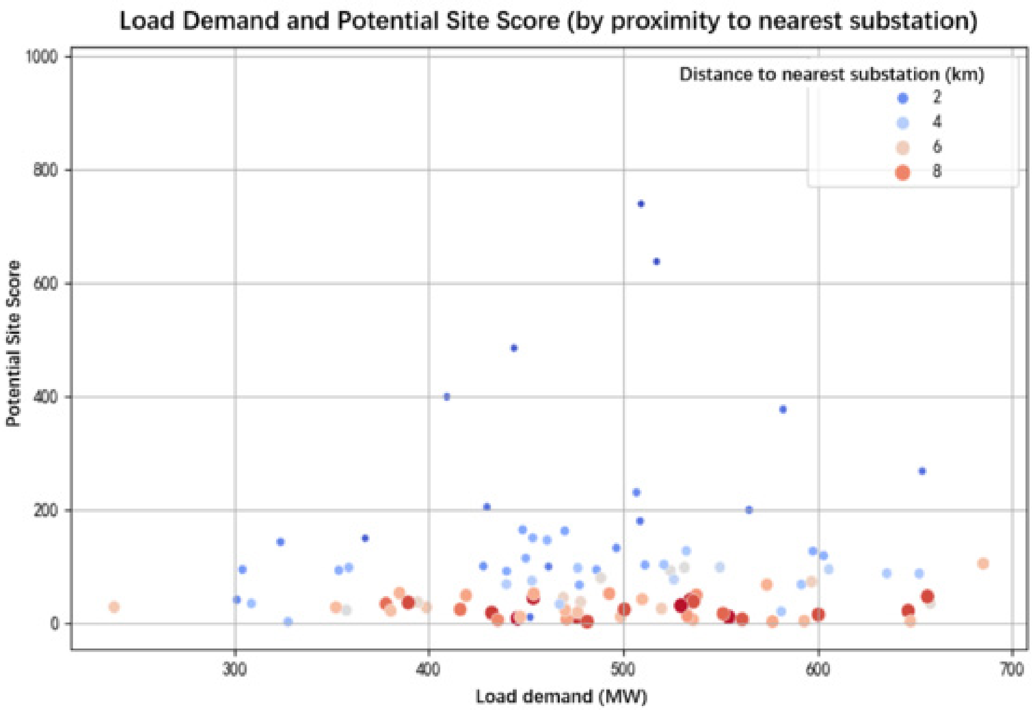
The full linkage clustering method considers the distance of the farthest pair of points between clusters as the distance between clusters. This method tends to form compact and clearly separated clusters, which is conducive to the identification of high-intensity and low-intensity areas in land use, and thus promotes the rational planning of land use and environmental protection. Complete linkage is shown in Figure 8.

Single linkage clustering methods use the distance between the nearest pairs of points between clusters as the distance between clusters. This method may lead to a chain effect in which clusters gradually expand through their nearest neighbors. In land-use optimization, the single-linkage approach can be used to identify land-use continuity and agglomeration, providing valuable insights for land-use planning, especially in protecting ecological connectivity and environmentally sensitive areas. Single linkage is shown in Figure 9.

The ward linkage clustering method aims to minimize the increase in within-cluster variation. This method focuses specifically on minimizing the total within-cluster variation after each merger, resulting in clusters that are relatively uniform in size and shape. In land use optimization, the ward linkage method helps to balance the distribution of land use, ensuring high efficiency and environmental protection while achieving the optimal allocation of land resources. Ward linkage is shown in Figure 10.

These four clustering methods reflect different aspects and objectives of land use optimization. By comparing the outputs of these methods, a more comprehensive understanding of the current state of land use and potential optimization schemes can be achieved, providing a scientific basis and decision support for substation siting and land use planning. This not only helps to maximize

Figure 7. Output of the land use optimization model by average linkage clustering method



the efficiency of land use but also promotes environmental protection and sustainable development while ensuring the stability and reliability of power supply.

