# Spatial Analysis in Ecology

## Mantel's Test

### Introduction

A central goal in ecology is to explain the distribution of species in terms of environmental variables presumed to be the operative constraints on the species. This goal, for example, underlies much of classical gradient analysis, niche theory, and much of biogeography. But this goal is confounded by two fundamental issues. First, environmental variables are intercorrelated among themselves, and so it may be difficult to ascribe causal mechanism to a variable even if it can be shown to be correlated with species distribution. Secondly, environmental variables have a characteristic spatial grain (autocorrelation), and so their influence is likely to be expressed only at particular scales of reference. These issues are in addition to the likelihood that the species itself may exhibit patchiness (autocorrelation) in its distribution due to biological processes such as dispersal.

In conventional statistical analyses, the former problem is addressed via multivariate methods that allow one to attend the correlations among predictor variables; partial regression is a familiar solution to this problem. Path analysis is an interpretative approach that allows one to conceptually separate causal relationships from spurious relationships engendered by coincidental correlations among variables. But conventional parametric approaches are confounded by the second issue, namely that autocorrelation in the variables violates the assumptions of parametric analysis.

### Mantel's Test

Mantel's (1967) test is an approach that overcomes some of the problems inherent in explaining species-environment relationships. Mantel's test is a regression in which the variables are themselves distance or dissimilarity matrices summarizing pairwise similarities among sample locations. For example, instead of "abundance of species *x* on plot *i*" the dependent variable might be "similarity in abundance of species *x* on plots *i* and *j*." Likewise, the predictor variable might be "similarity in soil type" between samples instead of "soil type" for a single sample. The operative question is, "Do samples that are similar in terms of the predictor (environmental) variables also tend to be similar in terms of the dependent (species) variable?" One important case that Mantel's test considers explicitly is the case where the predictor variable is space itself, measured as geographic distance. In this case, the question is "Are samples that are close together also compositionally similar?" Reciprocally, "Are samples that are spatially removed (or environmentally dissimilar) from each other also compositionally dissimilar?"

In fact, the power and versatility of Mantel's test stems from the various ways that the distance matrices or the regression itself can be framed.

#### *Distance Metrics*

One advantage of Mantel's test is that, because it proceeds from a distance (dissimilarity) matrix, it can be applied to variables of different logical type (categorical, rank, or interval-scale data). This is especially fortunate for ecologists, who often find themselves working with categorical variables (*e.g.*, soil type); converting these to distance (dissimilarity) metrics also makes the metrics better variables for use in regression. All that matters is that an appropriate distance metric be employed (see Orloci 1978, Legendre and Legendre 1998 for reviews of distance metrics). These metrics, likewise, can be univariate (*e.g.*, "similarity in maple abundance") or multivariate (*e.g.*, using a Sorenson's index or other index of overall species similarity).

Note that because dissimilarity (ecological distance) typically is equivalent to the additive inverse of similarity (*D=1-S*), using similarity (or closeness) instead of dissimilarity (distance) has no qualitative effect on the analysis: it merely changes the sign of the coefficients.

## *Calculations*

Mantel's statistic is based on a simple cross-product term:

$$z = \sum_{i=1}^{n}\sum_{j=1}^{n} x_{ij}y_{ij} \quad (1)$$

and is normalized:

$$r = \frac{1}{(n-1)}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{(x_{ij}-\bar{x})}{s_x} \cdot \frac{(y_{ij}-\bar{y})}{s_y} \quad (2)$$

where *x* and *y* are variables measured at locations *i* and *j* and *n* is the number of elements in the distance matrices (= *m(m-1)/2* for *m* sample locations), and the $s_x$ and $s_y$ are standard deviations for variable *x* and *y*. This standardized equation (2) allows one to consider variables of different measurement units within the same framework, rescaling the statistic to the range of a conventional correlation coefficient bounded on [-1,1]. In practice, a negative Mantel's correlation is rare. The magnitude of correlation is often comparatively small even when highly significant statistically (Dutilleul *et al*. 2000).

