

## Research Article

# Forecasting Passenger Flow Distribution on Holidays for Urban Rail Transit Based on Destination Choice Behavior Analysis

Enjian Yao <sup>1</sup>, Junyi Hong <sup>1</sup>, Long Pan <sup>2</sup>, Binbin Li <sup>3</sup>, Yang Yang,<sup>1</sup> and Dongbo Guo <sup>1</sup>

<sup>1</sup>School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China

<sup>2</sup>College of Metropolitan Transportation, Beijing University of Technology, Beijing 100124, China

<sup>3</sup>School of Traffic and Transportation, Chongqing Jiaotong University, Chongqing 400074, China

Correspondence should be addressed to Enjian Yao; [enjyao@bjtu.edu.cn](mailto:enjyao@bjtu.edu.cn)

Received 9 March 2021; Accepted 31 August 2021; Published 24 September 2021

Academic Editor: Steven Chien

Copyright © 2021 Enjian Yao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Passenger travel flows of urban rail transit during holidays usually show distinct characteristics different from normal days. To ensure efficient operation management, it is essential to accurately predict the distribution of holiday passenger flow. Based on Automatic Fare Collection (AFC) data, this paper explores the passengers' destination choice differences between normal days and holidays, as well as one-way tickets and public transportation cards, which provides support for variable selection in modeling. Then, a forecasting model of holiday travel distribution is proposed, in which the destination choice model is established for representing local and nonlocal passengers. Meanwhile, explanatory variables such as land matching degree, scenic spot dummy, and level of service variables are introduced to deal with the particularity of holiday passengers' travel behavior. The parameters calibrated by the improved weighted exogenous sampling maximum likelihood (WESML) method are applied to predict passenger flow distribution in different holiday cases with annual changes in the metro network, using the data collected from Guangzhou Metro, China. The results show that the proposed model is valid and performs better than the other comparable models in terms of forecasting accuracy. The proposed model has the capability to provide a more universal and accurate passenger flow distribution prediction method for urban rail transit in different holiday scenarios with network changes.

## 1. Introduction

With the development of the economic level, the travel activities and frequencies of urban residents continue to increase, which leads to the rapid growth of urban residents' demand for urban public transport. Urban rail transit has developed rapidly in recent years, and its superiority of traffic volume, speed, and punctuality are popular among people, which helps spur a boom in urban rail construction [1]. In recent years, a large number of new lines have opened and connected to the metro network, making the network operation effect of many cities particularly evident, significantly affecting regional accessibility and passenger flow distribution in the metro network. Furthermore, in regard to holidays, because of the exceptional flexibility of departure time and the diversity of destinations, the passenger travel characteristics are quite distinct from normal days, and the

spatiotemporal distribution of holiday travel demand presents complex characteristics [2, 3].

With the rapid change of the metro network, the operations have undergone quantitative and qualitative changes [4]. The particularity of holidays also aggravates travel demand's complexity, which poses a significant challenge to the metro system. Besides, the same holiday only occurs once a year, which is not conducive to study the characteristics in terms of lacking data sources. Therefore, to effectively organize the large passenger flow and alleviate traffic congestion during holidays, it is essential to accurately predict the distribution of holiday passenger flow, which is the basis of a reasonable train operation plan-making and the development of passenger flow induction strategy.

The traditional four-step methods and their modification models have been widely used in passenger flow distribution forecasting. It mainly includes the aggregate model method

based on statistical rules of historical data and the disaggregate model method based on behavior analysis.

In the study of the aggregate model methods, many researchers have investigated the gravity model by improving it in different contexts. Grosche et al. [5] proposed two gravity models to estimate the air passenger flow between city-pairs. They introduced geoeconomic variables describing the general economic activity and geographical characteristics as independent factors. Wang et al. [6] combined the gravity model considering the distance and free-flow travel time with the Fratar method to predict the seed O-D matrix of the expressway. More recently, Ren et al. [7] proposed three types of land-use function complementarity indices introduced into spatial interaction to improve the gravity model. In these studies, the appropriate variables are introduced to modify the model. Besides, the constrained gravity model is also used as a researching point. Tsekeris and Stathopoulos [8] used a doubly constrained gravity model that additionally incorporates the intraperiod evolution for forecasting the dynamic trip distribution. Jin et al. [9] proposed an O-D estimation model based on the doubly constrained gravity model, where the comparison of singly and doubly constrained models was made. However, the aggregate gravity model tends to overestimate when the distance-deterrence function is small, and the variables are usually less and simple, which cannot reflect the forming mechanism of passenger flow and travel behavior objectively.

The disaggregate model can reveal the internal mechanism of the passengers' destination choice from the perspective of behavior interpretation by establishing definable variables. Specifically, previous studies on the disaggregate model have focused on travel behavior analysis and demand forecasting. For example, in the research of influencing factors of travel behavior, Tsirimpa et al. [10] proposed a multinomial logit model and a mixed multinomial logit model to examine the impact of information acquisition on switching travel behavior. Yang et al. [11] proposed multinomial and nested logit models to analyze battery electric vehicle drivers' charging and route choice behaviors. Nguyen-Phuoc et al. [12] adopted a multinomial logit model to explore factors affecting changes in the event of major public transport disruptions. In addition, the discrete choice modeling technology based on random utility-based is mainly used for destination choice modeling. Faghih-Imani [13] used a multinomial logit model to study the decision process of identifying destination locations at a bicycle station. Kelly [14] built multinomial logit models to analyze the destination choice behaviors of pedestrians within an entire region. Orvin [15] developed a random parameter latent segmentation-based logit model to investigate trip destination choice behavior of the dockless bike-sharing users. These studies show that individual attributes and alternative factors influence passenger behavior and the decision process, assisting transit agencies in getting management guidance.

Focusing on the demand forecasting, Timmermans [16] proposed a model combining transportation mode selection and destination selection and predicted shopping-oriented

travel. To strengthen the forecasting power, Jovicic and Hansen [17] constructed a nested logit model, where logsums integrate generation, distribution, and mode choice models as submodels. Ashiabor et al. [18] developed the nested and mixed logit model to estimate county-to-county travel demand. Travel time, cost, and traveler's household income were used in the explanatory variables. Furthermore, recent studies [19] proposed a multistage demand forecasting model that considers the discrete choice approach, such as the binomial and multinomial logit model, for each decisional level. Moreover, Li [20] presented a new itinerary-based nonlinear demand estimator that estimates the distribution of demand based on a nested logit model. These studies contribute to the accurate prediction of travel demand with the improved disaggregate model. However, it is usually necessary to use questionnaires, such as the stated and revealed preference surveys, to obtain the data that include individual and alternative attributes for studying the behavioral characteristics. When applied to prediction, it is easy to be restricted by data conditions and difficult to use effectively.

In addition, many emerging data mining technologies and methods are used to study traffic or passenger flow demand. Ye and Wen [21] proposed a destination choice model based on link flows by constructing algorithms observing the detected data from part of the links. By using data mining, Wang et al. [22] developed cell phone location tracking algorithms to track cross-region traffic activities and derived the O-D traffic flow and travel demand. In the machine learning approaches, Wang et al. [23] designed a grid embedding network via graph convolution and established a multitask learning network for forecasting the demands of O-D pairs in ride-hailing. Although the prediction accuracy of the data-driven approach depending on long-term collection may be higher, it is hard to apply the network structure changes because of lacking the newly added stations' data in the metro. Moreover, it is often a black-box process that does not illustrate the internal behavior mechanism.

Generally, due to the holidays that occurred only once a year, it is not easy to continuously collect stable and long-term data. And with the rapid development of the metro system, the network structure of holidays usually changes every year, which makes it hard to use the statistical models for prediction. Previous relevant studies focused on researching the passenger flow on normal days. However, little work has explored passengers' choice behavior to construct special variables to effectively forecast the scenarios of holidays in the metro system. Besides, since the source of disaggregate data limits the forecasting application, new data sources are considered to replace the conventional questionnaires in this paper.

