

© <2021>. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
<http://creativecommons.org/licenses/by-nc-nd/4.0/>
The definitive publisher version is available online at [https://doi.org/
10.1016/j.physa.2021.126482](https://doi.org/10.1016/j.physa.2021.126482)

Journal Pre-proof

Spatial heterogeneity and migration characteristics of traffic congestion—A quantitative identification method based on taxi trajectory data

Xin Fu, Chengyao Xu, Yuteng Liu, Chi-Hua Chen, F.J. Hwang, Jianwei Wang



PII: S0378-4371(21)00755-X
DOI: <https://doi.org/10.1016/j.physa.2021.126482>
Reference: PHYSICA 126482

To appear in: *Physica A*

Received date: 27 January 2021
Revised date: 20 August 2021

Please cite this article as: X. Fu, C. Xu, Y. Liu et al., Spatial heterogeneity and migration characteristics of traffic congestion—A quantitative identification method based on taxi trajectory data, *Physica A* (2021), doi: <https://doi.org/10.1016/j.physa.2021.126482>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2021 Published by Elsevier B.V.

Spatial Heterogeneity and Migration Characteristics of Traffic Congestion-A Quantitative Identification Method Based on Taxi Trajectory Data

Xin Fu¹, Chengyao Xu^{2*}, Yuteng Liu², Chi-Hua Chen³, F.J. Hwang⁴, Jianwei Wang^{1*}

¹ College of Transportation Engineering, Chang'an University, China

² School of Economics and Management, Chang'an University, China

³ College of Mathematics and Computer Science, Fuzhou University, China

⁴ School of Mathematical and Physical Sciences, Transport Research Centre, University of Technology Sydney, Australia

*Correspondence: E-mail: 2018123013@chd.edu.cn; Tel. +86-180916821656(C.X.)

Abstract: It is of great reference significance to exploring spatial dependence of urban traffic activities and researching internal causes of regional traffic state changes for road network optimization and residents' travel behavior analysis. Based on trajectory data of taxis in Ningbo city of China, this study calculates average driving speed of taxis in different blocks during characteristic period and generates the global *Moran's I* and the *LISA* clustering diagram. On this basis, the spatial clustering characteristics of congestion on working days and non-working days are analyzed. Furthermore, in order to further characterize the changes of congestion from the perspective of spatial migration, a method of measuring geometric displacement is adopted to describe spatio-temporal migration trend of traffic states, four indicators designed to identify urban frequently congested areas, including migration direction, angle, distance, and low-value area. The results show that the high-clustering area are located urban fringe and the low-clustering area are located at geometric center of major urban areas. Spatial-temporal migration law of low-value areas in city-center is obvious. Difference between trend is compared with non-working days, the offset and azimuth of low-value area in downtown on working days are even bigger. The accurate capture of the characteristics of congestion space migration at the urban scale will help to formulate more targeted congestion management strategies.

Keywords: Traffic congestion; spatio-temporal migration; traffic status; GPS trajectory data; Spatial auto-correlation;

1. Introduction

As the artery of urban social and economic activities, urban traffic is the foundation to support the social and economic development of modern cities. An efficient and sustainable urban traffic environment plays an important role in promoting the economic and social development of cities. Urban travel patterns are consistently valued as they directly relate to the mobility of cities. Using vehicle GPS trajectory to analyze urban travel patterns has becoming a valuable perspective of research because of its fine spatio-temporal granularity. Researchers focus on spatial and temporal dimensions to analysis travel patterns, diversity of travel behavior, route selection behavior considering taxi trajectory data ^[1-5] and mobile data ^[6-7]. Meanwhile, urban traffic, which has unique spatial characteristics, is a complex structure composed of a large scale of traffic nodes and networks. Any local change in this structural system may cause a consequent change of traffic status in other regions. Furthermore, the traffic demand on the urban road network is changing dynamically all the time. When it rises in the same place at the same time, traffic congestion occurs and the operating efficiency of the road network is reduced accordingly. Considering the transmission mechanism and the adjacent-diffusion characteristics of the urban traffic congestion state, the spatial distribution of the traffic situation can provide only a static description of traffic congestion phenomenon which has happened without showing urban traffic sensitivity and real-time characteristics. Actually, the formation of traffic congestion, diffusion and convergence is a dynamic process evolving over time within an area. It thus cannot be sufficiently accurate to provide the description and recognition on the traffic congestion state only from one of time or spatial dimensions. Effective identification of urban traffic state and exploration of spatio-temporal evolution law have been the focus of research in the field of urban traffic management in recent years. Relevant studies have used mainly the GPS track data of taxis for the travel time and space analysis of residents ^[8-10], path analysis ^[11-12], road and regional traffic status analysis, local service level ^[13-14], etc. The GPS trajectory data of the floating car contains both the time attribute and spatial distribution attribute, and the data accuracy is relatively high. Hence it is effective and accurate to analyze the spatio-temporal evolution trend of regional traffic speeds with the help of GPS trajectory data of taxis, so as to identify effectively the traffic jam locations and jam state. Investigating traffic status using GPS track data of taxis, Lu et al. ^[15] estimated the probability density of real-time speed distribution of taxis with the kernel density estimation method and analyzed the congestion characteristics with the Gaussian mixture model. Altintasi et al. ^[16] converted the vehicular average speed in a certain time period into a specific traffic state parameter with the help of the track data of floating vehicles, so as to realize the identification of the traffic states of different sections of city. Weiming et al. ^[17] divided the whole urban road network into different sub-regions, converted the GPS track data of taxis into traffic flow data, and used the method of traffic flow matrix to detect abnormal areas of urban traffic. Li et al. ^[18] used the *k*-means clustering algorithm to extract hot urban areas and find the congested areas and periods according to the driving speeds of taxis. Wang et al. ^[19] proposed a visual analysis and interaction system for urban traffic congestion based on the GPS track data, which can automatically detect traffic jam events by calculating the speed of each section. Based on the GPS trajectory data, Chawla et al. ^[20] used the traffic flow model and OD matrix to infer the root cause of abnormal road traffic status.

Though most of the existing studies on traffic status identification are based on the track data of floating vehicles, there are significant differences in research methods. The kernel density, cluster analysis, and other methods have been widely used, but limited research has investigated urban traffic state from the perspective of spatial correlation. Traffic activities have a strong spatial auto-correlation, i.e., spatial dependence, and the traffic movement between adjacent areas of a city is correlated with each other. The traffic status of a certain area is transmitted to the surrounding areas through the road network and has a direct impact on adjacent areas. In addition, the traffic states of adjacent regions also interact with each other through the road network, which further makes an indirect impact on other regions.

Some related methodology of spatial econometrics in the literature have been used to identify the urban congestion area and situation, and the spatial and temporal migration of congestion is quantitatively described by the four sets of indicators, including migration direction, migration angle, migration distance, and low-value area in this paper. This study proposes a geometric procedure to identify the urban traffic congestion status and analyze the trend migration based on taxi track data. Considering Ningbo city of China as a case study, we analyze the spatio-temporal migration trend of traffic congestion as well as the agglomeration and diffusion effect of space and time by distinguishing weekdays and weekends. The conducted horizontal comparison is aimed at correctly understanding the temporal and spatial distribution characteristics and laws of urban traffic congestion, find the spatial structure of urban traffic congestion and the temporal and spatial distribution characteristics of residents' travel, and provide possibilities for traffic congestion prediction and prediction, and urban spatial planning. Therefore, this study aims to utilize the spatial autocorrelation test model^[21-22] for effective identification of urban traffic conditions, and on this basis, quantitatively describes the temporal and spatial migration law of the urban traffic congested areas with the help of the space geometry method. In this paper, global *Moran's I* statistics and local *Moran's I* aggregation map^[23,24] are used to quickly identify urban frequent congested areas. The advantages of our study are listed as follows:

(1) The cluster analysis, kernel density analysis, and threshold division have been widely applied in the identification of traffic congestion based on the fundamental traffic flow characteristics like speed and flow. Utilizing the space geometry method, this study recognizes a congested area not only by speed/flow but also by the spatial interaction and adjacency relationship.

