A Decision Method for the Catchment Area of Subway Station Under the Difference of Connection Modes

Zhaohua WANG, Xin ZHU*, Yanping DU, Shuihai DOU

Abstract: There is an interaction between subway station and the catchment area around the station, the catchment area of subway station on surrounding land is closely related to the connection modes. At present, subway passengers mainly reach subway stations by walking, cycling and bus, and the catchment area of the three connection modes is different. In order to clarify the catchment area of subway stations, it is necessary to study the catchment area of the three connection modes of subway stations. This study starts with a generalized connection cost for each connection mode, constructing the generalized connection cost models, especially considering the psychological changes of passengers with the increase of connecting distance. On this basis, we use Logit-SUE model to construct the probability density curve of the superior connection distance of different connection modes in different types and different area subway stations. Taking Beijing Metro Line 4 as the research case, through analyzing the probability density curve of each station of Line 4 we calculate the advantageous connection range of each subway station with different connection modes, and take the advantageous connection range as the catchment area for each type of connection mode.

Keywords: catchment; generalized connection cost; logit-SUE, subway connection

1 INTRODUCTION

Subway is the backbone of urban public transportation system; subway stations and surrounding catchment areas will have mutual influence. Subway station catchment area usually refers to the spatial area which has the greatest influence on the surrounding land use, and this area is also the main source area where most passengers come from [1, 2]. As the basic unit of rail transit system, subway stations are the connection points of rail transit system and urban system; the catchment area is also the core area where the urban residents interact with the subway system. The reasonable catchment area partitioning is the premise of the research on the different land use intensity, and the influence on subway station passenger flow by different intensity of land use. Catchment area for different types of connection modes is an important basis for station potential passenger flow prediction, connection service configuration and station area TOD planning. According to the unique situation of each subway station, this study provides a fine division method of the impact area of subway stations that can be applied to the built subway stations and the planned subway stations, which is closer to the real situation than the existing scheme. Subway station catchment area outlines the spatial area where subway passengers live. The range shape of catchment area may be affected by various factors, such as transportation network, station location, accessibility, arrival mode, land use, etc. In real life, when passengers choose stations, distance is not the only factor, but the number and distribution of entrances, connection time, total travel time, arrival frequency of vehicles and consumption cost are all factors that passengers need to consider [3-6]. For example, price-sensitive passengers may choose subway stations closer to their final destinations to save total travel costs. Comfort-sensitive passengers, on the other hand, may choose subway stations farther from their destinations to secure seats and improve travel comfort. Sanko et al. found that the size of the influence area of a station is influenced by the mode of travel, the frequency of departure and the interval between stations [7]. There is an interactive relationship between the subway station and the land around the station, and the interaction between them is mainly reflected in the process of subway connection. At present, the passengers are mainly connecting by walking, cycling and bus, the catchment area of three type modes is different. In the traffic scene of subway connection, the relationship between the three kinds of transportation connection modes is mainly reflected in the mutual competition. Each connection mode has its unique advantages and disadvantages, and the three connection modes of different subway stations correspond to different connection area range. The passenger flow distribution of the three connection modes can reflect each type's connection mode advantageous in the competition. So, the subway station catchment area question can be converted to the advantageous connection range for three connection modes. In this study, the reasonable connection range which passengers can accept to arriving at subway stations by walking, cycling and bus is used as the basis for dividing the catchment area of subway stations. According to that, we use the Logit-SUE model to calculate the passenger flow distribution between subway station and get the advantageous connection range of three kinds of connection modes, that is the catchment area.

2 LITERATURE REVIEW

The range of catchment area for subway station is diverse, and different researchers have different emphases on the catchment area, which also leads to great differences in the calculation of the catchment area. In mainstream studies, Proximity-only models are mostly used to classify edge morphology [8-11]. The main classification results of edge morphology of site influence domain include two types, respectively circle and polygon. The range shape of the circle is relatively simple, with the station as the center, and the time or distance of passengers arriving at the station as the radius. Different connection methods correspond to different radiuses, and finally form a nested ring shape. The polygon range shape depends on the built environment such as roads and buildings around the site, and the boundary of the catchment area is divided according to the environment. Thiessen polygon is a common polygon

partition method. Thiessen polygon was first proposed by [12], a Dutch scholar, for the purpose of studying the rainfall intensity around the weather station. As its method of dividing the boundary of the rainfall area is very similar to that of the catchment area of the station, it is widely used in the study of the boundary division of the influence area of subway stations. The characteristic of edge shape of Thiessen polygon site catchment area is that each polygon region contains only one site, and the distance of virtual points within the polygon range to the site is smaller than its distance to other sites. Although Thiessen polygon has been widely used in the research on the division of the catchment area of the site, its disadvantages in practical application are also very obvious. In many studies, the range shape of the catchment area divided by Tyson polygon appears an abnormally large range [13]. There are two types of catchment area partitioning method in existing research: one is that the government, experts, and scholars subjectively partition the catchment area of subway stations based on their work experience and research results; the other is the quantitative analysis study by building a model or analyzing the AFC data [14]. Governments all over the world have different rules on the partition of the catchment area of subway stations. China's "Guidelines for planning and design of areas along urban rail lines" stipulate that the area of rail transit may affect an area about 500 - 800 m from a subway station, where passengers can reach the entrance of the station within about 15 min of foot, and that this area shall be an area closely related to rail transit; the core area of the rail transit is an area of about 300 - 500 m from the subway station. This area is an area where subway stations can directly connect to surrounding buildings and public spaces [15]. In 2009, the American Transportation Association took a practical approach to publish a document, which defined the typical impact areas for different modes of transport [16]. In this file, it is defined that each mode of transportation has three impact areas the core impact area, the primary impact area, and the secondary impact area. According to field research data, the U.S. subway station's point impact area is usually defined as within 500 m of the station, which applies equally in South Korea [17]. There are also significant differences in the catchment area for subway stations in different countries or regions. The more mainstream catchment area is within 800 m radius near the subway station, which is an area accessible to passengers on foot of 10 min. Research by U. S. and Canadian experts shows that a North American typically does not walk more than 1750 m to urban rail transit sites. And 50% of the passenger walking distance is within 900 m, so North American experts and scholars will be near the rail station radius of 400 - 800 m coverage area as a direct attraction area of rail transit station. Due to the large and sparse nature of North America, some experts and scholars have used the area covered by a radius of 1600 m near the orbital site as a secondary attraction area [18-20]. In a case study in California, Ewing and Cervero showed that people who lived within half a mile of the transit station were four times more likely to use the transit station than those who lived between half and three miles from the transit station [22]. Korean experts [23] did a survey of 100 passers-by at each of Seoul's seven railway stations to determine the size of the station's catchment area, especially with regard to walking distance. They found that the average size and walking distance were 462 m and 376 m, respectively. [24] used a distance attenuation weight regression model to explore the factors influencing the flow of points at the Madrid metro station in Spain. Based on the Logic model to explore the Seoul metro station in South Korea to explore the impact factors of spot passenger flow [25]. Excavates variables that affect the passenger flow of subway stations through the multilinear regression and structural equation models [26]. Discussed the factors that influence the flow of passengers between points at Seoul subway stations [27]. Set up a direct forecast model of subway station passenger flow, and discussed the main factors impacting Nanjing passenger flow [28]. Used multiple linear regression methods to explore the impact factors of Beijing subway station spot passenger flow judging from the choice of connection mode and the distance of connection of different means [29]. Discussed the relationship between modes of transport and choice of activities [30]. Considers that travel options are determined by travel time and travel costs [31]. Considers that the main factor influencing the switch from low-cost modes of transport (e.g., bicycles) to high-cost modes of transport (e.g., public transport) is distance, i.e., each mode of transport has its most efficient travel distance [32]. Put forward that the main basis for urban residents to choose the mode of travel is the cost of this mode of travel [33]. puts forward the advantages based on the sharing rate function of Distance quantitative analysis method [34]. It is pointed out that the advantage distance refers to a certain distance. Some mode of transport compared with other methods of existence of the advantages, can be divided into absolute advantage distance and comparative advantage distance. The location of the subway has a great impact on the station's catchment area. When the surrounding environment of the subway station changes, the catchment area of the subway station changes. When the subway station is in the planning period or the surrounding environment is not yet mature, partition catchment area of the subway station veracity will become relatively difficult. In the past, the research on the reasonable connection distance of subway focused on the advantage analysis of one mode of connection, such as bus connection distance or walk connection distance, the main application of existing data for fitting, and did not form a common model system, and there are lots of limitations in application. In order to systematically analyze the reasonable connection range of various types of connection methods in the subway connection process this study aims at three types of connection: walking, cycling and bus, constructing a common system of calculation model of subway station catchment area, and calculating the walking, cycling and bus catchment area.

