



Optimization of 5G base station coverage based on self-adaptive mutation genetic algorithm

Jianpo Li, Jinjian Pang^{*}, Xiaojuan Fan

School of Computer, Northeast Electric Power University, China



ARTICLE INFO

Keywords:

Least square method
Signal propagation model correction
Self-adaptive mutation genetic algorithm (AMGA)
Base station coverage optimization

ABSTRACT

In communication network planning, a rational base station layout plays a crucial role in improving communication speed, ensuring service quality, and reducing investment costs. To address this, the article calibrated the urban microcell (UMa) signal propagation model using the least squares method, based on road test data collected from three distinct environments: dense urban areas, general urban areas, and suburbs. With the calibrated model, a detailed link budget analysis was performed on the planning area, calculating the maximum coverage radius required for a single base station to meet communication demands, and accordingly determining the number of base stations needed. Subsequently, this article proposed the Adaptive Mutation Genetic Algorithm (AMGA) and formulated a mathematical model for optimizing 5G base station coverage to improve the base station layout. Simulation experiments were conducted in three different scenarios, and the results indicate that the proposed AMGA algorithm effectively enhances base station coverage while reducing construction costs, thoroughly demonstrating the value of base station layout optimization in practical applications.

Funding

This study has received support from the China Natural Science Foundation. (61,501,106)

1. Introduction

With the rapid development of 5G mobile communication technology, the number of 5G users has significantly increased, leading to a corresponding expansion in network capacity [1]. To meet the growing user demand, researchers have begun to focus on improving the throughput of base stations (e.g. Refs. [2,3]). While enhancing the performance of individual base stations is crucial, the synergistic effect among all base stations is equally indispensable for further enhancing the overall performance of 5G communication systems. Therefore, addressing the challenges of 5G wireless network planning has become increasingly important [4]. The key lies in reducing the construction costs for network operators while ensuring user communication quality and network coverage [5].

Since 5G networks utilize higher frequencies and larger bandwidths compared to 4G, more base stations need to be deployed within the same area to achieve comprehensive coverage, undoubtedly increasing

operators' costs (e.g. Refs. [6,7]). However, the selection of 5G base station locations is also influenced by local terrain and population distribution, and obstacles such as streets, buildings, and trees can significantly impact signal propagation. Improper site selection may require increasing power output or optimizing signal detection technology to mitigate the impact on user communication quality [8]. Furthermore, the layout optimization process must comprehensively consider multiple aspects, balancing the relationship between construction costs, coverage range, and communication quality. Focusing solely on a single factor may lead to resource waste and unnecessary high costs [9]. Additionally, determining the appropriate number of base stations is crucial. Too many base stations can lead to overlapping coverage, cross-regional coverage, and interference between base stations, while too few can result in weak coverage and poor communication quality [10].

To address these issues, this article proposes a mathematical model for optimizing 5G base station coverage and introduces an innovative adaptive mutation genetic algorithm (AMGA) to tackle the challenges of 5G base station layout in different scenarios. Through carefully planned base station layouts, it can not only expand coverage but also improve resource utilization, reduce construction costs, and enhance user communication quality.

* Corresponding author.

E-mail addresses: jianpoli@163.com (J. Li), 2202200382@neepu.edu.cn (J. Pang), 3207672681@qq.com (X. Fan).

1.1. Relate work

The purpose of optimizing the layout of base stations is to reduce the construction cost of base stations and improve the communication quality for users. A majority of researchers have conducted extensive research and argumentation on this issue. To solve these problems, Chen et al. [11] employed the k-means clustering algorithm to cluster vulnerability data. They established a single-objective nonlinear programming model with the objective of minimizing the total construction cost of new base stations, while considering the minimum distance between adjacent base stations and the minimum coverage of communication traffic as constraint conditions. This model aims to optimize the layout of macro base stations and micro base stations. Liang et al. [12] evaluated the comprehensive performance of 5G base stations from aspects such as financial performance, operational performance, social impact, and environmental impact, to identify issues such as poor user experience and insufficient coverage in base station construction. They proposed a hybrid MCDM model based on the Bayesian best-worst method and DQ-GRA technology. The results indicate that signal coverage area and per capita investment cost are the most important indicators that affect the overall performance of 5G base stations. Wang et al. [13] proposed a mixed integer nonlinear optimization model that minimizes cost and maximizes coverage, and solved the location and positioning problems of 5G macro and micro base stations. The coverage-based positioning method proposed can achieve minimum deployment cost and maximum signal coverage in a balanced manner. Liu et al. [14] developed a random geometry-based analysis framework with various characteristics, including non-uniform base station deployment, biased user associations, flexible operation of macro and micro base stations, and resource allocation with a minimum rate guarantee. Dai et al. [15] proposed to extract the main features that determine the signal propagation strength from a large amount of received signal strength data, and then these features are fed into machine learning for modeling the wireless signal propagation, furthermore, to solve the non-convex optimization problem of base station deployment, multi-objective genetic algorithms and greedy algorithms are used to optimize the location of the base station deployments and the operational parameters, so that the number of base stations with coverage guarantee is kept to a minimum.

In the pursuit of optimizing base station layouts, numerous researchers have also accorded significant attention to the interplay between energy consumption and cost. (e.g. Refs. [16,17]). Dong et al. [18] proposed a minimum-cost millimeter wave base station deployment framework in Manhattan geometry, where the constraint function expression is composed of various factors such as obstacle size, base station transmission power, and base station density. They also designed a low-complexity algorithm to effectively find the optimal balance between base station density and transmission power, while ensuring network connectivity while generating the minimum base station deployment cost. Jee et al. [19] proposed that the use of incremental signaling and the combination of far-user and near-user (as well as successive interference cancelation) can provide significant gains in throughput and energy efficiency compared to non-incremental amplify-and-forward, coordinated direct and relay transmission, non-orthogonal multiple access and its relay orthogonal multiple access counterpart.

In recent years, with the aim of enhancing users' communication experience, researchers have conducted thorough and rigorous research into path loss prediction and channel capacity analysis. Nguyen et al. [20] proposed a feed-forward deep neural network (DNN) model for simultaneous prediction of path loss at 13 different frequencies from 0.8 GHz to 70 GHz in urban and suburban non-line-of-sight (NLOS) scenarios, and researchers also investigated various possible hyper-parameter values to find the optimal set of hyper-parameters to obtain the optimal structure of the proposed DNN model. Kola et al. [21] analyzed the capacity of 4×4 Multiple-Input Multiple-Output (MIMO)

channels at 28 GHz and 39 GHz for three regions (urban microcell, urban macrocell, and rural macrocell environments) using the NYUSIM simulator. The environment with line-of-sight (LOS) links has a very high channel capacity for urban areas and a very low channel capacity for rural areas. In NLOS, the urban macrocell (UMa) has the highest channel capacity while the urban microcell (UMi) has the lowest channel capacity and the climate parameter that has the greatest impact on the capacity of the MIMO channel is the rain rate. Raimundo et al. [22] studied the propagation of millimeter wave channels in various indoor scenarios using a broadband MIMO frequency-modulated continuous wave (FMCW) channel sounder. Dual polarization measurements in the frequency bands centered at 54 and 70 GHz were performed to study the temporal and angular small-scale statistics of the indoor radio channel, and the parameters of the logarithmic distance and proximity path loss models were estimated from the data.

1.2. Motivation and contribution

This article proposes an optimization approach for the deployment of 5G base stations. Initially, a continuous wave (CW) test is conducted in the planned area to acquire drive test data. These data, along with the least squares method, are utilized to calibrate the signal propagation model. Subsequently, the Link budget formula is employed to determine the half-path coverage of the base stations and the required number of base stations. Based on this, a mathematical model for optimizing 5G base station coverage and an AMGA are presented, serving as a reference for 5G network planners. The key contributions of this article are summarized as follows.

- Research on data preprocessing methods.

Utilizing the propagation loss of CW test signals, the collected data undergoes data filtering, discrete processing, and geographic averaging to yield accurate drive test data for the calibration of the signal propagation model.

- Research on the signal propagation model Correction.

Through simulation experiments, this article analyzed the path loss trends of different models in the planning area to obtain a signal propagation model that aligns with the environmental conditions of the planning region. The signal propagation model was then rectified using the least squares method based on road test data. Finally, the coverage radius and the number of base stations required in the planning area were determined through the link budget formula.

- Research on base station coverage optimization algorithm.

By analyzing the criterion for optimizing base station coverage, we propose a mathematical model for optimizing 5G base station coverage along with the AMGA. The problem of optimizing 5G base station coverage is thereby reformulated as the task of finding the optimal solution, and the optimal base station layout is achieved through the iterative process of the AMGA.

1.3. Structure of this article

The remainder of this article is organized as follows. In Section 2, we studied the base station coverage method, mainly comparing two signal propagation models. In Section 3, we mainly discussed the research on data preprocessing methods. In Section 4, we corrected the signal propagation model. In Section 5, we introduced the optimization of base station coverage based on the AMGA. In Section 6, we conducted relevant experiments and analyzed them. In Section 7, we summarized the article.

