Traffic Volume Prediction Methods in a Multi-Source Data Environment Based on the "Four-Step Method"

Jia Cao [1], Wenna Wang [2], Yiyi Cheng [3], Qian Yang [4]

email: [1]21509751@qq.com; email: [2]1161796377@qq.com; email: [3]chengyiyi@hpdi.com.cn; email: [4]18810903774@163.com

CCCC Highway Consultants Co., Ltd., Beijing 100010, China

Abstract. With the continuous updating of data collection methods in the field of highway transportation, the source of monitoring data is gradually expanding from a single static and intermittent dataset to a multi-source dataset that combines static, dynamic, and continuous data. Therefore, it is necessary to combine high-value social data and conduct research on traffic volume prediction methods based on continuous data environment, realizing the transformation of traffic volume prediction from intermittent and local analysis methods to continuous, multi-source, and gradually iterative optimization ideas, thereby improving the accuracy and reliability of traffic volume prediction.

Keywords: Traffic volume prediction; Multi source data; Model research; four-step method

1 Introduction

Traffic volume is one of the important indicators for feasibility study in the process of road construction, and it is necessary to use traffic volume when determining road grades and traffic facilities[1]. The development of smart transportation has greatly advanced in China, leading to improved levels of informatization in expressway monitoring, toll collection, and travel services. This progress has resulted in the accumulation of a significant amount of historical data as well as the generation of new real-time dynamic data. These developments offer a novel approach for utilizing multi-source, continuous, and large-sample data for traffic volume prediction, ultimately enhancing the efficiency of project investment and evaluation. Upon analyzing the current data conditions, it is evident that there has been a transition from limited types and low value density data to a multi-source, continuous, and large-sample data environment, The data situation is detailed in Table1. However, it is clear that the application of technical methods under these new data conditions is insufficient. Therefore, there is a need to adapt to this new data environment, assess the existing traffic volume prediction methods, and address challenges such as the collection and integration of large-volume data, model selection in complex association relationships, and parameter calibration during the traffic volume prediction process. By doing so, we can further enhance the accuracy and reliability of traffic volume prediction, thus providing a basis for project construction standards and scale, to determine the important parameters of highway service level and service level, as well as the basis for project construction scale and standards[2].

PMIS 2024, March 15-17, Changsha, People's Republic of China

DOI 10.4108/eai.15-3-2024.2346530

Table 1 Sources and classification of current data.

Data Source	Category
Manual reporting	Issuing work orders, statistical reports, transportation contracts, transportation documents, customs declaration and commodity inspection, etc.
Semi-automation	Traffic volume observation, detection data, weighing data, shift reporting data, gantry data,etc.
Full automation	V2X data, GPS monitoring, access video, road condition monitoring, bridge monitoring, etc.
Transaction data	Land, population, mobile phone signals, settlement data, etc.

2 Analysis of traffic volume prediction methods

In the field of highway and transportation planning, the most widely used and classic method is still the "four-step method"[3]. This method first emerged in the field of urban traffic planning and was initially formulated during the 1950s for the planning of Chicago in the United States. It was subsequently enhanced during the comprehensive planning of the Tokyo metropolitan area in the 1960s and gradually matured. This method presents challenges in accurately predicting future traffic volume for specific road segments in a single step. Generally, it needs to start from the actual regional economy, establish a road network model, and complete the traffic volume prediction of specific road segments through the four steps[4], including traffic generation, distribution, mode split, and the assignment of traffic flow, based on relevant analysis with the economic and social aspects. This study is built on the assumption that, in the absence of significant changes in the correlation between traffic volume and environmental factors, traffic volume development inertia, analogies with similar projects, and the correlation with regional socio-economic development are considered in order to predict future traffic volume through the analysis of the current OD. In comparison with traditional prediction methods, the four-step method is characterized by a more comprehensive definition and accurate portrayal of the prediction, and it can comprehensively take into account the social and economic factors of demand, service levels, and competition within the traffic and transportation system, making it the most widely utilized method in transportation planning, operation, organization and management.

