Research on Supply and Demand Matching of Urban Public Transport Based on IC Card and GPS Data

Qiuxia SUN1,a, Hao HE1,b, Lu SUN2,c*

1College of mathematics and Systems Sciences Shandong University of Science and Technology Jining, China
2College of Economics and Management Shandong University of Science and Technology Jining, China

e-mail addresses: qiuxiasun@sdust.edu.cn, b-mail addresses: 657167037@qq.com,
c*e-mail addresses: sunlu@sdust.edu.cn,

Abstract—A research method to establish the matching degree of supply and demand of bus stations by using Moran’s I, which describes the spatial agglomerating feature, was proposed based on IC card and GPS data mining. The bus boarding station information was deduced by the time matching method and the passenger demand of the station was represented by the passenger flow in the station, the vehicle data from GPS data were selected as the bus service supply, the matching situation of bus supply and demand in different time periods in the whole city was researched, the degree of station imbalance was identified, to assist the bus system to optimize supply. Take Qingdao city as an example, the empirical research indicated that: It was concluded that the demand and supply of the bus service in the research area matched well on the whole, and the supply of buses at the bus stations basically met the travel demand of residents. According to the LISA diagram of station clustering distribution, the stations with a mismatch of supply and demand in the area were identified through spatio-temporal difference analysis, and the mismatch was relatively more obvious in regions where a small number of stations were located.

1. Introduction
Urban public transportation is an efficient and intensive transportation mode. The establishment of a high-quality public transportation system can effectively alleviate urban congestion, reduce traffic pollution, improve the urban living environment, and enhance the quality of life of residents. To improve the service quality of urban bus operations, it is necessary to have a full understanding of the current situation of the supply and demand matching of bus stops. Bus stops are the key point. When researching the layout, it is generally explored from multiple perspectives such as improving the accessibility of the bus stops or the coverage of the stations, and the coupling relationship between supply and demand of bus services, so as to realize the evaluation or optimization of the layout of the bus stops. [1-5] Based on the IC card and GPS data and the coupling model in spatial geography to evaluate the matching degree between the layout of the bus station and the station demand. Zeng et al. [6] Discuss the design method of public transportation network based on the balance of supply and demand. Yao et al. [7] Related literature has discussed the demand and supply of public transportation services, and the models and methods are relatively mature. However, there are diversity and complexity in public transportation travel behavior and purpose. Existing research considers the matching degree of public transportation
supply and demand for residents in different time and space. Public transportation in China has received less attention and deserves more attention.

The GPS data and IC card data of public transportation vehicles have the attributes of truthfulness, accuracy, timeliness, and high exploitability, [8] which provide important data support for the management of public transportation operation and dispatch. Therefore, based on the bus IC card and GPS data, the research on the supply and demand matching degree of urban bus stations can provide services for bus planning and dispatching more realistically. Based on the credit card demand for public transportation travel and the supply of public transportation vehicles, this paper establishes appropriate evaluation indicators to study the extent to which urban public transport stations meet passenger travel demands, obtain the matching degree of supply and demand of public transport stations, and identify stations with unbalanced supply and demand. With a view to improving the quality of public transport services, it provides a focus for improving urban public transport.

2. Data and methods

The data used in this article consists of two parts: The first part is a set of data points representing the supply of bus services. The data set includes the location information of each bus stop and the number of bus vehicles that stop. This data set is regarded as the supply service of the bus station; The second part is a set of data points representing the needs of public transportation services. The data set includes passenger flow data of various stations with IC cards swiping and boarding. The attribute of this data set is the demand for public transportation services. In order to obtain the data of the passenger flow of the station, it is first necessary to obtain the information of the boarding station in the bus swiping data.

2.1. Derivation of boarding site

The GPS equipment of public transportation vehicles can collect information and data about the operation of public transportation vehicles in real time: Line name, station name, vehicle arrival time, departure time, speed, location information, etc. After that, the boarding station of each IC card data record is derived by matching the IC card data record of the corresponding line according to time. [8] The correlation between the bus IC card and GPS data is shown in Fig 1.

![Fig.1 Relation between data](image)

The derivation steps of the boarding station are as follows:

Step1: 1 The corresponding route and recorded time information of the vehicle-mounted POS machine number in IC card data were extracted and at the same time, the bus license plate number data of the corresponding route in the corresponding GPS data were also extracted to match the POS machine number and vehicle license plate number;

Step2: The information of the swipe card of a passenger $i$ to be matched was read;

Step3: For IC card swiping time, it was matched with the arrival time and departure time of the corresponding bus GPS data according to the information of the route number and the vehicles. The boarding station was identified according to the identification rules that the card swiping time of the passenger $i$ with a card must be between the arrival time and the departure time of the vehicle arrival station;

