International Journal of Advanced Engineering and Technology
ISSN: 2456-7655; Impact Factor: RJIF 5.54
Received: 23-01-2019; Accepted: 26-02-2019
www.newengineeringjournal.com
Volume 3; Issue 2; April 2019; Page No. 50-54

Station-level passenger flow demand forecast with a combination of ridership and land use data

YI Wang[1], LV Yitong[2], WANG Zhuoqun[3], MO Yihong[4], LUO Qin[5*]
[1] Thales SEC transport CO., Ltd. Shanghai, China
[2] Shenzhen Urban Transport Planning Center, Shenzhen, China
[3] College of Urban Transportation and Logistics Shenzhen Technology University Shenzhen, China
[4] College of Urban Rail Transit Shenzhen University, Shenzhen, China
[5] Shenzhen WeiKang Transport Information Technology Co., Ltd. Shenzhen, China

Abstract
With the continuous development of urbanization, some big cities continue to expand and the travel demand continues to grow. Among them, rail transit plays an increasingly prominent role in urban passenger transport system. As the fundamental information during the process of planning, construction and operation of rail transit, passenger flow is the key to the improvement and development of rail transit. Based on the inbound and outbound data of passenger flow of Shenzhen metro network, this paper applies the clustering algorithm to classify the stations, and uses decision tree model analyze the significant factors to the passenger flow. Finally, a multiple regression model is developed to forecast the passenger flow of each category of stations. After an example verification, the analysis method proposed in this paper can better reflect the relationship between land use and passenger flow demand and the results can provide theoretical guidance and useful reference for the reasonable determination of rail transit station scale, as well as the development of surrounding land use and the adjustment of operation organization plan.

Keywords: urban rail transit, time-varying characteristics, cluster analysis, decision tree model, passenger flow demand forecast

1. Introduction
As urban transportation demand continues to grow, rail transit operations tend to be networked. A large number of new lines were planned and in operation, which brought unavoidable impact on the existing lines on the network. Many problems have emerged, such as spatially and temporally unbalanced passenger flow and over loaded issues during peak hours. In such cases, how to select a more reasonable location for station and plan the routes of lines becomes a key issue to be solved urgently.

In the field of passenger flow scale, existing research usually focuses on the internal and external aspects of rail transit. Considering the network structure inside the rail transit, the passenger flow is predicted from the perspectives of passenger flow, travel route, and passenger flow transfer [1-2]; Si B F adopted RBF neural network mechanism algorithm to fit the evolution process of passenger flow in orbit network [3]; Li X N considered the walking environment factor, site size factor and site connection factor to classify the site [4]; Zou W Q analyzed passenger flow characteristics and passenger flow by station Connection method, using the travel mode chain probability selection model to predict the passenger volume and transfer volume of the station [5]; Liu L J, through the combination of deep neural network and automatic stacking encoder, based on historical passenger flow data training model, realized passenger flow forecast for Xiamen BRT [6]; Starting from the externally influential factors. For example, considering the geographical location of the station, the site function assumed, the type of site property around the station, and other factors, the station classification, and further combined with the residents' travel behavior to predict the passenger flow [7-8]; Ye Z X extracted the land use factors along the route, and combines the four-step method to study the relationship between line passenger flow and land use [9]; Guang Z R combined the historical passenger flow data, selected the land use property, development intensity and other indicators as the fuzzy characteristic indicators, and established the inbound and outbound passenger flow prediction model [10].

The existing passenger flow forecasting method is mainly based on historical data, extracting internal and external indicators that can reflect the characteristics of passenger flow characteristics, and predicting future passenger flows. The data on land use near the station, as well as passenger flow characteristics and passenger flow, are not combined and matched.

Therefore, this paper intends to match the land attribute indicators near the site, combined with the K-means clustering algorithm and the decision tree model, to match the station passenger flow characteristics within each rail transit with the external surrounding land use conditions. According to the obtained quantitative mapping results, the sites are classified, and based on this, multivariate regression prediction models for different station types are constructed to realize the purpose of passenger flow demand prediction of rail transit stations.

2. Station classification based on temporal patterns of passenger flow
2.1 Passenger flow characteristic features extraction
Taking the data of all-day inbound and outbound passenger flow of each station as the basic research unit, the data of a whole working week in Shenzhen Metro in March 2016 was

used for analysis. After pre-processing, a total of 36 dimensions of time-sharing traffic from 6:00 to 24:00 at each station of each line were obtained. Due to the disparity in the magnitude of the traffic between stations, the data is standardized by z-score method to improve the comparability of data.

