


## RESEARCH ARTICLE

# A site selection framework for urban power substation at micro-scale using spatial optimization strategy and geospatial big data

Yao Yao<sup>1,2</sup>  | Chenqi Feng<sup>1</sup> | Jiteng Xie<sup>1</sup> | Xiaoqin Yan<sup>1</sup> |  
Qingfeng Guan<sup>1</sup> | Jian Han<sup>3</sup> | Jiaqi Zhang<sup>1</sup> | Shuliang Ren<sup>5</sup> |  
Yuyun Liang<sup>1</sup> | Peng Luo<sup>4,6</sup>

<sup>1</sup>School of Geography and Information Engineering, China University of Geosciences, Wuhan, China

<sup>2</sup>Center for Spatial Information Science, The University of Tokyo, Chiba, Japan

<sup>3</sup>State Grid Pingxiang Power Supply Company, Pingxiang, China

<sup>4</sup>Chair of Cartography and Visual Analytics, Technical University of Munich, Munich, Germany

<sup>5</sup>Institute of Remote Sensing and Geographical Information Systems, Peking University, Beijing, China

<sup>6</sup>School of Geography and the Environment, University of Oxford, Oxford, UK

## Correspondence

Peng Luo, Chair of Cartography and Visual Analytics, Technical University of Munich, Munich, Germany.  
Email: [peng.luo@tum.de](mailto:peng.luo@tum.de)

## Funding information

Alibaba Innovative Research Project; National Key Research and Development Program of China; National Natural Science Foundation of China; State Key Laboratory of Resources and Environmental Information System; the Scientific Research Program of the Department of Natural Resources of Hubei Province

## Abstract

The world is facing more energy crises due to extreme weather and the rapidly growing demand for electricity. Siting new substations and optimizing the location of existing ones are necessary to address the energy crisis. The current site selection lacks consideration of spatial and temporal heterogeneity in urban power demand, which results in unreasonable energy transfer and waste, leading to power outages in some areas. Aiming to maximize the grid coverage and transformer utilization, we propose a multi-scene micro-scale urban substation siting framework (UrbanPS): (1) The framework uses multi-source big data and the machine learning model to estimate fine-scale power consumption for different scenarios; (2) the region growing algorithm is used to divide the power supply area of substations; and the (3) location set coverage problem and genetic algorithm are introduced to optimize the substation location. The UrbanPS was used to perform siting optimization of 110kV terminal substations in Pingxiang City, Jiangxi Province. Results show that the coverage and utilization rate of the optimization results under different power consumption scenarios are close to 99%. We also found that the power can be saved by dynamic regulation of substation operation.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Transactions in GIS* published by John Wiley & Sons Ltd.

## 1 | INTRODUCTION

The increasing economic growth and urban expansion have led to rapid growth in power consumption, especially in developing countries (Liu et al., 2017; Taghizadeh-Hesary et al., 2021). However, the unreasonable spatial distribution of substations and their limited number lead to power shortages and low utilization of current power grid equipment, making the mismatch between power energy supply and demand increasingly prominent (Fan et al., 2021). Most peak loads occur in cold winter and hot summer, especially in extreme climates. During this period, electricity loads are constantly at high levels, power equipment operates under high stress, and substations may be overloaded (Sun et al., 2018). This can lead to power shortages in urban areas, thus posing a considerable challenge to the power supply (Uddin et al., 2018). Therefore, it is urgent to relieve the pressure on grid operation and optimize the power supply structure during peak periods (Zeng et al., 2019). Future grid planning should expand the electricity supply network considering the spatial and temporal variation of the grid electricity load in different seasons based on the urban electricity demand (Chen & Zhang, 2021). The substation is an important part of the power system (Xia et al., 2018). Its reasonable location selection is of great significance for the safety and stability of electricity resources and regional power distribution (Ximing et al., 2019; Zeng et al., 2019).

However, most of the current power optimization is aimed at minimizing the total cost of siting and construction, which is difficult to cope with the power shortage problem during peak periods (Li, Yang, et al., 2018; Qin et al., 2019). Therefore, we aimed to maximize grid coverage and transformer utilization and proposed a multi-scenario micro-scale urban substation site selection optimization framework based on a genetic algorithm. Firstly, the fine-scale power consumption estimation models for different scenarios were constructed using the random forest method. Then, the region growing algorithm was used to assign the power supply area of each substation. Finally, the coverage model and genetic algorithm were used to solve the optimal solution of the substation site. The results of multi-scenario optimization were compared with the spatial distribution of existing substations to evaluate the effectiveness of the UrbanPS.

## 2 | RELATED WORKS

### 2.1 | Estimation of urban residential power demand

Knowing the detailed spatial distribution of power demand at the microscopic spatial (e.g., at every single house or 100m grid level) and temporal (i.e., at tight time intervals) scales is essential, especially for power system planning and the construction of substations (Yi et al., 2016). Some studies used spatial load forecasting methods to reveal the characteristics of electricity demand (Heymann et al., 2018; Ye et al., 2019). But these methods have low simulation accuracy and mainly focus on estimating future load growth. The development of remote sensing technology has brought new opportunities in this field (Yin et al., 2021). Some scholars have used nighttime light data to estimate power consumption at the grid level. For example, He et al. (2012) established a power consumption model using nighttime light data to map mainland China's spatial pattern of county-level power consumption. However, the method is vulnerable to data quality. Thus, the estimated result's accuracy and scale cannot meet the actual requirements.

In conclusion, the current research on predicting power consumption has low prediction accuracy and is only conducted on coarse spatial scales (1 km or 500m). Actual substation sitting in cities requires data support at finer plot scales (Zheng et al., 2014). In recent years, location-based service (LBS) technology has provided reliable data support for urban planning, energy, and economics (Hu et al., 2022; Yang et al., 2022). LBS data include points of interest (POI), mobile phone data, public transportation trajectory data, etc. They can reveal cities' spatial

structure and population activity characteristics at the microscopic scale (Ding et al., 2018), thus providing an opportunity for fine-scale urban power load distribution mapping (Cheng et al., 2022).

## 2.2 | Site selection of substation and other public facilities

Early research used traditional mathematical models to solve the optimal location of substations (Anjos & Vieira, 2017). However, as the power systems continue to expand and the grid structure becomes increasingly complex, traditional methods can no longer meet the requirements of optimal substation planning. Subsequently, some studies have used heuristic optimization models for urban substation siting. For example, Li, Yin, et al. (2018) used a combination of biogeographic optimization and population competition algorithms for substation site selection. They demonstrated that the method can be effective for complex substation site selection but is more applicable to a small number of substation sites. Kaewmamuang et al. (2019) used a dynamic bacterial foraging algorithm (DBFA) to search for the best binary variables and the best integer variables to determine the optimal substation location, capacity, power supply range, and construction time. The optimization objective of these two methods is cost minimization.

Recently, spatial analysis technology and geographic information systems (GIS) have become popular in site selection for many facilities, including hospitals, schools, and emergency shelters (Wang et al., 2020; Yu et al., 2020). Among them, the maximum coverage location problem (MCLP)-based models and location set coverage problem (LSCP)-based model have been widely used. For example, Yu et al. (2020) used the MCLP and the LSCP for macro and micro fire stations, respectively. Wang et al. (2020) used the MCLP-based model to maximize the network service coverage for the location optimization of base stations in 5G cellular networks. However, there is a lack of using the MCLP-based and LSCP-based models for the site selection of substations. Since electronic facilities are also essential public service facilities, this coverage model-based approach provides new opportunities to address the site selection of electric facilities.

To sum up, previous substation siting studies are rarely supported by the accurate fine-scale spatial distribution of residential electricity demand. Most of them have focused on the optimization objective of minimizing total cost. There is a lack of research that combines spatial analysis technology with the site selection of substations from the perspective of large-scale urban grid coverage. In response to the power supply shortage problems during peak periods in rapidly developing cities in China, the current site selection does not satisfy urban residents' primary production and daily power use, leading to low grid operation efficiency and even regional black-outs (Li, Yang, et al., 2018; Qin et al., 2019).

