Association Rules, Spatio-temporal

► Movement Patterns in Spatio-temporal Data

Atlas, Electronic

► Multimedia Atlas Information Systems

Atlas Information System

► Multimedia Atlas Information Systems

Atlas, Interactive

► Multimedia Atlas Information Systems

Atlas, Multimedia

► Multimedia Atlas Information Systems

Atlas, Virtual

► Multimedia Atlas Information Systems

Attribute and Positional Error in GIS

► Uncertain Environmental Variables in GIS

Autocorrelation, Spatial

CHANDANA GANGODAGAMAGE[1],
XIAOBO ZHOU[2], HENRY LIN[2]
[1] Saint Anthony Falls Laboratory,
Department of Civil Engineering,
University of Minnesota, Minneapolis, MN, USA
[2] Department of Crop and Soil Sciences,
The Pennsylvania State University,
University Park, PA, USA

Synonyms

Spatial correlation; Spatial dependence; Spatial inter-dependence

Definition

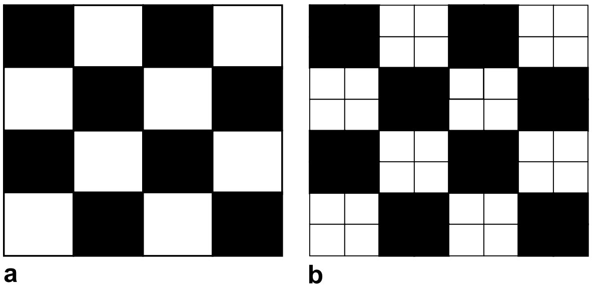
In many spatial data applications, the events at a location are highly influenced by the events at neighboring locations. In fact, this natural inclination of a variable to exhibit similar values as a function of distance between the spatial locations at which it is being measured is known as spatial dependence. Spatial autocorrelation is used to measure this spatial dependence. If the variable exhibits a systematic pattern in its spatial distribution, it is said to be spatially autocorrelated. The existence and strength of such interdependence among values of a specific variable with reference to a spatial location can be quantified as a positive, zero, or negative spatial autocorrelation. Positive spatial autocorrelation indicates that similar values or properties tend to be collocated, while negative spatial autocorrelation indicates that dissimilar values or properties tend to be near each other. Random patterns indicate zero spatial autocorrelation since independent, identically distributed random data are invariant with regard to their spatial location.

Historical Background

The idea of spatial autocorrelation is not new in the literature and was conceptualized as early as 1854, when nebula-like spatial clusters with distance-decay effects were readily apparent in mapped cholera cases in the city of London [1]. This led to the hypothesis that the systematic spatial pattern of Cholera outbreak decayed smoothly with distance from a particular water supply which acted as the source for the disease. This concept of spatial autocorrelation was also documented in the first law of geography in 1970 which states: "Everything is related to everything else, but near things are more related than distant things" [11].

Scientific Fundamentals

Spatial autocorrelation is a property of a variable that is often distributed over space [9]. For example, land surface elevation values of adjacent locations are generally quite similar. Similarly, temperature, pressure, slopes, and rainfall vary gradually over space, thus forming a smooth gradient of a variable between two locations in space. The propensity of a variable to show a smooth gradient across space aggregates similar values or properties adjacent to each other.
In classical statistics, the observed samples are assumed to be independent and identically distributed (iid). This assumption is no longer valid for inherently spatially autocorrelated data. This fact suggests that classical statistical tools like linear regression are inappropriate for spatial



Autocorrelation, Spatial, Figure 1 The strength of spatial autocorrelation as a function of scale using: **a** 4-by-4 raster and **b** 8-by-8 raster

data analysis. The inferences made from such analyses are either biased, indicating that the observations are spatially aggregated and clustered, or overly precise, indicating that the number of real independent variables is less than the sample size. When the number of real independent variables is less than the sample size, the degree of freedom of the observed data is lower than that assumed in the model.

Scale Dependence of Spatial Autocorrelation

The strength of spatial autocorrelation is often a function of scale or spatial resolution, as illustrated in Fig. 1 using black and white cells. High negative spatial autocorrelation is exhibited in Fig. 1a since each cell has a different color from its neighboring cells. Each cell can be subdivided into four half-size cells (Fig. 1b), assuming the cell's homogeneity. Then, the strength of spatial autocorrelation among the black and white cells increases, while maintaining the same cell arrangement. This illustrates that spatial autocorrelation varies with the study scale.