At present, the Automatic Fare Collection (AFC) system is widely adopted in the urban rail transit system, which is the main support data in this paper. Under the premise of ensuring validity, this paper applies the aggregate data obtained by the AFC to the disaggregate model by modifying the maximum likelihood estimation method, which overcomes the difficulty of getting the disaggregate model data.

Based on fully exploiting the holiday passenger travel rules and considering the differences in the choice behavior of different ticket type passengers, this paper constructs a holiday passenger flow distribution prediction model, in which some novel explanatory variables (such as land matching degree) are introduced. The proposed model structure can not only be suitable for the changes of urban rail transit network structure but also take into account the unique characteristics of holidays so as to have good interpretability.

The remainder of this paper is organized as follows. In the next section, the holidays' data collection effort and passenger flow characteristics are analyzed herein. Then, the modeling methodology and the explanatory variables of the utility function are described. After that, the proposed model is estimated and applied to the holiday distribution forecasting with comparisons of other traditional methods. Finally, concluding remarks are presented in the last section.

## 2. Data and Passenger Flow's Characteristic

**2.1. Data.** Urban rail transit adopts the AFC system to implement management methods such as ticketing, ticket checking, and billing. The data are gathered and transmitted into the center and automatically store passenger travel information. The data types are shown in Table 1. Under such limited data conditions and types, how to use them to construct a forecasting model suitable for the holiday scenario is the primary goal. In data processing, the data cleaning has been done by identifying outliers, such as judging whether the enter and exit stations are inconsistent, whether the enter and exit time, and the in-train time are reasonable. Besides, the stations are regarded as transportation analysis zones (TAZs) in the urban rail transit system. The boarding (origin) and alighting (destination) stations of passengers' trips can be obtained from the AFC system directly.

There are eight lines and 140 stations in Guangzhou Metro by the beginning of 2016. The daily average of raw data amounts to more than 4 million that need further processing. And more than one million passengers use one-way tickets per day during New Year's Day, which is almost 1.84 times the weekdays. Compared with January 1, 2016, there are seventeen new stations and three new lines connected to the network on January 1, 2017. The road network structure has tremendous changes.

**2.2. Passenger Flow's Characteristics.** Based on Guangzhou Metro's AFC data, the passenger flow of each station during the New Year's Day holiday from 2016 to 2017 is collected, and some travel characteristics have been found. The passenger flow, for instance, is closely related to the nature of land-use and the intensity of development around stations.

As shown in Figure 1, the entrance passenger flows of four typical stations from December 30, 2015, to January 4, 2016, are given. The passenger flow of Zhujiangxincheng station, which is dominated by office areas, declined significantly during the New Year's Day. Similarly, the passenger flow of Dashadong station also decreased, as

residential areas surround there. However, Guangzhouta and Beijinglu stations' passenger flow increased significantly during the holidays, with the main areas, respectively, surrounded by scenic spots and commercial districts.

Similarly, the passenger flow, from the origin station to the destination station (O-D station) during the holidays, shows different distinct characteristics, compared with weekdays. As shown in Figure 2, there are different passenger flow trends between O-D stations with different land-use types, and some of which increased significantly in holidays, while others, such as residential stations to office stations, dropped significantly.

From another perspective, there are also great differences in the distribution of people who use one-way tickets and public transportation cards during holidays. Generally, many one-way passengers are nonlocal passengers, who tend to go to scenic spots, business districts, and hub stations. In contrast, transportation card passengers are mostly local residents, whose travel purposes are diversified. This characteristic of choice behavior is especially evident during holidays. As shown in Figure 3, the passenger flow of one-way tickets and public transportation cards at Guangzhouta and Beijinglu stations has increased, while the growth rate of one-way ticket is significantly higher, indicating that the stronger attraction of one-way ticket passengers.

Furthermore, other characteristics can also be obtained by analyzing the passenger flow. For example, the O-D passenger flow on the same line is usually larger than that on different lines. And in the case of satisfying the purpose of passengers, they would give priority to the destination with a short ride time and transfer time. However, these features are influenced by many factors. They should be reflected in some explanatory variables to analyze how various factors jointly affect the behavior and improve the subsequent forecasting performance when modeling. Next, the approach considering passenger flow's characteristics of holidays is introduced in detail.

## 3. Methodology

Considering that the metro network scale is rapidly developing, the spatial passenger flow distribution of O-D stations also changes fast. New stations divert the passenger flow of old stations, and it is not easy to obtain the development data of all O-D pairs in time series, especially for newly added stations. Therefore, based on the above analysis of passenger flow's features, this paper constructs a destination choice model to describe the characteristics of passengers. Then, a forecasting model of holiday passenger flow distribution is developed, which is suitable for the structural change of the network and does not depend on long-term data collection. Meanwhile, considering different passengers' characteristics, the utility functions for passengers who use one-way tickets and public transportation cards are constructed separately.

**3.1. Model Structure.** The theory of random utility maximization refers to the alternatives in which traffic behavior decision-makers choose the most effective ones in their

TABLE 1: The Automatic Fare Collection data types.

Name	Field type	Form
Card type	Varchar	One-way ticket, public transportation card, etc.
Card number	Int	1000139***
Enter line	Varchar	Line 1, Line 2, etc.
Enter station name	Varchar	Guangzhou East, Donghu station, etc.
Enter time	Datetime	“2016-01-01 08:00:00”
Exit line name	Varchar	Line 1, Line 2, etc.
Exit station name	Varchar	Guangzhou East, Donghu station, etc.
Exit time	Datetime	“2016-01-01 08:00:00”

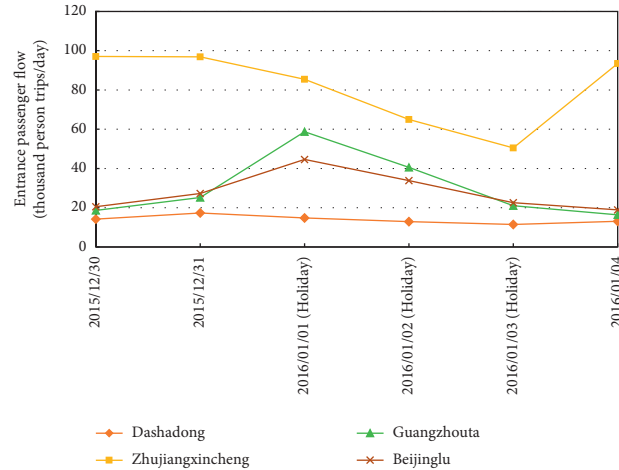


FIGURE 1: Entrance passenger flow of typical stations during New Year’s Day.

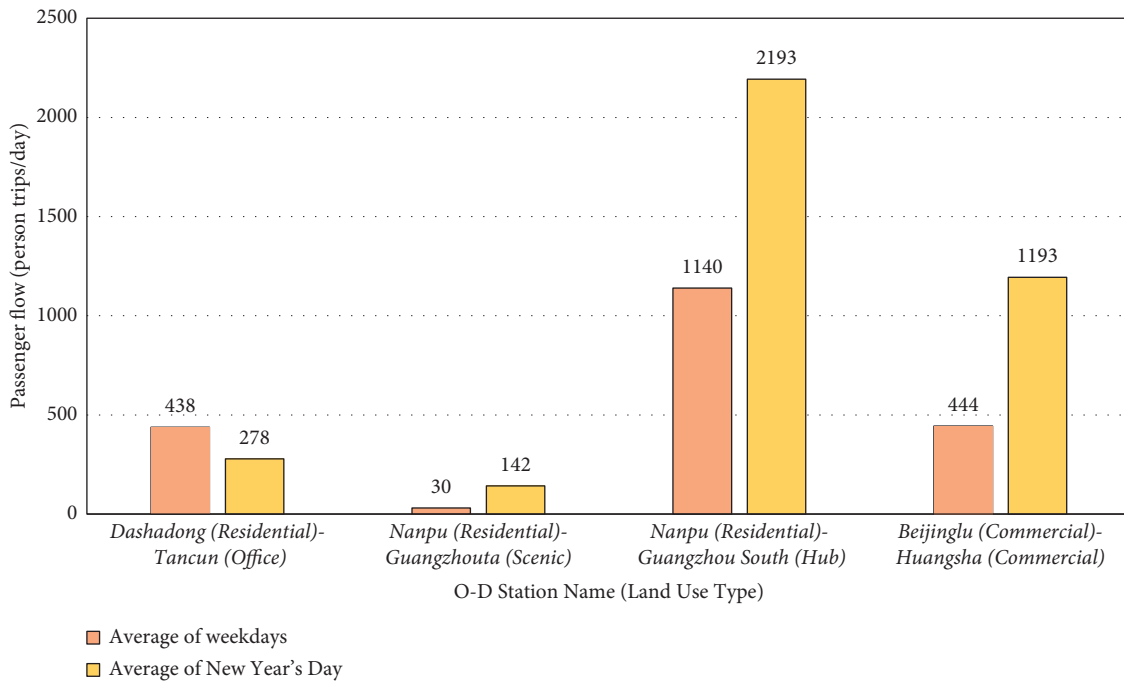


FIGURE 2: Comparison of O-D passenger flow during weekdays and New Year’s Day in 2016.