2. Research on base station coverage methods

Base station coverage optimization refers to the optimization of the number and placement of base stations to ensure comprehensive coverage of the wireless network, thereby enhancing the communication quality for users.

2.1. Problems

The planning of base station locations is undertaken to satisfy the needs of users within their respective coverage areas. To enhance the communication quality for users, it is essential to strategically lay out base stations within the planned territory [23]. In this process, In this process, several issues need to be addressed, including eliminating coverage vulnerability, mitigating the impact of areas with weak coverage, and preventing Cross-region and overlapping coverage.

Coverage vulnerability occurs when the signal strength is significantly reduced or no signal can be detected, preventing terminals from accessing the network [24].

Weak coverage arises when base stations require a large coverage area and there is excessive spacing between base stations, leading to weakened signals in border areas [25].

Cross-region coverage occurs when the coverage area of a base station extends beyond its planned range, forming a discontinuous dominant region within the coverage area of other base stations [26].

Overlapping coverage refers to the presence of significant overlap between multiple cells, where the reference signal receiving power (RSRP) may be strong but the signal-to-interference plus noise ratio (SINR) is degraded, or equipment experiences frequent handovers between multiple cells, resulting in a poor user experience [27].

2.2. Wireless signal propagation model

The signal propagation model is a crucial step in the layout and location selection of base stations, as it directly impacts the accuracy of base station coverage optimization. Among the commonly used signal propagation models are the SPM signal propagation model and the UMa signal propagation model, etc [28].

2.2.1. Standard propagation model (SPM)

The SPM model, also known as the Standard Signal Propagation Model, has a general frequency range of 150–3500 MHz and a communication distance of 1–2000 m. The model combines a large number of field test results and is suitable for a variety of wireless environments. Its model is shown in (1).

$$PL = K_1 + K_2 \cdot \log_{10}(d) + K_3 \cdot \log_{10}(h_{BS}) + K_4 \cdot Diff + K_5 \cdot \log_{10}(h_{BS}) \cdot \log_{10}(d) + K_6 \cdot h_{UT} + K_{clutter} \cdot f(clutter) \quad (1)$$

where PL represents path loss, K_1 represents a frequency-dependent offset constant of 17.4, K_2 indicates that the correction factor related to distance is 44.9, K_3 represents correction factor of 5.83 related to antenna height at the base station, K_4 indicates that the correction factor associated with diffraction loss is 1, K_5 denotes the correction factor related to the height and distance of the base station antenna is -6.55, K_6 indicates that the correction factor related to the height of the mobile terminal antenna is 0, $K_{clutter}$ denotes the correction factor for ground attenuation is 1, and d represents distance between the base station and the mobile terminal, h_{BS} represents the effective height of the base station antenna, $Diff$ represents the loss caused by diffraction on the blocking path, h_{UT} represents the effective height of the mobile terminal antenna, and $f(clutter)$ indicates landscape-weighted average depletion.

A simulation experiment was conducted on the SPM, and the parameters for the 5G mobile communication system operating in the 3500 MHz frequency band are defined in Table 1.

The SPM signal propagation model is simulated and analyzed using

Table 1

Definition of system parameters.

Frequency	Base station antenna height	Receiver antenna height
3500 MHz	29 m	1.5 m

MATLAB software in three different environments: dense urban areas, general urban areas, and suburbs. The results are shown in Fig. 1.

From Fig. 1, it can be concluded that when the antenna height parameters of the base station and receiver remain unchanged, the path loss increases with the complexity of the terrain environment. When the antenna height is the same, there are varying degrees of obstruction to the signal in three scenarios: dense urban areas, general urban areas, and suburban areas. The more dense the environment, the greater the path loss. As the distance between the base station and the receiving end gradually increases, the path loss also gradually increases.

2.2.2. UMa signal propagation mode

The UMa model is a new signal propagation model defined by the 3GPP organization, specifically designed for 5G high-speed and high-frequency trends. The universal frequency of this model is 0.5–100 GHz, and the effective distance of signal transmission is 10–5000 m. This signal propagation model possesses the capability to transmit signals at high velocities, with a maximum moving speed of 500 km/h (e.g. Refs. [29,30]). The model propagates in the LOS scenario as shown in (2), (3), and (4).

$$PL_{UMa-LOS} = \begin{cases} PL_1 & 10m < d_{2D} \leq d_{BP} \\ PL_2 & d_{BP} < d_{2D} \leq 5km \end{cases} \quad (2)$$

$$PL_1 = 28 + 22 \log_{10}\{d_{3D}\} + 20 \log_{10}\{f_c\} \quad (3)$$

$$PL_2 = 28 + 40 \log_{10}\{d_{3D}\} + 20 \log_{10}\{f_c\} - 9 \log_{10}(\{d'_{BP}\})^2 + ((\{h_{BS}\} - \{h_{UT}\})^2) \quad (4)$$

where $PL_{UMa-LOS}$ represents path loss under line-of-sight propagation. d_{2D} represents the horizontal distance between the receiving end and the transmitting end. d'_{BP} represents the breakpoint distance set by the signal propagation model. d_{3D} represents the linear distance between the receiving end and the transmitting end. f_c indicates the signal frequency used by the base station. h_{BS} represents the effective height of the base station antenna. h_{UT} represents the effective height of the mobile terminal antenna.

NLOS propagation conditions, the propagation model is shown in (5) and (6).

$$PL_{UMa-NLOS} = \max(PL_{UMa-LOS}, PL'_{UMa-NLOS}) \quad 10m \leq d_{2D} \leq 5km \quad (5)$$

$$PL'_{UMa-NLOS} = 13.54 + 39.08 \log_{10}\{d_{3D}\} + 20 \log_{10}\{f_c\} - 0.6(\{h_{UT}\} - 1.5) \quad (6)$$

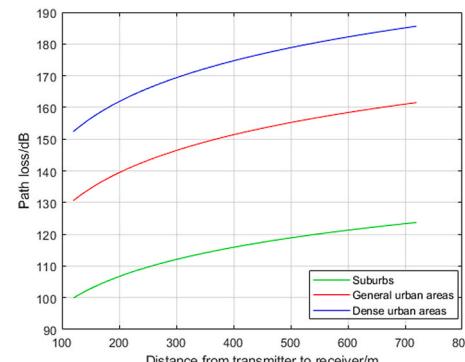


Fig. 1. SPM path loss in different environments.

where $PL_{UMa-NLOS}$ represents the path loss under NLOS propagation.

In the wireless signal transmission process, the NLOS UMa signal propagation model is used due to the influence of various objects such as occlusion in the planning area. By combining and converting (5) and (6), the generalized formula of the UMa signal propagation model can be obtained, as shown in (7).

$$PL = K_1 + K_2 \cdot \log_{10}\{d_{3D}\} + K_3 \cdot \log_{10}\{f_c\} - 0.6(\{h_{UT} - 1.5\}) \quad (7)$$

where K_1 represents the factor that needs to be corrected. K_2 represents the correction factor related to distance. K_3 represents the correction factor related to frequency.

In the process of wireless signal transmission, due to various factors such as object occlusion within the planned area, a NLOS UMa signal propagation model is adopted, as shown in formula (5). UMa signal propagation model was simulated, and the system parameter definition of this model is the same as in Table 1. The results are shown in Fig. 2.

From Fig. 2, the trend of path loss is similar to Fig. 1. Under identical parametric conditions, due to the sparse distribution of terrain features in suburban areas, low building heights, and low population density, the path loss is relatively lower compared to the other two types of regions. Moreover, as the distance between the base station and the receiver gradually increases, the path loss in different environments also gradually increases.

However, the SPM model is primarily designed for frequencies below 3.5 GHz and cannot satisfy the requirements of 5G communications, which operate in both the mid-low bands below 6 GHz and the high bands above 24 GHz. While the SPM model offers reliable propagation predictions in general scenarios, its accuracy is limited when applied to specific urban macrocell environments, particularly those involving high-speed mobile terminal devices. In contrast, the UMa model is tailored specifically to more accurately simulate the propagation of wireless signals in urban macrocell environments [31]. Furthermore, as depicted in Figs. 1 and 2, under the same environmental scenarios and parameters, the predicted path loss of the UMa model consistently remains lower than that of the SPM model. The purpose of adjusting the signal propagation model is to achieve a model that better reflects the actual environmental path loss. Therefore, this article will employ the UMa signal propagation model for model experimentation.

3. CW data testing and processing

Before correcting the signal propagation model, it is necessary to preprocess the collected data to obtain a more accurate signal propagation model.

3.1. Data acquisition

The data collection method is based on Li's Law. During signal propagation, when the test path uses 40 wavelengths, the number of

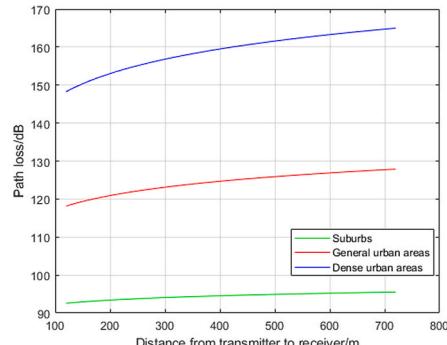


Fig. 2. UMa path loss in different environments.

signal reception strengths should be between 36 and 50, which can eliminate the impact of various fading and improve the accuracy of subsequent base station coverage optimization.