3 RESEARCH MODEL
3.1 Problem Description

Found that the advantageous connection range of different connection modes would be affected by travel mode, departure frequency and station interval. The object of this research is the subway station connection system which contains three connection modes, shown as Fig. 1. In the figure, *S* is the subway station, *D* is the origin or

destination area of passenger. Connection *k* contains three modes: walking, cycling and bus. Different passengers arrive at the subway station in different ways, each connection mode corresponding to the subway station catchment area is very different, the way passengers arrive at the subway station has a decisive influence on the catchment area range. Because this study focuses on the situation of direct access to the subway station, the alternating use cases between the three connection methods are not considered, and only the direct access situation is analyzed [7].
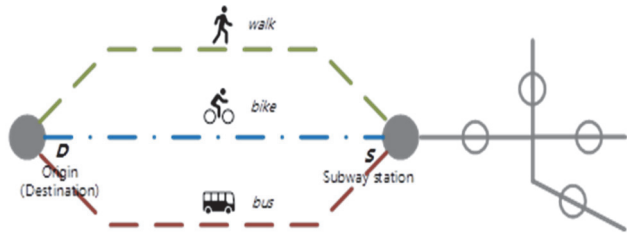

Figure 1 Subway station connection process

Passenger travel time cost and direct cost are mostly taken as the main components of generalized connection cost modeling in the existing research. Travel distance and time are simplified in the existing research; moreover, these kinds of research do not consider the change of passengers' perception as the distance and time increase during the connection process. To solve this problem, on the basis of existing research, this study fully considers the change of passengers' perception caused by the change of connecting distance, improve the calculation method of Generalized connection cost for subway connection.

3.2 Variable Description

Parameters and the definition of the model are shown in Tab. 1.

Table 1 Parameters and definition

Parameters	Definition
D	Origin or destination.
S	Subway station.
k	Connection modes, include walking, cycling, bus.
C_k	The generalized connection cost when passenger chooses k mode.
C_k^t	Connection time cost.
C_k^m	Connection direct cost.
C_k^e	Connection energy cost.
x_k	Ideal connection distance for connection mode k.
x	Actual connection distance for passengers.
α	Passenger's time value.
γ_k^e	Coefficient of energy expenditure.
λ_k	When the connection distance is beyond the ideal connection distance.
β_{walk}	Walking convenience level.
β_{bike}	Cycling activity level.
β_{bus}	Bus convenience level.

3.3 Bike Sharing Riding Data

The sharing riding data is from Mobike cup algorithm challenge in 2017; this database contains about 3 million cycling records data, consists of about 400 thousand shared bikes used by more than 300 thousand users in two weeks in Beijing 2017. There are seven fields including order ID, user ID, bike ID, bike type, start time, start loc, end loc; the data structure is shown in Tab. 2.

Table 2 Example of original data structure of shared bikes

Order ID	User ID	Bike ID	Bike type	Start time	Start loc	End loc
2157207	298025	235340	2	18:59:43	wx4ertw	wx4erv0
205764	258862	29610	1	18:59:40	wx4dyyn	wx4dyvq
2225354	575443	241859	2	18:59:40	wx4fepk	wx4fg12
2413836	491707	259634	2	18:59:39	wx4dy6s	wx4dy3g
3862160	147457	396042	1	18:59:39	wx4g36r	wx4g39d

Table 3 Bus AFC data structure

Card ID	SC ID	XC ID	SC time	XC time	SC name	XC name
C223DC90	00436A15	00436A12	2017030906	2017030906	Xiaocun	Xiaohongmen
69B6C6FA	24041A5	24041A4	2017030907	2017030907	Xiaocun	Xiaohongmen
CF1C72F3	24041A5	24041A4	2017030907	2017030907	Xiaocun	Xiaohongmen
C51727AC	24041A5	24041A3	2017030907	2017030907	Xiaocun	Xiaohongmen
99D33270	24041A5	24041A3	2017030907	2017030907	Xiaocun	Xiaohongmen

3.4 Bus AFC Data

The AFC dataset in this research is a week bus AFC data in 2017 from Beijing Bus Company. The AFC dataset contains seven fields: card number (Card ID), boarding station ID (SC ID), alighting station ID (XC ID), boarding time (SC time), alighting time (XC time), boarding station name (SCname), alighting station name (XCname). In order to filter out the bus station which has connections to the subway station among the many bus stops, here we add the bus network data. There are three fields we need: bus station name (station_name), bus station longitude (station_longitude), bus station latitude (station_latitude). The data structure is shown in Tab. 3.