The evaluation index of the passenger flow of the station is expressed as the distribution characteristics and morphological characteristics of the passenger flow. These stations demonstrate five types of passenger flow distribution: unimodal, bimodal, full-peak, spurt and non-peak; Passenger flow morphological characteristics are expressed by parameter values such as peak hours, peak hour passenger flow, peak hour factor, and peak hour factor.

This paper defines the peak hour factor as the ratio of the peak natural hour passenger flow to the sum of the daily passenger flow (Equation 1). In which means peak hour factor, means passenger flow of the peak natural hour and means the passenger flow of the hour.

In addition, the paper defines the super peak hour factor as the ratio of the maximum passenger flow during the peak hour to the average flow rate of one hour and the peak hour (Equation 2). In which means super peak hour factor, means the passenger flow of the 15 minutes of maximum passenger flow and means the average flow of an hour.

$$p_t = \frac{V_{peak}}{\sum V_i} \quad (1)$$

$$p_{st} = \frac{V_{15\min} \times 4}{\bar{V}} \quad (2)$$

2.2 Clustering Result Analysis

The cluster tree demonstrates the number of stations and the stations included in each category. The horizontal axis is the ID of 118 stations, the vertical axis is the distance between classes, and the tree branch structure reflects the aggregation of different categories, which is also the process of classification (Figure 1).

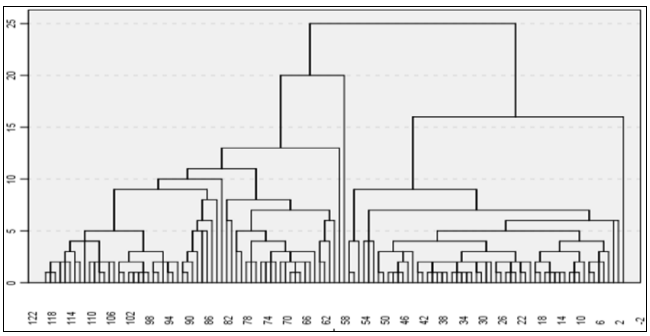


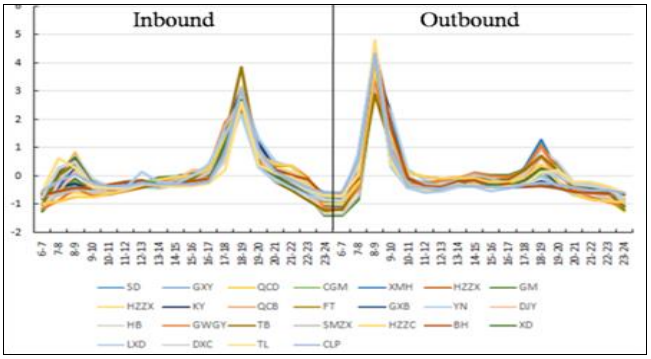
Fig 1: Cluster tree diagram.

The stations are classified into six categories, by analyzing the clustering results, and the distribution characteristics of each type of passenger flow are as follows.

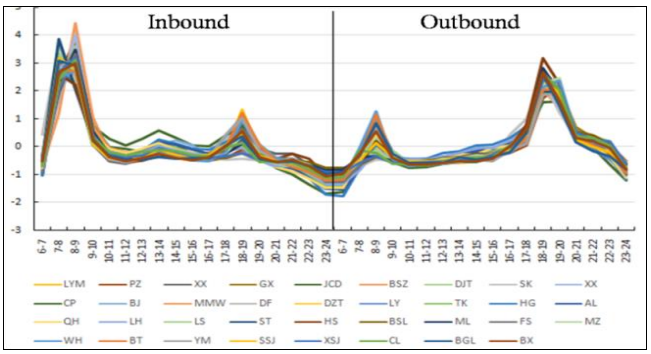
Unimodal. The peak usually occurs at the inbound passenger flow of the morning peak or the outbound passenger flow of the evening peak. Such stations are generally located in areas where land use functions are concentrated. The first type (Fig. 2a) and the fourth type (Fig. 2b) are unimodal, and the passenger flow characteristics are concentrated in the morning or evening peaks, it is speculated that the surrounding area is mainly residence and employment.

