Spatial Autocorrelation

D. A. Griffith, University of Texas at Dallas, Richardson, TX, USA

Glossary

Auto- A prefix literally meaning self; spatial autocorrelation means self-correlation, or values within a given variable are correlated, resulting in the variable being correlated with itself.

Correlation A description of the nature and degree of a relationship between a pair of quantitative variables.

Geary Ratio An index of spatial autocorrelation, involving the computation of squared differences of values that are geographic neighbors (i.e., paired comparisons), that ranges from 0 to 1 for negative, and 1 to approximately 2 for positive, spatial autocorrelation, with an expected value of 1 for zero spatial autocorrelation.

Geographic Connectivity/Weights Matrix An *n*-by-*n* matrix with the same sequence of row and column location labels, whose entries indicate which pairs of locations are neighbors.

Map Pattern The systematic organization of values for some variable across a map resulting in visually conspicuous texture that consists of global, regional, and local trends, gradients, swaths, or mosaics.

Moran Coefficient An index of spatial autocorrelation, involving the computation of cross-products of mean-adjusted values that are geographic neighbors (i.e., covariations), that ranges from roughly (–1, –0.5) to nearly 0 for negative, and nearly 0 to approximately 1 for positive, spatial autocorrelation, with an expected value of $-1/(n-1)$ for zero spatial autocorrelation, where n denotes the number of areal units.

Moran Scatterplot A scatterplot of standardized versus summed nearby standardized values whose associated bivariate regression slope coefficient is the unstandardized Moran coefficient.

Negative Spatial Autocorrelation For the geographic distribution of some variable across a map, high values tend to be geographic neighbors of low values, intermediate values tend to be geographic neighbors of intermediate values, and low values tend to be geographic neighbors of high values.

Positive Spatial Autocorrelation For the geographic distribution of some variable across a map, high values tend to be geographic neighbors of high values, intermediate values tend to be geographic neighbors of intermediate values, and low values tend to be geographic neighbors of low values.

Redundant Information Information in data that is duplicated, and hence unneeded for spatial statistical analyses; for georeferenced data, this duplication arises from locational closeness that results in mutually shared information, allowing an awareness of nearby values once the value for a given location is known.

Introduction

'Spatial autocorrelation' is the correlation among values of a single variable strictly attributable to their relatively close locational positions on a two-dimensional (2-D) surface, introducing a deviation from the independent observations assumption of classical statistics. Spatial autocorrelation exists because real-world phenomena are typified by orderliness, (map) pattern, and systematic concentration, rather than randomness. Tobler's first law of geography encapsulates this situation: "everything is related to everything else, but near things are more related than distant things." To this maximum should be added the qualifier: "but not necessarily through the same mechanisms." In other words, spatial autocorrelation means a dependency exists between values of a variable in neighboring or proximal locations, or a systematic pattern in values of a variable across the locations on a map due to underlying common factors.

Selected physical models portray the existence of spatial autocorrelation. First, and foremost, pictures (analogous to map pattern) would not be discernible on a television/computer screen without spatial autocorrelation. Nor could sliding tile or jigsaw puzzles be solved. Magnetic sculptures constructed from separate pieces of metal piled upon a magnetized base more completely illustrate spatial autocorrelation: when the pieces of metal are placed on their magnetic base, they can be sculpted into a 3-D figure; when the pieces of metal are moved far from their magnetic base (as well as any other magnetic source), they simply are independent pieces of metal that only can be brushed into a pile. The spatial autocorrelation mechanism here is the magnetic field that is transferred from piece to piece by metal pieces

touching. Meanwhile, a number of real-world examples also portray the existence of spatial autocorrelation. Most mineral deposits cluster at relatively few locations on the Earth's surface; they are not ubiquitous. House prices and house value assessments are established in the real estate market by comparisons between a house and similar nearby houses. Zoning ordinances force similar land-use types to group together in coterminous locations. Diseases, such as West Nile virus, creep across a landscape through contagion. Finally, unabated water (wind) polluters generate negative consequences for those downstream (downwind) of their locations; a similar but more fully 2-D example is furnished by defunct smelter superfund sites.

Conceptual Meanings of Spatial Autocorrelation

Spatial autocorrelation has many interpretations. The most dismissive is as a 'nuisance parameter'. Spatial autocorrelation is captured by a model specification because its presence is necessary for a good description, but it is not truly of interest and only interferes with the estimation of other model parameters that are of true interest. This interference tends to be for parameters such as variances, rather than means. Scientists increasingly are deciding that spatial autocorrelation should not be treated as a nuisance parameter.

Interpreting spatial autocorrelation as 'self-correlation' is literal. Correlation arises from the geographic context within which attribute values occur, and as such can be expressed in terms of the Pearson product–moment correlation coefficient (r) formula, but with neighboring values of variable y replacing those of variable x. Here, the correlation being ascertained is an average of that between location-specific time series of a variable for all possible pairs of locations. But these time series are not observable, and hence the assumption invoked is exchangeability (i.e., the set of time series can be permuted without affecting results – the order in which a time series mechanism generates values across a map is irrelevant).