3.5 Generalized Connection Cost Model
3.5.1 Connection Route Identification

The bike sharing riding data and bus AFC data are the full data on bike sharing riding and bus travel, so we need to discover the connection data in the full data. For bike sharing riding data, the shared bike we research are the bikes which are without parking piles; the parking position of these bikes is not fixed. When screening the travel data of subway shared bikes, we take the location of subway stations as the center and the travel data of unlocking and locking within a radius of 200 m as the ride data of connecting subway. We use the Haversine Eq. (1) and Eq. (2) to calculate the distance between the lockout point and the subway station.

$$h(\theta) = \sin^2\left(\frac{\theta}{2}\right) = \frac{1 - \cos(\theta)}{2} \tag{1}$$

$$h(\theta) = h\left(\frac{d}{R}\right) = h(\Delta\beta) + \cos(\beta_1)\cos(\beta_2)h(\Delta a) \tag{2}$$

R is the radius of the earth, $R = 6400$ km; d is the distance between the lockout point and the subway station, $d = 0.2$ km; θ is the radians between the two points and the center of the earth, $\theta = d/R$; $\alpha_i\beta_i$ is the longitude and latitude of the point. For bus AFC data, firstly we use the bus station name as an associated field, use the Haversine formula as a distance constraint, filter out the bus station within the 100 m of the subway station, and we define these stations as the bus-subway connect station. Secondly, filter out the passenger route data from the bus AFC dataset, which board or alight from the bus-subway connect station, use the filtered data as a passenger-route subset of the AFC dataset. Thirdly, use the card number as the associated data, associated subway AFC data and bus AFC data, filter out to connect route data from the passenger-route subset. The bus connect route data filtering logic is shown in Fig. 2.
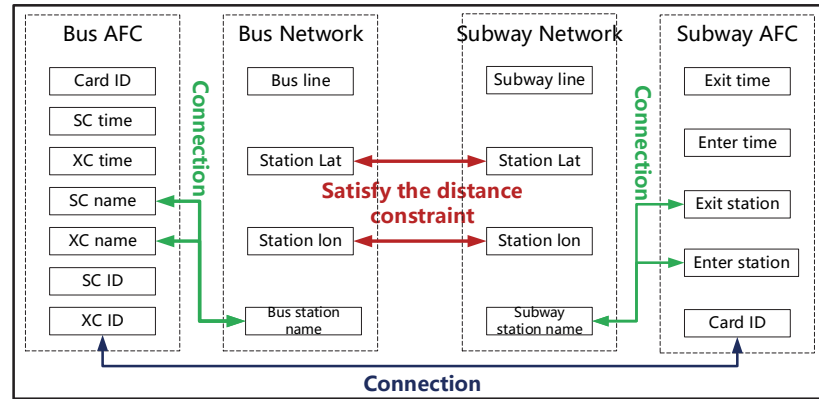

Figure 2 Bus-subway card data matching logic

3.5.2 Walking Generalized Connection Costs

There are no service providers when passengers choose walking to go to subway station, so there is no economic cost involved. The Generalized connection cost in the process of walking connection mainly includes time cost and energy cost. The walking Generalized connection cost $C_{walk} = C_{walk}^t + C_{walk}^e$. We take the passenger's walking speed as v_{walk}, there are four situations when passengers are walking to subway station, $x'_{walk} \le x, x_{walk} < x \le x_{bike}, x_{bike} < x'_{walk} \le x_{bus}$ and $x'_{walk} > x_{bus}$ respectively. Because the latter two situations are too far for walking, most passengers will not select walking to go to the subway station. So, we will not consider these two situations in this research. When $x'_{walk} \le x$, the passenger will be happy to walk to the subway station, the walking Generalized connection costs can be expressed as Eq. (3).

$$c_{walk}^t = \alpha\beta_{walk}\frac{x}{v_{walk}} \tag{3}$$

When $x_{walk} < x \le x_{bike}$, the passenger will not be satisfied with the distance which is beyond the expected walking connection distance. We take the punishment coefficient when the distance is beyond the expected walking distance as λ_{walk}, the walking Generalized connection costs can be expressed as Eq. (4).

$$c_{walk}^t = \alpha\beta_{walk}\left[\frac{x_{walk}}{v_{walk}} + \left(1 + \lambda_{walk}\right)\left(\frac{x - x_{walk}}{v_{walk}}\right)\right] \tag{4}$$

Passenger Generalized connection energy cost increases with the increase of distance, so the Generalized connection energy cost connected by walking can be expressed as Eq. (5).

$$c_{walk}^e = \gamma_{walk}^e \beta_{walk} \lambda_{walk} \frac{x}{v_{walk}} \tag{5}$$

We can get the Generalized connection cost of walking expressed as Eq. (6).

$$C_{walk} = \begin{cases} \left(a + \gamma_{walk}^e\lambda_{walk}\right)\dfrac{x\beta_{walk}}{v_{walk}}\left|\left(x \le x_{walk}\right)\right. \\[2mm] \beta_{walk}\left\{\left(\gamma_{walk}^e\lambda_{walk}\dfrac{x}{v_{walk}}\right) + \right. \\[2mm] \left. +\alpha\left[\dfrac{x_{walk}}{v_{walk}} + \left(1 + \lambda_{walk}\right)\left(\dfrac{x - x_{walk}}{v_{walk}}\right)\right]\right\}\left|\left(x_{walk} \le x \le x_{bike}\right)\right. \end{cases} \tag{6}$$

λ_{walk} increases with distance, λ_{walk} can be expressed as Eq. (7).

$$\lambda_{walk} = \frac{x}{x_{walk}} \tag{7}$$

3.5.3 Cycling Generalized Connection Costs

Cycling connection mode contains three types of costs, time cost, rent cost, and energy cost respectively. We take the Average cycling speed as v_{bike}, and define the shared bike riding activities within 3 km are connected cycling activities. Take the activity level of shared bikes as β_{bike}, β_{bike} is the number of shared bikes used within 3 km area around subway station accounted for the number of shared bikes using in the city. When value of β_{bike} is larger, it will be easier to obtain shared bikes near the subway station; otherwise, it will be more difficult to obtain shared bikes. When the passengers are cycling to the subway station, there will be three situations, $x_{bike} \le x, x_{bike} < x \le x_{bus}$, and $x > x_{bus}$ respectively. As

the third case is too far to ride, it is not considered here. We divide the connection mode of shared bikes into two processes, finding bikes and riding bikes respectively, as shown in Fig. 3. Referring to the report, assuming that the distance from the origin to the shared bike location is no more than 500 m, finding bike Generalized connection costs C_{bike}^{find} can be expressed as Eq. (8).
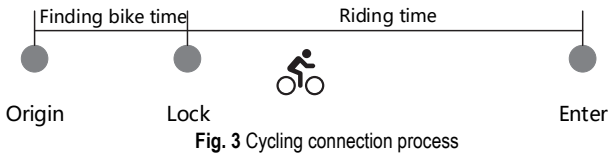

Fig. 3 Cycling connection process

$$C_{bike}^{find} = \alpha\beta_{walk}\left(\beta_{bike}\frac{500}{v_{walk}} + \gamma_{walk}^e\lambda_{walk}\frac{500}{v_{walk}}\right) \tag{8}$$

When $x \le x_{bike}$, the passenger is happy to choose to cycle to go to subway station, the cycling connection Generalized connection costs can be expressed as Eq. (9) and Eq. (10).

$$C_{bike}^t = \alpha\beta_{bike}\frac{x}{v_{bike}} \tag{9}$$

$$C_{bike}^e = \gamma_{bike}^e\frac{x}{v_{bike}} \tag{10}$$

When $x_{bike} < x'_{bike} \le x_{bus}$, the passenger will not be satisfied with the distance which is beyond the expected cycling connection distance. We take the punishment coefficient when the distance is beyond the expected cycling distance as λ_{bike}, cycling time cost and energy cost are shown in Eq. (11) and Eq. (12).