The frequency of the test signal in this article is 3500 MHz, with a wavelength of 0.09 m. According to Li's Law, the speed of collecting data at this time is between 0.072 m/s and 0.1 m/s, and because the speed of the test vehicle is about 28 km/h, the average time for collecting a single data is about 6 ms.

3.2. Road test data processing

Due to the non-standard testing environment or process, the data obtained by the CW testing method may be inaccurate. Therefore, it is necessary to preprocess the road test data [32].

- Data Filtering

During the data collection process, there may be some non-compliant road test data that needs to be filtered out.

- Data Discrete Processing

Since the speed of accurate positioning provided by the vehicle's GPS receiver is significantly slower than the data acquisition rate of the receiver, this can result in multiple identical data records being recorded for each test point within the planning area. Consequently, it is necessary to perform discrete processing on the collected data to ensure that the signal data is evenly distributed across the distance.

- Data Geographic Average

The purpose of geographical averaging is to eliminate fast fading and retain slow fading. Therefore, it is necessary to distribute the test path within the test area by 6 m per segment, average the road test data within this distance, and select a point within this distance as the location of the average data value.

The West Campus of Northeast Power University has been selected as one of the test sites for base station layout in general urban areas. Its main buildings, consisting mainly of playgrounds and teaching buildings with 4–6 floors, are relatively sparsely distributed. The red TX represents the base station transmitting location, while the blue RX indicates the receiving route. The overview of the study area is shown in Fig. 3.

The road test equipment is a HUAWEI mate30pro cell phone and Cellular Pro software. The test frequency is 3500 MHz and the test steps are as follows.

Step 1. Prepare geographic image data.



Fig. 3. Planning area study map.

Step 2. Obtain engineering data from relevant software, including basic information about the base station, latitude and longitude of the base station, antenna parameters, and other data.

Step 3. Install the receiving antenna on the roof of the testing vehicle and connect the testing antenna to the computer. And open the road test software for testing.

Step 4. Record the signal reception intensity, longitude and latitude, correlation distance, and other data mainly measured during CW testing.

In general, a base station adopts directional antennas, and a base station contains three sectors [33], with each sector corresponding to an antenna. According to theory, the angle of the antenna in each cell is 120° , which can achieve full coverage of the base station. However, in practical environments, due to factors such as user needs, the radiation angle of the base station antenna in each cell is different.

The relevant base station information of the testing site is shown in [Table 2](#).

After the parameter setting is completed, the CW test method is used to measure the road test data, and MATLAB is used to survey the signal strength distribution of the test data. The results are shown in [Fig. 4](#).

Upon observation of the data, it is found that there exist a significant amount of duplicate data within the test dataset, with extremely low signal intensity readings within a certain distance. Additionally, the antenna heights of some cell base stations are recorded as 0 or missing, and there are errors in the input of antenna-related parameters. Consequently, the following processing steps are required for the road test data.

Step 1: When there are multiple values at a certain position in the road test data, the average value is taken for calculation.

Step 2: Delete the data that does not match the base station used for the road data.

Step 3: Due to the impact of base station coverage or repeated coverage speed, data that is too close or too far away from the base station is deleted.

Step 4: Delete values with signal strength data greater than -110 dBm. In this case, it is generally considered that the base station has not covered this area.

After preprocessing the road test data, the data obtained from the CW test is converted to .txt format. Some of the road test data are shown in [Table 3](#).

The preprocessed data is represented using MATLAB as shown in [Fig. 5](#).

From [Fig. 5](#), it can be seen that as the distance from the base station to the receiving end increases, the overall signal strength of the receiving end also gradually weakens. Within 100–300 m, as the distance from the base station to the receiving end increases, the signal reception strength gradually weakens. However, within 300–400 m, due to being located in an open area with no buildings or trees blocking it, the signal strength has improved. Within 400–675 m, as the distance from the base station to the receiving end increases, the signal reception strength gradually weakens, and the receiver signal strength has reached around -110 dBm. At this point, stop testing.

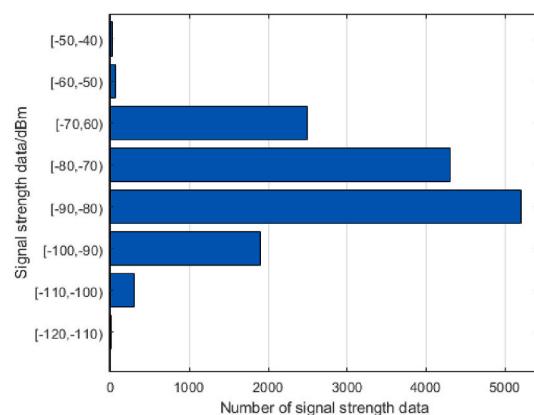


Fig. 4. Signal strength data distribution diagram.

Table 3
Partial test data.

Actual distance/ m	Receiving intensity/ dB	Actual distance/ m	Receiving intensity/ dB
163.73	-53	319.58	-69
171.05	-51	347.87	-72
180.97	-57	397.95	-81
182.97	-58	420.12	-86
225.05	-65	479.14	-89
233.47	-68	539.88	-89
252.21	-71	594.94	-92
275.32	-77	610.09	-94
301.01	-80	674.47	-106

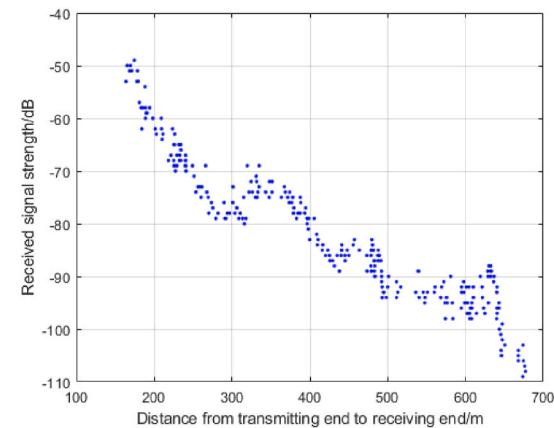


Fig. 5. Processed drive test data.

Then, the same method was used to measure the road survey data of a dense urban area and suburb in Jilin City, with an area of 25 square kilometers. The data obtained provide a theoretical basis for the correction of the subsequent signal propagation model. The method is the same as the above operation, and will not be repeated here.

4. Signal propagation model correction

Due to the complexity and diversity of the actual signal propagation environment, existing wireless signal propagation models cannot accurately reflect the signal propagation characteristics of the actual environment. Therefore, in the process of coverage optimization for base stations, it is necessary to correct the signal propagation model to make it more accurate [34].

Table 2
Relevant information of the test site.

Base Station	LAT and LON	Cell	Type	Azimuth	Transmit power
BS-1	126°29'46.75"	BS-1-1	macro	40	46
	43°49'20.06"	BS-1-2	macro	80	44
		BS-1-3	macro	270	48
BS-2	126°30'28.28"	BS-2-1	macro	60	45
	43°49'22.67"	BS-2-2	macro	120	40
		BS-2-3	macro	260	43
BS-3	126°30'20.88"	BS-3-1	macro	120	45
	43°49'28.27"	BS-3-2	macro	230	45
		BS-3-3	macro	290	45

4.1. Signal propagation model correction algorithm

The correction of signal propagation models is a process of using road test data to achieve model correction based on relevant algorithms, which mainly includes: setting the initial values of parameters to be corrected in the model and using the obtained path loss as the predicted value. Import the road test data into the model to obtain the path loss value as the measured value. Make a difference between the two values and iterate repeatedly until the average prediction error and mean square deviation of the two values meet the standard requirements. The corrected signal propagation model can better reflect the actual signal environment.

When correcting the signal propagation model, this article uses the least square method to iterate the data [35]. Its essence is the process of minimizing the sum of squared errors by subtracting the predicted and actual values of the model. Assuming that the initial values of the model parameters are substituted into the model and satisfy the linear regression equation between the dependent variable y_j and the independent variable $x_{1j}, x_{2j}, \dots, x_{nj}$, as shown in (8).

$$y_j = a_0 + a_1 x_{1j} + a_2 x_{2j} + \dots + a_n x_{nj} \quad (8)$$

where y_j represents the predicted path loss from the transmitting end to the terminal signal, a_1, a_2, \dots, a_n represents the initial coefficients K in the signal propagation model, x_1, x_2, \dots, x_n represents the known parameters of the base station (such as power, antenna height, azimuth, etc.).

Assuming that the linear regression equation between the dependent variable y and the independent variable x_1, x_2, \dots, x_n is satisfied after substituting the road test data into the model, as shown in (9).

$$\bar{y} = a + K_1 x_1 + \dots + K_n x_n \quad (9)$$

where \bar{y} represents the actual path loss of the transmitter to terminal signal, a, K_1, \dots, K_n represents the corrected model parameters obtained using the least squares method.

