



# A Local Line Optimization Model for Urban Rail Considering Passenger Flow Allocation

Peng He<sup>1</sup> · Hao Tang<sup>2</sup> · Feng Chen<sup>3</sup> · Zijia Wang<sup>1</sup> · Ying Sun<sup>4</sup> · Bobo Yang<sup>4</sup> · Jin Wang<sup>4</sup> · Na Li<sup>5</sup>

Received: 10 August 2023 / Revised: 22 November 2023 / Accepted: 13 December 2023 / Published online: 17 April 2024  
© The Author(s) 2024

**Abstract** It is important to strengthen the research on urban rail transit (URT) existing line renovation strategies. In this paper, we investigate the optimization of bottlenecks that are less attractive but have strong travel demand in existing URT networks. A URT local line optimization model is constructed. The maximum passenger flow and minimum project cost are chosen as the optimization objective for the benefit of both passengers and operators, and several actual constraints are considered in the proposed model, such as the station interval. In order to obtain higher computational efficiency and accuracy, a passenger flow allocation method is embedded in a genetic algorithm with elitist preservation. Taking the local network of the Beijing URT as a case study, the calculation results show that the designed algorithm can quickly and effectively obtain the optimal solution, and the generated local line scheme is able not only to meet the

regional travel demand, but also to optimize the connection relationship of the existing URT network. This study can provide a reference method for increasing the attraction of URT and optimization of existing URT networks.

**Keywords** Urban rail transit · Passenger flow distribution · Local network generation · Line optimization · Genetic algorithm

## 1 Introduction

Urban rail transit (URT) networks play a major role in facilitating residents' travel behavior and alleviating traffic congestion. However, with the complex spatial and temporal distribution of passenger flow in China's URT network, the mismatch between passenger demand and transportation supply during peak hours is becoming increasingly obvious. Passenger congestion is extremely serious at some key stations, and emergency management also faces huge challenges. These problems ultimately lie in the incongruity

---

✉ Jin Wang  
jerryW@csu.edu.cn

Peng He  
21114123@bjtu.edu.cn

Hao Tang  
tanghao1@bjucd.com

Feng Chen  
chenfengacademic@gmail.com

Zijia Wang  
zjwang@bjtu.edu.cn

Ying Sun  
214807005@csu.edu.cn

Bobo Yang  
boboya@csu.edu.cn

Na Li  
lina@mail2.gdut.edu.cn

Communicated by Baoming Han.

- <sup>1</sup> School of Civil Engineering, Beijing Jiaotong University, Beijing, China
- <sup>2</sup> Beijing Urban Construction Design and Development Group Co., Ltd, Beijing, China
- <sup>3</sup> Research Center of Beijing Rail Transit Line Security and Disaster-Resistance Technology, School of Civil Engineering, Beijing Jiaotong University, Beijing, China
- <sup>4</sup> MOE Key Laboratory of Engineering Structures of Heavy-Haul Railway, Center for Railway Infrastructure Smart Monitoring and Management, School of Civil Engineering, Central South University, Changsha 410000, China
- <sup>5</sup> Guangdong Meilong Railway Co., Ltd., Guangzhou, China

between the network structure and the spatial and temporal distribution of urban residents' travel, which limit the capacity of the URT [1]. Taking Beijing URT as an example, although there are many URT lines in the central city, only 16.5% of Beijing's central city residents traveled by URT on weekdays in 2019, while the percentage of trips by car was 24.3%. The attraction of URT is clearly not comparable to that of cars. However, in Tokyo, Japan [2], in 2008, the share of travel by car was only 11%, and that of URT transit was 48%. Thus, the Beijing URT network has great potential to be tapped.

There is currently no standardized and explicit definition for bottleneck in URT networks. In this study, we consider key areas that limit the effectiveness of URT networks as bottleneck regions. Specifically, from the perspective of matching residents' travel demands with URT supply, we believe that within existing rail network stations, there are origin–destination (OD) pairs with high demand for multimodal travel but extremely low usage of rail transit for commuting. These areas indicate a mismatch between the coverage and connectivity of URT and urban development, and we refer to them as bottlenecks in this study. We first outline the basic principles for the deployment of URT stations in bottleneck areas. A method for identifying potential rail transit stations is then proposed based on the urban hierarchy road network and conventional bus line network. The route between potential station locations in the identified bottlenecks is determined based on the urban road network, in preparation for subsequent optimization of local network route generation. Building upon this, we introduce two objective functions, passenger flow and estimated engineering investment, as well as five constraints including station interval and terminal stations. A single-objective route generation model considering passenger distribution is constructed. The improved genetic algorithm (GA) is employed to solve this model and optimize the routes in the bottleneck areas for enhancement. The study thus provides methodological references for enhancing the role of URT, and solving the traffic congestion problems in large modern cities.

The rest of this paper is organized as follows. In the “Literature Review” section, we review the literature related to the optimization and transformation of transportation networks. We present our optimization model in the “Model Formulation” section and express it as a mathematical model. Then, in the “Algorithm” section, we propose a GA-based algorithm. The performance of our proposed model and solution method is evaluated through the cases in the “Case Studies” section. Finally, we provide a summary in the “Conclusions” section.

## 2 Literature Review

The design and optimization of URT networks are important research areas within URT planning. Existing studies focus primarily on optimizing new URT networks before their construction and can be broadly classified into three categories. The first category of research involves adjusting and optimizing newly built network designs by constructing an evaluation index system for network schemes, but it lacks attention to network generation. The second category focuses on optimizing public transportation networks based on newly constructed URT lines. The third category involves optimizing train operation plans for newly constructed URT systems.

Many different approaches have been proposed in terms of network scheme evaluation index system methods. Liu and Qian [3] addressed the limitations of decision-makers' rationality in practical situations by introducing a multi-objective method. They constructed a multi-objective decision-making evaluation index system for the selection of URT network schemes and calculated the comprehensive weights of each indicator using a synthetic method. Yang et al. [4] proposed an improved TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) optimization decision model for URT network schemes based on the rational behavior of expert groups. Through proximity calculation, they ranked network schemes to achieve a comprehensive decision-making approach that ensures the rationality of traffic network planning for the entire community. Zhang et al. [5] presented an evaluation index weight calculation method based on the analytical hierarchy process (AHP) and entropy weight methods. They comprehensively considered the subjective factors of rich expert experience and the objective information of the indicators, conducting an evaluative analysis of the effectiveness of URT network planning.

With regard to adjusting public transportation networks based on newly constructed URT lines, some studies have focused primarily on optimizing the entire or partial bus networks. In the early stages of model construction, Ceder and Wilson [6] delineated the bus network design problem, summarized various approaches proposed to address this issue, and introduced two mathematical models for public transit network design based on passenger travel time. During the bus network design phase of the Rome transportation agency, Cipriani [7] discovered a linear relationship between passenger volume and speed. Pternea et al. [8] considered sustainable design goals and emission-free (electric) vehicles, introducing a direct route design model with route structure and direct control to address a sustainable public transit network design problem. Barahimi et al. [9] established a mathematical model to determine the capacity increase for links in a dual-mode public transportation

network and proposed a particle swarm optimization algorithm with dynamic parameters to solve a multi-objective bi-level model. Nayeem et al. [10] proposed a population-based traffic network design model that prioritizes maximizing the number of passengers served, minimizing the total number of transfers, and minimizing the total travel time for all passengers. They presented a GA-based optimization method for bus network design. As research deepens, models become increasingly complex, incorporating more factors. López-Ramos [11] considered passenger travel time and operational costs, proposing an optimization method to simultaneously address network design and frequency-setting issues for rapid rail transit networks. In a study by Liang et al. [12], the problem was formulated as a multi-objective model with two conflicting objectives: minimizing passenger costs and reducing operator costs. In the solution of such problems, Chakroborty [13] analyzed the reasons for the problems with traditional methods in solving the urban transit network design problem (UTNDP). Emphasizing the effectiveness of GA-based procedures in solving the UTNDP, Walteros et al. [14] used GAs, simulated annealing, and taboo search to provide a set of feasible solutions for the design of a public transportation line network as well as frequency setting in a reasonable amount of time.

Some studies have investigated the optimization of train operation plans for newly constructed URT lines. Canda et al. have conducted a series of ongoing studies that provide a solid foundation for integrating multiple transportation planning steps. In 2016, a mixed-integer nonlinear programming model [15] was proposed to optimize the line frequency and capacity of the URT network with the optimization objectives of minimizing the average travel time and the operation, maintenance, and vehicle purchase costs. It addressed issues such as track allocation for different routes and passenger route assignment. In 2017, with network construction cost, vehicle purchase, operation, and expected income as optimization objectives, an optimization model [16] for URT network topology and line planning were established. The authors innovatively introduced an adaptive large-neighborhood search algorithm for model solving. In 2019, a two-layer optimization model [17] of railway rapid transit (RRT) was constructed, with the line network and line generation on the upper layer. The lower layer of the model considered the passenger flow distribution and train selection and set the line frequency. The lower model layer was linearized as shown in Appendix A, and then was solved using Gurobi optimization. The obtained line frequency and other parameters are used as the input for the upper model for an iterative solution. In addition, researchers in this field have actively explored the introduction of novel objective functions, constraints, and solution

methods. Sun et al. [18] built a multi-objective model with maximum passenger volume and minimum passenger travel time for a single-corridor service area under a given rail transit route. In order to solve this model, a new algorithm based on GA was proposed. The feasibility and effectiveness of the proposed model and solution were verified by an example. Cadarso et al. [19] considered passenger demand, strategic costs, and uncertainty in network interruptions. They proposed a two-stage stochastic 0–1 programming model and a hybrid element heuristic algorithm to solve the optimization problem for the rapid transport network. Laporte et al. [20] proposed a network stepwise generation method with the optimization objectives of construction cost and population coverage. The main idea was to first determine an initial set of corridor strips through the planner's a priori knowledge, then fine-tune these line geometries by generating a set of non-dominated paths, and finally solve the bi-objective integer planning procedure to obtain the network solution. Zhao et al. [21] adopted a memetic algorithm to optimize the urban transit network. The mathematical model took the optimal route configuration and service frequency of the URT network as the objective function and aimed to minimize passenger travel costs and unmet passenger demands. The optimization algorithm was embedded with the local search operator based on the classical GA, which improved the computational performance of the algorithm.

In the present study, a comprehensive review of previous literature identified a lack of research on the actual operation and network optimization of existing URT systems. The optimization and transformation of existing URT networks require adjustments and capacity expansions under the constraints of network topology and passenger flow. Building upon the aforementioned methods for optimizing newly built networks, this study explores the construction of local network optimization models for existing URT systems. In order to meet passenger travel demands and reduce operational costs, two objective functions are proposed: passenger flow and estimated investment in line engineering. We also take into account practical constraints such as station interval and mandatory stops. Due to the complexity and specialization of the model, we convert the multi-objective functions into a single objective function, and the GA is improved through elite retention. By optimizing and solving the model considering network topology and passenger demand, local optimization and transformation of existing URT networks are achieved. A case study for optimizing the Beijing URT validates the effectiveness of this research model and algorithm. We also present some policy recommendations, which provide references for future adjustments and transformations of URT networks.

