An Efficient GIS-based Substation Placement and Sizing Strategy Using Semi-supervised Learning

Li Yu, Di Shi, *Senior Member*, *IEEE*, Xiaobin Guo, Zhen Jiang, Guangyue Xu, *Member*, *IEEE*, Ganyang Jian, Jinyong Lei, Chaoyang Jing

***Abstract*—As load and renewable penetration continue to grow, optimal placement and sizing of substation is becoming increasingly important in distribution system planning. This paper presents an improved methodology to solve the substation siting and sizing problem based on geographic information and supervised learning. The proposed approach can optimize the locations, capacities, power supply ranges of substations with minimum investment and annual operation cost. Capital cost of land adds complexity and difficulty to the substation placement problem, especially for highly developed urban areas. This paper presents a theoretical framework to determine the optimal location of substations considering the cost of land. The state-of-the-art parallel computing techniques are employed so that co-optimization for substations of multiple voltage levels can be directly conducted in a computational efficient way. Case studies are presented to demonstrate the effectiveness of the proposed approach.**

***Index Terms*—Distribution network planning, substation siting, substation sizing, GIS, parallel computing, cost of land.**

I. Introduction

China Southern Power Grid (CSG) operates one of the world's largest and most complex AC and DC hybrid power grid [1]. In recent years, with continuous increase in load and renewable penetration within CSG's footprint, the need for accurate and intuitive distribution planning tools continues to grow. Towards this end, a comprehensive Distribution Planning Package (DPP) has been developed at CSG-EPRI based on the geographic information system (GIS).

Distribution network planning refers to the process of determining major projects required to meet future load while maintaining system reliability criteria, with minimum investment and operational cost. It determines the optimal location, capacity, power supply range for each substation, and feeder number and routing. Typical planning model involves a huge number of variables, many of which are binary and nonlinear, and its solution, due to computation time required, can only be solved approximately or through heuristics. In its most complete form, the problem reduces to a nonlinear integer-programming problem, with a cost function including the capital cost of new equipment, losses in the network and O&M costs. In general, it is very difficult to guarantee that the solution found is the global optimal solution and trade-off is always needed between computation time and quality of the solution [2]. This paper focuses on the optimal placement and sizing of substation and discusses the major challenges encountered during the development of DPP.

Extensive research has been conducted in the field of optimal substation siting and sizing. In general, existing solutions can be divided into three major categories: mathematical optimization [3]-[4], intelligent optimization [5]-[6], and heuristic optimization [7]-[10]. Mathematical optimization models usually have strict optimality but are very difficult to solve when the size and complexity of the system grows substantially. In recent years, a growing number of intelligent optimization techniques have been applied to solve this problem. Simulated annealing algorithms have been applied to determine substations' location, capacity, and regional division with relatively good solution quality [11]-[12]. However, these approaches can hardly handle large-scale problems with reasonable computation time due to parameter selection and the various annealing requirements. Tabu search algorithms have also been applied to tackle the problem, as discussed in [13]-[14]. Besides computational inefficiency and slow convergence rate, these methods also suffer from low solution quality as they can easily get trapped in local solutions due to limited searching capability. Some Genetic Algorithms (GAs) are proposed in [15]-[16], and again these approaches typically suffer from uncertainty in computation time.

Another important factor which has been neglected in most existing literature is the cost of land, which becomes increasingly expensive in developed urban areas and starts to account for a significant portion of the total investment [17]-[19]. In addition, although GIS has been used in many power system applications, unfortunately distribution system planning is typically not among them. In practice, modern distribution system planning requires extensive details which takes significant amount of development efforts to handle. Without considering geographic information, substation locations, more often than not, may be placed in infeasible regions, such as in the middle of lakes, on mountains, on buildings, or in areas with very high land cost. This paper

L. Yu, X. Guo, Z. Jiang, G. Jian, and J. Lei are with Southern China Power Grid EPRI, Guangzhou, China.

D. Shi, G. Xu, and C. Jing are with eMIT, LLC., Pasadena, CA, United States. Email: c.jing@myemit.com.

presents a computationally efficient methodology to handle land cost in the substation siting problem. Assisted with GIS, the proposed approach is capable of modeling feasibilities of regions in the substation siting and sizing process and generates better solution as compared to existing approaches. Further, the state-of-the-art parallel computing technique can be used so that co-optimization for substations at multiple voltage levels can be directly conducted in a computationally efficient way. The proposed approach is also easy to implement with numerical stability.

The remainder of this paper is organized as follows. Section II discusses the substation siting problem with various types of constraints considered. Section III presents the substation sizing algorithm and the solution process of DPP. Two case studies are presented in section IV while conclusions are drawn in section V.

II. GIS Based Substation Siting Algorithm

This section presents the substation siting and sizing problem with various constraints considered as well as the proposed algorithm. The discussion starts with a simple single-substation placement problem and then generalizes its solution process to the case of multiple substations.

