Application of Integration of Spatial Statistical Analysis with GIS to Regional Economic Analysis

CHEN Fei DU Daosheng

ABSTRACT This paper summarizes a few spatial statistical analysis methods for measuring spatial autocorrelation and spatial association, discusses the criteria for the identification of spatial association by the use of global Moran Coefficient, Local Moran and Local Geary. Furthermore, a user-friendly statistical module, combining spatial statistical analysis methods with GIS visual techniques, is developed in Arcview using Avenue. An example is also given to show the usefulness of this module in identifying and quantifying the underlying spatial association patterns between economic units.

KEYWORDS spatial statistical analysis; spatial autocorrelation; spatial association; regional economic analysis

CLC NUMBER P208

Introduction

Nowadays, spatial statistics can be found in the fields of agriculture, geology, soils, water, the environment, economy and geography and so on. Many researchers have conducted comprehensive researches on spatial statistics in the last two decades or so [1-5], and new statistical approaches have been developed.

GIS technology, as an interactive visualization and decision-support tool, plays an important role in regional economic development plans as well as traditional analytical systems or approaches [4], especially in the decision-making procedure of economic development at local, regional and state levels. However, almost all the current commercial GIS packages are extremely limited in standard statistical. Many GIS provide only some of the most basic summary statistics about data and do not support spatial statistical modeling required by many decision-makers, let alone spatial statistical capabilities. Whereas, spatial statistical analysis is very necessary and helpful for a researcher who wishes to examine relationship between different variables, and to make statistical decisions with geo-referenced data. True spatial analysis in GIS is a much longer-term goal. In fact, since the late of 1980s, the statistical aspect of spatial analysis has received increasing attention from the GIS community. Since spatial statistical analysis becomes more necessary in many GIS analysis, researchers pay more attention to the integration of spatial statistical analysis and GIS. Nowadays, though there are still a lot of discussions and disagreements on the integration of GIS and spatial statistical analysis, it is generally agreed that combining at least some spatial statistical analysis with GIS is necessary. Different researchers put forward different opinions, but most of them believe that the integration can occur in two totally different but equally valid solutions: embedding spatial statistical analysis function into a GIS environment, or embedding selected GIS functions into a spatial statistical analysis environment [7].

Recently, researchers make a beneficial exploration of the use of the integration of GIS and...
spatial statistical analysis. The application in regional developing analysis reflects mainly the use in the field of social-economic development. For the reason that there are still some technical difficulty on the integration of spatial statistical analysis with GIS and the spatial statistical analysis capabilities of GIS are deficient up to now, the application of that integration in regional economic analysis is still limited.

1 Spatial statistical analysis approaches

Spatial statistics is concerned with the application of spatial sampling in geographic situations. In general, a geographic phenomena or an attribute value observed at one area unit is not independent of the same phenomena or the same attribute values observed at adjacency area units. Almost all kinds of spatial data have the feature of spatial dependence or spatial auto-correlation. The existing spatial dependence violates the basic assumption of independence among the observations in classical statistical analysis. Most of the classical statistical methods, when applied to geo-referenced data, fail to capture the spatial dependence of the data generally. However, most of urban and regional analysis is conducted with discretized data aggregated for different geographical areas or zones. Therefore, a set of spatial statistical analysis methods should be identified and introduced in order to handle those data efficiently. The "spatial statistics" in this paper is a narrow definition. It doesn't means all the statistical methods for analyzing spatial data, but means those methods suitable for handling discretized data of different geographical areas or zones. Thus, the core of spatial statistics is the explicit recognition of spatial dependence or spatial autocorrelation among geo-referenced data, the construction of spatial weight matrix, the measurement and test of spatial autocorrelation or spatial association, the identification of spatial association and so on. Not all of the classical statistical techniques are thrown away in spatial statistics, justly most of them are modified so that they can be used properly for spatial data analysis.

1.1 Spatial weight matrix

The topological information generated by GIS provides the basic measure of spatial linkages or proximity for spatial data analysis. A binary spatial weight matrix \( W(n \times n) \) is usually defined to represent the spatial proximity relations, which can be measured with adjacency or distance criterion. Besides, a general measure of the weighted spatial proximity can be defined in terms of the attribute value \( x_i \) observed and the binary spatial weight matrix. According to the adjacency criterion, the elements \( W_{ij} \) of the spatial weight matrix will be one when location \( i \) is adjacent to location \( j \), and zero otherwise. Similarly, according to the distance criterion, the elements \( W_{ij} \) of the spatial weight matrix will be one when the distance between location \( i \) and location \( j \) within a given distance \( d \) and zero otherwise. For convention, all the diagonal elements \( W_{ii} \) are set to zeros.

