

Research on Base Station Location Problem Solved Based on Simulated Annealing Algorithm

Lin Chen *

Pudong Huayao High School, Elpis, Shanghai, China

* Corresponding author Email: 3468316114@qq.com

Abstract. With the widespread adoption of 5G technology, the existing base stations, both in number and coverage, are insufficient to meet current societal needs. Therefore, establishing suitable new base stations within the current coverage area has become a significant challenge. To address this issue, this paper collects data on weak coverage areas and previously established base stations in those areas. The collected data undergoes a series of processing and analysis steps. Then, a suitable 0-1 programming mathematical model is established using methods for handling complex programming problems. Simulated annealing is then employed to solve the 0-1 programming model. Finally, further sensitivity analysis is conducted. The optimized solution yields a cost of 23.35 million yuan to establish new base stations in weak coverage areas, resulting in a total of 310 new base stations, including 225 macro base stations and 85 micro base stations. The sensitivity analysis in this paper effectively verifies the rationality of the 0-1 programming model and the accuracy of the simulated annealing algorithm. This study determines how to rationally establish new base stations in weak areas, which has important reference value for the future establishment of new base stations in weak areas.

Keywords: Establishment of a 0-1; SA; Sensitivity Analysis; Base Station Site Selection.

1. Introduction

To thoroughly implement the important directives of General Secretary Xi Jinping on accelerating the construction and development of 5G networks, and to fully advance the grand strategic deployment of empowering the digital transformation of all industries through 5G technology, a special plan to promote 5G nationwide and make 5G services available to the general public and various market entities has officially been placed on the national development agenda. This plan is not only an inevitable choice in line with the global trend of technological revolution and industrial transformation but also a key measure to strengthen the foundation of building a Digital China and enhance national core competitiveness. Its aim is to activate innovation vitality across all fields, and improve social productivity and public service levels through the widespread application of 5G technology.

However, in the process of mobile communication technology advancing rapidly towards high speed, low latency, and wide connectivity, new challenges have emerged with the comprehensive rollout and deep penetration of 5G technology. 5G technology itself features high frequency bands and large bandwidth, and compared with the previous 2G, 3G, and 4G networks, the bandwidth resources required for communication transmission have increased exponentially. At the same time, the propagation attenuation characteristics of high-frequency signals are more pronounced, leading to a significant reduction in the effective coverage area of traditional base stations — under the same conditions, the coverage radius of a 5G base station is only one-third to one-half that of a 4G base station. This change directly brings about a core contradiction in network construction: to achieve a level of coverage and service quality comparable to 4G networks a few years ago, the number of 5G base stations that need to be deployed must increase substantially, consequently triggering a series of related issues such as construction costs, operational and maintenance costs, and land resource occupation.

the current complex problem can be simplified as follows:

(1) A region in reality can be mapped to a number of adjacent squares. The center point of the square is taken as the representative of the square. That is, given any region, it can be divided into a finite number of points.

(2) Each mapped coordinate point has certain attribute values, including but not limited to: coordinates, traffic volume, and whether it is a weak coverage point.

(3) In order to minimize costs and interference between devices, the distance between any two base stations should meet a specific threshold requirement.

(4) Assuming the base station coverage area is d and the coordinates of the mapping point base station are $P_0(x_0, y_0)$, then any condition $\|P - P_0\| \leq d$ are all within the coverage area of the base station if $\|P - P_0\| \leq d$, otherwise they are considered to be outside the coverage area of the base station.

The planning model in this paper is established based on the following background: site planning for the total service coverage of weak coverage points within a 2500×2500 area, while also considering the cost of the constructed base stations, and ensuring that the distance between each base station is greater than 10. Specifically, macro base stations have a coverage range of 30 and a cost of 10; micro base stations have a coverage range of 10 and a cost of 1. Before establishing the model, the relationship between existing base stations and weak coverage points needs to be analyzed. Plotting reveals that many weak coverage points lie within a 10-radius area of existing base stations, and most of these points have high service volumes.[1]-[5] Therefore, base station planning not only includes the planning of new base stations but also the planning of existing base station types and costs.

