Convergence between GIS and Spatial Distribution via Statistical Analysis

Hossein Niavand[1], Shima Sajadi.J[2]

[1] Research scholar in Statistics Department at University of Mysore, Mysore
[2]Academic staff of Surveying Engineering Department at University of Eshragh, Iran

Abstract: *GIS and spatial analysis have enjoyed a long and productive relationship over the past decades. GIS has been seen as the key to implementing methods of spatial analysis, making them more accessible to a broader range of users, and hopefully more widely used in making effective decisions and in supporting scientific research. It has been argued (e.g. Goodchild 1988) that in this sense the relationship between spatial analysis and GIS is analogous to that between statistics and the statistical packages.*
Keywords: *Spatial Distribution, GIS, Statistical Analysis, Poisson distribution.*

I. Introduction

A spatial distribution is the arrangement of a phenomenon across the Earth's surface and a graphical display of such an arrangement is an important tool in geographical and environmental statistics. A graphical display of a spatial distribution may summarize raw data directly or may reflect the outcome of more sophisticated data analysis. Many different aspects of a phenomenon can be shown in a single graphical display by using a suitable choice of different colours to represent differences. One example of such a display could be observations made to describe the geographic patterns of features, both physical and human across the earth.

The information included could be where units of something are, how many units of the thing there are per units of area, and how sparsely or densely packed they are from each other.
In geography, the term spatial distribution refers to the act of being spread or scattered over a range, area or volume of a place.

A geographic information system (GIS) is a computer-based tool for mapping and analyzing things that exist and events that happen on Earth. GIS technology integrates common database operations such as query and statistical analysis with the unique visualization and geographic analysis benefits offered by maps. These abilities distinguish GIS from other information systems and make it valuable to a wide range of public and private enterprises for explaining events, predicting outcomes, and planning strategies. (ESRI 1997)

A GIS provides a powerful collection of tools for the management, visualization and analysis of spatial data. These tools can be even more powerful when they are integrated with statistical methods for spatial data analysis and many GIS users are requesting this integration.

Most users use GIS only for mapping purposes or, at most, for buffering or overlaying various thematic data layers (Lee and Wong 2001). However, increasingly GIS users are recognizing that to make the best use of the spatial data, it should be analyzed statistically and mapped in the same environment. The present trend is to use classical statistical tools in packages such as SAS, SPSS and then map the results in GIS.

Lee and Wong (2001) make the point that by following this approach, data sets which should be explored on different geographic scales are only examined on one scale, as it is very difficult to do any more than this with conventional statistical packages. They also warn that using

classical statistical methods in this way and ignoring spatial relationships within data sets can potentially violate the assumptions of independence that underpin classical statistical methods. Recognition of this has resulted in new developments in GIS, where statistical analysis and mapping are carried out in the same environment.

II. Methods

Measuring the distribution of the set of features allow you to calculate a value which represents a characteristics of the distribution such as the center, compactness, or orientation. You can use this value to track the change in the distribution or to compare distributions of different features. There are some tools (tool set) for measuring spatial distribution. The Measuring Geographic Distributions toolset addresses questions such as:

i. Where is the center?
ii. What is the shape and orientation of the data?
iii. How dispersed are the features?

III. Centrographic Statistics

The most basic type of descriptors for the spatial distribution is Centrographic Statistics. These are indices which estimate basic parameter s about the distribution (Lefever , 1926; Furfey, 1927; Bachi, 1957; Neft, 1962, Hultquist , Brown and Holmes, 1971; Ebdon, 1988). They include:

I. Mean Center
II. Median Center
III. Standard Deviation
IV. Standard Distance
V. Standard Deviational Ellipse

They are called centrographic in that they are two dimensional correlates to the basic statistical moments of a single-variable distribution - mean, standard deviation, skewness, and kurtosis (Bachi, 1957).

Mean Center

The mean center is the average X and Y coordinate of all the features in the study area. It is useful for taking the changes in the distribution or for comparing the distribution of different type of features. The Mean Center Tool creates a new point feature class where each feature represents a mean center. This tool is also called as Center of Gravity. The mean center formula is shown as below:

$$\overline{X} = \sum_{i=1}^{N} \frac{X_i}{N} \qquad \overline{Y} = \sum_{i=1}^{N} \frac{Y_i}{N}$$

Median Center

The median center is the intersection between the median of the X coordinate and the median of the Y coordinate. Median Center is robust to outliers. It uses an algorithm to find the point that minimizes travel from it to all other features in the dataset.   At each step (t) in the algorithm, a candidate Median Center is found (X^t, Y^t) and refined until it represents the location that minimizes Euclidian Distance d to all features (i) in the dataset. The median center utilizes the below formula:

$$d_i^t = \sqrt{(X_i - X^t)^2 + (Y_i - Y^t)^2}$$

Standard Deviation

The standard deviations of the X and Y coordinates indicate the degree of dispersion of the distribution. The below formula represents the Standard Deviation process.

$$S_x = \sqrt{\sum_{i=1}^{N} \frac{(X_i - \overline{X})^2}{N-1}} \qquad S_y = \sqrt{\sum_{i=1}^{N} \frac{(Y_i - \overline{Y})^2}{N-1}}$$