### 3 Model Formulation

In response to the issue of supply–demand imbalance between existing URT networks and the travel demands of city residents, this study constructs a nested passenger flow allocation model for URT line generation. To consider the mutual influence between passenger flow changes and line generation models, the network passenger flow allocation method is first introduced. Next, in order to construct a local line generation model, the selection process for potential URT stations is determined based on the principles of station selection. This model establishes connections between potential URT stations, aiming to obtain an optimal line considering multiple objectives including passenger travel demands and operational costs. To ensure the practical application of the generated new line scheme, this model not only considers basic constraints such as station interval and passenger volume but also takes into account constraints in practical situations such as OD points and mandatory stops. The following is a detailed introduction to the proposed model.

#### 3.1 Network Traffic Allocation

The generation of a new feasible line scheme in the local line generation model means the generation of a new URT network. Different line schemes inevitably lead to different URT networks, making it possible to change the route choice of passenger flows on different line networks with fixed origins and destinations. When considering the response of passenger flows to changes in the topology of the network, it is necessary to perform a passenger flow allocation on the network after route optimization. Then the data such as urban rail traffic flow between stations in the network, linear distance between stations, and network distance between stations are updated, and these data are fed to the line generation model for calculating relevant objective functions and constraints, and determining the optimization direction of the model. Therefore, the route generation model must consider passenger flow allocation.

The existing research on traffic flow allocation is relatively mature. When the famous scholar Wardrop put forward the first and second theorems of traffic network equilibrium allocation, he began to study the traffic flow passenger allocation under traffic congestion with system analysis and equilibrium analysis methods. This enabled a great leap forward in the theory of traffic flow allocation. Based on classical methods, subsequent studies have improved the practicability and accuracy of the passenger flow allocation model by improving the generalized travel cost function [22], the selection method of effective path set [23], and the probability selection model of effective path selection [24].

Similar to the traffic flow allocation in the road network, passenger flow allocation in the URT network involves the allocation of the obtained passenger flow to each zone in the network according to certain rules and then determining the passenger flow for each zone in the network.

Considering the calculation complexity of the constructed line generation model and the granularity of routing and station locations in the planning stage of the network, the classical all-or-none passenger flow allocation method is adopted in this study. This means that all passenger flows on an OD will be allocated to the path with the lowest cost. The network distance traveled is used as the cost index, and the cost function can be expressed as

$$C_k = \sum_{i \in I} l_i, \quad (1)$$

where  $C_k$  denotes travel cost for path  $k$ ,  $I$  represents the set of intervals that path  $k$  passes through,  $i$  is the  $i$ th interval in set  $I$ , and  $l_i$  is the network distance of the  $i$ th interval.

#### 3.2 Potential Rail Transit Station Selection

In order to construct the local line generation model, after identifying the bottlenecks of the existing URT network, some potential URT stations should be selected in the bottleneck area.

The URT network is connected by lines between stations, and the impact of lines on the city is more than that of stations, because the station is only a local point area, while the line is a long strip of influence on the city, and the direction of the line and the layout of the station are mutually affected. Therefore, from the perspective of reducing the disturbance to the city and reducing the investment, it is necessary to restrict the route before determining the potential URT stations.

The general principle of line design is to lay the line as far as possible along the city road, and try not to cross the planning red line on either side. In order to reduce the disturbance to the built-up area, reduce the demolition, and control the project cost, the line plane should be laid along the road with a width of at least 20 m as far as possible. According to the construction standards of Chinese roads, the two-way expressways, main roads, and secondary roads with more than four lanes in large and medium-sized cities can be used as the routing constraints between alternative stations of URT lines.

Combined with the availability of data, this study selects city-level road network and conventional bus network to select potential rail transit stations in the region. The basic selection process is as follows:

1. Route determination: Routes should be preferentially laid on the urban road network with four or more vehicle lanes.
2. Preliminary determination: Locate stations at the intersections of expressways and trunk roads, trunk roads and main roads, and trunk roads and secondary roads that are four lanes or more. This is because locating the station at a road intersection with sufficient space allows for less demolition and disturbance to the surrounding area, as well as meeting the demand for travel in multiple directions.
3. Station determination: By visualizing the conventional bus network and counting the number of conventional bus line stations within 800 m around the preliminary alternative station in step 2, we ensure that there are certain travel demands around the selected station.

### 3.3 Model Assumptions

1. The potential stations are also ready for operation and the line generated by the proposed model is an extension, branch, liaison line of an existing line, or the first section of a new line.
2. All the generated line schemes adopt the same traffic organization mode, purchase the same number of trains, and use the same standard vehicles as the connected existing lines.
3. Assume that all new stations are underground stations.
4. Urban rail transit stations mainly include underground stations, ground stations, and elevated stations, and the costs of different laying methods are also different. In addition, the engineering environment of the station has a great impact on the cost, and the cost of different construction methods is also different. Since the model constructed in this work is mainly for the early stage of planning, it is assumed that a uniform estimate is used for all new stations, and another standard uniform estimate is used for a few existing stations that need to be renovated.
5. All passengers travel according to the shortest path. When there are multiple shortest paths, one shortest path is selected at random.

### 3.4 Objective Functions

#### 3.4.1 Passenger Flow

Generally, URT is built to serve the travel needs of citizens and relieve the pressure of urban road traffic. The volume of URT is the largest among all urban public transportation. Taking subway, for example, the one-way peak passenger volume can reach 30,000–60,000 per hour. Therefore, the size of passenger flow must be considered when building a

new URT line. In addition, from the perspective of bringing convenience to residents' travel, the greater the number of people covered by the station service area, the more residents can enjoy the travel convenience provided by URT.

In this study, the all-mode passenger flow between ODs is taken as the benchmark, and the URT passenger flow between ODs related to potential stations is obtained according to a certain proportion. According to the Annual Report of Beijing Transport Development in 2021 released by the Beijing Transport Development Research Institute, the travel share rate for URT in 2019 before COVID-19 was 16.5%. Here, 16.5% is taken as the URT share of all-mode passenger flow.

In this research, the number of people within 800 m around the station is taken as the substitute index of passenger flow, for two reasons. Firstly, the generated feasible line connections to the existing network in the model can result in changes in the network structure and the reconstruction of passenger flow distribution. Traditional methods for predicting passenger flow distribution in URT networks mostly rely on historical data patterns. However, after the network optimization, there is no historical passenger flow distribution data available for the new stations or between the new stations and existing ones. Moreover, the detailed data related to land use and socioeconomic characteristics associated with the lines are not easily obtainable. Therefore, accurately predicting passenger flow in the URT network after line optimization is a challenge. Secondly, based on research on the walkability of URT stations [25], residents consider a suitable walking distance to nearby stations to be around 750 m. This means that places that require a walking distance of more than 750 m are considered far, and in such cases, people tend to choose other modes of transportation or select alternative destinations. Considering the variability in the data used as the basis for calculation, such as height and walking speed mentioned in this paper, the suitable walking distance is appropriately relaxed to 800 m. Within this distance range, passengers are more likely to choose the station as their origin or destination for their travels. The mathematical form is shown as follows:

$$f_1 = \sum_{i=1}^{m_K} P_{S_{A_i}} + \sum_{j=1}^{n_K} P_{S_{E_j}}, \quad (2)$$

where  $f_1$  is passenger flow,  $S_A$  is the abbreviation for alternative station,  $S_{A_i}$  represents the  $i$ th alternative station,  $S_E$  is the abbreviation of existing station,  $S_{E_j}$  represents the  $j$ th existing station,  $P_{S_{A_i}}$  represents the number of people that can be covered within 800 m of the  $i$ th alternative station,  $P_{S_{E_j}}$  represents the number of people that can be covered within 800 m of the  $j$ th existing station,  $K$  represents the  $K$ th new line scheme,  $m_K$  represents the number of alternative stations

on the  $K$ th optimized new line, and  $n_K$  represents the number of existing stations on the  $K$ th optimized new line.

### 3.4.2 Project Construction Cost

URT engineering construction requires high project investment, and its design should be conducive to reducing investment and other costs without affecting safety and reliability and other conditions. URT is a huge and complex system engineering, which is composed of a series of related facilities and equipment, including lines, vehicles, stations, power supply system, communication, signal, fire protection, and environmental control systems. All systems cooperate to provide services for passengers. From the perspective of URT project budgets, the main costs of building a URT line include four types [26]: (a) station, section, communication, signal, track, and other engineering costs; (b) compensation

where  $f_2$  is the engineering construction cost;  $\eta_i$ ,  $\mu_j$ ,  $\gamma$ , and  $\lambda$  are all 0–1 variables, that is, 1 if selected for construction or 0 otherwise;  $C_{S_{A_i}}$  represents the cost of constructing the  $i$ th alternative station;  $C_{S_{E_j}}$  is the abbreviation for existing station,  $S_{E_j}$  represents the  $j$ th existing station;  $P_{S_{A_i}}$  represents the cost of building a new interchange station near the  $i$ th existing station;  $l_{a,b}$  represents the interval line length between two adjacent stations  $a$  and  $b$ ;  $C_{\text{section}}$  represents the construction cost of unit length interval;  $C_{\text{stabling yard}}$  represents the construction cost of a URT parking lot; and  $C_{\text{depot}}$  represents the construction cost of a depot on a URT line.

### 3.4.3 Linear Weighted Objective Function

Based on previous engineering experience, when the passenger flow is 10,000 and the construction cost is 1 billion yuan, the value of each objective function is similar. Therefore, this study chooses the linear weighting method to convert the two objectives into a single objective [27].

$$\min f = -z_1 \times \left( \sum_{i=1}^{m_K} P_{S_{A_i}} + \sum_{j=1}^{n_K} P_{S_{E_j}} \right) + z_2 \times \left( \sum \left( \eta_i C_{S_{A_i}} + \mu_j C_{S_{E_j}} \right) + \sum l_{a,b} C_{\text{section}} + \gamma C_{\text{stabling yard}} + \lambda C_{\text{depot}} \right), \quad (4)$$

for land acquisition, construction (structure) relocation, pipeline relocation, traffic relief, commercial compensation, and other project construction expenses; (c) reserve expense, also known as unforeseeable expense of project construction, which refers to the expense that should be reserved in advance for unpredictable expenses that may occur in the process of project implementation; and (d) special expenses such as vehicle purchase fee and loan interest. The statistical data for existing URT projects show that, generally speaking, the engineering cost of the project accounts for about 60% of the total cost, expenses for demolition and other engineering construction account for about 20%, reserve expenses account for about 3–4%, and special expenses such as vehicle purchase account for about 16%. The civil construction, installation of signal facilities, and compensation costs for land expropriation and demolition in the station area occupy about 80% of the budget estimate of a URT project, among which the budget estimate for the station constitutes 20% of the total budget estimate. Considering the availability of cost data and the small impact of special expenses such as reserve expense and vehicle purchase expense on investment, it is ignored here. The total project investment function is expressed as follows:

where  $z_1$  and  $z_2$  represent the weight coefficients of the first and second objective functions, respectively.

## 3.5 Constraints

### 3.5.1 Station Interval Constraint

As the distance within the city is generally short, in order to ensure the attractiveness of the line, the energy consumption of the train, and the operational efficiency, the interval between stations should not be too large or too small.