2. Data Preprocessing

Before building the model, we first drew a traffic distribution map of the original weak coverage points. The result is shown in the figure below:

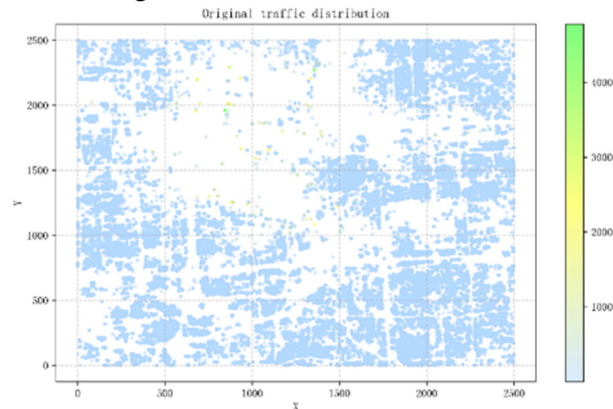


Figure 1. Distribution of traffic volume in the original weak coverage areas

As can be seen from the above figure, the original weak coverage point traffic distribution is not very uniform, and most of the traffic is below 10,000.

Table 1. Distribution of traffic volume between original weak coverage areas

range	points	percentage	interval	percentage
<1	56925	31.14%	18642	0.26%
1-10	64487	35.28%	259891	3.68%
10-30	28072	15.36%	500445	7.09%
30-50	10268	5.62%	398594	5.65%
50-100	10450	5.72%	735032	10.42%
100-500	10828	5.92%	2202007	31.21%
>500	0776	0.97%	2941620	41.69%

Next, this paper conducts statistical analysis on the data of weak coverage points, and the distribution of traffic volume intervals of weak coverage points is shown in the table 1.

The first column 'range' in Table 1 is an abbreviation for 'business volume range'. The second column 'points' in Table 1 is an abbreviation for 'number of weak coverage points'. The fourth column 'interval' in Table 1 is an abbreviation for 'total business volume within the interval'.

Table 1 shows that there are 56,925 weak coverage points with a traffic volume less than 1, accounting for 31.14% of all weak coverage points. However, these 31.14% of weak coverage points only account for 0.26% of the total traffic volume. Similarly, weak coverage points with a traffic volume between 1 and 10 account for 35.28%, but their traffic volume only accounts for 3.68% of the total traffic volume. From the above comparison, it is easy to conclude that over 96% of the traffic volume is concentrated in the first 1/3 of weak coverage points, while the other 2/3 of weak coverage points provide only 4% of the traffic volume. To simplify the computation of the subsequent model, we removed weak coverage points with a traffic volume less than 10, leaving 96% of the total traffic volume. The distribution of traffic volume for the remaining weak coverage points is shown in Figure 2.

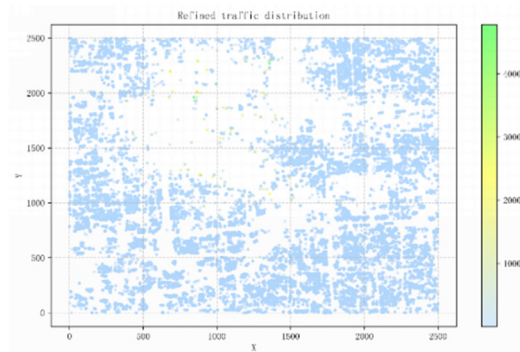


Figure 2. Distribution of traffic in remaining weak coverage areas

As shown in the figure above, the distribution of traffic volume in the remaining weak coverage points is more even than before, but the overall traffic volume is still below 10,000.

