Spatial Statistics: An Overview of Uses

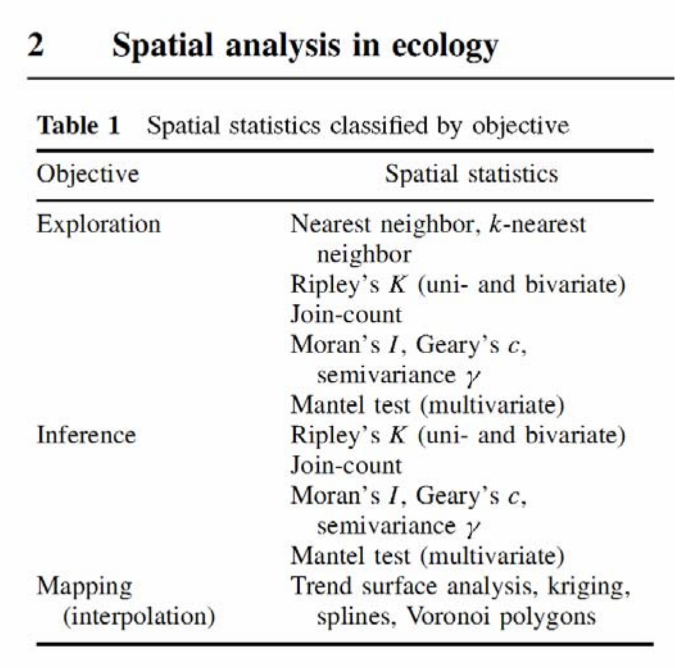
Gavin Archbald

Spatial statistics is the collection of statistical methods in which spatial locations play an explicit role in the analysis of data (Ribeiro and Diggle, 2001). Most often, spatial statistics are used to detect, characterize, and make inferences about **spatial patterns**, primarily in ecology and geography.

Spatial statistics are primarily used to:

1. Describe spatial patterns (**exploration**).
2. Test hypothesis about patterns (**inference**).
3. Predict patterns based on (mapping, or **interpolation**).

The methods of spatial statistics frequently employed in ecological studies are summarized in Table 1, an excerpt from *Spatial Analysis in Ecology* (Fortin et al, 2002):



2 Spatial analysis in ecology

Table 1 Spatial statistics classified by objective

Objective	Spatial statistics
Exploration	Nearest neighbor, *k*-nearest neighbor Ripley's *K* (uni- and bivariate) Join-count Moran's *I*, Geary's *c*, semivariance γ Mantel test (multivariate)
Inference	Ripley's *K* (uni- and bivariate) Join-count Moran's *I*, Geary's *c*, semivariance γ Mantel test (multivariate)
Mapping (interpolation)	Trend surface analysis, kriging, splines, Voronoi polygons

The following discussion expands a selection of the spatial statistics described above.

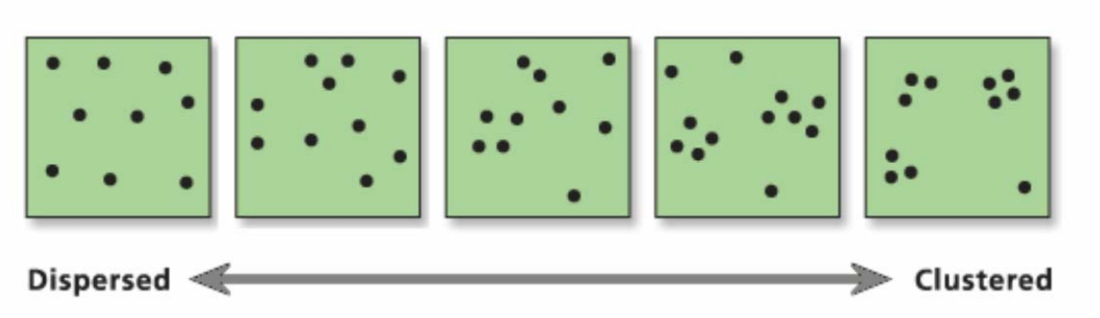
1. Methods to describe spatial patterns (exploration):

Methods have been developed to describe **spatial patterns** of point locations or events. Events are any sort of phenomena which occur at specific locations at a specific time. For example, locations of disease outbreaks or locations of bird nesting sites are events.