The minimum allowed value of the distance between stations in the model is  $l_{\min}$ , and the maximum allowed value is  $l_{\max}$ , and  $l_{a,b}$  represents the network distance between station  $a$  and station  $b$ . The mathematical form is as follows:

$$l_{\min} < l_{a,b} < l_{\max} \quad (5)$$

$$f_2 = \sum \left( \eta_i C_{S_{A_i}} + \mu_j C_{S_{E_j}} \right) + \sum l_{a,b} C_{\text{section}} + \gamma C_{\text{stabling yard}} + \lambda C_{\text{depot}}, \quad (3)$$

### 3.5.2 Single line length constraint

Similar to station intervals, the length of a URT line should not be too long or too short. In general, URT lines rarely have the problem of being too short. In addition, urban residents usually travel for short and medium distances, and URT lines have high investment in the early stage, so a URT line is usually not very long. According to existing studies, a reasonable length for a single line of URT ranges from 14.7 to 45.7 km and increases with the increase in passenger flow intensity. In the optimization model constructed in this study, the line upper limit value is set as  $L_{max}$ , and the total length of the line  $K$  is set as  $L_K$ .

$$L_{min} < L_K < L_{max}$$

$$L_K = \sum_{a=1}^{N_K} l_{a,a+1} \quad (6)$$

where  $N_K$  is the total number of stations for the  $K$ th line scheme,  $a$  is the  $a$ th station,  $a + 1$  is the next station after the  $a$ th station, and  $l_{a,a+1}$  is the distance between adjacent intervals ( $a, a + 1$ ).

### 3.5.3 Quantity Constraint

The objective function value is different under the condition of generating different numbers of lines, so it is necessary to constrain the number of optimized lines as well. The number can be set to 1, 2, or the desired number.

$$N_l = 1 \text{ or } N_l = 2 \text{ or } N_l = 3 \dots, \quad (7)$$

where  $N_l$  is the number of lines.

### 3.5.4 Origin and destination constraint

In order to enable the optimized line to access the existing URT network, it is generally agreed that the stations at both ends of the optimized line should belong to the set of existing URT stations:

$$k[0] \in \{S_E\} \text{ and } k[-1] \in \{S_E\}, \quad (8)$$

where  $k[0]$  is the origin of the lines and  $k[-1]$  is the destination.

### 3.5.5 Obligatory Station Constraint

During actual line selection, there may be some important anchor stations to pass through. In order to be closer to the actual situation of line selection, the obligatory station constraint is added:

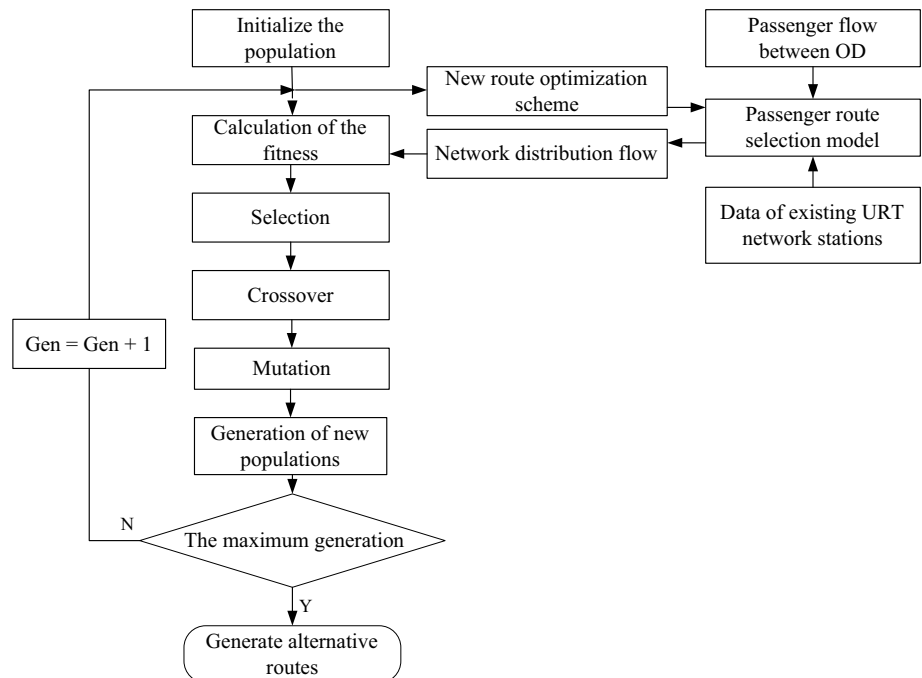
$$i \in \{k[0], k[1], \dots, k[-1]\}, \quad (9)$$

where  $k[i]$  is the  $i$ th station of a line.

## 4 Algorithm

In view of the complexity and specialization of the model considering passenger flow allocation constructed in this

**Fig. 1** Basic process of E-GA nested with passenger flow allocation method



study, the existing GA with elitist preservation (E-GA) is modified, and a method using E-GA nested with passenger flow allocation is proposed. The basic flow chart of this algorithm is shown in Fig. 1, and the steps are outlined in the following.

*Step 1 Coding.* According to the characteristics of the model constructed in this study, the number of candidate stations is set as the length of the gene, and the bases on the genes represent stations. Specifically, since there are 19 candidate stations, the length of an individual’s genotype is set as 19. The 19 bases on the gene are encoded using integers from 0 to 19. Here, 0 indicates that no station is selected at that position, while an integer  $n$  between 1 and 19 indicates that station  $n$  is selected at that position.

*Step 2 Initialization.* Initialize the chromosome population through random generation, then calculate the fitness of each individual. Next, select  $N$  individuals to form a new population via the elitist selection algorithm. And by substituting the first parent population into the passenger flow distribution model, the data for passenger flow and network distance between ODs in the network are updated to calculate the value of the new objective function and the value of each constraint condition. A static penalty function is introduced to ensure that the generated solutions satisfy the constraints.

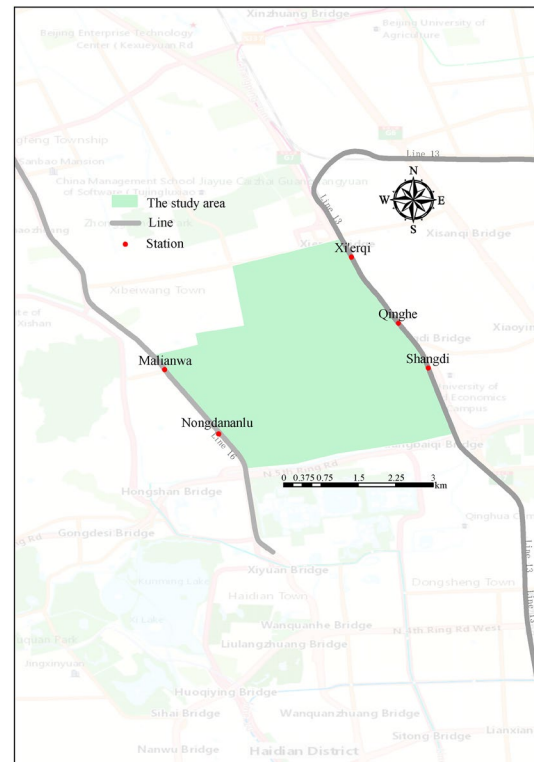


Fig. 3 Schematic of the study area

*Step 3 Crossover and mutation.* The new population can obtain a new parent population through operations of two-point crossover and uniform mutation. The crossover and mutation coefficients in this paper are selected based on empirical values. The crossover coefficient is set as 0.9, and the mutation coefficient is set as 0.1.

*Step 4 Calculate the fitness of each individual in the new parent population, and then obtain the new parent population through elitist selection, two-point crossover, and uniform variation.* And by substituting the parent population into the passenger flow distribution model, the data for passenger flow and network distance between ODs in the network are updated to calculate the value of the new objective function and the value of each constraint condition. A static penalty function is used to ensure that the generated solutions satisfy the constraints.

*Step 5 Repeat step 3 and step 4 until the maximum generation is reached; then the algorithm is terminated.*

## 5 Case Studies

### 5.1 Case Overview

We obtained a bottleneck in the Beijing URT network from relevant departments and selected its local area to verify the

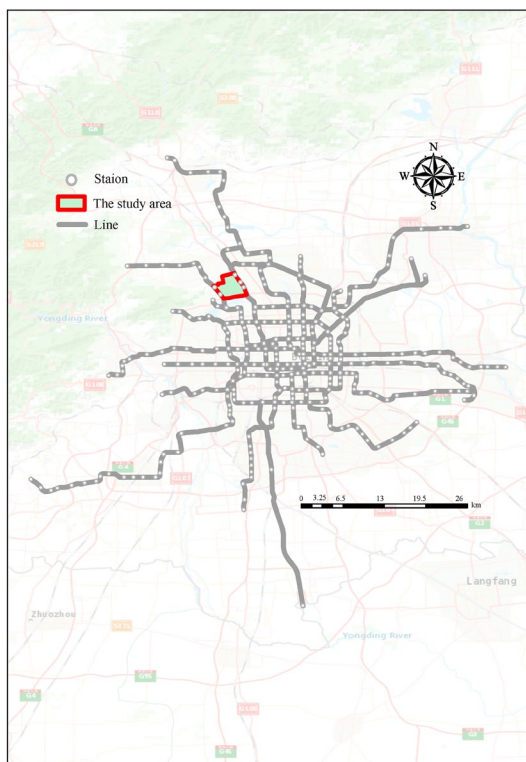
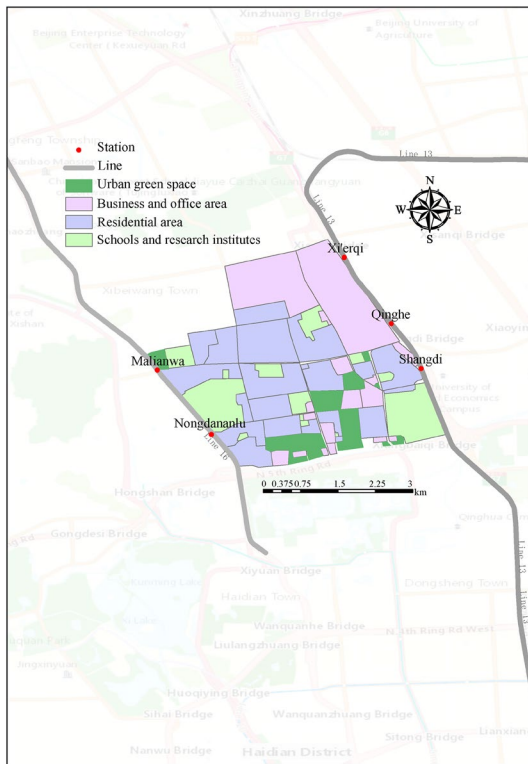


Fig. 2 Location of the study area with respect to the whole network



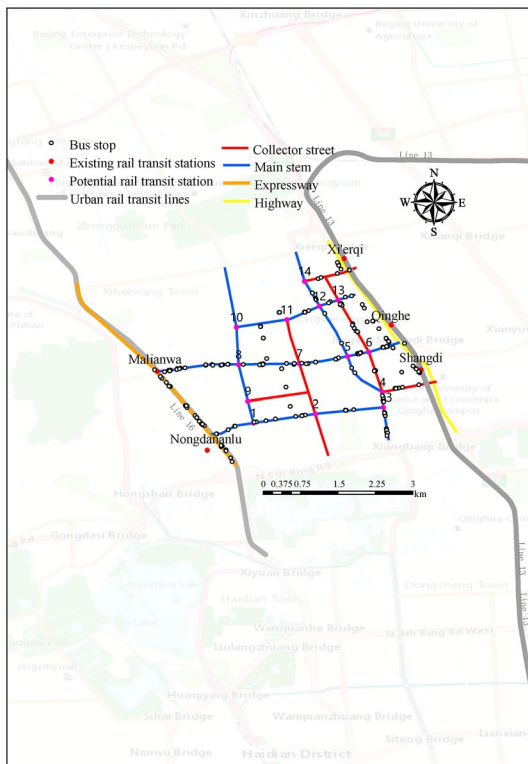


**Fig. 4** Land use characteristics in the area

line generation model constructed in this study. The location diagram of the research area with respect to the whole URT network is shown in Fig. 2. In order to reduce the complexity of the model, the part of the interaction between Line 16 and Line 13 in the region is selected for example analysis. The research area is shown in Fig. 3.