1.2 Measurement of spatial autocorrelation and spatial association

Goodchild (1987) thought that in its most general sense spatial autocorrelation or spatial dependence concerns the degree to which objects or activities at some place are similar to other objects or activities located nearby. Spatial dependence can be measured on two different scales: global indicators and local indicators.

The Moran Coefficient (MC) and Geary Ratio (GR) are two well-known global indicators of spatial autocorrelation. MC reflects attribute-similarity among area units that are located nearby. Spatial dependence can be measured on two different scales: global indicators and local indicators.

The Moran Coefficient (MC) and Geary Ratio (GR) are two well-known global indicators of spatial autocorrelation. MC reflects attribute-similarity among area units that are located nearby. Spatial dependence can be measured on two different scales: global indicators and local indicators.
value to describe autocorrelation within a certain given area\(^2\), so the patterns of spatial association existing in different local map areas are difficult to be detected. \(G(d)\) statistics, Local Moran statistics and Local Geary statistics are alternative local indicators.

According to Reference [3], a spatial statistics \(G_i(d)\) can be defined as follows,

\[
G_i(d) = \left( \sum_{j \neq i} w_{ij} x_j \right) / \sum_{j \neq i} x_j
\]

(1)

where \(x\) denotes the observed attribute value at location \(j\), and the construction of symmetric spatial weight matrix \(W\) is based on the distance criterion. For ease of interpretation, \(Z(G_i)\), a standardized form of \(G_i(d)\), can be defined\(^4\).

According to Reference [5], Local Moran statistics and Local Geary statistics for each location \(i\) can be defined respectively as follows:

\[
n \sum W_{ij} Z_j
\]

\[
(2)
\]

\[
n \sum W_{ij} (Z_i - Z) \bar{j}
\]

\[
(3)
\]

where \(Z\) and \(Z_i\) are with deviations from the mean, such as \(Z = x - \bar{x}\), \(W_{ij}\) are the elements of binary symmetric spatial weight matrix, \(S^2 = \sum (x_j - \bar{x})^2/(n - 1)\) and \(j \neq i\), \(W_{ij} Z_j\) is the weighted average of the deviations in the surrounding locations. Unlike Local Moran statistics, Local Geary statistics is the weighted sum of the squared differences between the deviation at location \(i\) and surrounding locations.

### 1.3 Identification of spatial association

The inverse relationship between Moran’s \(I\) and Geary’s \(C\) is basically linear in nature\(^2\), with one index we can express the other. The MC is more popular and powerful statistically. Hence, we implement MC to measure the global spatial autocorrelation in this article. Under a normality assumption, the test statistics for MC is:

\[
Z(\bar{I}) = (\bar{I} - E(\bar{I})) / \sqrt{\text{Var}(\bar{I})}
\]

(4)

where \(E(\bar{I})\) and \(\text{Var}(\bar{I})\) denote the expected value and variance of MC respectively\(^1\).

On the basis of the calculated test statistics, the significant test on the null hypothesis \(H_0\) (there is no spatial autocorrelation between observed values over the \(n\) area units) can be conducted. \(MC = -1/(n - 1)\) or \(GR = 1\) indicates a random map pattern, \(MC > -1/(n - 1)\) or \(0 < GR < 1\) (MC or GR is significant too) indicates that similar values tend to cluster on a map (positive spatial autocorrelation), \(MC < -1/(n - 1)\) or \(GR > 1\) (MC or GR is significant too) indicates that dissimilar values tend to cluster on a map (negative spatial autocorrelation). When \(n\) is large, the expected value of MC converges to zero, and a positive value is associated with positive spatial autocorrelation, while a negative value is associated with negative spatial autocorrelation.

The standardized form of \(G_i(d)\) can be applied to either positive or negative attributes. A \(t\)-test can be conducted on the null hypothesis of \(H_0\): \(G_i = 0\). \(Z(G_i)\) does not include the observation \(i\) from the index. The \(G\) statistics can be used to identify spatial clustering patterns with high-values or low-values. However, the \(G\) statistics cannot detect spatial patterns of positive association or negative association.