Identifying the pattern of events is the first step to understanding the process responsible for that pattern. Events may be **dispersed** from one another, **clustered** together, or occur at **random** (Figure 1). For example, a disease, such as lung cancer, may occur in spatially clustered locations, suggesting a localized environmental factor may be responsible. Linking pattern to process is often the purpose of identifying spatial patterns of events (Legendre, 1993).

Spatial patterns can be identified using **average nearest neighbor** analysis (Fortin et al, 2002). This method measures the mean nearest distance for all points and assumes all points in the study area have been surveyed. Then, the observed mean distance is compared to the expected mean distance under the null hypothesis that the distribution of points is random. The degree to which the observed mean distance varies from the expected mean indicates whether points are dispersed, random or clustered. A z-score is used to test whether the spatial pattern is significantly different from random. A succinct explanation of the math underpinning average nearest neighbor can be found in *Spatial Analysis in Ecology* (Fortin et al, 2002).

Figure 1: The nearest neighbor tool tests whether a particular spatial pattern is significantly different from random. Spatial patterns of events can range from dispersed to clustered.



2. Methods to test hypotheses about patterns (inference):

Tests have been developed to evaluate the degree to which spatial data are autocorrelated. Two measures of spatial autocorrelation of univariate data are commonly used (**Moran's I** and **Geary's c**). Both measure the correlation between the spatial location of sample points and the value at those points. The **Mantel test** can be used to evaluate the degree of spatial autocorrelation of multivariate data, or more generally, to test for correlation between any two similarity matrices. Moran's I and Geary's c are discussed more fully in the sections "Spatial Autocorrelation" and "Spatial Statistics: the Mantel Test and Partial Mantel Test" on this website.

If a data set is determined to be spatially autocorrelated by the methods above, and one wishes to perform a parametric statistical hypothesis test, a correction for spatial autocorrelation is required. An example of how to do this in regression analysis is described in detail in the article *Spatial Autocorrelation and Autoregressive Models in Ecology* (Lichstein et al, 2002). In this paper, the authors demonstrate how basing regression analysis on a model that specifically accounts for spatial autocorrelation increases the degree to which independent variables explain the variation in the dependent variable being evaluated.

Corrections for other parametric statistical tests are discussed in *Spatial Autocorrelation: Trouble or New Paradigm?* (Legendre,1993).

3. Methods to predict patterns (mapping or interpolation):

In contrast to methods developed to analyze the pattern of specific events, a set a tools have been developed to predict the distribution of a continuous variable, such as temperature, from a limited number of sample points. These methods result in maps or surfaces of the continuous variable in which unknown values are **interpolated**, or predicted, between known values.

Interpolation is based on the assumption that the values of the continuous phenomenon being mapped are **spatially autocorrelated** (see “Spatial Autocorrelation”) with their locations. In other words, values at locations nearby are expected to be more similar than values of locations far apart. A mathematical model is “fit” to observed values at known locations. This model can then be used to interpolate values between the known values.

ArcGIS offers a number of methods to interpolate unknown values of a continuous variable. Methods include **kriging, splining**, **inverse distance weighting**, and **trend surface analysis**. These tools all require known values at point locations as inputs, and predict values at unknown locations. These methods differ in that:

The **kriging** method of interpolation is based on a variogram (see “Spatial Autocorrelation”) that measures the degree of spatial autocorrelation between points and uses that measure to predict unknown values. Because predictions are based on a measure of spatial autocorrelation, predicted values have known standard errors.

By contrast, **inverse distance weighting** and **splining** use specified functions to predict how values change with distance. They assume generic spatial autocorrelation, rather than relying on a measure of the specific degree of spatial autocorrelation of the dataset.