Analysis of Neighborhood Satisfaction Model Results

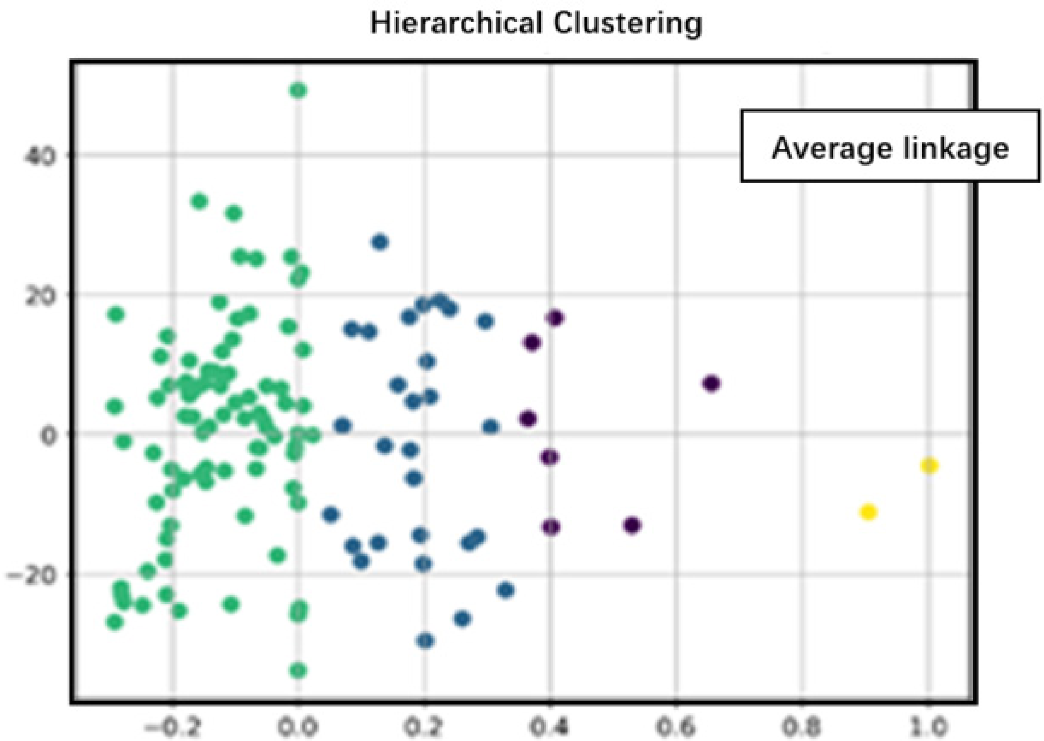
In this part, the hierarchical analysis method was used to establish the satisfaction index system of the residents near the selected site, the indexes of environmental impact, noise level, air quality, and transportation accessibility were taken as the levels of the hierarchical structure, and the weights of the indexes were determined through the feedback from the residents' survey or the experts' evaluation. Finally, a composite index of residents' satisfaction was calculated by weighted average. The final results are shown in Table 2.

Analysis of Construction Cost Model Results

In this analysis, we explored a substation construction cost model that takes into account the effects of variables such as demand point generation, number of people, material requirements, material transportation difficulty level, and transportation costs on the total cost and cost of the substation. By applying a GA, we optimized these variables to find the lowest cost substation construction solution.

Table 3 is part of the results of the construction cost model to show the cost optimization results for different substation construction scenarios. Each row represents a possible scenario configuration, and each column corresponds to a specific construction metric, such as location choice, expected number of workers, material requirements, or transportation costs. The term "configuration" refers to a modeling process that optimizes the cost of a substation based on the above metrics and refers to the different combinations of these metrics used in the modeling process to evaluate the total cost of substation construction. Each row represents a particular scenario or combination of solutions, where 1 means that the indicator in the corresponding column was selected in the scenario, and 0

Figure 8. Output of the land use optimization model by complete linkage clustering method



means that it was not selected. This representation allows the analyst to clearly see which factors were considered and how they interacted in the search for a cost-optimized solution. By comparing these different configurations, it is possible to identify the lowest-cost construction option that meets all the necessary conditions and constraints.

We define the variables:

Columns 1-14 represent different construction indicators, including:

Location choice: based on geographic location, proximity to the point of need, or land cost.

Expected number of workers: the total number of workers expected to be required based on project size and duration.

Material requirements (civil materials): for example, cement, steel, sand, and gravel requirements.

Material requirements (electrical materials): for example, cables, transformers, or insulation materials.

Transportation cost: the cost of transportation of materials from the place of supply to the construction site.

Construction equipment costs: rental or purchase costs of construction equipment such as excavators or cranes.

