Analysis of spatial autocorrelation for point objects based on line buffer

Jiangping Chen
School of remote sensing Information Engineering
Wuhan University
Wuhan Hubei, PR China
chenjp_lisa@163.com

Jingchao Jiang
School of Remote sensing Information Engineering
Wuhan University
Wuhan Hubei, PR China
77068177@qq.com

Abstract—**In the light of Tobler's first law: everything is related to everything else, but near things are more related than distant things. This kind of relations called spatial autocorrelation. Spatial autocorrelation occurs when values of a variable sampled at nearby locations are more similar than those sampled at locations more distant from each other. Moran's I maybe one of the oldest and best indices for spatial autocorrelation. The ways of weight matrix definition will affect the value of Moran's I greatly. Since some weight matrix cannot differentiate the strength of spatial linkages between adjacent locations, some more complex spatial weight matrices are proposed for more precise spatial linkages. There are many ways to define a spatial weight matrix. For example: Rook's weight matrix, Queen's matrix, Binary connectivity matrix, K nearest matrix and distance threshold matrix etc.**

For point objects in the real world, the distance between the point objects is not the only feature affect their autocorrelation which were more affected by their accessibility. For example the autocorrelation of two cities, it's more affected by roads and the accessibility of the two cities than the distance between. It's naturally to consider the spatial autocorrelation for point objects based on the line between them.

One new form of spatial weight matrix for point objects based on line buffer is provided in this paper. There are 3 steps in the paper. First, the line buffer range of the point objects according to the theme knowledge of the research fields is set. Second, the relationship of two point objects is defined by line buffer. If the point objects are both within the line buffer then they are spatial related and will be measured by a formula which offered in the paper otherwise their spatial autocorrelation is considered zero. Third, the method is applied in Moran's I.

The experiment on spatial autocorrelation of the prices of buildings were applied in the new method, it shows that the method is effective.

Keywords: Spatial autocorrelation;Spatial weight matrix; Moran's I; Point objects; Line buffer

I. INTRODUCTION

In spatial data, it is often the case that some or all outcome measures exhibit spatial autocorrelation (Cliff and Ord, 1973). This occurs when the relative outcomes of two points is related to their distance(Cliff and Ord 1981). When analyzing spatial data, it is important to check for autocorrelation(Jingfeng Wang,2006). In the light of the first law Geography, put forward by Tobler, the concept of spatial autocorrelation is proposed. The degree of relationship that exists between two or more spatial data variables (e.g. amount of organic matter in soil, gradient, suitability for agriculture) and it is described as similarity between a spatial entry and the around spatial entries(Cliff and Ord, 1973). As a hotspot of spatial statistics, spatial autocorrelation has been utilized and achieved good results in many fields, such as in biology science, in regional economy, in urban development, in epidemiologic research, in real estate field(Chen Fei, 2002; He TiaraHua, 1999; Overmans 2003;Yong Tu, 2007). Classic spatial autocorrelation statistics include Moran's I and Geary's C. Moran's I maybe one of the oldest and best indices for spatial autocorrelation for its excellent quality(Jingfeng Wang,2006). Moran's I is consisted of spatial weight matrix and attributes of spatial data, so how to define appropriate spatial weight matrix is always vital in the research of spatial autocorrelation(Cliff and Ord, 1973). The spatial weight matrix depicts the relation between a geographical cell and its surrounding geographical cell(Cliff and Ord 1981). Since binary contiguity weight is proposed by Moran in 1948, a variety of methods have been used to imply the spatial weight matrix, such as based Cliff-Ord weight, Rook, Queen, K-nearest point and so on(Moran, 1948; Cliff and Ord 1981). In these approaches, the definition of spatial weight matrix is built upon either contiguity relations or distance between cells. These spatial weight matrices just take into the Euclidean distances between two objects into account, and another problem of using conventional weight matrix is that these definitions ignore the influence of accessibility between two point objects. In this paper, it will try to provide a method about line cells' buffer of the point objects to solve the problem. Distance and topology may be affected by line cells' buffer in some fields. This paper propose a new method which is based on the line buffer and can obey the First Law of Geography well, then apply it to depict spatial weight relations between real estates as the case study in Wuchang District of Wuhan City. The objective of this paper is to inspect whether the method based on line buffer is better to reflect spatial relationship than conventional methods. Section 2 presents the methodology and study area. Section 3 reports the results and compares the method defined in this paper with the conventional methods used in Wuchang

National Natural Science Foundation of China--- Youth Science Fund Project(No. 40801152)

District. Section 4 will respectively summarize the main findings and conclusions..