## 3 | METHODOLOGY

The workflow of the proposed UrbanPS is shown in Figure 1, which includes three steps: (1) Estimate the urban power demand under different scenarios in 100m spatial resolution. Based on multi-source spatial data and power data, the relationship between power load and spatial variable factors in four scenarios is constructed using a random forest model fitted to get the fine-scale spatial distribution of power load. Grids with power load are extracted and transformed to urban residential power demand points; (2) Assign the power supply area of each substation. Select the power supply index of the terminal substation, take the substation as the initial seed point, use the regional growing algorithm for neighborhood power diffusion, and gradually assign the surrounding power demand points to get the actual power supply area of each substation; (3) Substation site selection optimization. Based on the urban residential power demand and substation supply area of different scenarios obtained in the above two steps, the selection of the optimal substation location is abstracted as a LSCP according to the actual urban grid coverage characteristics, coupled with genetic algorithm to solve the optimal location and construct a substation location optimization model.

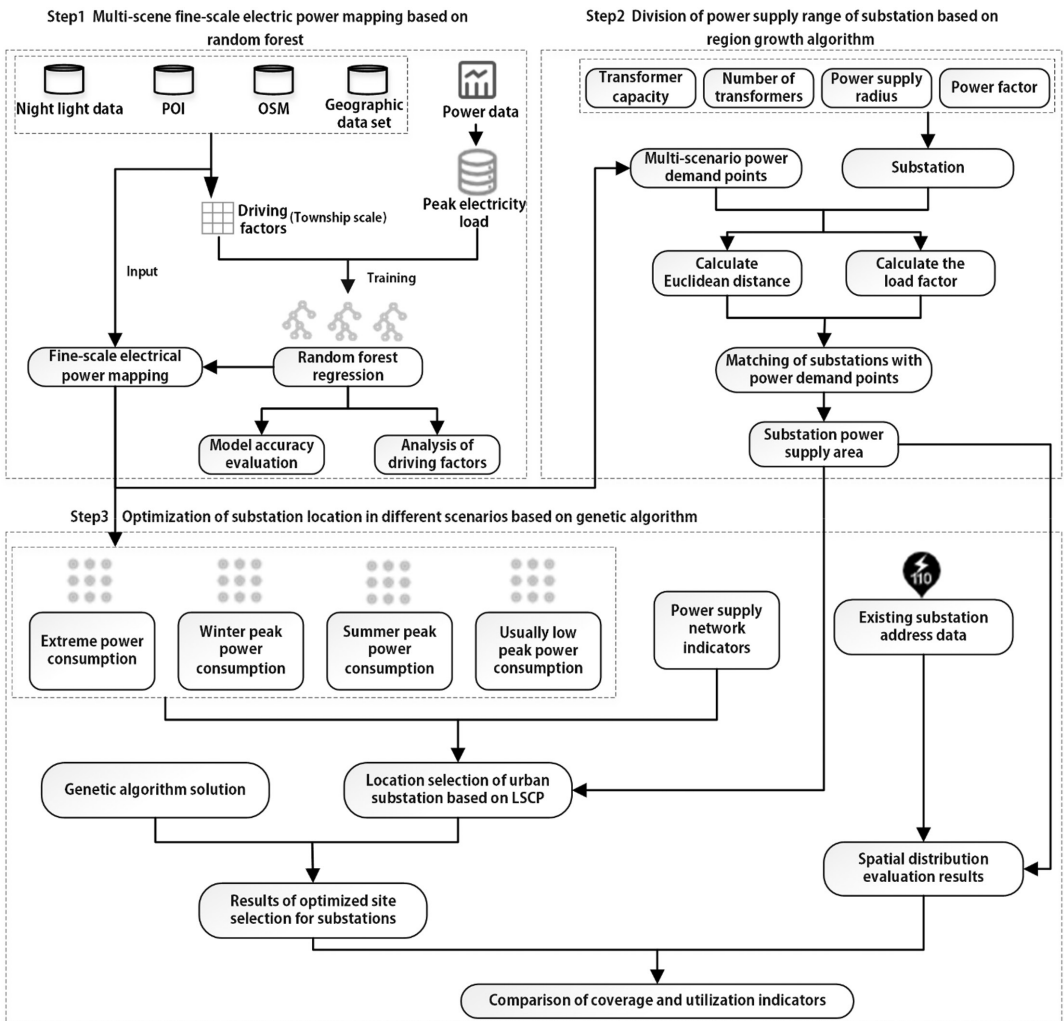


FIGURE 1 UrbanPS framework: A micro-scale urban substation site selection optimization framework based on geospatial big data and spatial optimization strategies.

### 3.1 | Multi-scenario fine-scale power mapping

Accurate power demand mapping is a key step for substation site selection optimization. Four scenarios of urban power consumption were considered in the UrbanPS, considering the urban production and living power load vary with the seasonal climate change (Li, Yin, et al., 2018; Senthil Kumar et al., 2018). They are extreme power consumption, winter peak power consumption, summer peak power consumption, and usually low power consumption. The extreme power consumption scenario is the sum of the maximum power consumption of all distribution transformers in each township of the city at different times. The winter peak power consumption scenario is the maximum power value obtained during January–February. The summer peak power consumption scenario is the maximum power value obtained during June–August. And the usual low peak power consumption scenario is the maximum power values obtained during March and October each year. Considering the efficiency and applicability of the urban substation siting and referring to previous studies (Zheng et al., 2014), a 100 m land parcel was chosen as the basic analysis unit for the residential electricity demand estimation.

In the UrbanPS framework, the random forest model was selected to characterize the complex relationship between the residential power load and explanatory variables. The random forest model was used for each power consumption scenario to characterize the nonlinear relationships and interactions between explanatory variables and the power load at the township level. Through the random forest model, we modeled the urban electricity load and multi-source spatial data-driven factors in four different scenarios, and constructed power fitting models for various electricity consumption scenarios. The  $R^2$ , Pearson's  $r$ , and root mean square error (RMSE) was used to evaluate the fitting accuracy of the model (Yao et al., 2019). After fitting models, the spatial distribution of power load at 100m resolution in four scenarios was mapped using the explanatory variables. After mapping the power at the 100m grid level, grids with non-zero power demands were selected. They were transformed into points and were regarded as the power demand points.

### 3.2 | Substation power supply area division based on region growing algorithm

Estimating the supply area of each substation is a crucial step in optimizing the substation location. The electrical energy is transmitted from the substations to the demand points stepwise from near to far (Shu et al., 2020), and different substations have their specific supply area. In this study, we assume that the transmission process can be described by region growing. Therefore, the region growing algorithm was used in the UrbanPS framework to assign each substation's power supply region based on the power mapping result. Region growing refers to the gradual addition of neighboring pixels from a specific pixel according to a certain criterion, and the region growing terminates when certain conditions are satisfied.

There are two region growing conditions considered in the UrbanPS: (1) the Euclidean distance between the power demand point and the substation is less than or equal to the power supply radius; (2) the load rate of the substation is less than the maximum load rate.

The steps of substation supply area division using the region growing method area are as follows: First, the demand point located in the substation was designated as the seed point. The nearest power demand point to the substation was judged by the growth condition and was assigned to this substation if the condition was satisfied. Second, we considered the growth conditions of the grid of 8 neighboring demand points around the seed point and assigned the demand point that satisfies the conditions. When judging the growth condition of the power demand point, the Euclidean distance from all substations to this demand point was calculated. If the nearest substation meets the condition of continued growth, the demand point is assigned to the substation closest to it. Otherwise, a step-by-step judgment from near to far assigns that demand point to other substations. The above steps are cycled until one of the termination conditions is reached, which means the area growth is terminated, and the substation power supply range division is completed. In UrbanPS, there are three termination conditions: the substation reaching its maximum load, all power demand points being allocated, or the power supply radius of that substation being exceeded.

### 3.3 | Substation location optimization for different scenarios

The load mapping results at different scenarios could provide important data support for optimizing the substation location. The substation location optimization process is based on four different scenarios: extreme power consumption, winter peak power consumption, summer peak power consumption, and low power consumption. After the power mapping and substation supply area division, the final step of the UrbanPS framework is location optimization based on LSCP and genetic algorithm, consisting of two parts: problem construction and the optimization of the solution. As for the problem construction, the LSCP was used in UrbanPS to cover all service demand points in a certain area with a minimum number of facilities (Yu et al., 2020). The LSCP-based model is often used

for siting emergency facilities, such as fire stations and ambulances (Yu et al., 2020). The objective is to determine the minimum number of service facilities or construction costs while satisfying the need to cover all emergency points. Given that low coverage in the urban grid currently leads to local area shortage (Yadav et al., 2021), we consider that substations also have some emergency facility nature.