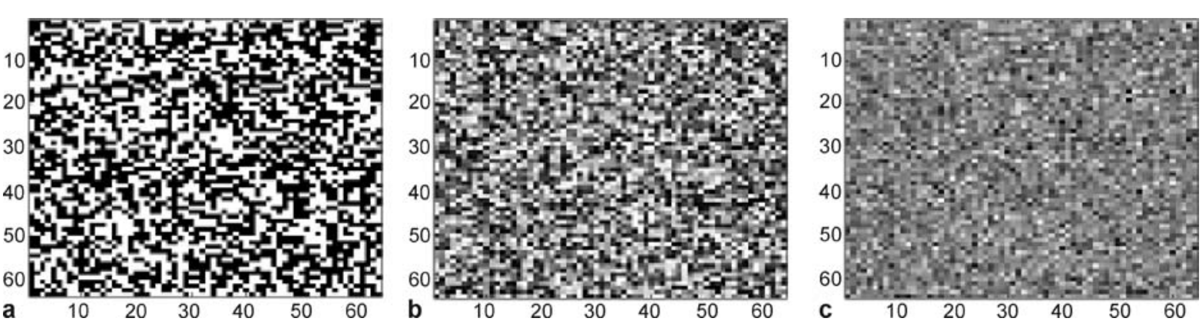
Differentiating Random Data from Spatial Data

Consider three different random distributions and three lattice grids of 64-by-64 cells (see Fig. 2) of each distribution: The first lattice data-set (Fig. 2a) is generated from a binary distribution, the second data-set (Fig. 2b) is generated from a uniform distribution, and the third data-set is generated from a normal distribution. The value at pixel (i, j), $P(i, j)$ $\{P(i, j);\ i = 1, \ldots, 64,\ j = 1, \ldots, 64\}$ is assumed to be independent and identically distributed. As shown in Fig. 2, the non-clustering or spatial segregation of the data suggests that the value $P(i, j)$, where $i, j \in R$, has no correlation (zero correlation) with itself in space.

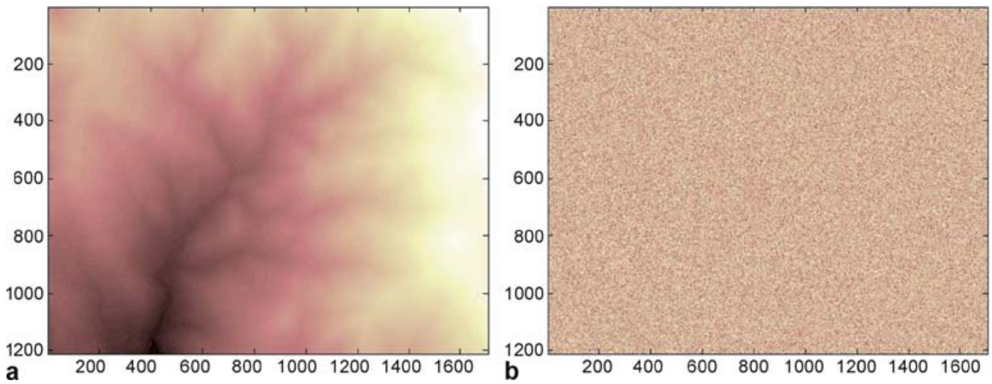
Each pixel (i, j) has eight neighborhoods and each neighborhood also has its own eight adjacent neighborhoods except the cells located on the boundary. The variability of $P(i, j)$ in one direction will not be the same in other directions; thus forming an anisotropic system, indicating the spatial autocorrelation varies in all directions. The quantification of this directional spatial autocorrelation is computationally expensive; thus, the average of each direction at distance k is used to quantify the spatial autocorrelation. The distance k (e. g., k pixel separation of (i, j) in any direction) is called lag distance k. The spatial autocorrelation from each spatial entity to all other entities can be calculated. The average value over all entities of the same lag distance is expressed as a measure of spatial autocorrelation. The above three data-sets are illustrative examples, demonstrating the nonexistence of spatial autocorrelation in randomly generated data-sets.