Next, the following figure shows the current website's traffic distribution:

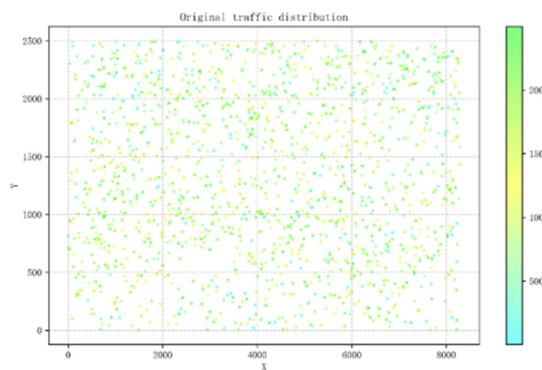


Figure 3. Distribution of Business Volume at Current Website Address

As shown in the figure above, the current website's business volume is relatively evenly distributed.

Next, the current website address traffic volume distribution data is statistically analyzed and presented in Table 2, which shows the current website address traffic volume distribution intervals.

The first column 'range' in Table 2 is an abbreviation for 'business volume range'. The second column 'points' in Table 1 is an abbreviation for 'number of weak coverage points'. The fourth column 'interval' in Table 1 is an abbreviation for 'total business volume within the interval'.

As shown in the table above, the number of coverage points and the total business volume within the range are mostly in the range of 1000-2000, and there are very few ranges below 100.

Table 2. shows the distribution of business volume at current website addresses and the distribution of business volume between different intervals.

range	points	percentage	interval	percentage
0-10	6	0.41%	27	0.001%
10-50	14	0.95%	480	0.02%
50-100	28	1.90%	2003	0.10%
100-500	172	11.67%	50787	2.48%
500-1000	219	14.86%	163830	8.00%
1000-2000	700	47.49%	1079287	52.70%
>2000	335	22.73%	751627	36.70%

3. Establishment of a 0-1 Mathematical Programming Model

0-1 programming is a special form of integer programming. In this type of programming, the decision variables only take the values 0 or 1, hence they are called 0-1 variables or binary variables, because any non-negative integer can be represented using several 0-1 variables in binary notation. 0-1 variables can quantitatively describe the logical relationships, sequential relationships, and mutually exclusive constraints among discrete variables that reflect phenomena such as on and off, take or abandon, have or not have. Therefore, 0-1 programming is very suitable for describing and solving various problems of interest, such as network design, factory location, production scheduling, traveling salesman problems, knapsack problems, personnel scheduling, code selection, and reliability.

The establishment of new base stations requires satisfying the following conditions: the distance threshold between new base stations is 10; the planned new base stations can cover 90% of the total traffic in weakly covered areas; this paper needs to explore how to minimize the total cost of establishing base stations based on the above assumptions. Therefore, a reasonable and effective 0-1 mathematical programming model needs to be established.[6][10] The specific modeling steps are as follows:

(1) Objective function

The objective function of this paper is to minimize the total cost of building new base stations, as follows:

$$\min f = \sum_{i=1}^n (10v_i a_i + v_i b_i) \quad (1)$$

Where $v_i, a_i, b_i = 0$ or 1 , the cost of a macro base station is 10, and the cost of a micro base station is 1. $v_i = 0$ means no new base station will be built at candidate base station i , and $v_i = 1$ means a new base station will be built; $a_i = 0$ means no macro base station will be built at candidate base station i , and $a_i = 1$ means a macro base station will be built; $b_i = 0$ means no micro base station will be built at candidate base station i , and $b_i = 1$ means a micro base station will be built. $v_i a_i$ represents the case where candidate base station i is selected as a new base station and a macro base station is chosen. Candidate base station i is established as a macro base station if and only if $v_i a_i = 1$; similarly, candidate base station i is established as a micro base station if $v_i b_i = 1$. The meanings of v_i, a_i , and b_i , and their interrelationships, are as follows:

$$\begin{cases} v_i = \begin{cases} 0 & \text{(Candidate base station } i \text{ was selected)} \\ 1 & \text{(not selected.)} \end{cases} \\ a_i = \begin{cases} 0 & \text{(Establish macro base stations)} \\ 1 & \text{(Not establish macro base stations)} \end{cases} \\ b_i = \begin{cases} 0 & \text{(Establish micro base stations)} \\ 1 & \text{(Not Establish micro base stations)} \end{cases} \end{cases} \quad (2)$$

(2) Decision variables: Coordinates of the new base station (x, y) .