3 Idea of traffic volume prediction

Based on the traffic survey and data collection, in conjunction with the analysis of the traffic impact zone, the initial forecast for the total volume of passenger and freight within the impact zone is derived from the projected development of the economy and society. Subsequently, the projected traffic generation volume for passenger and freight in the forecast year is determined based on the correlation between the passenger and freight volume and traffic volume. This determination is coupled with the selection of the OD in the base year and distribution pattern to establish the distribution of traffic volume. Furthermore, taking into account the historical apportionment of traffic volume on expressways and national/provincial highways over the years, the traffic volume for expressways and provincial highways is ascertained. Ultimately, a

road impedance function is formulated to calculate the traffic assignment volume for road segments. The detailed prediction framework is depicted in the figure below.

3.1 Transportation volume prediction

Cluster recognition is conducted to identify key factors with the greatest impact on highway transportation and total traffic volume generation within the impact range, such as economic industry, land use, and population distribution[5]. To predict the future passenger and freight transportation volume in the region providing a validation path for relying on historical traffic volume to predict traffic generation volume, as compared to the prediction methods directly relying on economic indicators, this method provides a more comprehensive consideration of transportation factors and can thereby improve the accuracy of traffic generation volume prediction. The process of predicting transportation volume is shown in Figure1.

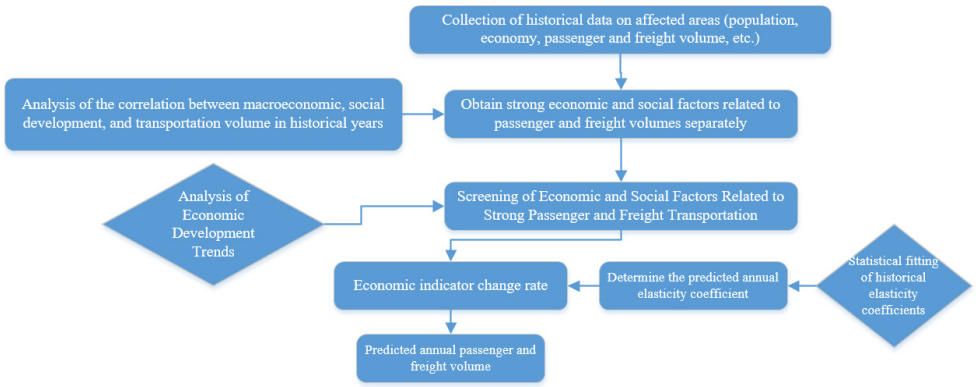


Figure 1 Prediction process of transportation volume within the region.

3.2 Traffic generation prediction

The traffic generation volume is a quantifiable indicator in the process of predicting travel volume through external factors such as economic industry, investment, trade, and population. The volume of traffic generation serves as a quantifiable indicator in predicting travel volume, taking into account external factors such as economic industry, land use, and population distribution. Traffic generation prediction currently encompasses two types: vehicle-based and passenger/goods-based. Urban traffic volume prediction generally relies on a per capita basis, given that the distribution of urban traffic flow is significantly influenced by individual behavioral factors. In comparison to highways, predicting urban traffic volume often necessitates the incorporation of economic and social predictions within the impact zone. This information is then transmitted to intermediate variables, such as transport intensity or travel frequency, in order to calculate the traffic generation volume and establish a more coherent logic. The prediction of total traffic volume generation hinges on establishing a set of functional relationships from key impact factors, such as economy and society, to the generation of total highway traffic volume. Therefore, it is essential to begin with the key impact factors and strategically utilize machine learning and other big data technology methods to determine the relationship between these factors and the total volume of regional highway traffic generation[6].

By establishing a correlation analysis model of "highway traffic volume generation-economic and social impact factors," we can optimize the total traffic volume generation model.