Consider a digital elevation model (DEM) that shows an array of elevations of the land surface at each spatial location (i, j) as shown in Fig. 3a. The values of this data-set do not change abruptly, whereas in Fig. 3b, the difference of the elevations between the location (i, j) and its neighborhoods changes abruptly as shown in its corresponding color scheme.

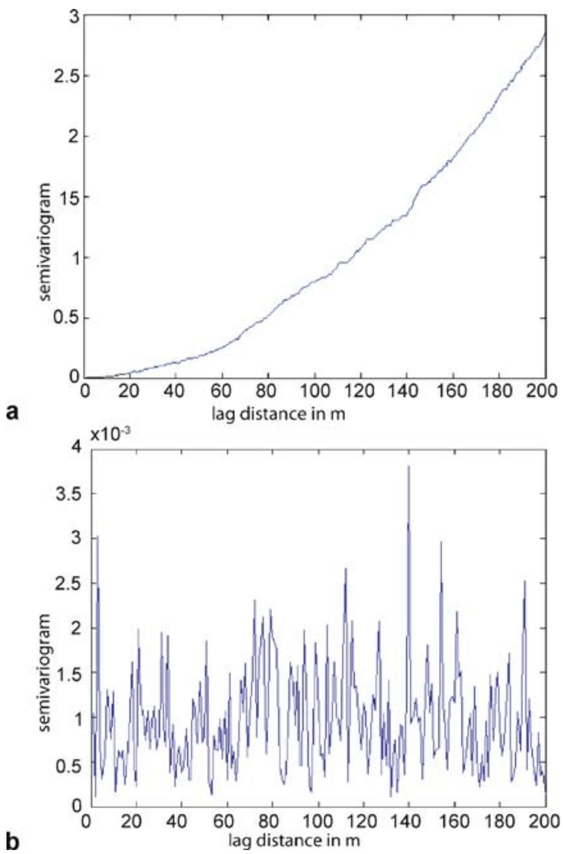
The variogram, a plot of the dissimilarity against the spatial separation (i. e., the lag distance) [12] in spatial data, quantifies spatial autocorrelation and represents how spatial variability changes with lag distance [2]. In Fig. 4a, the semi-variogram value of the DEM surface is zero at the zero lag distance and increases with the lag dis-



Autocorrelation, Spatial, Figure 2 Three different data distributions. **a** Binary distributed data in space. **b** Random uniformly distributed lattice data. **c** Random normally distributed lattice data in space



Autocorrelation, Spatial, Figure 3 **a** One meter spatial resolution LIDAR DEM for south fork Eel, California. **b** One meter normally distributed DEM reconstructed for same statistics (i. e., mean and variance) as LIDAR DEM in **a**



Autocorrelation, Spatial, Figure 4 **a** Variogram for spatial data in Fig. 3a. **b** Variogram for the random data in Fig. 3b

tance, whereas in Fig. 4b, the semi-variogram value of the random surface varies erratically with the increasing lag distance. Contrary to spatial autocorrelation, the semi-variogram has higher values in the absence of spatial correlation and lower values in the presence of spatial correlation. This indicates that spatial autocorrelation gradually disappears as the separation distance increases [10] (Fig. 4a). These variogram figures are generated at a point (x_i) by comparing the values at its four adjacent neighbors such that:

$$\gamma(h) = \frac{1}{N(h)} \sum_{i=1}^{n} (z(x_i) - z(x_i + h))^2 \ , \tag{1}$$

where $z(x_i)$ and $z(x_i + \mathrm{h})$ are the values of the function z located at x_i and $(x_i + \mathrm{h})$, respectively. The four-adjacent-average of the squared difference values along the X and Y axes at lag distance h are used in these variogram clouds. The semi-variogram values in Fig. 4a (generated from lattice data) increase with increasing lag distance whereas the semi-variogram values generated from point data reach a steady state with increasing lag distance.

How to Quantify Spatial Autocorrelation

Several indices can be used to quantify spatial autocorrelation. The most common techniques are Moran's I, Geary's C, and spatial autoregression. These techniques are described in the following sections.