(3) Constraints:

1) The formula for selecting either a macro base station or a micro base station is as follows:

$$a_i + b_i = 1 \quad (3)$$

The above formula represents a candidate base station i selected as a new base station site, which can be chosen from two types of base stations: macro base station and micro base station.

2) The distance threshold between newly built base stations is 10, and its calculation formula is as follows:

$$\sqrt{(x - x_i)^2 + (y - y_i)^2} > 10 \quad (4)$$

The above formula states that the distance between two newly built base stations should be greater than 10.

3) The distance between a weak coverage point covered by a macro base station and the macro base station is no greater than 30, and the calculation formula is as follows:

$$\sqrt{(x - x_h)^2 + (y - y_h)^2} \leq 30 \quad (5)$$

4) The distance between a weak coverage point covered by a micro base station and the micro base station is no greater than 10, and the calculation formula is as follows:

$$\sqrt{(x - x_h)^2 + (y - y_h)^2} \leq 10 \quad (6)$$

The above formula must satisfy $a_i + b_i = 1$, meaning all weak coverage points that can be covered by micro base stations.

5) The total traffic volume that the newly built base station can cover in weakly covered areas is greater than 90% of the total traffic volume. The mathematical expression for this is as follows:

$$\sum w_h \geq 6350607 \quad (7)$$

The above formula represents the total traffic volume of weak coverage points that can be covered by the newly built base station, which must be greater than or equal to 90% of the total traffic volume of 7,056,230, i.e., 6,350,607; where H is the set of weak coverage points that can be covered by the newly built base station, w_h and is the traffic volume of the h th weak coverage point ($h \in H$).

$$\begin{aligned} \min f &= \sum_{i=1}^n (10v_i a_i + v_i b_i) \\ \text{s. t. } &\begin{cases} a_i + b_i = 1 \\ \sqrt{(x - x_i)^2 + (y - y_i)^2} > 10 \\ \sqrt{(x - x_h)^2 + (y - y_h)^2} \leq 30 \\ \sqrt{(x - x_h)^2 + (y - y_h)^2} \leq 10 \\ \sum w_h \geq 6350607 \end{cases} \end{aligned} \quad (8)$$

$$\begin{aligned} \text{s. t. } &\begin{cases} \sqrt{(x - x_i)^2 + (y - y_i)^2} > 10 \\ \sqrt{(x - x_h)^2 + (y - y_h)^2} \leq 30 \\ \sqrt{(x - x_h)^2 + (y - y_h)^2} \leq 10 \\ \sum w_h \geq 6350607 \end{cases} \end{aligned} \quad (9)$$

4. Solving Using the Simulated Annealing Algorithm

Genetic Algorithm (GA) was first proposed by John Holland in the United States in the 1970s. The algorithm was designed based on the evolutionary laws of organisms in nature and is a computational model that simulates the biological evolution process of natural selection and genetics according to Darwin's theory of evolution. It is a method for searching for optimal solutions by simulating the natural evolution process. Through mathematical methods and computer simulation, the algorithm transforms the problem-solving process into processes similar to biological evolution, such as crossover and mutation of chromosomes and genes. When solving relatively complex combinatorial optimization problems, compared to some conventional optimization algorithms, this algorithm can generally obtain better optimization results more quickly. Genetic algorithms have been widely applied in the fields of combinatorial optimization, machine learning, signal processing, adaptive control, and artificial life.