The purpose of using the least squares method to calibrate the signal propagation model in this project is to minimize the sum of the squared deviations of the predicted path loss value y_j and the actual path loss value \bar{y} , which satisfies the formula shown in (10).

$$\Delta = \sum_{j=1}^n (y_j - \bar{y})^2 = \sum_{j=1}^n (y_j - a - K_1 x_{1j} - \dots - K_n x_{nj})^2 = 0 \quad (10)$$

according to the Extreme value theorem of Differential calculus functions, if you want to make Δ reach the minimum value, you should derive the coefficients a and K_i . The formulas are shown in (11) and (12).

$$\frac{\partial \Delta}{\partial a} = -2 \sum_{j=1}^n (y_j - a - b_1 x_{1j} - \dots - b_n x_{nj}) = 0 \quad (11)$$

$$\frac{\partial \Delta}{\partial K_i} = -2 \sum_{j=1}^n (y_j - a - b_1 x_{1j} - \dots - b_n x_{nj}) x_{ij} = 0 \quad (i=1, 2, \dots, n) \quad (12)$$

Convert equations (11) and (12) into formulas as shown in (13) and (14).

$$\sum_{j=1}^n y_j = na + b_1 \sum_{j=1}^n x_{1j} + \dots + b_n \sum_{j=1}^n x_{nj} \quad (13)$$

$$\sum_{j=1}^n x_{ij} y_j = a \sum_{j=1}^n x_j + b_1 \sum_{j=1}^n x_{1j}^2 + \dots + b_n \sum_{j=1}^n x_{nj}^2 \quad (14)$$

Therefore, from formula (10) to formula (14), obtain the regression coefficient a , as shown in (15).

$$a = \bar{y} - b_1 \bar{x}_1 - b_2 \bar{x}_2 - \dots - b_n \bar{x}_n \quad (15)$$

after obtaining the coefficient a , use formula (9) to obtain the coefficient

K_1, \dots, K_n , which is the corrected value of the signal propagation model.

Formula (7) has already described the path loss of the UMa model in the planning area. The working frequency within the planned area is known to be 3.5GHZ. Based on the experience of base station deployment engineering, K_3 is a correction factor related to frequency, so determining the frequency can determine the value of K_3 . Therefore, $K_3 = 20$. Now, only correcting K_1 and K_2 is needed to obtain the corrected model formula.

4.2. Criteria for correcting signal propagation models

The purpose of signal propagation model correction is to obtain a propagation model that conforms to the actual environment. In actual engineering, Root mean squared error (RMSE) and mean prediction error (MAPE) are used to test whether the corrected signal propagation model conforms to the standard. For the error of the correction result, if the corrected Mean squared error is less than and the average prediction error is about equal to, then the corrected propagation model is in line with the standard.

Root mean squared error (RMSE) is the difference between the actual path loss value and the predicted loss value, which can reflect the change of the difference during the correction process. The formula is shown in (16).

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n}} \quad (16)$$

Mean Absolute Percentage Error (MAPE) refers to the average difference between the actual path loss value and the predicted loss value, which can reflect the size of the actual prediction error. The formula is shown in (17).

$$MAPE = \frac{1}{n} \sum_{j=1}^n \left| \frac{y_j - \bar{y}}{y_j} \right| \times 100\% \quad (17)$$

The signal propagation model is corrected by the least square method, and the root mean squared error and average prediction error are checked to see if they meet the correction standard. If both are true, the correction is over. If either of the two does not meet the standard, continue to iterate for correction. Through continuous iteration, a model that conforms to the planned regional environment can be obtained and the model correction is completed.

4.3. Practical application of model correction

The road test data and UMa signal propagation model are calibrated by the least square method. The specific steps are as follows: the factor of correction parameters in the UMa model is set to the default value in each environment, and the predicted loss value is calculated. The actual loss value is obtained by using the road measurement data. The difference between the predicted value and the measured value is iterated continuously until the average prediction error and mean square error reach the minimum and meet the standard, then the corrected model is considered to be in line with the actual environment.

After the above process of correcting the model, the values of mean square error and average prediction error before and after correction under different environments are obtained, as shown in Tables 4 and 5.

According to engineering experience, if the corrected mean square error is less than 8dB and the average prediction error is approximately

Table 4
Comparison of mean square error before and after model correction.

	Dense urban areas	General urban areas	Suburbs
Before correction	19.7101	2.2954	17.2592
After correction	4.8040	3.8945	3.9250

Table 5

Comparison of average prediction error before and after model correction.

	Dense urban areas	General urban areas	Suburbs
Before correction	0.1413	0.1522	0.1271
After correction	0.0365	0.0284	0.0285

0dB, it indicates that the corrected propagation model meets the standard [36]. At this time, the obtained model correction coefficients are shown in Table 6.

The fitting diagram of the road test data obtained simultaneously is shown in Fig. 6.

After completing the calibration of the signal propagation model, the path loss values of the model before and after calibration are compared as shown in Fig. 7.

From Fig. 7, it can be seen that the corrected path loss value of the signal propagation model is more in line with the actual road test data, indicating that the results of this correction are in line with the characteristics of the actual environment.

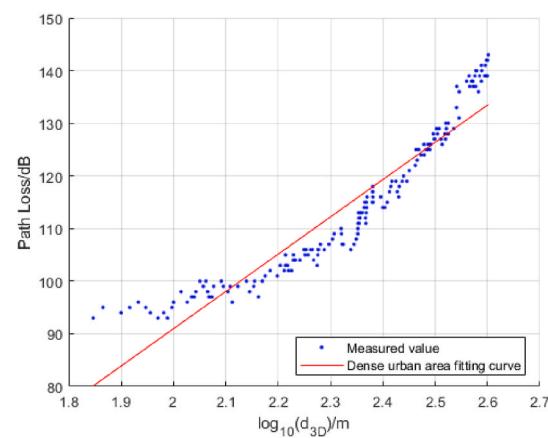
4.4. Correction of signal propagation model

The link budget is an essential prerequisite for wireless network coverage planning, and its main objective is to determine the maximum allowable path loss in a given scenario to satisfy coverage demands, taking into account factors such as margin, loss, gain, and other computational components throughout the signal transmission process from the transmitter to the receiver. Through the propagation model in line with the current environment, the coverage radius of a single macro base station is calculated. The coverage capacity of the wireless network system is then initially estimated by reducing coverage in blind and weak coverage areas based on the requirements of the regional station planning 5G network deployment. It is then deployed according to the 5G site planning requirements to reduce the coverage blind zones and weak coverage zones, to initially estimate the coverage capacity of the wireless network system [37]. The link budget is shown in (18).

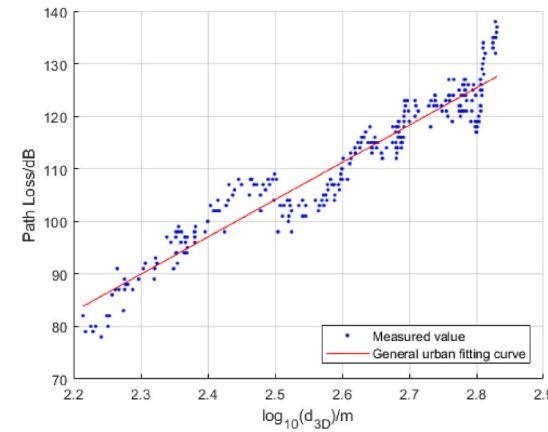
$$MAPL = P_{\max} + G_{TX} - L_f - L_p - L_b - M_I - M_f - S_{BS} + G_{RX} \quad (18)$$

where, $MAPL$ represents the maximum path loss. P_{\max} denotes the maximum transmit power of the transmitter, which is generally taken as 53 dBm at the base station and 23 dBm at the user side. G_{TX} denotes the antenna gain at the transmitter side. G_{RX} denotes the antenna gain at the receiving end. L_f represents the feeder line loss, and the general value range is 0 dB to 1 dB, L_p indicates the penetration loss, generally refers to the building penetration loss, dense urban areas take the value of the range of 26 dB – 32 dB, general urban areas take the value of the range of 18 dB – 23 dB, suburban areas take the value of the range of 7 dB – 10 dB. L_b represents human body loss, the value is 3 dB, M_I denotes the interference margin, which takes the value of 6 dB for dense urban areas, 4 dB for general urban areas, and 2.8 dB for suburban areas. M_f denotes the shadow fading margin, which takes the value of 10.3 dB for dense urban areas, 8.5 dB for general urban areas, and 4.2 dB for suburban areas. S_{BS} represents receiver sensitivity, generally, it is the minimum signal strength data that the receiver can receive and ensure normal operation.