(2) The main idea of traditional clustering methods is sub-dividing the investigated area, measuring the speed or flow of each sub-area, and then using a specific clustering algorithm to divide it into different congestion levels. However, one of its major drawbacks is that the frequently congested areas cannot be accurately identified. If (i) a certain sub-region is congested due to some sudden traffic situation within the region, (ii) the congestion occurs in the peak period of congestion, and (iii) the congestion is located in the boundary of the city's frequently congested area, then conventional methods such as cluster analysis or kernel density analysis cannot effectively distinguish the congested region. However, the spatial autocorrelation model can effectively avoid this issue because it can identify the areas with abnormal traffic status based on the statistical methods. Hence the procedure proposed by this study has the advantage of identifying accurately the urban frequently congested areas.

(3) The threshold segmentation takes the traffic state of each sub-region as the research object, obtains the speed and flow distribution under normal conditions through the historical data, and then calculates the deviation of the speed or flow from the stably observed(average) value in each region at the current moment for identifying the areas with abnormal traffic state. However, it requires high quality of data. Without sufficient data, the measured normal speed and flow rates are prone to extreme values. The technique used in this study does not have the issue and just needs data of working or non-working days.

2. Data Preparation and Preprocessing

2.1. Study area description

The study area in this paper is the Ningbo city, an important port city of China southeast coast. The latitude and longitude ranges of the city are from 120.55°E to 122.16°E, 28.51°N to 30.33°N. In this study, Zhenhai district and Beilun district within the belt expressway of Ningbo city are the main research area. The road network structure of the main research area is shown in Figure 1.

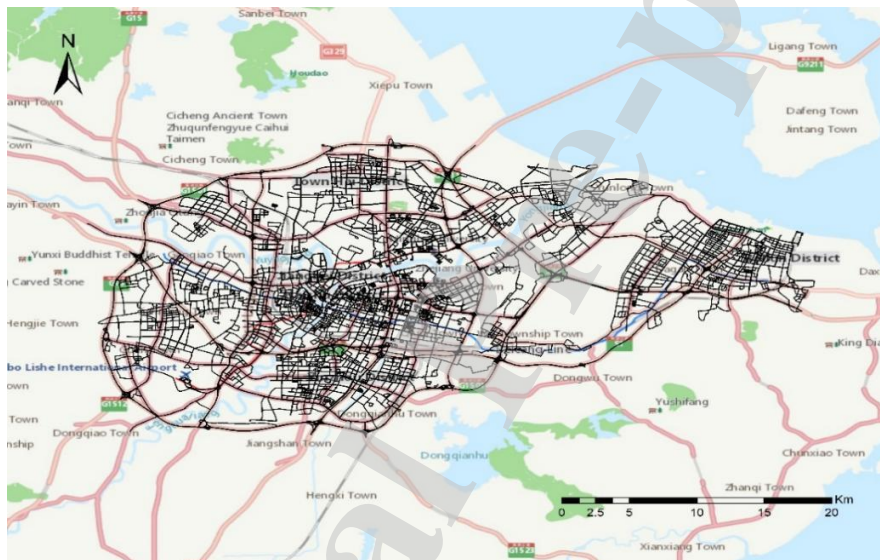


Figure 1. Ningbo road network

In this paper, the main research area is divided into the square grids of 500m×500m. It should be pointed out that most research work took the traffic area as the basic spatial unit in the spatial-related research of taxi track data. The basis of choosing the grid as the basic research unit in this paper lies in the following reasons. Firstly, the vehicle speed has a strong spatio-temporal sensitivity, which is similar to the spatial division of the traffic area, and it is difficult to accurately reflect the actual vehicular speed change in each zone due to the large difference in the area. In addition, the speed within the grid based on trajectory extraction is the average speed in one unit. The spatial autocorrelation test method adopted in this paper has a high sensitivity to the change of vehicle speed of adjacent land units, so it is necessary to minimize the individual research area and normalize the speed of each research object to ensure the rationality of the study.

2.2. Taxi trajectory data disposal and cleaning

The original data used in this paper are a series of location points collected from 4,000 taxi vehicles. The collection duration are one working day and one non-working day in June 2017. The specific information of trajectory data set is shown in Table 1 and Table 2. This track data has a total of 28,988, 152 records, with an average sampling interval of 15s.

Table 1. Information for raw data of taxi trajectory

Field Name	Field Type	Field Content	Remarks
Taxi ID	VARCHAR2(50)	Taxi id	-
Time Created	TIMESTAMP	Time	-
Latitude	NUMBER	Taxi GPS latitude	-
Longitude	NUMBER	Taxi GPS longitude	-
Speed	INT32	Driving speed	current instantaneous speed
Status	INT32	Passenger status	0 means empty car, 1 means passenger
Heading	INT32	Direction	0~359° is 0° in the north, clockwise

In order to ensure more accurate analysis results, abnormal GPS data should be preprocessed. This paper deals with the following exceptions:

(1) Eliminating the fields unrelated to this study

In view of the fact that this study is based on the GPS track data of taxis to study and mine the spatio-temporal law of urban residents' travel, the fields irrelevant to the research content are excluded according to actual needs.

(2) Excluding exception data for the related fields

The data of "ACC pass", "unlocated", "out of service" and other information are removed from the field "tflagDesc". The data in the field "aflagDesc" that shows "failure" and other information are removed. In addition, since the study involves the actual speed of taxi, the data of abnormal speeds are removed to ensure the reliability of the study content and results.

(3) Eliminate missing data

The instability of receiving the GPS signals leads to data loss, such as time data loss, GPS data loss, etc. In order to ensure the integrity of the research content, the data with missing values are removed.

Table 2. Part of the dataset after cleaning for a specific daytime

Vehicle Id	Longitude(°)	Latitude(°)	GPS Upload Timestamp	Vehicle Status	Speed(km/h)
4654137002	121.5526	29.8716	2017/06/06 00:00:03	0	20.8
4654137003	121.5521	29.8710	2017/06/05 00:00:04	0	22.2
4654137004	121.5313	29.7645	2017/06/06 03:23:09	1	34.8

2.3. Trajectory map matching

When the GPS track points are directly projected into the basic map, it is difficult to ensure whether the GPS track points and the road network in the basic map are perfectly matched due to errors in the GPS track data and differences in the geographic coordinates of various data sources. To address this matching issue, the so-called map matching method is utilized to match the GPS track data with the basic map data for determining the correct vehicular traffic status on the road map. In this paper, the affine transformation^[17] is adopted to conduct the spatial calibration of track data of taxis. Here a section of Jiangnan highway is used to demonstrate the matching result. As shown in Figure 2(a) and 2(b), the effect of the map matching method is obvious.



Figure 2. The distribution of GPS track points of taxis before (as shown in (a)) and after (as shown in (b)) the map matching

2.4. Matching of track points and block units

The method of matching trajectory points with spatial units (geographic grids) was used in this paper to determine the spatial relationship between vehicle trajectories and land parcels. The geographic coordinates of left bottom and top right corner of each plot unit are selected as the boundary points of the plot unit. The latitude and longitude information of boundary points and track points is used to determine land unit to which each track point belongs. The range of the plot unit to which it belongs to complete the matching of track points with each plot unit. The following is the implementation method in the python environment:

```
Range = [[lat1,lon1,lat2,lon2]1,[lat1,lon1,lat2,lon2]2,..., [lat1,lon1,lat2,lon2]i]
```

```
# The longitude and latitude information of the lower left and upper right corners of i plot units
```

```
point = [[lat1,lon1]1,[lat1,lon1]2,..., [lat1,lon1]j]
```

```
# The longitude and latitude information of j trajectory points
```

```
v = [[] for i in range(len(Range))]
```

```
#i represents a collection of all points falling in the i-th parcel unit
```

```
for i in range(len(Range)):
```

```
    for j in range(len(point)):
```


if (Range[i][0] <= point[j][0] <= Range[i][2]) & (Range[i][1] <= point[j][1] <= Range[i][3]):

$k = (\text{point}[j], \text{point}[j])$

$v[i].\text{append}(k)$

2.5. Definition of contiguous low-value area

The low-value land unit nearest to the geometric center of the main research area is selected as the low-value central area, denoted as K_0 . If the land units K_i and K_j have the common boundary and both of them are the low-value areas, then K_i and K_j are called the adjacent lands with low values, denoted as $K_i \leftrightarrow K_j$. Adjacent plots of low values can conduct such proximity. If $K_1 \leftrightarrow K_2 \leftrightarrow \dots \leftrightarrow K_h$, then any two block units in K_1, K_2, \dots, K_h are adjacent blocks with low values. The set of all land units in the main research area is denoted as U , and its low-value central area is clearly some land unit K_l in U , i.e. $K_0 = K_l \in U$. The contiguous low-value areas in downtown is the collection of low-value plot units adjacent to K_0 and defined by $\bar{U} = \{K_i | K_i \in U \text{ and } K_i \leftrightarrow K_0\}$.