This study introduced the LSCP model based on the spatial distribution of power loads and characteristics of distribution networks (transformer capacity, transformer units, power factor, supply radius, maximum load factor). Considering the actual substation power supply range division, we aimed to maximize urban grid coverage and utilization of grid equipment (Honarparvar et al., 2021; Ye et al., 2018), thus establishing the LSCP-based urban substation site selection problem. The problem can be formulated as follows: within the urban area, the minimum number of terminal substations was used to cover all power demand points, and the objective was to maximize the grid coverage and equipment utilization.

In this study, the grid coverage is the percentage of demand points supplied with electricity (Wang et al., 2020). The utilization rate is defined as the ratio of the current load factor to the maximum load factor of the substation (Ye et al., 2018). At the same time, UrbanPS can also consider the optimization of the transformer's capacity by setting the capacity parameter  $C_j$ . The specific mathematical expression is as follows.

Objective function:

$$\text{Max} \left( W_1 \times \frac{\sum Y_i}{T_i} + W_2 \times \frac{\sum \frac{E_j X_j}{C_j L_j}}{T_j} \right) \quad (1)$$

The constraints on the variables:

$$E_j \leq C_j L_j \quad (2)$$

$$d_{ij} \leq R \quad (3)$$

$$W_1 + W_2 = 1 \quad (4)$$

where constraint (2) ensures that the substation provides power to the demand point within the power supply's safety and that no overload occurs. Constraint (3) limits the Euclidean distance between the demand point and the substation to be less than or equal to the supply radius of the substation. Constraint (4) ensures that the coverage weight and utilization weight sum to 1. The symbols associated with the equations are shown in Table 1. Among them,  $T_i$ ,  $L_j$ ,  $R$ ,  $W_1$ , and  $W_2$  are constants.  $C_j$  can be optimized by the algorithm or assigned by a constant according to the practical application scenarios. Other symbols represent variables that are calculated by the optimization model.

The next step is the optimization of the solution. The problem of selecting  $P$  optimal substation locations from  $N$  candidate locations to maximize the grid coverage and distribution utilization is NP-hard. Previous studies have shown that the heuristic algorithm effectively solves the NP-hard problem (Seyhan et al., 2018). Commonly used heuristics in site-selection problems include genetic algorithms (Yu et al., 2020), greedy search algorithms (Vasko et al., 2016), forbidden search (Silva & Cunha, 2017), and Lagrangian relaxation methods (Zhang, Beltran-Royo, et al., 2019). Genetic algorithms have been proven to effectively solve combinatorial optimization problems (Comber et al., 2011).

This framework used a genetic algorithm to find the optimal substation location for different scenarios. We used a parameter tuning approach to determine the optimal parameters to ensure that the algorithm obtains the global optimal solution (Saif-Eddine et al., 2019; Zhang, Zhang, et al., 2019). The parameters to be determined for the genetic algorithm are the number of iterations, the population size, and the chromosome length. By setting different parameter values within a specific range, the optimal parameters are determined based on the accuracy

TABLE 1 Relevant symbols in mathematical expressions.

Symbols	Definition
$i$	Index of power demand point
$j$	Index of substation
$T_i$	Total number of power demand points
$T_j$	Total number of selected substations
$X_j$	If substation $j$ is selected, $X_j=1$ ; otherwise, $X_j=0$
$Y_i$	If power demand point $i$ is assigned, $Y_i=1$ ; otherwise, $Y_i=0$
$L_j$	Maximum load factor of substation $j$
$C_j$	The total capacity of the main transformer in substation $j$
$E_j$	The actual active power output of substation $j$
$d_{ij}$	Distance between demand point $i$ and substation $j$
$R$	Power supply radius of the substation
$W_1$	Weight of coverage, $0 < W_1 < 1$
$W_2$	Weight of utilization, $0 < W_2 < 1$

of the results. The other parameters were kept constant while solving the relationship between one parameter and the approximate optimal solution.

### 3.4 | Evaluation of power supply capacity of urban substations

In order to evaluate the effectiveness of our proposed UrbanPS for site selection, it is crucial to perform an evaluation of substations before and after the optimization. Based on the substations' location and the electricity load mapping results, a region growing algorithm was used to divide the power supply area of each substation. The demand points covered and not covered by the substation were revealed. Finally, the grid coverage  $C_{\text{exist}}$  and transformer utilization  $D_{\text{exist}}$  were calculated as follows:

$$C_{\text{exist}} = \frac{\sum \gamma_k}{D} \quad (5)$$

$$D_{\text{exist}} = \frac{\sum W_k}{\sum M_k} \quad (6)$$

where  $k$  denotes the index number of the existing substation,  $\gamma_k$  denotes the power demand points covered by the substation,  $D$  denotes all the power demand points,  $W_k$  denotes the actual active power output of the substation, and  $M_k$  denotes the maximum active power that can be output by the substation.

## 4 | CASE STUDY

### 4.1 | Study area and data

Pingxiang is located in the western part of Jiangxi Province (Figure 2), with complex geomorphology. The total area of Pingxiang is 3823.99 km<sup>2</sup>, and there are 54 towns. The contribution of tourism to the economic growth of Pingxiang reached 48.6%, and the domestic tourism revenue in 2019 was 52.048 billion RMB (Pingxiang Bureau

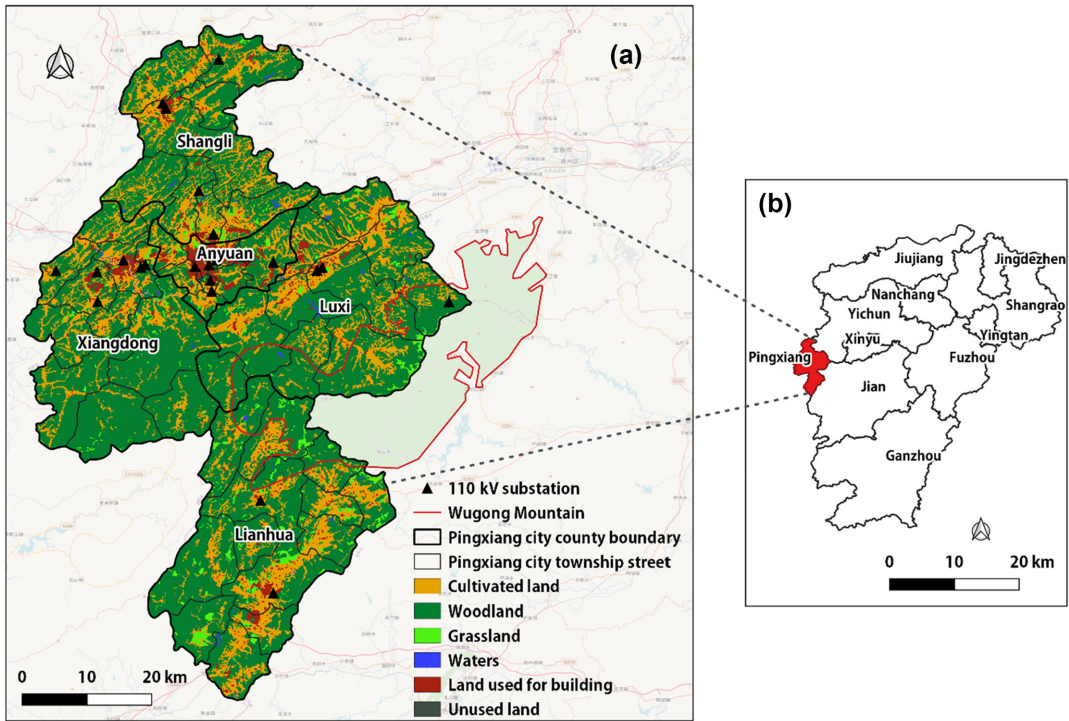


FIGURE 2 (a) Spatial distribution of land use types and existing 110kV substations in the study area, (b) Jiangxi Province, China.

of Statistics and Pingxiang General Survey Team of National Bureau of Statistics, *n.d.*). Tourism can indirectly affect the spatial pattern of power consumption in the city, thus impacting the planning layout of substations (Lai et al., 2008). Currently, there are 23,110kV terminal substations distributed in these five districts and counties.

We collected electricity consumption data recorded every half hour at substations from June to October 2017 and from January to March 2018. After cleaning and aggregating the data, the zonal statistics were used to obtain each township-level administrative district's average daily electric power. Such electric power was used as the power consumption data in our experiment. The power data collection information is shown in Table 1. The substation information contains transformer capacity, load factor, power supply radius, power factor, and other power data collection information is shown in Table 2.