A. Single Substation Siting with Constraints

Assuming there are a total N loads in the system, with the ith load L_i located at coordinate (x_i, y_i) consuming power W_i. Objective function of the single substation placement problem is to minimize a weighted distance/cost function $d(x,y)$ defined as:

$$d(x,y)=\sum_{i=1}^{n}\left(W_i\sqrt{(x-x_i)^2+(y-y_i)^2}\right) \tag{1}$$

Without considering cost of land, the optimal location of this substation in 2-D space, (x_q, y_q), can be obtained as:

$$d(x_q,y_q)=min(d(x,y)) \tag{2}$$

It can be proved that $d(x,y)$ is a convex function with a global optimum satisfying:

$$\frac{\partial d}{\partial x}=\sum_{i=1}^{n}\left(W_i\frac{(x-x_i)}{\sqrt{(x-x_i)^2+(y-y_i)^2}}\right)=0 \tag{3}$$

$$\frac{\partial d}{\partial y}=\sum_{i=1}^{n}\left(W_i\frac{(y-y_i)}{\sqrt{(x-x_i)^2+(y-y_i)^2}}\right)=0 \tag{4}$$

An iterative approach can be utilized to solve (3)-(4) based on a set of given initial estimates using:

$$x^{(k+1)}=\sum_{i=1}^{n}\frac{W_i x_i}{d_i^{(k)}}\Big/\sum_{i=1}^{n}\frac{W_i}{d_i^{(k)}} \tag{5}$$

$$y^{(k+1)}=\sum_{i=1}^{n}\frac{W_i y_i}{d_i^{(k)}}\Big/\sum_{i=1}^{n}\frac{W_i}{d_i^{(k)}} \tag{6}$$

where $x^{(k)}$ and $y^{(k)}$ are values of coordinates, x and y, at the kth update, and $d_i^{(k)}=\sqrt{(x^{(k)}-x_i)^2+(y^{(k)}-y_i)^2}$. Note that speedup factors and Matrix Splitting techniques can be employed to speed up and parallelize the solution process.

With a line constraint, the corresponding objective function becomes:

$$d=\sum_{i=1}^{n}\left(W_i\sqrt{(x-x_i)^2+(y-y_i)^2}\right) \quad s.t.\ \ ax+by\geq c \tag{7}$$

Assuming the optimal solution without this line constraint is (x_p, y_p), there are two possibilities for this constrained problem. If $ax_q+by_q\geq c$, the optimal solution is (x_p, y_p) itself; otherwise, it can be proved that the optimal solution must be on the line as the objective function d with line constraint is still convex. For the second case, solution of the problem can be found, again, using an iterative approach:

$$x^{(k+1)}=\sum_{i=1}^{n}\left(\frac{W_i}{d_i^{(k)}}(x_i-\beta y_i-\delta)\right)\Big/\sum_{i=1}^{n}\left(\frac{(1+\beta^2)W_i}{d_i^{(k)}}\right) \tag{8}$$

$$y^{(k+1)}=(c-ax^{(k+1)})/b \tag{9}$$

where $\beta=a/b$, and
$d_i^{(k)}=\sqrt{(x^{(k)}-x_i)^2+(\beta x^{(k)}+y_i+\delta/\beta)^2}$.

In the rare case that $b=0$, solution of the problem can be obtained as:

$$y^{(k+1)}=\sum_{i=1}^{n}\frac{W_i y_i}{d_i^{(k)}}\Big/\sum_{i=1}^{n}\frac{W_i}{d_i^{(k)}} \tag{10}$$

$$x_q=\frac{c}{a} \tag{11}$$

where $d_i^{(k)}=\sqrt{(c/a-x_i)^2+(y^{(k)}-y_i)^2}$.

A similar observation can be further extended to the case with a line segment constraint. If the optimum (x_p, y_p) obtained without considering the constraint is not on the line segment, then the point on the line segment which is closest to (x_p, y_p) should be the optimal solution. Otherwise, if (x_p, y_p) is on the line segment, then (x_p, y_p) is the solution.

A more complicated case is the polygon constraint. A polygon constraint is typically defined by a set of vertices. Assuming we have a polygon with m vertices, with coordinates of $(p_1, p_2, ..., p_m)$. This polygon can either be a convex polygon or a concave one. Except for the vertex, any edge of the polygon cannot intersect with any other edges, and each vertex connects two edges. The solution process to the single substation placement problem with polygon constraint can be divided into several steps. The first step is to find the optimal location, p_0, without any constraint. If p_0 lies within the polygon, then p_0 is solution of the problem. Otherwise, it can be proved that the solution must lie on the edge/boundary of the polygon. Therefore, the solution process can be decomposed into m sub-problems, each of which is to solve the placement problem with line segment constraint. Upon solving the m sub-problems, the optimal solution can be obtained by comparing values of the objective functions of the m sub-problems. The one with the minimum value is the optimal solution of the original problem.

B. GIS-based Single Substation Siting

When geographic information is not considered, a substation basically can be placed/built anywhere within the planned area as long as the spot gives the lowest total cost. When geographic information is considered, as certain regions within the planned area are taken, the siting algorithm has to avoid these areas. There are also areas that cannot be utilized, such as lakes, rivers, canyons, etc. The substation siting algorithm must avoid these areas as well. When setting up the mathematical models, the aforementioned constraints can be modeled as regions with different land prices. To stay away from one particular region, a very high land price can be assigned to it. In addition, ordinary areas can also be distinguished by land prices. For example, if we want to avoid the central business district, we can assign a very high land price to it; otherwise, if we prefer to place the substation close to an industrial area, we can make the corresponding land price negative.

With geographic information considered, the simplest case is to divide the planned area with one line, such as low land price in the south and high land price in the north, as shown in Fig. 1. If the optimum obtained without considering land cost is located in area II, then this point is the global optimum. Otherwise, if the optimum without considering land cost, p_1, is located in the north, then we can solve the problem again with line segment constraint (the boundary) to find the optimum p_2. Assuming the land costs of area I and II are L_1 and L_2, respectively, if $d(p_1) + L_1 < d(p_2) + L_2$, then p_1 is still the global optimum with land cost considered; otherwise, p_2 is the global optimum.