Step4: All IC card numbers were read until the boarding station recorded by the bus IC card was deduced.
2.2. Study area
Qingdao is an important national coastal center city, an international port city, a major node city in the Belt and Road New Asia-Europe Continental Bridge Economic Corridor, and a strategic fulcrum for maritime cooperation. Qingdao is located on the southeast coast of the Shandong Peninsula, with a total area of 11,282 square kilometers, and a total permanent population of nearly 9,499,800 at the end of 2019. Following the SCO Qingdao Summit in 2018, Qingdao, which has attracted worldwide attention, has been put under the spotlight of the media time and time again, attracting a large influx of talents, the urban population is growing, and the demand for public transportation is also increasing. Therefore, it is of great significance to analyze the matching degree of supply and demand of Qingdao bus stations. The research area in this paper is selected as the six main jurisdictions of Qingdao: Shimin District, Shibe District, Licang District, Laoshan District, Chengyang District, and West Coast New District. Based on the bus station data of 6 jurisdictions, the bus stations with the same name on both sides of the road and the same side of the road are merged into one station. The total number of bus stations in the data set is 2,430.

2.3. Data preprocessing results
This article selects the IC card and GPS data of public transportation in Qingdao for a complete working week from September 16 to September 22, 2019 as the basic data. A total of 9,962,693 pieces of data were successfully matched using the time matching method, and the matching rate of the passenger boarding station was judged to be 96.02%, of which a total of 412,950 pieces of incomplete data that appeared during the collection process were eliminated. Part of the calculation results derived from the passenger boarding station is shown in Fig 2.

| CARDID | POSOPRID   | TIME          | STATIONNAME          | TYPE  | ROUTENAME | LONGITUDE | LATITUDE |
|--------|------------|---------------|----------------------|-------|-----------|-----------|----------|
| 2660023010013713 | 370020026026 | 2019-09-16 16:51:00 | Xueajiaozhi | down  | Road 26   | 120.156421 | 36.023733 |
| 2660000004400261 | 370020008299 | 2019-09-16 16:51:30 | Dongyantian community | uplink | East Road 17 | 120.162724 | 36.03337 |
| 2660022000030091 | 370020090189 | 2019-09-16 16:51:30 | Jimiya | down  | Road 27   | 120.166162 | 35.92503 |
| 266000000100278876 | 370020008636 | 2019-09-16 17:01:30 | Hongshan | down  | Road 26   | 120.102777 | 36.091365 |
| 2660032000038673 | 370020017282 | 2019-09-17 17:01:50 | No.5 Middle School | down  | Road 20   | 120.006436 | 35.873705 |
| 2660001342004226 | 370020042342 | 2019-09-17 17:02:00 | Fagou countries | down  | Road 46   | 120.146778 | 35.954064 |
| 26600000004863058 | 370020016705 | 2019-09-17 17:03:00 | Qinghi College | uplink | Road 30   | 120.169967 | 35.975501 |

Fig.2 Partial calculation results of passenger boarding station judgment

After obtaining the information of the passenger boarding station from 1.1, the overall trend of the bus passenger flow is calculated by day, as shown in Fig 3. It can be clearly seen from the figure that the number of card swiping on workdays is significantly higher than that on weekends. This is due to the difference in travel time and behavioral purpose of passengers on weekdays and weekends. Therefore, the average passenger flow of each period of weekdays and weekends is calculated separately, and the time distribution of passenger flow of each period shown in Fig 4 is obtained. Through the comparison of the passenger flow in each time period, it is obtained that the bus operation will be stopped from 1 am to 5 am. The morning peak on working days is 6-8 o'clock, and the evening peak is 16-19 o'clock; the morning peak on weekends is 7-9 o'clock, and the evening peak is 17-19 o'clock. According to the operating hours of Qingdao’s bus lines and the daily travel characteristics of residents, the working day is divided into the following 5 time periods: 00:00-06:00, 06:00-09:00, 09:00-16:00, 16:00-19:00, 19:00-23:59, divide the weekends into the following 5 time periods: 00:00-07:00, 07:00-10:00, 10:00-17:00, 17:00-20:00, 20:00-23:59.
3. Basic model
Spatial data generally includes spatial location information and attribute information. When analyzing its attribute information through location information, there is often autocorrelation, which is called spatial autocorrelation. Spatial autocorrelation refers to the correlation of the same variable at different spatial locations. The purpose is to analyze its attribute information based on location information, so as to test whether the attributes of a spatial unit and its neighboring spatial units are similar. Spatial autocorrelation is divided into global autocorrelation and local autocorrelation. Global autocorrelation is used to reflect the spatial distribution pattern of a phenomenon in the entire study area or the average concentration of similar attributes; The local spatial autocorrelation is used to reflect the correlation between a certain attribute value on a local unit and the same attribute value on adjacent local units in the entire area.