Moran's *I* Method Moran's I index is one of the oldest (Moran, 1950) methods in spatial autocorrelation and is still the de facto standard method of quantifying spatial autocorrelation [8]. This method is applied for points or zones with continuous variables associated with them. The value obtained at a location is compared with the value of other locations. Morgan's I method can be defined as:

$$I = \frac{N \sum_i \sum_j W_{i,j}(X_i - \bar{X})(X_j - \bar{X})}{(\sum_i \sum_j W_{i,j}) \sum_i (X_i - \bar{X})^2} \ , \tag{2}$$

where N is the number of cases, $\bar{X}$ is the mean value of the variable X, X_i and X_j are the values of the variable X at

location i and j, respectively, and $W_{i,j}$ is the weight applied to the comparison between the values at i and j.
The same equation in matrix notation can also be represented as [9]:

$$I = \frac{zWz^t}{zz^t}\,, \tag{3}$$

where $z = (x_1 - \bar{x}, x_2 - \bar{x}, \ldots, x_n - \bar{x})$, z^t is the transpose of matrix z and W is the same contiguity matrix of n- by-n that has been introduced in Eq. 2.
An important property in Moran's I is that the index I depends not only on the variable X, but also on the data's spatial arrangement. The spatial arrangement is quantified by the contiguity matrix, W. If a location i is adjacent to location j, then this spatial arrangement receives the weight of 1; otherwise the value of the weight is 0. Another option is to define W based on the squared inverse distance $(1/d_{ij}^2)$ between the locations i and j [6]. There are also other methods to quantify this contiguity matrix. For example, the sum of the products of the variable x can be compared at locations i and j and then weighted by the inverse distance between i and j.
The value of I is close to 1 or -1 when spatial autocorrelation is high or low, respectively [6].

Geary's *C* Method Geary's method (Geary, 1954) differs from Moran's method mainly in that the interaction between i and j is measured not as the deviation from the mean, but by the difference of the values of each observation [4]. Geary's C can be defined as:

$$C = \frac{(N-1)\left[\sum_i \sum_j W_{ij}(X_i - X_j)^2\right]}{2\sum_i \sum_j W_{ij}(X_i - \bar{X})^2}\,, \tag{4}$$

where C typically varies between 0 and 2. If the value of one zone is spatially unrelated to any other zone, the expected value of C will be 1. If the value of C is less than 1, a negative spatial autocorrelation is inferred [6]. Geary's C values are inversely related to Moran's I values.
Geary's C and Moran's I will not provide identical inference because the former deals with differences and the latter deals with covariance. The other difference between these two methods is that Moran's I gives a global indication while Geary's C is more sensitive to differences in small neighborhoods [6].

Spatial Autoregression

The disadvantage of linear regression methods is that they assumed iid condition, which is strictly not true for spatial data analysis. Research in spatial statistics has suggested many alternative methods to incorporate spatial dependence into autoregressive regression models, as explained in the following section.

Spatial Autoregressive Regression Model The spatial autoregressive regression model (SAR) is one of the commonly used autoregressive models for spatial data regression. The spatial dependence is introduced into the autoregressive model using the contiguity matrix. Based on this model, the spatial autoregressive regression [9] can be written as:

$$Y = \rho WY + X\beta + \varepsilon\,, \tag{5}$$

where

- Y Observation or dependent variable,
- ρ Spatial autoregressive parameter,
- Y Observation or depend-ant variable,
- W Contiguity matrix,
- β Regressive coefficient,
- α Unobservable error term (N (0,σI)),
- X Feature values or independent variable.

When $\rho = 0$, this model is reduced to the ordinary least square regression equation.
Solution for Eq. 5 is not straightforward and the contiguity matrix W gets quadratic in size compared to the original size of data sample. However, most of the elements of W are zero; thus, sparse matrix techniques are used to speed up the solution process [9].

Illustration of SAR Using Sample Data-Set Consider the following 2-by-2 DEM grid data-set.

