ANALYZING THE SPATIOTEMPORAL DISTRIBUTION OF DIFFERENT INDUSTRIES IN WUHAN CITY USING ENTERPRISE REGISTRATION DATA

Zousen Cui [a], Gang Xie [a], Zhipeng Gui [a,b,c, *], Huayi Wu [b,c]

[a] School of Remote Sensing and information Engineering, Wuhan University, Wuhan, China - billcui4511@gmail.com, mesryxiegang@sina.cn, zhipeng.gui@whu.edu.cn
[b] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China - (zhipeng.gui, wuhuayi)@whu.edu.cn
[c] Collaborative Innovation Center of Geospatial Technology, Wuhan, China - (zhipeng.gui, wuhuayi)@whu.edu.cn

Commission V, WG V/4

KEY WORDS: Spatial Autocorrelation Analysis, Enterprise Registration Data, Spatiotemporal Distribution, Spatial Agglomeration of Industry

ABSTRACT:

Enterprises are the basis of urban economic development and an essential factor that affects urban structure. It is of great significance to study the spatiotemporal distribution of enterprises to benefit cities planning and development. Based on the enterprises registration data of Administrator for Industry and Commerce of Wuhan from 1996 to 2007, in this paper, we analysed the spatiotemporal distribution of enterprises in Wuhan city. We divided Wuhan into 2,356 square grids with a length of 2 km and counted the number of enterprises of different industry categories in each grid. Then we calculated the aggregation intensity index of enterprises in each grid. With the method of global spatial autocorrelation analysis, we discovered the change of the degree of enterprise agglomeration for different industry categories in different periods. We also used local autocorrelation analysis to find the hotspots of enterprise distribution areas furtherly and discover the changes of the location of the hotspots during each period. Based on the two analyses and context information of Wuhan city, we can know the features of spatiotemporal distribution of enterprises. It can be concluded that the spatial distribution of enterprises could be reflected directly by spatial autocorrelation analysis. Moreover, the enterprises registration data with fine-grained enterprises spatiotemporal information can support the research of spatial distribution and evolution of enterprises. This study can show how different industry categories developed and provide reference to urban geography to help cities plan the distribution of enterprises and help enterprises decide its location.

1. INTRODUCTION

Enterprise, the basis of urban economic development, is an important area-organising institution. As a branch of economic geography, enterprise geography focuses on the spatial behaviour and spatial structure of enterprises. Urban geography also shows that the spatial distribution of enterprises of different industries can make huge difference on the usage of urban land and layout of function. Therefore, it is important to study the spatiotemporal distribution of enterprises to benefit cities planning and development. From the existing studies of the enterprises' spatial distribution of domestic cities, we found that many researches have been conducted on the first-tier cities of China, e.g., Shanghai and Shenzhen. Using the traditional regional density model and geo-statistics methods based on the digital city data of 2012, Zhang (2015) found that Shenzhen enterprises distribution basically showed strong step-down trend from the city center to the periphery. Besides, using the spatial GINI coefficient and the Kriging interpolation methods, Chen (2015) found that it has been formed industry ecological circle led by the financial industry and there is structural difference in the agglomeration capacity of enterprises among the different kinds of commercial buildings in Xiaolujiazui district, Shanghai. These two researches want to fit the distribution of enterprises by a geography model to get a current circumstance. But urban planners may care more about the tendency of development when they want to make planning and try to improve the industrial structure. Spatial

autocorrelation analysis can show the correlation within variables across georeferenced space. Using spatial autocorrelation analysis, Cao (2011) found that generally the logistics enterprise distribution has been in a state of agglomeration in recent twenty years, but as time goes by, the decline of BPI (Balanced-Polarization Index) marks a weakening tendency. It shows that we can know how enterprise distribution change along with time by comparing the distribution pattern during each period.