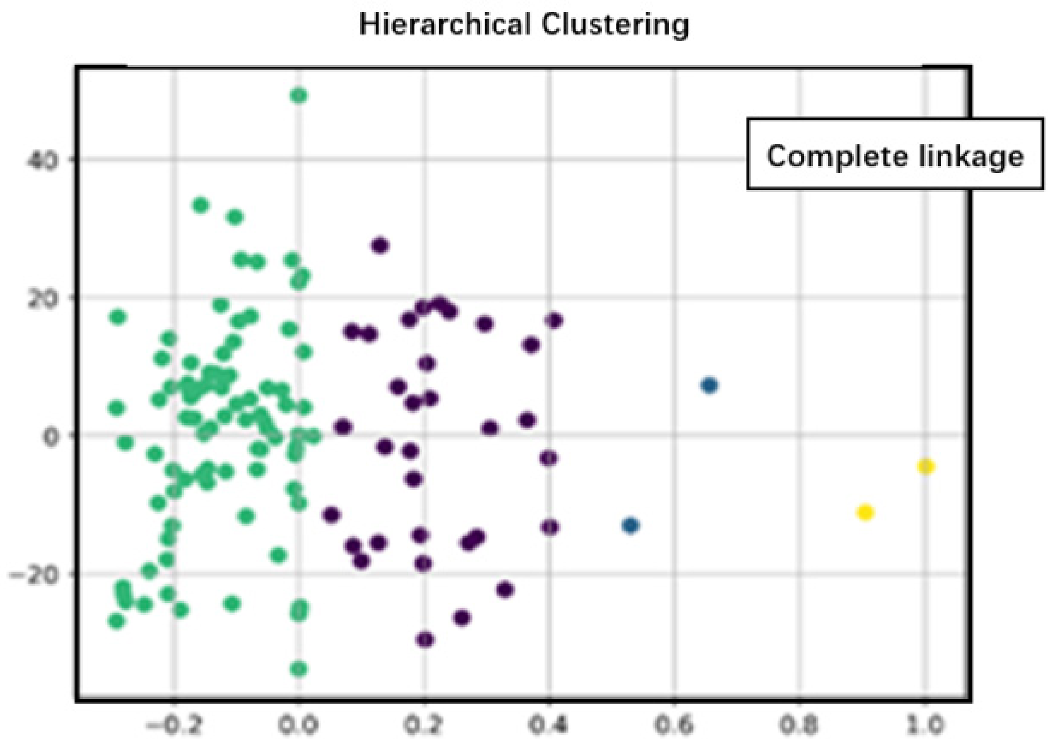
Construction difficulty level: evaluation of construction difficulty based on factors such as terrain and weather conditions.

Environmental impact assessment: costs of the project's potential impact on the surrounding environment and the protective measures that need to be taken.

Safety management costs: costs of measures to ensure safety management during construction.

Quality control costs: costs of ensuring construction quality, including material testing or process supervision.

Figure 9. Output of the land use optimization model by single linkage clustering method



Duration: the length of time required to complete the project.

Legal and license costs: costs of obtaining necessary construction licenses and complying with local laws and regulations.

Provisioning costs: funds set aside for unforeseen events (e.g., natural disasters, material price fluctuations, etc.).

Project management costs: salaries, office expenses, and other administrative costs of the project management team.

Rows 1-18 represent different scenario configurations.

These data were derived through a GA optimization process aimed at minimizing construction costs while satisfying all constraints.

This representation allows us to visualize which factors are considered in each scenario and how they interact to find the cost-optimized solution. This helps to compare the total costs of different configurations, thus identifying the least costly construction solution that satisfies all necessary conditions and constraints. This approach is very useful because it allows a large number of possible combinations of solutions to be systematically analyzed and compared, ensuring that the selected solution is both economical and practical.

DISCUSSION

In this study, we conducted a multidimensional analysis of substation siting by integrating load distribution, land use, neighborhood satisfaction, and construction cost models. The findings highlight several key insights and areas for further exploration.

Table 2. Indicators and weights for assessing residents near substations in S region of Western China

Indicators of Level 1	Indicators of Level 2	Weighting
External influences	Environmental Impact	0.3124
	Noise Level	0.0789
	Accessibility	0.1954
Internal influences	Electromagnetic Field Exposure	0.0485
	Safety Concerns	0.2684
	Maintenance and O&M activities	0.0964

The load distribution model reveals a significant impact on substation siting, with a positive correlation between load demand and potential site scores. This finding aligns with existing power system planning theories, emphasizing the central role of load demand in siting decisions. However, the accuracy of this model is contingent on high-quality load forecasting data, which may vary due to regional differences and temporal changes.