The land use properties within the region are shown in Fig. 4, and the urban development maturity is relatively high. A number of high schools and universities are located in the area. Many technological parks and office areas are located along the highway. A dense residential area is present within the region. In addition, there are large shopping malls and a large integrated transportation hub with suburban railways, subways, and high-speed railways in the area. In general, the area is densely populated and has a high demand for travel.

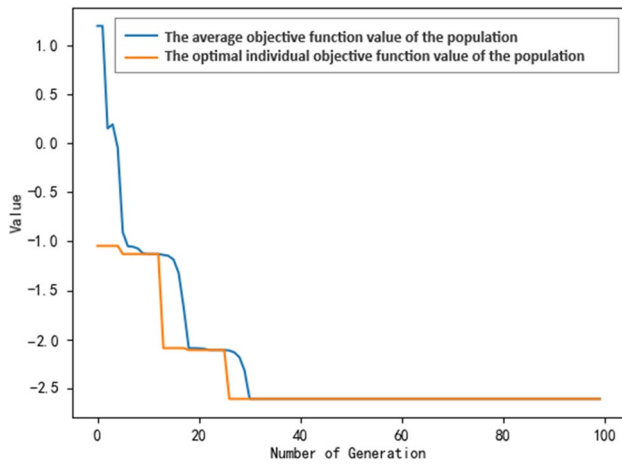
URT stations according to their location and role in the network can be divided into the intermediate stations and transfer stations. Transfer stations are larger than typical intermediate stations, and the investment is usually larger. Referring to the existing document [26], the investment of an intermediate station is taken as 100 million yuan, the investment of a transfer station is taken as 150 million yuan, and the cost of intervals is 1 billion yuan per kilometer.



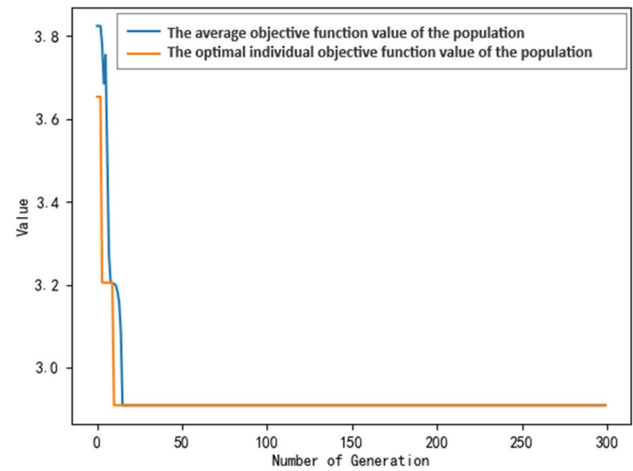
**Fig. 5** Distribution of bus stations and URT stations in the region

**Table 1** URT passenger flow in candidate stations

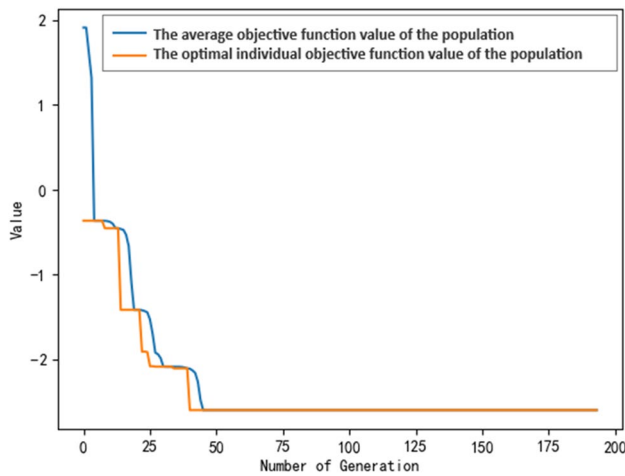
Station	Subway traffic within 800 m/ person
Nongdananlu	4044
Malianwa	5303
1	2470
2	5423
3	4038
4	2776
5	1929
6	1497
7	1038
8	1055
9	175
10	5986
11	6184
12	1224
13	1988
14	7987
Shangdi	14,901
Qinghe	19,115
Xi'erqi	71,993



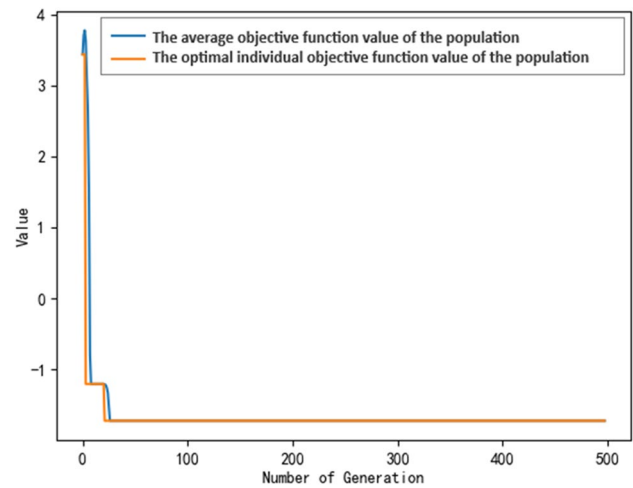
**Fig. 6** Evolution of optimal/average objective function value for each generation of population (100 generations)



**Fig. 8** Evolution of optimal/average objective function value for each generation of population (300 generations)



**Fig. 7** Evolution of optimal/average objective function value for each generation of population (200 generations)



**Fig. 9** Evolution of optimal/average objective function value for each generation of population (500 generations)

### 5.2 Selecting Alternative Sites

Based on the method above, alternative stations are selected through the urban road network and the conventional bus network. As shown in Fig. 5, there are 12 secondary arterial roads in the region. Dense white dots represent bus stations, red dots represent existing URT stations, and pink dots represent potential stations. Finally, an alternative set of 19 stations including five existing URT stations is formed. It should be noted that the Nongdananlu station is represented as 1 in the algorithm and analysis, which is different from the representation in the figure.

Demand for rail traffic at existing stations can be obtained from URT card-reading data. The rail traffic of the potential station is obtained by discounting the all-mode traffic around the station according to a certain percentage. The proportion

**Table 2** Analysis of results under different evolutionary generations

Evolutionary generations	Converge or not	Optimal objective function value	Optimal solution
100	Yes	-2.85582	2-10-12-13-16-19
200	Yes	-2.85582	2-10-12-13-16-19
300	Yes	2.86139	1-4-5-17
500	Yes	-1.77252	2-9-13-16-19

can be obtained by referring to the URT trip share rate in the Annual Report on Transportation Development in Beijing 2021. To avoid the impact of the COVID-19 pandemic, the 2019 trip share of 16.5% is taken as the conversion factor for

all-mode passenger flow to rail traffic. Finally, the average daily rail transit passenger flow for each candidate station is shown in Table 1.

### 5.3 Results and Analysis

#### 5.3.1 Algorithm Analysis

The proposed algorithm is a modified algorithm based on the GA. Therefore, the process of searching for the optimal solution has some similarities with the GA. If the evolutionary generation is too small, the algorithm does not converge easily, and the population is not yet mature. On the other hand, if the evolutionary generation is too large, the algorithm is already proficient or the population is too premature to converge, and it is meaningless to continue the evolution.

In order to explore the convergence problem of the single-objective line generation model and set up appropriate evolutionary generation, we first designed four experimental schemes of 100 generations, 200 generations, 300 generations, and 500 generations according to a weight of 1:1 and conducted several experiments. The evolution of the average objective function value of the population and the optimal objective function value of the population along with the evolutionary generations under the four typical schemes are shown in Figs. 6, 7, 8, and 9, respectively. The orange line

represents the population optimal individual objective function value, and the blue line represents the average individual objective function value. The optimal objective function value and the optimal route under each scheme are shown in Table 2.

By combining Table 2 and the evolution process of the optimal objective function value of the population and the average objective function value of the population under the four schemes, it can be found that when the evolutionary generation is set to 100 or 200 generations, the results converge around the 35th and 40th generations, respectively, and the optimization results are the same. When the generation of evolution increases to 300 or 500 generations, the evolution process converges in less than 20 generations, indicating that it is too premature to continue the evolution of the situation. Therefore, the number of evolutionary generations set to 100 or 200 is more suitable for the model in this research. Several experimental results show that the single-objective line generation model converges within 50 generations in most cases, and the latest convergence is within 60 generations. This indicates that the constructed E-GA with nested passenger flow allocation method is fast and effective for solving the model in this study.

The optimized “best individual” in the four cases is 2-10-12-13-16-19, which is translated as Malianwa–station 8–station 10–station 11–station 14–Xi’erqi, and the objective function value is  $-2.85582$ . Fig. 10 shows the optimized line in the study area.