$$C_{bike}^t = \alpha\beta_{bike}\left[\frac{x_{bike}}{v_{bike}} + (1+\lambda_{bike})\left(\frac{x-x_{bike}}{v_{bike}}\right)\right] \tag{11}$$

$$C_{bike}^e = \gamma_{bike}^e\frac{x_{bike}}{v_{bike}} + \lambda_{bike}\gamma_{bike}^e\frac{x-x_{bike}}{v_{bike}} \tag{12}$$

Cycling cost C_{bike}^m mainly refers to the cost of passengers using shared bikes. The data in this study is the data of shared bikes in Beijing in 2017, so it is calculated according to the charging standard of Beijing in 2017 of 1.5 yuan per 30 minutes, as the connecting cycle takes less than 30 minutes, $C_{bike}^m = 1.5$. λ_{bike} increases with distance, λ_{bike} can be expressed as Eq. (13).

$$\lambda_{bike} = \frac{x}{x_{bike}} \tag{13}$$

We can get the Generalized connection cost of cycling expressed as Eq. (14).

$$C_{bike} = C_{bike}^{find} + \begin{cases} \left(\alpha\beta_{bike} + \gamma_{bike}^e\right)\dfrac{x}{v_{bike}} + 1.5 \ \left|(x_{walk} < x \le x_{bike})\right. \\[2ex] \left(\alpha\beta_{bike} + \gamma_{bike}^e\right)\dfrac{x_{bike}}{v_{bike}} + \\[1ex] \quad + (1+\lambda_{bike})\left(\alpha\beta_{bike} + \gamma_{bike}^e\right)\dfrac{x-x_{bike}}{v_{bike}}\right| + \\[2ex] \quad + 1.5\left(x_{bike} < x \le x_{bus}\right) \end{cases} \tag{14}$$

3.5.4 Bus Connection to Generalized Connection Costs

The Generalized connection cost of bus connection contains three parts of costs. We take the bus average speed as v_{bus}.

Bus connection time cost is C_{bus}^t:

$$C_{bus}^t = \alpha\beta_{bus}\frac{x}{v_{bus}} \tag{15}$$

Bus energy cost of bus connections is C_{bus}^e:

$$C_{bus}^e = \alpha\beta_{bus}\frac{x}{v_{bus}} \tag{16}$$

Ticket cost C_{bus}^m is referred to public transportation fares in Beijing, starting at 2 yuan within 10 km, the connecting journey will not exceed 10 km, so $C_{bus}^m = 2$. Define k as the travel mode between origin or destination point D and the target subway station S, $k = \{walk, bike, bus\}$. Assume the Generalized connection cost of passenger from $D \rightarrow S$ or from $S \rightarrow D$ is C_k, based on passenger travel choice considerations, divide passenger Generalized connection costs into two parts: C_k^t is the time costs, C_k^m is the direct costs, Generalized connection costs $C_k = C_k^t + C_k^m$. Time cost C_k^t is positively correlated with distance, the time costs in the existing research are mostly fixed values. But in the real world, the coefficient of time cost increasing with distance is not constant. When passengers weigh the cost of travel time, the coefficient of time cost increasing with distance should be divided into several sections with distance increasing, and the critical distance of the step change is related to the connection range of different competing connections in the network. Direct cost C_k^m is the bus ticket or sharing bike rent, it is fixed, the value needs to be assigned according to the specific connection situation.

3.6 Subway Station Passenger Flow Assignment

In a system of three connection modes, each passenger will choose the one he thinks is the most cost-effective; passenger's selection behavior conforms to Wardrop's first principle. Assume the probability where the passenger connection mode k is δ_k, each type of the flow assignment is q_k, $q_k = \delta_k Q$, $\forall_k$, $\sum_{k=1}^m \delta_k = 1$, $\sum_{k=1}^m q_k = Q$. The flow

assignment model when the probability of traveler's choice of three connection modes reaches equilibrium state shown in Eq. (17):

$$q_k = Q \frac{\exp(-\eta C_k)}{\sum_{j=1}^{m} \exp(-\eta C_j)} \qquad (17)$$

η is the passengers' perception coefficient of connection cost, the larger the value of η is, the smaller the passenger flow assignment randomness will be; when $\eta = 100$, SUE passenger flow assignment tends to UE passenger flow assignment, refer to the research of Li, construct the subway connection Logit-SUE flow assignment model shown in Eq. (18) to Eq. (20).

$$\min Y = \sum_{k=1}^{m} \int_0^{q_k} C_k(q)\,\mathrm{d}q + \frac{1}{\theta} \sum_{k=1}^{m} q_k \ln q_k \qquad (18)$$

$$s.t. \sum_{k=1}^{m} q_k = Q \qquad (19)$$

$$q_k \geq 0 \qquad (20)$$

Construct generalized Lagrange functions to prove the equivalence between the model and SUE conditions, as shown in Eq. (21).

$$L(q,k) = Y(q) - k \left(\sum_{k=1}^{m} q_k - Q \right) \qquad (21)$$

The K-T condition of Logit-SUE optimized model is shown in Eq. (22) to Eq. (25):

$$q_k \left(C_k + \frac{1}{\theta}(\ln q_k + 1) - \kappa \right) = 0 \qquad (22)$$

$$C_k + \frac{1}{\theta}(\ln q_k + 1) - \kappa \geq 0 \qquad (23)$$

$$\sum_{k=1}^{m} q_k = Q \qquad (24)$$

$$q_k \geq 0 \qquad (25)$$

κ is the K-T multiplier of Eq. (17). When $q_k > 0$, $C_k + \frac{1}{\theta}(\ln q_k + 1) - \kappa = 0$, based on that, we can get the q_k, and we can get the Logit-SUE model as Eq. (26).