The Simulated Annealing Algorithm (SA) is a global optimization algorithm based on simulating the annealing process of metals.[11][14] Its function is to search for a globally optimal or near-optimal solution among multiple candidate solutions. The idea behind the simulated annealing algorithm comes from the behavior of atomic thermal vibrations during the annealing process of solid materials. During annealing, the material gradually reaches its lowest energy point as the temperature decreases,

making the crystal structure more stable. Analogously, in problem-solving, the simulated annealing algorithm controls the state transition probabilities in the search space, sometimes accepting suboptimal solutions, thereby searching around the globally optimal solution. The algorithm's process is as follows:

1) Encoding strategy selection

To simplify encoding and decoding operations, this paper adopts binary encoding with a precision of 4 decimal places. The number of binary strings for a single variable is calculated using the following formula:

$$2^{k-1} \leq (U - L) * 10^4 \leq 2^k \quad (10)$$

Here, k represents the length of the binary number. In this algorithm, L=0 and U=20 are selected, resulting in K=11, which means that the solution to the problem is represented by a binary number of length 11.

2) Population initialization:

In practice, the group size can be set to n, depending on the situation, and each individual will be selected using $L + random(k) * (U - L)$ this random number. We randomly select n data points from U, where U is the maximum value and L is the minimum value. In this algorithm, cost-plus pricing and replenishment quantity are set as variables to be determined, and the remaining parameters are set as shown in the table below:

Table 3. Simulated Annealing Algorithm Parameter Settings

temperature	rate	iterations	threshold
1200	0.992	8000	0.00001

The first column 'range' in Table 3 is an abbreviation for 'initial temperature'. The second column 'rate' in Table 3 is an abbreviation for 'cooling rate'. The fourth column 'threshold' in Table 3 is an abbreviation for 'minimum temperature threshold'.

3) Determine the fitness function

The fitness function for this problem is to maximize the number of mountains visited and minimize the total cost of entrance fees and transportation. The fitness function is the objective function of the optimization model, as follows:

$$\begin{cases} \text{Min } cost = cost_{travel} + cost_{tickets} \\ \text{Max } Citys_{mntns} = \sum_{j=1}^n C_{mntns_{ij}} \end{cases} \quad (11)$$

4) Select Operation

This paper uses the fitness ratio method, and the formula is as follows:

$$f_i = \frac{k}{F_i}, P_i = \frac{f_i}{\Pi}, \Pi = \sum_{j=1}^N f_j \quad (12)$$

Wherein, the fitness value of individual i is F, k is a coefficient, and N is the total number of individuals in the population.

5) Cross operations

The solution reveals differences in distance between different base stations, suggesting a potentially high correlation between their coverage areas. Therefore, the crossover operation in the genetic algorithm must also consider this correlation. The crossover operation is as follows:

Based on the crossover rate set during initialization, select a specific chromosome and list all crossover combinations.

The correlations among subspecies are ranked, and the chromosome crossover coverage rate is set to 0.7-1. The top 70% -100 % combinations from this ranking are selected for compilation. The j-th genes of the i-th individual are then selected for mutation, using the following formula:

$$a_{ij} = \begin{cases} a_{ij} + (a_{ij} - a_{min})f(g) & r \geq 0.5 \\ a_{ij} + (a_{max} - a_{ij})f(g) & r < 0.5 \end{cases} \quad (13)$$

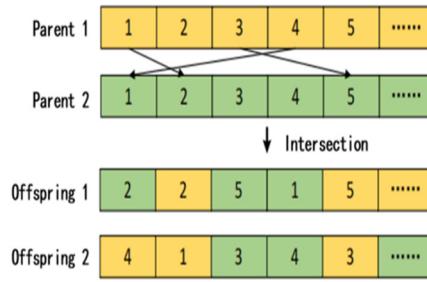


Figure 4. SA Schematic diagram of algorithm crossover

Where b is any number between $[0, 1]$.