The most generalized form of interpolation is **trend analysis** which fits a polynomial to sample points to create a smooth surface. Only coarse scale patterns are captured.

These interpolation methods are described in greater detail and can be preformed easily in ArcGIS.

Resources:

Many spatial statistics operations can be preformed using a **Geographic Information System** (ie. ArcGIS- proprietary or GRASS- free open source). In this overview, all spatial statistics tools available in ArcGIS are described except for the Mantel test. Free trial versions of ArcGIS, which contains an extensive and clearly written searchable help section, are available from the Geography Department at San Francisco State University (http://bss.sfsu.edu/geog/).

Additional spatial statistics software packages, including software to perform the Mantel test can be found at:

A web resource for geostatistic and spatial statistics:
http://www.ai-geostats.org/

A compendium of tutorials, lessons and code for R and more:
http://www.sal.uiuc.edu/

Spatial stats for MATLAB and FORTRAN:
http://www.spatial-statistics.com/software_index.htm

Free GIS!
http://grass.itc.it/

References:

Fortin, Marie-Josee, M. Dale and J. Hoeff, 2002. *Spatial Analysis in Ecology*. Encyclopedia of Environmetrics, 4: 2051-2058.

Legrendre, Pierre, 1993. *Spatial Autocorrelation: Trouble or New Paradigm?* Ecology, 74(6): 1659-1673.

Legrendre, Pierre and M.J. Fortin, 1989. *Spatial pattern and ecological analysis*. Vegetation 80: 107-138.

Lichstein, Jeremy, T.R. Simons, S.A. Shriner and K.E. Franzreb, 2002. *Spatial Autocorrelation and Autoregressive Models in Ecology.* Ecological Monographs 73(3): 445-463.

Ribeiro, Paulo and P.J. Diggle, 2001. *geoR: A Package for Geostatistical Analysis*. http://spatial.nhh.no/R/Rgeo/rnews1.2.15-18.pdf

Spatial Statistics: the Mantel Test and Partial Mantel Test

Spatial statistics is a field of statistics that includes many different tests and techniques. However, a great majority of the techniques are focused on determining the extent to which data are spatially autocorrelated, or to performing hypothesis tests after accounting for spatial autocorrelation. Spatial autocorrelation occurs when observations are not independent of one another because of their spatial arrangement. When data are spatially autocorrelated the value of one observation can be predicted based on adjacent observations. Spatial autocorrelation violates the assumption of independence of observations which is a serious concern for traditional hypothesis tests. Traditionally, random assignment of subjects to treatments is the technique that is most often employed to neutralize the effects of spatial autocorrelation.

ANOVA generally is a common method employed by ecologists to test for treatment effects. If we consider a field experiment examining the response of plants to some treatment and analyzed by ANOVA, then there are two possible outcomes; detection of significant treatment effects or an absence of significant treatment effects. However, when there are significant treatment effects and spatial autocorrelation is believed to be present, then there are three possible reasons for the significant effect: 1) there is no significant spatial autocorrelation and the treatments did affect plant response, 2) the degree of spatial autocorrelation is significant and is introducing spurious treatment effects, or 3) both the degree of spatial autocorrelation and the treatment effects are significant. The Mantel test and Partial Mantel test allow one to distinguish among these three cases by assessing the extent of spatial autocorrelation among subjects.

Mantel Test

The Mantel test computes a correlation between two *n* by *n* distance matrices, where one matrix might represent spatial distances, for example, while the other represents *differences between pairs of plants* in some measure of plant status (e.g., mass). The null hypothesis is that the observed relationship between the two distance matrices could have been obtained by any random arrangement in space (or time, or treatment assignment) of the observations through the study area.

Thus, the results of the test will reveal whether small plants are located near other small plants, while large ones will have large neighbors. The null hypothesis is no relationship between spatial location and plant size.

The matrices will be square and thus the calculations for the test are carried out on only the upper or lower diagonal of the matrices. Again, the values of the two matrices are distance measurements between pairs of values measured in the field (location, mass, height, etc.).