100	101
102	103

The contiguity matrix (neighborhood matrix) W can be written as follows:

$$\begin{array}{c|c} & \begin{array}{cccc} 100 & 101 & 102 & 103 \end{array} \\ \begin{array}{c} 100 \\ 101 \\ 102 \\ 103 \end{array} & \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix} \end{array}$$

$$\xrightarrow{\text{Normalized contiguity matrix}} \begin{bmatrix} 0 & 0.5 & 0.5 & 0 \\ 0.5 & 0 & 0 & 0.5 \\ 0.5 & 0 & 0 & 0.5 \\ 0 & 0.5 & 0.5 & 0 \end{bmatrix}$$

The normalized contiguity matrix is shown in the right panel. We assumed that $\rho = [0.1]$, $\beta = 1.0 * [1; 2; 3; 4]$, and column vector ε **is** equal to a column vector *(0.01*rand(4,1))*. Then, Eq. 5 can be written as:

$$Y = (QX)\,\beta + \varepsilon\,, \tag{6}$$

where $Q = (I - \rho W^{-1})$.

Demonstration Using Mathworks Matlab Software Matlab software [7] is used to demonstrate this example. The following five matrices are defined for W, ρ, ε, β, and X as W = *[0, 0.5, 0.5, 0; 0.5, 0, 0, 0.5; 0.5, 0, 0, 0.5; 0, 0.5, 0.5, 0]*, $\rho = [0.1]$, $\varepsilon = 0.01*rand(4,1)$, $\beta = 1.0$, $X = [100; 101; 102; 103]$. The above defined values are substituted into Eq. 6, which can be shown in Matlab notation as $y = inv\ (eye(4,4)\text{-}\rho.\ *W)*(\beta\ *X{+}\varepsilon)$. The solution provides an estimation of y = *[111.2874, 112.2859, 113.2829, d 114.2786]*.

Key Applications

The key application of spatial autocorrelation is to quantify the spatial dependence of spatial variables. The following are the examples from various disciplines where spatial autocorrelation is used:

Sampling Design

The defined spatial autocorrelation among the contiguous or close locations can be used to answer how large of an area does a single measurement represent. The answer to such questions allows estimates of the best places to make further observations and the number of the samples required in accuracy assessment, and provides useful information for interpolation to estimate values at unobserved locations [13].

Cartography

A main assumption on which statistical estimates of uncertainty are usually based is the independence of the samples during mapping processes. A spatial autocorrelation analysis can be used to test the validity of such an assumption and the related mapping errors. Adjacent elevation differences are usually correlated rather than independent and errors tend to occur in clusters. In addition, the level of accuracy of GIS output products depends on the level of spatial autocorrelation in the source data-sets.

Soil Science

Spatial autocorrelation has been used to study the domain that a soil water content or soil temperature measurement can represent. The distinctive spatial autocorrelations of soil solutes manifests the different reaction and migration patterns for solutes in soil. With a high-resolution soil sampling, a spatial autocorrelation analysis provides another means to delineate boundaries between soil series.

Biology

Patterns and processes of genetic divergence among local populations have been investigated using spatial autocorrelation statistics to describe the autocorrelation of gene frequencies for increasing classes of spatial distance. Spatial autocorrelation analysis has also been used to study a variety of phenomena, such as the genetic structure of plant and animal populations and the distribution of mortality patterns.

Ecology

Ecologists have used spatial autocorrelation statistics to study species–environment relationships. Spatial autocorrelation analysis is a useful tool to investigate mechanisms operating on species richness at different spatial scales [3]. It has shown that spatial autocorrelation can be used to explore how organisms respond to environmental variation at different spatial scales.

Environmental Science

The physical and chemical processes controlling the fate and transport of chemicals in the environment do not operate at random. All measurable environmental parameters exhibit spatial autocorrelation at certain scales [5]. The patterns of spatial autocorrelation in stream water quality can be used to predict water quality impaired stream segments. The spatial autocorrelation test of environmental variables provides important information to the policy-makers for more efficient controls of environmental contaminants.

Risk Assessment

It is often the case that the occurrence of natural hazardous events such as floods and forest fires shows spatial dependence. Spatial autocorrelation allows risk assessment of such undesirable events. It can be used to estimate the probability of a forest fire, as an example, taking place at a specific location. Spatial autocorrelation analysis is also useful in geographical disease clustering tests.