3. Methodology

In previous studies, there are usually a variety of methods to characterize traffic congestion, and the most common one is to use the carrying capacity of the road to characterize traffic flow. After the acquisition of large-scale floating car data becomes possible, more and more practices have been made to use the track point velocity of floating cars to characterize the congestion level of its location. This paper adopts such a research methodology: based on the matching between the track points of the floating cars and the block units, this paper analyzes the spatio-temporal characteristics of the congestion of the block unit by using the speed level of the track point set in the block unit.

The measures of spatial autocorrelation usually include the global spatial autocorrelation and local spatial autocorrelation. The global spatial autocorrelation represents the global distribution of a phenomenon. It can be determined whether the phenomenon has the clustering characteristics in a specific region, but it is difficult to determine the exact location of clustering. The local spatial autocorrelation, which is used to determine the clustering of a local space, can not only spot the location of clustering but also detect spatial anomalies. Global spatial autocorrelation describes the speed of the spatial dependence degree in a total spatial range. Local spatial autocorrelation describes the speed of similarity degree between a spatial unit and its domain, can indicate the degree to which each local unit obeys the global general trend (including direction and magnitude), and prompts spatial heterogeneity, showing how spatial dependence changes with location. In this paper, *Moran's I* statistics (Moran, 1948) are used to investigate the spatio-temporal migration characteristics of speeds in Ningbo.

3.1. Spatial autocorrelation

(1) Global Moran's I statistics

Moran's I statistics reflect the similarity degree of attribute values of spatial adjacency. If Y is the observed value of the position (region), the global *Moran's I* value of the variable is calculated by the following formula

$$Moran's\ I = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (Y_i - \bar{Y})(Y_j - \bar{Y})}{S^2 \sum_{i=1}^n \sum_{j=1}^n w_{ij}} \quad (1)$$

where

$$S^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2 \quad (2)$$

and

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i. \quad (3)$$

Y_i represents the observation of region i , n stands for the total number of districts, and w_{ij} is the spatial weight matrix for region i and j . The standardized *Moran's I* statistics are:

$$Z = \frac{Moran's\ I - E(I)}{\sqrt{VAR(I)}} \quad (4)$$

where

$$E(I) = -\frac{1}{n-1} \quad (5)$$

$$VAR(I) = \frac{n^2 w_1 + n w_2 + 3 w_0^2}{w_0^2 (n^2 - 1)} - E^2(I) \quad (6)$$

$$w_0 = \sum_{i=1}^n \sum_{j=1}^n w_{ij} \quad (7)$$

$$w_1 = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (w_{ij} + w_{ji})^2 \quad (8)$$

and

$$w_2 = \sum_{i=1}^n \sum_{j=1}^n (w_{i.} + w_{.j})^2 \quad (9)$$

Notice that $w_{i.}$ and $w_{.j}$ are the sum of the rows and columns in the space weight matrix, respectively. Under the null hypothesis that there is no spatial correlation, Z obeys the standard normal distribution.

The value of *Moran's I* generally exists between $[-1, 1]$. A value of *Moran's I* less than 0 represents a negative correlation, a value equal to 0 means irrelevance, and that greater than 0 means a positive correlation. The closer to 1 it is, the greater the difference between cells, i.e., the less concentrated the distribution is. The closer to 0 it is, the closer the relationship between units is and the more similar the nature is (high-value clustering or low-value clustering). Approaching 0 indicates that the cells are not related. The Z value can test its significance.

(2) Local Moran's I statistics

The local *Moran's I* statistics are shown as follows:

$$I_i = \frac{(x_i - \bar{x})}{S^2} \sum_{j=1, j \neq i}^n w_{ij} (x_j - \bar{x}) \quad (10)$$

$$S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (11)$$

and

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (12)$$

where I_i is the local *Moran's I* of unit i , w_{ij} is the spatial weight between unit i and unit j , n is the number of units of space adjacent to unit i . The variable x_i represents the average speed of unit i , and \bar{x} is the mean of variables x_i . A positive value of I_i indicates that the space unit is similar to the attributes of adjacent units ("high-high" or "low-low"), while a negative value of I_i indicates that the space unit is not similar to the attributes of adjacent units ("high-low" or "low-high").

Since the road network is the carrier of the vehicle speed and the road network has spatial continuity, the paper selects the widely used binary adjacency matrix for spatial autocorrelation analysis. The element of the binary adjacency matrix, w_{ij} is defined as follows:

$$w_{ij} = \begin{cases} 1, & \text{if the adjacent region } i \text{ and } j \text{ have a common boundary;} \\ 0, & \text{otherwise.} \end{cases} \quad (13)$$

3.2 Method for defining azimuth angle and offset distance

The azimuth angle, it is one of the ways to measure the angle difference between objects in a plane. The horizontal angle between a point's pointing line clockwise and the target's pointing line. The azimuth angle was used to measure the real-time speed of every vehicle object in this paper. Assume that the azimuth of B relative to A is calculated, i.e., A is the current position and B is the target position. Notice that the north latitude is positive and the south latitude is negative. The east longitude is positive, and the west longitude is negative. The azimuth starts with the true north at 0 degrees and rotate 360 degrees clockwise from east to south to west.

$$\cos \angle AOB = \cos \angle AOC \times \cos \angle BOC + \sin \angle AOC \times \sin \angle BOC \times \cos \angle C \quad (14)$$

$$L = R \times \frac{\arccos(\cos \angle AOB)}{180^\circ} \times \pi \quad (15)$$

$$Azimuth = \arcsin\left(\frac{\sin \angle BOC \times \sin \angle C}{\sin \angle AOB}\right) \quad (16)$$

where R is the average radius of the earth; *Azimuth* is the Azimuth Angle, that is, the dihedral Angle $\angle B - OA - C$ in figure 3 and $\angle C$ is the dihedral Angle $\angle B - OC - A$ in Figure 3. L is the offset distance.

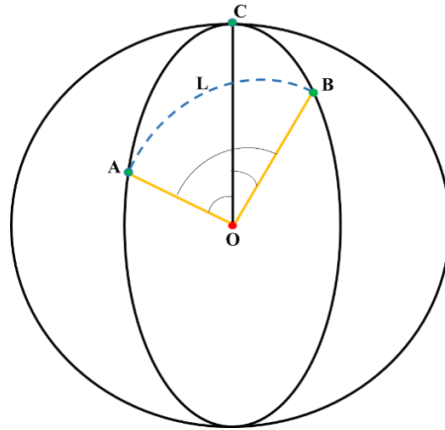


Figure 3. Schematic diagram of azimuth and offset distance

4. Results and Interpretation

The spatial correlation analysis can detect the randomness of event distribution in space, and the detection of non-randomness indicates that the event is correlated in space. Real-time vehicle speed has certain characteristics of spatial agglomeration or distribution. Specifically, it is shown that the vehicular speed of a region is not only related to the influencing factors of the region, e.g., road network density, road grade, land attributes, etc., but also related to the vehicular speed of adjacent regions. Through the grid unit division of the real-time speed, this section aims to use *Moran's I* statistics to detect the non-randomness of spatial clustering or distribution of the real-time speed of adjacent grid cells and prove its spatial relevance.

4.1. Global spatial autocorrelation characteristics

Based on the GPS track data of taxis in Ningbo in June 2017, the space matrix adopted R adjacent matrix, and calculated the global space-time *Moran's I* of Ningbo from 16:00-21:00 on June 4, 2017 (non-working day) and 16:00-21:00 on June 5, 2017 (working day) through Geoda software. The inspection results are shown in Table 3 and 4. We can notice that the global *Moran's I* of the real-time speed of block unit in the study area was positive and had passed the significance test of 5%, which indicates that the real-time speed in the study area showed the positive correlation characteristics in space. It can be seen that the real-time speed in study area was affected by the spatial correlation factors and showed a trend of agglomeration in the spatial distribution. From 16:00-21:00 on weekdays to 16:00-21:00 on non-weekdays, the global *Moran's I* of the block unit real-time speed in study area was basically greater than 0.3, the *Moran's I* of some time periods was more than 0.4, and the corresponding Z value was greater than 4. The real-time speed showed the obvious spatial aggregation. Besides, the spatial aggregation of working days is significantly greater than that of non-working days, which is mainly due to the unconstrained travel time of non-working days, relatively small area of congestion aggregation and stochastic travel destinations of the morning peak of non-working days.