choice sets under certain conditions. If the destination choice sets of passengers from station  $i$  are  $A_i$  and the utility of the alternative  $n$  is  $U_{in}$ , the requirement that the

passengers select the destination  $j$  from  $A_i$  is  $U_{ij} > U_{in}$ . Among them, the utility function  $U$  has divided into two parts: a deterministic term  $V_{ij}$  and an error term  $\epsilon_{ij}$ .

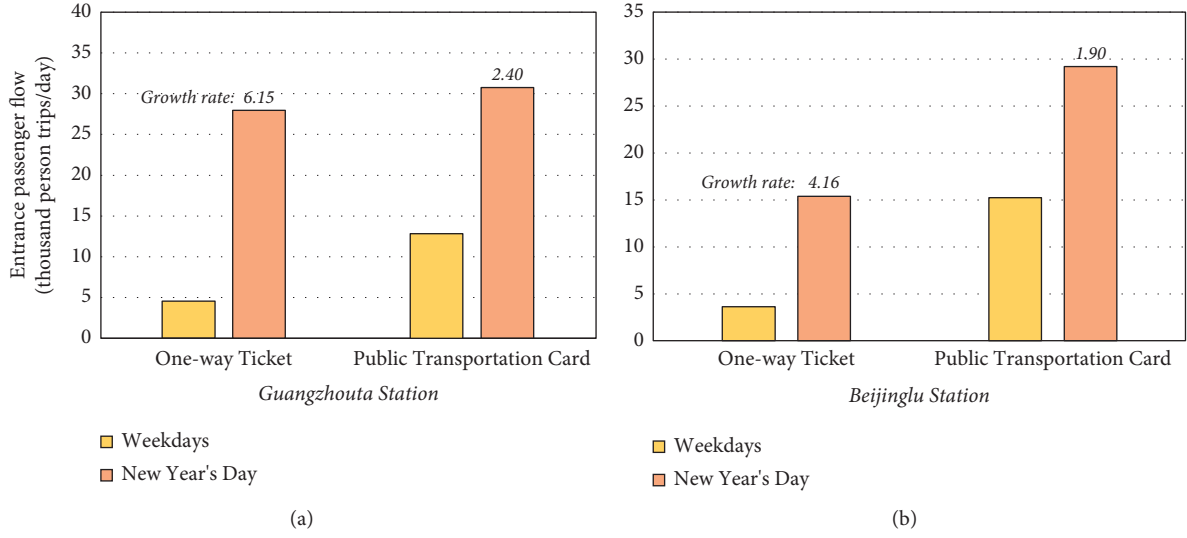


FIGURE 3: Comparison of passenger flow between one-way ticket and public transportation card in 2016. (a) Guangzhouta station. (b) Beijinglu station.

Therefore, the utility function  $U_{ij}$  can be formulated as follows:

$$U_{ij} = V_{ij} + \varepsilon_{ij} = \sum \beta^k X_{ij}^k + \varepsilon_{ij}, \quad (1)$$

where  $\beta^k$  is an estimable parameter of attribute  $k$ ;  $X_{ij}^k$  is an observable attribute as the explanatory variable;  $\varepsilon_{ij}$  is the error term that is used to address the unobserved factors that influence the choices taken by the passengers.

The researcher observes some attributes of the alternatives as faced by the decision maker, labeled  $X_{ij}$ , and can specify a function that relates these observed factors to the decision maker's utility [24]. The term  $\varepsilon_{ij}$  is treated as random, and it captures the factors that affect utility but are not included in  $V_{ij}$ . When the error term  $\varepsilon_{ij}$  obeys the independent Gumbel distribution, multinomial logit (MNL) models can be derived. For the origin station  $i$ , the probability for choosing  $j$  is calculated as follows:

$$P_{ij} = \frac{\exp(V_{ij})}{\sum_{n \in A_i} \exp(V_{in})}, \quad j \in A_i. \quad (2)$$

Equation (2) is the destination choice model. The probability that a passenger chooses another station as the destination can be calculated. The production trips from each station are then distributed to all other stations based on the choice of probability destination. That is, the passenger flow distribution  $q_{ij}$  from the origin station  $i$  to the destination station  $j$  is computed. The formula is shown as follows:

$$q_{ij} = Q_i \cdot P_{ij}, \quad (3)$$

where  $Q_i$  is the entrance passenger flow in station  $i$ .

Considering the different levels of sensitivity of travel characteristics of different types of passengers in the particularity of holidays, two utility functions in the proposed model are constructed with passengers who use one-way tickets and public transportation cards for representing local

and nonlocal passengers. The trip distribution is applied separately for each ticket type of passengers who have characteristic travel behavior, with different model parameters. Then, the distribution results of the two ticket types are added together. The formula is shown as follows:

$$q_{ij} = q_{ij}^{\text{one}} + q_{ij}^{\text{ptc}}, \quad (4)$$

where  $q_{ij}^{\text{one}}$  is the one-way ticket passengers' distribution prediction;  $q_{ij}^{\text{ptc}}$  is the public transportation card passengers' distribution prediction.

Equation (4) is the forecasting model of passenger flow distribution. However, it is a singly constrained model so far. There is no guarantee that the sum of the passenger flow from each station to the destination station  $j$  is equal to the attracted trips of station  $j$ . Therefore, it is necessary to modify the travel flow to enforce constraints between total origins and destinations. The Fratar method is widely used in distribution adjustment due to its fast convergence speed and high calculation accuracy. The idea of the Fratar method is a distribution of horizon year trips from a zone that is proportional to the base year trip distribution pattern modified by the growth factors of the zones under consideration [25, 26]. Therefore, this paper uses the Fratar method for equalization processing. The approach is shown as follows:

$$q_{ij}^{m+1} = q_{ij}^m \cdot F_{O_i}^m \cdot F_{D_j}^m \cdot \left( \frac{L_i + L_j}{2} \right), \quad (5)$$

$$L_i = \frac{O_i^m}{\sum_j q_{ij}^m \cdot F_{D_j}^m},$$

$$L_j = \frac{D_j^m}{\sum_i q_{ij}^m \cdot F_{O_i}^m}$$



where  $q_{ij}$  is the passenger flow of station  $i$  to station  $j$ ;  $F_{O_i}$  is the growth rate of the entrance passenger flow in station  $i$ ;  $F_{D_j}$  is the growth rate of the exit passenger flow in station  $j$ ;  $L_i$  is the adjustment coefficient of station  $i$ ;  $L_j$  is the adjustment coefficient of station  $j$ ;  $O_i$  is the entrance passenger flow in station  $i$ ;  $D_j$  is the exit passenger flow in station  $j$ ;  $m$  is the  $m$ -th iteration.