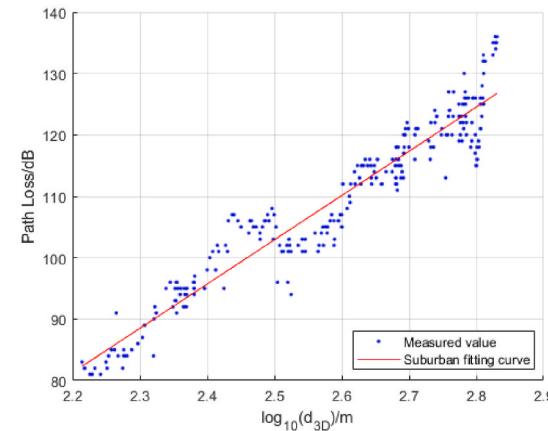
The base station coverage radius and the required number of base stations can be calculated based on the link budget formula and the corrected signal propagation model formula. The approach to deter-



(a)Fitting diagram of road survey data in densely populated urban areas



(b)Fitting diagram of general urban road survey data



(c)Fitting diagram of suburban road test data

Fig. 6. Fit map of road test data in different environments.

mining the required range of base stations in the planning area is as follows: The minimum number of base stations is calculated by positioning the diameter of the base station's coverage circle within the planning area, while the maximum number is determined using formula (19), which is presented below.

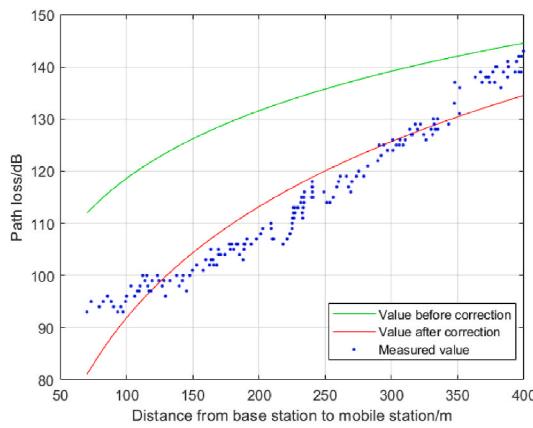
$$N = X/S \quad (19)$$

where S represents the coverage area of the cell base station, and the base station coverage is circular. X represents the total area of the planned area.

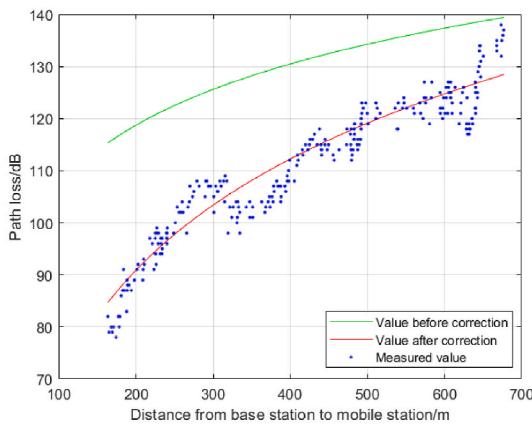
Table 6

Correction factors of models in different environments.

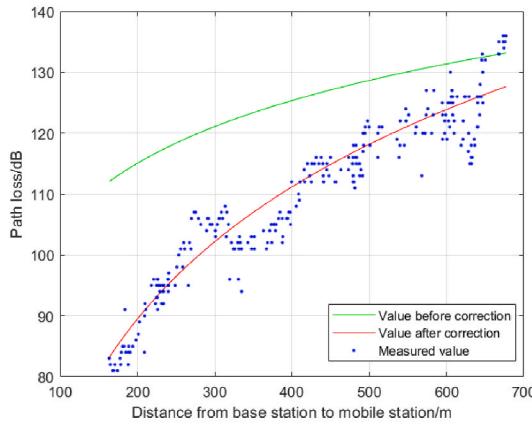
Coefficient	Dense urban areas	General urban areas	Suburbs
K_1	-60.7	-83.34	-87.45
K_2	70.84	70.98	72.15



(a) Comparison of loss before and after correction in dense urban areas



(b) Comparison of loss before and after correction in general urban areas



(c) Comparison of loss before and after correction in suburbs

Fig. 7. Comparison of loss before and after correction in different environments.

- Coverage Radius and Required Number of Base Stations in Dense Urban Areas

Using the link budget formula, the maximum allowable path loss in dense urban areas is $131.97dB$. Combining the corrected signal propagation model, the base station coverage radius $R1 = 0.368km$ is obtained. Given that the total area of dense urban areas is approximately 25 square kilometers, the number of base stations required ranges from 48 to 62.

- Coverage Radius and Required Number of Base Stations in General Urban Areas

Using the link budget formula, the maximum allowable path loss in general urban areas is $125.25dB$. Combining the corrected signal propagation model, the coverage radius of base stations in general urban areas is calculated to be $R2 = 0.61km$. Given that the total area of general urban areas is approximately 25 square kilometers, the number of base stations required ranges from 20 to 36.

- Coverage Radius and Required Number of Base Stations in Suburbs

Using the link budget formula, the maximum allowable path loss in the suburbs is $146.21dB$. Combining the corrected signal propagation model, the coverage radius of the suburban base stations in suburbs is calculated to be $R3 = 1.23km$. Given that the total area of the suburbs is approximately 25 square kilometers, the number of base stations required ranges from 3 to 10.

5. Research on base station coverage optimization based on self-adaptive mutation genetic algorithm

Base station coverage optimization is a multi-objective optimization problem. The construction cost, coverage, and wireless propagation environment of a base station are important factors that affect base station coverage optimization. This article proposes an AMGA based on genetic algorithms, which can effectively improve the accuracy of the algorithm while maintaining the diversity of the genetic algorithm population.

5.1. Optimization objective function construction

5.1.1. Cost objective function

Assume that the set of 5G candidate base stations is $T = \{1, 2, \dots, t\}$, the construction cost of 5G base station i is $d_i (i \in T)$, and the probability of base station i being selected is $x_i \in \{0, 1\} (i \in T)$. The set of existing 3/4G base stations in the planning area is $S = \{1, 2, \dots, s\}$. The probability of selecting an existing 3/4G base station is $y_j \in \{0, 1\} (j \in S)$, and the construction cost of base station J is $S_j (j \in S)$.

The selected base station can be divided into two situations, namely, the original 3/4G shared base station and the new base station. Therefore, the target function of the base station construction cost can be expressed as (20).

$$f_1 = \sum_{i=1}^t d_i x_i + \sum_{j=1}^n S_j y_j \quad (20)$$

5.1.2. Coverage objective function

Assume that d_{ik} is the distance from user k to base station i in the planning area, and r is the coverage radius of base station i . When $d_{ik} \leq r$, it indicates that the test point is within the coverage range, and vice versa, it is not within the coverage range. Therefore, the case where the test point is covered is $g_{ik} (i \in T, k \in N)$, which can be expressed as (21).

$$g_{ik} = \begin{cases} 1 & d_{ik} \leq r \\ 0 & \text{other} \end{cases} \quad (21)$$

Therefore, the coverage objective function can be expressed as (22).

$$f_2 = t \left(1 - \sum_{k=1}^n \frac{g_{ik}}{n} \right) \quad (22)$$

where t represents a 1% coverage vulnerability loss per occurrence.

5.1.3. Overall optimization objective function

Based on the above analysis, in the genetic algorithm, the coverage optimization of a base station mainly considers two optimization objectives: the base station construction cost and the coverage goal. The overall optimization function is (23).

$$F = m_1 f_1 + m_2 f_2 \quad (23)$$

where, m_1 represents the weight coefficient of construction cost, m_2 represents the weight coefficient of coverage loss, and $m_1 + m_2 = 1$. By default, $m_1 = m_2 = 0.5$. In actual projects, the value can be adjusted appropriately based on the actual situation.

5.2. Implementation of AMGA algorithm

Combining the basic principles of genetic algorithms [38], the specific steps for implementing the AMGA algorithm are as follows.

Code: The method of real number coding is used to optimize the location of a base station. In the algorithm, each individual represents the specific location of a base station, expressed as $a = \{g_1, g_2, \dots, g_c\}$, where $g_i = (x_i, y_i)$ represents the specific location of the i base station, x_i represents the abscissa of the i base station, and y_i represents the ordinate of the i base station.

Initialize population: The initial population is determined by randomly placing the initial base station location, and then an initial solution is generated. Each initial solution is the specific location of a base station. The target area is set as $\{S(x, y) : 1 \leq x \leq L, 1 \leq y \leq W\}$, where L and W are the length and width of the planned area, respectively.

Construct a fitness function: The fitness function is set based on the actual problem that needs to be solved, and its selection directly affects whether the AMGA algorithm can find the optimal base station location. The fitness function is shown in formula (23). This function is an indicator used to determine the degree of superiority or inferiority of base station location in the process of optimizing base station coverage. The ultimate optimization goal of this model is to achieve comprehensive coverage of the planned area as much as possible with the minimum number of base stations.

Select operation: The roulette wheel method is used for the selection operation of 5G base station optimization so that the probability of each base station location being selected is directly proportional to the optimization objective function value, as shown in (24) and (25).

$$p(x_i) = \frac{f(x_i)}{\sum_{j=1}^N f(x_j)} \quad (24)$$

$$q(x_i) = \sum_{j=1}^i p(x_j) \quad (25)$$

where $f(x_i)$ represents the individual fitness function value, $p(x_i)$ represents the probability of the individual being selected, and $q(x_i)$ represents the individual accumulation probability. Randomly generate a random number between $[0, 1]$. If the number is less than $q(x_i)$, it indicates that the base station is selected. Otherwise, it indicates that the base station is not selected. Compare the next base station directly until the optimal individual is selected.