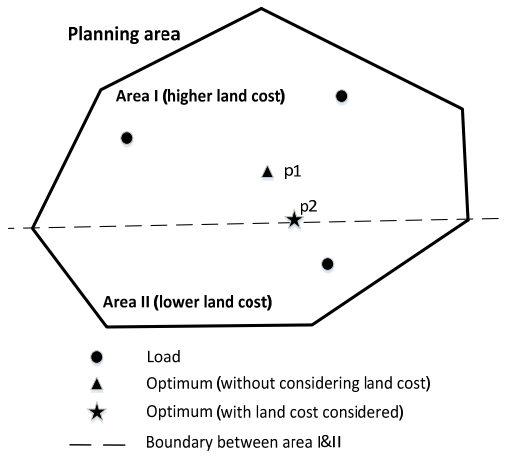


Fig. 1 An illustrative example for GIS-based substation siting

Further, a planning region may be divided into many pieces, each with a different land price. For example, Fig. 2 shows a planning area which is divided into five pieces. Without loss of generality, we assume that a planning area is divided into m pieces and costs of constructing the same-sized substation in these sub-areas are (L_1, L_2,…,L_m). As each land can be modeled as a polygon, we can apply the aforementioned algorithm for polygon constrained single substation siting for each of the m lands to obtain the local optimum for each sub-problem, denoted by $(p_1, p_2, \dots, p_m)$, and the corresponding d's, denoted by $(d_1, d_2, \dots, d_m)$. We can then evaluate the total cost for each sub-problem by adding the distance cost, which is d, and the corresponding land cost, as:

$$c_i = d_i + L_i \tag{12}$$

where $i = (1,2,\dots,m)$.

Among all the c_i's, we can select the smallest one, denoted as c_k, where $c_k = \min(c_i)$, and therefore c_k is the global optimum for the planning area with land cost taken into consideration. It is noted that this solution process includes solving multiple sub-problems which are independent to each other, and therefore parallel computing can be applied to speed up this process.

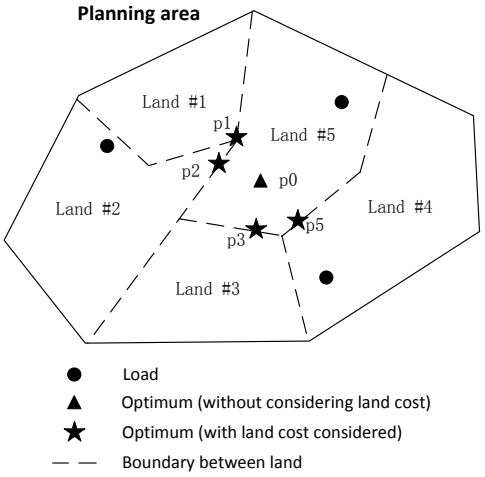


Fig. 2 A general case of GIS-based substation siting

C. GIS-based Multiple Substation Siting

Similar to the single substation siting problem, objective of multiple substation siting is to find locations of M substations for N loads so that the total investment can be minimized. The objective function becomes:

$$min \sum_{j=1}^{M} \left(\sum_{i=1}^{N} W_i a_{ij} d_{ij} + L(p_j) + B(c_j) \right) \tag{13}$$

where p_j is the location of the jth substation, d_{ij} is the distance between the ith load and the jth substation, W_i is the weight of the ith load, and a_{ij} indicates the connectivity between the ith load and the jth substation. Variable a_{ij} is 1 if substation j supplies power to load i, and 0 otherwise. Variable L is a function of the geographic information, c_j is capacity of substation j, and B is the construction cost of substation j.

Although this problem is NP-hard and it's very difficult to get its global optimal solution, a reasonably accurate solution can be obtained using the proposed iterative approach. The proposed approach divides the substation siting problem into two sub-problems, which can be solved in an iterative way, as shown below:

Step 1) Select M substations and assign N loads to these substations. Each load can be connected to only one substation.

The objective of the assignment is to minimize the following objective function:

$$min \sum_{j=1}^{M} \left(\sum_{i=1}^{N} W_i a_{ij} d_{ij} \right) \quad (14)$$

Step 2) For each group of loads, calculate their centroid subject to constraints based on geographic information. Update the location of the corresponding substation and then go back to step 1.

Step 3) Continue solving 1) and 2) iteratively until convergence is reached.

III. Substation Sizing Algorithm

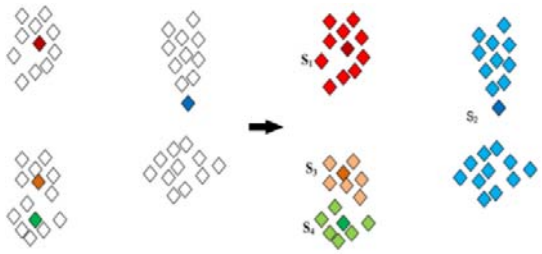
A. Load Clustering using Semi-Supervised Learning

The load clustering problem is essentially to group the loads according to their characteristics (size, geographical location, etc.). Clustering analysis or classification is one of the most popular machine learning problem. For classification, the input training data is characterized by a label. The essence of the learning process is to find the relationship between features and labels. Semi-supervised learning (SSL) is a key problem in the field of pattern recognition and machine learning. This section presents a semi-supervised learning approach and discusses its application in load clustering.