Wuhan locates in the central region of China. It is the transportation intersection of east-west and north-south traffic lines. Because of that Wuhan plays an important part in China's H-shaped economic development pattern in the coastal areas, the Yangtze River Economic Belt and the western region have taken shape（Government of Wuhan, 2012a）. Wuhan is also an important industrial base. It has a solid foundation in both hi-tech manufacturing industry and traditional manufacturing industry. In recent years, Wuhan has established a new economic growth belt featuring the integration of hi-tech and traditional manufacturing industries, including over 30,000 enterprises such as iron and steel, automobile, equipment manufacturing, petroleum and petrochemical, optoelectronic communication, western and traditional Chinese medicine, biological engineering, textile, clothes and food（Government of Wuhan, 2012b）. Different industries show distinct spatial features when the enterprises are increasing. By analysing the spatiotemporal distribution of different industries in Wuhan City, we can reveal

---

* Corresponding author

and utilize these spatiotemporal patterns to help government optimize the industrial layout and make urban planning.

2. DATA COLLECTION AND PREPROCESS

2.1 Data collection

The original data is collected by local Administration for Industry and Commerce and produced after processing by a workflow of data imputation based on machine learning methods and geocoding tools that completes the missing values of industry level and spatial location (Li et al., 2017). Each record in this dataset includes multiple attribute fields, such as enterprise name, address, industrial category code, registered capital and registration time. The industrial category is coded by the rules of Classification of National Economy (GB/T 4754-2011). We used the registration time, industrial category code and address to conduct spatiotemporal analysis. After data cleaning, we got 8,159,640 enterprise records of year 1996 to 2007 which have relatively complete attribute fields.

With a view to facilitate the research, we choose the example of three industrial categories which are the closest to daily life and urban development. They are health and sports facilities and social welfare industry (Q of GB/T 4754-2011), scientific research and comprehensive technical services industry (M of GB/T 4754-2011) and wholesale and retail trade and catering industry (F&H of GB/T 4754-2011).

2.2 Data divided by period

As spatiotemporal distribution need to be analysed along with time, data must be divided into different periods and seek how distribution changes with time goes by. In consideration of the data size and industry development process, the 12-years period is split into four intervals of 3 years each. As a result, the number entries of the three industrial categories in each period as Table 1, and the spatial distributions of data are illustrated Figure 1.

Category	Time period			
	1996-1998	1999-2001	2002-2004	2005-2007
HSSW	112428	400809	336958	154051
SRCT	30121	96618	104622	92427
WRC	396577	1230666	791303	359622

Table 1. Enterprise number entries of each 3-years-period between year 1996 to 2007(where HSSW denotes health and sports facilities and social welfare; SRCT denotes scientific research and comprehensive technical services; and WRC denotes wholesale and retail trade and catering)
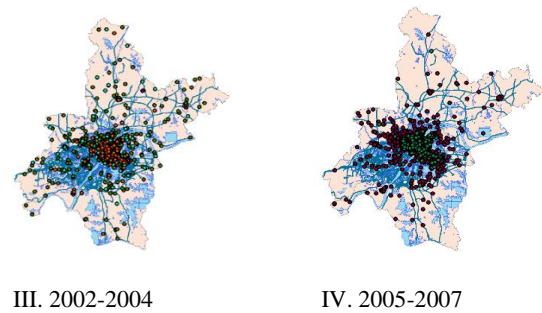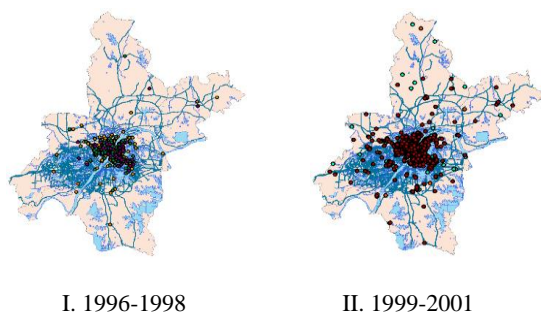

I. 1996-1998 II. 1999-2001

III. 2002-2004 IV. 2005-2007

Figure 1. Spatial distribution of enterprises registered in Wuhan among different time periods