Cross and mutation operations: Using the overall arithmetic cross-over method to improve population diversity, as shown in (26).

$$\begin{cases} x_{i+1} = w \cdot x_i + (1 - w) \cdot y_i \\ y_{i+1} = (1 - w) \cdot x_i + w \cdot y_i \end{cases} \quad (26)$$

where x_{i+1} and y_{i+1} represent the new parent individuals that have been cross generated, x_i and y_i represent the two parent chromosomes before the operation, and w represents a random number between $[0, 1]$. The specific steps are as follows: For two parent individuals, randomly

generate a random number between $[0, 1]$. If the number is less than the crossover probability, a crossover operation is performed to obtain two new parent individuals. Otherwise, please continue with the above operation.

The so-called mutation is to carry out gene mutation on individuals to prevent them from falling into local optimization. This article proposes an adaptive mutation method to achieve global optimization of the algorithm. After selecting the individual to be mutated according to the mutation probability, perform adaptive mutation to obtain a new mutated individual, as shown in (27).

$$x'_i = \begin{cases} x_i + \frac{1}{e^{u_i - l_i}} & (r \geq 0.5) \\ x_i - \frac{1}{e^{u_i - l_i}} & (r < 0.5) \end{cases} \quad (27)$$

where, x'_i represents the individual after mutation, x_i represents the individual before mutation, u_i represents the upper bound of the function value, l_i represents the lower bound of the function value, and r represents a random number between $[0, 1]$.

6. Simulation experiment and analysis

This topic will use MATLAB simulation software to conduct simulation experiments and analysis on the AMGA based on 5G base station coverage optimization. Jilin City will be selected as the area to be studied for base station coverage optimization. The base stations in the planning area will use directional antennas, and the coverage area will be an ideal circular shape. The planning fields are divided into three types: dense urban areas, general urban areas, and suburbs. The algorithm's experimental parameter settings are shown in Table 7.

6.1. Simulation experiment results in dense urban areas

Dense urban areas are generally commercial centers composed of dense high-rise buildings, and their personnel are relatively dense, which can easily cause problems such as signal blocking, overlapping coverage, and weak coverage. These problems pose enormous pressure on 5G communication networks. As shown in Fig. 8, an area covering an area of 25 square kilometers in Jilin City is selected as the location for dense urban areas to optimize the coverage of 5G base stations.

As can be seen from Fig. 4, to reduce the cost of establishing a station for operators, this layout of base stations fully utilizes the existing 3/4G shared base stations. It is known that there are 32 3/4G shared base stations in this area.

According to Section 5, the number of base stations in dense urban areas ranges from 48 to 62. Therefore, in the simulation experiment, the optimal results of the base station layout are shown in Table 8.

From the simulation comparison results in Tables 8 and it can be seen that when $m_1 = 0.3, m_2 = 0.7$, although the coverage target function result is slightly lower than the 92.8 % coverage result, the result saves the cost of building the station. Although the cost of building a station with this weight coefficient is higher than the latter six cases, its coverage effect is the best. Therefore, when $m_1 = 0.3, m_2 = 0.7$, it is more suitable for the problem of base station layout in this topic. The variation of the number of base stations with the fitness value of the

Table 7
System parameter settings.

Parameter	Numerical value
Maximum Iterations	1000
Population size	200
Cross probability	0.8
Variation probability	0.08
5G candidate base station sites	1000
Test point	1200

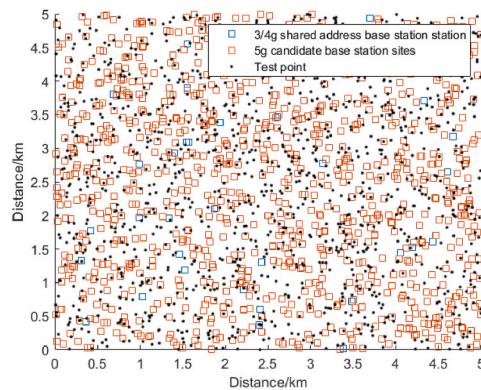


Fig. 8. Distribution of initial base stations in dense urban areas.

Table 8
Layout results of 5G base station in dense urban areas.

Weight coefficient m_1 , m_2	Station construction cost	Coverage rate	Number of 3/4G base stations	Number of 5G base stations
0.1,0.9	585	92.8 %	3	57
0.2,0.8	570	91.6 %	6	54
0.3,0.7	490	92.5 %	13	43
0.4,0.6	475	86.2 %	15	40
0.5,0.5	380	82.3 %	26	25
0.6,0.4	350	77.3 %	26	22
0.7,0.3	320	69.5 %	32	16
0.8,0.2	320	69.3 %	32	16
0.9,0.1	320	69.1 %	32	16

objective function is shown in Fig. 9.

As can be seen from Fig. 9, when the number of base stations is 56, the fitness value of the target function reaches the minimum, resulting in the optimal base station coverage map, which reduces the cost of the base station while achieving a wide range of coverage. The optimal base station coverage is shown in Fig. 10.

As can be seen from Fig. 10, after optimizing the coverage of 5G base stations, 43 new 5G base stations and 13 3/4G shared base stations are included, resulting in a base station coverage rate of 92.5 %.

This algorithm is compared with the traditional genetic algorithm and the multi-objective optimization algorithm NSGA-II. The results are shown in Table 9.

As shown in Table 9, the algorithm proposed in this topic reduces the site construction cost by at least 7 %, improves the coverage by at least 3.3 %, and reduces the number of base stations by at least 7 % compared to other algorithms.

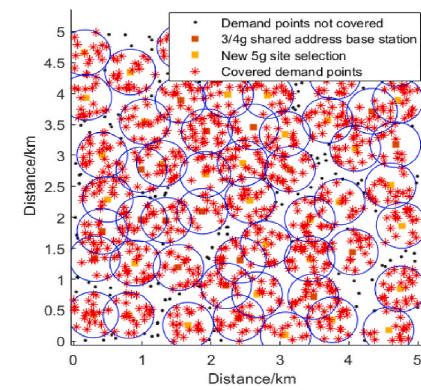


Fig. 10. Optimal layout of base stations in dense urban areas.

Table 9
Comparison results of 5G base station optimization in dense urban areas.

	Population size	Station construction cost	Coverage rate	Number of 5G base stations
NSGA-II	200	555	89.2 %	64
GA	200	525	88.6 %	60
AMGA	200	490	92.5 %	56

6.2. Simulation experiment results in general urban areas

General urban area refers to a relatively small population density and relatively low building height compared to dense urban areas. Fig. 11 shows the distribution of base stations in general urban areas of Jilin City.

As can be seen from Fig. 11, to reduce the cost of establishing a station for operators, this layout of base stations fully utilizes the existing 3/4G shared base stations. It is known that there are 20 3/4G shared base stations in this area.

According to Section 5, the number of base stations in general urban areas ranges from 20 to 36. Therefore, in the simulation experiment, the optimal results of the base station layout are shown in Table 10.

According to the simulation and comparison results in Table 10, when $m_1 = 0.3$, $m_2 = 0.7$, although the coverage objective function results are slightly lower than those with coverage rates of 99.4 % and 98.8 %, this result saves a lot of station construction costs. Although the cost of using this weighting coefficient to build a station is higher than the latter six cases, its coverage effect is the best. Therefore, when $m_1 = 0.3$, $m_2 = 0.7$, it is more suitable for the base station layout problem in this article. At this time, the number of base stations varies with the fitness value of the target function as shown in Fig. 12.

As can be seen from Fig. 12, when the number of base stations is 28,

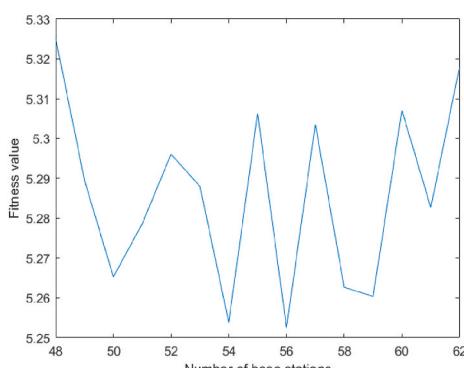


Fig. 9. Variation of fitness value with the number of base stations.