II. Methodology And Study Area

A. Methods

The spatial weight matrix can be based, for example, on contiguity relations or distance, and K-nearest point for point objects' autocorrelation (Basu,1998). One new form of spatial weight matrix for point objects based on line buffer is provided in this paper. There are three steps in the paper. First, the line buffer range of the point objects according to the theme knowledge of the research fields is set. Second, the spatial relationship of two point objects is defined based on line buffer. The Fig1 shows this spatial relationship, so if the point objects are both within the line buffer then they are spatial related and will be measured by a formula (F) which offered in the paper otherwise their spatial autocorrelation is considered zero. Formula (F) is defined as:

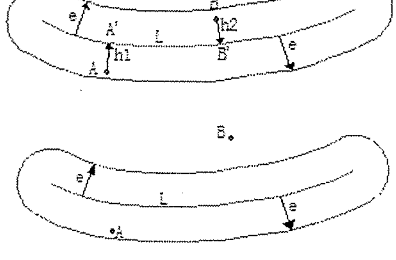
$$w(A,B;L) = a\times\frac{1}{l}\times\left(\frac{e-h_1}{e}\right)\times\left(\frac{e-h_2}{e}\right),\quad (h_1 < e, h_2 < e)\ .\ \text{(F)}$$

l is the length between the point A' and B', A' is the point of A projected on line L, B' is the point of B projected on line L , e denotes the line buffer's width range, h_1, h_2 denotes the distance of point object A,B projected on line L.

Third, the method is applied in Moran's I, the definition of the Global Moran's I is as below:

$$I = \frac{n\bullet\sum_i^n\sum_j^n w_{ij}\bullet(y_i-\bar{y})(y_j-\bar{y})}{(\sum_i^n\sum_j^n w_{ij})\bullet\sum_i^n(y_i-\bar{y})^2}$$

In this formula, n is the number of sample size, y_i and y_j are the values of the observed variable at the site i and j. y is the average of the observed variable at n sites. The values of w_{ij} are the weights. The weights w_{ij} are written in a $n*n$ spatial weight matrix W. Negative (positive) values indicate negative (positive) spatial autocorrelation. Values range from -1(indicating perfect dispersion) to +1(perfect correlation) (Cliff and Ord, 1973). A zero value indicates a random spatial pattern. In this paper, after the spatial weight matrices are computed, we apply them into global Moran's I to test which kind of spatial weight matrix is best. As a contrast, the K nearest matrix, distance threshold matrix and voroni matrix are used in this paper to compare different Moran's I values.

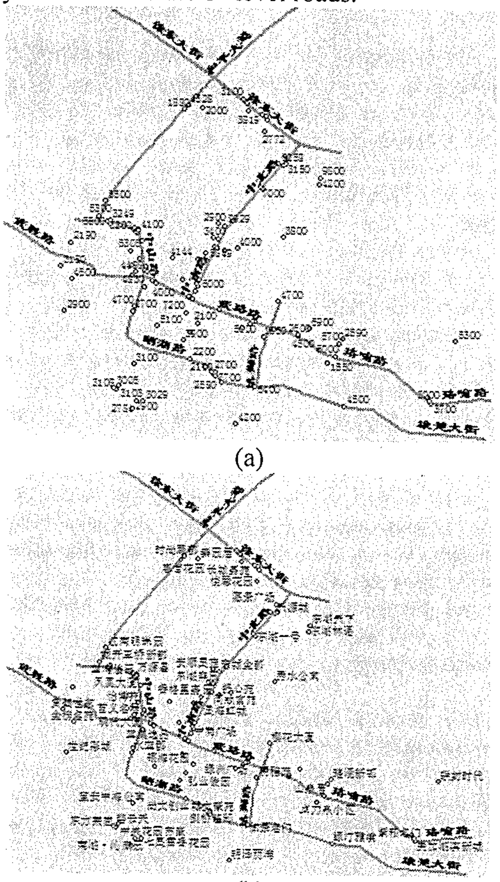


(a) (b)

Figure 1. Two discrete points within buffer of line (a) and one discrete point outside buffer of line (b)

B. Study Area and data

In this paper, the spatial autocorrelation of the real estate prices of Wuchang district is computed to verify the superiority of this new weight matrix. Wuchang district, which is part of Wuhan city in Hubei province of China, locates on the south of the Yangtze River Basin. The data includes the roads of 1st level (line objects) and real estates (point objects), which are showed in Fig.2. There are fourteen roads of 1st level. Each road of 1st level generates a line buffer. There are 121 different sites' real estate prices included in our study dataset. Most real estate sites are distributed along the 1st level roads. The real estates are spatially in relation to the 1st level roads.