This article uses Bivariate Moran's I to measure. Moran's I is divided into Bivariate Global Moran's I and Bivariate LISA. Bivariate Global Moran's I judges the totality and judges whether there is agglomeration in the space. Bivariate LISA identifies the area where the agglomeration occurs. In this study, Bivariate Global Moran's I is used to explore the global coupling relationship between the density of public transportation stations and the distribution density of public transportation vehicles. Bivariate LISA is used to analyze the matching degree of public transportation service demand and supply at the station scale, and evaluate it based on this. The basic expression is as follow.
Global Moran’s $I = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(x_i - \bar{y})(x_j - \bar{y})}{S^2 S_0}$

(1)

local Moran’s $I_i = (x_i - \bar{y}) \sum_{j=1, j\neq i}^{n} w_{ij}(x_j - \bar{y})$

(2)

Where $I$ is the Bivariate Moran’s $I$, $I_j$ is the local Moran’s $I$ of bus stop $k$, $n$ is the number of bus stops, $x_i$ is the observation value of unit $i$, $\bar{y}$ is the mean value of the observed value of the tested variable, $(x_i - \bar{y})$ and $(x_j - \bar{y})$ represents the amount of bus card swiping and the standardized value at bus stops $k$ and $I$, $S^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{y})^2$ a is the sample variance with $n - 1$ degrees of freedom, $S_0 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}$ is the sum of the spatial weights of all variables. The spatial weight $w_{ij}$ refers to the spatial weight between positions $i$ and $j$. This paper uses the Queen adjacency weight rule to construct the spatial weight matrix.

4. Experimental results and analysis

4.1. Experimental results

The calculation of Bivariate Global Moran’s $I$ could reflect the spatial coupling of bus service demand and supply in the global space. From the perspective of the global spatial influence of each period in the research region (Table 1), the swiping card demand and the vehicle supply at the bus stations in the research area presented a relatively significant spatial dependence, and this tendency would become stronger with the increase of passenger flow. The Moran’s $I$ of five periods on business days and weekends has passed the significance test at the level of 5%, and the Moran’s $I$ all were in the range of $(0,1)$, which indicated that there was a significant positive spatial correlation between the supply and demand of bus stations in the research area. It could actually be expressed as: within the urban area of Qingdao city, the tendency of bus service with certain intensity was distributed in the bus stations with corresponding intensity. It indicated that from the perspective of supply and demand of bus service, the coupling of bus station setting and demand was excellent in the global space.

| Time | Moran’s $I$ | Z value | Moran’s $I$ | Z value |
|------|-------------|---------|-------------|---------|
| 1    | 0.209       | 1.803   | 0.223       | 1.945   |
| 2    | 0.253       | 2.059   | 0.240       | 1.999   |
| 3    | 0.247       | 2.019   | 0.249       | 2.002   |
| 4    | 0.254       | 2.061   | 0.219       | 1.901   |
| 5    | 0.202       | 1.799   | 0.199       | 1.747   |

The demand and supply of bus service represented different spatial agglomerating features in different periods. According to the measurement of Bivariate LISA, the Bivariate LISA was displayed in the form of the LISA graph (Fig 5). The local clustering trend of bus stations with statistical significance was given on the basis of random replacement of 1000, with a significant level of 5%. Each point in the Figure corresponded to one bus station in the research area, and the bus stations were divided
into 5 categories on the basis of the spatial agglomering feature of bus service demand and supply. High-High represents the high demand and high supply category for this site; Low-Low represents the low demand and low supply category for this site; Low-High represents the low demand and high supply category for this site; High-Low stands for high demand and low supply for this site; Not Significant sites.
4.2. Result analysis

According to the results of Fig. 5, the analysis of the temporal and spatial characteristics obtains: There are a large number of High-High sites in the morning and evening peak hours on weekdays and weekends, mainly in the North District of Shinan City. Low-Low sites are mainly distributed in sub-hot spots and urban fringe areas, and the number in each period is relatively stable. The number of Low-High sites is more weekends than workdays, especially during the sub-peak hours, which are mainly concentrated in the North District of Shinan City. There are more high-low types of stations in weekdays and weekends than in other periods, mainly concentrated in the West Coast New Area and Licang District, and there are more High-Low stations on weekends and evenings, mainly concentrated in Licang District.
(1) High-high cluster bus stations are mainly concentrated in areas with strong economic vitality and greater demand for employment and housing. Qingdao Station in Shina District, around Fushansuo Subway, Taitung Commercial Pedestrian Street, one of the five major business districts on Weihai Road in Shibe District, and Qingdao North Station. Among them, Haier Information Park (Metro Mansion Station) is around the morning and evening peaks on weekdays. The number of this type of stations is prominent in the morning and evening peak of the surrounding workdays, and the commuting phenomenon is obvious. Jinggangshan Road in West Coast New District, the area surrounded by Yangtze River Middle Road and along the Jiangshan South Road, as well as the Liuting Airport in Chengyang District. There are many high-high types of stations along the Heilongjiang Middle Road in Licang District and near Licun Subway Station throughout the day. This is the most prosperous commercial area in Licang District. The morning peak hours on weekends are mainly concentrated near Licang District, the number of areas around the railway station is reduced, and the stations during the sub-peak hours are mainly concentrated near the Taitung business district.