Because the elements of a distance matrix are not independent, Mantel's test of significance is evaluated via permutation procedures. In this, the rows and columns of the distance matrices are randomly rearranged. Mantel statistics are recomputed for these permuted matrices, and the distribution of values for the statistic is generated via many iterations (~1000 for $\alpha$=0.05, ~5000 for $\alpha$=0.01, ~10,000 for greater precision; Manly 1991, Legendre 2000).

Note that the Mantel test is based on linear correlation and hence is subject to the same assumptions that beset a common Pearson correlation (*i.e.*, nonlinear relationships between variables may be degraded or lost in the linear correlation). Moreover, the test of spatial dependence is averaged over all distances in the simple Mantel's test, and so this test cannot discover changes in the pattern of correlation at different distances (scales). The Mantel correlogram (below) overcomes this problem.

## *Mantel's Tests: Cases*

Because Mantel's test is merely a correlation between distance matrices and the distance matrices can be variously defined, the test can assume a variety of forms as special cases. These are, in fact, variants of the same case but are interpreted somewhat differently. There are at least six variants.

*Case 1. Simple Mantel's Test on Geographic Distance.* If the dependent distance matrix is species similarity and the predictor matrix is geographic distance ("spatial dissimilarity"), the research question is "Are samples that are close together also compositionally similar?" This is equivalent to testing for overall autocorrelation in the dependent matrix (*i.e.*, averaged over all distances).

*Case 2. Simple Mantel's Test on a Predictor Matrix.* If the dependent matrix is again species similarity and the predictor matrix is a dissimilarity matrix based on a set of environmental variables, then the simple test is for correlation between the two matrices. Such correlation would indicate that locations that are similar environmentally tend to be similar compositionally. This, of course, is one of the fundamental questions in ecology.

*Case 3. Simple Mantel's Test between an Observed Matrix and One Posed by a Model.* As a formal hypothesis test, Mantel's test can be used to compare an observed dissimilarity matrix to one posed by a conceptual or numerical model. Here, the model is provided as a user-provided matrix of similarities or distances, and the test is to summarize the strength of the correspondence between the two matrices. The model distance matrix might be provided as a simple binary matrix of 0's and 1's, or a matrix derived from a more complicated model.

A simple example of a Mantel's test using a model matrix would be a case where the samples are each assigned to a group (*e.g.*, community type) and the predictor variables are measured environmental variables. The question is, "Are samples in the same group (community type) also similar in terms of the environmental variables?" In this case two samples are similar (distance=0) if they are both assigned to the same group, otherwise they are dissimilar (distance=1). A simple Mantel's using this matrix tests group means by comparing among- to within-group dissimilarities—much like an *F*-ratio . This is analogous to the test developed by Clark (1993) and provided in the program ANOSIM.

*Case 4. The Mantel Correlogram*. A special case of case 1 (above) is to partition or subset the analysis into a series of discrete distance class. That is, a first distance matrix is evaluated for all pairs of points within the first distance class; then a second matrix is scored for all pairs of points within the second distance interval, and so on. The result of this analysis is a Mantel's correlogram, completely analogous to an autocorrelation function but performed on a (possibly multivariate) distance matrix.

*Case 5. Partial Mantel's Test on Three Distance Matrices.* The idealized Mantel's test is a partial regression on three distance matrices: species dissimilarity, environmental dissimilarity, and geographic distance ("space") (Legendre and Fortin 1989). Here, the research questions are, "How much of the variability in species composition is explained by the environmental matrix?" and "Is there residual variability in species composition that is spatially structured, after removing the effects of the environmental variables?" The analysis in this case is partial regression (Smouse *et al.* 1986), and both partial correlation (or regression) coefficients are of interest: $r_{YX|S}$ and $r_{YS|X}$ where $Y$ is the dependent matrix (species similarity), $X$ is the predictor matrix (environmental variables), and $S$ is space itself.