The land use model demonstrates varying emphases on land use efficiency and environmental protection depending on the clustering method employed. This provides decision makers with diverse planning options. However, the dynamic nature of land use and environmental factors, such as urban expansion and climate change, necessitates continuous monitoring and model adjustments.

Figure 10. Output of the land use optimization model by ward linkage clustering method

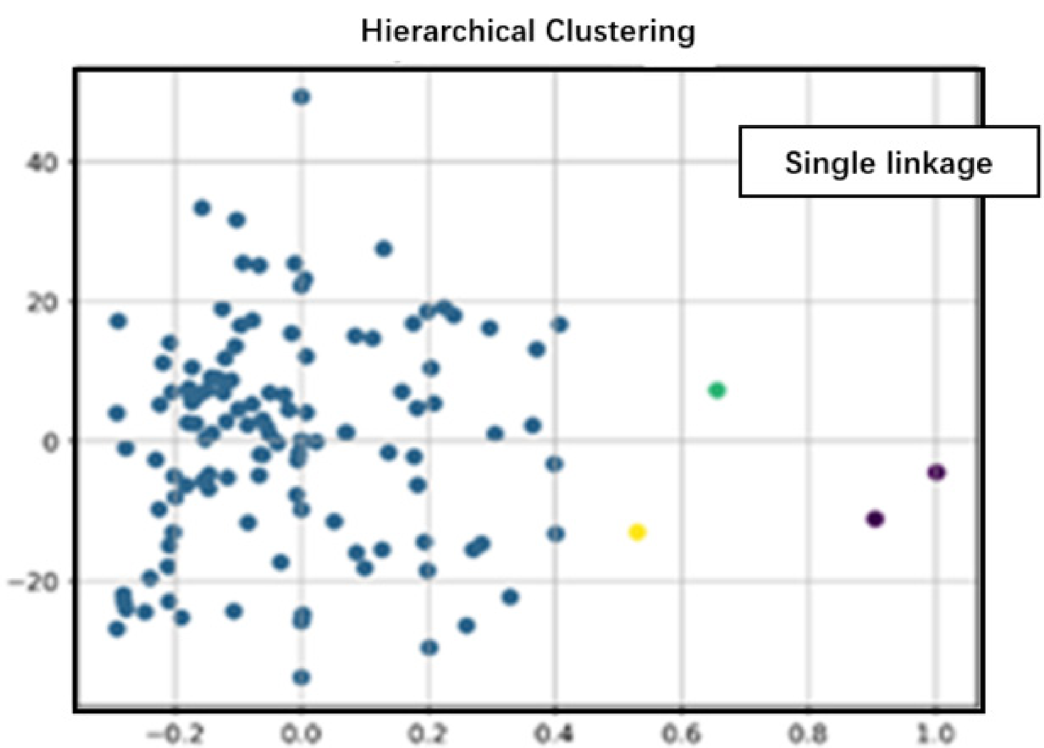


Table 3. Construction indicators and scenario configurations for cost optimization

	1	2	3	4	5	6	7	8	9	10	11
1	0	0	0	0	0	0	0	0	1	0	0
2	0	0	0	0	1	0	0	0	0	0	0
3	0	0	1	0	0	0	0	0	0	0	0
4	0	1	0	0	0	0	0	0	0	0	0
5	0	0	0	0	1	0	0	0	0	0	0
6	1	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	1
8	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	1	0	0	0	0
10	0	0	0	0	0	0	0	1	0	0	0
11	0	0	0	0	0	0	0	0	1	0	0
12	0	0	0	0	0	1	0	0	0	0	0
13	0	0	0	1	0	0	0	0	0	0	0
14	0	0	0	0	0	1	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	1	0
16	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	1	0	0

The neighborhood satisfaction model underscores the importance of social factors in infrastructure projects. The composite satisfaction index derived from hierarchical analysis offers a quantitative tool for assessing and improving the substation's impact on the surrounding community. Nevertheless, biases in survey responses and variations in community expectations can affect the reliability of the satisfaction scores.