Table 3. Working day global space auto-correlation indicator results

Time Slot	I	E[I]	Mean	Sd	Z-value
Time 1	0.3382	-0.0007	0.0001	0.0171	19.7533
Time 2	0.3279	-0.0007	-0.0002	0.0160	20.5636
Time 3	0.3358	-0.0007	0.0002	0.0159	21.0873
Time 4	0.3536	-0.0008	-0.0014	0.0160	22.1619
Time 5	0.3532	-0.0008	-0.0011	0.0170	20.8747
Time 6	0.3326	-0.0008	-0.0003	0.0174	19.0931
Time 7	0.3955	-0.0009	-0.0007	0.0177	22.3544
Time 8	0.4027	-0.0009	-0.0031	0.0172	23.6436
Time 9	0.3614	-0.0010	-0.0011	0.0200	18.1626
Time 10	0.3522	-0.0009	0	0.0176	19.9585

Table 4. Non-working day global space auto-correlation indicator results

Time Slot	I	E[I]	Mean	Sd	Z-value
Time 1	0.3652	-0.0007	-0.0013	0.0149	24.5807
Time 2	0.3415	-0.0007	-0.0010	0.0152	22.5445
Time 3	0.3783	-0.0007	-0.0007	0.0157	24.0763
Time 4	0.2915	-0.0007	-0.0010	0.0160	18.2890
Time 5	0.3793	-0.0008	0.0002	0.0162	23.4175
Time 6	0.3769	-0.0008	-0.0015	0.0169	22.3700
Time 7	0.3391	-0.0008	0.0001	0.0167	20.3470
Time 8	0.3563	-0.0008	0.0006	0.0186	19.1420
Time 9	0.3182	-0.0008	-0.0019	0.0176	18.1897
Time 10	0.3167	-0.0009	0.0007	0.0181	17.4246

It can be seen from Figure 4 that the global Moran index of time period 2 was lower than that of time period 1, and the spatial clustering of real-time speed was weak during weakened. It shows a continuous upward trend, and gradually decreases after time period 4 until time period 6. When the global Moran index was reduced to the lowest value, it then ushered in a second rise, and finally began to decline at the beginning of the period 8 and stabilized. Therefore, it can be concluded that the blocks' real-time speed spatial agglomeration effect is most obvious during the two periods 17:30-18:30 and 19:00-20:00 on the working day night. It can be observed from Figure 4 that the spatial agglomeration effect of real-time speed on non-weekday nights was mainly concentrated in the following three periods, namely: 17:00-17:30, 18:00-19:00 and 19:30-20:00. On the whole, the global Moran's I value shows higher value and relatively small change amplitude in the working day period.

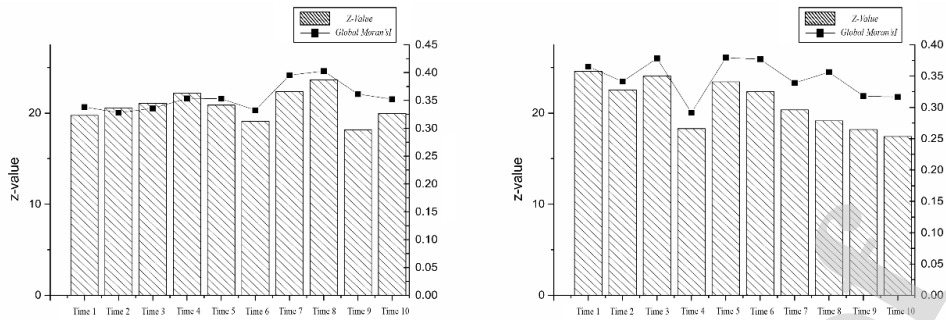


Figure 4. Working day (left) and non-working day (right) global Moran index and Z-value

4.2. Local spatial autocorrelation characteristics

Using Geoda software, the Moran scatter map and LISA (local indicators of spatial association) cluster map were calculated, respectively. The local spatial autocorrelation of block units' real-time speed was conducted one working and non-working days. The variation trend of local *Moran's I*, specific spatial location of agglomeration area and heterogeneous area, and spatio-temporal evolution rule of passenger flow were analyzed.

4.2.1. Stability of spatial autocorrelation

The first quadrant is defined as the H-H (high-high) quadrant. If a certain block unit is in the H-H quadrant, the unit of the plot and the neighboring unit have the relatively fast real-time speed. The second quadrant is defined as the L-H (low-high) quadrant. If a certain block unit is in the L-H quadrant, the real-time speed of the block unit in the plot is slow but the real-time speed of the surrounding unit is relatively fast. The third quadrant is the L-L (low-low) quadrant. If a certain block unit is in the L-L quadrant, the speed of the block itself is slow and the real-time speed of the surrounding unit is also slow. The fourth quadrant is the H-L (high-low) quadrant. If a block unit is in the H-L quadrant, the unit itself has a fast speed but the real-time speed of surrounding units is relatively slow. Also, the H-H and L-L quadrants correspond to the typical positive spatial autocorrelation with obvious spatial dependence and are respectively called the high-value clustering and low-value clustering, which are typical manifestations of spatial clustering. If a unit is of the H-H or L-L type, it indicates that there is a little spatial difference between different observation values. The L-H and H-L quadrants are associated with the typical negative spatial autocorrelation, showing the obvious spatial heterogeneity, and are also known as the spatial outliers. If a unit is of the L-H and H-L, it indicates that there is a large spatial difference between different observed values. If all the units are uniformly distributed in the four quadrants of the local scatter diagram, this situation indicates that there is no spatial correlation between the real-time speeds of any two different blocks in the research area.

Through the analysis of Figure A1 and Figure A2 (as shown in Appendix A), the following conclusion can be drawn. Most land units were concentrated in the third quadrant, which indicates that the spatial dependence between the real-time speed of block units in the research area of Ningbo is very strong. The local correlation of the real-time vehicle speeds of each local block unit in the research area of Ningbo city was changing with time, but the change amplitude was small, which indicates that the spatial

correlation of the real-time vehicle speed of each local block unit in the research area of Ningbo city was relatively stable.

4.2.2. Evolutionary trend of L-L aggregation area

Local spatial autocorrelation analysis of traffic conditions during non-working days (16:00-21:00) In the non-weekday evening period, the local spatial autocorrelation in the main research area was basically good, generally presenting a state of concave down in the middle. The H-H accumulation area was mainly distributed in the east and southeast of the main urban area and the peripheral high-speed area around the city. The L-L clusters were mainly distributed in northeast and geometric center of the research area.

It can be seen from Figure 5 that the L-L aggregation area was distributed in the geometric center of the main urban area from 16:00 to 19:00 during a non-weekday evening. According to the distribution of POI location on the online map, it can be found that most of the properties of this area are commercial, with dense road network and large traffic flow. From 19:00 to 21:00, the low-speed accumulation area gradually shrunk to the geometric center of the city. Compared with the main urban areas, east and northeast places showed a similar situation. The L-L aggregation area was concentrated in the larger period from 16:00 to 19:00, and then gradually shrunk.