$$q_k = Q \frac{\exp(-\theta C_k)}{\sum_{k=1}^{m}(-\theta C_k)} \qquad (26)$$

4 RESULTS
4.1 The Location and Function Difference of Subway Station

The location and function of subway station have

direct influence on the station advantageous connection range. About the station location, Beijing subway takes Tian 'anmen as the center. The closer it is to the center, the denser the subway stations will be. The farther away from the center, the sparser the subway stations will be. About the station function, we classify the Beijing subway station as employment station, shopping station, travel spot station, living station, and transportation hub station.
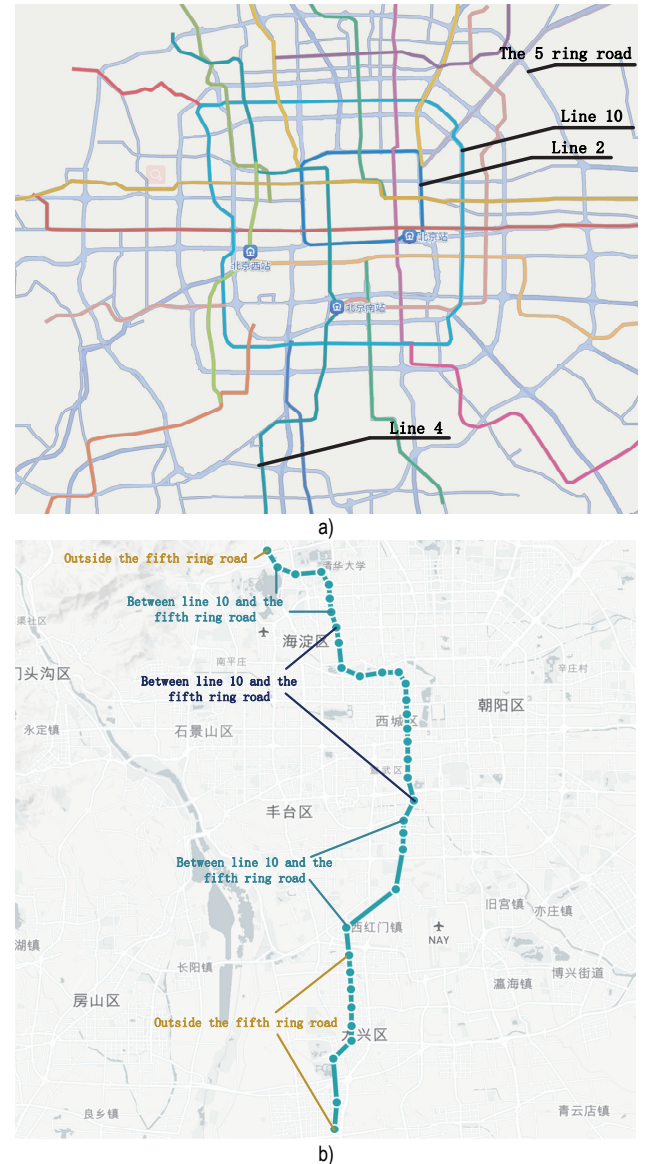

a)

b)

Figure 4 a) The location of line 2, line 10 and the fifth ring road, b) three parts of line 4 subway station

In Fig. 4a, there are two ring lines, line 2 and line 10, and a ring road, the fifth ring road. We divide the subway stations into three parts, inside line 10, between line 10 and the fifth ring road, outside the fifth ring road, respectively. Inside line 10, subway station is intensive, between line 10 and fifth ring road, subway station is not as intensive as stations inside line 10, but more intensive than the stations outside the fifth ring road. In this research, we choose line 4 as the object. In Fig. 4b, line 4 is a subway line through Beijing from south to north, over 50 km. The line 4 through all three kinds of area, and line 4 has all types of stations, the function types for each station are shown in Tab. 4.

Table 4 The function type of line 4 station

Function type	Station name		
	Inside line 10	Between line 10 & 5th	Outside the 5th
Employment station	Renming Univ., Weigongcun, Lingjing Hutong, Xuanwu Men, Caishi Kou	Haidian Huangzhuang, Peking Univ. East Gate, Zhongguancun	Anheqiao Bei(N), Biomedical Base
Shopping station	Xidan	Xihong Men	Gaomidian Nan(S), Tiangong Yuan
Travel spot station	National Library, Xinjie Kou, Ping'an Li, Xsi	Beigongmen, Yuanmingyuan Park	Qingyuan Lu
Living station	Taoranting, Majiapu, Gongyi Xiqiao, Xin Gong	Xi Yuan	Gaomidian Bei(N), Zaoyuan, Huangcun Xidajie, Yihezhuang,
Transportation hub station	Beijing Zoo, Xizhi Men, Beijing South Railway Station	Jiaomen Xi(W)	Huangcun Railway Station

4.2 Initial Value of the Model

According to "Guidelines for Planning and Design of Areas along Urban Rail Lines" we take the subway stations which are inside the line 10 $x_{walk} = 0.5, x_{bike} = 0.9, x_{bus} = 1.2$, and the subway stations which are between the line 10 and the fifth ring road $x_{walk} = 0.7, x_{bike} = 1.1, x_{bus} = 1.5$, outside the fifth ring road $x_{walk} = 0.8, x_{bus} = 1.8$, because in this area, there are no bikes sharing serve, so we do not consider the x_{bike} in this area. We use the GDP calculation method to calcite the passenger's time value α (yuan/min), $\alpha = GDP/N \times T$, N is the annual employment number, T is the average number of hours worked annually by the employee. For the coefficient of energy expenditure, walking is the most physically expensive, followed by cycling; taking a bus is the easiest, and in a different area, the coefficient will be different. Initial value of the model is shown in Tab. 5. It is necessary to clarify the boundary scope of subway connection. Among the three kinds of connection modes, the distance of bus connection mode is farthest. In this research, we use the bus connection route data in 3.2 to define the border of subway station connection range. We select the top 80% of passenger flow from the bus connection route data, and take the average distance of this route as the connection range border. According to that, we take 3 km as the connection range border. Then we calculate the Generalized connection cost of every subway station in line 4, and calculate the probability density curve of each connection mode 5.

Table 5 Initial value of β_{walk}, β_{bike} and β_{bus}

Station name	β_{walk}	β_{bike}	β_{bus}	Station name	β_{walk}	β_{bike}	β_{bus}
Anheqiao Bei (N)	0.34	0	0.51	Caishi Kou	0.23	0.34	0.46
Beigongmen	0.4	0.3	0.61	Taoranting	0.23	0.35	0.58
Xi Yuan	0.28	0.43	0.5	Beijing South Railway Station	0.56	0.39	0.54
Yuanmingyuan Park	0.4	0.65	0.49	Majiapu	0.29	0.39	0.8
Peking Univ. East Gate	0.34	0.43	0.56	Jiaomen Xi (W)	0.21	0.39	0.58
Zhongguancun	0.15	0.72	0.34	Gongyi Xiqiao	0.15	0.62	0.44
Haidian Huangzhuang	0.26	0.3	0.67	Xin Gong	0.40	0.64	0.49
Renming Univ.	0.28	0.31	0.22	Xihong Men	0.46	0.54	0.4
Weigongcun	0.21	0.33	0.38	Gaomidian Bei (N)	0.34	0	0.31
National Library	0.28	0.43	0.5	Gaomidian Nan (S)	0.28	0	0.46
Beijing Zoo	0.24	0.32	0.48	Zaoyuan	0.34	0	0.3
Xizhi Men	0.25	0.1	0.37	Qingyuan Lu	0.4	0	0.35
Xinjie Kou	0.28	0.12	0.35	Huangcun Xidajie	0.34	0	0.36
Ping'an Li	0.28	0.2	0.32	Huangcun Railway Station	0.34	0	0.39
Xsi	0.30	0.17	0.34	Yihezhuang	0.34	0	0.57
Lingjing Hutong	0.30	0.11	0.27	Biomedical Base	0.46	0	0.61
Xidan	0.27	0.19	0.26	Tiangong Yuan	0.34	0	0.6
Xuanwu Men	0.28	0.1	0.33				