Given a set of n loads and their locations ($\boldsymbol{x_1}$, $\boldsymbol{x_2}$,…, $\boldsymbol{x_n}$), the objective of the load clustering problem is to divide loads into k ($k \le n$) clusters, denoted by $\mathbf{S}=\{S_1, S_2,\ldots, S_k\}$, so that the sum of all distances between each load and its corresponding centroid μ's is minimized, defined as:

$$\arg min \sum_{i=1}^{k} \sum_{x \in S_i} dist(x, \mu_i) \quad (15)$$

Note that distance function *dist*(.) can have different forms. The most popular clustering algorithm is the k-means algorithm, which belongs to unsupervised learning. As it's unsupervised, k-means is not reliable and convergence starts to become an issue as the number of sample increases [21]. In addition, k-means is very sensitive to initial seeding, as demonstrated by Fig. 3.



(a) Initial seeding (b) Clustering result

Fig. 3 An illustrative example showing sensitivity of K-means to initial seeding [21]

It is observed that by carefully selecting the initial seeds and supervising the clustering process, limitations of aforementioned algorithm can be overcome and satisfactory clustering results can be obtained. To solve the convergence problem, we tested the k-means++ [21] and the bisecting k-means methods [23]. The k-means++ algorithm is an extension of the k-means algorithm. The basic idea of this algorithm is that by carefully selecting the initial seeds (centroids), convergence can be sped up and better results can be obtained as measured by the Sum of Square Errors (SSE), which is the sum of the squared difference between each sample and the corresponding cluster's centroid. The steps used are listed below:

1) Step 1: Randomly select a load as the first centroid for load cluster S_1 from all loads, denoted as μ_1.
2) Step 2: Calculate the Euclidean distance *dist*(.) between the remaining loads and μ_1, denoted as (d_1, d_2, …d_i,…). Let $D(x_k)$ be the furthest distance between x_k and the existing centroids so that $D(x_k)$=max{d_1, d_2, …d_i,…}.
3) Step 3: For all the remaining (non-centroid) x's, select the one that has the largest $[D(x_k)]^2$ and make it a new centroid.
4) Step 4: Go back to step 2 if number of the selected centroids has not reached the number desired; otherwise go to step 5.
5) Step 5: Use the k-means algorithm to cluster the loads with the selected centroids as the initial seeds.

The bisecting k-means algorithm is a straightforward extension of the k-means approach. The basic idea of the bisecting k-means is to first split the entire system into two clusters using the k-means algorithm and then choose the worse cluster (in terms of the SSE) and split it into two new clusters. Repeat this process until the number of total clusters meets the target. Bisecting k-means guarantees the convergence of the clustering process since, in each step, the algorithm only splits the cluster into two.

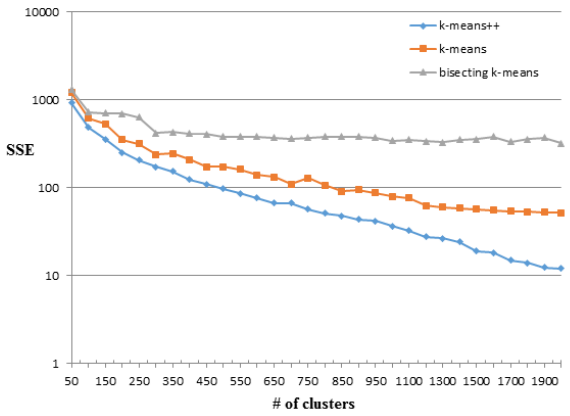


Fig. 4 Performance comparison between different load clustering algorithms

Performance of the three clustering algorithms has been tested numerically and their performance are compared as a function of number of clusters, as shown in Fig. 4. Note that the y-axis in this figure is on logarithm scale. The proposed algorithm based on k-means++ has the best performance in terms of SSE.

Assuming a total of N_c clusters are obtained with each cluster represented by S_t (t=1,2,…,N_c), the SSE is calculated as:

$$SSE = \sum_{t=1}^{N_c} \sum_{x_j^{(t)} \in S_t} \left\| x_j^{(t)} - \mu^{(t)} \right\| \quad (16)$$

where $x_j^{(t)}$ is location of load j which belongs to cluster S_t, and $\mu^{(t)}$ is centroid of cluster S_t.

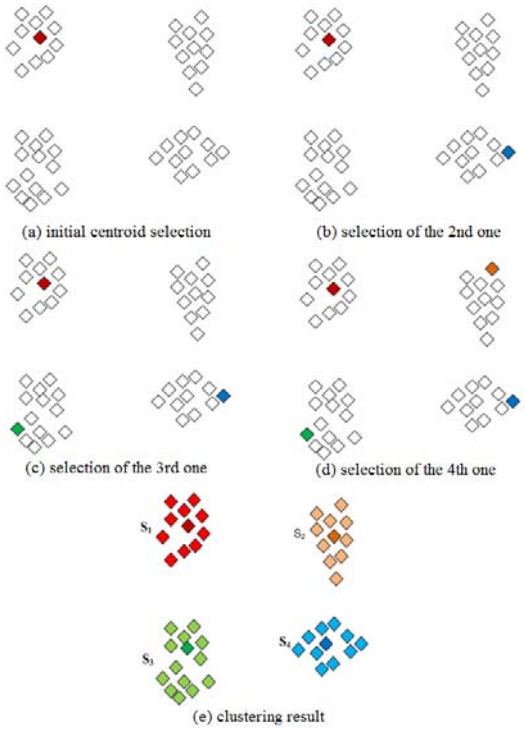


Fig. 5 Load clustering using the *k*-means++ algorithm [22]

B. Substation Sizing

In addition to supervising the centroid selection process, the clustering algorithm needs to ensure that the total capacity of a single load group cannot exceed capacity of the corresponding substation. Therefore, the load clustering approach needs to consider the total capacity of each substation. As the *k*-means++ clustering algorithm itself does not consider the capacity at each substation, we further improve the *k*-means++ algorithm to incorporate this factor. Assuming that the capacity of the substation(s) under planning is a series of discrete values, the maximum capacity is c_{max}, and the improved load clustering algorithm is below.