A total of 18 spatial drivers (Figure 3) were used as the explanatory variables of power consumption in this study. They are derived from multi-source data, including POI, OSM road network, and nighttime light data. Among them, POI data and road network density strongly correlate with urbanization levels (Wu et al., 2021). Therefore, they are widely used for urban functional area identification (Yao et al., 2017; Yu et al., 2022). In this study, POI data in 2018 were obtained from the GAODE map ([www.amap.com](http://www.amap.com)), including residential communities, living services, educational institutions, corporate enterprises, entertainment venues, financial institutions, medical institutions, accommodation services, dining venues, shopping districts, and transportation facilities. The kernel density analysis was conducted to obtain the spatial distribution of the POI density of each category.

We extracted five types of roads, highways, primary roads, secondary roads, tertiary roads, and railroads, from the OSM road network data in 2018. The accessibilities to five types of roads at the grid level were calculated according to Euclidean distance. They were regarded as the line driver of the power load. Nighttime light intensity is commonly used to estimate regional socio-economic parameters, which has potential applications in energy consumption pattern analysis (He et al., 2012). The nighttime light data in 2018 were obtained from publicly available



TABLE 2 Collection information table of distribution transformer power data.

Station ID	Start time of collection	End time of collection	Electrical power (W)	Name	Longitude (°)	Latitude (°)
23**30	2018/1/1 0:00	2018/1/2 15:51	0.242	Liushi town	117.271	20.311
69**55	2018/1/1 0:00	2018/1/2 15:51	0.537	Paishang town	112.606	27.535
...	...	...	...	...	...	...
81**51	2018/1/1 0:00	2018/1/2 15:51	0.212	Shangbu town	114.058	27.575

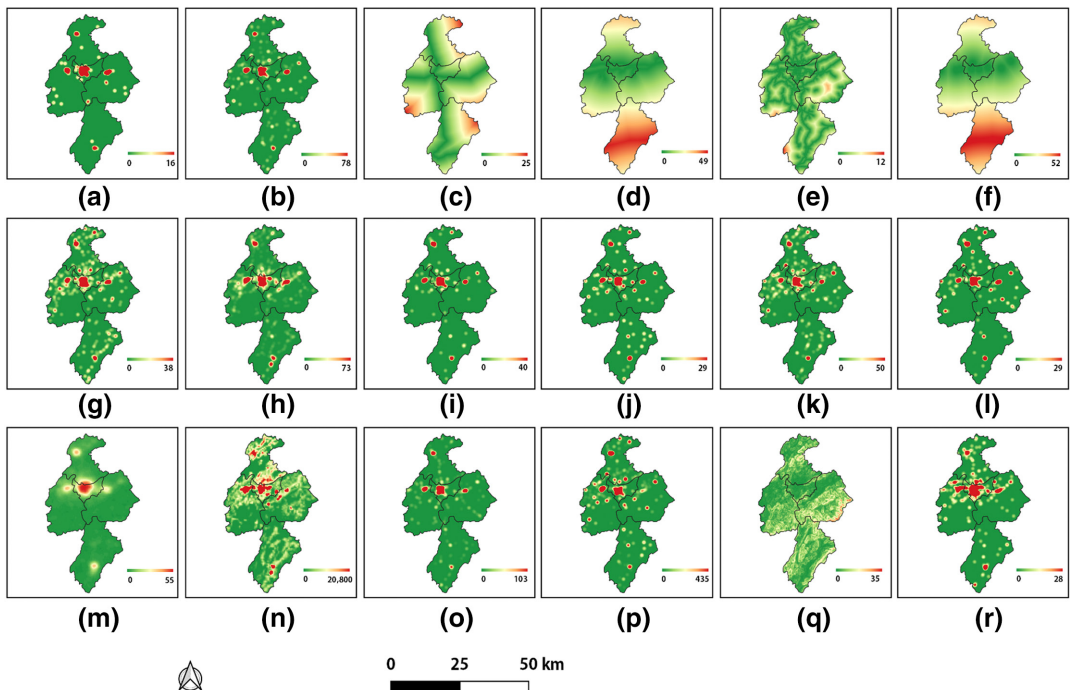


FIGURE 3 Explanatory variables for power demand: (a) Community density. (b) Density of amenities. (c) Distance to highway. (d) Distance to railroad. (e) Distance to main roads. (f) Distance to train station. (g) Density of educational institutions. (h) Density of companies. (i) Density of entertainment venues. (j) Density of financial institutions. (k) Density of medical institutions. (l) Density of hotels. (m) Light index at night. (n) Density of population. (o) Density of restaurant establishments. (p) Density of shopping areas. (q) Slope. (r) Density of transportation facilities.

data on the High-Resolution Earth Observation System of Systems website ([www.hbeos.org.cn](http://www.hbeos.org.cn)), with system-level geometric correction, radiometric correction, and resampling statistics to obtain the average light intensity in the region. The final acquired raster data were spatially interpolated to obtain a 100m spatial resolution driver dataset.

## 4.2 | Spatial distribution of fine-scale power load

The multi-source spatial data was divided into a training set (80%) and a test set (20%) in this study. The constructed power consumption estimation model had high accuracy, with fitting accuracy  $R^2$  of 0.453, RMSE of

0.010, and Pearson's  $r$  of 0.694. The importance analysis of the random forest model shows that the slope factor has the most significant influence on the power consumption load in Pingxiang, followed by the hotel density and night light index (Figure S4).

Multi-scenarios fine-scale power was estimated, and the grids with power load value greater than 0 were converted into power demand points (Figure 4). The total load of four power consumption scenarios shows a significant seasonal difference in Pingxiang. In the extreme power consumption scenario, the power consumption load in each region was relatively large, especially in the central city, which reached 991.686 MW. This was followed by 665.377 MW in the satellite city (Lianhua County) and 642.611 MW in the industrial area. The overall demand for power during the winter peak period declined, and the power load in the central city was 752.045 MW. Central urban areas, satellite cities, and industrial areas had the most significant electrical loads. The summer peak power demand was relatively small compared with the usual low peak power demand except for four areas: central city, industrial area, tourist attractions, and satellite city (Shangli County), which was due to the increase in the use of cooling equipment in the hot summer (Lai et al., 2008). In contrast, the satellite city (Lianhua County) had a summer power load of 316.682 MW and a usual power load of 375.379 MW, which showed a special situation in that the usual power consumption is higher than the summer power consumption. The overall residential power consumption in Pingxiang obtained by micro-scale fitting matches the actual urban power consumption pattern (Guo et al., 2018), with only very few regions having differential patterns.

### 4.3 | Spatial distribution evaluation analysis of existing substations

The coverage and utilization rates of the existing 23 substations in Pingxiang were evaluated. The results show that power supply shortage and low utilization of grid equipment exist in all scenarios. The coverage rate range from 71 to 95%, and the utilization rates range from 75 to 99%. Figure 5 and Table 3 show that Pingxiang suffers from serious power shortages under extreme power consumption and winter peak power consumption. The grid coverage rate was only 71.39% during the extreme power consumption period, prone to large area blackouts, especially in central cities, satellite cities (Shangli County), and scenic tourist areas. In the winter peak period, the grid coverage was 89.66%, and there was a shortage of power supply in the southern part of the central city and the northeastern part of the tourist area. These two scenarios with relatively high power demand had very low grid coverage and had a high utilization rate of nearly 99%. In the summer peak and usual low peak scenarios, the grid coverage was close to 95% to meet most of the regional power supply. Only the satellite city (Lianhua County) had a power shortage. The utilization rate of substations in these two scenarios was relatively low, with 91.41% in the summer peak period and 75.51% in the usual low peak period.

Among all the regions, the power supply situation in the satellite city (Lianhua County) was the tensest, with a serious power shortage phenomenon in all four scenarios. The results show that 58.27% of the power demand area in the satellite city cannot be supplied under the extreme scenario, 42.07% during the winter peak period, and 29.30 and 30.53% during the summer peak period and the usual low peak period, respectively.