Table 6 Catchment area of five types stations in different area

Station type	Station name	location	Walking	Cycling	Bus
Employment station	Xuanwu Men	Inside line 10	0.61	1.42	1.62
	Peking Univ. East Gate	Between line 10 & 5th ring road	0.82	1.28	1.65
	Anheqiao Bei(N)	Outside 5th ring road	0.93	/	2.16
Shopping station	Xidan	Inside line 10	0.65	1.18	1.25
	Xihongmen	Between line 10 & 5th ring road	0.71	1.09	1.49
	Gaomidian Nan(S)	Outside 5th ring road	1.03	/	1.78
Travel spot station	Xinjie Kou	Inside line 10	0.59	1.38	1.42
	Beigongmen	Between line 10 & 5th ring road	0.75	1.35	1.86
	Qingyuan Lu	Outside 5th ring road	0.89	/	1.98
Living station	Taoranting	Inside line 10	0.85	1.19	1.38
	Xi Yuan	Between line 10 & 5th ring road	0.92	1.38	2.15
	Huangcun Xidajie	Outside 5th ring road	1.05	/	2.38
Transportation hub station	Xizhi Men	Inside line 10	0.62	1.42	1.82
	Beijing South Railway Station	Between line 10 & 5th ring road	0.35	1.12	1.28
	Huangcun Railway Station	Outside 5th ring road	1.15	/	1.38

4.3 Advantageous Connection Range of Different Stations

In this part, we focus on the difference of each type of stations which are in different area. Firstly, we choose three stations from each type of stations in different area. Secondly, we compare the advantageous connection range of three stations in the same type and take the intersection point of Logit-SUE function as the advantageous connection range of walking and cycling. And in reference to 5.2, we take the result of connection range as the bus connection limit, calculate three connection modes' advantageous connection range of five types stations in different area, as shown in Tab. 6. And the advantageous connection range corresponds to the catchment area of each connection mode.

5 DISCUSSION
5.1 Employment station

The catchment area for Xuanwu Men, Peking Univ. East Gate, Anheqiao Bei (N) is shown in Fig. 5. For walking and bus connection modes, the farther the station is from the city center, the larger the catchment area is. Furthermore, bus catchment area of the station which is outside the fifth ring road is much larger than the station in other two areas, and the bus catchment area of the station in the other two areas is almost the same. For cycling connection modes it is opposite, the catchment area of station inside line 10 is larger than the station between line 10 and the fifth ring road. For employment station, when the station is inside line 10, catchment area of cycling and bus is almost the same, when the station is between line 10 and fifth ring road, the range is great difference.
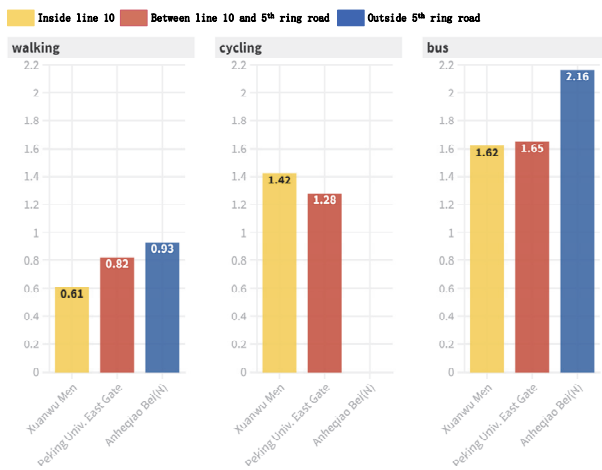

Figure 5 Employment station catchment area

5.2 Shopping Station

The catchment areas for Xidan, Xihongmen, Gaomidian Nan (S) are shown in Fig. 6. For walking and bus connection modes, the basic laws are the same as for employment station, but walking catchment area of the station which is outside the fifth ring road is much larger than the station in the other two areas, and the walking catchment area of the station in the other two areas is almost the same. For cycling, connection modes are opposite, are also the same as for employment station. Shopping station catchment area of cycling and bus is
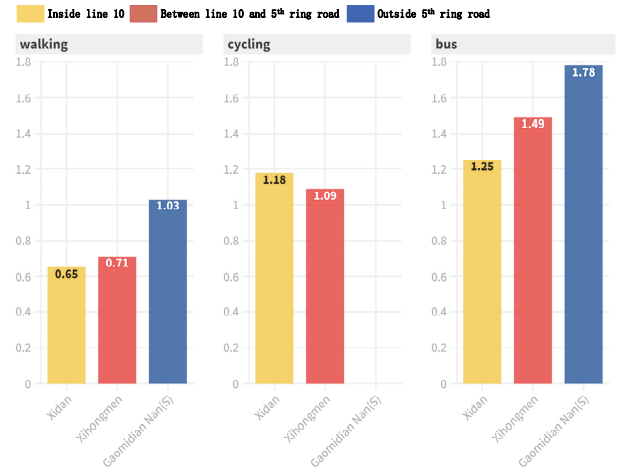
similar as for the employment station.

Figure 6 Shopping station catchment area

5.3 Shopping station

The catchment area for Xinjie Kou, Beigongmen, Qingyuan Lu is shown in Fig. 7. For walking and bus connection modes, the basic laws are the same as for employment and shopping station. The difference with employment and shopping station is that the bus catchment area of the station which is inside line 10 is smaller than other two areas. Furthermore, cycling catchment area of travel spot station is almost the same, it is different from employment and shopping station. The catchment area of travel spot station is basically the same as employment and shopping station.
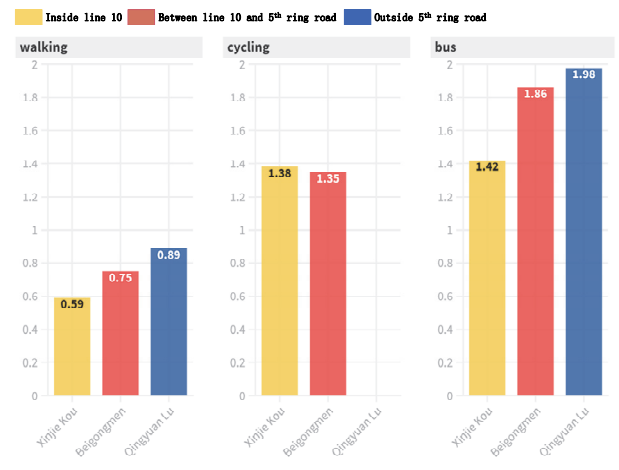

Figure 7 Travel spot station catchment area

5.4 Living Station

The catchment areas for Taoranting, Xi Yuan, Huangcun Xidajie are shown in Fig. 8. The catchment area for walking and bus connection modes will be basically same as for travel spot station and shopping station. The catchment area for cycling mode is different from the first three stations. The cycling catchment area for the station inside line 10 is smaller than the station between line 10 and fifth ring road. The catchment area of travel spot station is basically the same as employment and shopping station.
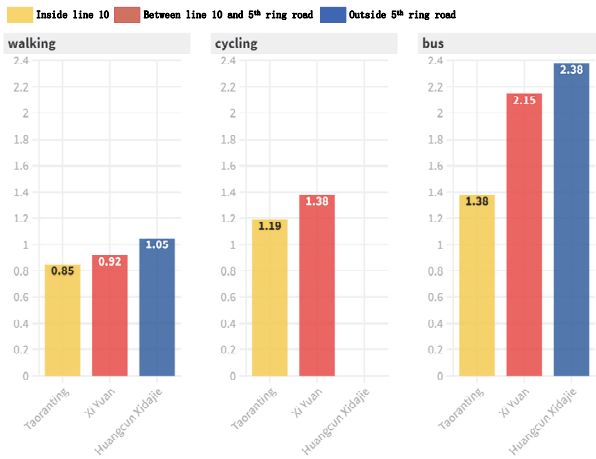
Figure 8 Living station catchment area