1) Regardless of the capacity constraints of substations, apply the *k*-mean++ clustering algorithm to divide *N* loads into *M* groups.

2) Traverse each group and select the substation capacity for each group of loads. The selected substation capacity should be one of the discrete values that is greater than the total load of each group. If the total load of each group is lower than c_{max}, the algorithm ends.

3) If the total load of a group exceeds the capacity of a substation, sort loads of this group according to their distances to the substation. Remove loads from this group one by one starting from the farthest load, until the sum of the remaining load falls below the capacity of the substation.

4) Cluster loads that are removed from step 3). These loads can only be allocated to clusters whose total load is well below the capacity of the corresponding substation.

The proposed clustering algorithm does not require uniform load distribution. When the load distribution is extremely uneven, the proposed load clustering algorithm can determine the substation capacity required for each load group according to its density without the need to specify the capacity of each substation in advance. This is also an advantage of the proposed algorithm as compared to existing approaches.

In practice, capacity of a substation is typically not chosen arbitrarily but instead from a lookup table. Given the total load of a group, b, determining the combination of substations to satisfy the load is a knapsack problem or rucksack problem, a problem in combinatorial optimization. Given the rated capacities of N substations and their construction costs (a_1, c_1), (a_2, c_2), …, (a_n, c_n), the objective is described below:

$$min(\sum_{i=1}^{n} c_i x_i) \quad (17)$$

$$\textit{subject to} \quad \sum_{i=1}^{n} a_i x_i \geq b$$

This is essentially a bounded knapsack problem with limited searching space. According to the capacity of each substation and the total load, the following relationship exists for the kth substation.

$$0 \leq x_k \leq \lceil b/a_k \rceil \quad (18)$$

and the upper limit for the search space is:

$$\prod_{i=1}^{n} (1 + \lceil b/a_i \rceil) \quad (19)$$

In general, this problem can be solved through mixed-integer linear programming (MILP) but its performance is not always satisfactory in terms of computational efficiency and memory cost. In DPP, this problem is actually converted into a 0/1 knapsack problem whose solution can be found in a recursive manner.

IV. Experimental Results

Two case studies are presented in this section to demonstrate the performance and effectiveness of the proposed approach. The first case study uses simulated data while the second one is based on real data collected for Gaoming District from Foshan Power Supply Company.

A. An Illustrative Example Using Simulated Data

A polygon area is created with five different shapes of convex and concave sub-polygons with a total of 200 loads, as shown in Fig. 6, where the red '+' signs show locations of loads.

Costs of land of the five sub-polygons are all different from each other with the following set of values: 10 million (m), 15 m, 8 m, 2 m, and 7m, in ascending order of land ID.

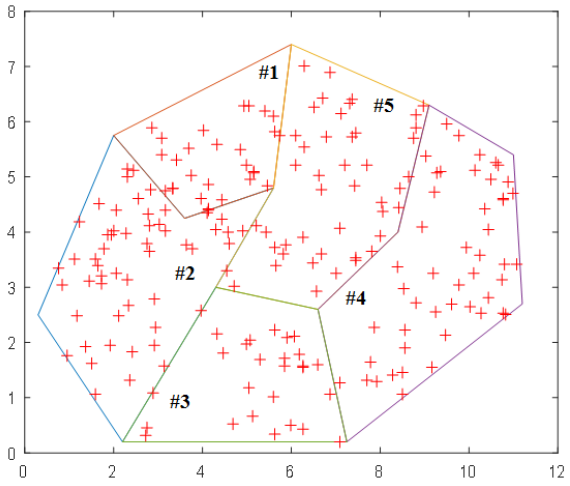


Fig. 6 Planning area and load distribution

Two case studies are conducted: one with GIS and land cost neglected and the other with both considered. Using the proposed approach, results obtained from the two case studies are shown in Fig. 7 and Fig. 8, respectively. As the figures show, three substations are planned and loads they supply are in the same color. These two figures show both the initial locations of substations, their trajectories (in each iteration) and their final locations. Darker marks show the final location of the three substations.

Basically it can be seen that locations of substations can be very different when geographic information and cost of land are considered. For example, the substation which supplies power to load in land #2 is originally located in the middle of land #2 while it moves to the boundary of land no. 2 and no. 5 when land prices are considered. With GIS and land price information included, the proposed approach significantly reduces the investment.

As the next step, both the construction cost of lines and substations are considered. Regarding the construction cost of lines, two types of line with four capacity numbers are considered as shown in Table I. Assuming a total of 5 substations are to be built, the results with and without cost of land considered are shown in Fig. 9. The trajectories of substations at each iteration are also shown in this figure to demonstrate evolution of algorithm.