The construction cost model, optimized using a GA, efficiently minimizes the economic costs of substation construction. While the GA shows strong optimization capabilities, its parameter settings and convergence need further verification and refinement to ensure robustness across different scenarios.

However, there are some limitations in this paper. First, the accuracy of the model is limited by data quality and availability, especially in load forecasting and resident satisfaction assessment. Second, the dynamic changes of environmental and social factors, such as the impact of climate change on electricity demand and the long-term impact of socio-economic development on land use and residents' satisfaction, are not sufficiently considered in this paper. Third, although the GA shows good optimization ability in the cost model, its parameter setting and convergence still need to be further verified and improved. This study acknowledges the limitations of not fully integrating all relevant aspects into a unified decision-making model. Future research should focus on developing a comprehensive framework that concurrently considers load forecasts, land usage patterns, neighborhood impacts, and economic costs. By adopting this integrated approach, substation siting decisions can be optimized across all critical dimensions, thereby improving the overall efficacy and sustainability of the siting process. To achieve this, we plan to implement multicriteria decision analysis (MCDA) techniques, which would allow for a systematic evaluation and balancing of these factors. MCDA provides a robust methodological foundation to incorporate diverse criteria, enabling us to weigh their relative importance and interactions comprehensively. Through the application of MCDA, we aim to enhance the robustness and reliability of our substation siting framework, ensuring that future implementations are more responsive to the complex and multifaceted nature of real-world scenarios. This methodological enhancement would contribute significantly to the field, offering a more holistic and practical tool for decision makers.

To further improve the decision-making process, we recognize the need to incorporate a sensitivity analysis into the framework. This addition would allow us to assess how changes in input parameters, such as load forecasts and cost estimates, influence siting decisions. By understanding the sensitivity of the framework to these variables, we can ensure that it remains robust and adaptable across various scenarios.

Additionally, we acknowledge the potential to modify the framework to include multiobjective optimization strategies. By integrating these strategies, the framework can address multiple, often conflicting objectives simultaneously, such as minimizing costs while maximizing environmental sustainability and community acceptance. This enhancement would provide a more comprehensive decision-support tool, aligning with the complex and multifaceted nature of substation siting challenges.

CONCLUSIONS

In this paper, we achieved innovative results in the field of substation siting and constructed a multidimensional comprehensive assessment framework by integrating four key factors, namely, load distribution, land use, residents' satisfaction, and construction cost. The framework not only strengthens the understanding of the multidimensional influencing factors of substation siting but also proposes a scientific and refined siting methodology in order to achieve the best balance between economic benefits and social and environmental responsibilities.

Through this paper, it is clarified that in the context of modern power grid development, substation siting needs to be responsive to the growth trend of load demand while considering the rational use and protection of land resources, ensuring broad community acceptance and support, and maximizing cost-effectiveness. In particular, the results of the study reveal the interactions between various factors and their combined effects on siting decisions, providing new perspectives for more accurate and efficient power system planning.

The study also identifies several key practical guidelines including, but not limited to, the use of advanced data analytics to optimize load forecasting, the adoption of diverse land planning strategies for ecological preservation and optimal utilization, and the strengthening of communication and cooperation with the community to ensure the social acceptability of substation projects.

At the same time, the study recognizes the challenges of data quality and dynamics during the implementation process and calls for follow-up work to innovate in data collection and model adaptation to enhance the accuracy and usefulness of the model. In addition, future research should explore new techniques and methods to cope with the environmental changes and the evolution of social needs, so as to further enhance the scientific and practical effectiveness of substation siting.

In conclusion, this paper provides a comprehensive and systematic analytical framework and decision support tool for substation siting, which is expected to contribute to the sustainable development of power systems. Through continuous optimization and improvement, it is expected to better cope with the complex challenges of power system planning and construction in the future and to promote the harmonious development of socioeconomics and the environment.

CONFLICTS OF INTEREST

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

FUNDING STATEMENT

No funding was received for this work.