**3.2. Model Specifications.** Although personal characteristics affect destination choice, it is unable to obtain personal attributes data from the AFC directly. Therefore, seven indexes as the utility function of characteristic variables that could be extracted from the urban rail transit network are considered in the destination choice model, including in-vehicle travel time, transfer time, station position relationship, and matching degree of land-use types. The seven variables are mainly used to characterize three categories of explanatory attributes, namely, the accessibility of the destination, the attractiveness of the destination, and matching degree of O-D stations, through which the choice behavior mechanism of passengers can be characterized.

According to the choice behavior characteristics of the one-way ticket and the public transportation card passengers, and through the multiple calibration experience of the model, the utility functions  $V_{ij}^{\text{one}}$  and  $V_{ij}^{\text{ptc}}$  of the destination choice model are constructed, as shown in equations (6) and (7), respectively:

$$V_{ij}^{\text{one}} = \beta_1 D_j + \beta_2 Z_{ij} + \frac{\beta_3 T_{ij}^{\text{train}}}{3600} + \frac{\beta_4 N_{ij}^{\text{trans}}}{3600} + \beta_5 G_{ij} + \beta_6 S_{ij} + \beta_7 L_j, \quad (6)$$

$$V_{ij}^{\text{ptc}} = \beta_8 D_j + \beta_9 Z_{ij} + \frac{\beta_{10} T_{ij}^{\text{train}}}{3600} + \frac{\beta_{11} N_{ij}^{\text{trans}}}{3600} + \beta_{12} G_{ij} + \beta_{13} S_{ij}, \quad (7)$$

where  $\beta_1 - \beta_{13}$  is the parameter to be calibrated for each variable;  $D_j$  is the exit passenger flow of destination station  $j$ , ten thousand person trips;  $Z_{ij}$  is the matching degree of land-use type;  $T_{ij}^{\text{train}}$  is the in-vehicle travel time from the origin station  $i$  to the destination station  $j$ , second;  $N_{ij}^{\text{trans}}$  is the transfer time from the origin station  $i$  to the destination station  $j$ , second;  $G_{ij}$  is a dummy variable, and if the sum of trip generation at origin station  $i$  and the attraction at destination station  $j$  is larger than a specific scale, the value is 1;  $S_{ij}$  is a dummy variable, and if the origin station  $i$  and destination station  $j$  are in the same line, the value is 1;  $L_j$  is a dummy variable, and if the land-use type of destination station  $j$  is scenic, commercial, or hub, the value is 1.

For one thing, these variables are introduced to facilitate data acquisition, and for another, the characteristics of holidays are considered so as to improve the interpretability and prediction effect of the model further. It should be noted that the travel cost is a sensitive variable to influence the choice behavior, which was included in the variable sets at the beginning. However, when the variables are checked for multicollinearity, the travel cost shows a strong correlation with the travel time. Therefore,

the travel cost was eliminated in the utility functions. Compared to one-way ticket passengers, the public transportation card utility functions do not have the variable  $L_j$ , as adding this variable would reduce the model's accuracy.

Moreover, the acquisition of the matching degree of land-use types  $Z_{ij}$  and the scenic destination station variables  $L_j$  need to be additionally explained. The distribution of passenger flow between stations is closely related to land-use nature around the station, especially the significant difference between holidays and normal days. It is necessary to quantify the impact of land-use interaction. Therefore,  $Z_{ij}$  is constructed to describe the degree of attraction between different types of stations. Based on this, the metro stations need to be clustered to determine the category of the station first.

Due to the land-use properties are a relatively stable indicator and it usually shows a certain relationship with the passenger flow characteristics, the  $K$ -means clustering method is used to classify the stations of the whole network of Guangzhou Metro.  $K$ -means is a vector quantization method that is popular for cluster analysis in data mining [27]. Through the analysis of passenger flow characteristics, the morning and evening peak flow has a greater correlation with the nature of land-use around stations. And the proportion of one-way tickets and all-day passenger flow at comprehensive transportation hubs is usually larger, while the passenger flow at commercial and scenic stations tends to increase significantly during holidays. Therefore, the five variables are used as inputs for clustering as shown in Table 2. In the clustering research of metro stations, the stations are usually divided into five categories according to weekday travel data [28, 29]. However, since the research scenarios are aimed at holidays, we set eight cluster numbers as preset categories according to the land-use and application requirements of the model. The clustering results are shown in Table 3 (figures in brackets denote the sum number of clustering stations), and they are representative and matched with the preset types.

Therefore, the value of  $L_j$  can be obtained directly through the clustering results. Besides, the matching degree of land-use type  $Z_{ij}$  needs further processing. Based on the above clustering results, the average O-D passenger flow with different cluster types could be calculated. Then, the logarithm function is used to normalize the values of various types to differentiate passenger flow better. The formula is as follows:

$$Z_{ij} = \frac{(\ln Q_{ij} - \ln \min_{i,j} Q_{ij})}{(\ln \max_{i,j} Q_{ij} - \ln \min_{i,j} Q_{ij})}, \quad (8)$$

where  $Z_{ij}$  is land matching degree from type  $i$  to type  $j$ ;  $Q_{ij}$  is the average O-D passenger flow of the stations from type  $i$  to type  $j$ .

A case result of  $Z_{ij}$  is shown in Table 4 (the vertical column indicates the type of the origin station, and the horizontal row indicates the type of the destination station), where the value from Type1 to Type 1 is zero. This means that the passenger flow is the lowest of all type pairs, mainly because the attraction between residential stations is less during holidays in all type

TABLE 2: The variables for  $K$ -means.

Variables	Description
Morning peak hour factor	The passenger flow of morning peak hour (7:00–9:00) divided by all-day passenger flow
Evening peak hour factor	The passenger flow of evening peak hour (17:00–19:00) divided by all-day passenger flow
Proportion of one-way ticket	The proportion of passengers using one-way ticket
All-day passenger flow	The all-day passenger flow in the stations
Passenger flow growth rate	The passenger flow of the holiday divided by the weekdays before the holiday

TABLE 3: Station types' clustering results based on the surrounding area's land-use.

Type	Attribute	Description	Clustering results
Type 1	Residential	The station is surrounded by residential areas	Nanpu, Sanxi, Dongpu, etc. (30)
Type 2	The majority of residential area	The station is surrounded by majority of residential areas	Shiqiao, Meihuay, Xicun, etc. (25)
Type 3	Office	The station is surrounded by office areas	Haizhuguangc, Quzhuang, etc. (20)
Type 4	The majority of office area	The station is surrounded by majority of office areas	Donghu, Ximenkou, etc. (21)
Type 5	Comprehensive	The station is surrounded by various types of land-use	Shibi, Fangcun, Shiergong, etc. (28)
Type 6	Commercial	The station is surrounded by commercial areas	Jinzhou, Jinronggaoxinqu, etc. (14)
Type 7	Scenic	The station is close to scenic spot	Diyong, Guangzhouta, etc. (14)
Type 8	Hub center	The station functions as transportation hub center	Guangzhou South, Airport South, etc. (5)

TABLE 4: Land matching degree values of different types of O-D stations.

	Type 1	Type 2	Type 3	Type 4	Type 5	Type 6	Type 7	Type 8
Type 1	0.00	0.08	0.24	0.08	0.10	0.18	0.28	0.55
Type 2	0.09	0.15	0.28	0.15	0.16	0.24	0.29	0.63
Type 3	0.09	0.17	0.35	0.18	0.18	0.20	0.27	0.61
Type 4	0.22	0.25	0.28	0.29	0.22	0.22	0.29	0.64
Type 5	0.11	0.15	0.24	0.16	0.22	0.22	0.36	0.60
Type 6	0.17	0.23	0.25	0.18	0.23	0.14	0.26	0.59
Type 7	0.28	0.26	0.25	0.19	0.34	0.22	0.39	0.59
Type 8	0.51	0.59	0.61	0.55	0.57	0.56	0.58	1.00

pairs. In contrast, the connections between transportation hubs are strengthened, reflected in the maximum value from Type 8 to Type 8.