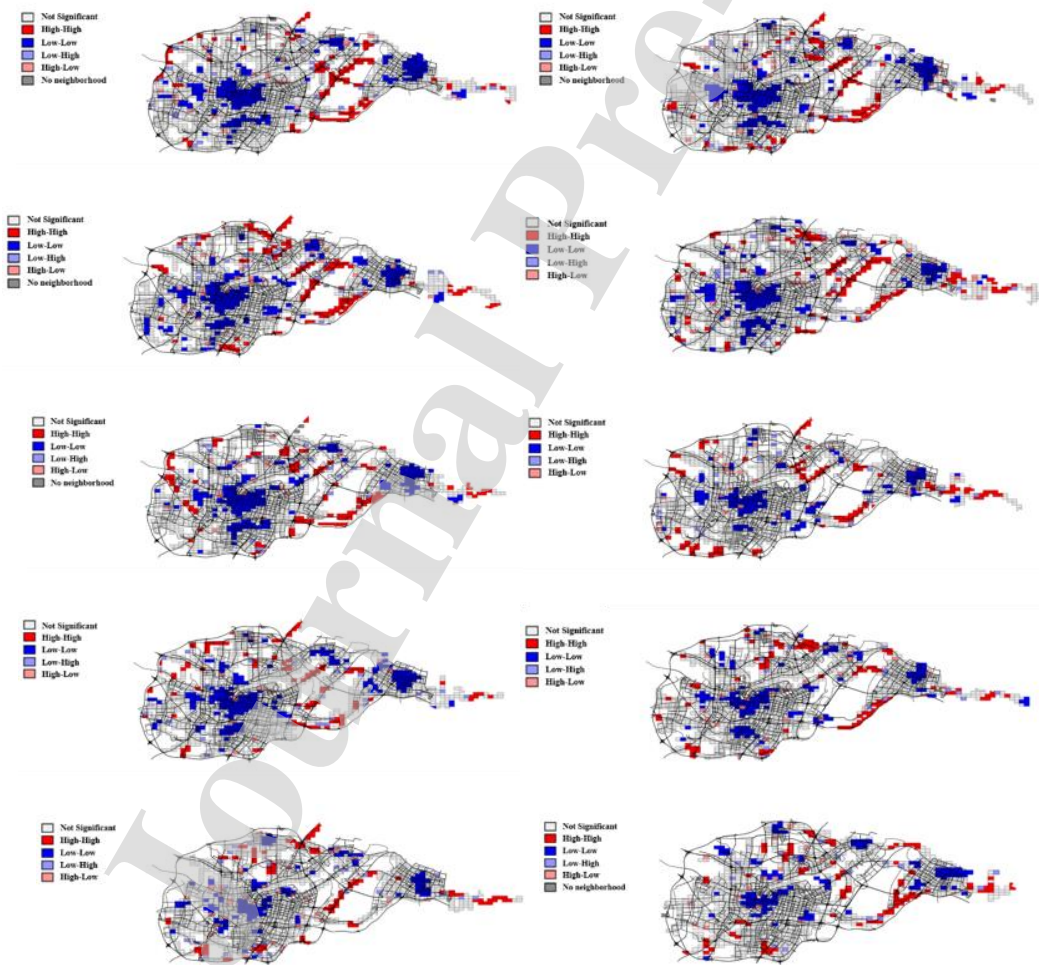


Figure 5. Non-working days (16:00-21:00) LISA aggregation

Considering the actual road network, the congestion points in the city center first appeared in the plot unit in the southwest of prosperous area in the city center, which was characterized by the relatively large proportion of secondary trunk roads. Then, the congestion point was shifted to the northeast with a relatively short distance. The center of the new congestion area was located in the plot unit with secondary trunk roads accounting for more than 96%. Between 17:00 and 19:00, the center of the congestion area moved back and forth along the north-south direction. The proportion of roads of all levels in the block unit was roughly the same, and the road network structure presented diversified features. Subsequently, the center of the congestion area moved eastward, with a large deviation distance. After 7:30, the congestion center moved in the north-south direction, and the secondary arterial roads in the block units accounted for more than 60%. To sum up, the congestion transmission characteristics of non-working day downtown area are as follows. The congestion center started from the area with dense secondary trunk roads, and then moved to the area with diversified road network structure. It finally returned to the area with dense secondary trunk roads again.

Table 5. Related indicators of non-working day low value area migration trend

Time Slot	Geometric Center Offset Direction	Geometric Center Azimuth	Geometric Center Offset Distance	Number of Continuous Low Value Areas in the City Center
Time 1	-	-	-	60
Time 2	North east	52.9°	66m	56
Time 3	North east	28.1°	550m	64
Time 4	South west	2.1°	610m	65
Time 5	North east	19.9°	404m	67
Time 6	South west	32.9°	165m	54
Time 7	South east	60.3°	507m	69
Time 8	North east	55.4°	373m	52
Time 9	South west	57.0°	844m	46
Time 10	South east	18.5°	382m	37

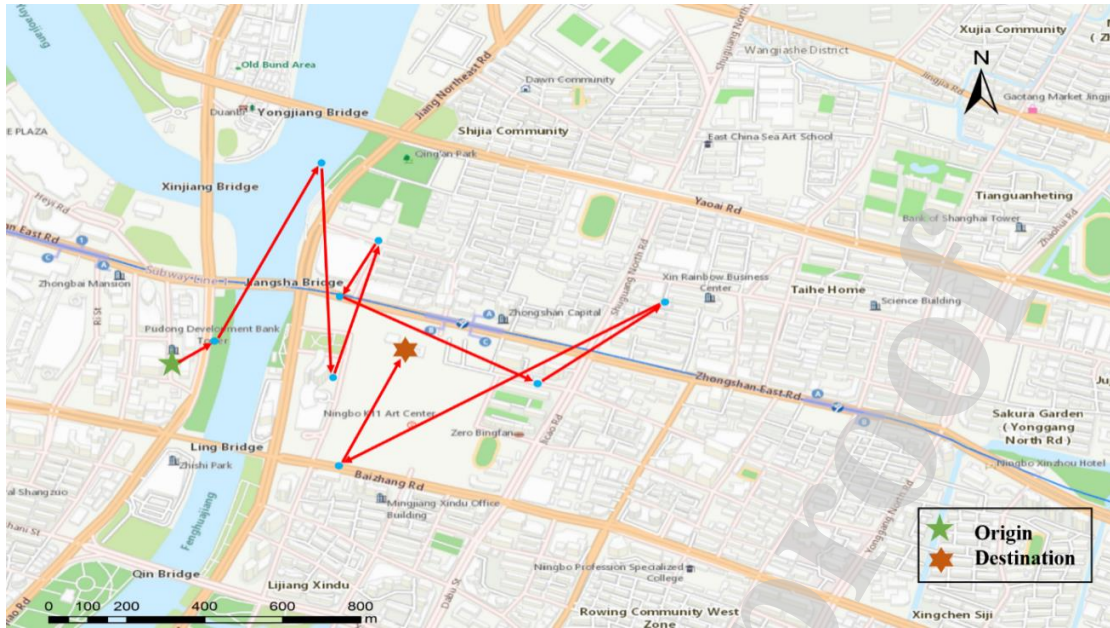


Figure 6. Spatial position change of the geometric center of the low-value area in the city center during non-working days

In this paper, the spatio-temporal migration characteristics of low-value areas in city center of non-working days are further quantitatively analyzed. As shown in Table 5 and Figure 6, the results are summarized as follows:

- (1) The L-L clustering area in the city center was large during the following periods: 16:00-16:30, 17:00-18:30 and 19:00-19:30;
- (2) Before 18:30, the low-value clusters in the city center continued to shift to the north of the city, and between 17:00 and 18:30, the offset distance of the L-L gathering area was larger;
- (3) After 18:30, the low-value clusters in the city center continued to shift to the south of the city, and during the period of 19:00-19:30 and 20:00-20:30, the offset distance of the L-L gathering area was larger.

As to local spatial autocorrelation analysis of traffic state during working night (16:00-21:00), during the working night period, the local spatial autocorrelation of local blocks and units was well, and the distribution rule was basically similar to that of the non-working day period. In general, the spatial concentration performed concave down in the middle and convex on around. Moreover, the change rule was obvious. It can be seen from Figure 7 that from 16:00 to 17:00 on weekday night, the L-L aggregation areas were distributed in the geometric center of the main urban area with a small number. From 17:00 to 18:00, the L-L aggregation area in the city center spread around and then became stable. During 18:30-19:00, the L-L aggregation area in the central region rapidly spread outward between 19:00-19:30 after a short period of shrinkage. From 19:30 to 21:00, the low-value accumulation area in the downtown gradually shrank, and the number of plots in the H-H accumulation area near the expressway around the city in the southeast of the main city increased.