Bimodal. The peaks generally occurring in the morning peak hours and the evening peak hours. In both periods, the passenger flow inbound and outbound are significantly increased. The second type (Fig. 2c) and the fifth type (Fig. 2d) are bimodal stations. Especially, the inbound passenger flow of morning peak and the outbound passenger flow of evening peak of the second type station are significantly larger than other two peaks. Therefore, the second type of station is partially surrounded by residence-type land; similarly, the fifth types are partially surrounded by employment-type land.

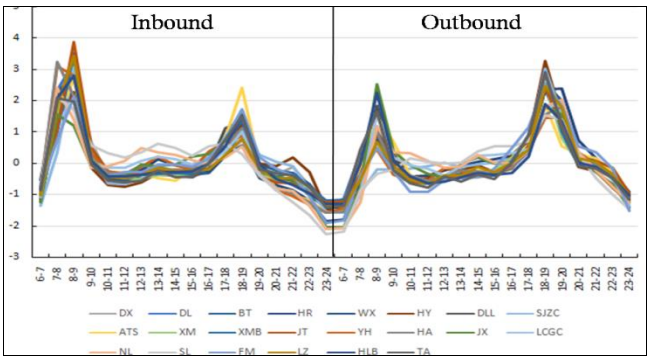
Multimodal. Multimodal stations are those stations whose passenger flow shows agglomeration characteristics during multiple time periods throughout the day. For example, the passenger flow of the third type of station (as shown in Figure 2e) has multiple large peaks throughout the day; it is because that the land around the station is comprehensive, which is composed by various land use features such as commercial, residential, and leisure squares. The sixth type of station (as shown in Figure 2f) has two large peaks throughout the day, as well as a number of small peaks. Therefore, land around the station is mainly commercial land use and the station is commercial-type station.



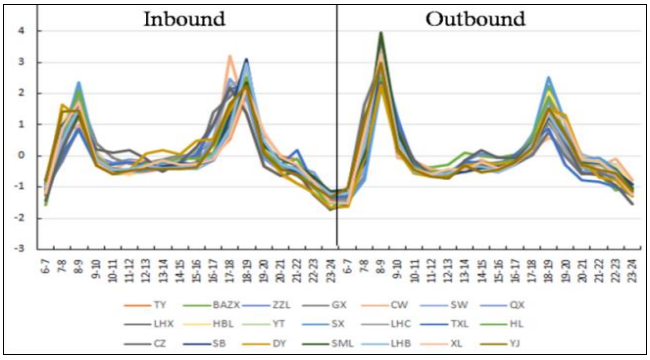
(a). Type 1: Employment unimodal station



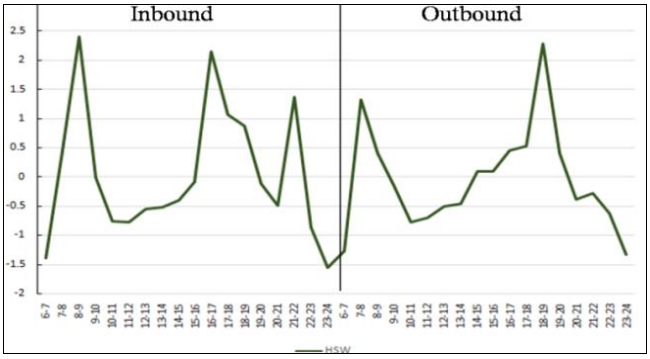
(b). Type 4: Residential unimodal station



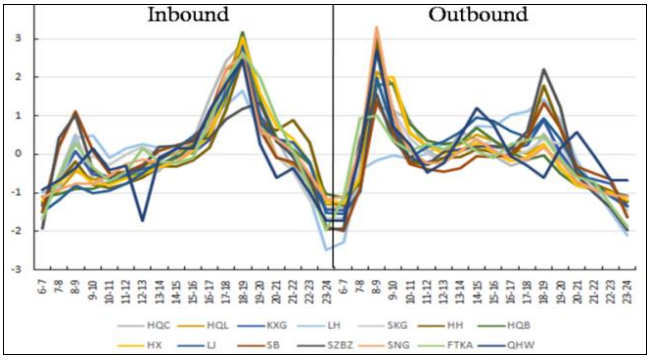
(c). Type 2: Biennial residential bimodal station



(d). Type 5: Bi-employment bimodal station



(e). Type 3: Integrated multimodal station



(f). Type 6: Commercial multimodal station

Fig 2: Clustering results of various types of stations.