(2) The low-low cluster type bus station has low demand for public transportation services, but the supply of public transportation vehicles is not oversaturated, the supply and daily demand are well matched, and the supply of station vehicles is reasonable. This type of bus stops are mainly concentrated in the west and north of the West Coast New Area, and the fringe areas of the main urban areas of Chengyang District and Laoshan District. They are not the traditional core areas of the city. However, along the Yan'erdao Road Station with strong economic vitality, there are sporadic low-low type stations around Belle, Hisense, and Mackay shopping malls, which are related to the impact of tourist attractions, private vehicles and subway stations in the region.

(3) The service supply of low-high cluster bus stations is significantly higher than the average level, mainly concentrated in the Affiliated Hospital of Qing University and Binhai College in the West Coast New District, as well as many colleges and universities in Licang District. The number of stations of this type in the vicinity of the Wanxianghui, Liqun, and Jiajiayuan business districts in Chengyang District is relatively large at time 3 hours. This is related to the living habits of residents.

(4) The supply of public transportation services at stations with high-low clustering types is insufficient, and demand is saturated. There are mainly Beichuan Workers' Apartment Station in West Coast New District, University of Science and Technology Station and Baitongxinyuan Station in Licang District. The vicinity of Licun Station is the main gathering place for entertainment venues, leading to an increase in the number of this type of stations during the weekend evening peak.

Most of Qingdao's bus stations are of the type with a good match between supply and demand, and the high-high bus stations are mainly located in the traditional core area of the city, especially in the morning and evening peaks. The low-low bus stations are mainly located in the urban areas with weak urban vitality in the peripheral sub-hot spots, there is a big gap between the infrastructure construction and economic vitality of this area and the core area. The classification results are basically consistent with common sense and actual conditions, and the supply of vehicles at bus stations follows the basic demand distribution at each time period.

5. Conclusion

Business days and the weekends were divided into 5 time periods in the Paper on the basis of the bus IC and GPS data of a complete business week in Qingdao city and according to different travel times and purposes of residents. After the Bivariate Moran’s I was used to carry out the measurement, the classification result of each station was shown on the LISA clustering map. The analysis of the supply and demand of bus stations showed that the overall matching degree of supply and demand of bus services is excellent in the main urban area of Qingdao; insignificant and high-high, low-low types of stations accounted for the vast majority, and the supply and demand matching of bus stations is relatively reasonable. The method of evaluating bus station supply and demand and selecting optimization target was formed. For low-high type and high-low type stations, the frequency of departures at stations could be adjusted according to the features of each time period. For the places with more low-high stations in period 1, the departure time could be advanced to meet the demand.
Acknowledgment
This research was supported in part by the National Key R&D Plan (No. 2018YFC0831002) and Qingdao Social Science Planning Research Project (QDSKL1901121).

References:
[1] Manju Manohar Manjualavil, Gitakrishnan Ramadurai. Topological properties of bus transit networks considering demand and service utilization weight measures [J]. Physica A: Statistical Mechanics and its Application. 2020: 124683.
[2] Yan, X. a; Levine, J. a; Marans, R. b. The effectiveness of parking policies to reduce parking demand pressure and car use (Article)[J]. Transport Policy. 2019: 41-50.
[3] Bai T Z, Cai L, Zhu J Z, Lin X. Spatial difference analysis of coordinated development of rail transit and city: a case study of Beijing [J]. Transportation system engineering and information, 2020, 20 (03): 14-19.
[4] Quan W, Sun C. Research on accessibility of urban regional public transport based on hot spot detection [J]. Transportation system engineering and information, 2020, 20 (02): 231-236.
[5] Li Miaoyi, Long Ying. Evaluation of the service scope and spatial characteristics of major urban public transport stations in China [J]. Journal of urban planning, 2015 (06): 30-37.
[6] Zeng X. Evaluation of bus service capability based on IC card and GPS data [J]. Xi'an: Changan University, 2019.
[7] Yao H Y, Wang Y G, Zhang D D. A study on the design of multi-level bus network in mountainous cluster cities based on supply and demand coordination [J]. Journal of Transportation Engineering and Information, 2016, 14(2): 1-11.
[8] Li H B, Chen X W, Chen Z R. A method for estimating origin-destination matrix of public transit based on smart card and AVL data [J]. Journal of Transport Information and Safety, 2015, 33(6): 33-39.