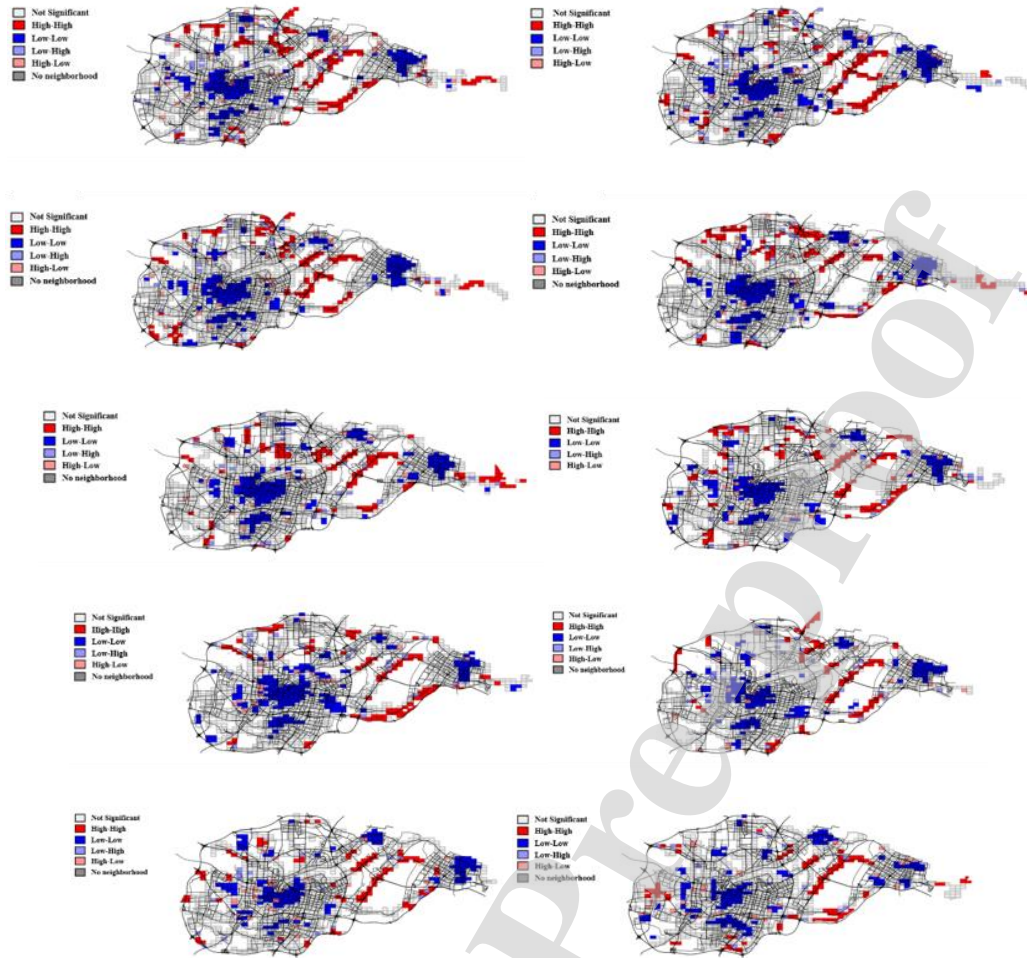


Figure 7. Working day (16:00-21:00) *LISA* aggregation chart

The spatio-temporal migration rule of the low-value accumulation area in the main part of east district was basically similar to that in the central part of the main city. However, between 19:30 and 20:30, the low-value accumulation area in east district expanded for the second time and then shrank. Northeast position showed a relatively different change rule, reflected in: L-L zones expanding time in northeast position obviously lag the main and east districts. From 8-21:00 period, northeast position has been appeared large areas of low-value clusters, meanwhile northeast district did not appear same situation, so it can deduce the area residents' travel behavior and the travel law is strongly influenced by their working time.

The spatio-temporal deviation of the central position of the congestion area in the corresponding period of the working day showed a completely different regularity. There has a large proportion of sub-trunk road in the unit of this plot, and there is no trunk road. Then, before 17:30, the congestion area center moved back and forth in the east-west direction, with a large offset distance. In the above period, the ratio of different types of roads in the block units located in the congestion area center was roughly the same, and the road network structure showed diversified features. From 17:30 to 20:00, the center of the congestion area moved repeatedly along the north-south direction, with a relatively short deviation distance. The central point of the congestion area was always located in the block unit with a large

proportion of the secondary trunk road. Subsequently, the central location of the congestion area was shifted eastward for a long distance, and finally fell into the block unit with roughly the same proportion of all types of roads. To sum up, the congestion conduction characteristics of the downtown area on weekdays are described as follows: The congestion center started from the area with dense secondary arterial road, and then moved to the block units with diversified features in the road network structure. During the evening rush hour, the congestion center moved to the area with dense secondary trunk roads and lack of main roads and the migration distance was relatively short. Then it was migrated eastward for a long distance and finally returned to the block units with diversified structural features of the road network structure.

Table 6. Relevant indicators of the migration trend of low-value areas in working days

Time Slot	Geometric Center Offset Direction	Geometric Center Azimuth	Geometric Center Offset Distance	Number of Continuous Low Value Areas in the City Center
Time 1	-	-	-	60
Time 2	North east	82.4°	601m	56
Time 3	South west	69.7°	517m	64
Time 4	North east	33.0°	153m	65
Time 5	North east	50.4°	55m	67
Time 6	South west	53.6°	274m	54
Time 7	South east	82.0°	33m	69
Time 8	North west	25.7°	332m	52
Time 9	South east	71.4°	846m	46
Time 10	South west	4.3°	284m	37

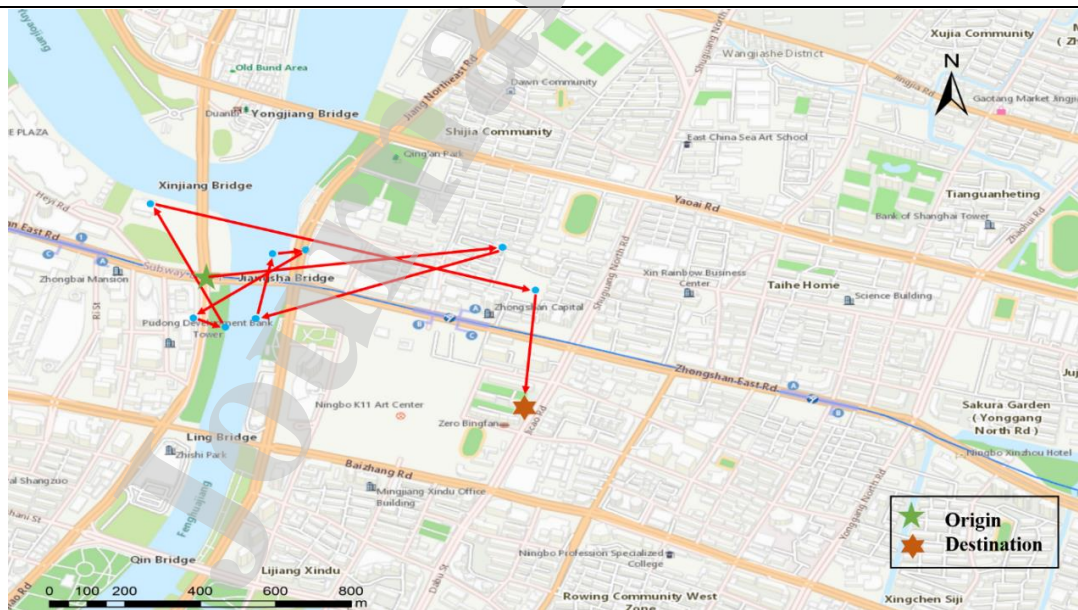


Figure 8. Changes in the spatial position of the geometric center of the low-value area in the city center during the working day

Further quantitatively analysis was conducted on the temporal and spatial migration characteristics of the low-value area in the city center during the working days. As shown in Table 6 and Figure 8, the results are summarized as follows:

- (1) The L-L cluster area in the city center was large during the following periods: 16:00-16:30, 17:00-18:30 and 19:00 -19:30. The rule was consistent with that of non-working days;
- (2) Before 18:30, the low-value clusters in the city center continued to shift to the north of the city, and between 16:30-17:30, the offset distance of the L-L gathering area was larger. Compared with the non-working days, L-L low-value zone was offset to the north for the same time, but the start and end times of the larger offset were slightly advanced;
- (3) After 18:30, the low-value clusters in the city center continued to shift to the south of the city. During the period of 20:00-20:30, the offset distance of the L-L gathering area was larger;
- (4) Compared with non-working days, the offset distance of the low-value area in the downtown area was larger overall, and the range of the low-value area was unstable;
- (5) Between 16:00 and 18:30, the deviation angle of the geometric center of the low-value area in the city center of working days was larger, while that of the low-value area in the city center of non-working days was smaller. In other words, during the above period, the L-L aggregation area in the city center not only moved in the north-south direction, but also moved in the east-west direction. However, the L-L concentration area, which was not in the city center on weekdays, was only offset in the north-south direction.