#### 5.3.2 Analysis of Different Weights

After determining that the reasonable evolutionary generation of the model constructed in this study is 100 or 200 generations, the optimal lines under different weights were

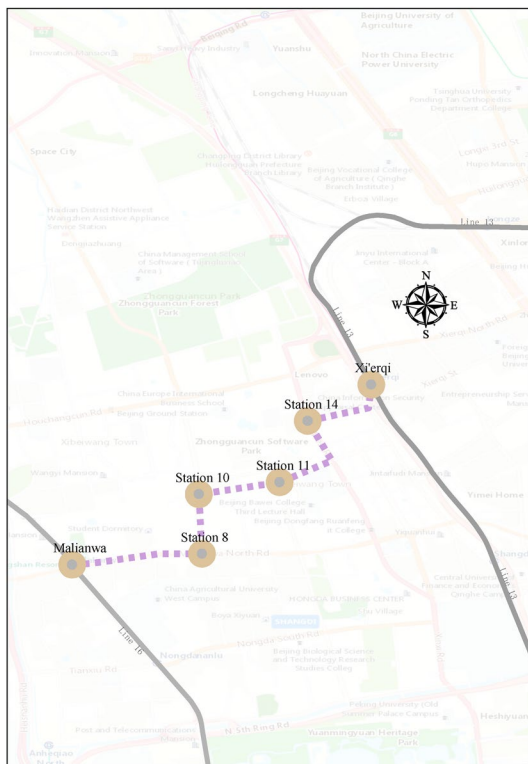


Fig. 10 Schematic of the optimal scheme for the model

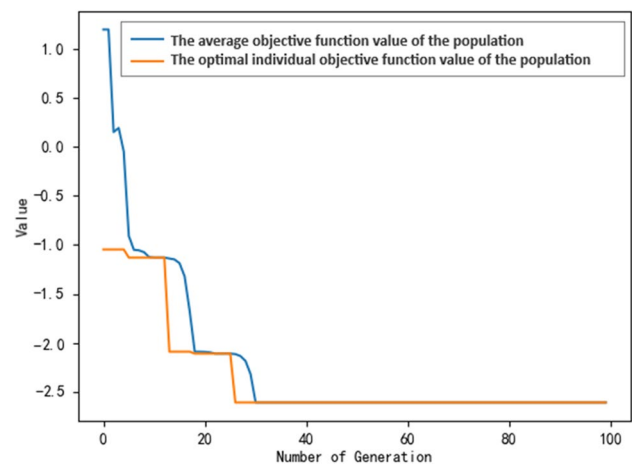
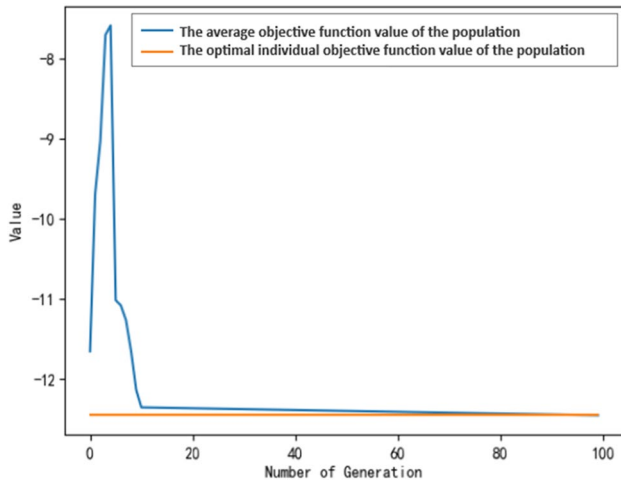


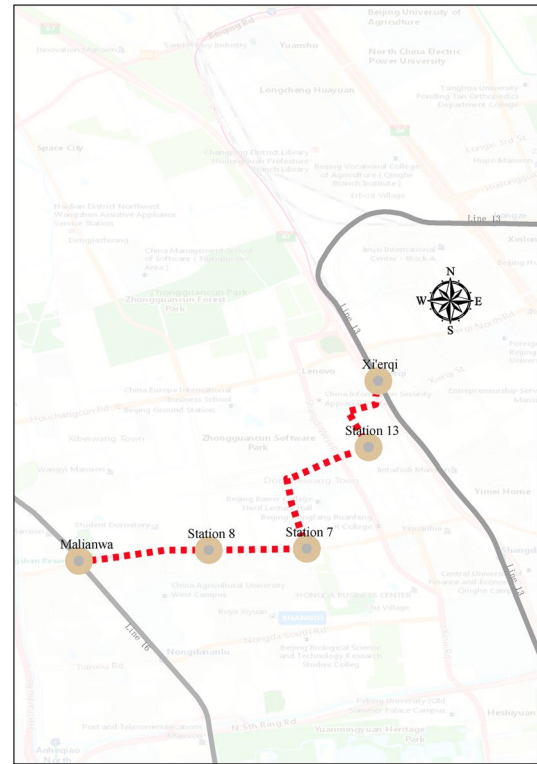
Fig. 11 Evolution of optimal/average objective function value of population with weight 1:1



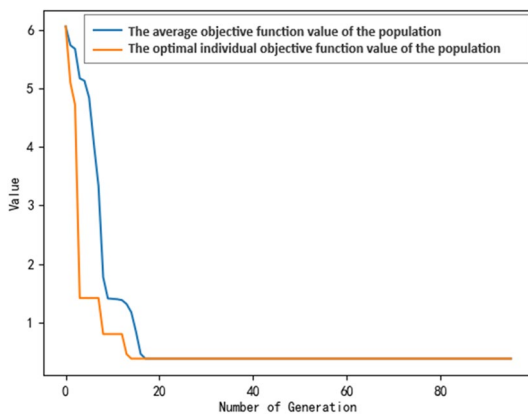
**Fig. 12** Evolution of optimal/average objective function value of population with weight 2:1

**Table 3** Optimal route scheme and objective function value under different weights

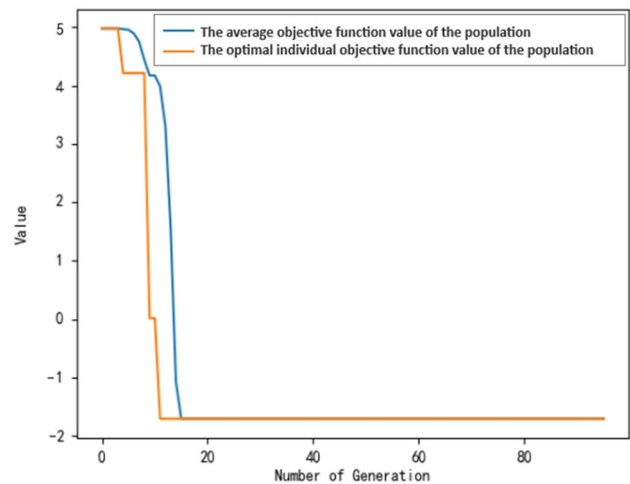
Weight	Route plan	Objective function value
1:1	2-10-12-13-16-19	-2.85547
2:1	2-10-12-13-16-19	-12.70647



**Fig. 14** Schematic of optimal route scheme under the constraint of station 7



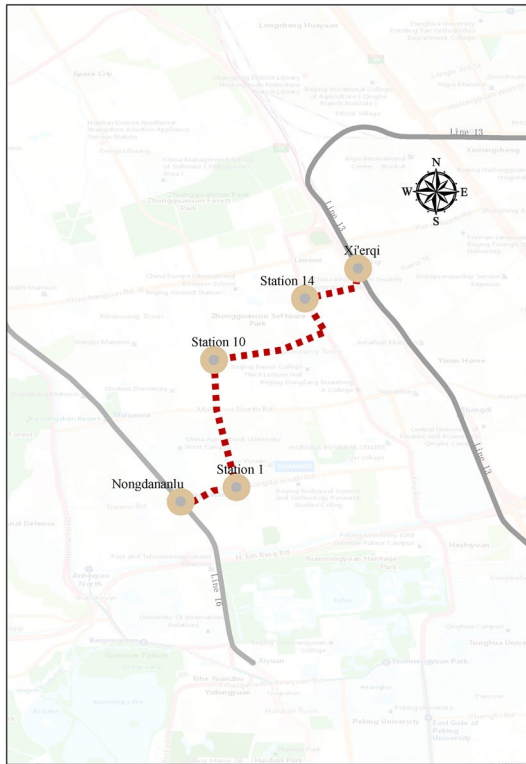
**Fig. 13** Evolution of optimal/average objective function value of population with site constraints



**Fig. 15** Evolution of optimal/average objective function value of population with origin–destination constraints

studied. The weight settings are based on the practical requirements of the problem and can be readily adjusted according to the decision-maker’s needs. As URT networks are essential for public welfare, this paper primarily considers two scenarios: maximum number of people served and minimum project cost are equally important, and maximum number of people served is more important than minimum project cost. For these two scenarios, we design two cases

with weight ratios of 1:1 and 2:1. Generally, since the total weight sums to 1, the weights are set as 1/2 and 1/2, as well as 2/3 and 1/3. This takes into account the impact of different weights on the results, providing decision-makers with flexibility in choosing weights.



**Fig. 16** Schematic of optimal route scheme under origin and destination constraints

The evolution of the average objective function value of the population and the optimal objective function value of the population along with the evolutionary generation under the two schemes are shown in Figs. 11 and 12, respectively. The optimal objective function value and the optimal route scheme under each scheme are shown in Table 3.

As can be seen from the above figure, regardless of which weight method is used, the algorithm converges within 100 generations. The optimal route optimization scheme for 1:1 and 2:1 weight is 2-10-12-13-16-19, and the location diagram is shown in Fig. 10.

### 5.3.3 Constraints on Different Required Sites

In addition, we designed two experimental schemes with weight of 1:1 and evolution of 100 generations, i.e., experiments with mandatory station constraints and experiments with fixed starting and finishing station constraints.

Station 7 is taken as the necessary station, and the proposed algorithm is adopted for calculation. The evolution diagram of population optimal/average objective function value is shown in Fig. 13. The optimal route scheme is 2-10-9-15-19, and the objective function value is 0.3807. The location diagram is shown in Fig. 14.

Nongdananlu station and Xi'erqi station are taken as the fixed OD constraints. The evolution of the population optimal/average objective function values is obtained, as shown in Fig. 15. The optimal route scheme is 1-3-12-16-19, and the objective function value is  $-1.51897$ . The location diagram is shown in Fig. 16.

## 5.4 Policy Suggestions for Resolving Bottlenecks

Resolving bottlenecks is beneficial for strengthening the overall URT capacity, improving transportation services, and enhancing operational management efficiency. This study proposes policy recommendations for resolving bottlenecks from three perspectives: network structure, planning and operations, and industrial layout.

### 5.4.1 Enhancing Transport Coordination

**Expand the URT network:** Consider adding new lines or branch lines to the URT network to cover underserved areas and ensure that more areas can benefit from URT services.

**Coordinate public transportation services:** Coordinate the public transportation system such as bus system with the URT network to meet the travel demands in the region and enhance the synergy between the two.

### 5.4.2 Focusing on Planning and Operations

Relevant planning, design, and operations departments should focus on the underlying causes of bottleneck areas. In future planning and design processes, these issues should be addressed to improve service quality.

### 5.4.3 Optimizing Industrial Layout

Arrange industries or chains of industries that are highly reliant on URT commuting on the same rail line. This helps increase rail passenger flow, especially in areas with low passenger volume on the URT network periphery, thus maximizing URT utilization.

## 6 Conclusions

As the URT network in metropolises like Beijing continues to expand, the focus has shifted from construction to renovation and improvement. The transformation and enhancement of existing URT networks have become urgent requirements. However, research on optimizing existing URT network operations lacks attention to actual operational conditions, and only a few studies have considered optimizing existing networks under the constraints of a large-scale URT system.

This study proposes a method for selecting potential URT stations in bottleneck areas and constructs a line generation model to optimize and improve the routes in these bottleneck areas. Additionally, a GA method with nested passenger distribution is designed to solve this problem. Taking the existing URT network in Beijing as an example, the study analyzes the interconnections between local networks to optimize them. The results show that the algorithm converges within 100 iterations, indicating its rapid and effective performance. The generated solutions can meet regional travel demands and optimize the relationships within the URT network. To enhance the attractiveness of URT, this study also provides some policy recommendations.