PROCESS DATES

08, 2024

This manuscript was initially received for consideration for the journal on 06/14/2024, revisions were received for the manuscript following the double-anonymized peer review on 08/15/2024, the manuscript was formally accepted on 08/20/2024, and the manuscript was finalized for publication on 08/29/2024

CORRESPONDING AUTHOR

Correspondence should be addressed to Jinming Yang; 202115030224@sust.edu.cn

REFERENCES

Atrigna, M., Buonanno, A., Carli, R., Cavone, G., Scarabaggio, P., Valenti, M., Graditi, G., & Dotoli, M. (2023). A machine learning approach to fault prediction of power distribution grids under heatwaves. *IEEE Transactions on Industry Applications*, *59*(4), 4835–4845. DOI: 10.1109/TIA.2023.3262230

Azhar, N. A., Mohamed Radzi, N. A., Mohd Azmi, K. H., Samidi, F. S., & Muhammad Zainal, A. (2022). Criteria selection using machine learning (ML) for communication technology solution of electrical distribution substations. *Applied Sciences (Basel, Switzerland)*, *12*(8), 3878. DOI: 10.3390/app12083878

de Almeida, F. A., Romão, E. L., & Gomes, G. F. (2022). Combining machine learning techniques with kappa–Kendall indexes for robust hard-cluster assessment in substation pattern recognition. *Electric Power Systems Research*, *206*, 107778. DOI: 10.1016/j.epsr.2022.107778

Deka, D., & Chertkov, M. (2024). Topology detection in distribution networks with machine learning. In *Big data application in power systems* (pp. 89-108). DOI: 10.1016/B978-0-443-21524-7.00017-7

Gangwar, A. K., & Gafoor Shaik, A. (2022). Fault detection, location in hybrid distribution network using machine learning algorithm. SSRN *Electronic Journal*. DOI: 10.2139/ssrn.4170512

Gao, Z. N. (2022). Auxiliary decision-making for substation distribution transformer site selection in urban distribution networks. *Journal of Shanghai University of Electric Power*, *25*(4), 2095–0020.

Garcia, D. R., Poudel, B., Bidram, A., & Reno, M. J. (2022). *Substation-level circuit topology estimation using machine learning. In 2022 IEEE power & energy society innovative smart grid technologies conference*. ISGT.

Hou, G. Z. (2021). Analysis of the design characteristics of full-indoor GIS substations in inner Mongolia power grid. *Power Grid Design*, *11*, 18–23. DOI: 10.13500/j.dlkcsj.issn1671-9913.2021.11.004

Joga, S. R. K., Sinha, P., & Maharana, M. K. (2023). A novel graph search and machine learning method to detect and locate high impedance fault zone in distribution system. *Engineering Reports*, *5*(1), e12556. DOI: 10.1002/eng2.12556

Kang, J. Y., & Kim, Y. S. Application and comparison of a machine learning-based fault location estimation in a power distribution network under restricted conditions. *Available at SSRN* 4102834. DOI: 10.2139/ssrn.4102834

Li, Y. H. (2022). Prediction and analysis of noise impact of 500kV outdoor substation. *China New Technology and New Products*, *117-119*. Advance online publication. DOI: 10.13612/j.cnki.cntp.2022.19.038

Livani, H. (2024). Supervised learning-based fault location in power grids. In *Big data application in power systems*. Elsevier., DOI: 10.1016/B978-0-443-21524-7.00018-9

Luo, X., Liu, X. G., & Guo, W. B. (2021). Site selection research of a 750 kV substation based on analytic hierarchy process. *Power Grid Design*, *3*, 38–44. DOI: 10.13500/j.dlkcsj.issn1671-9913.2021.03.008

Meng, S. C., Zhang, S. J., Ding, H. X., Zhu, S. C., Hu, Y., Duan, J. B., & Liu, H. (2022). Research on 5G base station deployment in substations considering radiation and conducted interference. *Power Information and Communication Technology*, *20*(7), 10–17. DOI: 10.16543/j.2095-641x.electric.power.ict.2022.07.002

Mirshekali, H., Keshavarz, A., Dashti, R., Hafezi, S., & Shaker, H. R. (2023). Deep learning-based fault location framework in power distribution grids employing convolutional neural network based on capsule network. *Electric Power Systems Research*, *223*, 109529. Advance online publication. DOI: 10.1016/j.epsr.2023.109529

Nhung-Nguyen, H., Girdhar, M., Kim, Y. H., & Hong, J. (2024). Machine-learning-based anomaly detection for GOOSE in digital substations. *Energies*, *17*(15), 3745. DOI: 10.3390/en17153745

Ponukumati, B. K., Sinha, P., Maharana, M. K., Kumar, A. P., & Karthik, A. (2022). An intelligent fault detection and classification scheme for distribution lines using machine learning. *Engineering, Technology &. Applied Scientific Research*, *12*(4), 8972–8977.