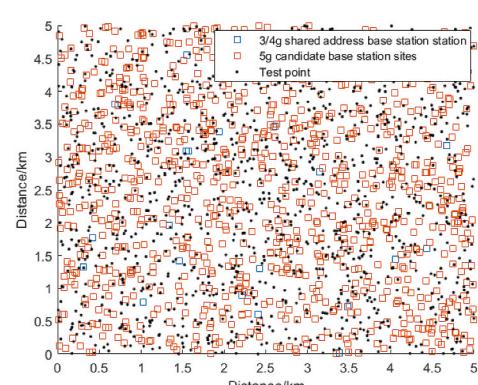
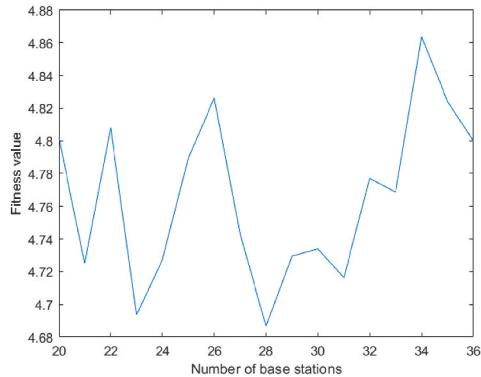


Fig. 11. Distribution of initial base stations in general urban areas.

Table 10

Layout results of 5G base station in general urban areas.

Weight coefficient m_1 , m_2	Station construction cost	Coverage rate	Number of 3/4G base stations	Number of 5G base stations
0.1,0.9	320	99.4 %	4	30
0.2,0.8	270	98.8 %	4	25
0.3,0.7	235	97.5 %	9	19
0.4,0.6	220	94.8 %	10	17
0.5,0.5	210	92.3 %	8	17
0.6,0.4	150	84.8 %	12	9
0.7,0.3	140	81.5 %	12	8
0.8,0.2	120	73.3 %	16	4
0.9,0.1	105	62.6 %	19	1

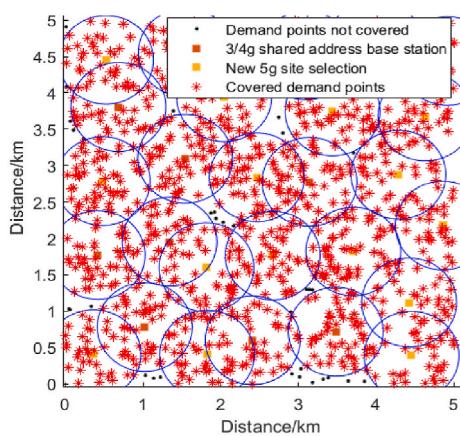
**Fig. 12.** Variation of fitness value with the number of base stations.

the fitness value of the target function reaches the optimal value, which means reducing the construction cost of the base station while achieving a wide range of coverage. The optimal base station coverage diagram is shown in Fig. 13.

As can be seen from Fig. 13, after optimizing the coverage of 5G base stations, 19 new 5G base stations, and 9 3/4G shared base stations are included, resulting in a base station coverage rate of 97.5 %.

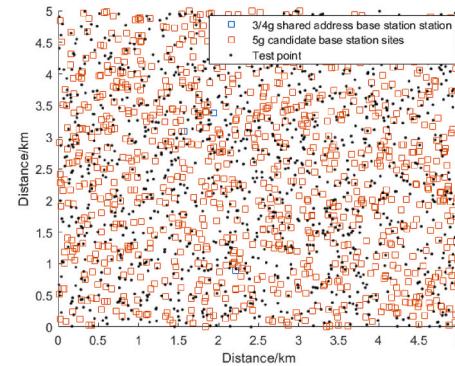
This algorithm is compared with the traditional genetic algorithm and the multi-objective optimization algorithm NSGA-II. The results are shown in Table 11.

As shown in Table 11, the algorithm proposed in this topic reduces the site construction cost by at least 13 %, improves the coverage by at least 5.4 %, and reduces the number of base stations by at least 17.6 % compared to other algorithms.

**Fig. 13.** Optimum layout of base stations in general urban areas.**Table 11**

Comparison results of 5G base station optimization in general urban areas.

	Population size	Station construction cost	Coverage rate	Number of 5G base stations
NSGA-II	200	280	92.1 %	36
GA	200	270	89.3 %	34
AMGA	200	235	97.5 %	28

**Fig. 14.** Distribution of initial base stations in suburbs.

6.3. Simulation experiment results in suburbs

Suburbs generally refer to areas with relatively sparse population density, relatively wide areas, and relatively few buildings. The village is known to have a distribution area of 25 square kilometers. As shown in Fig. 14, the distribution map of base stations in a suburb of Jilin City is shown.

According to Section 5, the number of base stations in suburbs ranges from 3 to 10. Therefore, in the simulation experiment, the optimal results of the base station layout are shown in Table 12.

According to the simulation comparison results in Table 12, when $m_1 = 0.4, m_2 = 0.6$, although the coverage target function results are slightly lower than the first three coverage results, this result saves a lot of site construction costs; Although the cost of using this weighting coefficient to build a station is higher than the latter four cases, its coverage effect is the best. Therefore, when comprehensively considering the weight coefficient $m_1 = 0.4, m_2 = 0.6$, it is more suitable for the base station layout problem in this article. At this time, the change in the number of base stations with the fitness value of the objective function is shown in Fig. 15.

As can be seen from Fig. 15, when the number of base stations is 8, the fitness value of the target function meets the requirements, that is, to reduce the cost of base stations while achieving a wide range of coverage. The optimal base station layout is shown in Fig. 16.

As can be seen from Fig. 16, after optimizing the coverage of 5G base stations, including 7 newly built 5G base stations and 1 3/4G shared

Table 12

Layout results of 5G base station in suburbs.

Weight coefficient m_1 , m_2	Station construction cost	Coverage rate	Number of 3/4G base stations	Number of 5G base stations
0.1,0.9	95	98.9 %	1	9
0.2,0.8	90	98.5 %	0	9
0.3,0.7	85	98.8 %	1	8
0.4,0.6	75	98.4 %	1	7
0.5,0.5	75	97.5 %	1	7
0.6,0.4	70	96.4 %	0	7
0.7,0.3	70	96.1 %	0	7
0.8,0.2	55	89.3 %	1	5
0.9,0.1	45	82.8 %	3	3

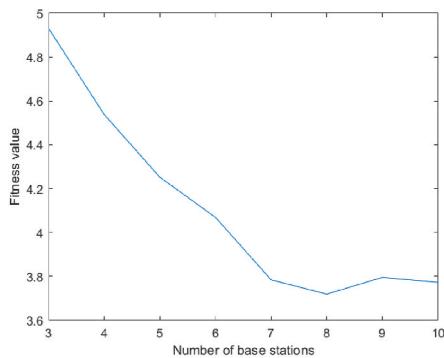


Fig. 15. Variation of fitness value with the number of base stations.

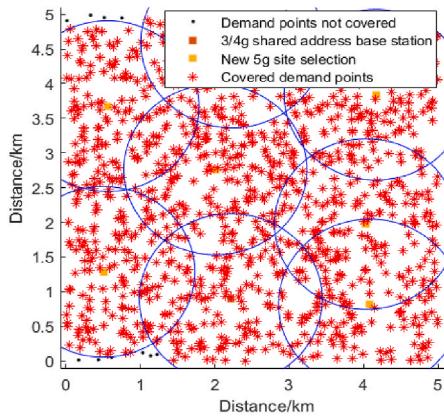


Fig. 16. Optimal layout of suburban base stations.

Table 13
Comparison results of 5G base station optimization in suburban.

	Population size	Station construction cost	Coverage rate	Number of 5G base stations
NSGA-II	200	85	91.2 %	11
GA	200	90	92.3 %	12
AMGA	200	75	98.4 %	8

base station, the base station coverage rate reaches 98.4 %, which greatly improves the communication quality of users.

This algorithm is compared with the traditional genetic algorithm and the multi-objective optimization algorithm NSGA-II. The results are shown in Table 13.

As shown in Table 13, the algorithm proposed in this topic reduces the site construction cost by at least 11.8 %, improves the coverage by at least 6.1 %, and reduces the number of base stations by at least 27.2 % compared to other algorithms.

7. Conclusion

This article investigates the optimization of 5G base station coverage in dense urban, general urban, and suburban areas, considering a mathematical model that aims to minimize costs while maximizing coverage. Road test data from the target regions are acquired and pre-processed using the CW testing method. These data are then utilized to enhance the signal propagation model, and the coverage radius and required number of base stations are determined through link budget analysis. An AMGA algorithm is proposed to optimize the layout of 5G base stations, taking advantage of existing infrastructure to further reduce costs. Simulation results demonstrate that the AMGA algorithm outperforms traditional genetic algorithms and NSGA-II, achieving

coverage rates of 92.5 %, 97.5 %, and 98.4 % in dense urban, general urban, and suburban areas, respectively. Moreover, the AMGA algorithm improves coverage by at least 3.3 %, 5.4 %, and 6.1 %, while reducing the number of base stations by 7 %, 17.6 %, and 27.2 % in each respective area.

Next, we aim to determine the optimal micro base station layout based on macro base stations, simplifying deployment complexity for combined macro-micro deployment. This aims to boost throughput and reduce energy consumption. Second, lacking actual user data, we simulated uniform distribution. We'll gather more actual user data to better reflect network conditions and facilitate algorithm application. Third, we'll consider more economic factors like deployment and site rental costs in our multi-objective model to provide a more detailed and practical solution.