3.3 Traffic distribution prediction

The relevant parameters of the traffic distribution prediction model need to be calibrated using the OD in the base year. Therefore, the determination of the OD in the base year is the pre-process of traffic distribution prediction, and its accuracy is directly related to the prediction accuracy. Using the internal OD and transit OD within the project as the OD matrix in the base year, combined with the data characteristics of mobile phone signals, further verification and analysis of the OD in the base year can be achieved, The OD determination process is shown in Figure2. The data of mobile phone signals can determine the spatial location of users, and can relatively accurately record the spatiotemporal trajectory of population movement. It selects the OD of resident's driving trips based on the distance, time, and average speed from the starting point to the destination.

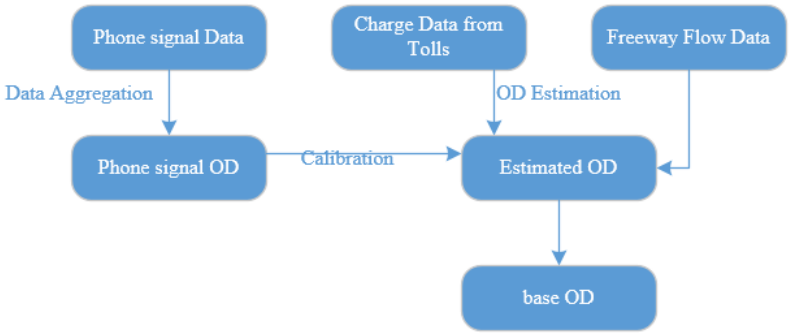


Figure 2 Process for Determining the OD in the Base Year.

The distribution of traffic volume is influenced by various factors, including the state of traffic sources (such as economic and demographic conditions), and the condition of transportation facilities (traffic impedance on traffic sections, like travel time, distance, and cost). The impact of costs on traffic distribution becomes increasingly apparent, especially with the adjustment of toll policies. Passenger and freight traffic is allocated in the system defined by traffic sources and transportation facilities in accordance with inherent laws, while also ensuring supply-demand equilibrium (occurrence and attraction). Economic activities and traffic convenience significantly influence traffic occurrence and attraction[7]. The gravity model method is an analysis method that simulates the law of universal gravitation. This method considers the travel impedance between two transportation communities, such as the distance and time of travel, while also taking into account the socio-economic factors of the area where the transportation community is located[8]. The fundamental formula is:

$$t_{ij} = k \frac{G_i^{\alpha} \cdot A_j^{\beta}}{R_{ij}^{\gamma}} \tag{1}$$

t_{ij}: Traffic volume prediction in the future distribution between Zone i and Zone j

G_i: The predicted value of occurrence traffic volume for Zone i in the coming year

A_j: The predicted value of attraction traffic volume for Zone j in the coming year

R_{ij}: The impedance value (cost, time, distance) between Zone i and Zone j

α, β, k: The coefficients of models

By the obtained OD in the base year, the parameters related to the gravity model can be calibrated and determined using the least square method.

3.4 Mode split

The traffic mode split in highways mainly aims to select the travel mode for expressways and ordinary national and provincial highways within the impact zone. In general, the primary factor defining the choice between expressways and ordinary national and provincial highways is the general cost:

This research uses the general minimum travel cost model as the primary applied model for mode split, and the model formula is as follows:

$$P_{ijm} = \frac{e^{-r_{ijm}}}{\sum_k e^{-r_{ijm}}} \quad (2)$$

P_{ijm}: The share ratio of traffic mode from Zone i to Zone j

r_{ijm}: The traffic impedance of traffic mode m from Zone i to Zone j

$\sum_k e^{-r_{ijm}}$: The traffic impedance of traffic mode k from Zone i to Zone j

It reflects the various factors and their importance that users consider when choosing traffic modes. It is generally defined as:

$$r_{ijm} = \sum_n \alpha_n y_{ijmn} + \alpha_0 \quad (3)$$

α_n: Weight coefficient

y_{ijmn}: The value of the nth impedance factor for the mth traffic mode from Zone i to Zone j

α_0: Non-quantifiable factors

- The tolls of expressway are expensive, but the speed is fast, but the accessibility of its end is relatively weak

- Ordinary national and provincial highways are free, but the speed of mixed traffic flow is slow, and the accessibility is great

It is concluded that the most important impedance factors that affect the choice of traffic modes are cost and time, thus can be calibrated. Please refer to the detailed process as shown in Figure 3.