2.3 Dividing grids

In order to analyse spatial interaction of a set of objects in an area, we need to divide the area into grids. Based on the dividing rule proposed in previous research (Cao, 2007; Li, 2008; Li, 2010), we divide Wuhan into 2,356 square grids with a length of 2 kilometers. The grids are shown in Figure 2.
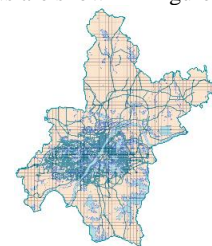

Figure 2. The divided square grids of Wuhan City, China

2.4 Calculating the aggregation intensity index

The aggregation intensity index (AII) is a normalized increasing rate of enterprise number in each grid of each period. It can be defined by the equation (1) (Cao, 2011):

$$AII_i = \frac{n_i}{N_j \times \Delta t} \times k \qquad (1)$$

Where AII_i is the aggregation intensity index of grid i

 n_i is the number of enterprises in grid i of period j

 N_j is the total number of enterprises of period j

 Δt is the time span of period j

 k is a constant

By using aggregation intensity index, we can eliminate the incomparability because of enterprises' unequal distribution over time.

3. METHODOLOGY

Spatial autocorrelation can shows the correlation within variables across georeferenced space. Given a set geographical units, spatial autocorrelation refers to the relationship between some variable observed in each of the geographical units and a measure of geographical proximity defined for all pairs chosen from all geographical units (Hubert et al., 1981).

Spatial autocorrelation or spatial dependence can be defined as a particular relationship between the spatial proximity among observational units and the numeric similarity among their values.

Spatial autocorrelation analysis is based on the spatial weight. By measuring the similarity and difference of enterprises distribution intensity at the level of neighbouring, spatial autocorrelation analysis can reveal the layout of enterprise distribution and its structural characteristics. In our research, we regard the square cells which are divided in quadrant analysis as evaluation units and the enterprise aggregation intensity index (AII) as variables. According to the measurement of Getis-Ord General G (Getis, 1988) and Getis-Ord Gi* (Getis, 1992), we can know the global and sectional spatial characteristics of enterprises distribution. The former is used to detect the correlation structure of the whole study areas, the latter is used to identify the spatial distribution of high-value clusters and low-value clusters, hot spots and cold spots in different spatial locations. Furthermore, we can confirm the general trend and characteristics of enterprise spatial distribution.

3.1 Global autocorrelation analysis

Global autocorrelation analysis is a kind of tool measures spatial autocorrelation based on both feature locations and feature values simultaneously. Given a set of features and an associated attribute, it evaluates whether the pattern expressed is clustered, dispersed, or random. There are two most commonly used indices for global spatial autocorrelation analysis, Moran's Index and Getis-Ord General G. Zhang (2007) has found that Getis-Ord general G is more sensitive to high clusters. To analyzing the spatial agglomeration of enterprises, we chose the value of Getis-Ord General G as the judgement of the data in this research.

The General G statistic of overall spatial association is given as formula 2:

$$G = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,j} x_i x_j}{\sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j}, \ \forall j \neq i \tag{2}$$

Where: x_i and x_j are the attribute values for features i and j
$w_{i,j}$ is the spatial weight between feature i and j.

The Z_G-score for the statics is computed as formula 3:

$$Z_G = \frac{G - E(G)}{\sqrt{V(G)}} \tag{3}$$

Where:

$$E(G) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,j}}{n(n-1)}, \forall j \neq i \tag{4}$$

$$V(G) = E(G^2) - E(G)^2 \tag{5}$$

3.2 Local autocorrelation analysis

In the 1990s, the idea of spatial autocorrelation was extended to local conditions. Getis and Ord showed how, by a relatively simple variation on a basic autocorrelation statistic that they called G, one could focus on the possible spatial association of designated observations to a single observation i. They developed a local statistic called Gi and another called Gi*. The first considers the ith observation but does not include it in its calculations, while the second includes the ith observation in the analysis (Getis, 1992). Getis-Ord Gi* statistic can identify statistically significant spatial clusters of high values (hot spots) and low values (cold spots). The local statistics help to distinguish the more general statistics like General G.