(a)

(b)

Figure 2. Distribution map of roads and real estate prices (a) and Distribution map of roads and real estate names (b)

III. Results

A. Results of Spatial Autocorrelation

The global spatial autocorrelation indices of four different spatial weights are calculated by Moran's I. For the K nearest spatial weight matrix, K is 4, because most real estate sites are dispersed, so the K cannot be set bigger (Liu Jisheng, 2007). For distance threshold weight matrix, the threshold is 2.5km, the threshold is determined by the road's significance of importance, more important the road is, bigger the threshold is (Yong Tu, 2007). For the line buffer weight matrix, in the paper the buffer range is 1.0km, the range is also depended on the road's significance, more important the road is, bigger the range is. For the voronoi matrix, Queen adjacency is used to get weight matrix. After the spatial weight matrix is computed, applying them in the Moran's I, the global spatial autocorrelation indices are calculated. Table 1 shows global spatial autocorrelation indices and Z-statistics of four spatial weights matrix for all real estates. All the global spatial autocorrelation indices are positive; the overall condition of Wuchang real estate prices is clustered. If the value of Z-statistic is more than 1.96 or less than -1.96 at the significance level p=0.05, then it is regarded to be of high significance under the hypothesis of stochastic (Cliff and Ord, 1973). It is mentioned that real estate prices has positive spatial autocorrelation and the significance do not distribute stochastically, but cluster together.

TABLE I. (A) SPATIAL WEIGHT MATRIX BASED ON K- NEAREST (PART)

ID	10	11	12	13	14	15	16
10	0	0.50	0.25	0.20	0.86	0.20	0.50
11	0.50	0	0.28	0.12	0.42	0.16	0.39
12	0.25	0.28	0	0.50	0.45	0.09	0.12
13	0.20	0.12	0.50	0	0.77	0.55	0.39
14	0.86	0.42	0.45	0.77	0	0.10	0.30
15	0.20	0.16	0.09	0.55	0.10	0	0.25
16	0.50	0.39	0.12	0.39	0.30	0.25	0

(B) SPATIAL WEIGHT MATRIX BASED ON DISTANCE THRESHOLD (PART)

ID	10	11	12	13	14	15	16
10	0	1	0	0	0	1	1
11	1	0	0	0	0	0	0
12	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0
15	1	0	0	0	0	0	0
16	1	0	0	0	0	0	0

(C) SPATIAL WEIGHT MATRIX BASED ON VORONOI (PART)

ID	10	11	12	13	14	15	16
10	0	0	1	0	0	0	1
11	0	0	0	0	0	0	0
12	1	0	0	0	0	0	0
13	0	0	0	0	0	0	1
14	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0
16	1	0	0	1	0	0	0

(D) SPATIAL WEIGHT MATRIX BASED ON LINE BUFFER (PART)

ID	10	11	12	13	14	15	16
10	0	1.80	2.63	0.55	0	2.50	3.70
11	1.80	0	0	0	0	0	1.80
12	2.63	0	0	0	0	0.20	0
13	0.55	0	0	0	0	0	4.70
14	0	0	0	0	0	0	0
15	2.50	0	0.20	0	0	0	0
16	3.70	1.80	0	4.70	0	0	0

B. *Comparison Analysis*

While calculating Moran's I using the matrix based on line buffer and other weights matrix respectively, different Moran's I value can be get. The global Moran's I of K-nearest point weight and voronoi weight is bigger, and the global Moran's I of threshold distance weight is smaller. So the global Moran's I is influenced by spatial weight matrix. And all Z-statistics are more than 1.96 at the significance level p=0.05, the global Moran's I of four spatial weights are all positive. It can be seen from Table.1 that the global Moran's I which is calculated by spatial weight matrices defined in the way of spatial adjacency (such as voroni weight and K-nearest point weight) is bigger, and the global Moran's I which is calculated by spatial weight matrices defined in the way of spatial distance (such as Threshold distance weight and Line buffer weight) is smaller. The reasons for this phenomenon are to be discussed and analyzed in the next section.