Interpreting spatial autocorrelation as 'map pattern' emphasizes conspicuous trends, gradients, swaths, or mosaics across a map. Consider a constant, which is the degenerate case (i.e., a constant has no variance) of perfect positive spatial autocorrelation: once the value of a constant is known at a single location, it is known at all locations. Next, consider a variable that portrays a north–south (or east–west) linear trend across a map. If this variable has a mean of 0, then it is geometrically independent of any constant. These north–south and east–west oriented linear trend variables also are independent. A variable with mean 0 whose values' magnitudes form a 3-D symmetric hill (or valley) in the center of a map constitutes yet another mutually independent map pattern. These three variables display maximum levels of positive spatial autocorrelation when geographic variance is present, and may be described as global geographic patterns. Alternating sequences of moderately large hills and valleys with either an east–west or a north–south orientation portray moderate positive spatial autocorrelation, and constitute regional map patterns. Alternating sequences of small hills and valleys with either an east–west or a north–south orientation portray weak positive spatial autocorrelation, and constitute local map patterns. This fragmentation continues through randomness (zero spatial autocorrelation) to arrangements of increasingly alternating values (i.e., single value hills and valleys), which portray increasing negative spatial autocorrelation. Most substantive variables have geographic distributions that can be described by combinations of some subset of these mutually independent varying hill–valley cluster size map patterns.

As a 'diagnostic tool', spatial autocorrelation plays a crucial role in model-based inference, whose foundation is a set of valid assumptions rather than a scientific (i.e., random) sampling design. Detected spatial autocorrelation can signify model misspecification, including treating nonlinear relationships with a linear specification, such as an exponential relationship between a predictor variable, x, and a response variable, y, that is described with a straight trend line. It also can signify 'missing variables' for a regression equation, and as such serves as a 'surrogate' for variation otherwise unaccounted for because these variables are missing. This surrogate role occurs when autocorrelation map patterns displayed by predictor variables align with autocorrelation map patterns displayed by y. Moreover, the same set of distinct map patterns is common to both a set of missing variables and y. Accounting for spatial autocorrelation in this context can do a surprisingly good job of representing missing variables in an equation specification. In contrast, spatial autocorrelation that remains unaccounted for tends to distort classical correlation coefficient interpretations.

The term correlation alludes to the notion of 'redundant information'. If x and y are perfectly correlated, then knowing x means exactly knowing y. In other words, the information content of y is perfectly duplicated in x; this degree of duplication decreases as the correlation coefficient moves toward 0. Spatial autocorrelation extends this notion of redundant information to georeferenced data. The value at a given location can be predicted with some degree of accuracy from the values at nearby locations; this spatial data feature constitutes the foundation of cartographic interpolation. Recalling the example of housing prices, such redundant information results from 'spatial spillovers': house values at

one location spill over to impact upon house values at nearby locations. Building an expensive house near an inexpensive house tends to reduce the value of the expensive house while increasing the value of the inexpensive house, all other things being equal. This mechanism is similar to a river that floods its banks into its flood plain. Another example, at the micro-level of geographic resolution, is second-hand smoke generated by smokers. And yet another example is the noxious smell generated by such facilities as sewage treatment or rendering plants that permeates their surrounding neighborhoods.

Spatial autocorrelation can materialize from some course of action operating over a geographic landscape, such as contagious diffusion of a disease, resulting in it being interpreted as a 'spatial process mechanism'. The diffusion of West Nile virus across the coterminous United States (US) illustrates this interpretation. This disease quickly became a serious problem in the US, extremely rapidly diffusing from Long Island in 1999 throughout the remainder of the country following northeast-to-west and -south paths. A weather front moving across a geographic landscape can be viewed in a similar way because it results in highly spatially autocorrelated local weather conditions.

Finally, interpreting spatial autocorrelation as an 'outcome of areal unit demarcation' relates it to the modifiable areal unit problem (MAUP), whereby results from statistical analyses of georeferenced data can be varied at will simply by changing the surface partitioning used to demarcate areal units. For example, a standard eight-by-eight checkerboard pattern exhibits negative spatial autocorrelation between the red and black colors of its squares. But if these squares are aggregated into compact clusters of four (i.e., two-by-two groupings), and the red and black colors averaged (resulting in a constant dark red color across the checkerboard), then the spatial autocorrelation becomes maximally positive. Political redistricting involving gerrymandering (i.e., electoral district or constituency boundaries are manipulated in order to achieve a prespecified geographic aggregation result, such as some political advantage) exemplifies this general situation in practice.

Illustrations of Spatial Autocorrelation

Most empirical spatial autocorrelation cases involve moderate, positive relationship tendencies between nearby values on a map. Remotely sensed satellite images are one exception, almost always displaying a very strong positive relationship. Most socioeconomic/demographic data display a moderate positive relationship. And, negative spatial autocorrelation rarely is encountered in practice.

Strong Positive Spatial Autocorrelation

Remotely sensed images are one exception to the georeferenced data norm of moderate positive spatial autocorrelation, frequently displaying very strong positive spatial autocorrelation. This feature partly results from light reflectance scattering, rather than being neatly contained in pixel boundaries, which are imaginary, hence spilling over into nearby pixels measured by a satellite's sensors.

A 1000-by-1000 pixel subset was extracted from a satellite image of the Florida Everglades for illustrative purposes (**Figure 1a**). The amount of green vegetation in this region can be quantified with a normalized difference vegetation index (NDVI), which is calculated with the spectral reflectance values for the near-infrared (B5) and visible red (B3) bands as follows:

$$\frac{B5 - B3}{B5 + B3}$$

Positive values beyond about 0.3 tend to represent green vegetation, whereas negative values tend to represent swamp areas; positive values closer to 0 tend to represent soil. The 1 000 000 selected pixels yield only 3033 different NDVI values. This measurement was modified mathematically (i.e., transformed) in order to better align it with a bell-shaped curve (**Figure 1b**).