CRediT authorship contribution statement

Jianpo Li: Conceptualization, Funding acquisition, Resources, Supervision, Writing – review & editing. **Jinjian Pang:** Conceptualization, Methodology, Software, Validation, Visualization, Writing – original draft. **Xiaojuan Fan:** Formal analysis, Investigation, Project administration, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

References

- [1] C. Casetti, 5G Standalone deployments are on the rise mobile radio, *IEEE Veh. Technol. Mag.* 18 (1) (2023) 5–11.
- [2] C.C. Kao, H.C. Young, Opportunities, challenges, and solutions in the 5G era, *IEICE Trans. Commun.* E105B (11) (2022) 1291–1298.
- [3] A. Jee, S. Prakriya, A coordinated direct AF/DF relay-aided NOMA framework for low outage, *IEEE Transactions on Communications* 70 (3) (2022), 1599–1579.
- [4] G. Liu, Y. Huang, Z. Chen, et al., 5G Deployment: standalone vs. non-standalone from the operator perspective, *IEEE Commun. Mag.* 58 (11) (2020) 83–89.
- [5] Y.J. Xu, G. Gui, H. Gacanin, F. Adachi, A survey on resource allocation for 5G heterogeneous networks: current research, future trends, and challenges, *IEEE Communications Surveys and Tutorials* 23 (2) (2021) 668–695.
- [6] D. Ficere, P. Varga, A. Wippelhäuser, et al., Large-scale cellular vehicle-to-everything deployments based on 5G-critical challenges, solutions, and vision towards 6G: a survey, *Sensors* 23 (16) (2023) 33–56, 7031.
- [7] J. Navarro-ortiz, P. Romero-diaz, S. Sendra, et al., A survey on 5G usage scenarios and traffic models, *IEEE Communications Surveys and Tutorials* 22 (2) (2020) 905–929.
- [8] A. Jee, K. Janghel, S. Prakriya, Performance of adaptive multi-user underlay NOMA transmission with simple user selection, *IEEE Transactions on Cognitive Communications and Networking* 8 (2) (2022) 871–887.
- [9] A. Jee, K. Agrawal, S. Prakriya, Performance of CDRT-Based underlay downlink NOMA network with combining at the users, *IEEE Transactions on Cognitive Communications and Networking* 9 (2) (2023) 414–434.
- [10] E. Eren, B.Y. Katanalp, Fuzzy-based GIS approach with new MCDM method for bike-sharing station site selection according to land-use types, *Sustain. Cities Soc.* 76 (2022) 103434.
- [11] J. Chen, J. Tian, S. Jiang, H. Li, J. Xu, The allocation of base stations with region clustering and single-objective nonlinear optimization, *Mathematics* 10 (13) (2022) 2257.
- [12] M. Liang, W. Li, J. Ji, Z. Zhou, Y. Zhao, H. Zhao, S. Guo, Evaluating the comprehensive performance of 5G base station: a hybrid MCDM model based on bayesian best-worst method and DQ-GRA technique, *Math. Probl Eng.* 2022 (2022) 4038369.
- [13] C.H. Wang, C.J. Lee, X.J. Wu, A coverage-based location approach and performance evaluation for the deployment of 5G base stations, *IEEE Access* 8 (2020) 123320–123333.
- [14] K.H. Liu, T.Y. Yu, Performance of off-grid small cells with non-uniform deployment in two-tier hetnet, *IEEE Trans. Wireless Commun.* 17 (9) (2018) 6135–6148.

- [15] L.C. Dai, H.T. Zhang, Propagation-model-free base station deployment for mobile networks: integrating machine learning and heuristic methods, *IEEE Access* 8 (2020) 83375–83386.
- [16] N. Gonzalez-Prelcic, A. Ali, V. Va, R.W. Heath, Millimeter-wave communication with out-of-band information, *IEEE Commun. Mag.* 55 (12) (2017) 140–146.
- [17] Z. Xiao, L. Zhu, J. Choi, P. Xia, X.-G. Xia, Joint power allocation and beamforming for non-orthogonal multiple access (NOMA) in 5G millimeter wave communications, *IEEE Trans. Wireless Commun.* 17 (5) (May 2018) 2961–2974.
- [18] M.M. Dong, T. Kim, J.J. Wu, E.W.M. Wong, Cost-efficient millimeter wave base station deployment in manhattan-type geometry, *IEEE Access* 7 (2019) 149959–149970.
- [19] A. Jee, S. Prakriya, Performance of energy and spectrally efficient AF relay-aided Incremental CDRT NOMA-Based IoT network with imperfect SIC for smart cities, *IEEE Internet Things J.* 10 (21) (2023) 18766–18781.
- [20] C. Nguyen, A.A. Cheema, A deep neural network-based multi-frequency path loss prediction model from 0.8 GHz to 70 GHz, *Sensors* 21 (2021) 5100.
- [21] A.F. Kola, C. Kurnaz, A.A. Cheema, A. Rahimian, Millimeter-wave dual-band MIMO channel capacity analysis based on climate data: a Samsun province case study, *Electronics* 12 (2023) 2273.
- [22] X. Raimundo, S. Salous, A. Cheema, Indoor dual polarised radio channel characterisation in the 54 and 70 GHz bands, *IET Microw., Antennas Propag.* 12 (8) (2018) 1287–1292.
- [23] J. An, K. Yang, J. Wu, N. Ye, S. Guo, Z. Liao, Achieving sustainable ultra-dense heterogeneous networks for 5G, *IEEE Commun. Mag.* 55 (12) (2017) 84–90.
- [24] P. Singh, Y.C. Chen, Sensing coverage hole identification and coverage hole healing methods for wireless sensor networks, *Wireless Network* 26 (3) (2020) 2223–2239.
- [25] J.F. Jiang, H.J. Ding, Uplink spectrum overlay coverage enhancement algorithm in 5G network, *Math. Probl. Eng.* 2021 (2021) 9724623.
- [26] L.Y. Liu, Z.G. Ma, Y. Xue, W.B. Yan, Y.Y. Li, Research on coverage probability in ultra-dense 5G heterogeneous cellular networks based on Poisson clustered process, *Wireless Pers. Commun.* 95 (3) (2017) 2915–2930.
- [27] L. Liu, Y.Q. Zhou, W.H. Zhuang, J.H. Yuan, L. Tian, Tractable coverage analysis for hexagonal macrocell-based heterogeneous UDNs with adaptive interference-aware CoMP, *IEEE Trans. Wireless Commun.* 18 (1) (2019) 503–517.
- [28] R.N. Zhang, H.C. Xu, X.J. Du, D.Y. Zhou, M. Guizani, Dual-polarized spatial-temporal propagation measurement and modeling in UMa O2I scenario at 3.5 GHz, *IEEE Access* 7 (2019) 122988–123001.
- [29] M. Sousa, A. Alves, P. Vieira, M.P. Queluz, A. Rodrigues, Analysis and optimization of 5G coverage predictions using a beamforming antenna model and real drive test measurements, *IEEE Access* 9 (2021) 101787–101808.
- [30] S. Salous, J. Lee, M.D. Kim, M. Sasaki, W. Yamada, X. Raimundo, A.A. Cheema, Radio propagation measurements and modeling for standardization of the site general path loss model in International Telecommunications Union recommendations for 5G wireless networks, *Radio Sci.* 55 (1) (2020) 6924.
- [31] T.Y. Li, S.C. Hua, L. Kang, S.H. Chang, Research on TD-LTE wireless communication network propagation model optimization based on visual simulation and GIS, *EURASIP J. Wirel. Commun. Netw.* 2021 (1) (2021) 161.
- [32] H. Aghasi, P. Heydari, Millimeter-wave radars-on-chip enabling next-generation cyberphysical infrastructures, *IEEE Commun. Mag.* 58 (12) (2020) 70–76.
- [33] R. Hussain, M.S. Sharawi, 5G MIMO antenna designs for base station and user equipment: some recent developments and trends, *IEEE Antenn. Propag. Mag.* 64 (3) (2022) 95–107.
- [34] C.C. Fan, X.F. Zhong, J. Wei, BS-to-ground channel reconstruction with 3D obstacle map based on RSS measurements, *IEEE Access* 7 (2019) 99633–99641.
- [35] D. Casillas-Perez, C. Camacho-Gomez, S. Jimenez-Fernandez, J.A. Portilla-Figueras, S. Salcedo-Sanz, Weighted ABG: a general framework for optimal combination of ABG path-loss propagation models, *IEEE Access* 8 (2020) 101758–101769.
- [36] X.Z. Xu, Q.T. Fang C, L. Zheng, Propagation model correction and link budget based on 5G new scenario, *Journal of Nanjing University of Posts and Telecommunications (Natural Science Edition)* 40 (2) (2020).
- [37] T. Patra, S.K. Mitra, Link budget analysis for 5G communication in the tropical regions, *Wireless Communications & Mobile Computer* 2020 (2020) 6669965.
- [38] Z. Rosic, O. Mihic, D. Aleksic, D. Drajic, Novel method for optimal synthesis of 5G millimeter wave linear antenna array, *Int. J. Antenn. Propag.* 2017 (2017) 6848234.