6) Mutation operation

The j -th gene of the i -th individual a_{ij} is selected for mutation operation, and the formula is:

$$a_{ij} = \begin{cases} a_{ij} + (a_{ij} - a_{min})f(g) & r \geq 0.5 \\ a_{ij} + (a_{max} - a_{ij})f(g) & r < 0.5 \end{cases} \quad (14)$$

Wherein, a_{ij} the upper limit of the number of genes is a_{max} , a_{ij} the lower limit of the number of genes is a_{min} , $f(g) = r_2(10g/G_{max})$, r_2 is an arbitrary number, g is the current iteration number, G_{max} is the maximum number of evolutions, and r is an arbitrary number between $[0, 1]$. the total cost of building a new base station using the simulated annealing algorithm to solve the 0-1 planning model is shown in the figure below:

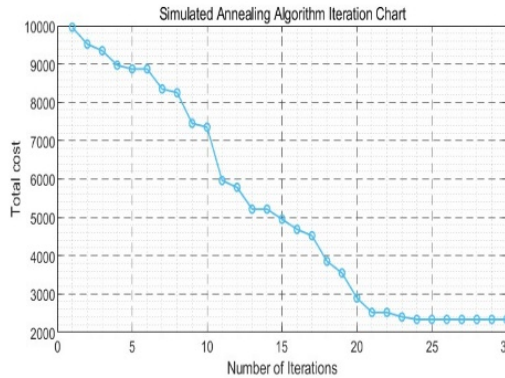


Figure 5. Simulated Annealing Algorithm Iteration Chart

The results obtained from solving the model are shown in the table below:

Table 4. Solution results of simulated annealing algorithm

total cost	new base	macro base	micro base
2335	310	225	85

The second column 'new base' in Table 4 is an abbreviation for 'total number of new base stations'. The third column 'macro base' in Table 4 is an abbreviation for 'number of macro base stations'. The fourth column 'micro base' in Table 4 is an abbreviation for 'number of micro base stations'.

The results of the new base station are visualized in this article, as shown in the figure below

As can be seen from the figure above, the number of macro base stations is much greater than the number of micro base stations.

the new base station and the weak coverage area is shown in the following figure 7:

As shown in the figure above, the new base station can effectively cover areas with weak coverage, indicating that the 0-1 planning model and simulated annealing algorithm established in this paper are very reasonable.