### 4.4 | Multi-scenario optimization results

The substation location optimization was performed using the genetic algorithm with the optimal parameters set. We used Visual Studio software for the optimization, and the entire optimization process took 3039.952 s. We used parameter tuning to determine the optimal values of the sensitive parameters of the genetic algorithm (Zhang, Beltran-Royo, et al., 2019). The parameters used in this study were the initial number of iterations of 100, the population size of 20, and the chromosome length of 2839 (Figures S1 and Table S1). We set the same power capacity for all substations. Each substation was equipped with two transformers of 50 MVA capacity, that is,

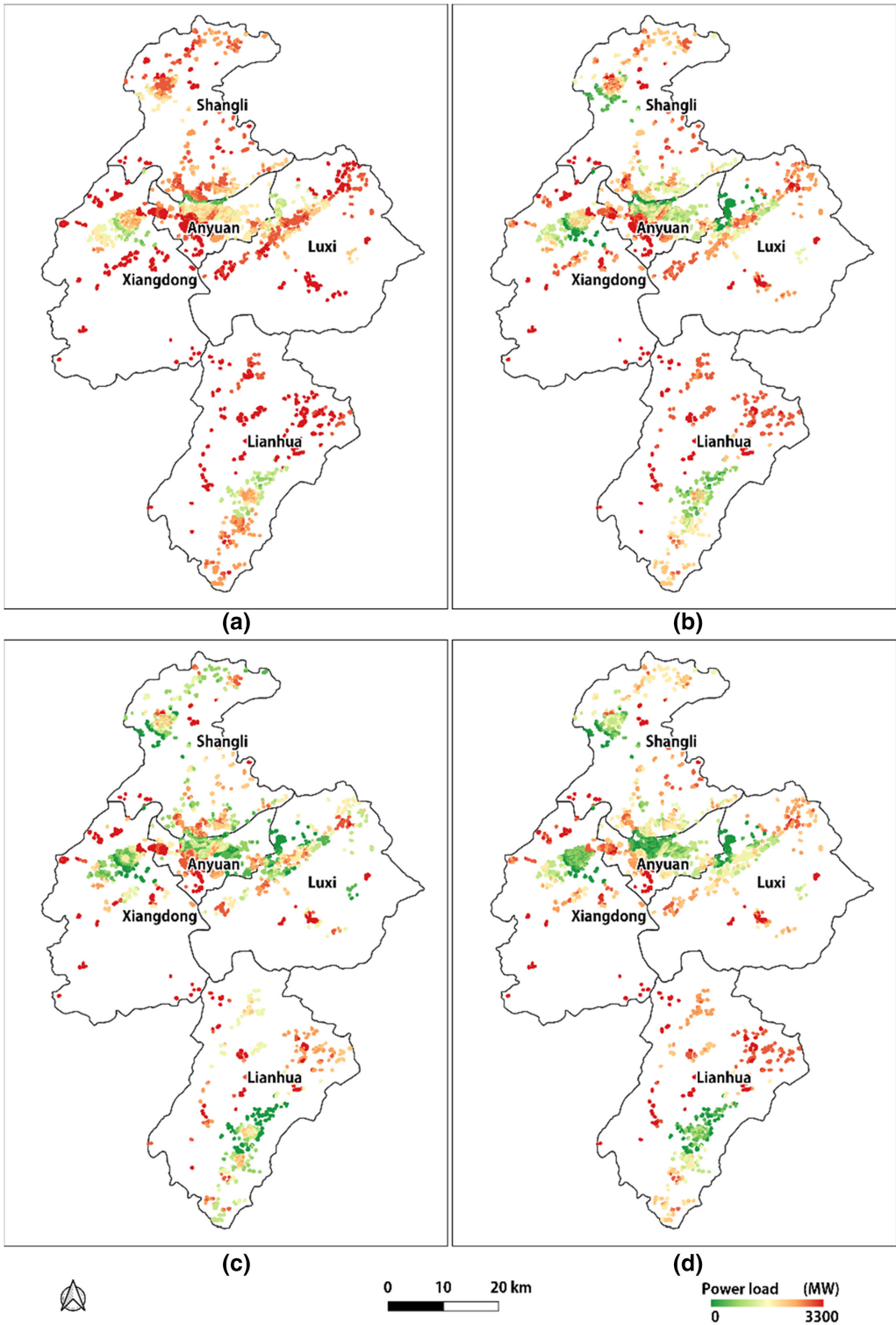


FIGURE 4 Fine-scale power consumption distribution for four different scenarios. (a) Extreme power consumption; (b) winter peak power consumption; (c) summer peak power consumption; (d) usual low peak power consumption.

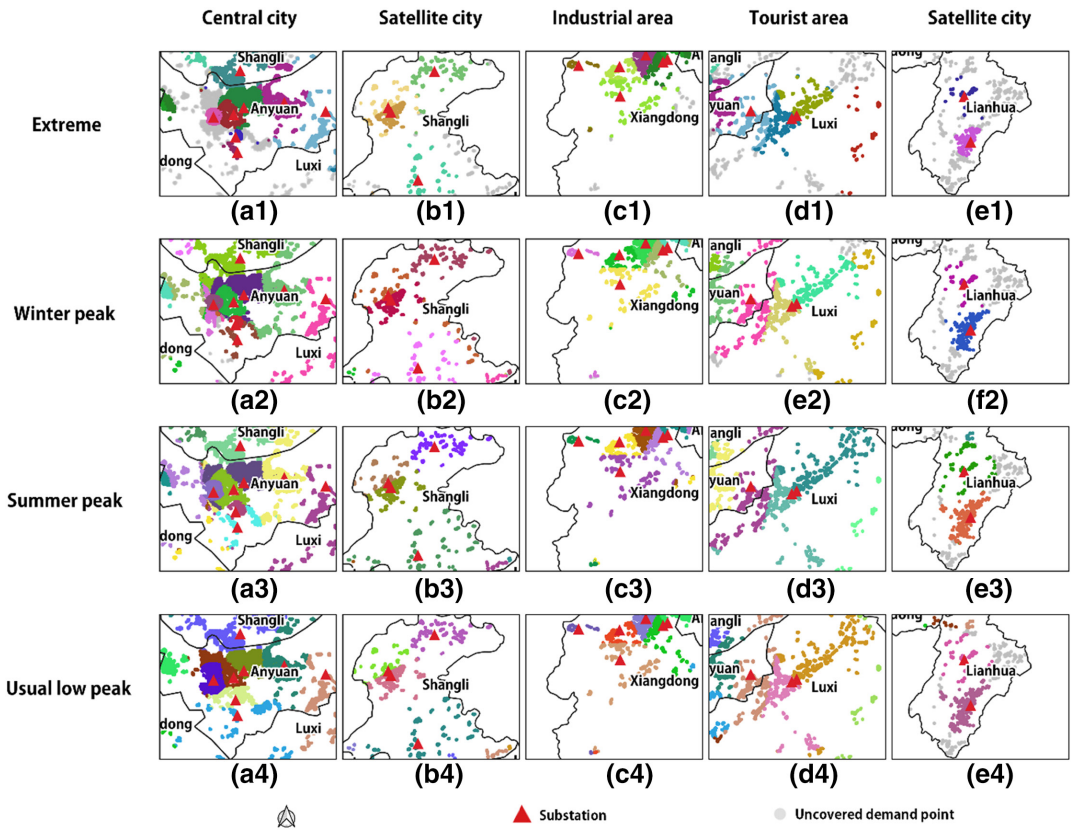


FIGURE 5 Demand points are covered and not covered by existing 110kV substations in different scenarios in Pingxiang. Demand points of the same color are supplied by the same substation. Demand points of the same color are supplied with electricity by a single substation.

$C_j$  is 100 MVA. We set a maximum load factor of 80%, that is,  $L_j=0.8$ , a power factor of 100%, and a maximum power supply radius of 30km from the substation (China Academy of Urban Planning and Design, 2014). In order to balance the grid coverage and utilization in the optimization objective (Ma et al., 2018), this study set equal weights for both, that is,  $W_1=0.5$  and  $W_2=0.5$ .

As shown in Table 3 and Figure 6, optimization results' coverage and utilization rate under four scenarios could reach 99–100%, which means the substation can cover all the power demand points and reduce transformer losses. Under the extreme scenario, the current coverage rate of the existing 23 substations was only 71.39%, and it was increased by nearly 30% to 99.47% after the optimization. Results show that 19 new substations were needed to meet the power demand under this scenario. They were suggested to build in the north, northeast, and southwest of the satellite city (Lianhua County) and the southern industrial zone. Under the winter peak power consumption scenario, the current coverage rate of the existing 23 substations was 89.66%, and it would increase nearly 10% to 99.46% after the optimization. To meet the power demand, five new substations were suggested to build in the satellite city and the southern part of the tourist attractions. In the summer peak power consumption scenario, the existing 23 substations cover 94.49 and 91.41% utilization rates, while after the optimization, they can cover nearly 100% for both indicators. Results show that adjusting the spatial location of substations can improve grid coverage and equipment utilization. In this scenario, two to three substations in the central city were suggested to move to the southern part of the industrial area and the southern part of the satellite city (Lianhua County). In the usual low peak scenario, the existing 23 substations had 94.49% coverage

TABLE 3 Comparison of the coverage and utilization rate of existing substations and optimized substations.