5.5 Transportation Hub Station

The catchment area for Xizhi Men, Beijing South Railway Station, Huangcun Railway Station is shown in Fig. 9. The difference between these three stations is not only in location, but also in the function. Xizhi Men is a large-scale comprehensive transportation hub, connecting the trains, subways, and bus system. For Beijing South Railway Station, it is a subway and train connect station. And train station is the main functions of it. Except that, Beijing South Railway Station is far from other urban functions. The Huangcun Railway Station is just a small train station, the size is nowhere near the other two stations. According to the basic information, it is not hard to understand, the walking catchment area is only just 350 m. About the catchment area of cycling, the range of Xizhi Men is larger than the range of Beijing South Railway Station. Even in the bus catchment area, the range of Xizhi Men is larger than the other two stations.
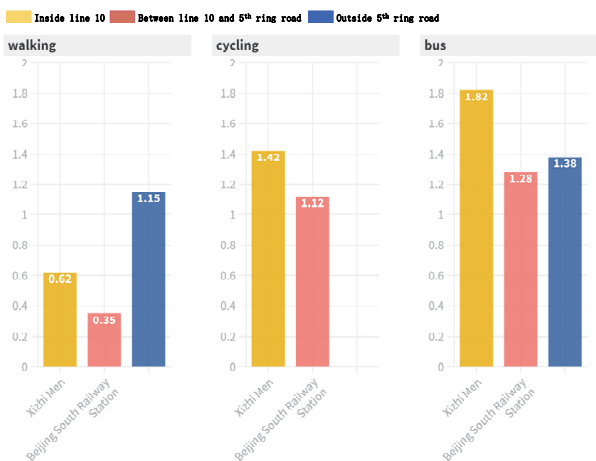

Figure 9 Transportation hub station catchment area

By comparing the catchment area, we come to the conclusion as follows:

(1) There is a strong substitutability between the cycling and bus connection modes. Inside line 10, shared bike serves are relatively perfect, can basically satisfy the needs of passengers. Except for the transportation hub station Xizhimen, the cycling and bus catchment areas of the station in this area are basically the same. Xizhimen is different because the bus serves are better than at other types of stations.

But the situation for the stations between line 10 and the fifth ring road is different. The service of shared bike in this area is not good as at the station inside line 10. So, the catchment area of bus is larger a lot than the range of cycling.

(2) The catchment area of transportation hub station is greatly influenced by station functions. For the transportation hub station whose main function is train station, in most instances the catchment area is smaller than other types of stations in the same area. But for the transportation hub station whose main function is comprehensive transportation hub, the catchment area is larger than other types of stations in the same area.

(3) In Beijing, large residential communities are mostly located in the suburbs, so when the stations are outside line 10, connection range for living station is larger than the other four types of stations. For the stations inside line 10, even the bus and shared bike serves are relatively perfect, but it still cannot satisfy the needs of living station passengers inside line 10.

(4) Almost all the shopping stations are near the subway station. We can find that the connection range of shopping stations is almost smaller than other types of stations.

6   CONCLUSIONS

Subway is the most effective way to improve the efficiency of urban transportation. Subway stations have different attractions for passengers in different areas. It is necessary to study the attraction range of subway stations. This study starts from the connecting modes of passengers, then constructs the generalized connection cost model of three kinds of connection modes and calculates the generalized connection cost for three modes. On this basis, construct the connection Logit-SUE model, calculate the passenger connection utility for each type of connection modes, according to the passenger connection utility, delineate the catchment area for each type of connection mode. Thus, determine the passenger flow attraction range of subway stations. Finally, we chose Beijing subway line 4 as the research object, studied the catchment area of three connection modes in different regions and different types of subway stations. Compared with traditional research of generalized connection cost, this paper considers the psychological changes of passengers with the increase of connecting distance, and the generalized travel cost model is closer to the actual situation. When determining the attraction range, we consider the difference of station location, station type and connection modes, refine the research on the passenger flow attraction range of subway station. The shortcoming of this study is that due to the lack of data support that directly represents the population size attracted or generated by each land unit, it was not possible to incorporate population as a direct indicator into the evaluation system. Instead, we had to rely on other indirect indicators to assess the scale of population for each land unit. And there is no further discussion of the interactive relationship between subway stations and surrounding environment. In the follow-up study, subway passenger flow will be taken as the main line to further explore the influence mechanism of subway stations on the surrounding environment, to provide theoretical and practical basis for subway TOD construction.

7 REFERENCES

[1] Dolega, L., Pavlis, M., & Singleton, A. (2016). Estimating attractiveness, hierarchy and catchment area extents for a national set of retail Centre agglomerations. *Journal of Retailing and Consumer Services*, *28*, 78-90. https://doi.org/10.1016/j.jretconser.2015.08.013

[2] Neeley, B. C., Niazi, F. A., Ebbert, M. A., Forman, A. G., Hobbs, G. R., & Riggs, J. E. (2021). Using catchment population to estimate sporadic Creutzfeldt Jakob disease incidence. *Military Medicine*, 34865142. https://doi.org/10.1093/milmed/usab510

[3] Ryan, M., Lin, T. G., Xia, J. C., & Robinson, T. (2016). Comparison of perceived and measured accessibility between different age groups and travel modes at greenwood station, Perth, Australia. *European Journal of Transport and Infrastructure Research*, *16*(2), 406.

[4] Shaoa, C. & Jianghong, T. (2015). Logistic regression models for the nearest train station choice: A comparison of captive and non-captive stations. *Case Studies on Transport Policy*, *3*(4), 382-391. https://doi.org/10.1016/j.cstp.2015.06.002

[5] Aqib, M., Mehmood, R., Alzahrani, A., Katib, I., & Altowaijri, S. M. (2019). Rapid Transit Systems: Smarter Urban Planning Using Big Data, In-Memory Computing, Deep Learning, and GPUs. *Sustainability*, *11*(10), 2736. https://doi.org/10.3390/su11102736

[6] Ahmad, I., Alqurashi, F., Abozinadah, E., & Mehmood, R. (2022). Deep journalism and deep journal v1.0: a data-driven deep learning approach to discover parameters for transportation. *Sustainability*, 14. https://doi.org/10.20944/preprints202203.0245.v1

[7] Sanko, N. & Shoji, K. (2009). Analysis of the structural characteristics of the station catchment area in Japan. *International Conference Series on Competition and Ownership in Land Passenger Transport 2009 Delft, The Netherlands Thredbo 11.*

[8] Lin, T. G., Xia, J. C., Robinson, T. P., Olaru, D., Smith, B., Taplin, J., & Cao, B. (2016). Enhanced huff model for estimating park and ride (pnr) catchment areas in perth, wa. *Journal of Transport Geography*, *54*(jun.), 336-348. https://doi.org/10.1016/j.jtrangeo.2016.06.011

[9] Tsai, Y. H. (2009). Impacts of self-selection and transit proximity on commute mode choice: evidence from taipei rapid transit system. *Annals of Regional Science.* https://doi.org/10.1007/s00168-008-0233-z

[10] Namiot, D. & Sneps-Sneppe, M. (2019). On proximity versus geo-information systems. *Computer Science.*

[11] Mahrsi, M., Côme, E., Oukhellou, L., & Verleysen, M. (2017). Clustering smart card data for urban mobility analysis. *IEEE Transactions on Intelligent Transportation Systems.* https://doi.org/10.1109/TITS.2016.2600515.