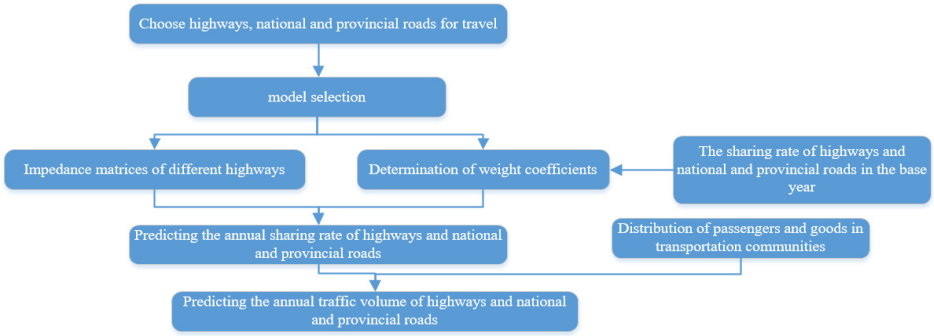


Figure 3 Flowchart for Method Selection.

3.5 Traffic assignment prediction

The user equilibrium assignment method is employed for allocating traffic volume in this context. The fundamental principle is as follows: when there are multiple roads between two points within the zone, the shortest path is selected based on the regional traffic volume. As the traffic volume grows, the volume on the shortest path also increases. Once the traffic volume attains a certain level, congestion leads to an increase in travel time on the path, rendering it no longer the shortest. Consequently, a new shortest path is determined based on the traffic volume. Correspondingly, as the traffic volume between two points continues to rise, an equilibrium state is reached, where the time required for each road between the two points becomes equivalent. When selecting the impedance function, it is essential to take into account factors such as traffic volume, capacity, and travel time. There is often extensive research on the adjustment of BPR function[9] at this step, while there is relatively less research on the modification of traffic capacity. The research takes into full consideration various factors influencing road capacity, and re-calibrates and analyzes road capacity, considering the level of road development and characteristics specific to our country, The parameter calibration process is shown in Figure4. Key factors include lane width, lateral clearance, longitudinal gradient, inadequate time spacing, road conditions, and degree of vehicle merging. The main formula for capacity is:

$$C = C_0 \times f_{Cw} \times f_{HV} \times f_{DIR} \times f_{FRIC} \tag{4}$$

C: Capacity under actual conditions

C_0: Basic capacity

f_{Cw}: Adjustment factor of lane width to traffic capacity

f_{HV}: Adjustment factor of traffic composition to traffic capacity

f_{DIR}: Adjustment factor of directional distribution to traffic capacity

f_{FRIC}: Adjustment factor of transverse interference to traffic capacity

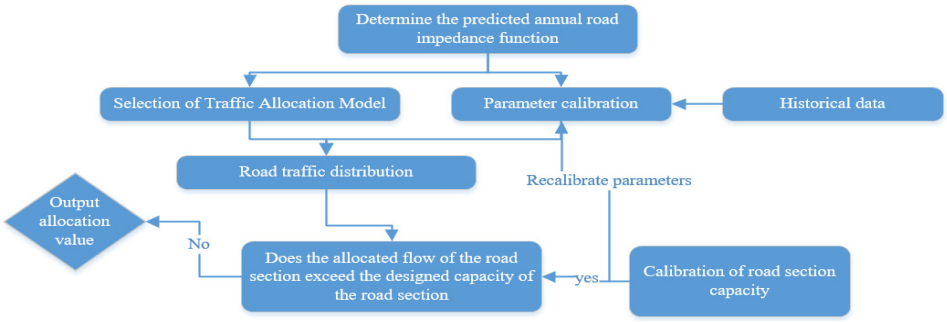


Figure 4 Parameter Calibration Process for Traffic Assignment.