The image appearing in **Figure 1a** displays very strong positive spatial autocorrelation. The outline of the Everglades is apparent, as is the southeast coast of Florida, south of Miami. The absence of autocorrelation would result in this image appearing as a shuffling of the set of colors, similar to the picture-distorting momentary white specks that appear on a television screen when, for example, atmospheric static is dominant when, say, a cable connection is lost.

Moderate Positive Spatial Autocorrelation

Maps of population density tend to display moderate positive spatial autocorrelation, in part due to urbanization at a regional or national scale, and zoning at a local scale. Data from the 1993 census of Peru furnish population counts by districts across the Cusco department; an ArcGIS shapefile furnishes area measures for these 108 districts. Population density tends to be skewed, with a natural lower bound of 0, and few areal units with relatively sizeable concentrations.

Population density by district in the Cusco department ranges from 0.8 to 11 579.2 per unit area. Its inverse square root, after adding 11 to each density, better mimics a bell-shaped curve (see **Figure 2**). This transformed population density displays moderate positive spatial autocorrelation, forming an elongated mound map pattern with a single peak. The highest density is in the city of Cusco, with the next highest densities stretching

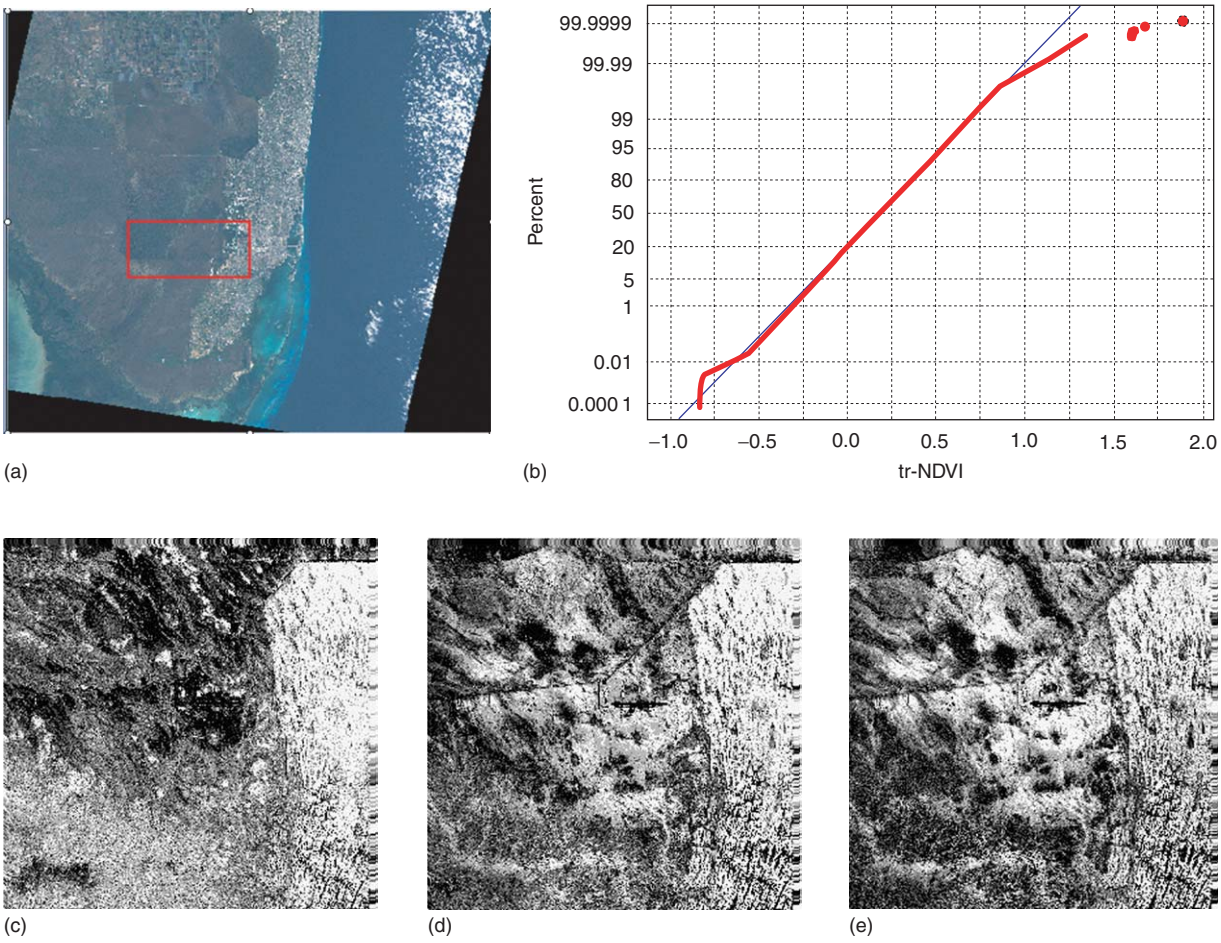

Figure 1 A Landsat 7 Enhanced Thematic Mapper Plus (ETM+) image of the Florida Everglades for 1 January 2002. (a) A composite image constructed using all spectral bands, with the subregion of analysis outlined with a red rectangle and the nonimage areas blacked out; (b) a quantile plot of the 1 000 000 transformed NDVI values for the subregion; (c) visible red spectral band (wavelength 0.63–0.69 microns (i.e., nanometers/1000)) for the subregion, with data values in the range 13–175; (d) mid-infrared spectral band (wavelength 1.55–1.75 microns) for the subregion, with data values in the range 1–173; and (e) transformed NDVI, with values in the range –0.84 to 1.89. Note: the gray scale for the maps is directly proportional to the pixel values.

along an economic corridor formed by the Vilcanota River valley; the lowest densities are in the most rural areas of this department.