Although this study has made certain progress in the theory and practice of optimizing existing URT networks, there are still some issues that need to be resolved due to a lack of knowledge and time. Considering the complexity of the research problem, some factors have been simplified in the modeling process. For example, assumptions have been made to simplify the analysis, population coverage within an 800-m range has been used as a proxy for passenger flow indicators, and a shortest path-based all-or-none passenger distribution method has been employed, without taking into account factors such as transfer times and congestion levels that may affect route selection. In the next phase of research, it is necessary to consider more comprehensive factors and indicators and establish more accurate models. It is also important to compare the performance of the proposed method with other algorithms.

were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## Appendix A. Lower-Level Submodel Linearization

### Lower-Level Objective Function Linearization

As stated in Sect. 3, at this level, the network design variables are fixed according to the values given by the current network. However, the objective function contains two nonlinear terms, variable operation cost and rolling stock acquisition cost, as a consequence of the product of the fleet size (FS) variables for each line, and the train model choice variables for each line  $\delta_\ell^m$ ,

$$-\lambda h_{\text{year}} \sum_{\ell \in \hat{\mathcal{L}}} FS_\ell \left( \sum_{m \in M} \delta_\ell^m c_{\text{train}}^m \right),$$

and

$$-\frac{\chi}{(t_f - t_o)} \sum_{\ell \in \hat{\mathcal{L}}} FS_\ell \left( \sum_{m \in M} I_{\text{train}}^m \delta_\ell^m \right).$$

A new set of binary variables  $\alpha_\ell^m$  is introduced in order to linearize these products, obtaining a linear objective function expression:

$$\text{Maximize } \sum_{k=t_o}^{t_f-1} \frac{1}{e^{rk}} \left[ (\eta + \xi) h_{\text{year}} \sum_{w \in \hat{W}} g_w f_w^{\text{RRT}} - \lambda h_{\text{year}} \sum_{\ell \in \hat{\mathcal{L}}} \left( \sum_{m \in M} c_{\text{train}}^m \alpha_\ell^m \right) - c_{\text{crew}} \sum_{\ell \in \hat{\mathcal{L}}} FS_\ell - \frac{\chi}{(t_f - t_o)} \sum_{\ell \in \hat{\mathcal{L}}} \left( \sum_{m \in M} I_{\text{train}}^m \alpha_\ell^m \right) \right]. \quad (10)$$

**Acknowledgements** This work was supported by the Fundamental Research Funds for the Central Universities (2022JBZY039).

#### Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes

The next new sets of constraints have to be included:

$$FS_\ell - \mathcal{C}_1 (1 - \delta_\ell^m) \leq \alpha_\ell^m, \quad \ell \in \hat{\mathcal{L}}, m \in M, \quad (11)$$

$$\alpha_\ell^m \leq FS_\ell, \quad \ell \in \hat{\mathcal{L}}, m \in M, \quad (12)$$

$$\alpha_\ell^m \leq \mathcal{C}_1 \delta_\ell^m, \quad \ell \in \hat{\mathcal{L}}, m \in M, \quad (13)$$

where

$$\mathcal{C}_1 = \frac{2\mathcal{H}^{\max}}{\lambda} \sum_{\{i,j\} \in \hat{E}} d_{ij} \bar{x}_{ij}^\ell, \quad \ell \in \hat{\mathcal{L}}$$

represents the maximum allowed value for the fleet size of line  $\ell \in \hat{\mathcal{L}}$ , and  $\bar{x}_{ij}^\ell$  are the fixed values of the network design variables at the current iteration.

### Generalized Cost Linearization

In order to linearize constraints (18), (19), (20), and (21), we proceed as follows. Constraints (21) are multiplied by  $f_w^{\text{RRT}}$ , yielding

$$f_w^{\text{RRT}} U_w^{\text{RRT}} = f_w^{\text{RRT}} \eta + \beta_{tw} f_w^{\text{RRT}} u_w^{\text{RRT},tt} + \beta_{tr} f_w^{\text{RRT}} u_w^{\text{RRT},tr} + \beta_{wt} f_w^{\text{RRT}} u_w^{\text{RRT},tw}, \quad w \in \hat{W}. \tag{14}$$

(1) The linearization process starts by considering the right-hand side of (14). For a given  $\eta$ , the value of the first term is linear. The second term  $f_w^{\text{RRT}} u_w^{\text{RRT},tt}$ , corresponding to the ride time, now becomes linear. After multiplying (21) by  $f_w^{\text{RRT}}$ , constraints (18) become

$$f_w^{\text{RRT}} u_w^{\text{RRT},tt} = \frac{60}{\lambda} \sum_{\ell \in \hat{\mathcal{L}}} \sum_{(i,j) \in \hat{A}(\ell)} f_i^{w\ell} d_{ij}, \quad w \in \hat{W}. \tag{15}$$

We can replace the term  $f_w^{\text{RRT}} u_w^{\text{RRT},tt}$  in (14) by (15), thus removing constraints (18) from the linearized model.

(2) The third term on the right-hand side of (14), the result of multiplying (19) by  $f_w^{\text{RRT}}$ , is linearized as follows:

$$\begin{aligned} f_w^{\text{RRT}} u_w^{\text{RRT},tr} &= \sum_{\ell \in \hat{\mathcal{L}}} \sum_{\ell' \in \hat{\mathcal{L}}} \sum_{i \in \hat{N}(\ell, \ell')} f_i^{w\ell\ell'} \left( \frac{\zeta_{\ell'}}{2} + uc_i \right) \\ &\quad \ell' \neq \ell \\ &= \sum_{\ell \in \hat{\mathcal{L}}} \sum_{\ell' \in \hat{\mathcal{L}}} \sum_{i \in \hat{N}(\ell, \ell')} f_i^{w\ell\ell'} \frac{\zeta_{\ell'}}{2} + \sum_{\ell \in \hat{\mathcal{L}}} \sum_{\ell' \in \hat{\mathcal{L}}} \sum_{i \in \hat{N}(\ell, \ell')} f_i^{w\ell\ell'} uc_i \\ &\quad \ell' \neq \ell \qquad \qquad \qquad \ell' \neq \ell \\ &= \frac{1}{2} \sum_{\ell \in \hat{\mathcal{L}}} \sum_{\ell' \in \hat{\mathcal{L}}} \sum_{\substack{i \in \hat{N}(\ell, \ell') \\ \ell' \neq \ell}} \mathcal{H}^p f_i^{w\ell\ell'} \gamma_{\ell'}^p \\ &\quad + \sum_{\ell \in \hat{\mathcal{L}}} \sum_{\ell' \in \hat{\mathcal{L}}} \sum_{i \in \hat{N}(\ell, \ell')} f_i^{w\ell\ell'} uc_i, \quad w \in \hat{W}, \\ &\quad \ell' \neq \ell \end{aligned}$$

where the variables  $\zeta_{\ell'}$  in the first summation are first expressed using the set of admissible headway values and the binary variables  $\gamma_{\ell'}^p$ , used to select a specific headway value for each line  $\ell \in \hat{\mathcal{L}}$ :

$$\zeta_{\ell} = \sum_{p \in \mathcal{H}} \mathcal{H}^p \gamma_{\ell}^p, \quad \ell \in \hat{\mathcal{L}}, \tag{16}$$

where

$$\sum_{p \in \mathcal{H}} \gamma_{\ell}^p = 1, \quad \ell \in \hat{\mathcal{L}}. \tag{17}$$

It is also convenient to define the line frequencies using the variables  $\gamma_{\ell'}^p$ .

$$\psi_{\ell} = \sum_{p \in \mathcal{F}} \mathcal{F}^p \gamma_{\ell}^p, \quad \ell \in \hat{\mathcal{L}}. \tag{18}$$

The new products  $f_i^{w\ell\ell'} \gamma_{\ell'}^p$  are then replaced with a new set of variables  $\chi_{ip}^{w\ell\ell'}$ , obtaining for the third term in (14) the equivalent linear formulation:

$$\begin{aligned} f_w^{\text{RRT}} u_w^{\text{RRT},tr} &= \frac{1}{2} \sum_{\ell \in \hat{\mathcal{L}}} \sum_{\ell' \in \hat{\mathcal{L}}} \sum_{i \in \hat{N}(\ell, \ell')} \sum_{p \in \mathcal{H}} \mathcal{H}^p \chi_{ip}^{w\ell\ell'} \\ &\quad \ell' \neq \ell \\ &\quad + \sum_{\ell \in \hat{\mathcal{L}}} \sum_{\ell' \in \hat{\mathcal{L}}} \sum_{i \in \hat{N}(\ell, \ell')} f_i^{w\ell\ell'} uc_i, \quad w \in \hat{W}, \\ &\quad \ell' \neq \ell \end{aligned}$$

These expressions are included in (14) by replacing the product  $f_w^{\text{RRT}} u_w^{\text{RRT},tr}$ , and consequently, constraints (19) are removed from the model. In order to represent the behavior of variables  $\chi_{ip}^{w\ell\ell'}$ , the next sets of constraints have to be added to the linearized model:

$$f_i^{w\ell\ell'} - (1 - \gamma_{\ell'}^p) \leq \chi_{ip}^{w\ell\ell'}, \quad w \in \hat{W}, \ell, \ell' \in \hat{\mathcal{L}}, i \in \hat{N}(\ell, \ell'), p \in \mathcal{H}, \tag{19}$$

$$\chi_{ip}^{w\ell\ell'} \leq f_i^{w\ell\ell'}, \quad w \in \hat{W}, \ell, \ell' \in \hat{\mathcal{L}}, i \in \hat{N}(\ell, \ell'), p \in \mathcal{H}, \tag{20}$$

$$\chi_{ip}^{w\ell\ell'} \leq \gamma_{\ell'}^p, \quad w \in \hat{W}, \ell, \ell' \in \hat{\mathcal{L}}, i \in \hat{N}(\ell, \ell'), p \in \mathcal{H}. \tag{21}$$

(3) The fourth term on the right-hand side of (14),  $f_w^{\text{RRT}} u_w^{\text{RRT},tw}$ , corresponding to the product of (20) by  $f_w^{\text{RRT}}$ , is linearized as follows:

$$\begin{aligned} f_w^{\text{RRT}} u_w^{\text{RRT},tw} &= \frac{1}{2} \sum_{\ell \in \hat{\mathcal{L}}} \sum_{j \in \hat{S}(w^o)} \zeta_{\ell} f_{w^o j}^{w\ell} \\ &= \frac{1}{2} \sum_{\ell \in \hat{\mathcal{L}}} \sum_{j \in \hat{S}(w^o)} \sum_{p \in \mathcal{H}} \mathcal{H}^p f_{w^o j}^{w\ell} \gamma_{\ell}^p, \quad w \in \hat{W}. \end{aligned}$$

The headways  $\zeta_{\ell}$  are first replaced with Eq. (16), and a new set of variables  $\sigma_{pj}^{w\ell}$  representing the products  $f_{w^o j}^{w\ell} \gamma_{\ell}^p$  is then introduced, yielding the equivalent linear formulation:

$$f_w^{\text{RRT}} u_w^{\text{RRT},tw} = \frac{1}{2} \sum_{\ell \in \hat{\mathcal{L}}} \sum_{j \in \hat{S}(w^o)} \sum_{p \in \mathcal{H}} \mathcal{H}^p \sigma_{pj}^{w\ell}, \quad w \in W.$$