For the first, fifth and sixth type of employment-oriented stations, the passenger flow at the early peak of the working day and the passenger flow at the evening peak are more prominent than those of other stations. For the second and fourth types of residential stations, the passenger flow at the early peak and the passenger flow at the evening peak are higher than those at other stations.

3. Prediction of station type and passenger flow based on its time-varying characteristics

Considering the characteristic attributes and quantity of the sample stations, it is concluded that the results of the third and sixth types of stations include fewer stations, and the statistical description is less meaningful. Therefore, the remaining four types of sites are considered, and they are classified into residential-based stations and employment-oriented stations according to the type of land use.

3.1 Extraction of main factors affecting passenger flow scale

Data related to social economy, land use, traffic environment and site characteristics inside an area are extracted in an area covered by a distance of 800 meters away from the station, as shown in Table 1.

FAR is an indicator of the extent of land use development (Equation 4).

$$FAR = \frac{gross\ floor\ area}{area\ of\ the\ plot} \tag{4}$$

ENT is a quantitative indicator of land use equilibrium. The larger the entropy index, the higher the land use diversity and the degree of land mixing (Equation 5).

$$ENT = -\frac{\sum_{j=1}^{k} P_j \ln(P_j)}{\ln(k)} \tag{5}$$

Where P_j is the percentage of a certain type of land area to the total study area, and k is the number of land type.

In addition, whether it is a transfer station and whether it is adjacent to a large hub are two dummy variables that characterize the site's own properties. If the station is a transfer station or a large hub docking station, it is represented by 1; otherwise, it is represented by 0.

Table 1: Influential factors of passenger flow scale

Socioeconomic indicators	Number of residents
	Number of positions
	Post population ratio
Land use indicators	Office land area
	Industrial land area
	Public supporting land area
	Transportation land area
	Residential land area
	Residential supporting land area
	commercial and residential land area
	Construction area of municipal facilities
	Private residential land area
	Comprehensive land area
	Commercial land area
	FAR (Floor area ratio)
	Mixed entropy
Traffic environment indicators	Number of bus lines connected per unit area
	Road network density
Station's own characteristic indicators	Whether it is a transfer station
	Whether it is adjacent to a large hub

3.2 Multiple regression prediction model

This paper tries to build a model for passenger flow scale prediction mainly focusing on the employment-oriented station and the residential-based stations. The AFC data of passenger flow of Shenzhen Metro during a working week in March 2016 was extracted, and the passenger flow of each station was calculated. The verification shows that there is a nonlinear relationship between the passenger flow of the station and different influencing factors. Therefore, convert nonlinear logical relationships to linear logical expressions by logarithmic operations.

The multivariate regression prediction model of the inbound

and outbound traffic of each station is as follows:

$$\ln(S_{i,in}) = \sum \alpha_{in} \ln(LAND_{ij}) + \beta_{in}(FAR_i) + \gamma_{in}(\mathrm{LUMIX}_i) + \delta_{in}(\mathrm{ROAD}_i) + \theta_{in}(\mathrm{BUSLINE}_i) + \mu_{in}(\mathrm{TRANS}_i) + \sigma_{in}(\mathrm{HUB}_i) + \varphi_{in}(\mathrm{BUS\_STOP}_i) + \omega_{in} \quad (6)$$

$$\ln(S_{i,\mathrm{out}}) = \sum \alpha_{out} \ln(LAND_{ij}) + \beta_{out}(FAR_i) + \gamma_{out}(\mathrm{LUMIX}_i) + \delta_{out}(\mathrm{ROAD}_i) + \theta_{out}(\mathrm{BUSLINE}_i) + \mu_{out}(\mathrm{TRANS}_i) + \sigma_{out}(\mathrm{HUB}_i) + \varphi_{out}(\mathrm{BUS\_STOP}_i) + \omega_{out} \quad (7)$$