TABLE I
TYPES OF LINE, CORRESPONDING CAPACITIES AND COSTS

Type of Line	Circuit	Capacity (MVA)	Construction Cost (Million Dollar)
I	Single-circuit	100	1.47
I	Double-circuit	200	1.76
II	Single-circuit	150	2.21
II	Double-circuit	300	2.64

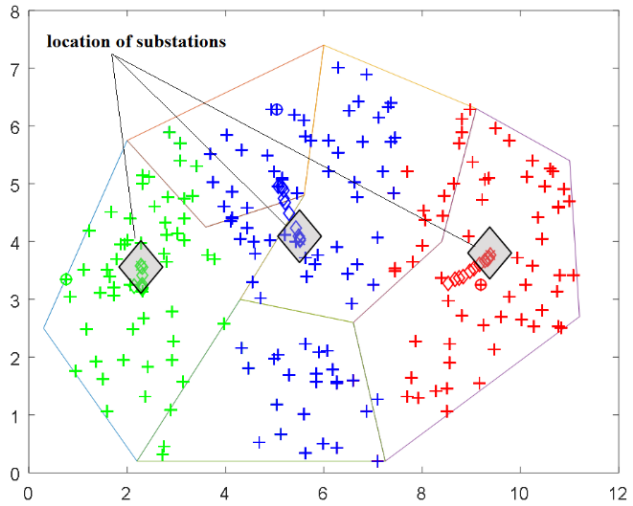


Fig. 7 Location of substations without considering GIS and land price

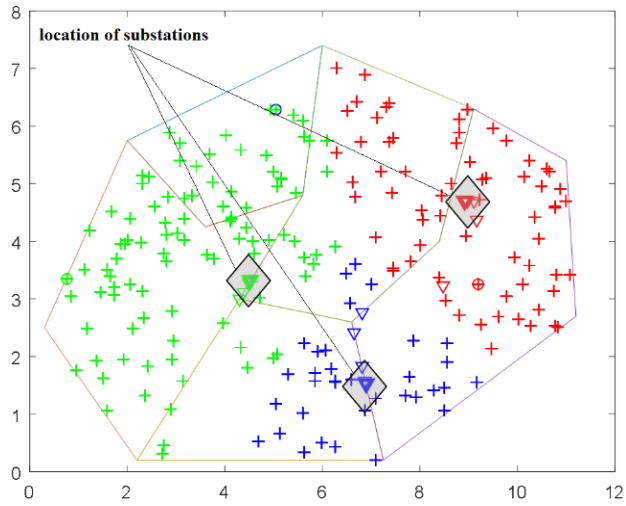


Fig. 8 Location of substations with GIS and land price considered

Due to high computational efficiency of the proposed approach, fast screening can be achieved with parallel computing to examine the total cost as a function of the number of substations. The average cost per year (for a total of 20 years) as a function of the number of substations is shown in Table II, from which it can be concluded that the optimal number of substations is 7, which gives the least cost of 11.85 million.

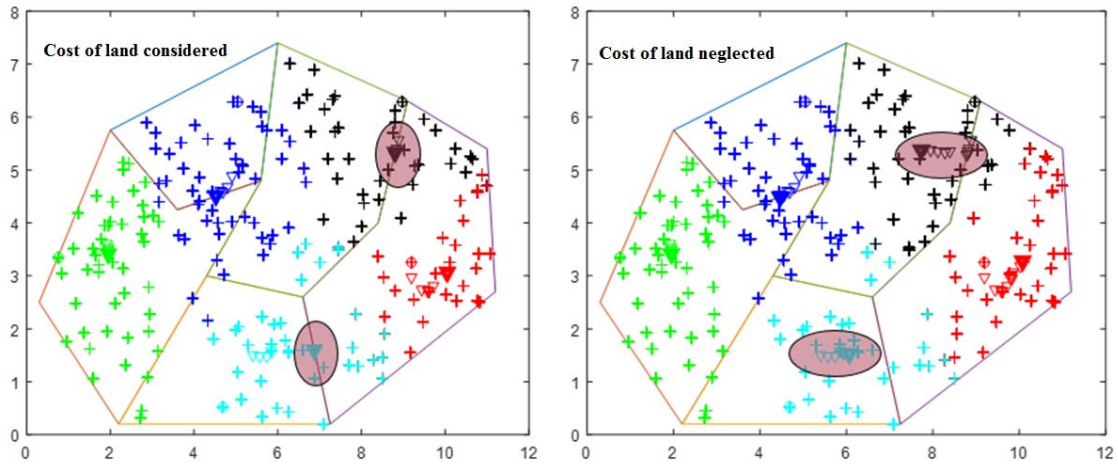


Fig. 9 Locations of substations with and without GIS and cost of land considered

TABLE II
TOTAL COST AS A FUNCTION OF NUMBER OF SUBSTATIONS

No. of Substations	Average Investment Cost per Year Over 20 Years (10^4 RMB)
1	1984.2530
2	1520.7004
3	1393.6341
4	1290.3858
5	1239.1312
6	1211.1400
7	1185.5609
8	1225.8762
9	1265.2456
10	1298.5525

B. *Results Using Real Data*

In the second case study, real data collected for Gaoming District of Foshan in Guangdong are used. Fig. 10 shows 305 loads, aggregated based on feeders, with X axis showing the latitude and Y axis showing the longitude. Loads shown in Fig.10 sums up to 414.11MW.