Moderate Negative Spatial Autocorrelation

Few empirical examples of negative spatial autocorrelation are reported in the literature. Usually this phenomenon is discussed conceptually in terms of geographic competition. In other words, if a finite amount of land is available, gains in land size of one territory can occur only through the loss of land size in nearby territories. The World Wars fought on the European continent illustrate this situation.

Continental Europe principally in terms of the European Union is partitioned into 22 countries. Treating the capital of each country as its seat of power, and hence its focal point, this portion of the continent also can be partitioned into Thiessen polygons (**Figure 3b**). The ratio of each country's actual to corresponding Thiessen polygon land size gives an index of local competition, whose normal quantile plot (**Figure 3a**) implies a frequency distribution that conforms well to a bell-shaped curve. This areas ratio displays negative spatial autocorrelation: Switzerland, Luxembourg, Slovenia, and the Czech Republic are islands of very low ratio values that are completely surrounded by countries that have the highest ratio values (**Figure 3c**).

Estimators of Spatial Autocorrelation

Historically, once the spatial autocorrelation concept was established and widely recognized, spatial scientists became interested in quantifying it, and then testing hypotheses about it, with an ultimate goal of incorporating it into models. The two most commonly used

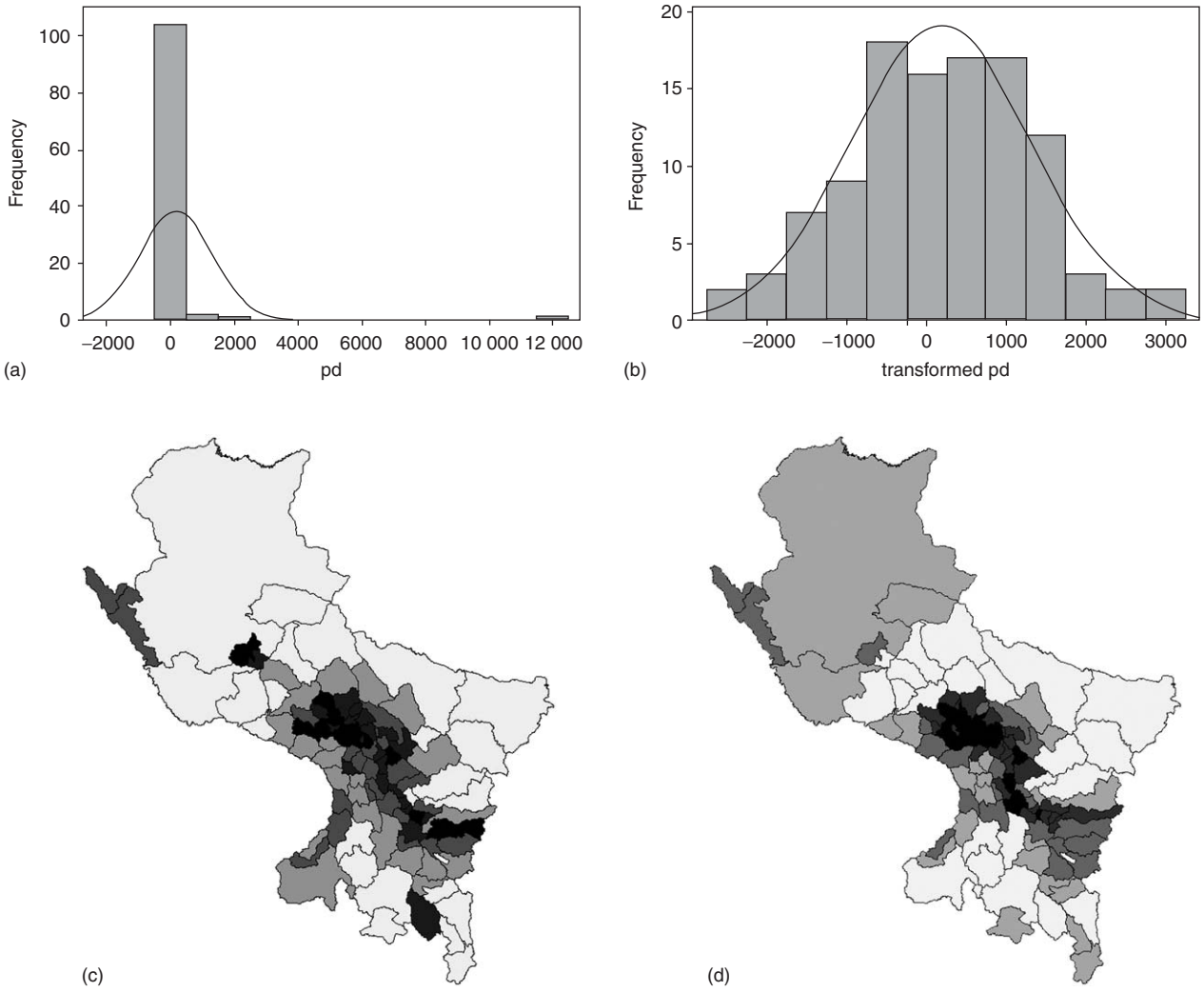

Figure 2 The 1993 population density across the department of Cusco, Peru. (a) A histogram constructed with the 108 raw population densities; (b) a histogram constructed with the 108 transformed population densities; (c) a quintile map of the geographic distribution of transformed population density values; and (d) a quintile map of the composite spatial autocorrelation map pattern latent in the geographic distribution of transformed population density values. Note: the gray scale for the maps is directly proportional to population density.

quantitative indices are the Moran coefficient (MC) and the Geary ratio (GR).