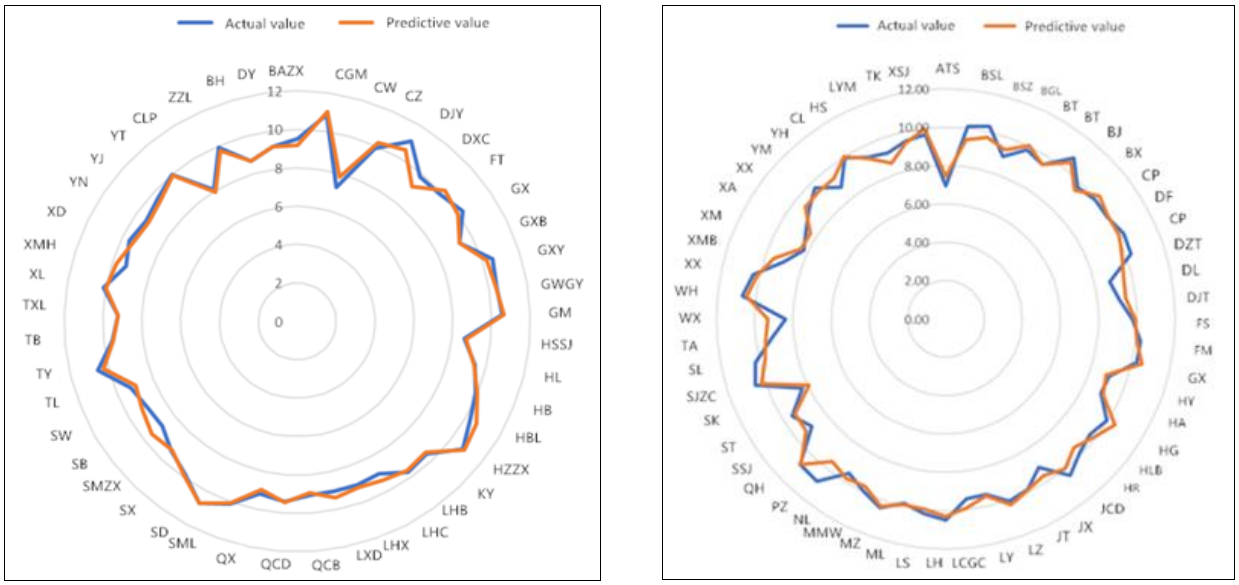
$LAND_{ij}$, FAR_i $LUMIX_i$, $ROAD_i$, $BUSLINE_i$, $TRANS_i$, HUB_i and BUS_STOP_i respectively represent the building area, FAR, ENT, road density, density of connected bus lines within the unit range, and whether it is a transfer station, whether to connect to a large passenger hub site, and whether to connect to a bus station, the above is the independent variable in the regression model, and the unit is consistent with the above. $\alpha, \beta, \gamma, \delta, \theta, \mu, \sigma, \varphi$ are the regression coefficients of each independent variable, and ε is a constant.

3.3 Results and analysis

In the regression model, the R^2 of the inbound and outbound models of the first and fifth types of stations are both 0.835, which means that the model interpretation ability is better. The R^2 of the inbound and outbound models of the first and fifth types of stations are 0.676 and 0.694, respectively, which shows that the model interpretation ability is acceptable. According to statistics, the scale of inbound and outbound passenger flow at each station is basically the same and the significant influence factors are also basically the same.

The construction area of the office and residential land around the first and fifth stations is the basis of the passenger flow scale. The transfer station has higher traffic accessibility than the non-transfer station; so it is easier to have larger passenger flow. In addition, due to various transportation modes, if there is a bus station around the site, the passenger flow scale will be weakened. The second and fourth types of stations are mainly surrounded on the private residential land and the commercial land area as the basis of the passenger flow scale. The land use is relatively balanced, and the surrounding large-scale passenger transportation hub is more conducive to attracting passengers.

The actual value of daily average inbound and outbound traffic of each type of station and the predicted value of the regression model are shown (Figure 3). The analysis shows that the difference between the true value and the predicted value is small, and the prediction result is good.



(a). Outbound passenger flow at employment stations

(b). Inbound passenger flow at residential stations

Fig 3: Comparison of real and predicted values.

4. Example analysis

Based on the above analysis, this paper predicts the basic characteristics of the passenger flow of the stations through many factors. Based on the historical data of passenger flow at each station of the No.1-No.5 lines of the Shenzhen Railway Phase II project, the passenger flow of the Chiwei Station (opened in October 2016) of the 7th Line is predicted. The indicators around the Chiwei station are extracted.