[12] Thiessen, A. H. (1998). Precipitation average for large areas. *Mon.wea.rev*, 39.

[13] Reyer, M., Fina, S., Siedentop, S., & Schlicht, W. (2014). Walkability is only part of the story: walking for transportation in Stuttgart, Germany. *International Journal of Environmental Research and Public Health*, *11*(6), 5849-5865. https://doi.org/10.3390/ijerph110605849

[14] Ma, X., Liu, C., Wen, H., Wang, Y., & Wu, Y. J. (2017). Understanding commuting patterns using transit smart card data. *Journal of Transport Geography*, *58*(JAN.), 135-145. https://doi.org/10.1016/j.jtrangeo.2016.12.001

[15] Ministry of Housing and Urban-Rural Development of the People's Republic of China (2015). Guidelines for planning and design of areas along urban rail lines.

[16] American Public Transportation Association (2009). Defining transit areas of influence, *APTA SUDS-UD-RP-001-09*, Washington, D.C.

[17] Korea Development Institute. (2008). The fifth edition of transportation feasibility studies. Korea Development Institute

[18] Schlossberg, M., Brown, N., Bossard, E. G., & Roemer, D. (2004). Using spatial indicators for pre- and post-development analysis of TOD areas: a case study of Portland and the Silicon Valley. *Brit.j.appl.phys.* https://doi.org/10.1088/0508-3443/10/1/307

[19] Reusser, D. E., Loukopoulos, P., Stauffacher, M., & Scholz, R. W. (2008). Classifying railway stations for sustainable transitions balancing node and place functions. *Journal of Transport Geography*, *16*(3), 191-202. https://doi.org/10.1016/j.jtrangeo.2007.05.004

[20] Wibowo, S. S. (2005). Modelling walking accessibility to public transport terminals. *Nanyang Technological University.* https://doi.org/10.11175/easts.6.147

[21] A., L. X., B., M. R., & C., M. O. (2021). Woodland's role in natural flood management: Evidence from catchment studies in Britain and Ireland. *Science of The Total Environment*, *813*, 151877. https://doi.org/10.1016/j.scitotenv.2021.151877.

[22] Ewing, R. & Cervero, R. (2010). Travel and the built environment - a meta-analysis. *Journal of the American Planning Association*, *76*(3), 265-294. https://doi.org/10.1080/019443610003766766

[23] Kim, H. & Nam, J. (2013). The size of the station influence area in Seoul, Korea: based on the survey of users of seven stations. *International Journal of Urban Sciences*, *17*(3), 331-349. https://doi.org/10.1080/12265934.2013.810463

[24] Taylor, B. D., Miller, D., Iseki, H., & Fink, C. (2009). Nature and/or nurture? analyzing the determinants of transit ridership across us urbanized areas. *Transportation Research Part A Policy & Practice*, *43*(1), 60-77. https://doi.org/10.1016/j.tra.2008.06.007

[25] Jun, M. J., Choi, K., Jeong, J. E., Kwon, K. H., & Kim, H. J. (2015). Land use characteristics of subway catchment areas and their influence on subway ridership in Seoul. *Journal of Transport Geography*, *48*(OCT.), 30-40. https://doi.org/10.1016/j.jtrangeo.2015.08.002

[26] Sohn, K. & Shim, R. (2010). Factors generating boarding at metro stations in the Seoul metropolitan area. *Cities*, *27*(5), 358-368. https://doi.org/10.1016/j.cities.2010.05.001

[27] Choi, J., Yong, J. L., Kim, T., & Sohn, K. (2012). An analysis of metro ridership at the station-to-station level in Seoul. *Transportation*, *39*(3), 705-722. https://doi.org/10.1007/s11116-011-9368-3

[28] Zhao, J., Wei, D., Yan, S., & Zhu, Y. (2013). What influences metro station ridership in china? insights from Nanjing. *Cities*, *35*(dec.), 114-124. https://doi.org/10.1016/j.cities.2013.07.002

[29] Sun, L. S., Wang, S. W., Yao, L. Y., Rong, J., & Ma, J. M. (2016). Estimation of transit ridership based on spatial analysis and precise land use data. *Transportation Letters*, 140-147. https://doi.org/10.1179/1942787515Y.0000000017

[30] Krygsman, S., Arentze, T., & Timmermans, H. (2007). Capturing tour mode and activity choice interdependencies: a co-evolutionary logit modelling approach. *Transportation Research Part A*, *41*(10), 913-933. https://doi.org/10.1016/j.tra.2006.03.006

[31] Javanmardi, M., Langerudi, M. F., Shabanpour, R., & Mohammadian, K. (2015). Mode choice modelling using personalized travel time and cost data. *International Conference on Travel Behaviors Research.* https://doi.org/10.5038/CUTR-NCTR-RR-2015-07

[32] Mueller, S., Tscharaktschiew, S., & Haase, K. (2008). Travel-to-school mode choice modelling and patterns of school choice in urban areas. *Journal of Transport Geography*, *16*(5), 342-357. https://doi.org/10.1016/j.jtrangeo.2007.12.004

[33] Hua, X. D. (2016). Transportation modal split optimization of multimodal traffic network based on supply-demand equilibrium. *Southeast University.*

[34] Xiao, J., Chen, B. Y., Sun, X. H., Yang, H. Q., & Hong-Qing,

L. (2019). Storm rainfall-runoff quality analysis of typical catchment using swmm. *DE Stech Transactions on Computer Science and Engineering*. https://doi.org/10.12783/dtcse/iteee2019/28714

Contact information:

Zhaohua WANG, PhD
School of Mechanical and Electrical engineering,
Beijing Institute of Graphic Communication,
No.1, Xinghua Street, 102699 Daxing District, Beijing, P. R. China
E-mail: wzh@bigc.edu.cn

Xin ZHU, PhD
(Corresponding author)
Management School, Beijing Union University,
No.97, East North Fourth Ring Road, 100101 Chaoyang District, Beijing,
P. R. China
E-mail: 18515525865@163.com

Yanping DU, Prof.
School of Mechanical and Electrical engineering,
Beijing Institute of Graphic Communication,
No.1, Xinghua Street, 102699 Daxing District, Beijing, P. R. China
E-mail: duyanping@bigc.edu.cn

Shuihai DOU, Assoc. prof.
School of Mechanical and Electrical engineering,
Beijing Institute of Graphic Communication,
No.1, Xinghua Street, 102699 Daxing District, Beijing, P.R. China
E-mail: doushuihai@bigc.edu.cn