From *r* to Moran Coefficient

Exploiting the interpretation of self-correlation, spatial autocorrelation can be expressed in terms of the formula for r, but with neighboring values of variable y replacing those of x

$$\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})/n}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2/n}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2/n}}$$

becomes

$$\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}c_{ij}(y_i-\bar{y})(y_j-\bar{y})/\sum_{i=1}^{n}\sum_{j=1}^{n}c_{ij}}{\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2/n}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2/n}}$$

The left-hand expression converts to the right-hand one by replacing y's with x's in the right-hand side, by computing the numerator term only when areal units i and j are nearby (c_{ij} is an indicator variable – often called a spatial weight – whose value is 1 for neighbors, and 0 otherwise), and by averaging the numerator cross-product terms over the total number of pairs denoted as being nearby. The denominator of this revised expression is the sample variance of Y, s_Y^2. But unlike the values ± 1 for r, the extreme values of the MC are determined by sophisticated mathematical quantities called eigenvalues that are computed when the set of n indicator variables is

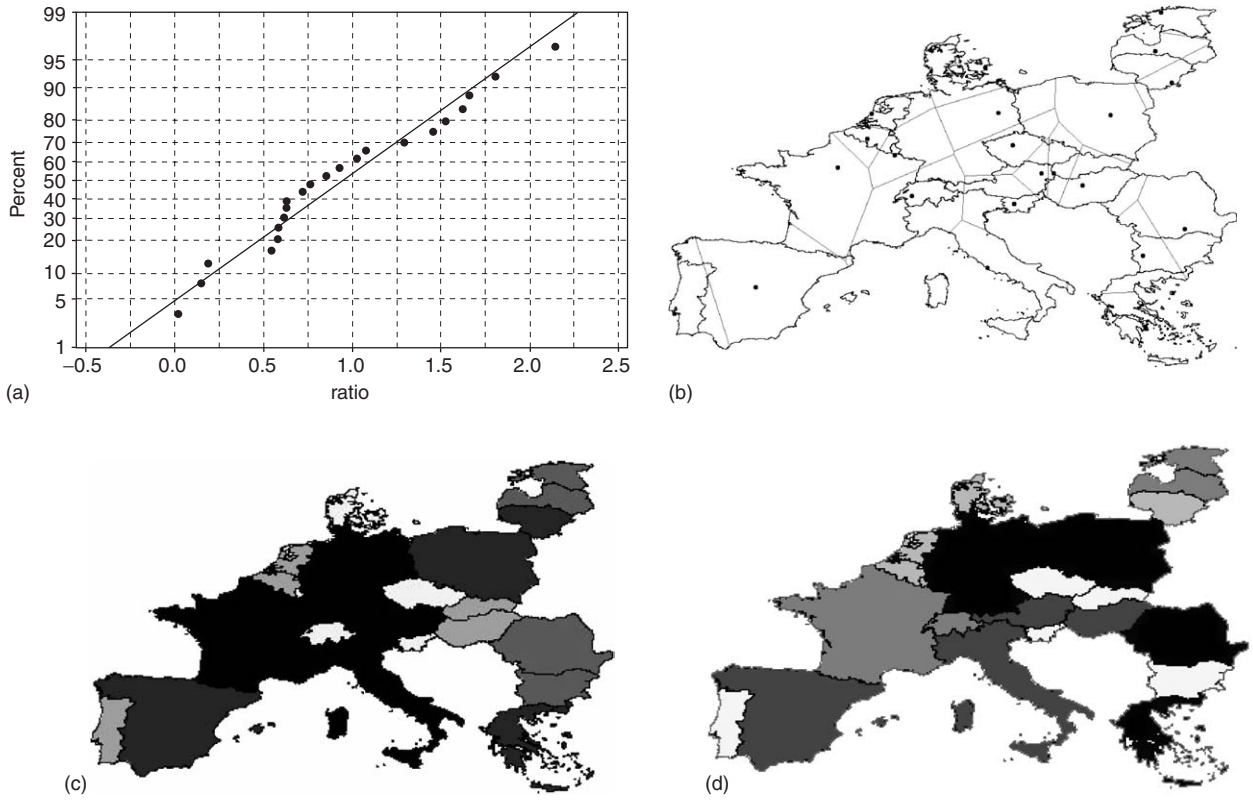

Figure 3 The geographic distribution of hypothetical land size competition across continental Europe. (a) A quantile plot of the ratio between actual and Thiessen polygon areas, by country; (b) the Thiessen polygon partitioning of the continent (gray lines), with country capitals as focal points; (c) a quintile map of the geographic distribution of the areas ratio values; and (d) a quintile map of the composite spatial autocorrelation map pattern latent in the geographic distribution of the areas ratio values. Note: the gray scale for the maps is directly proportional to the size of the values mapped.

organized into a table, much like a Microsoft Excel spreadsheet, called a matrix. These values rarely are ± 1; more often, the lower bound is between –1 and –0.5, and the upper bound is between 1 and 1.1. Nevertheless, like the relative values for r, MC values greater than $-1/(n-1)$ indicate positive spatial autocorrelation (e.g., using the ratio MC/MC_{max}, where MC_{max} denotes the maximum possible MC value, 0.25 to 0.50 denotes a weak, 0.50 to 0.70 denotes a moderate, 0.70 to 0.90 denotes a strong, and 0.90 to 1.00 denotes a marked degree), whereas values less than $-1/(n-1)$ indicate negative spatial autocorrelation.