Scenario	Optimized result			Current result		
	Substation number	Coverage rate (%)	Utilization rate (%)	Substation number	Coverage rate (%)	Utilization rate (%)
Extreme power consumption	42	99.47	99.69	23	71.39	99.98
Winter peak power consumption	28	99.46	99.95	23	89.66	98.37
Summer peak power consumption	23	99.83	99.97	23	94.49	91.41
Usual low peak power consumption	20	99.50	99.97	23	94.49	75.51

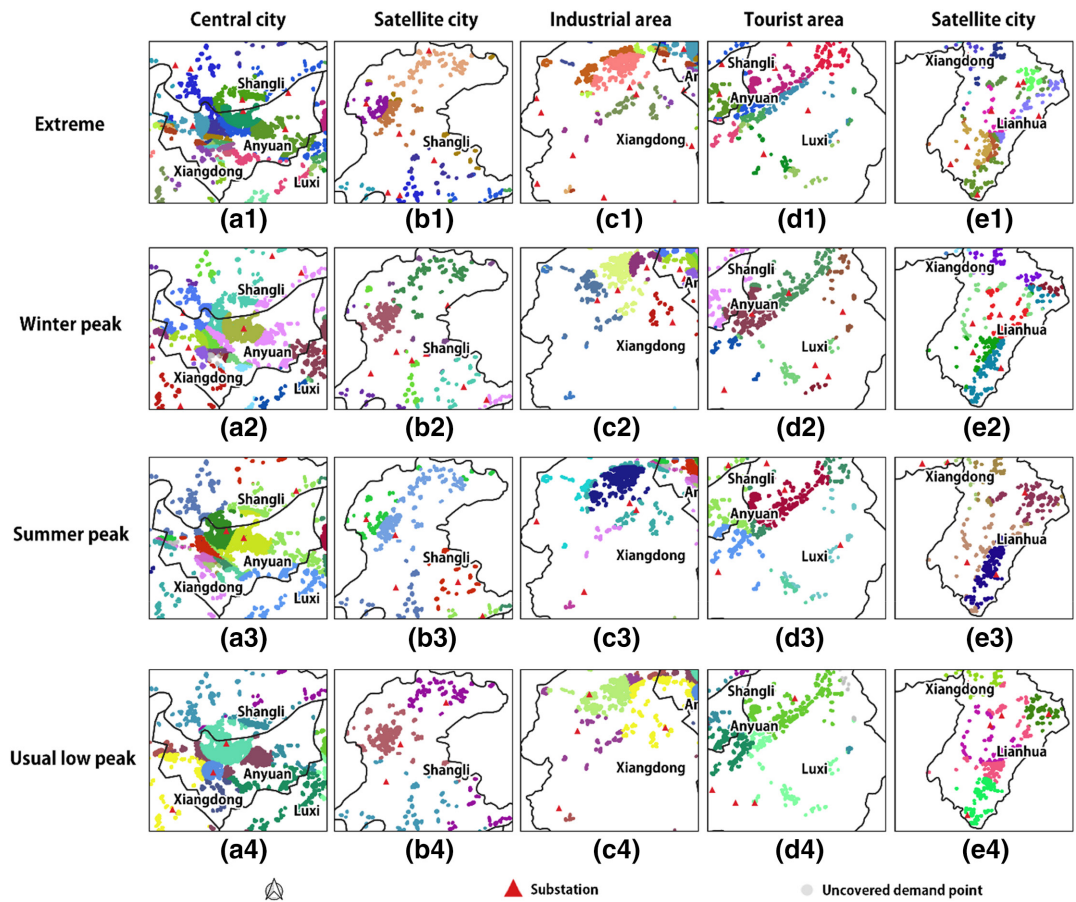


FIGURE 6 Spatial distribution of substation sites and their covered and uncovered demand points in different scenarios in Pingxiang, with demand points of the same color supplied by the same substation. Demand points of the same color are supplied with electricity by a single substation.

and a 75.51% utilization rate. The optimization results show that only 20 substations were needed to achieve close to 100% coverage and utilization in this scenario. Additional substations were required to provide power in satellite cities (Lotus County) and tourist attractions, and some substations in other areas were suggested to close to save energy.

## 5 | DISCUSSION

### 5.1 | Interpretation of the findings

We proposed the UrbanPS framework from the perspective of grid coverage to optimize the location of large-scale substations in newly built urban areas. The framework can accurately simulate the spatial distribution of urban residential electricity demand and effectively improve cities' grid coverage and equipment utilization. The results show that the power consumption estimate model has high accuracy ( $R^2=0.453$ ,  $RMSE=0.010$ , Pearson's  $r=0.694$ ). The weight analysis of driving factors of power load verified the feasibility and effectiveness of the model for urban residential power demand simulation. The optimized grid coverage and equipment utilization rate for different scenarios were close to 99%. The results demonstrate that the UrbanPS framework is a reliable method for parameter tuning and can effectively optimize the substation location.

The UrbanPS framework introduced multi-source spatio-temporal variables to reveal the drivers of electricity consumption. Population activity is the most significant factor affecting the distribution of electricity assumption. However, topographical conditions such as slopes can mainly limit electricity consumption in some areas. Previous studies on estimating electricity demand had insufficient fitting accuracy and large analysis units (1 km or 500 m) (Lu et al., 2019; Wang & Lu, 2021). The UrbanPS provides a new opportunity to explore the prevalent spatio-temporal heterogeneity of electricity consumption at a fine scale.

### 5.2 | Electricity consumption is influenced by urban spatial structure and seasons

The residential electricity load of fast-growing Chinese cities has spatial and temporal heterogeneity at the microscopic scale. The central city has a dense and uneven distribution of demand points (Figure 4). The value difference of neighboring demand points can be up to 330%. The high electricity loads are mainly located around hospitals, schools, and companies, indicating a high electricity consumption in areas with dense populations. The demand points in industrial areas show a regional aggregation, with high electricity consumption in various industrial parks, while residential electricity consumption is relatively low. In addition, the electricity load of tourist attractions and surrounding satellite cities is relatively high due to the influence of crowd gathering.

The residential electricity load also shows a strong seasonal effect. The seasonal effect is particularly pronounced in the central city, where winter and summer electricity loads are twice and 1.5 times higher than usual. This phenomenon is related to spiking electricity loads due to temperature control equipment such as air conditioners (Li, Yang, et al., 2018). The seasonal effect of electricity consumption in industrial areas is relatively weak, with load differences between seasons averaging 1.35 times. The differences between these two typical functional areas suggest that electricity loads have a strong seasonal pattern in population-dense areas. However, there are some exceptions. For example, the electricity load in tourist attractions during the tourist season (spring and autumn) is 1.05 times the summer electricity load, which is very close to both. This is due to the higher number of tourists in mountainous areas during the peak tourist season.

### 5.3 | Substation siting and refinement management

The proposed UrbanPS framework uses the LSCP model for substation siting optimization. The optimization results can alleviate the problems of power supply constraints and low utilization of electrical facilities and can be used to refine the management of grid resources. We found that the coverage and utilization rate of the optimization results increased by 5–30% compared to the existing distribution. Traditional substation siting studies are mainly limited to planning a minimal number of substations to achieve the optimization goal of minimizing the

total cost. Compared to the substation siting models proposed in recent years (Li, Yin, et al., 2018), the proposed UrbanPS framework considers the electricity consumption patterns of multiple urban scenarios and adopts a region growing algorithm to more finely delineate the power supply area of the substation. At the same time, UrbanPS innovatively uses grid coverage and utilization as the optimization objectives for substation siting. The model helps to optimize the location of future substations in new urban areas, relieving the pressure on the grid during peak periods and meeting the basic electricity needs of urban residents.