TABLE II. GLOBAL SPATIAL AUTOCORRELATION INDEX AND Z-STATISTIC

Type	Global Moran's I	Z-statistics
K-nearest point weight	0.23	5.2
Threshold distance weight	0.11	4.5
Voronoi weight	0.15	5.4
Line buffer weight	0.09	4.8

IV. DISCUSSION AND CONCLUSION

From the experimental results, a conclusion that the spatial autocorrelation indices based on different spatial weight matrices are unequal can be made, but the value of Z-statistics of these indices are all bigger than 1.96, so they all can be accepted. That means the real estate prices of Wuchang district and the 1st level road are spatial autocorrelated.

K-nearest point weight's matrix and threshold distance weight matrix are two very good weight matrix for the variation. Since the moran's I computed by them are very small. So they are selected for comparing. The voronoi weight matrix is Queen in essentially, which only considers spatial contiguity between two spatial objects. The voronoi weight matrix ignores the influence of distance, and it is based on the assumption that two spatial objects are relational if they are adjacent, though they may be far away from each other. So the global spatial autocorrelation index based on these spatial weights is big. Line buffer weight matrix take the distance and adjacent and accessibility into account, it reflects actual situations well. Line buffer weight matrix is based on the assumption that if two point objects are more adjacent on the 1st level road, the autocorrelation between each other is bigger; if the two point objects are located in more than one 1st roads' buffer, the autocorrelation is bigger. The weight of two point objects can be superimposed, which other spatial weight matrix doesn't take it into account.

The reason of the spatial autocorrelation of the first level road and the price of real estate is very small in the experiments maybe that the there are a lot of other factors effect the price of the real estate (Liu Jisheng, 2007; Sabyschi,1998;Yong, 2007). Constructing contiguous spatial weight matrix (W) is one of the most important task within spatial autocorrelation analysis. The spatial autocorrelation of real estate sites is influenced by many geography factors such as main roads, parks, schools, rivers terrains and so on (Liu Jisheng, 2007). The influence of parks and schools and other features will be take into accounted in subsequent studies.

In summary, the findings may need to be further verified by more fields, although during this work several other spatial weight matrix were tested and consistent results were obtained. Future research needs to focus on how to build the spatial weight matrix of topographic relationship and direction into it such as the 1st level road near the main commerce area on the effect of topographic autocorrelation is more important.

REFERENCES

Chen Fei, and Du Daosheng (2002). Application of the Integration of Spatial Statistical Analysis with GIS to the Analysis of Regional Economy. *Geomatics and Information Science of Wuhan University.* 27 (4), 391-396.

Cliff, A. D., and Ord, J. K. (1973) . *Spatial autocorrelation.* London: Pion.

Cliff A. D., and Ord, J. K. (1981). *Spatial processes.* London: Pion.

He Tiara Hua , Yang Ji, and Rao Guang Yuan (1999). Spatial autocorrelation analysis of plant population genetic variation. *Chinese Bulletin of Botany.*, 16 (6), 636-641.

Wang, J. (2006). *Spatial analysis.*Beijing: Science Press.

Liu Jisheng, and Chen Yanguang (2007) spatial autocorrelation and localization of urban development. *Chinese Geographical Science.* China, 17(1), 34-39.

Moran, P. (1948). The interpretation of statistical maps. *Journal of the Royal Statistical Society* B, 10, 243-51.

Overmans, K. P., De Koning, C. H. J., and Veldkamp, A. (2003). Spatial autocorrelation in multi-scale land use models. *Ecological Modelling.* 164, 257-270.

Tiefelsdorf, M. (2000). *Modeling spatial processes: the identification and analysis of spatial relationships in regression residuals by means of Moran's i.* Berlin: Springer.

Tobler, W. R. (1973). A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, 46(2),

Basu, S. and Thibodeau, T.G. (1998). Analysis of Spatial Autocorrelation in House Prices. *Journal of Real Estate Finance and Economics*, 17 (1), 61-85.

Yong Tu, Hua Sun, and Shi-Ming Yu (2007). Spatial Autocorrelations and Urban Housing Market Segmentation. *J Real Estate Finan Econ.* 34, 385–406.