**3.3. Parameter Estimation.** For the parameters  $\beta$  in equation (1), the personal travel survey is generally performed by the simple random sampling method to obtain the disaggregate type data of the individual choice, thereby using the maximum likelihood estimation method to calibrate the parameters. However, in this paper, the aggregate data obtained by the AFC should be transformed into the disaggregate form for application in the destination choice model. When being applied, it needs methods to deal with the original aggregate data. Yao and Takayuki [30] proposed an integrated model that combines estimation across multiple data sources such as SP, RP, and aggregate data. Therefore, the maximum likelihood estimation method is improved by introducing a weight factor to realize the application of AFC data in the destination choice model' calibration.

Manski and Lerman [31] proposed a weighted exogenous sampling maximum likelihood (WESML) method, introducing weights into log-likelihood functions to calibrate the bias between the sample and population data. It can be expressed as follows:

$$L(\theta) = \sum_{i \in A_n} \sum_n \delta_{in} w_i \ln(P_{in}), \quad (9)$$

$$w_i = \frac{Q_i}{H_i}, \quad i \in C, \quad (10)$$

where  $\delta_{in}$  is 1 if the passenger  $n$  chooses selected branch  $i$  as destination and 0 otherwise;  $w_i$  is weights;  $Q_i$  is the proportion of the selected branch  $i$  in the population;  $H_i$  is the proportion of  $i$  in the sample.

To improve the practicability of the method, Cosslett's research [32] proves that it can be transformed as follows:

$$w_i = \frac{Q_i}{N_i/N}, \quad (11)$$

where  $N_i$  is the data amount of the selected branch  $i$ ;  $N$  is the sum of the data amounts of the respective selected branch.

However, in terms of urban rail transit, passengers with the same origin and destination station have the same characteristics; that is, they all make the same choice for the destination. Thus, the amount of O-D passenger flow can be expressed as the selection result of individuals. The weight factor is suited for adjusting the likelihood function of the dataset. Therefore, according to the characteristics of the data that can be extracted, equation (11) is corrected as follows:

$$w_i = \frac{q_i}{\sum q_i} \cdot R, \quad (12)$$

where  $q_i$  is the O-D passenger flow in the selected branch  $i$ ;  $R$  is the number of individuals, that is, the sum of O-D station pairs.

## 4. Results and Analysis

**4.1. Model Estimation and Analysis.** In the construction of the selection set, there are 140 stations in the 2016 New Year's Day. That is to say, 139 stations should be put into the alternative set except for the real choice of each traveler. However, for general disaggregate models, the size of the alternatives is too large, which would affect the speed of model estimation and is not conducive to application. Ben-Akiva and Lerman [33] demonstrated that the consistency of model parameters is not lost when extracting subselective branches for parameter estimation in the selection set. Therefore, this paper constructed the subselection set by randomly extracting nine stations from the alternative set. It could reduce the difficulty of calibration and increase the operability while ensuring the consistency of the model's calibrated parameters.

In the process of parameter calibration, the values of seven variables are obtained in combination with the network topology and train operation plan of Guangzhou Urban Rail Transit. By using the parameter estimation method described in the section before, the undetermined parameters of the utility function are calibrated. Especially, after several tests, the scale dummy variable  $G_{ij}$  was set to 1 if it exceeds 7,000 person trips. The calibration results of the New Year's Day are shown in Table 5 as a study case. All absolute  $t$ -values are greater than 1.96, indicating statistical significance and variables' validation. Moreover, the adjusted  $\rho^2$  of this model is over 0.2, which can be regarded as a satisfactory goodness-of-fit [34].

The estimated parameters are provided with practical significance and expected signs in the sense of explaining passenger destination choice behavior in either the one-way ticket model or public transportation card model. An obvious example is that the parameter of destination attraction variable is positive, indicating that the greater attraction of destination station is, the more passengers choose.

As for the negative parameters of travel time and transfer time, the longer the travel time and transfer times are, the less probability of destination station would be chosen, which is consistent with common sense and inversely proportional to destination choice preference. Moreover, the units are the same, but the estimated parameters are not close, which means the travelers have different perceptions. The trade-off between travel and transfer time shows that an increase of 10 minutes in transfer time is equivalent to an increase of 68 minutes in travel time for one-way ticket passengers and 55 minutes for public transportation card passengers in the case of New Year's Day. It reveals that travelers have a significant negative impact on lengthy transfer times. For public transportation card passengers, the absolute parameters of travel time and transfer time are both

larger than the one-way ticket passengers, indicating that the passengers who used the card care more about the time when other variables remain unchanged.

Besides, the land matching degree's parameter is positive, which indicates that when the relationship between O-D station's land-use types is strong, the destination stations will be more likely to be chosen. The scale and collinear variable's parameters are positive, revealing that when the origin and destination stations' travel scale is more extensive, or the O-D station stands on the same line, the probability of the destination station being chosen is greater.

For the scenic variable in the one-way ticket model, its parameter is positive. It is also in line with the characteristics of passengers traveling on holidays because there are plenty of tourists who use the one-way tickets. In general, the estimated results are statistically significance and can explain the choice behavior mechanism on the New Year's Day to some degree. However, it is worth emphasizing that the parameters should be recalibrated so as to regain the travel behaviors when applying other different holidays.

**4.2. Model Application and Comparison.** To test the predictive effect of the proposed forecasting model, the calibrated results are used to predict the New Year's Day of Guangzhou Metro on January 1, 2017, where the data of the predicted year are used as the test-set and do not participate in the calibration. There are seventeen new stations and three new lines connected to the network. Meanwhile, the singly constrained gravity (SCG) model in the traditional statistical model, the support vector machine (SVM), the back propagation (BP) neural network, and radial basis function (RBF) neural network in machine learning model are selected for comparison under the same data source and conditions. And the traffic impedance function in the form of the exponential function is used in the gravity model, as shown in equation (13). The least-square method is used to transform it into a linear form for parameter estimation [35]:

$$f_{ij} = \exp(-\mu T_{ij} - \tau n_{ij}) T_{ij}^{-\gamma}, \quad (13)$$

where  $T_{ij}$  and  $n_{ij}$  are travel time and transfer time from the origin station  $i$  to the destination station  $j$ , respectively;  $\mu$ ,  $\tau$ , and  $\gamma$  are the coefficients to be determined.

As shown in Figure 4(a)–4(e), the predicted values are compared with the actual passenger travel data, and the prediction deviation graph is drawn. The error fluctuation of the singly constrained gravity model and the other three machine learning models is larger than the proposed forecasting model established in this paper. The mean absolute error of the whole network in the gravity model is 130.2 person trips, the SVM model is 140.9, the BP neural network is 139.1, and the RBF neural network is 157.3, while the proposed model is 54.6 that is far better.

Furthermore, the detailed prediction error statistics of the five models, in this case, are shown in Table 6. Compared with the other four models, the mean absolute error of the proposed model is reduced by 58.05%, 61.21%, 60.72%, and



TABLE 5: Model estimation results.