This study optimized the siting of substations under different scenarios by analyzing the residential electricity consumption patterns under multiple scenarios. The results might solve the power shortage problem during seasonal peak consumption periods. For example, during the seasonal peak power consumption, substations in the central city, industrial areas, and satellite cities can be turned on, thus preventing power outages due to overload operation of the distribution network. In contrast, several substations can be turned off to reduce power losses during the low power consumption period of spring and autumn. The study suggests that the operation of different substations can be dynamically adjusted by evaluating the actual local power load to balance the electrical energy supply and consumption.

## 5.4 | Limitations and future works

The present framework still has some limitations. First, in future research, we will dynamically simulate the opening and closing of substations in different city areas according to the seasonal power consumption pattern and conduct the dynamic dispatching distribution of electric energy among regions. Second, capacity planning is another essential part of urban power grid planning, which is not addressed in this study. In the future, we will establish a multi-objective optimization model for substation siting and capacity by adjusting the capacity  $C_j$  in the model and optimizing the capacity and location of substations simultaneously. Third, the disadvantage of heuristic optimization algorithms for solving NP-hard problems is it is difficult to obtain a global optimal solution (Deng et al., 2019; Wang & Sobey, 2020). In the future, we will consider further introducing reinforcement learning methods to enhance the algorithm's optimization performance.

## 6 | CONCLUSION

Aiming at the spatial planning of substations in new urban areas and dealing with the power shortage problem during peak periods, we proposed the UrbanPS framework from the perspective of large-scale urban grid coverage. The maximum coverage and utilization rate were regarded as the optimization objective. The substation siting optimization was conducted in Pingxiang, China, and the results verify the proposed methods' theoretical and practical contributions. Results show that the UrbanPS helps allocate power resources and effectively alleviates the power shortage problem during peak hours in a fast-growing city. It is suggested that power resources can be reasonably distributed by dynamically adjusting the operation of different substations. Findings from this study can guide grid planning in new urban areas in China.

## ACKNOWLEDGMENTS

We would like to thank editor and anonymous reviewers for their constructive suggestions and comments for improving this manuscript. Open Access funding enabled and organized by Projekt DEAL.

## CONFLICT OF INTEREST STATEMENT

No conflict of interest exists in this article, and the article was approved by all authors for publication.

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## ORCID

Yao Yao  <https://orcid.org/0000-0002-2830-0377>

## REFERENCES

- Anjos, M. F., & Vieira, M. V. (2017). Mathematical optimization approaches for facility layout problems: The state-of-the-art and future research directions. *European Journal of Operational Research*, 261(1), 1–16. <https://doi.org/10.1016/j.ejor.2017.01.049>
- Chen, Y., & Zhang, D. (2021). Theory-guided deep-learning for electrical load forecasting (TgDLF) via ensemble long short-term memory. *Advances in Applied Energy*, 1, 100004. <https://doi.org/10.1016/j.adapen.2020.100004>
- Cheng, J., Zhang, X., & Huang, J. (2022). Optimizing the spatial scale for neighborhood environment characteristics using fine-grained data. *International Journal of Applied Earth Observation and Geoinformation*, 106, 102659. <https://doi.org/10.1016/j.jag.2021.102659>
- China Academy of Urban Planning and Design. (2014). *Code for planning of urban electric power GB/T 50293-2014*.
- Comber, A. J., Sasaki, S., Suzuki, H., & Brunson, C. (2011). A modified grouping genetic algorithm to select ambulance site locations. *International Journal of Geographical Information Science*, 25(5), 807–823. <https://doi.org/10.1080/13658816.2010.501334>
- Deng, W., Xu, J., & Zhao, H. (2019). An improved ant colony optimization algorithm based on hybrid strategies for scheduling problem. *IEEE Access*, 7, 20281–20292. <https://doi.org/10.1109/ACCESS.2019.2897580>
- Ding, L., Zhou, J., Zhang, X., Liu, S., & Cao, R. (2018). Downscaling of surface air temperature over the Tibetan Plateau based on DEM. *International Journal of Applied Earth Observation and Geoinformation*, 73, 136–147. <https://doi.org/10.1016/j.jag.2018.05.017>
- Fan, J., Wang, J., Hu, J., Yang, Y., & Wang, Y. (2021). Will China achieve its renewable portfolio standard targets? An analysis from the perspective of supply and demand. *Renewable and Sustainable Energy Reviews*, 138, 110510. <https://doi.org/10.1016/j.rser.2020.110510>
- Guo, Z., Zhou, K., Zhang, C., Lu, X., Chen, W., & Yang, S. (2018). Residential electricity consumption behavior: Influencing factors, related theories and intervention strategies. *Renewable and Sustainable Energy Reviews*, 81, 399–412. <https://doi.org/10.1016/j.rser.2017.07.046>
- He, C., Ma, Q., Li, T., Yang, Y., & Liu, Z. (2012). Spatiotemporal dynamics of electric power consumption in Chinese Mainland from 1995 to 2008 modeled using DMSP/OLS stable nighttime lights data. *Journal of Geographical Sciences*, 22(1), 125–136. <https://doi.org/10.1007/s11442-012-0916-3>
- Heymann, F., Melo, J. D., Martínez, P. D., Soares, F., & Miranda, V. (2018). On the emerging role of spatial load forecasting in transmission/distribution grid planning. *11th Mediterranean Conference on Power Generation, Transmission, Distribution and Energy Conversion, Dubrovnik, Croatia* (pp. 1–6). IET.
- Honarparvar, S., Malek, M., Saeedi, S., & Liang, S. (2021). IPAWL: An integrated power aware wireless sensor network and location-based social network for incidence reporting. *International Journal of Applied Earth Observation and Geoinformation*, 104, 102540. <https://doi.org/10.1016/j.jag.2021.102540>
- Hu, S., Ge, Y., Liu, M., Ren, Z., & Zhang, X. (2022). Village-level poverty identification using machine learning, high-resolution images, and geospatial data. *International Journal of Applied Earth Observation and Geoinformation*, 107, 102694. <https://doi.org/10.1016/j.jag.2022.102694>
- Kaewmamuang, K., Siritarativat, A., Surawanitkun, C., Khunkitti, P., & Chatthaworn, R. (2019). A novel method for solving multi-stage distribution substation expansion planning. *Energy Procedia*, 156, 371–383. <https://doi.org/10.1016/j.egypro.2018.11.109>
- Lai, T. M., To, W. M., Lo, W. C., & Choy, Y. S. (2008). Modeling of electricity consumption in the Asian gaming and tourism center—Macao SAR, People's Republic of China. *Energy*, 33(5), 679–688. <https://doi.org/10.1016/j.energy.2007.12.007>
- Li, L., Yang, Y., Wang, C., & Lin, K. (2018). Biogeography-based optimization based on population competition strategy for solving the substation location problem. *Expert Systems with Applications*, 97, 290–302. <https://doi.org/10.1016/j.eswa.2017.12.039>
- Li, Z. H., Yin, X., Yang, H., Ma, R., Shi, G., & Zhao, W. (2018). Seasonal load characteristics analysis based on improved k-means clustering algorithm. *Power System and Clean Energy*, 34(2), 53–59.
- Liu, S., Du, Z., Wang, Q., & Zhou, T. (2017). Analysis on the development ideas of electric power industry in the period of the 13th five-year. *Electric Power*, 50(1), 7–12.