The Getis-Ord statistic is given as:

$$G_i^* = \frac{\sum_{j=1}^{n} w_{i,j} x_j - \bar{X} \sum_{j=1}^{n} w_{i,j}}{S\sqrt{\frac{n\sum_{j=1}^{n} w_{i,j}^2 - \left(\sum_{j=1}^{n} w_{i,j}\right)^2}{n-1}}}, \forall j \neq i \tag{6}$$

Where: x_j is the attribute values for features j
$w_{i,j}$ is the spatial weight between feature i and j
n is equal to the total number of features

$$\bar{X} = \frac{\sum_{j=1}^{n} x_j}{n} \tag{7}$$

$$S = \sqrt{\frac{\sum_{j=1}^{n} x_j^2}{n} - (\bar{X})^2} \tag{8}$$

4. DATA ANALYSING

From the enterprises data, we chose three main industries to analyse how they developed in the period of year 1996-2007. The three industries are health and sports facilities and social welfare (HSSW), scientific research and comprehensive technical services (SRCT), and wholesale and retail trade and catering industry (WRC).

4.1 Global autocorrelation analysis

For each industry in each time period, we measures the degree of clustering using the Getis-Ord General G statistic based on ArcGIS Desktop 10.5. The Z_G-score of Getis-Ord General G statistic is shown in Table 2.

Category	Time period			
	1996-1998	1999-2001	2002-2004	2005-2007
HSSW	66.08842	61.62084	68.72985	72.12713
SRCT	42.62105	59.79373	65.53019	67.85271
WRC	82.51389	88.39436	80.71793	83.43746

Table 2. Z_G-score of General G statistic (where HSSW denotes health and sports facilities and social welfare; SRCT denotes scientific research and comprehensive technical services; and WRC denotes wholesale and retail trade and catering)

In the results, all the confidence level is 99 percent, indicating that the pattern was created by random chance is really small (less than a 1 percent probability). A positive Z_G-scores indicates clustering of high values. For the positive value, the higher the Z_G-score, the stronger the intensity of the clustering. A Z_G-score near zero indicates no apparent clustering within the study area.