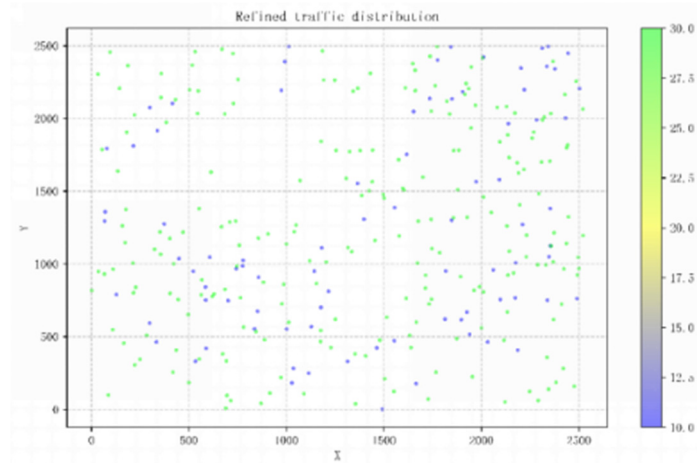


Figure 6. Scatter plot of the distribution of new base stations

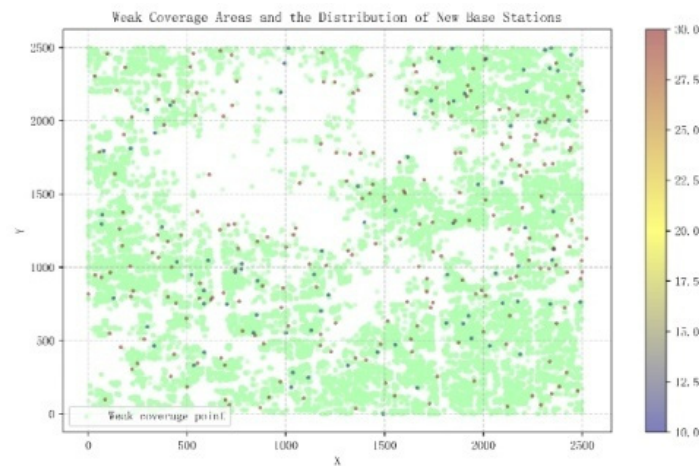


Figure 7. Distribution map of new base stations and areas with weak coverage.

5. Sensitivity Analysis

Sensitivity analysis aims to understand the extent to which model outputs respond to input parameters or variables. It can help us identify key parameters of the model, understand how the model behaves, and assess the robustness of the model. Sensitivity analysis is usually carried out by calculating the rate of change of the model output with respect to the input parameters, in order to evaluate the model's sensitivity to different parameters.

In simulated annealing algorithms, the cooling rate is adjusted to change the temperature, thus affecting the search process. More specifically, the behavior of annealing in the search space can be adjusted by changing the cooling rate. Increasing the cooling rate accelerates annealing's ability to explore new areas, while decreasing the cooling rate makes annealing more focused on the region surrounding the current optimal solution. Of course, this method of adjusting the cooling rate requires reasonable parameter settings to ensure that the algorithm can fully explore and utilize the search space.

Therefore, this paper applies a 5% perturbation to the original simulated annealing algorithm based on the self-cooling rate of a single temperature.[15][18] The simulated annealing continuously searches the solution space and selects the objective function value with the minimum fitness as the iteration value in each round. After 50 iterations, the convergence value at the current cooling rate is found, and the sensitivity analysis is plotted as shown in the figure below:

After repeatedly modifying the cooling rate parameter in the simulated annealing algorithm, the algorithm consistently converged to the same value. Therefore, the algorithm is reasonable and

effective for solving the aforementioned 0-1 programming model, and the results are highly convincing.

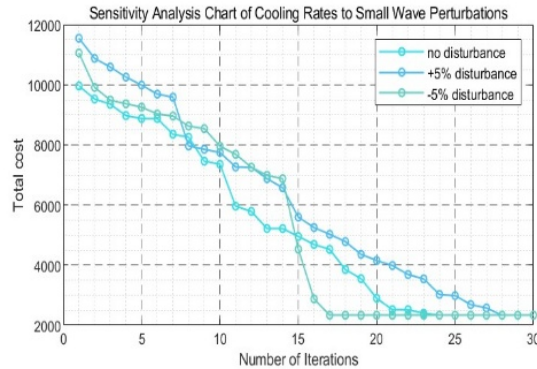


Figure 8. Sensitivity analysis of temperature cooling rate

6. Conclusion

The 0-1 programming mathematical model and simulated annealing algorithm established in this paper provide a valuable reference for establishing new base stations in areas with weak base station coverage. The model effectively addresses the challenge of establishing new base stations in such areas. Furthermore, to further verify the model's accuracy, a sensitivity analysis was conducted, further validating the rationality of the established 0-1 programming model and the simulated annealing algorithm.

In the future, new base stations will be built in areas with different coverage types, requiring the integration of more data and the establishment of more complex mathematical programming models and solution methods. However, this article still has great initial reference value.