These equations replace the fourth term on the right-hand side of (14). In addition, constraints (20) are removed from the linearized model. In order to represent the behavior of variables  $\sigma_{pj}^{w\ell}$ , the next sets of constraints have to be included in the linearized model:

$$\sigma_{pj}^{w\ell} \leq f_{w^o j}^{w\ell}, \quad w \in \hat{W}, \ell \in \hat{\mathcal{L}}, p \in \mathcal{H}, j \in \hat{S}(w^o), \quad (22)$$

$$f_{w^o j}^{w\ell} - (1 - \gamma_p^\ell) \leq \sigma_{pj}^{w\ell}, \quad w \in \hat{W}, \ell \in \hat{\mathcal{L}}, p \in \mathcal{H}, j \in \hat{S}(w^o), \quad (23)$$

$$\sigma_{pj}^{w\ell} \leq \gamma_p^\ell, \quad w \in \hat{W}, \ell \in \hat{\mathcal{L}}, p \in \mathcal{H}, j \in \hat{S}(w^o). \quad (24)$$

(4) To linearize the left-hand side of Eq. (14),  $f_w^{\text{RRT}}$  is discretized considering a set of real values equally distributed in the interval  $[0,1]$ , denoted as  $\mathcal{J}$ . For instance,  $\mathcal{J}$  could be the set  $\{0, 0.005, 0.01, \dots, 0.095, 1\}$ , where each element will be represented by  $\mathcal{J}^i, i = 1, \dots, |\mathcal{J}|$ . A new set of binary variables  $\rho_{iw}, i = 1, \dots, |\mathcal{J}|, w \in \hat{W}$ , is considered. Then,

$$f_w^{\text{RRT}} = \sum_{i \in \mathcal{J}} \mathcal{J}^i \rho_{iw}, \quad w \in \hat{W}, \quad (25)$$

where

$$\sum_{i \in \mathcal{J}} \rho_{iw} = 1, \quad w \in \hat{W}. \quad (26)$$

The left-hand side of (14) then becomes:

$$f_w^{\text{RRT}} U_w^{\text{RRT}} = \sum_{i \in \mathcal{J}} \mathcal{J}^i \rho_{iw} U_w^{\text{RRT}}, \quad w \in \hat{W}.$$

The products  $\rho_{iw} U_w^{\text{RRT}}$  are replaced with new variables  $\varsigma_{iw}$ , so that

$$f_w^{\text{RRT}} U_w^{\text{RRT}} = \sum_{i \in \mathcal{J}} \mathcal{J}^i \varsigma_{iw}, \quad w \in \hat{W}.$$

New sets of constraints have to be included in the linearized model to achieve the correct behavior of the new variables.

$$\varsigma_{iw} \leq U_w^{\text{RRT}}, \quad w \in \hat{W}, i \in \mathcal{H}, \quad (27)$$

$$U_w^{\text{RRT}} - \mathcal{C}_2^w (1 - \rho_{iw}) \leq \varsigma_{iw}, \quad w \in \hat{W}, i \in \mathcal{H}, \quad (28)$$

$$\varsigma_{iw} \leq \mathcal{C}_2^w \rho_{iw}, \quad w \in \hat{W}, i \in \mathcal{H}, \quad (29)$$

where  $\mathcal{C}_2^w$  is an upper bound of  $U_w^{\text{RRT}}$  that can be set to the next value independently of the pair considered,

$$\mathcal{C}_2^w = \mathcal{C}_2 = \eta + \beta_{tr} \frac{60}{\lambda} \sum_{\ell \in \hat{\mathcal{L}}} \sum_{(ij) \in \hat{\mathcal{A}}(\ell)} d_{ij} + \beta_{tr} |\hat{\mathcal{L}}| + \beta_{wt} \frac{\mathcal{H}^{\max}}{2}, \quad w \in \hat{W}.$$

Note that the discretized expression of  $f_w^{\text{RRT}}$  in Eq. (15) gives rise to an approximation of Eq. (14), so that the maximum discrepancy between the left- and right-hand sides will depend on the length of intervals between consecutive values in  $\mathcal{J}$ , i.e.,

$$f_w^{\text{RRT}} U_w^{\text{RRT}} \approx f_w^{\text{RRT}} \eta + \beta_{tr} f_w^{\text{RRT}} U_w^{\text{RRT},tr} + \beta_{tr} f_w^{\text{RRT}} U_w^{\text{RRT},tr} + \beta_{wt} f_w^{\text{RRT}} U_w^{\text{RRT},tw}, \quad w \in \hat{W}.$$

Now, introducing the linearized version of the left- and right-hand sides of (14), and taking into account the possible discrepancy as a consequence of using a discrete set of feasible values for  $f_w^{\text{RRT}}$ , Eq. (14) is replaced with two new sets of constraints:

$$\sum_{i \in \mathcal{J}} \mathcal{J}^i \varsigma_{iw} - \epsilon \leq f_w^{\text{RRT}} \eta + \beta_{tr} \frac{60}{\lambda} \sum_{\ell \in \hat{\mathcal{L}}} \sum_{(ij) \in \hat{\mathcal{E}}} f_{ij}^{w\ell} d_{ij} + \beta_{tr} \left( \frac{1}{2} \sum_{\ell \in \hat{\mathcal{L}}} \sum_{\ell' \in \hat{\mathcal{L}}} \sum_{i \in \hat{\mathcal{N}}} \sum_{p \in \mathcal{H}} \mathcal{H}^p \chi_{ip}^{w\ell\ell'} + \sum_{\ell \in \hat{\mathcal{L}}} \sum_{\ell' \in \hat{\mathcal{L}}} \sum_{i \in \hat{\mathcal{N}}} f_i^{w\ell\ell'} u_{ci} \right)_{\ell' \neq \ell} + \beta_{wt} \frac{1}{2} \sum_{\ell \in \hat{\mathcal{L}}} \sum_{j: \{w^o, j\} \in \hat{\mathcal{E}}} \sum_{p \in \mathcal{H}} \mathcal{H}^p \sigma_{pj}^{w\ell}, w \in \hat{W} \quad (30)$$

$$\sum_{i \in \mathcal{J}} \mathcal{J}^i \varsigma_{iw} + \epsilon \geq f_w^{\text{RRT}} \eta + \beta_{tr} \frac{60}{\lambda} \sum_{\ell \in \hat{\mathcal{L}}} \sum_{(ij) \in \hat{\mathcal{E}}} f_{ij}^{w\ell} d_{ij} + \beta_{tr} \left( \frac{1}{2} \sum_{\ell \in \hat{\mathcal{L}}} \sum_{\ell' \in \hat{\mathcal{L}}} \sum_{i \in \hat{\mathcal{N}}} \sum_{p \in \mathcal{H}} \mathcal{H}^p \chi_{ip}^{w\ell\ell'} + \sum_{\ell \in \hat{\mathcal{L}}} \sum_{\ell' \in \hat{\mathcal{L}}} \sum_{i \in \hat{\mathcal{N}}} f_i^{w\ell\ell'} u_{ci} \right)_{\ell' \neq \ell} + \beta_{wt} \frac{1}{2} \sum_{\ell \in \hat{\mathcal{L}}} \sum_{j: \{w^o, j\} \in \hat{\mathcal{E}}} \sum_{p \in \mathcal{H}} \mathcal{H}^p \sigma_{pj}^{w\ell}, w \in \hat{W} \quad (31)$$

for a certain non-negative value of  $\epsilon$  which can be experimentally fixed or included in the lower-level objective function conveniently penalized in order to ensure feasibility.

### Linearization of the Modal Split

The proportion of the demand captured by the RRT,  $f_w^{\text{RRT}}$ , bounded by constraint (17), is computed using a nonlinear logit function that compares, for each pair  $w \in \hat{W}$ , the generalized cost of using the RRT network and the alternative transportation mode. In order to replace the logit function, we consider a piecewise linear approximation following a multiple-choice modeling schema (Jeroslow and Lowe 1984) as follows. Let  $z = U_w^{\text{ALT}} - U_w^{\text{RRT}}$  be the variable representing the difference between the generalized cost of the competing mode and that of the RRT mode, and let  $\mathbf{Lf}(z) = 1/(1 + \exp(-\beta z))$  be the logit function written in terms of  $z$ . We define a uniform partition of the  $z$  axis by means of a set of real points  $\{z_1, z_2, \dots, z_k, z_{k+1}, z_{k+2}, \dots, z_{2k+1}\}$ , so that  $z_{k+1} = 0$ . Therefore, the uniform partition  $\mathcal{I}$  is defined by a total of  $2k + 2$  intervals:  $\mathcal{I} = \{\mathcal{I}_0, \mathcal{I}_1, \mathcal{I}_2, \dots, \mathcal{I}_k, \mathcal{I}_{k+1}, \mathcal{I}_{k+2}, \dots, \mathcal{I}_{2k}, \mathcal{I}_{2k+1}\}$ ,



where  $\mathcal{I}_0 = (-\infty, z_1), \mathcal{I}_s = [z_s, z_{s+1}), s = \{1, \dots, 2k\}$  and  $\mathcal{I}_{2k+1} = [z_{2k+1}, \infty)$ . For the sake of readability, interval  $\mathcal{I}_s$  will be represented by its index  $s$  if this creates no confusion. In order to represent the possible values of  $z$  taking into account the intervals in partition  $\mathcal{I}$ , for each OD pair  $w \in \hat{W}$ , we introduce two new sets of variables: a set of free variables  $U_w^s$  that will represent the value of  $z$  depending on the specific interval ( $U_w^s = z$  if  $z \in \mathcal{I}_s, s \in \mathcal{I}$ ), and a new set of binary variables  $\mu_w^s$  equal to 1 if  $z \in \mathcal{I}_s, s \in \mathcal{I}$ . The selection of the appropriate interval in the  $z$  axis is governed by the following sets of constraints:

$$U_w^{\text{RRT}} - U_w^{\text{RRT}} = \sum_{s \in \mathcal{I}} U_w^s, \quad w \in \hat{W}, \tag{32}$$

$$U_w^0 \leq z_1 \mu_w^0, \quad w \in \hat{W}, \tag{33}$$

$$z_s \mu_w^s \leq U_w^s \leq z_{s+1} \mu_w^s, \quad w \in \hat{W}, \quad s \in \mathcal{I}, \quad s \neq \{0, 2k+1\}, \tag{34}$$

$$z_{2k+1} \mu_w^{2k+1} \leq U_w^{2k+1}, \quad w \in \hat{W}, \tag{35}$$

$$\sum_{s \in \mathcal{I}} \mu_w^s = 1, \quad w \in \hat{W}. \tag{36}$$

In addition, the value of  $\mathbf{Lf}(z)$ , which acts as an upper bound of the fraction of demand captured by the RRT network for each OD pair, corresponds to the right-hand side of the following sets of constraints:

$$f_w^{\text{RRT}} \leq 1 + \sum_{s=1}^{2k} \mathbf{Lf}(z_s) \mu_w^s + m_s (U_w^s - z_s \mu_w^s), \quad w \in \hat{W}, \tag{37}$$

where  $m_s = \mathbf{Lf}(z_{s+1}) - \mathbf{Lf}(z_s) / z_{s+1} - z_s, s = 1, \dots, 2k$ , is the slope of the line that approximates  $\mathbf{Lf}(z)$  at segment  $s$ . Constraints (37) replace old constraints (17).