Table 2 show that, for all the three industrial categories, registered enterprises in each time period shows strong intensity of clustering. For health and sports facilities and social welfare industry, the clusting value doesn't grows conspicuously. As for scientific research and comprehensive technical services industry, the intensity of clustering became higher and higher as time goes by. For the wholesale and retail trade and catering industry, whose intensity of clustering was highest among three industrial categories, did not show an increasing trend but fluctuated.
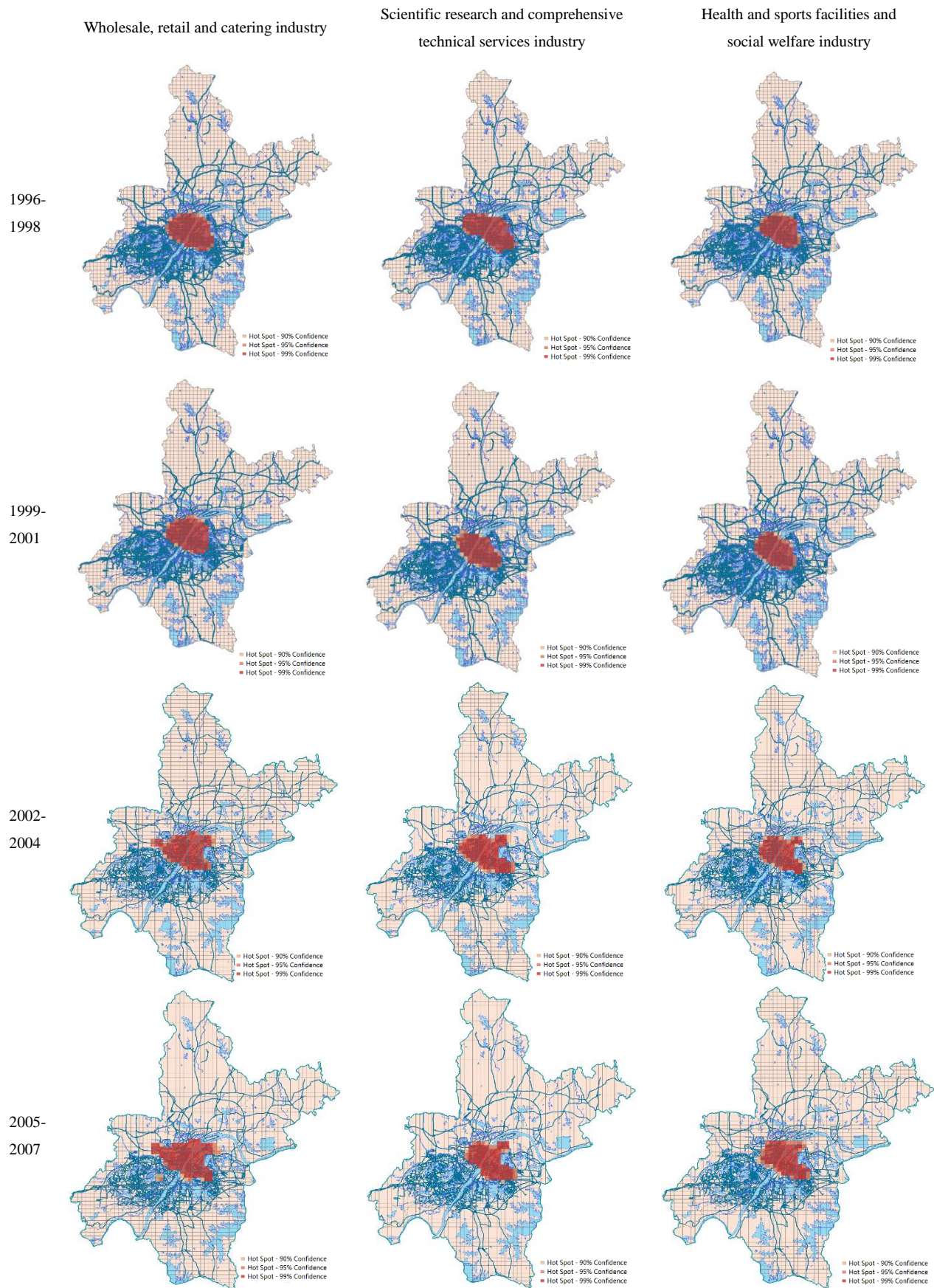
Figure 3. Spatial distribution of Industry "Hotspots" in different time periods

We can find out that wholesale and retail trade and catering industry shows more obvious industrial agglomeration, because enterprises of this industry rely on customer flows. As a result, they densely locate near commercial districts and cause a high clusting. However, health and sports facilities and social welfare industry is related to residential areas which continue to expand from central city to outskirts. So, the clusting value of it grows unobviously. For the scientific research and comprehensive technical services industry, it mostly determined by the government policy. With the growth of Wuhan East Lake High-Tech Development Zone, it has been developing well near the zoon and shows increasing agglomeration.

4.2 Local autocorrelation analysis

Using the Getis-Ord Gi* statistic based on ArcGIS Desktop 10.5, we can identify statistically significant hot spots. The hotspot grids for the different periods can be seen in figure 3.

We can learn from figure3 that the hotpots of the three industries all located in the central part of the city and linked several major commercial areas, such as Jianghan Road, Jiedaokou area, Hongshan Square, Tan Hualin, the Optical Valley area and so on. During the period of 1996-2007, the hotpots have all expended but show different characteristic with variation in industries and time periods.

Comparing different industry categories, the development processes of different industrial enterprises are not the same. The wholesale, retail and catering industry expanded from the center of the Jiang Han District and the center of the Wuchang District to nearly main road and subway lines. There are also some independent gathering points in the surrounding area in period 2005-2007. Health and sports facilities and social welfare industry gradually expanded from the affluent city center to the surrounding areas, especially the Optical Valley area.

As shown in the second column of Figure 3, scientific research and comprehensive technical service industry gradually expanded from the central Jiang Han District to the Optical Valley area along the Wu Luo Road and Luo Yu Road. The gathering center moved from Jiang Han District to Optical Valley area. Because of that, the aggregation degree firstly declined but then increased. There are also some enterprises near the universities. The enterprises have formed a pattern concentrated in the Optical Valley area and axially distributed around the main road.