4 Application of prediction model

Based on the actual road network conditions, this study assumes a certain region as the research object. In this network model, the research region is divided into 7 traffic zones and 12 connecting highways. There is travel demand between each traffic zone, with a total of 49 OD demands. All roads in the network are put into use for traffic. The basic field information of the road network is shown in Table 2, and the split of traffic zones, road network structure, current road segment traffic volume and saturation is shown in Figure 5.

Table 2 Basic Field Information of Road Network.

Field Name	Unit	Field Description
Length	km	Length of road segment
Level	Dimensionless	Expressway, Level 1, Level 2, Level 3, Level 4
Designed speed	Km/h	Speed reference value under the set technical level
Designed traffic capacity	Pcu	The traffic capacity of road segments under the set technical level and service level
Saturation	Dimensionless	The ratio of actual traffic volume to the designed traffic capacity on road segments

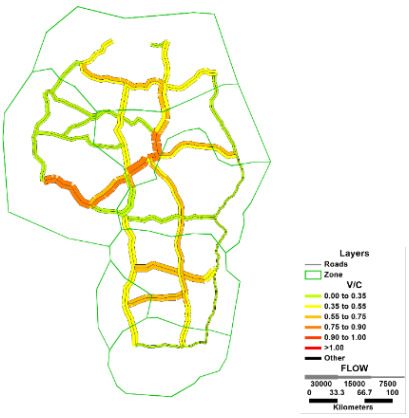


Figure 5 Road network and zone split of the research object.

Based on the prediction idea in Section 3, first predict the transportation volume and traffic occurrence. Based on considerations of data availability, economic, transportation, and other influencing factors[10], this study selects 18 initial indicators in four aspects such as national economic indicators, population and people's livelihood, retail investment and trade, etc. The relevant indicators are shown in Figure 6. Through correlation analysis, select the economic indicators affecting freight traffic generation volume with the secondary industry output and the comprehensive value of the primary industry output; the economic indicators affecting passenger traffic generation volume with the comprehensive value of regional gross domestic product and the permanent population. Due to data sensitivity, only the growth rate of the predicted transportation volume and the results of traffic distribution are shown here. The value for the growth rate of each traffic zone are shown in Table 3.

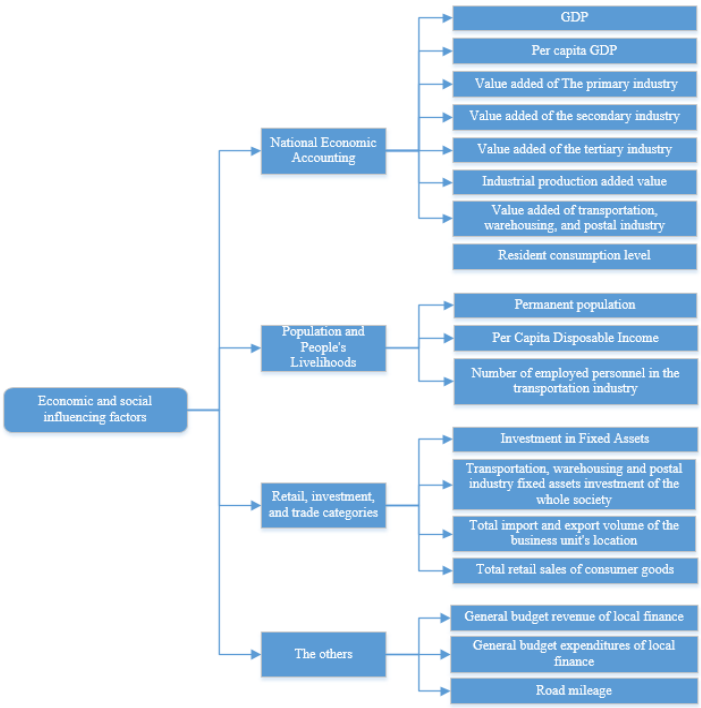


Figure 6 Initial selection of economic indicators.