Characteristic variable	Case1: the New Year's Day				Case2: the National Day			
	One-way ticket		Public transport card		One-way ticket		Public transport card	
	$\beta_i$	$t$ -value	$\beta_i$	$t$ -value	$\beta_i$	$t$ -value	$\beta_i$	$t$ -value
Destination attraction ( $D_j$ )	0.301	37.989	0.084	26.818	0.290	38.791	0.211	28.646
Land matching degree ( $Z_{ij}$ )	0.311	3.641	0.424	6.802	0.257	2.367	0.532	8.481
Collinear variable ( $S_{ij}$ )	0.533	18.190	0.549	18.759	0.493	17.179	0.553	18.812
Scale variable ( $G_{ij}$ )	0.638	29.106	0.533	23.835	0.629	29.519	0.517	23.696
Travel time ( $T_{ij}^{train}$ )	-0.815	-18.286	-1.172	-25.162	-0.508	-11.884	-1.094	-23.556
Transfer time ( $N_{ij}^{trans}$ )	-5.507	-18.324	-6.449	-20.833	-5.975	-20.316	-6.766	-21.519
Scenic variables ( $L_{ij}$ )	0.084	3.971	—	—	0.092	3.350	—	—
Observations	17954		18314		17534		17747	
Model summary	$L(\theta_{max})$	-32640.37	-33441.36	-31654.83	-32615.35			
	$L(0)$	-41340.61	-42169.54	-40373.53	-40863.98			
	Adjusted $\rho^2$	0.210	0.207	0.216	0.202			

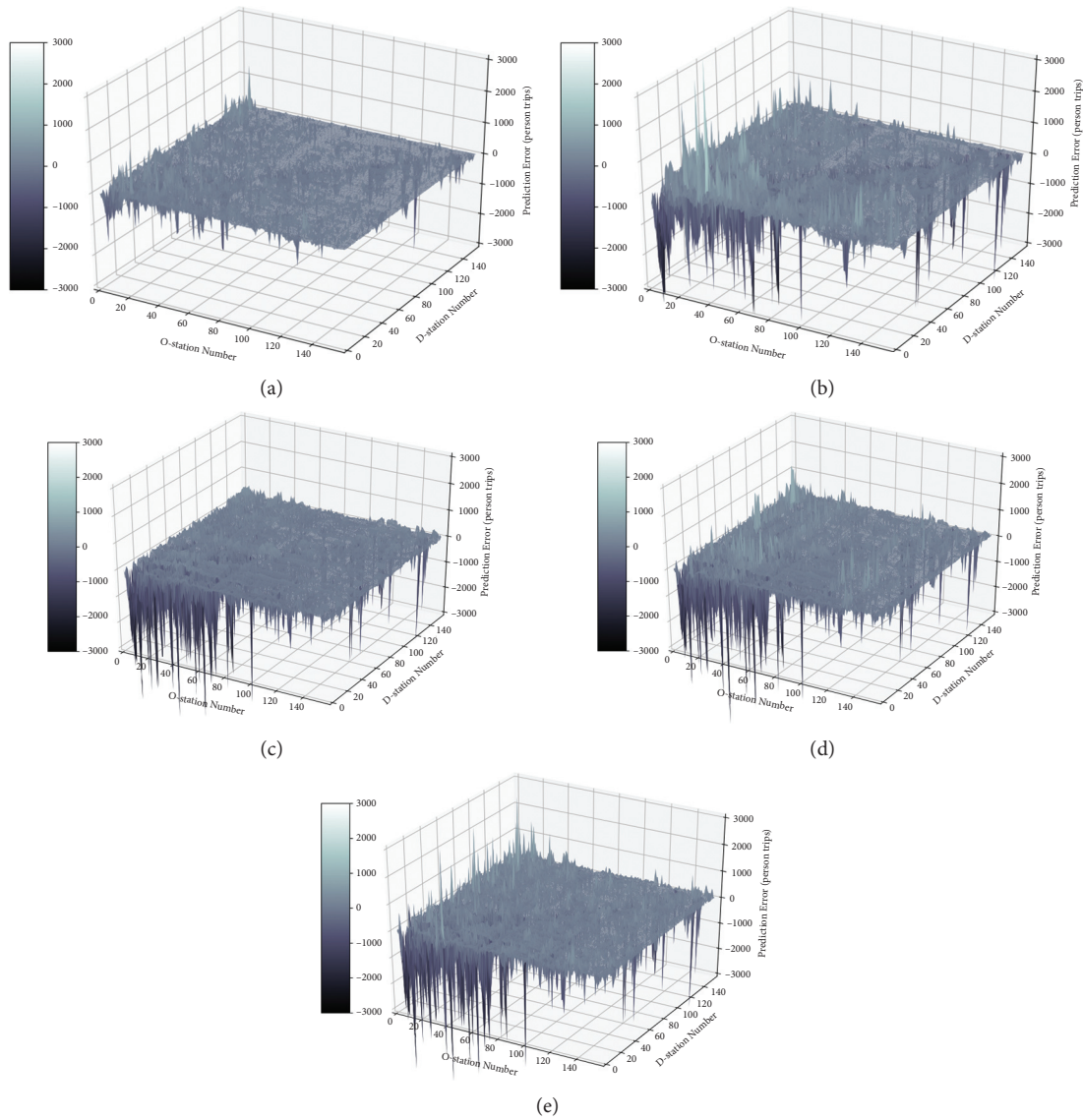


FIGURE 4: Prediction deviation of the proposed model and the comparison models on the New Year's Day. (a) The proposed model. (b) The gravity model. (c) The support vector machine model. (d) The back propagation neural network model. (e) The radial basis function neural network model.

TABLE 6: Statistics of model deviation in the New Year's Day.

Case1: the New Year's Day	SCG	SVM	BP	RBF	The proposed model
Maximum of absolute error (person trips)	5071	6099	6240	5932	2506
Mean absolute error (person trips)	130.22	140.86	139.08	157.27	54.63
Proportion of absolute errors over 50 trips (%)	50.13	56.12	57.57	66.3	26.90
Proportion of absolute errors over 100 trips (%)	31.31	31.83	35.71	39.29	14.78
Proportion of absolute errors over 200 trips (%)	16.27	15.17	14.74	17.15	5.90
Proportion of relative error over 20% (%)	84.60	85.11	84.00	85.95	62.74
Proportion of relative error over 50% (%)	66.78	64.34	64.08	68.37	33.17
Proportion of relative error over 100% (%)	40.45	37.55	45.70	46.07	16.39

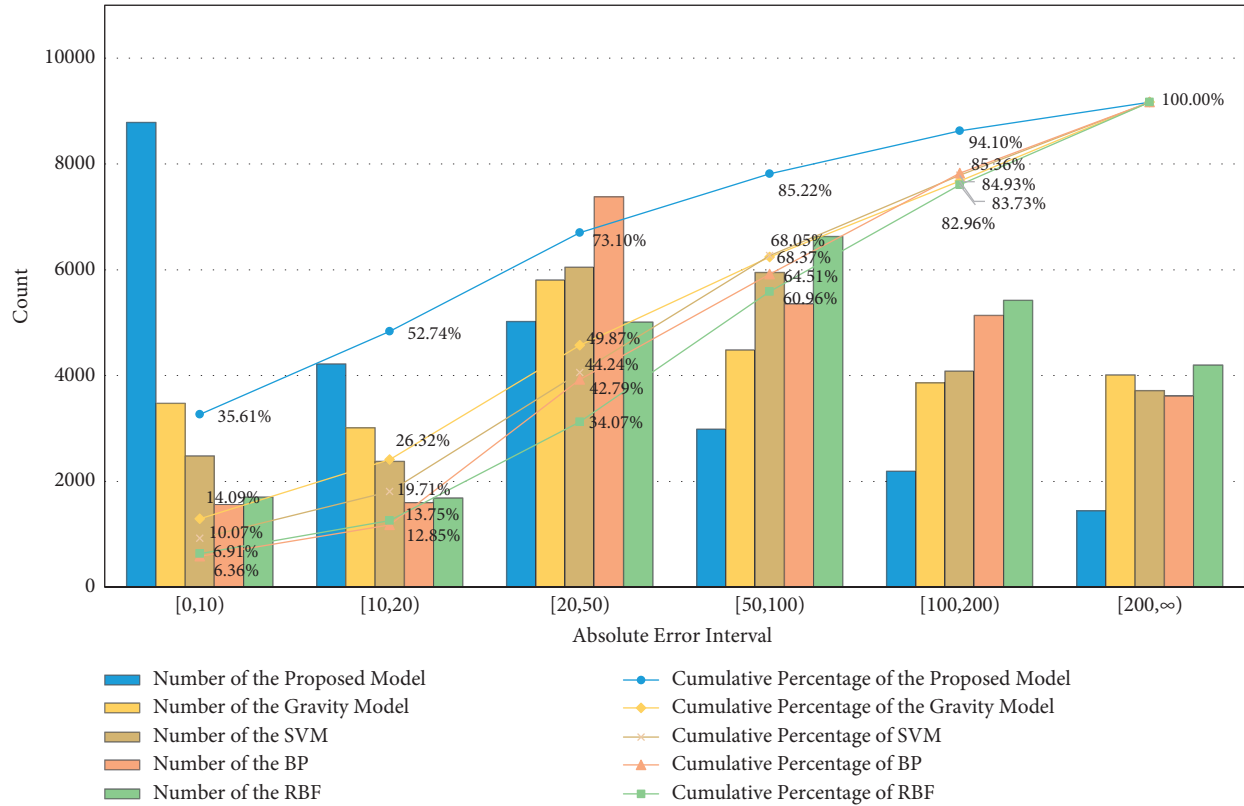


FIGURE 5: Comparison of absolute error and cumulative percentage.