Economics

Because of the heterogeneity across regions and a large number of regions strongly interacting with each other, economic policy measures are targeted at the regional level. Superimposed spatial structures from spatial autocorrelation analysis improve the forecasting performance of non-spatial forecasting models. The spatial dependence and spatial heterogeneity can be used to investigate the effect of income and human capital inequalities on regional economic growth. Spatial autocorrelation analysis is also a useful tool to study the distribution of unemployment rate and price fluctuation within a specific area.

Political Science

After spatial autocorrelation has been defined, geographic units (countries, counties, or census tracts) can be used as predictors of the political outcomes. For example, spatial autocorrelation methods can use geographic data coordinates to check if a location has a significant impact on the voting choice.

Sociology

Spatial autocorrelation has been used to study the correlation between population density and pathology. The spatial interaction has been taken into consideration to study the relationship between population density and fertility. Spatial autocorrelation can also be used to investigate the variations in crime rates and school test scores.

Future Directions

A good knowledge and understanding of spatial autocorrelation is essential in many disciplines which often need predictive inferences from spatial data. Ignoring spatial autocorrelation in spatial data analysis and model development may lead to unreliable and poor fit results. The result of spatial autocorrelation analysis can guide our experiment design, trend analysis, model development, and decision-making. For example, long-term field data monitoring is tedious and costly. Spatial autocorrelation analysis would benefit the design and sampling strategies development of optimal field monitoring sites. Spatial autocorrelation analysis for variables of interest can also assist in the selection of a supermarket location or of a new school.

Cross References

► Semivariogram Modeling

Recommended Reading

1. Moore, D.A., Carpenter, T.E.: Spatial analytical methods and geographic information systems: Use in health research and Epidemiology. Epidemiol. Rev. **21**(2), 143–161 (1999)
2. Devary, J.L., Rice, W.A.: Geostatistis software user's manual for geosciences research and engineering. Pacific Northwest Laboratory, Richland, Washington (1982)
3. Diniz, J.A.F., Bini, L.M., Hawkins, B.A.: Spatial auto-correlation and red herrings in geographical ecology. Glob. Ecol. Biogeogr. **12**, 53–64 (2003)
4. Geary, R.C.: The contiguity ratio and statistical mapping. Inc. Stat. **5**, 115–145 (1954)
5. Haining, R.P.: Spatial data analysis in the social and environmental sciences. Cambridge University Press, Cambridge (1993)
6. Lembo, J.: Lecture 9, Spatial Autocorrelation. http://www.css.cornell.edu/courses/620/lecture9.ppt#256,1. Accessed 10 Apr 2006
7. MATLAB, The Student Edition of MATLAB, The Language of Technical Computing, Version 5, User's Guide. The Mathworks, Inc., New Jersey (1997)
8. Moran, P.A.P.: Notes on continuous stochastic phenomena. Biometrika **37**, 17–23 (1950)
9. Shekhar, S., Chawla, S.: Spatial databases: A tour. Pearson Education Inc., Prentice Hall, Upper Saddle River, NJ (2003)
10. Spatial Autocorrelation, Natural resources Canada. http://www.pfc.forestry.ca/profiles/wulder/mvstats/spatial_e.html. Accessed 10 Apr 2006
11. Tobler, W.R.: A computer movie simulating urban growth in the Detroit region. Econ. Geogr. **46**(2), 234–240 (1970)
12. Wackernagel, H.: Multivariate Geostatistics: An introduction with applications, 3rd edn., pp. 45–58. Springer-Verlag, New York, LLC (2003)
13. Wei, H., Chen, D.: The effect of spatial auto-correlation on the sampling design in accuracy assessment: A case study with simulated data. Environ. Inform. Arch. **2**, 910–919 (2004)
14. Pace, R.K., Barry, R., Sirmans, C.F.: Spatial Statistics and Real Estate. J. Real Estate Finance Econ. **17**, 5–13 (1998)

Automated Map Compilation

► Conflation of Features

Automated Map Generalization

► Feature Extraction, Abstract

Automated Vehicle Location (AVL)

► Intergraph: Real Time Operational Geospatial Applications

Automatic Graphics Generation

► Information Presentation, Dynamic

Automatic Information Extraction

► Data Acquisition, Automation

Autonomy, Space Time

► Time Geography

Autoregressive Models

► Hierarchical Spatial Models

A