Local Moran and local Geary statistics have some advantages over the \(G_i(d)\) statistic. For a randomization hypothesis, the test statistics for \(I_i\) is:

\[
Z(I_i) = (I_i - E[I_i]) / \sqrt{\text{Var}[I_i]}
\]

(5)

where \(E[I_i]\) and \(\text{Var}[I_i]\) denote the expected value and variance of \(I_i\) respectively\(^5\). On the basis of the calculated test statistics similarly, the significant testing on local spatial association can be conducted.

A pseudosignificance level of \(I\) can be obtained by a “conditional” randomization or permutation approach\(^6\). The experimental \(p\)-value also provides a basis for the test on the null hypothesis \(H_0\) (all observed values are randomly distributed over the space).

The interpretation of local Moran is similar to the \(G\) statistics. A small \(p\)-value (such as \(p < 0.05\)) indicates that observation \(i\) is associated with relatively high values in surrounding observations, while a large \(p\)-value (such as \(p > 0.95\)) indicates that observation \(i\) is associated with
relatively low values in surrounding observations.

The calculation of pseudosignificance level *p*-value of the local Geary is similar to that of Local Moran\(^{(92)}\). A large *p*-value (such as *p* > 0.95) indicates a small *C_i*, in extremes, which indirectly suggests a positive spatial association (+ + or − −) of observation *i* with its surrounding observations, while a small *p*-value (such as *p* < 0.05) indicates a large *C_i*, in extremes, which indirectly suggests a negative spatial association (− − or − +) of observation *i* with its surrounding observations.

2 Integration of spatial analysis statistical with GIS

The key feature of GIS is that it can link different geo-referenced data to geographic locations, that is to say, GIS represents and analyzes different data from a spatial perspective. Currently, researchers pay more attention to the integration of spatial statistical analysis and GIS, and make a beneficial exploration into the use of the integration in the field of social-economic development. The application in regional developing analysis reflects the use in the field of social-economic development mainly\(^{10}11\).

Different researchers put forward different opinions on the integration of GIS and spatial statistical analysis\(^{10}11\), and most of them believe that this integration can occur in two totally different but equally valid solutions: embedding spatial statistical analysis function into a GIS environment, or embedding selected GIS functions into a spatial statistical analysis environment. Attention has been focused almost exclusively on the former so far, while the latter has been largely ignored in the past\(^{11}\). However, almost any module developed can only offer limited functions of spatial statistical analysis.

With the research on regional economic analysis, embedding spatial statistical analysis approaches mentioned above into a GIS environment can meet the need of the analysis on underlying spatial association patterns between area units, and it is possible for a user to execute spatial statistical analysis and visualization analysis in the same environment. Consequently, we develop a user-friendly interactive spatial statistical analysis module with ArcView as developing environment, and realize the combination of spatial statistical technologies with a regional analysis procedure in a GIS environment. The module provides a flexible and convenient tool for the decision-making in regional economy. The following example illustrates its application in the field of regional economic analysis.

3 Example analysis

We take Xinjiang Uyger Autonomous Region as research area, and utilize mean annual GDP increasing velocity from 1978 to 1999 in different counties as an analytical indicator, then calculate global MC for research area and local MC for each county. Those spatial outliers can also be identified by the use of Local MC scatterplot. The results show the usefulness of the module in identifying and quantifying the underlying spatial association patterns among economic units.

To construct an adjacency spatial weight matrix is the first step in the procedure of spatial statistical analysis, in Fig. 1 for details. A two-dimensional matrix can be expressed as a one-dimensional array by use of the “List” class. In the module developed by authors, a spatial neighbor list table is used to represent spatially adjacent relations among different regional units\(^{11}\). Fig. 2 is one section of the spatial neighbor list created in practice.

On the basis of the spatial neighbor list created and the results calculated in Fig. 3, the significant testing on spatial autocorrelation can be implemented. If choose \(a = 0.05\), then \(Z_{0.025} = 1.96\). Because \(Z = 4.202 > 1.96\) in this example, so null hypothesis \(H_0\) is not true and should be refused. The test shows that there is a significant positive spatial autocorrelation among the values of analytical indicator of all counties.

Furthermore, we can investigate the underlying local patterns of spatial economic association
among counties in Xinjiang by calculating the local Moran statistics $I$, at county level. On the basis of the module developed and that criteria for the identification of Spatial Association mentioned above, we can do well for completing relevant calculations and analyses. See Table 1 (only those significant testing scores listed) and Fig. 4 for details. According to Local Moran Coefficient $I$, and its Testing Z-Score in Table 1, there are underlying positive or negative significant spatial association existing among different counties. For instance, there is a positive significant spatial association between the mean annual GDP increasing velocity of Korla City and those indexes of Korla’s surrounding counties, a negative significant spatial association between Shanshan and its surrounding counties. An examination on original data gives further help for explaining those underlying local spatial association patterns. According to those linked windows in Fig. 4, a local Moran scatterplot window and a table of local spatial statistics can be used to help identify spatial outliers and underlying patterns of local spatial clustering.