Table 3 The Growth Rate of Passenger and Freight Traffic in Traffic Zones.

Zone No.	1	2	3	4	5	6	7
Growth Rate of Passenger Traffic	4.20%	4.70%	5.20%	5.00%	5.50%	5.00%	3.70%
Growth Rate of Freight Traffic	3.20%	3.70%	4.20%	5.40%	4.50%	6.00%	6.30%

Traffic distribution prediction is then carried out. During the verification process of the OD in the base year, the main fields for the data of the mobile phone signals selected in this study are shown in Table 4. Due to data sensitivity, only the OD in the base year obtained inversely, OD generated by data of mobile phone signals, and the OD in the base year verified using mobile

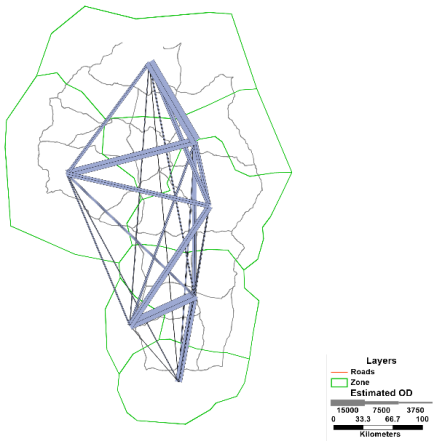
phone signals are shown here, respectively as shown in Figure 7 (1), (2), and (3). Using the OD in the base year for parameter calibration of the gravity model, with the main parameter values shown in Table 5. Using the calibrated gravity model for traffic distribution prediction, the OD in the future year is obtained as shown in Figure 7 (4).

Table 4 Basic Field Information of Mobile Phone Signals.

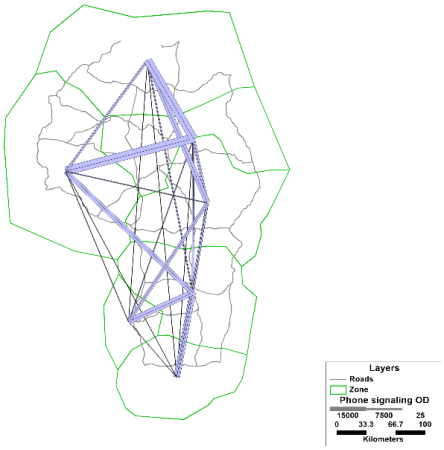
Field Name	Field Description
IMSI (International Mobile Subscriber Identity)	Identify and distinguish subscribers
LAC (Location Area Code)	Area covered by the base station
CI (Cell Identity)	Base station cell ID (can determine the accurate subscriber location through LAC and CI)
Timestamp	Record the time when subscriber trajectory occurs, accurate to seconds
Signal Time Type	Mobile service types, such as switching on/off, updating, receiving information, etc.
Basic Information Attribute	Basic information such as subscriber location

Table 5 Parameter Value of the Gravity Model.

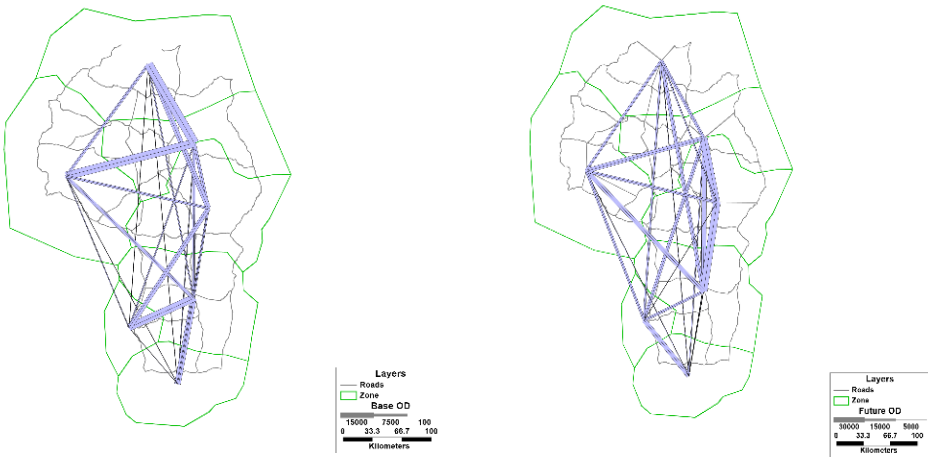
Parameter Name	α	β	γ	k
Parameter Value	1.220513	1.141748	0.543255	3.6E-08