- Lu, L., Weng, Q., Xie, Y., Guo, H., & Li, Q. (2019). An assessment of global electric power consumption using the Defense Meteorological Satellite Program-Operational Linescan System nighttime light imagery. *Energy*, 189, 116351. <https://doi.org/10.1016/j.energy.2019.116351>
- Ma, L., Liu, W., Chen, H., Cui, Y., Wang, Y., & Yuan, S. (2018). Operation efficiency evaluation frame and its criteria for distribution network based on annual load duration curve. *IEEE Innovative Smart Grid Technologies-Asia, Singapore* (pp. 373–378). IEEE.
- Pingxiang Bureau of Statistics and Pingxiang General Survey Team of National Bureau of Statistics. (n.d.). *Pingxiang City 2019 National Economic and Social Development Statistical Bulletin*.
- Qin, C., Yan, Q., & He, G. (2019). Integrated energy systems planning with electricity, heat and gas using particle swarm optimization. *Energy*, 188, 116044. <https://doi.org/10.1016/j.energy.2019.116044>
- Saif-Eddine, A. S., El-Beheiry, M. M., & El-Kharbotly, A. K. (2019). An improved genetic algorithm for optimizing total supply chain cost in inventory location routing problem. *Ain Shams Engineering Journal*, 10(1), 63–76. <https://doi.org/10.1016/j.asej.2018.09.002>
- Senthil Kumar, J., Charles Raja, S., Srinivasan, D., & Venkatesh, P. (2018). Hybrid renewable energy-based distribution system for seasonal load variations. *International Journal of Energy Research*, 42(3), 1066–1087. <https://doi.org/10.1002/er.3902>
- Seyhan, T. H., Snyder, L. V., & Zhang, Y. (2018). A new heuristic formulation for a competitive maximal covering location problem. *Transportation Science*, 52(5), 1156–1173. <https://doi.org/10.1287/trsc.2017.0769>
- Shu, J., Cao, M., Zhang, M., Wang, X., Cao, W., Fang, X., & Cao, M. (2020). Molecular patching engineering to drive energy conversion as efficient and environment-friendly cell toward wireless power transmission. *Advanced Functional Materials*, 30(10), 1908299. <https://doi.org/10.1002/adfm.201908299>
- Silva, M. R., & Cunha, C. B. (2017). A tabu search heuristic for the uncapacitated single allocation p-hub maximal covering problem. *European Journal of Operational Research*, 262(3), 954–965. <https://doi.org/10.1016/j.ejor.2017.03.066>
- Sun, M., Wang, Y., Strbac, G., & Kang, C. (2018). Probabilistic peak load estimation in smart cities using smart meter data. *IEEE Transactions on Industrial Electronics*, 66(2), 1608–1618. <https://doi.org/10.1109/TIE.2018.2803732>
- Taghizadeh-Hesary, F., Rasoulinezhad, E., Shahbaz, M., & Vo, X. V. (2021). How energy transition and power consumption are related in Asian economies with different income levels? *Energy*, 237, 121595. <https://doi.org/10.1016/j.energy.2021.121595>
- Uddin, M., Romlie, M. F., Abdullah, M. F., Abd Halim, S., & Kwang, T. C. (2018). A review on peak load shaving strategies. *Renewable and Sustainable Energy Reviews*, 82, 3323–3332. <https://doi.org/10.1016/j.rser.2017.10.056>
- Vasko, F. J., Lu, Y., & Zyma, K. (2016). What is the best greedy-like heuristic for the weighted set covering problem? *Operations Research Letters*, 44(3), 366–369. <https://doi.org/10.1016/j.orl.2016.03.007>
- Wang, Q., Zhao, X., Lv, Z., Ma, X., Zhang, R., & Lin, Y. (2020). Optimizing the ultra-dense 5G base stations in urban outdoor areas: Coupling GIS and heuristic optimization. *Sustainable Cities and Society*, 63, 102445. <https://doi.org/10.1016/j.scs.2020.102445>
- Wang, J., & Lu, F. (2021). Modeling the electricity consumption by combining land use types and landscape patterns with nighttime light imagery. *Energy*, 234, 121305. <https://doi.org/10.1016/j.energy.2021.121305>
- Wang, Z., & Sobey, A. (2020). A comparative review between genetic algorithm use in composite optimisation and the state-of-the-art in evolutionary computation. *Composite Structures*, 233, 111739. <https://doi.org/10.1016/j.composit.2019.111739>
- Wu, H., Lin, A., Xing, X., Song, D., & Li, Y. (2021). Identifying core driving factors of urban land use change from global land cover products and POI data using the random forest method. *International Journal of Applied Earth Observation and Geoinformation*, 103, 102475. <https://doi.org/10.1016/j.jag.2021.102475>
- Xia, S., Chan, K. W., Luo, X., Bu, S., Ding, Z., & Zhou, B. (2018). Optimal sizing of energy storage system and its cost-benefit analysis for power grid planning with intermittent wind generation. *Renewable Energy*, 122, 472–486. <https://doi.org/10.1016/j.renene.2018.02.010>
- Ximing, C., Dian, B., Chaojun, Q., & Shao-Jie, Z. (2019). Research on construction practice of assembled substation—A case in Anhui Xuancheng Xinhang 110 kV substation project. *E3S Web of Conferences*, 136, 2038.
- Yadav, M., Pal, N., & Saini, D. K. (2021). Resilient electrical distribution grid planning against seismic waves using distributed energy resources and sectionalizers: An Indian's urban grid case study. *Renewable Energy*, 178, 241–259. <https://doi.org/10.1016/j.renene.2021.06.071>
- Yang, M., Kong, B., Dang, R., & Yan, X. (2022). Classifying urban functional regions by integrating buildings and points-of-interest using a stacking ensemble method. *International Journal of Applied Earth Observation and Geoinformation*, 108, 102753. <https://doi.org/10.1016/j.jag.2022.102753>
- Yao, Y., Li, X., Liu, X., Liu, P., Liang, Z., Zhang, J., & Mai, K. (2017). Sensing spatial distribution of urban land use by integrating points-of-interest and Google Word2Vec model. *International Journal of Geographical Information Science*, 31(4), 825–848. <https://doi.org/10.1080/13658816.2016.1244608>

- Yao, Y., Liang, Z., Yuan, Z., Liu, P., Bie, Y., Zhang, J., Wang, R., Wang, J., & Guan, Q. (2019). A human-machine adversarial scoring framework for urban perception assessment using street-view images. *International Journal of Geographical Information Science*, 33(12), 2363–2384. <https://doi.org/10.1080/13658816.2019.1643024>
- Ye, C., Ding, Y., Wang, P., & Lin, Z. (2019). A data-driven bottom-up approach for spatial and temporal electric load forecasting. *IEEE Transactions on Power Systems*, 34(3), 1966–1979. <https://doi.org/10.1109/TPWRS.2018.2889995>
- Ye, L., Hu, Z., Li, C., Zhang, Y., Jiang, S., Yang, Z., & Zhang, D. (2018). The reasonable range of life cycle utilization rate of distribution network equipment. *IEEE Access*, 6, 23948–23959. <https://doi.org/10.1109/ACCESS.2018.2803840>
- Yi, B., Xu, J., & Fan, Y. (2016). Inter-regional power grid planning up to 2030 in China considering renewable energy development and regional pollutant control: A multi-region bottom-up optimization model. *Applied Energy*, 184, 641–658. <https://doi.org/10.1016/j.apenergy.2016.11.021>
- Yin, J., Dong, J., Hamm, N. A., Li, Z., Wang, J., Xing, H., & Fu, P. (2021). Integrating remote sensing and geospatial big data for urban land use mapping: A review. *International Journal of Applied Earth Observation and Geoinformation*, 103, 102514. <https://doi.org/10.1016/j.jag.2021.102514>
- Yu, W., Chen, J., & Wei, C. (2022). A hierarchical learning model for inferring the labels of points of interest with unbalanced data distribution. *International Journal of Applied Earth Observation and Geoinformation*, 108, 102751. <https://doi.org/10.1016/j.jag.2022.102751>
- Yu, W., Chen, Y., & Guan, M. (2020). Hierarchical siting of macro fire station and micro fire station. *Environment and Planning B: Urban Analytics and City Science*, 48(7), 1972–1988. <https://doi.org/10.1177/2399808320958424>
- Zeng, Y., Zhang, R., Wang, D., Mu, Y., & Jia, H. (2019). A regional power grid operation and planning method considering renewable energy generation and load control. *Applied Energy*, 237, 304–313. <https://doi.org/10.1016/j.apenergy.2019.01.016>
- Zhang, H., Beltran-Royo, C., Wang, B., & Zhang, Z. (2019). Two-phase semi-Lagrangian relaxation for solving the uncapacitated distribution centers location problem for B2C E-commerce. *Computational Optimization and Applications*, 72(3), 827–848. <https://doi.org/10.1007/s10589-019-00061-5>
- Zhang, H., Zhang, Q., Ma, L., Zhang, Z., & Liu, Y. (2019). A hybrid ant colony optimization algorithm for a multi-objective vehicle routing problem with flexible time windows. *Information Sciences*, 490, 166–190. <https://doi.org/10.1016/j.ins.2019.03.070>
- Zheng, Y., Capra, L., Wolfson, O., & Yang, H. (2014). Urban computing: Concepts, methodologies, and applications. *ACM Transactions on Intelligent Systems and Technology*, 5(3), 1–55. <https://doi.org/10.1145/2629592>

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Yao, Y., Feng, C., Xie, J., Yan, X., Guan, Q., Han, J., ... 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, 1662–1679. <https://doi.org/10.1111/tgis.13093>