Table 1 Local Moran coefficient and its test statistics for different counties (one section)

| Counties      | ID | $I$       | Z-Score |
|---------------|----|-----------|---------|
| Hejing County | 41 | 17.16114  | 8.03215 |
| Shanshan County | 50  | -7.03272  | -3.25338 |
| Korla City    | 56 | 14.83759  | 6.94830 |
| Wei County    | 66 | 8.17691   | 3.84134 |
| Heshou County | 81 | 13.87901  | 6.50116 |
| Bohu County   | 82 | 15.73555  | 7.36716 |
| Moryu County  | 83 | 10.96643  | 5.14255 |
| Luopu County  | 84 | 7.61284   | 3.57823 |
| Pishan County | 86 | 6.78195   | 3.53973 |
| Hotan City    | 88 | 6.71225   | 3.50360 |

Notes: With an overall significance level of $a=0.05$, the individual significance level $a_{n}=0.00057$ based on a Bonferroni criteria.

Fig. 4 A sketch map of multi-window linkages for the results of spatial statistical analysis

4 Conclusions

In many regional studies, relatively independent economic areas comprise an important basis
for regional economic analysis. GIS is well suited for the change of research area in level, like moving from small level geographical units to larger units, and it is convenient to capture the effect of a larger region on one of its components too. Nowadays, the integration of spatial statistical analysis with GIS provides a more sophisticated approach to evaluate the role that space plays in both economic and environment, and spatial data analysis problems become more manageable. However, the use of that integration in regional economic studies is still in relatively low level currently.

On the basis of the division of Xinjiang discussed in Reference [12] and the integration of the types of spatial association with the economic concepts, we will make a further research on results calculated. The followings are those underlying spatial association patterns that can be detected between core and its adjacent area units\(^8\):

1) Spread Through Growth (\(+ +\)\): Growth in adjacent area units is associated with rapid growth in the economic core;

2) Spread Through Decentralization (\(- +\)\): Growth in adjacent area units is associated with slow growth in the economic core;

3) Backwash (\(+ -\)\): growth in economic core adjacent is associated with slow growth or decline in adjacent area units;

4) Independence; growth in adjacent area units is not closely associated with changes in economic activity in the economic core.

REFERENCES

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

2 Goodchild M F (1986) *Spatial autocorrelation*. Norwich: GeoBooks.

3 Getis A, Ord J K (1992) The analysis of spatial association by the use of distance statistics. *Geographical Analysis*, 24(3): 189-206

4 Ord J K, Getis A (1995) Local autocorrelation statistics; Distributional issues and an application. *Geographical Analysis*, 27(4): 286-306

5 Anselin L (1995) Local indicators of spatial association-LISA. *Geographical Analysis*, 27(2): 93-115

6 Drummond W J (1993) GIS as a visualization tool for economic development. *Comput. , Environ. and Urban Systems*, 17(5): 469-179

7 Goodchild M, Haining R, et al. (1992) Integrating GIS and spatial data analysis; problems and possibilities. *Int. J. Geographical Information Systems*, 6 (5): 407-423

8 Ding Y M, Fotheringham A S (1992) The integration of spatial analysis and GIS. *Comput. , Environ. and Urban Systems*, 16(1): 3-19

9 Bao S, Henry M S, Barkley D (1995) RAS~ a regional analysis system integrated with arc/Info. *Comput. , Environ. and Urban Systems*, 19(1): 37-56

10 Zhang Z Q, Griffith D A (1997) Developing user-friendly spatial statistical analysis modules for GIS; an example using ArcView. *Comput. , Environ. and Urban Systems*, 21(1): 5-29

11 Zhang Z Q, Griffith D A (2000) Integrating GIS components and spatial statistical analysis in DBMSs. *Int. J. Geographical Information Science*, 14 (6): 543-566

12 Gao Z G (2001) Study on the regional economic difference and its pre-warning in Xinjiang; [Ph. D dissertation]. Beijing: Institute of Geographic Sciences and Natural Resources, Chinese Academy of Sciences. (in Chinese)