From the data of various indicators around the Chiwei station, combined with the results above, the time-varying characteristics of the passenger flow at the Chiwei station are predicted to be the fourth category. The inbound passenger flow of morning peak and the outbound passenger flow of evening peak are large and the passenger flow at hollow time is small, the overall performance is unimodal. According to Figure 4, the decision tree model and the extracted factors used in this paper are correct, and the type and time-varying characteristics of the station can be predicted well.

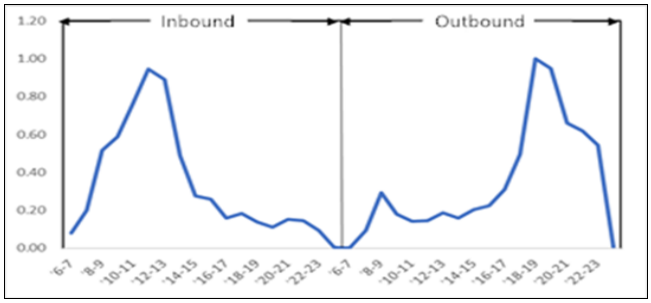


Fig 4: Time-varying characteristics of passenger flow of Chiwei Station.

According to the type of station at Chiwei Station, the regression coefficients of the second and fourth type of station prediction models are used for full-day passenger flow prediction (as shown in Table 1). Since the magnitude of the data of land construction area is different from that of other data, the land data needs to be logarithmically normalized, and finally the value of the independent variables in the model is obtained.

By solving the regression model, the average daily inbound and outbound passenger flow predicted is 9136 and 8778. According to statistics, the actual daily average inbound and outbound passenger flow is 9159 and 9483. The prediction error of the inbound and outbound passenger flow of the model is 2% and 5%, the prediction result is good.

5. Conclusion

This paper applied the K-means algorithm to identify the station categories on the basis of maximally retaining the original passenger flow information through the full-time AFC historical data of different stations in Shenzhen metro network. The passenger flow forecasting model based on employment and residential land was constructed. At last, the prediction results were verified by an example in Shenzhen.

After the example verification of Chiwei station, it can be concluded that the analysis method of cluster analysis and decision tree proposed in this paper can better reflect the relationship between land use and passenger flow demand.

In addition to the factors that have been considered, the future studies will focus on passenger flow forecasting model with the resident's subjective travel attitude, and construct more effective model prediction methods to achieve higher precision time-sharing passenger flow forecasting.

6. Acknowledgments

This paper is supported by the Research Projects of Natural Science Foundation of Guangdong Province under grant No. 2018A030313119, the Research Projects of the Social Science and Humanity on Young Fund of the Ministry of Education under grant No.15YJCZH108, and the Research Project of Shenzhen Technology University.

7. References

1. Wang GP. Passenger Flow Forecast of Intercity Railway Network in Central Yunnan Urban Agglomeration. Railway Standard Design, 2017; 61(07):5-10.
2. Liu Y. Cheng M J. Research on Estimation Method of Passenger Flow Forecast for Rail Transit. Journal of Railway Engineering Society. 2017; 32(12):92-96.
3. Si BF, He JR, Ren H, *et al*. Urban railway traffic passenger flow forecast based on the timing characteristics. Journal of Beijing Jiaotong University. 2014; 38(3):1-6.
4. Li XN. Research on clustering method for classification of urban rail transit stations [J]. Railway Standard Design, 2015.
5. Zou WQ. Analysis and forecast of passenger flow in urban rail transit stations. Chang'an University, 2013.
6. Liu LJ, Chen RC. A novel passenger flow prediction model using deep learning methods, Transportation Research Part C: Emerging Technologies, 2017; 84:74-91, ISSN 0968-090X.
7. Liu JX. Analysis of the impact of land use and development on passenger flow around urban rail transit stations. Chengdu: Southwest Jiaotong University, 2016; 6-8.
8. Li YW. Controlling the Design Scale of Subway Based on Passenger Flow Forecast Results. Urban Rapid Transit Traffic, 2015; 3-6.
9. Ye ZX. Research on the relationship between urban rail transit line passenger flow and land use along the route. East China Jiaotong University, 2014.
10. Guang ZR. Passenger flow forecast of urban rail transit in and out based on land use and accessibility. Beijing Jiaotong University, 2013; 55-58.