Poudel, B. P., Bidram, A., Reno, M. J., & Summers, A. (2022). Zonal machine learning-based protection for distribution systems. *IEEE Access : Practical Innovations, Open Solutions*, *10*, 66634–66645. DOI: 10.1109/ACCESS.2022.3184865

Rizeakos, V., Bachoumis, A., Andriopoulos, N., Birbas, M., & Birbas, A. (2023). Deep learning-based application for fault location identification and type classification in active distribution grids. *Applied Energy*, *338*, 120932. DOI: 10.1016/j.apenergy.2023.120932

Sun, K., Han, H. Y., Zhang, Q. M., Ye, G. F., & Zhang, X. D. (2020). A substation planning method for urban dense areas based on grid division: Taking the substation planning in the main urban area of Hangzhou as an example. *Modern Urban Research*, *5*, 60020020506307. DOI: 10.1109/ACCESS.2021.3093563

Tavoosi, J., Shirkhani, M., Azizi, A., Din, S. U., Mohammadzadeh, A., & Mobayen, S. (2022). A hybrid approach for fault location in power distributed networks: Impedance-based and machine learning technique. *Electric Power Systems Research*, *210*, 108073. DOI: 10.1016/j.epsr.2022.108073

Ţboacă-Ciupăgeanu, M. E., & Ţboacă-Ciupăgeanu, D. A. (2024). Optimal substation placement: A paradigm for advancing electrical grid sustainability. *Sustainability (Basel)*, *16*(10), 4221. DOI: 10.3390/su16104221

Townsend, W. E., & Reid, A. J. (2022). A 3D fault location program for use in high-voltage substations using multilateration, geolocation and machine learning. In 2*022 17th international conference on probabilistic methods applied to power systems (PMAPS)* (pp. 1-4).

Wu, L. J. (2019). Application of GIS in substation site selection planning. *Housing and Real Estate*, (8), 1006–6012.

Xiang, X. M., Dong, X. C., He, J. Q., Zheng, Y. K., & Li, X. Y. (2023). A study on the fault location of secondary equipment in smart substation based on the graph attention network. *Sensors (Basel)*, *23*(23), 9384. DOI: 10.3390/s23239384 PMID: 38067757

Xu, L. K., Ding, L. R., & Liu, S. M. (2022). Application of GIS and variable weight model in substation site selection research. *Energy & Environment*, *44*(10), 110–114. DOI: 10.19389/j.cnki.1003-0506.2022.10.018

Yao, Y., Feng, C., Xie, J., Yan, X., Guan, Q., Han, J., Zhang, J., Ren, S., Liang, Y., & Luo, P. (2023). A site selection framework for urban power substation at micro-scale using spatial optimization strategy and geospatial big data. *Transactions in GIS*, *27*(6), 1662–1679. DOI: 10.1111/tgis.13093

Yazdanpanah, Z., Rastegar, M., & Jooshaki, M. (2024). Determining target levels of power distribution system reliability indices using machine learning. *Electric Power Systems Research*, *233*, 110456. DOI: 10.1016/j.epsr.2024.110456

Zeng, Z. H. (2022). *Substation site selection method considering the whole life cycle*. Shanxi Electric Power., DOI: 10.1109/ACCESS.2021.3093563

Jing Xiong, Lecturer, Doctor,Graduated from Xidian University in 2016. Worked in Shaanxi University of Science & Technology. Her research interests include Computer Application & Visual Computing

Jinming Yang, Bachelor, Undergrad,Study in Shaanxi University of Science & Technology University in 2021. His research interests include Computer Application.

Yueling Deng, Bachelor, Undergrad,Study in Shaanxi University of Science & Technology University in 2023. Her research interests include Machine learning.

Zixuan Chen,Bachelor, Undergrad,Study in Shaanxi University of Science & Technology University in 2022. Her research interests include Machine learning.

Jihaoyu Yang, Bachelor, Undergrad,Study in Shaanxi University of Science & Technology University in 2023. Her research interests include Machine learning.

Chen Xu,Bachelor, Undergrad,Study in Shaanxi University of Science & Technology University in 2022. Her research interests include Machine learning.

Yong Qi,Associate professor, Doctor,Graduated from Northwestern Polytechnical University in 2011. Worked in Shaanxi University of Science & Technology. His research interests include Computer Application.