Five pieces of lands, each with a different cost, are identified as shown in Fig. 11. The cost of land, in ascending order of land IDs, are 10, 15, 8, 2, and 7 million RMB, respectively. Apply the proposed approach to co-optimize the substation locations and capacities for both the 35kV and 110kV substations via fast screening by varying the number of substations for each voltage level. The results are shown in Fig. 12, with the blue bar showing the cost for constructing 35kV substations and lines and the yellow bar showing the corresponding construction cost of the 110kV substations and lines. From Fig. 12 it is identified that the optimal number of 35kV substations is 14 while the optimal number of 110kV substations is 3. Optimal locations of the substations are shown in Fig. 13, where circles are the locations of 110kV substations and triangles are the locations of 35kV substations.

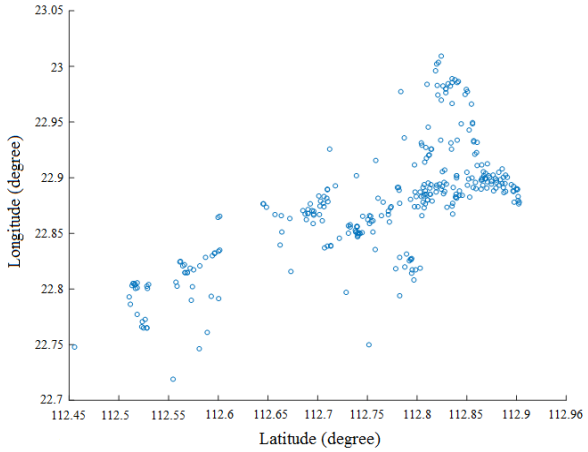


Fig. 10 Scatter plot for loads in Foshan Gaoming District

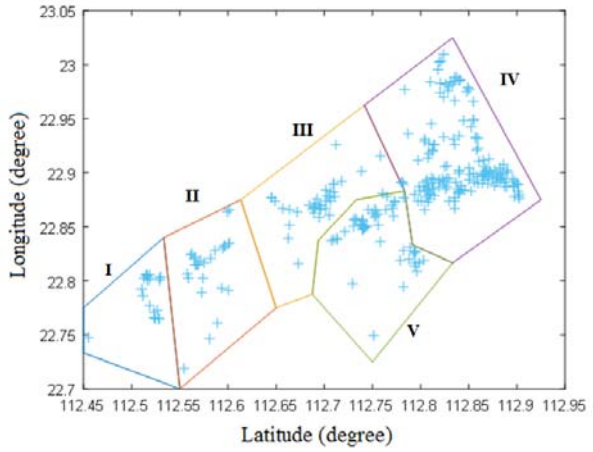


Fig. 11 Five pieces of land with different costs

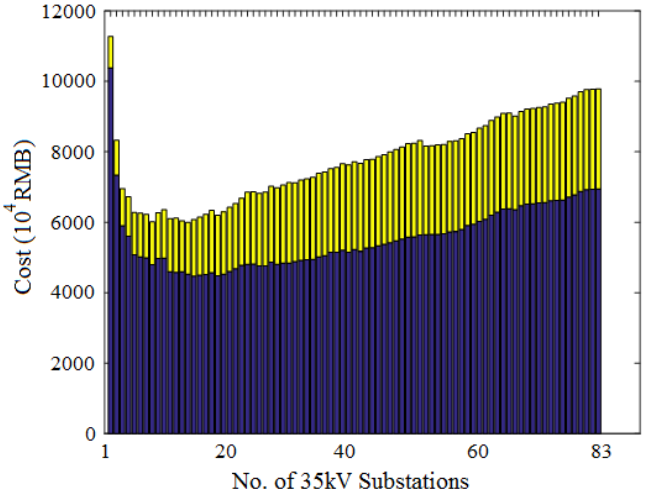


Fig. 12 Average yearly investment as a function of number of substations

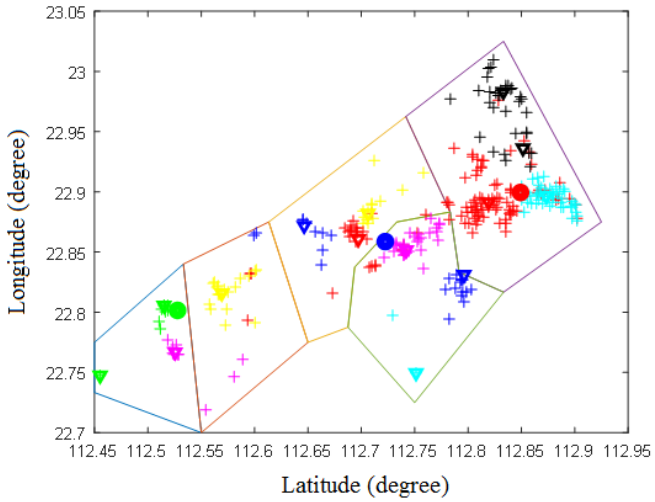


Fig. 13 Locations of substations

V. Conclusion

This paper presents a novel and highly efficient methodology for solving the substation siting and sizing problem based on geographic information and semi-supervised learning. The proposed approach can optimize the locations, capacities, power supply ranges of substations with minimum investment. Geographic information and cost of land are taken into consideration in DPP so that it is especially useful for highly developed urban areas. The proposed approach ~~are~~ is also very useful when it comes to cases when there are certain regions the planners prefer or want to avoid.