The 'MC' is a covariation (i.e., pairwise products of deviations from the mean) index, and is defined by the preceding right-hand expression. Its sampling distribution can be constructed in one of two ways: randomly sampling from a hypothetical probability distribution or taking the set of observed values as given and constructing all possible permutations of their values across a given surface partitioning. In the former case, present distributional results assume that a normal probability model underlies variable y. In either case, the expected value of the sampling distribution is $-1/(n-1)$, the boundary separating positive and negative spatial autocorrelation.

The standard error formulae for the MC are rather complicated, but can be well approximated by

$$\sqrt{\frac{2}{\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}}}$$

when n is at least 20; the sum $\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}$ counts the number of 1's in the n indicator variables, which equals twice the number of neighbors. This standard error result is not surprising, recalling that the standard error of a sample correlation coefficient relates to n, and that all neighbors are double counted (both i to j and j to i).

The MC values for the preceding empirical examples are 0.78 for the Everglades transformed NDVI, 0.51 for the Cusco transformed population density, and –0.33 for the European areas ratio. Their respective standard deviations are: 0.058 78, 0.000 71, and 0.171 18.

The 'GR', the other popular spatial autocorrelation index, is based upon paired comparisons (i.e., pairwise squared differences), and may be defined as follows:

$$\frac{\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}(y_i - y_j)^2 / \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}}{2\sum_{i=1}^{n}(y_i - \bar{y})^2/(n-1)}$$

This expression, which involves an average squared differences term, also involves the unbiased variance estimate for variable y. A '2' appears in the denominator because of double counting

$$(y_i - y_j)^2 = (y_i - y_j + \bar{y} - \bar{y})^2 = (y_i - \bar{y})^2 + (y_j - \bar{y})^2 - 2(y_i - \bar{y})(y_j - \bar{y})$$

The expected value of GR is 1. It ranges between roughly 0 and 2, with 0–1 signifying positive spatial autocorrelation (i.e., $(y_i - y_j)^2$ goes to 0), and with 1–2 signifying negative spatial autocorrelation (i.e., $(y_i - y_j)^2$ is increasing in magnitude). Again, the actual extreme values are a function of eigenvalues affiliated with the matrix constructed with their spatial weight indicator variables. In addition, 0 can never be reached, because it is associated with the degenerate case of a constant across a geographic landscape, and hence yields a division by 0.

The GR values for the preceding empirical examples are 0.22 for the Everglades transformed NDVI, 0.41 for the Cusco transformed population density, and 1.67 for the European areas ratio.

Relationships between the Moran Coefficient and the Geary Ratio

The MC and GR values tend to give consistent implications when y is a normal random variable. In this case, a useful rule of thumb is that the two coefficients should sum to approximately 1. Dramatic deviations from 1 by this sum suggest that y may be non-normal. Meanwhile, most – but not all – empirical cases yield roughly the same MC and GR values for raw data as well as data transformed to better mimic a bell-shaped curve. And, the MC is the statistically most powerful of the two indices: it does the better job, on average, of differentiating between null and alternative hypotheses. The algebraic relationship between the MC and the GR highlights this feature

$$\mathrm{GR} = \frac{n-1}{n}\left[\frac{\left[\sum_{i=1}^{n}\left(\sum_{j=1}^{n} c_{ij}\right)(y_i - \bar{y})^2\right] / \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}}{\sum_{i=1}^{n}(y_i - \bar{y})^2/n} - \mathrm{MC}\right]$$

This equation reveals that the GR incorporates locational information in addition to that included in the MC. This additional information is the ratio of squared deviations times their number of neighbors, divided by the sample variance. If outliers are present, then this numerator can become excessively large; if an areal unit has a large number of neighbors, then this numerator can be markedly influenced by the corresponding deviation. If variable y conforms to a bell-shaped curve, then this additional information term is approximately 1; as n becomes increasingly large, the ratio $(n-1)/n$ converges on 1. Therefore, adding MC to both sides of this equation would result in $(\mathrm{MC} + \mathrm{GR}) \approx 1$.

The sign for the MC term in this preceding equation is negative, emphasizing the negative relationship between the MC and the GR: MC values approaching 1 correspond to GR values approaching 0 (positive spatial autocorrelation), and MC values approaching −1 correspond to GR values approaching 2 (negative spatial autocorrelation). Zero spatial autocorrelation corresponds to an expected value for the MC of $-1/(n-1)$, which asymptotically converges on 0, resulting in the corresponding GR value converging on 1, its expected value for zero spatial autocorrelation. This negative relationship can be seen simply by constructing a scatterplot of MC and GR values for the three preceding empirical examples.

Graphical Portrayals of Spatial Autocorrelation

Spatial autocorrelation is a concept that lends itself to visualization, including scatterplots and mappings. Various graphical portrayals of this concept can be constructed, including a Moran scatterplot, a semivariogram plot (based upon the GR and interareal unit distances), and a spatial correlogram (based upon MC and GR values for 1st, then 2nd, then 3rd, and so forth, nearest neighbors). The first of these, a Moran scatterplot, can be constructed by: converting georeferenced data values to z-scores (i.e., subtract the mean and then divide by the standard deviation), summing the surrounding z-scores for each areal unit (i.e., $\sum_{j=1}^{n} c_{ij} z_j$), and then plotting these pairs of z-scores (horizontal axis) versus sums of surrounding z-scores (vertical axis). The slope of the resulting trend line is proportional to the MC (i.e., it needs to be divided by $\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}$). Moran scatterplots for each of the three preceding empirical examples appear in **Figure 4**.