65.26%, respectively. The proportion of absolute errors of the proposed model under 50 person trips reaches 73.1%, and the relative error less than 50% is 66.83%, where the errors are better than that of the other four models.

A detailed comparison of absolute error and its cumulative percentage can be seen in Figure 5. The statistics also show that the proposed model accuracy is better than the conventional gravity model, SVM, and the two neural network models as a whole. However, the proposed model has a slightly weak performance in terms of relative error, mainly because there are many O-D stations with small basic flow, leading to a large relative error. For example, the proportion of relative error more than 200% is 6.78%, where the average absolute error is 41.0 person trips. Moreover, the proportion of relative error more than 500% is 1.70%, where the average absolute error is 35.40 person trips, which is below the total average absolute error. Therefore, it does not

mean that the poorer the relative error, the larger the absolute error, and the worse the prediction performance. The prediction effect of the proposed model can still be guaranteed.

Besides that, the error results of different categories between new lines and existing lines in the models are shown in Table 7. The proposed model's mean absolute error results are relatively low when predicting the new line, namely, only 23.14 and 23.26 person trips. In the prediction performance of the existing line to existing line, the error is relatively larger than that of others, mainly because of the large basic flow between existing stations.

In this case, the holiday of New Year's Day is chosen for analysis. However, other holidays might be a little longer in time, and passenger flow patterns and choice behavior would be different in some ways. The proposed destination choice model could be used to reflect the choice behavior

TABLE 7: Mean absolute error of prediction results.

Category	Mean absolute error (person trips)				
	SCG	SVM	BP	RBF	The proposed model
Existing to new lines	49.92	99.81	67.26	82.96	23.14
New line to existing line	48.07	76.62	119.80	105.64	23.26
New line to new line	90.03	141.72	96.56	131.55	41.30
Existing line to existing line	154.94	152.50	152.12	171.68	64.10
<b>Whole network</b>	<b>130.22</b>	<b>140.86</b>	<b>139.08</b>	<b>157.27</b>	<b>54.63</b>

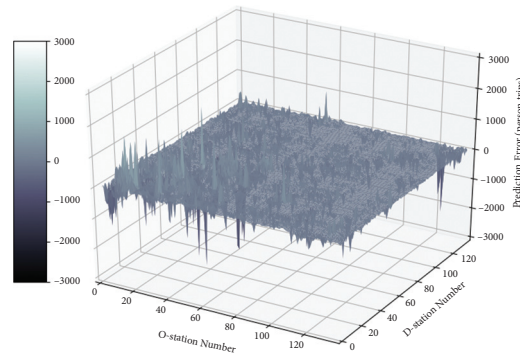


FIGURE 6: Prediction deviation of the proposed model on the National Day.

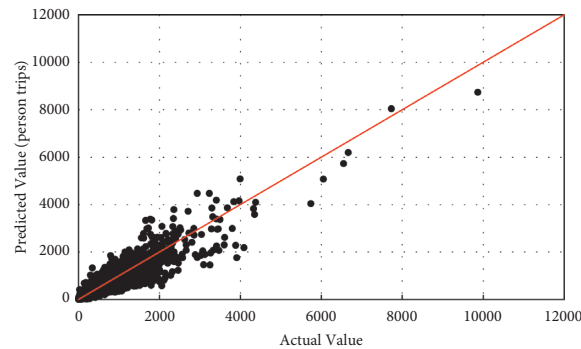


FIGURE 7: Scatter plot of predicted and actual value result presentation.

TABLE 8: Statistics of model deviation in the National Day.

Case2: the National Day	SCG	SVM	BP	RBF	The proposed model
Maximum of absolute error (person trips)	5681	8931	6730	4320	2467
Mean absolute error (person trips)	129.07	110.29	109.25	79.58	51.63
Proportion of absolute errors over 50 trips (%)	51.58	43.63	48.34	37.53	23.49
Proportion of absolute errors over 100 trips (%)	32.44	26.16	26.99	20.68	12.55
Proportion of absolute errors over 200 trips (%)	16.53	12.58	12.40	9.70	5.47
Proportion of relative error over 20% (%)	81.10	84.66	82.45	73.06	65.15
Proportion of relative error over 50% (%)	55.40	62.93	59.80	43.66	31.61
Proportion of relative error over 100% (%)	31.86	35.73	38.52	26.56	14.06

characteristics and passenger flow rules, so the methodology applies to all holidays. Considering the validation and portability of the proposed method, this study supplemented a case of the National Day (seven-day holiday) for a relatively comprehensive experimental design. One day of the National Day in 2014 was randomly selected for model

estimation (that is, October 2, 2014), and the proposed method was used to predict the passenger flow distribution on the same day of next year.

There is one small difference in the calibrated parameters as shown in Table 5 above, which reflects the slight distinction of the travel characteristics in different holidays. However, all

absolute  $t$ -values are still greater than 1.96, and the adjusted  $\rho^2$  is over 0.2, which indicates that the model is still applicable and reliable. The prediction deviation graph is also drawn to show the overall error, as shown in Figure 6. For the convenience of reading, the scatter plot is given as shown in Figure 7, which is consistent with the meaning expressed in Figure 6. The graph plots values for the modeled prediction along the  $Y$ -axis and the corresponding actual count along the  $X$ -axis. If all of the predictions match the actual value exactly, the points on the graph would match up with the red line (45 degree line) drawn in the graph. The prediction results of the proposed model are mostly close to the red line, illustrating that the prediction performs well. The National Day's comparison models' error statistics are shown in Table 8. In summary, the prediction effect and accuracy are still ideal than the other models. The validation and applicability in other holiday scenarios can still be guaranteed. And it can be more effectively applied to practical engineering.

## 5. Conclusions

This paper utilizes AFC data to propose a forecasting model for passenger flow distribution for urban rail transit, which is suitable for network structure and the unique characteristics of holidays. The weighted exogenous sampling maximum likelihood (WESML) estimation method is used to calibrate the parameters. The aggregate data extracted from AFC are transformed into the disaggregate form, which realizes the valid calibration of the parameters. It reduces the difficulty of data acquisition and enhances the applicability of the model, meanwhile ensuring acceptable accuracy.

In the proposed model, the destination choice model defines destination attraction, land matching degree, and others as explanatory variables. This is the main advantage of the model's interpretability and predictive power. The model presents reasonable performance because  $t$ -values are all greater than 1.96, and the moderately adjusted  $\rho^2$  is over 0.2. Moreover, the calibration results show that both travel and transfer time have significant negative effects on passengers' destination choice, while other variables such as destination attraction and land matching degree have a positive influence. The results also show that the public transportation card passengers care more about both travel and transfer time when other variables remain unchanged. The dummy variables used to describe the attractiveness and accessibility of the destination also have reasonable interpretability and significance. The proposed model is applied to predict two cases of Guangzhou Metro on New Year's Day and National Day. Compared with the gravity model, SVM, BP, and RBF models, the proposed model's error is greatly reduced, which proves the validation and applicability of the prediction model in different holiday scenarios with network changes.

As more cities rely on metro systems, accurately forecasted holiday passenger flow distribution could provide important primary data for the metro operation management department to develop a useful organization scheme before the holiday period, which is conducive to easing congestion and improving holiday emergency response capabilities.