5. Discussion

Based on the obtained research results as well as the characteristics of Ningbo's urban structure and road network, we have the following discussion.

Firstly, the real-time speed of taxis in the study area is confirmed to show a positive spatial correlation. And in terms of time distribution, the spatial clustering of working days is obviously greater than that of non-working days. Resident travel on weekdays mainly involves locations in commercial areas, and the spatial clustering of workdays is consistent with the time period for residents to commute to and from get off work. On non-working days, residents travel more freely and there is no specific high concentration area. Therefore, residents' travel behavior is relatively scattered. In the evening hours of non-working days, the crowded areas are mainly distributed in the east and southeast of the main urban area and the surrounding urban expressway areas. In the evening hours on weekdays, from 19:30 to 21:00, the number of plots in the H-H gathering area near the Ring Expressway in the southeast of the main urban area increased.

Secondly, by comparing the LISA aggregation maps on weekdays and non-weekdays, it can be found that the urban road congestion is concentrated in the center of the main urban area, the center of northeast district, and the main location of east district. Through the study, it can be found that the causes of congestion in the three low-value aggregation areas are different. Although Ningbo has a "double-core"

structure, the Euclidean distance from the east to the west of the city is long and that from the north to the south is short. Besides, the double cores are far away from each other and thus cannot play a role in relieving the traffic pressure. As city center has been the main commercial center of Ningbo for a long time, all kinds of public service facilities are available in this area. Hence the city center is extremely crowded and the transportation infrastructure is insufficient. East district is another core area of the city. Although the traffic pressure is not as considerable as that of the central area of the city, the transportation infrastructure of this area is not sufficient. Although the road network is dense, it lacks trunk roads. Thus, it becomes the main congestion area of the city. As for northeast district, the main reason for its congestion is the low density of road network and the poor road accessibility. Therefore, the following three suggestions are proposed for the three congestion areas:

- (1) In the future urban layout planning, Ningbo City should try to disperse the service facilities in the downtown area, promote the expansion of commercial centers, medical and educational services to the periphery, and guide the population of the city center to migrate to other areas of the city, eventually forming a “dual-core and multi-center” structure. The urban structure thus effectively alleviates the traffic pressure in the downtown area;
- (2) In order to alleviate the traffic congestion in the main area of east district, the main roads should be added on the basis of the existing network structure. The narrow roads should be widened, and the traffic infrastructure in this area should be improved to solve the congestion problem from the root;
- (3) Northeast district has a special geographical position and is located in the middle of the core areas of the two major cities. Compared with the other two regions, the cause of congestion in Northeast district is relatively simple, i.e., the regional road network density does not match its actual traffic capacity. In response to this problem, the overall road network density and coverage of the road network in Northeast district should be improved.

6. Conclusions

In this paper, the grids unit is taken as the smallest research areas based on Ningbo city taxi GPS trajectory data, we have calculated and analyzed the range blocks around the average speed, global Moran and local Moran index, azimuth, offset direction, offset distance and scope change of quantitative analysis of four groups of space-time migration law of low accumulation. The research conclusions are shown as follows:

- (1) On both weekdays and non-weekdays, the real-time block unit speed in the study area shows a strong positive correlation in space. The study has confirmed that the real-time speed is affected by the spatial correlation factors, and a tendency of agglomeration in the spatial distribution has been demonstrated.
- (2) The global Moran index of all time periods fluctuates with time. Compared with working days, the block unit speed in the spatial aggregation period of non-working days was delayed, but the aggregation begins to dissipate earlier.
- (3) The LISA aggregation figure has shown that the spatial distribution of the H-H or L-L aggregation area is relatively stable. The H-H aggregation area is distributed near the circum-city expressway on the edge of the city, while the L-L aggregation area is mainly concentrated in the center of the main urban area, the center of Northeast district and the main area of east district.

(4) The spatio-temporal migration of low-value areas in the city center is obvious, and there are significant differences between working days and non-working days. Compared with non-working days, the offset and azimuth of the low-value area of the downtown on working days are generally large. Besides, the range of the low-value area on working days is unstable.