References

- [1] Zhao Zhenyu, Lin Shan, Ma Qianxin. Integrated energy system shared energy storage planning based on split-blob bar optimization [J]. *Electric Power Construction*, 2025, 46(11): 58-70.
- [2] Gao Xuejun, Yang Lu. Regional sensitivity study of bogie critical speed [J]. *Journal of Hunan University (Natural Science Edition)*, 2025, 52(10): 50-57.
- [3] Zhou Xiaodong, Sun Shuai, Du Longkun, et al. Sensitivity analysis of mid-infrared single-photon correlation imaging (invited) [J]. *Acta Optica Sinica*, 2025, 45(20): 377-389.
- [4] Zhu Yalan. Sensitivity and specificity analysis of cardiac color Doppler ultrasound in the diagnosis of coronary heart disease complicated with heart failure [J]. *Practical Medical Imaging Journal*, 2025, 26(05):393-397.
- [5] Huang Fengshou, Li Jiacheng, Sun Xing, et al. Lightweight improvement design of bus frame structure based on sensitivity analysis [J]. *Bus Technology and Research*, 2025, 47(05):40-46.
- [6] Xie Yunquan, Zhu Weixing, Gu Min, et al. Sensitivity Analysis Method for Military Software Capabilities Based on Neural Networks [J/OL]. *Command Control and Simulation*, 1-12 [2025-11-13].
- [7] Li Minghao, Niu Hao, Fan Jiayi, et al. Sensitivity Analysis and Optimization Design of Coal Mining Machine Spiral Drum Based on Discrete Element Method [J]. *Journal of Henan Polytechnic University (Natural Science Edition)*, 2025, 44(06):165-171.
- [8] Li Hongwei, Ying Changjie, Guan Guan. [9] Li Ming, Liu Huali, Xing Ying, et al. BP neural network blasting fly rock distance prediction model based on simulated annealing optimization [J/OL]. *Command, Control and Simulation*, 1-11 [2025-11-13].
- [9] Chen Jie, Zhuang Bailiang, Yang Baohua, et al. NSGA-SA: Simulated Annealing Enhanced Multi-Objective Optimization Algorithm and Its Application in Heat Treatment [J]. *Software Guide*, 2025, 24(09): 87-98.

- [10] Zhou Xingjian, Zhang Zuhan, Ai Zhen. Research on Site Selection and Layout of Public Charging Stations for New Energy Vehicles Based on Bi-Objective 0-1 Programming [J]. *Logistics Technology*, 2025, 44(07): 60-72.
- [11] Li Congqing. Research on Vehicle Pickup and Delivery Transportation Path Based on Simulated Annealing Algorithm [J]. *Journal of Zhengzhou Railway Vocational and Technical College*, 2025, 37(01): 49-53.
- [12] Wang Yun. Research on Last-Line Vehicle-Mounted UAV Logistics Delivery Route Planning Based on Simulated Annealing Algorithm [J]. *Automation and Instrumentation*, 2025, (02): 247-251.
- [13] Zhuang Yuxi. Research on 0-1 Quadratic Programming Problem Solving Algorithm Based on Graph Neural Network [D]. Shanghai University, 2023.
- [14] Liao Yifei. Research on Ship Domestic Sewage Disposal Strategy Based on 0-1 Programming [J]. *Ship Materials and Market*, 2023, 31(04): 107-111.
- [15] He Kun, Ren Shuo, Guo Zijie, et al. Local Dynamic Programming Algorithm for Solving Complete 0-1 Knapsack Problem Based on Greedy Backtracking [J]. *Journal of Huazhong University of Science and Technology (Natural Science Edition)*, 2024, 52(02): 16-21.
- [16] Liu Shangyi, Wu Tao. 0-1 Linear Programming Model and Application Based on Mixed Variables [J]. *Science and Technology Innovation*, 2023, (01): 92-95.
- [17] Wang Maoping, Pan Dazhi. An improved dynamic programming algorithm for solving the set-valued discount $\{0-1\}$ knapsack problem [J]. *Computer Applications and Software*, 2022, 39(09):274-277.
- [18] Qian Haochen, Jiang Xintong, Li Xuanhao, et al. Landscape planning and design of resettlement community based on POE theory - a case study of Nanjing Zijin Moxiangyuan [J]. *Urban Architecture*, 2022, 19(13):15-18.