### Capacity and Fleet Size Linearization

Using Eqs. (18), the right-hand side of constraints (22) can be reformulated as

$$\sum_{w \in \hat{W}} f_{ij}^{wl} g_w \leq \sum_{p \in \mathcal{F}} \sum_m K_{\text{train}}^m \xi_p^{ml}, \quad \ell \in \hat{\mathcal{L}}, \tag{38}$$

where variables  $\xi_p^{ml}$  represent the product  $\delta_m \gamma_\ell^p$ . The following two set of constraints must be added:

$$\delta_m + \gamma_\ell^p \leq 1 + \xi_p^{ml}, \quad \ell \in \hat{\mathcal{L}}, p \in P, m \in M, \tag{39}$$

$$2 \xi_p^{ml} \leq \delta_m + \gamma_\ell^p, \quad \ell \in \hat{\mathcal{L}}, p \in P, m \in M. \tag{40}$$

Concerning the fleet size of each line, constraints (25), since  $\text{FS}_\ell$  can only take integer values which are closely related to the value of headways  $\text{FS}_\ell$ , taking advantage of Eqs. (16), the fleet size of each line can be expressed as

$$\begin{aligned} \text{FS}_\ell &= \sum_{p \in \mathcal{H}} \frac{120}{\lambda \mathcal{H}^p} \left[ \sum_{\{ij\} \in \hat{E}} d_{ij} \bar{x}_{ij}^\ell \right] \gamma_\ell^p \\ &= \sum_{p \in \mathcal{F}} \frac{2 \mathcal{F}^p}{\lambda} \left[ \sum_{\{ij\} \in \hat{E}} d_{ij} \bar{x}_{ij}^\ell \right] \gamma_\ell^p, \quad \ell \in \hat{\mathcal{L}}, \end{aligned} \tag{41}$$

where the variables  $\bar{x}_{ij}^\ell$  correspond to the fixed values of network design variables at the current iteration.

### Frequency on Shared Segments

Although in the INDLPP formulation, constraints (24) are nonlinear, once fixed in a network solution, they becomes linear and no further linearization is needed.

### The Lower-Level Submodel

The lower-level submodel can then be finally defined as follows:

Maximize (10)  
subject to (14)–(16), (23), (24), (11)–(13), (16)–(41),  
where

$$f_{ij}^{w\ell} \in [0, 1], \quad w \in \hat{W}, \ell \in \hat{\mathcal{L}}, (i, j) \in \hat{A}(\ell),$$

$$f_w^{\text{RRT}} \in [0, 1], \quad w \in \hat{W},$$

$$f_i^{w\ell\ell'} \in [0, 1], \quad w \in \hat{W}, \ell, \ell' \in \hat{\mathcal{L}}, i \in \hat{N}(\ell, \ell'), i \neq \{w^o, w^d\},$$

$$\psi_\ell \in \{\mathcal{F}^1, \mathcal{F}^2, \dots, \mathcal{F}^{|\mathcal{F}|}\},$$

$$\zeta_\ell \in \{\mathcal{H}^1, \mathcal{H}^2, \dots, \mathcal{H}^{|\mathcal{H}|}\},$$

$$\delta_\ell^m \in \{0, 1\}, \quad \ell \in \hat{\mathcal{L}}, \quad m \in M,$$

$$\gamma_\ell^p \in \{0, 1\}, \quad \ell \in \hat{\mathcal{L}}, \quad p \in \mathcal{H}$$

$$\text{FS}_\ell \text{ integer}, \quad \ell \in \hat{\mathcal{L}},$$

$$\xi_p^{m\ell} \in \{0, 1\}, \quad \ell \in \hat{\mathcal{L}}, \quad p \in \mathcal{H}, \quad m \in M,$$

$$\chi_{ip}^{w\ell\ell'} \in [0, 1], \quad w \in \hat{W}, \quad i \in \hat{N}(\ell, \ell'), \ell, \ell' \in \hat{\mathcal{L}}, p \in \mathcal{H},$$

$$\sigma_{pj}^{w\ell} \geq 0, \quad w \in \hat{W}, \quad \ell \in \hat{\mathcal{L}}, \quad p \in \mathcal{H}, \quad j \in \hat{S}(w^o),$$

$$\varrho_{iw} \geq 0, \quad w \in \hat{W}, \quad i \in \mathcal{I},$$

$$\varsigma_{iw} \geq 0, \quad w \in \hat{W}, \quad i \in \mathcal{I},$$

$$U_w^s \text{ free}, \quad w \in \hat{W}, \quad s \in I(U),$$

$$\mu_w^s \in \{0, 1\}, \quad w \in \hat{W}, \quad s \in I(U),$$

$$\alpha_\ell^m \geq 0, \quad \ell \in \hat{\mathcal{L}}, \quad m \in M,$$

$$U_w^{\text{RRT}} \text{ free}, \quad w \in \hat{W}.$$

## References

- Gao JC, Lan YJ, Wang ZQ, Xu SW, Zeng H, Lu H, Qian LB, Long JR, Sun XL, Li T, Zhu H, Zheng M, Wang GT (2022) Discussion on the key issues of metro urban construction—the 29th symposium of China urban transport development forum. *Urban Transp* 20(01):110–127. <https://doi.org/10.13813/j.cn11-5141/u.2022.0107>
- Maria NC, Medeiros-Sousa AR, Slovic AD (2019) An environmental health typology as a contributor to sustainable regional urban planning: the case of the metropolitan region of São Paulo (MRSP). *Sustainability* 11(20):5800. <https://doi.org/10.3390/su11205800>
- Liu YZ, Qian BY (2011) A multi-objective lattice-order decision making method for selection of rail transit network schemes. *J Traffic Transp* 11(05):76–82. <https://doi.org/10.19818/j.cnki.1671-1637.2011.05.012>
- Yang ZJ, Zhang FL, Chen YJ, Chen C (2018) Research on Group decision optimization model of urban rail transit network scheme. *J Railw Sci Eng* 15(07):1893–1900. <https://doi.org/10.19713/j.cnki.43-1423/u.2018.07.035>
- Zhang K, Qin BB, Liu YS, Zhang Q (2014) Research on evaluation of urban rail transit network. *J Railw Eng* 03:97–101
- Ceder A, Wilson NHM (1986) Bus network design. *Transp Res B Methodol* 20(4):331–344. [https://doi.org/10.1016/0191-2615\(86\)90047-0](https://doi.org/10.1016/0191-2615(86)90047-0)
- Cipriani E, Gori S, Petrelli M (2012) Transit network design: a procedure and an application to a large urban area. *Transp Res C Emerg* 20(1):3–14. <https://doi.org/10.1016/j.trc.2010.09.003>
- Pternea M, Kepaptsoglou K, Karlaftis MG (2015) Sustainable urban transit network design. *Transp Res A Policy* 77:276–291. <https://doi.org/10.1016/j.tra.2015.04.024>
- Barahimi AH, Eydi A, Aghaie A (2021) Multi-modal urban transit network design considering reliability: multi-objective bi-level optimization. *Reliab Eng Syst Safe* 216:107922. <https://doi.org/10.1016/j.ress.2021.107922>
- Nayeem MA, Rahman MK, Rahman MS (2014) Transit network design by genetic algorithm with elitism. *Transp Res C Emerg* 46:30–45. <https://doi.org/10.1016/j.trc.2014.05.002>
- López-Ramos F, Codina E, Marín A, Guarnaschelli A (2017) Integrated approach to network design and frequency setting problem in railway rapid transit systems. *Comput Oper Res* 80:128–146. <https://doi.org/10.1016/j.cor.2016.12.006>
- Liang M, Wang W, Dong C, Zhao D (2020) A cooperative coevolutionary optimization design of urban transit network and operating frequencies. *Expert Syst Appl* 160:113736. <https://doi.org/10.1016/j.eswa.2020.113736>
- Chakraborty P (2003) Genetic algorithms for optimal urban transit network design. *Comput Aided Civ Inf* 18(3):184–200. <https://doi.org/10.1111/1467-8667.00309>
- Walteros JL, Medaglia AL, Riaño G (2015) Hybrid algorithm for route design on bus rapid transit systems. *Transp Sci* 49(1):66–84. <https://doi.org/10.1287/trsc.2013.0478>
- Canca D, Barrena E, De-Los-Santos A, Andrade-Pineda JL (2016) Setting lines frequency and capacity in dense railway rapid transit networks with simultaneous passenger assignment. *Transp Res B Methodol* 93:251–267. <https://doi.org/10.1016/j.trb.2016.07.020>
- Canca D, De-Los-Santos A, Laporte G, Mesa JA (2017) An adaptive neighborhood search metaheuristic for the integrated railway rapid transit network design and line planning problem. *Comput Oper Res* 78:1–14. <https://doi.org/10.1016/j.cor.2016.08.008>
- Canca D, De-Los-Santos A, Laporte G, Mesa JA (2019) Integrated railway rapid transit network design and line planning problem with maximum profit. *Transp Res E Logist* 127:1–30. <https://doi.org/10.1016/j.tre.2019.04.007>
- Sun Y, Sun X, Li B, Gao D (2013) Joint optimization of a rail transit route and bus routes in a transit corridor. *Procedia Soc Behav Sci* 96:1218–1226. <https://doi.org/10.1016/j.sbspro.2013.08.139>
- Cadarso L, Escudero LF, Marín A (2018) On strategic multistage operational two-stage stochastic 0–1 optimization for the rapid transit network design problem. *Eur J Oper Res* 271(2):577–593. <https://doi.org/10.1016/j.ejor.2018.05.041>
- Laporte G, Pascoal MMB (2015) Path based algorithms for metro network design. *Comput Oper Res* 62:78–94. <https://doi.org/10.1016/j.cor.2015.04.007>
- Zhao H, Jiang R (2015) The memetic algorithm for the optimization of urban transit network. *Expert Syst Appl* 42(7):3760–3773. <https://doi.org/10.1016/j.eswa.2014.11.056>
- Shen JY (2008) Basic line shape and intersection calculation of urban rail transit line network planning. *Res Urban Rail Transit* 06:5–10
- Liu JF (2010) Study on distribution model of urban rail transit network based on IC card data. *Logist Technol* 29(16):64–67
- Sun L, Lu Y, Jin JG, Lee DH, Axhausen KW (2015) An integrated Bayesian approach for passenger flow assignment in metro networks. *Transp Res C Emerg* 52:116–131. <https://doi.org/10.1016/j.trc.2015.01.001>
- Li MD (2008) A study on pedestrian accessibility and the service scope of subway stations. In: 2008 China urban planning annual conference. China Association of City Planning, Dalian
- Ma JH, Cheng YP (2019) Study on the influence of different construction methods on project cost of subway station. *Build Econ* 40(05):64–68. <https://doi.org/10.14181/j.cnki.1002-851x.201905064>
- Jeroslow RG, Lowe JK (1984) *Modelling with integer variables*. Springer, Berlin