The Moran scatterplot portrays $\sum_{j=1}^{n} c_{ij} z_j$ versus z_i, whose trend line highlights the global trend across a given geographic landscape. These sums of neighboring values' quantities also can be visualized with a map. Doing so produces local indices of spatial autocorrelation (LISA) statistics, which enable clusterings on a map to become more conspicuous. Again, hills (i.e., clusters of surrounding values above a mean) and valleys (i.e., clusters of surrounding values below a mean) constitute the patterns of interest. LISA quantities highlight local trends across a given geographic landscape, emphasizing any clusterings in the deviations from the global trend line. These individual contributions to an MC reveal whether spatial autocorrelation essentially is the same in all parts,

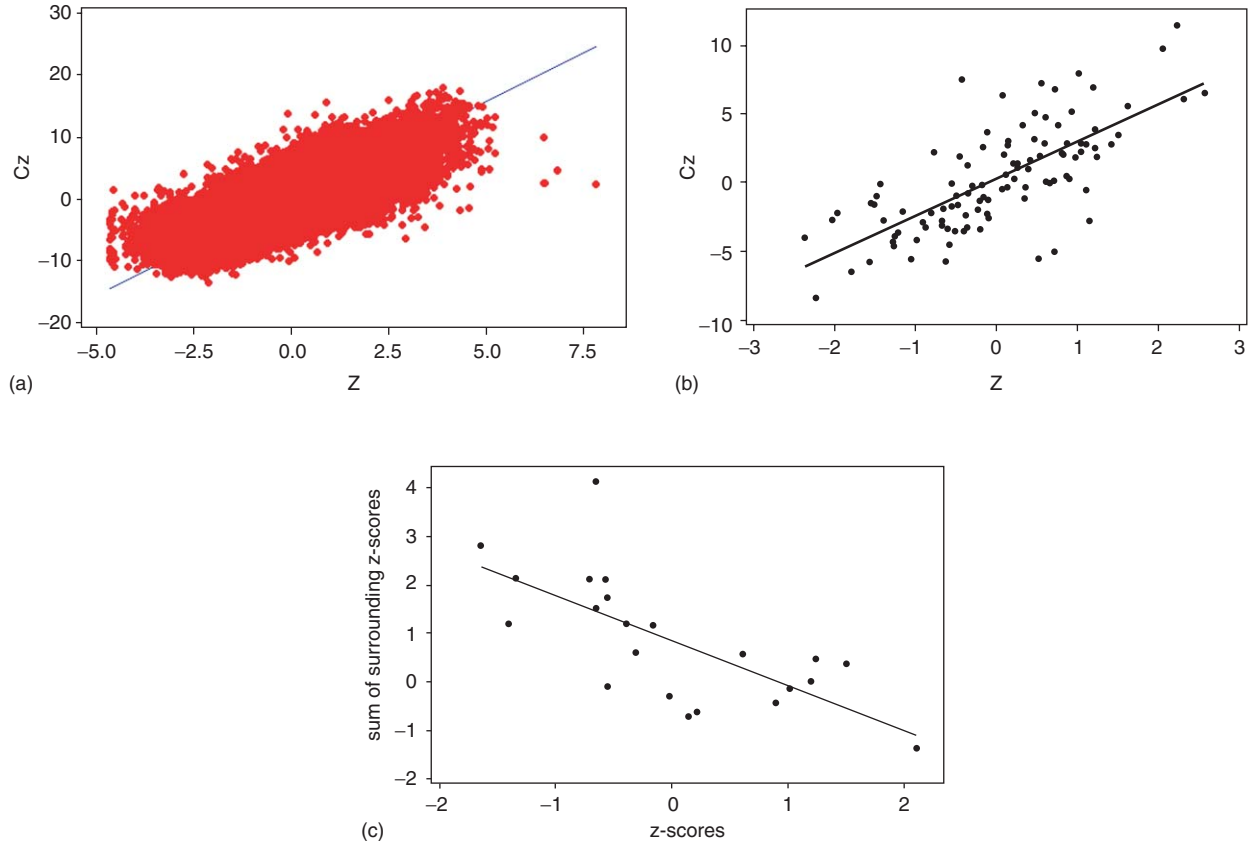

Figure 4 Moran scatterplots for the three empirical examples. (a) Everglades transformed NDVI subregion; (b) Cusco transformed population density; and (c) European areas ratio.

or differs from one part to another part, of a geographic landscape. Local indices can be constructed with a global GR, too, as well as with statistics computed from geostatistical quantities (e.g., the pair of Getis–Ord statistics).

Inspection of a map often suggests the nature and degree of the spatial autocorrelation it contains. This composite map pattern can be viewed in terms of global, regional, and local map pattern components. For example, the composite map pattern associated with the geographic distribution of population density across the Cusco department appears in **Figure 2d**, and comprises 14 individual distinct map patterns, of which 11 are global and regional (strong-to-marked positive spatial autocorrelation), and three are local (weak-to-moderate positive spatial autocorrelation). The most prominent among this selected set of map patterns is a northeast–southwest trend, followed by the hill-shaped map pattern; the most fragmented of these map patterns (i.e., a local pattern) includes nine visually detectable clusters scattered across the department, many of which are partly merged. This composite map pattern accounts for roughly 62% of the geographic variation (i.e., redundant information), and for all but a trace amount of the positive spatial autocorrelation in population density across the department. Similar descriptions can be furnished for the Everglades NDVI and the European areas ratio maps. These local map patterns relate to LISAs. The totality of these map patterns relates to geographically weighted regression (GWR).