(1) OD in the base year obtained inversely



(2) OD of the mobile phone signals



(3) Verified OD in the base year

(4) Predicted OD

Figure 7 Traffic Distribution Prediction.

Finally, conduct mode split and traffic assignment. As this step already has mature calculation software, and this study does not include improvements or optimizations for these processing steps, only the final results are presented, with the final traffic volume of the road segment and saturation shown in Figure 8.

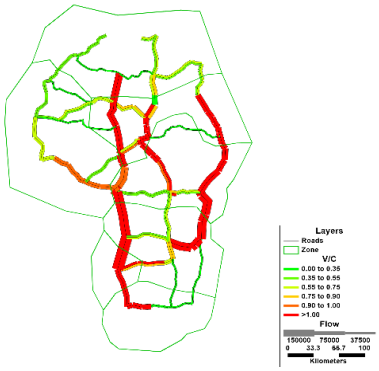


Figure 8 Traffic Allocation Results in the Future Year.

5 Conclusion

This research is grounded in the traditional "four-step method" and takes into full consideration the limitations of predicting each step. It combines the new data conditions at the current step in an attempt to study and analyze the prediction models and parameters for each step. This provides new ideas and methods for subsequent research and application of the "four-step method".

References

[1] Xu Hongling, Zhao Yanming. Research on Traffic Volume Prediction in Feasibility Study of Highway. Heilongjiang Transportation Technology, 2020(1):224-225.
[2] Sha Aimin, Wang Xiaodong. Research on the Application of Four Stage Method in Highway Traffic Volume Prediction. Northern Transportation, 2022(04):70-79.DOI:10.15996/j.cnki.bfjt.2022.04.020.
[3] Sha Aimin, Lv Fanren, Wang Xiaodong. Research on the Forecast of the Highway Traffic Volume Based on Four Stages Model. Beifang Jiaotong, 2016(12):94-98.DOI:10.15996/j.cnki.bfjt.2016.12.024.
[4] Wang Wei, Chen Xuewu. Transportation Planning. Beijing: China Communications Press, 2012:7.
[5] Cook County Highway Dept. Expressway influence on parallel routes: a study on Edens Expressway traffic diversion and generation trends.Chicago,1995.
[6] Li Xinghui, Zeng Bi, Wei Pengfei. Real time traffic flow prediction based on flow computing and big data platforms Compater Engineering and Designing, 2024(2):553-561.
[7] Holguín-Veras J, Kalahasthi L, Ramirez-Rios D G. Service trip attraction in commercial establishments[J]. Transportation Research Part E: Logistics and Transportation Review, 2021, 149: 102301.
[8] Cao Yindi Prediction of Traffic Volume for Four Stage Development and Expansion of Railway Lines. Science and Technology Innovation and Productivity. Science and Technology Innovation and Productivity, 2022(09),128-131+134.
[9] Bureau of Public Roads. Traffic Assignment Manual. Washington DC: Urban Planning Division,US Department of Commerce, 1964.
[10] Cham, 2020: 595-610. Networks[C]//International Online Conference on Intelligent Decision Science. Springer, to Improve the Technical and Economic Impacts of Urban Interchanges on Traffic Hosseini S H, Mehrabian A, Ebrahimi Z D, et al. A New Approach for Macroscopic Analysis.