The computation yields a Z statistic:

$$Z = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} B_{ij}$$

where **A** and **B** are the distance matrices.

More commonly this statistic is normalized via a standard normal transformation where the mean of the matrix is subtracted from each element and then each element is divided by the standard deviation. This yields a ***r*** statistic:

$$r = \frac{\sum \sum stdA_{ij} stdB_{ij}}{n-1}$$

The significance of the test statistic can be assessed by one of two methods. A permutation test is recommended with small sample sizes (n< 20). In this case, the rows and columns of the distance matrices are randomly rearranged and the distribution for the test statistic generated though many iterations of this procedure. However, with small sample sizes it is difficult to detect significant spatial patterns. Large sample sizes (n>40) can be tested for significance by an asymptotic *t*-approximation where the test statistic is transformed into a *t* statistic. A significant result indicates spatial autocorrelation.

In addition to testing for spatial autocorrelation of a data set, the Mantel test can be used to test for correlation between other types of multivariate data sets. In these cases, the distance matrix need not be a matrix of geographic distances, but any sort of similarity or dissimilarity matrix. The correlation matrix can also contain any measure of similarity between subjects. For example, if you wanted to ask whether similar physical environments have similar species compositions, the correlation matrix would be a matrix of species similarity and the distance matrix a dissimilarity matrix based on a set of environmental variables. The resulting test statistic would determine the degree to which groups of species composition are correlated with groups of environmental variables. Thus, the Mantel test can be used as a test for multivariate correlation.

A discussion of these additional uses for the Mantel test can be found at:
http://www.nceas.ucsb.edu/scicomp/Dloads/SpatialAnalysisEcologists/SpatialEcologyMantelTest.pdf

Partial Mantel Test

While the Mantel test only allows a comparison among two variables, a Partial Mantel test can be used to compare three or more variables. Essentially, the Partial Mantel test allows a comparison to be made among two variables while controlling for the third. To accomplish this, a third matrix (**C**) is created that can be another variable or a design matrix that refers to the experimental design (such as treatment or not or which of 3 treatments an observation receives). Remember, this matrix is also a matrix of distance (or difference) measurements.

The test statistic is calculated by constructing a matrix of residuals, **A'**, of the regression between **A** and **C**, and a matrix of residuals, **B'**, of the regression between **B** and **C**. The two residual matrices, **A'** and **B'**, are then compared by a standard Mantel test.

Calculations of these test statistics is laborious and is usually carried out by a computer. Pierre Legendre has written a relatively comprehensive spatial statistics program for the Macintosh, "R package: multidimensional analysis, spatial analysis" that is available for free for download from the website listed below. Numerous other programs are also available for download that will also compute a Mantel test.

References:

The spatial statistics literature is notoriously difficult to understand and penetrate. During the course of preparing this section, I consulted many sources, very few of which were of use. The references listed below are the most readable to the sources I consulted.

Cressie, N. A. C., 1993. Statistics for Spatial Data. John Wiley and Sons, Inc.

Fortin, M., and J. Gurevitch. 1993. Mantel Tests: Spatial Structure in Field Experiments: *in* Design and Analysis of Ecological Experiments by Scheiner, S.M., and J. Gurevitch. Chapman & Hall.

Legendre, P. and M. Fortin. 1989. Spatial pattern and ecological analysis. Vegetatio 80:107-138.

Ver Hoef, J.M., and N. Cressie. 1993. Spatial Statistics: Analysis of Field Experiments: *in* Design and Analysis of Ecological Experiments by Scheiner, S.M., and J. Gurevitch. Chapman & Hall.

Websites:

Pierre Legendre's website for spatial statistics and R-package:
http://alize.ere.umontreal.ca/~casgrain/en/labo/R/v3/index.html

Richard Davis' website for his course on Spatial Statistics at Colorado State. Has a very good bibliography. In addition, there is a reader for the course.
http://www.stat.colostate.edu/~rdavis/st523/index.html