Since it is difficult to obtain real land-use data around stations, this paper clusters the stations with similar passenger flow characteristics and defines new variables describing the land-use connection into the model. Nonetheless, the impact of significant land-use changes on passenger flow is hard to capture accurately. Furthermore, the dynamic characteristics of traffic flow distribution could be an extending study, which has not yet been considered in this paper. In future research, more land-use attributes and dynamic traffic distribution could be taken into account to develop the distribution forecasting model.

## Data Availability

The data used to support the findings of this study were supplied by Guangzhou Metro under license and so cannot be made freely available. Access to these data should be considered by the corresponding author upon request, with permission of Guangzhou Metro.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This research was supported by the Fundamental Research Funds for the Central Universities (no. 2020YJS080) and the National Natural Science Foundation of China (no. 71931003).

## References

- [1] A. Salzberg, S. Mehndiratta, and Z. Liu, "Urban rail development in China," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2275, no. 1, pp. 49–57, 2012.
- [2] C. Cai, E. Yao, S. Liu, Y. Zhang, and J. Liu, "Holiday destination choice behavior analysis based on AFC data of urban rail transit," *Discrete Dynamics in Nature and Society*, vol. 2015, Article ID 136010, 2015.
- [3] S. Liu and E. Yao, "Holiday passenger flow forecasting based on the modified least-square support vector machine for the metro system," *Journal of Transportation Engineering, Part A: Systems*, vol. 143, no. 2, Article ID 04016005, 2016.
- [4] L. He, Q. Liang, and S. Fang, "Challenges and innovative solutions in urban rail transit network operations and management: China's Guangzhou metro experience," *Urban Rail Transit*, vol. 2, no. 1, pp. 33–45, 2016.
- [5] T. Grosche, F. Rothlauf, and A. Heinzl, "Gravity models for airline passenger volume estimation," *Journal of Air Transport Management*, vol. 13, no. 4, pp. 175–183, 2007.
- [6] Y. Wang, L. Yang, Y. Geng, and M. Zheng, "OD matrix estimation for urban expressway," *Journal of Transportation Systems Engineering and Information Technology*, vol. 10, no. 2, pp. 83–87, 2010.
- [7] M. Ren, Y. Lin, M. Jin, Z. Duan, Y. Gong, and Y. Liu, "Examining the effect of land-use function complementarity on intra-urban spatial interactions using metro smart card records," *Transportation*, vol. 47, no. 4, pp. 1607–1629, 2019.



- [8] T. Tsekeris and A. Stathopoulos, "Gravity models for dynamic transport planning: development and implementation in urban networks," *Journal of Transport Geography*, vol. 14, no. 2, pp. 152–160, 2006.
- [9] P. J. Jin, M. Cebelak, F. Yang, J. Zhang, C. M. Walton, and B. Ran, "Location-based social networking data," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2430, no. 1, pp. 72–82, 2014.
- [10] A. Tsirimpia, A. Polydoropoulou, and C. Antoniou, "Development of a mixed multi-nomial logit model to capture the impact of information systems on travelers' switching behavior," *Journal of Intelligent Transportation Systems*, vol. 11, no. 2, pp. 79–89, 2007.
- [11] Y. Yang, E. Yao, Z. Yang, and R. Zhang, "Modeling the charging and route choice behavior of BEV drivers," *Transportation Research Part C: Emerging Technologies*, vol. 65, pp. 190–204, 2016.
- [12] D. Q. Nguyen-Phuoc, G. Currie, C. De Gruyter, and W. Young, "Transit user reactions to major service withdrawal - a behavioural study," *Transport Policy*, vol. 64, pp. 29–37, 2018.
- [13] A. Faghih-Imani and N. Eluru, "Analysing bicycle-sharing system user destination choice preferences: chicago's Divvy system," *Journal of Transport Geography*, vol. 44, pp. 53–64, 2015.
- [14] J. Kelly, A. Patrick, D. Christopher, and J. Robert, "Development of destination choice models for pedestrian travel," *Transportation Research Part A: Policy and Practice*, vol. 94, pp. 255–265, 2016.
- [15] M. Mehadil Orvin and M. Rahman Fatmi, "Modeling destination choice behavior of the dockless bike sharing service users," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2674, no. 11, pp. 875–887, 2020.
- [16] H. J. P. Timmermans, "A stated choice model of sequential mode and destination choice behaviour for shopping trips," *Environment and Planning A: Economy and Space*, vol. 28, no. 1, pp. 173–184, 1996.
- [17] G. Jovicic and C. O. Hansen, "A passenger travel demand model for Copenhagen," *Transportation Research Part A: Policy and Practice*, vol. 37, no. 4, pp. 333–349, 2003.
- [18] S. Ashiabor, H. Baik, and A. Trani, "Logit models for forecasting nationwide intercity travel demand in the United States," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2007, no. 1, pp. 1–12, 2007.
- [19] A. Nuzzolo and A. Comi, "Urban freight demand forecasting: a mixed quantity/delivery/vehicle-based model," *Transportation Research Part E: Logistics and Transportation Review*, vol. 65, pp. 84–98, 2014.
- [20] T. Li, "A demand estimator based on a nested logit model," *Transportation Science*, vol. 51, no. 3, pp. 918–930, 2017.
- [21] P. Ye and D. Wen, "A study of destination selection model based on link flows," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 1, pp. 428–437, 2013.
- [22] M.-H. Wang, S. D. Schrock, N. V. Broek, and T. Mulinazzi, "Estimating dynamic origin-destination data and travel demand using cell phone network data," *International Journal of Intelligent Transportation Systems Research*, vol. 11, no. 2, pp. 76–86, 2013.
- [23] Y. Wang, H. Yin, H. Chen, T. Wo, J. Xu, and K. Zheng, "Origin-destination matrix prediction via graph convolution: a new perspective of passenger demand modeling," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1227–1235, Anchorage, AK, USA, July 2019.
- [24] K. E. Train, *Discrete Choice Methods with Simulation*, Cambridge University Press, Cambridge, UK, 2009.
- [25] T. J. Fratar, "Vehicular trip distributions by successive approximations," *Traffic Quarterly*, vol. 8, no. 1, pp. 53–65, 1954.
- [26] P. K. Sarkar, V. Maitri, and G. J. Joshi, *Transportation Planning: Principles, Practices and Policies*, PHI Learning Pvt. Ltd, New Delhi, India, 2017.
- [27] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognition Letters*, vol. 31, no. 8, pp. 651–666, 2010.
- [28] M.-K. Kim, S.-P. Kim, J. Heo, and H.-G. Sohn, "Ridership patterns at subway stations of Seoul capital area and characteristics of station influence area," *KSCE Journal of Civil Engineering*, vol. 21, no. 3, pp. 964–975, 2017.
- [29] X. Zhao, Y. Wu, G. Ren, K. Ji, and W. Qian, "Clustering analysis of ridership patterns at subway stations: a case in Nanjing, China," *Journal of Urban Planning and Development*, vol. 145, no. 2, Article ID 04019005, 2019.
- [30] E. Yao and T. Morikawa, "A study of an integrated intercity travel demand model," *Transportation Research Part A: Policy and Practice*, vol. 39, no. 4, pp. 367–381, 2005.
- [31] C. F. Manski and S. R. Lerman, "The estimation of choice probabilities from choice based samples," *Econometrica*, vol. 45, no. 8, pp. 1977–1988, 1977.
- [32] S. R. Cosslett, "Efficient estimation of discrete-choice models," *Structural Analysis of Discrete Data with Econometric Applications*, vol. 3, pp. 51–111, 1981.
- [33] M. E. Ben-Akiva and S. R. Lerman, *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press, Cambridge, MA, USA, 1985.
- [34] D. McFadden, *Quantitative Methods for Analyzing Travel Behaviour of Individuals: Some Recent Developments*, Cowles Foundation for Research in Economics, New Haven, CT, USA, 1977.
- [35] M. Nakanishi and L. G. Cooper, "Parameter estimation for a multiplicative competitive interaction model-least squares approach," *Journal of Marketing Research*, vol. 11, no. 3, pp. 303–311, 1974.