Standard Distance

Measuring the compactness of a distribution provides a single value which represents the dispersion of the features around the center. The value is a distance, so compactness can be represented on a map by drawing a circle with the radios equal to the value. The Standard Distance Tool creates a new feature class containing a circle polygon centered on the mean center. The concern formula of standard distance is represented as below:

$$S_{xy} = \sum_{i=1}^{N} \frac{(d_{iMC})^2}{N-2}$$

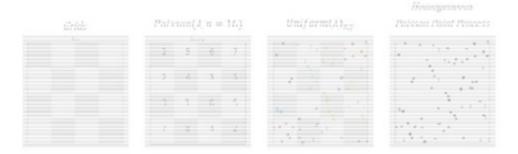**Standard Deviational Ellipse (Directional Distribution)**

A common way to measure the trend of a group of features is to calculate the standard distance separately in the X and Y directions. These two measures define the axis of an ellipse which encompassing the distribution of features. The ellipse is referred to as the Standard Deviational Ellipse, since the method calculates the standard deviation of the X coordinates and Y coordinated from the mean center to define the axes of the ellipse. The ellipse allows you to see if the distribution of features is elongated and hence has a particular orientation. The directional distribution uses a formula as mentioned below:

$$Distribution = \sqrt{\frac{\delta_x^2 + \delta_y^2}{2}}$$

Where δ are the two standard deviations in the x and y direction and orthogonal to each other to define an ellipse.

IV. Poisson distribution

Poisson point processes (PPP) are widely discussed in the literature. In the following figure a framework to generate two-dimensional PPP is demonstrated. First the area being studied (part of space which can be in 1D, 2D, 3D, ..., in our example is a 2D shape i.e., square) is divided into cells (gridding). Second, for each cell a random number n is drawn from a Poisson distribution with density of λ. Then within each cell n points are uniformly distributed. The resulting point pattern is a homogeneous point process.

1. Sub- divide the region R into m sub-regions, A_1, A_2, …, Am, where A_1, A_2, …, A_m are disjoint. Evaluate the mean of the Poisson distribution for each sub- region as $\mu_i = \lambda \cdot \nu(A_i)$, where λ is the (constant) intensity. R is partitioned by a user- specified grid system which results in rectangular sub- regions in 2D and cuboid block sub- regions in 3D applications.
2. For each of the m sub- regions A_i, generate a random variable, N_i, from the Poisson distribution.

$$P(N_i = n) = e^{-\mu_i} \frac{\mu_i^n}{n!}$$

3. For each of the n events in sub- region A_i, generate two (2D) or three (3D) values from uniform distributions and use them as the coordinates of the point simulated inside the sun- region.
4. Repeat step 2 and 3 until all sun- regions in R have been visited.

V. Conclusion

Geographical information systems and spatial statistics got to a crossroads about a decade ago. The special nature of spatially referenced data gave rise to intensive scientific research, widespread educational and knowledge dissemination efforts and active market extension.

Spatial distribution is referred to the locations of features or measurements observed in geographic space. A GIS provides a powerful collection of tools for the management, visualization and analysis of spatial data. These tools can be even more powerful when they are integrated with statistical methods for spatial data analysis and many GIS users are requesting this integration. The spatial distribution measurement tools determine the center, compactness or orientation of a set of features.

According to the methods discussed above, we conclude that the value of mean and variance are equal to each other in Poisson distribution, hence, if the considered study area has a homogeneous shape such as square or rectangular shapes, the PPP method could be used to investigate the spatial distribution pattern. Versus, for a study area having a heterogeneous shape, the centrographic statistics method would be utilized, since an erratic geometric shape creates a digressed space which increases the percentage of Standard Error of Mean (SEM). Therefore, it is better to use the centrographic statistics method to access the central point easily.

References

[1]. Bailey, T. C. and Gatrell, A. C. (1995). Interactive Spatial Data Analysis. Essex: Addison Wesley Longman Limited.
[2]. Berry B J L, Marble D F (eds) (1968). Spatial analysis: a reader in statistical geography. Englewood Cliffs, Prentice-Hall.
[3]. Cliff, A. D. and Ord, J. K. (1981). Spatial Processes: Models and Applications, London: Pion Limited.
[4]. Cressie, N.A.C., (1993). Statistics for spatial data. J. Wiley and Sons, New York.
[5]. DeGroot, M. H., (1986)Probability and Statistics, Addison-Wesley, Reading, Mass.
[6]. Fotheringham, S. and Rogerson, P. (eds) (1994). Spatial analysis and GIS. Taylor & Francis, London.
[7]. Fotheringham S. (2000). Quantitative Geography - Perspectives on Spatial Data Analysis. Trowbridge, Great Britain 2000, Sage Publications.
[8]. Isaaks E. and R. Mohan Srivastava (1989). Applied Geostatistics. New York, USA 1989, Oxford University Press.
[9]. Lee J. and D. Wong (2001). Statistical Analysis with ArcView GIS. USA 2001, John Wiley & Sons.