References

[1] X. Zhao, H. Zhou, etc., "On-Line PMU-Based Transmission Line Parameter Identification," *CSEE Journal of Power and Energy Systems*, vol. 1, no. 2, June 2015.

[2] S. Wang, Z. Lu, etc., "An Improved Substation Locating and Sizing Method Based on the Weighted Voronoi Diagram and the Transportation Model," *Journal of Applied Mathematics*, vol. 2014, pp. 1-9, May 2014.

[3] D. L. Wall, G. L. Thompson, etc., "An Optimization Model for Planning Radial Distribution Networks," *IEEE Trans. Power App. and Syst.*, vol. 98, no.3, pp. 1061–1068, 1979.

[4] M. Ponnavaikko, K. S. P. Rao, and S. S. Venkata, "Distribution System Planning through a Quadratic Mixed Inger Programming Approach," *IEEE Trans. Power Delivery*, vol. 2, no.4,pp. 1157–1163, 1987.

[5] H. L. Willis, H. Tram, etc., "Optimization Applications to Power Distribution," *IEEE Computer Applications in Power*, vol. 8, no. 4, pp. 12-17, Oct. 1995.

[6] H. L. Willis, H. Tram, etc., "Selecting and Applying Distribution Optimization Methods," *IEEE Computer Applications in Power*, vol. 9, no. 1, pp. 12-17, Jan. 1996.

[7] Y. Y. Hsu, and J. L. Chen, "Distribution Planning Using a Knowledge-based Expert System," *IEEE Trans. Power Delivery*, vol. 5, no. 3, pp. 1514-1519, July 1990.

[8] K. L. Lo, and I. Nashid, "Interactive Expert System for Optimal Design of Electricity Distribution Systems," *IEE Proc. Gen., Trans. and Distr.*, vol. 143, no. 2, pp. 151-156, Mar. 1996.

[9] D. E. Bouchard, M. M. A. Salama, etc., "Optimal Distribution Feeder Routing and Optimal Substation Sizing and Placement Using Evolutionary Strategies," *1994 Canadian Conference on Electrical and Computer Engineering*, Canada, Aug. 2002.

[10] M. Skok, D. Skrlec, and S. Krajcar, "Genetic Algorithm and GIS Enhancement Long Term Planning of Large Link Structured Distribution Systems," *2002 Large Engineering Systems Conference on Power Engineering*, Halifax, Canada, 2002.

[11] C.-M. Liu, R.-L. Kao, etc., "Solving Location Allocation Problems with Rectilinear Distances by Simulated Annealing," *Journal of the Operational Research Society*, vol. 45, no. 11, pp. 1304–1315, 1994.

[12] G. J. Chen, K. K. Li, and L. Wang, "Distribution System Planning by Tabu Search Approach," *Automation of Electric Power Systems*, vol. 25, no. 7, pp. 40–44, 2001.

[13] A. Navarro, and H. Rudnick, "Large-scale Distribution Planning-Part II: Macro-optimization with Voronoi's Diagram and Tabu Search," *IEEE Trans. Power Systems*, vol. 24, no. 2, pp. 752–758, 2009.

[14] T. H. M. El-Fouly, H. H. Zeineldin, etc., "A New Optimization Model for Distribution Substation Siting, Sizing, and Timing," *Inter. Journal of Electrical Power and Energy Systems*, vol. 30, no. 5, pp. 308–315, 2008.

[15] E. Miguez, E. Diaz-Dorado, and J. Cidras, "An Application of An Evolution Strategy in Power Distribution System Planning," *Proc. of the IEEE World Congress on Computational Intelligence, International Conference on Evolutionary Computation (ICEC'98)*, pp. 241–246, May 1998.

[16] M. R. Haghifam, and M. Shahabi, "Optimal Location and Sizing of HV/MV Substations in Uncertainty Load Environment Using Genetic Algorithm," *Electric Power Systems Research*, vol. 63, no. 1, pp. 37-50, 2002.

[17] A. Phayomhom, N. Rugthaicharoencheep, and S. Chaitusaney,, "GIS Application to Distribution Substation Planning in MEA's Power System," *12th Inter. Conf. on Electr. Engineering/Electronics, Computer, Telecommunications and Infor. Tech.*, Thailand, June 2015.

[18] Z. Liu, and J. Zhang, "Optimal Planning of Substation of Locating and Sizing Based on GIS and Adaptive Mutation PSO Algorithm," *International Conference on Power System Technology*, 2006.

[19] W. M. Lin, M. T. Tsay, and S. W. Wu, "Application of Geographic Information System for Substation and Feeder Planning," *International Journal of Electrical Power & Energy Systems*, vol. 18, no. 3, pp. 175-183, Mar. 1996.

[20] A. Minot, Y. M. Lu, and N. Li, "A Distributed Gauss-Newton Method for Power System State Estimation," *IEEE Transactions on Power Systems*, vol. 31, no. 5, Sept. 2016.

[21] D. Shi, and D. J. Tylavsky, "A Novel Bus-aggregation-based Structure–Preserving Power System Equivalent," *IEEE Trans. Power Systems*, vol. 30, no. 4, 2015.

[22] D. Shi, "Network Reduction for Engineering and Economic Analysis," Ph.D. Dissertation, Arizona State University, Tempe, AZ, USA, 2012.

[23] C. C. Aggarwal, S. C. Gates, and P. S. Yu, "On the Merits of Building Categorization Systems by Supervised Clustering," *Proc. 5th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, pp. 352–356, 1999.