Theoretical Statistical Properties of Spatial Autocorrelation

Classical statistics furnishes criteria to distinguish useful from useless spatial autocorrelation measures, including unbiasedness, efficiency, sufficiency, and consistency.

The arithmetic mean of the sampling distribution for an 'unbiased estimator' equals its corresponding population parameter. This property has been used extensively to evaluate the case of zero spatial autocorrelation measures, and the impact of nonzero spatial autocorrelation on conventional sample statistics. For many, but not all, conventional statistical models, sample means, and regression coefficients tend to be unbiased, whereas sample variances and correlation coefficients tend to be biased, by nonzero spatial autocorrelation. Thus, the principal impact of spatial autocorrelation is on standard

error calculations. Because a correlation coefficient is the ratio of a covariance and two standard deviations, spatial autocorrelation impacts in both of these calculations largely cancel each other when the ratio is calculated.

An 'efficiency estimator' is both unbiased and has the smallest possible standard error. Because spatial autocorrelation mostly impacts upon variance calculations, which in turn are used to compute standard errors, this statistical property is the one most affected by nonzero spatial autocorrelation. In the presence of positive spatial autocorrelation, a variance tends to be inflated. The net result is that the corresponding sampling distribution is flatter than traditional statistical theory indicates. Redundant information introduced by spatial autocorrelation results in, for example, sample sizes being misleadingly large. Consequently, more geographic samples are needed to acquire a given margin of error.

A 'sufficient estimator' utilizes all information contained in a sample that is relevant to a particular parameter. Conventional statistics applied to georeferenced data overlook locational information contained in nearby values. Recognition of this information is required for sufficiency to be preserved. For example, an arithmetic mean not only needs $\sum_{i=1}^{n} y_i$, but also the cross-product term $\sum_{i=1}^{n} y_i \sum_{j=1}^{n} c_{ij} y_j$, which is used to measure spatial autocorrelation.

Finally, a 'consistent estimator's' sampling distribution concentrates at the corresponding parameter value as n increases. The efficiency criterion implies that this will occur for unbiased statistics, such as the arithmetic mean, but at a slower rate when spatial autocorrelation is present than is suggested by conventional statistics. Asymptotic concentration will occur even with variance calculations, although not necessarily at the correct value. If geographic sampling intensifies in a given region (infill sampling), then sample points increasingly become closer as n increases, resulting in spatial autocorrelation increasing, and hence concentration increasingly slowing down. If geographic sampling involves expanding a region while maintaining the same average spacing between sample point (increasing domain sampling), then the rate of convergence will be constant rather than decreasing, and thus consistency relies only on the finiteness of all or part of the globe.

Summary and Contemporary Issues

Spatial autocorrelation has many faces, with its most common interpretations being expressed in terms of self-correlation, map pattern, and redundant information. Its preferred measure is the MC, and one of its two most popular graphical portrayals is the associated Moran scatterplot. Few empirical examples of negative autocorrelation have been found, with most empirical examples involving moderate positive spatial autocorrelation; remotely sensed images tend to display strong positive spatial autocorrelation. Impacts of spatial autocorrelation on conventional statistics can be assessed with standard mathematical statistics criteria, such as unbiasedness, efficiency, sufficiency, and consistency.

Today, spatial scientists routinely compute measures of spatial autocorrelation, and rather than test hypotheses about its presence, automatically include it in their model specifications. Doing so often costs only 1 degree of freedom. Spatial autocorrelation with linear models is well understood, and has yielded spatial autoregressive tools used in spatial statistics and spatial econometrics. Spatial autocorrelation with generalized linear (mixed) models is not well understood, with only a few cumbersome tools available to handle it. Spatial filtering, which is in its infancy and exploits the map pattern interpretation, offers an approach that spans both linear and nonlinear statistical models with tools that account for the presence of spatial autocorrelation. This methodology decomposes an underlying composite map pattern into global, regional, and local components of spatial autocorrelation.

The frontiers of spatial autocorrelation research entail fuller development of contemporary techniques such as spatial filtering, and efficient extensions of existing techniques to massively large datasets. For example, the Everglades remotely sensed image contains 41 611 007 land coverage pixels, whereas the simple preceding analysis was challenged by dealing with only 1 000 000 of these pixels. Furthermore, a need still exists for development of quality spatial autocorrelation measures, especially robust ones, for non-normal data.

See also: Regression, linear and non-linear; Segregation indices; Spatial clustering, detection and analysis of; Spatially autoregressive models; Statistics, Spatial.

Further Reading

Anselin, L. (1995). Local indicators of spatial association – LISA. *Geographical Analysis* 27, 93–115.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society B* 36, 192–225.

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

Griffith, D. (1987). *Spatial Autocorrelation: A Primer*. Washington, DC: Association of American Geographers Resource Publication.

Griffith, D. (1992). What is spatial autocorrelation? Reflections on the past 25 years of spatial statistics. *l'Espace Géographique* 21, 265–280.

Griffith, D. (1996). Spatial autocorrelation and eigenfunctions of the geographic weights matrix accompanying geo-referenced data. *The Canadian Geographer* 40, 351–367.

Mardia, K. and Marshall, R. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika* 71, 135–146.

Richardson, S. and Hémon, D. (1981). On the variance of the sample correlation between two independent lattice processes. *Journal of Applied Probability* 18, 943–948.
Tiefelsdorf, M. and Boots, B. (1995). The exact distribution of Moran's I. *Environment and Planning A* 27, 985–999.

Relevant Websites

http://www.ecoevol.ufg.br/sam/
http://www.geoda.uiuc.edu
http://ncg.nuim.ie/ncg/GWR/
http://spatialfiltering.com/