Appendix A

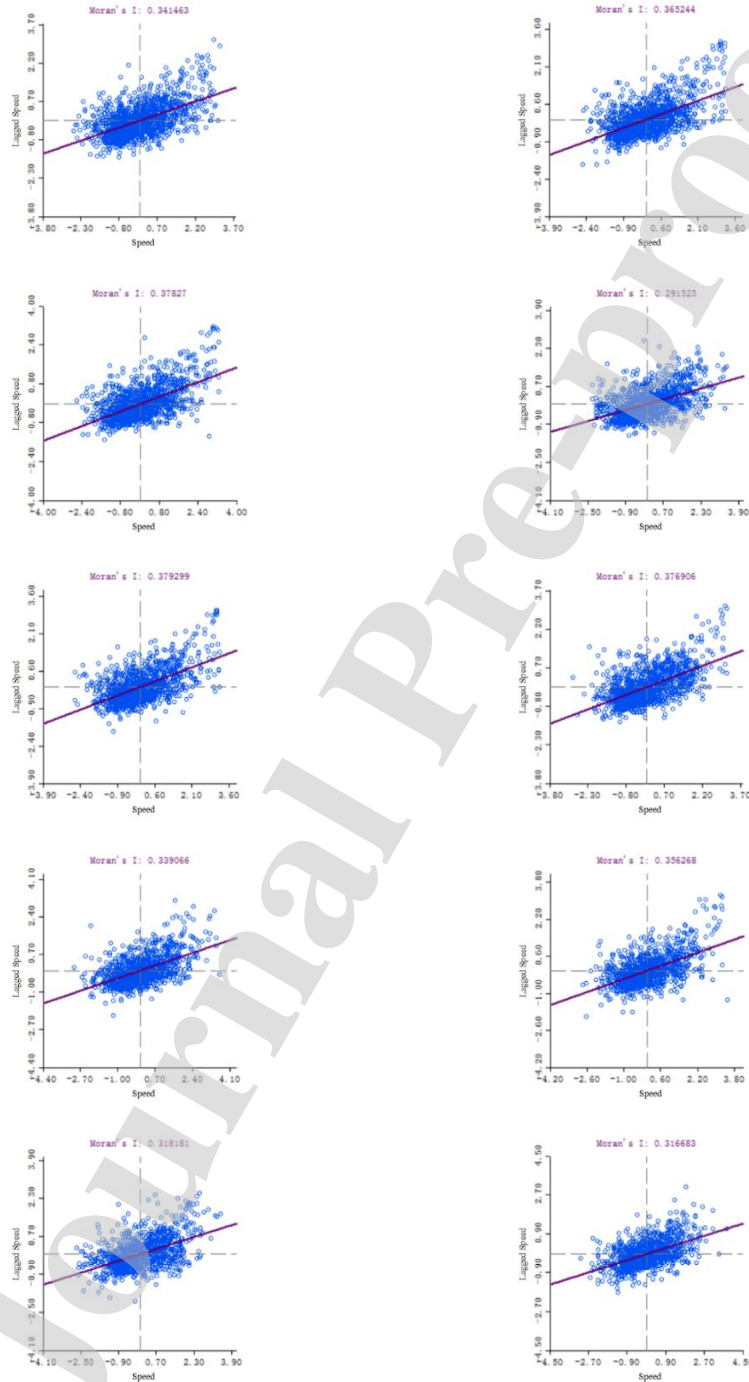


Figure A1. Non-working day Moran scatter plot

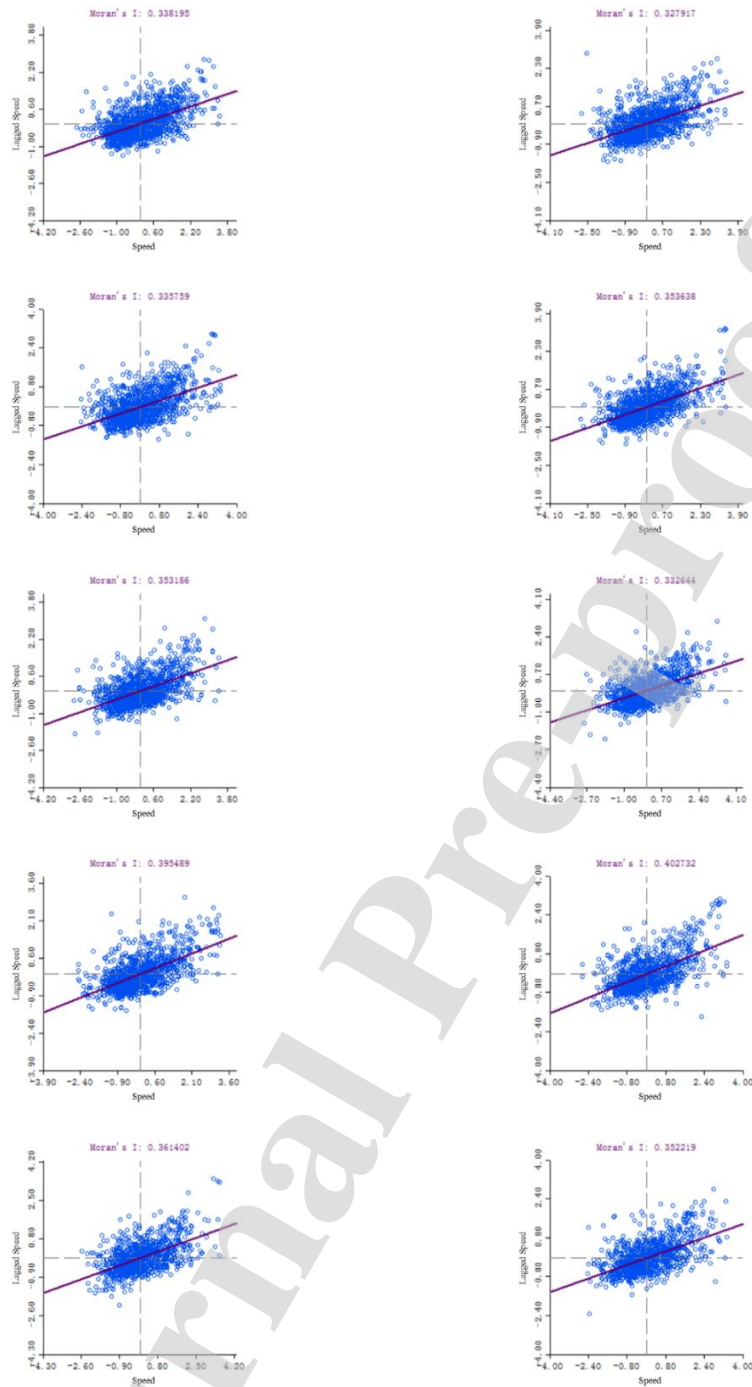


Figure A2. Working day Moran scatter plot

Acknowledgements

Funding: This work was supported in part by the Fundamental Research Funds for the Central Universities under Grant 300102230501, and in part by Chang'an University (Xi'an, China) through the National Key Research and Development Program of China (No. 2020YFC1512000).

References

1. Exploring urban travel patterns using density-based clustering with multi-attributes from large-scaled vehicle trajectories"[J]. *Physica A*. Vol.561, 2020, 125301.
2. Zheng L , Xia D , Zhao X , et al. Spatial-temporal travel pattern mining using massive taxi trajectory data[J]. *Physica A: Statistical Mechanics and its Applications*, 2018:S0378437118301419.
3. Tang J , Liu F , Wang Y , et al. Uncovering urban human mobility from large scale taxi GPS data[J]. *Physica A Statistical Mechanics & Its Applications*, 2015, 438:140-153.
4. A mixed path size logit based taxi customer-search model considering spatio-temporal factors in route choice"[J]. *IEEE Transactions on Intelligent Transportation Systems*, 21(4), 1347-1358. 2020.
5. Taxi Trips Distribution Modeling Based on Entropy-maximizing Theory: A Case Study in Harbin City - China". *Physica A*[J], Vol.493, 430-443, 2018.
6. Shan, Jiang, Joseph, et al. Activity-Based Human Mobility Patterns Inferred from Mobile Phone Data: A Case Study of Singapore[J]. *IEEE Transactions on Big Data*, 2017.
7. Wang Z , Hu Y , Zhu P , et al. Ring aggregation pattern of metro passenger trips: A study using smart card data[J]. *Physica A Statistical Mechanics & Its Applications*, 2017, 491:471-479.
8. Jiang B , Yin J , Zhao S . Characterizing the human mobility pattern in a large street network[J]. *Physical Review E*, 2009, 80(2):021136.
9. Liu L , Andris C , Ratti C . Uncovering cabdrivers' behavior patterns from their digital traces[J]. *Computers Environment & Urban Systems*, 2010, 34(6):541-548.
10. Scholz R W , Lu Y . Detection of dynamic activity patterns at a collective level from large-volume trajectory data[J]. *International Journal of Geographical Information Science*, 2014, 28(5):946-963.
11. Tang J , Liang J , Zhang S , et al. Inferring driving trajectories based on probabilistic model from large scale taxi GPS data[J]. *Physica A: Statistical Mechanics and its Applications*, 2018:S0378437118305016.
12. Yulong W , Kun Q , Yixiang C , et al. Detecting Anomalous Trajectories and Behavior Patterns Using Hierarchical Clustering from Taxi GPS Data[J]. *ISPRS International Journal of Geo-Information*, 2018, 7(1):25-.
13. Jiang S , Guan W , He Z , et al. Measuring Taxi Accessibility Using Grid-Based Method with Trajectory Data[J]. *Sustainability*, 2018, 10(9).
14. Cui, Jianxun, Liu, et al. City-wide examining transport network accessibility using taxi GPS data[C]// *Cota International Conference of Transportation Professionals*. 2015.
15. Lu S F , Mai Y H , Liu X M . The Analysis of Characterization of Urban Traffic Congestion Based on Mixed Speed Distribution of Taxi GPS Data[J]. *Applied Mechanics and Materials*, 2012, 241-244:2076-2081.
16. Altintasi O , Tuydes-Yaman H , Tuncay K . Detection of urban traffic patterns from Floating Car Data (FCD)[J]. *Transportation Research Procedia*, 2017, 22:382-391.
17. Weiming K , Shi A , Huifu J . Detecting Traffic Anomalies in Urban Areas Using Taxi GPS Data[J]. *Mathematical Problems in Engineering*, 2015, 2015:1-13.
18. Li M , Li R T , Xia Y , et al. Analysis of Urban Traffic Based on Taxi GPS Data[J]. *Lecture Notes in Electrical Engineering*, 2014, 279:1007-1015.
19. Wang Z , Lu M , Yuan X , et al. Visual Traffic Jam Analysis Based on Trajectory Data[J]. *IEEE Transactions on Visualization and Computer Graphics*, 2013, 19(12):2159-2168.
20. Chawla S , Zheng Y , Hu J . Inferring the Root Cause in Road Traffic Anomalies[C]// *IEEE International Conference on Data Mining*. IEEE, 2013.
21. Cheng Z , Zu Z , Lu J . Traffic Crash Evolution Characteristic Analysis and Spatiotemporal Hotspot Identification of Urban Road Intersections[J]. *Sustainability*, 2018, 11(1).
22. Toran Pour A , Moridpour S , Tay R , et al. Influence of Pedestrians' Age and Gender types on Spatial and Temporal Distribution of Pedestrian Crashes[J]. *Traffic Injury Prevention*, 2017:0-0.

23. Chen Yanguang. Development and method improvement of spatial autocorrelation theory based on Moran statistic[J]. Geographical Research,2009,28(06):1449-1463.
24. Guo Jinyun, Zhu Mingfa, Xu Yulin. Comparative Analysis of Affine Transformation and Similar Transformation in Geometric Correction of Map Data[J]. Bulletin of Surveying and Mapping, 2001(04): 23-24+27.

Journal Pre-proof

Highlights:

- 1) Quantitative analysis of azimuth, offset direction, offset distance and scope change of Taxi average speed block.
- 2) Confirmation of the real-time taxi speed that is affected by the spatial correlation factors, and a tendency of agglomeration in the spatial distribution.
- 3) Demonstration of the spatial and temporal evolution law of low-value areas in the city center.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Journal Pre-proof