Comparing different time periods, the hotpots of three industries did not change a lot in year 1996-2001. Nevertheless, the hotpots showed that the Qingshan district and the Optical Valley area have been greatly developed in year 2002-2007. These two areas became new centers for business and technology.

5. CONCLUSION

In order to analyse the spatiotemporal distribution of enterprises in Wuhan city, this paper take use of the spatial autocorrelation analysis to find out the distribution feature of different time period. Using enterprises registration data of Wuhan Administrator for Industry and Commerce from 1996 to 2007, we analysed the spatiotemporal distribution of three industry categories in Wuhan city, by taking health and sports facilities

and social welfare industry (Q of GB/T 4754-2011), scientific research and comprehensive technical services industry (M of GB/T 4754-2011) and wholesale and retail trade and catering industry (F&H of GB/T 4754-2011) as example. From the result of global autocorrelation analysis and local autocorrelation analysis, we can know how these industrial categories developed from 1996 to 2007.

Based on the discoveries, we can come to the conclusion that enterprises registration data with fine-grained enterprises spatiotemporal information is valuable for the research of spatial distribution and its evolution of enterprises. The development of spatial distribution of enterprises could be detected and analysed by using spatial autocorrelation analysis methods. This study can provide reference to urban geography by helping us find out how urban enterprise developed. It may also helpful in urban planning and location selection of business site selection.

REFERENCES

Cao W., 2011, Spatial pattern and location evolution of urban logistics enterprises: Taking Suzhou as an example[J]. *Geographical Research*, (11), pp. 1997-2007.

Chen Y., Chen L., 2015, Spatial distribution of enterprises in Xiao Lujiazui, Shanghai. *Urban Problems*, 2015, (02), pp. 87-94.

Getis A, Ord J K., 1992, The Analysis of Spatial Association by Use of Distance Statistics[J]. *Geographical Analysis*, 24(3), pp. 189–206.

Getis A., 1988, Spatial Autocorrelation: A Primer[J]. *Economic Geography*, 64(1), pp. 88.

Getis A., 2008, A History of the Concept of Spatial Autocorrelation: A Geographer's Perspective[J]. *Geographical Analysis*, 40(3), pp. 297–309.

Government of Wuhan, 2012a, Wuhan Overview, http://english.wh.gov.cn/whgk_3581/zrdl_3582/ (23 May 2017)

Government of Wuhan, 2012b, Wuhan's Advantages, http://english.wh.gov.cn/whgk_3581/whys/ (23 May 2017)

Hayter R, Watts H D., 1983. The Geography of Enterprise: A Reappraisal[J]. *Progress in Human Geography*, 7(2), pp. 157-181.

Hubert L J, Golledge R G, Costanzo C M., 1981, Generalized Procedures for Evaluating Spatial Autocorrelation[J]. *Geographical Analysis*, 13(3), pp. 224-233.

Khan G, Xiao Q, Noyce D A., 2014, Spatial analysis of weather crash patterns in Wisconsin[J]. *American Society of Civil Engineers*, 134(5), pp. 191-202.

Li, F., Gui, Z., Wu, H., Gong, J., Wang, Y., Tian, S., Zhang, J., 2017. Big enterprise registration data imputation: Supporting

spatiotemporal analysis of industries in China. Computers, Environment and Urban Systems.

Li L., Zheng X., Xiang W., 2008, A study on Site Coverage Spatial Distribution Pattern of Beijing City Based on GIS, *China Population Resources and Environment*, (01), pp. 122-127.

Li R., Guo F., Zhang J., 2010, Spatial pattern of health clubs in Beijing at various scale, *Progress in geography*, (02), pp. 232-240.

Zhang S., Zhang K., 2007, Comparision between General Moran's index and Getis-Ord General G of Spatial Autocorrelation, *Acta Scientiarum Naturalium Universitatis Sunyatseni*, 46(4), pp. 93-97

Zhang W., Zhang L., Wang J., 2015, Study on spatial distribution pattern of Shenzhen enterprise, *Sicence of Surveying and Mapping*, (09), pp. 65-68