*Case 6. Partial Mantel's on Multiple Predictor Variables.* Often, knowing that the environment has some relationship with the dependent variable of interest is not sufficiently satisfying: we wish to know *which* variables are actually related to the dependent variable. The logical extension of Mantel's test is to multiple regression, in which the predictor variables are entered into the analysis as individual distance matrices (Smouse *et al*. 1986, Manly 1986). As a partial regression technique, Mantel's test provides not only an overall test for the relationships among distance matrices, but also tests the contribution of each predictor variable for its pure partial effect on the dependent variable. If geographic location is included as one of the predictor matrices, then the test returns the pure spatial residuals (the effect of "space itself") as well as the partials for each of the predictor variables.

Thus, the flexibility of Mantel's test provides for a wide range of reasonably explicit hypothesis tests. The onus is on the investigator to pose these hypotheses and interpret the analysis in a meaningful way.

## Presentation and Interpretation

By convention, Mantel's test is presented in the framework of path analysis (Leduc *et al.* 1992). In this, the underlying conceptual hypothesis is made explicit: space "causes" environmental variation, environmental variables may "cause" species distribution, and there may be residual spatial variation in the species that is not accounted by the measured environmental variables ("pure spatial" residuals). In fact, these spatial residuals are unaccountable and as such are thus fodder for further study.

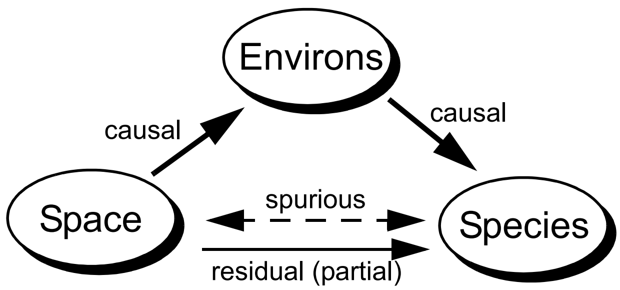


Figure 1. Path diagram of relationships among the physical environment, which is spatially articulated, and the relative abundance of a given species. Here some of the variation in the physical environment governs measured environmental variables, which may influence species abundance. But part of the correlation between space itself and species abundance is spurious, reflecting the correlation caused by the environment; the partial correlation controlling for environment removes this. Finally, the residual partial correlation between species and space may be caused either by species processes or by other unmeasured spatial variables; its source is unknown.

This same information is often presented in tabular form, in which the tabled matrix is split at the diagonal into simple and partial correlations (Legendre and Fortin 1989). The matrix representation of a Mantel's test for three environmental variables would take the form illustrated in Table 1. In this, the upper-diagonal elements are simple correlations and the lower-diagonal elements are partials. In practice, one would table the coefficients as well as their significance levels (*P*-values). A predictor might have a high simple correlation but a much lower (even nonsignificant) partial if it was itself correlated with another predictor variable (a spurious correlation in path analysis), while a variable that maintained a high value even as a partial would be interpreted as a causal factor (note well that this is termed "causal" for interpretation-- correlation cannot ascribe causality under any circumstances).

In the case of multiple predictor matrices, the path diagrams and corresponding tables can get a bit more cumbersome, but the idea is the same. An analysis with three independent predictor matrices plus space itself, presented as a path diagram, would appear as in Figure 2. The corresponding table of coefficients would be a 5x5 matrix (*Y, X1, X2, X3*, and *S* = space), but in practice we are mostly concerned with the relations between the *X*'s and *Y*, the *X*'s and *S*, and the partials ($r_{YX|S}$ and $r_{YS|X's}$).

Table 1. Tabular form for a partial Mantel test with three matrices. Elements in the upper diagonal are simple correlations; the lower, partial correlations.

| | Y | X | space |
|---|---|---|---|
| **Y** | - | $r_{YX}$ | $r_{YS}$ |
| **X** | $r_{YX\|S}$ | - | $r_{XS}$ |
| **space** | $r_{YS\|X}$ | $r_{XS\|Y}$ | - |

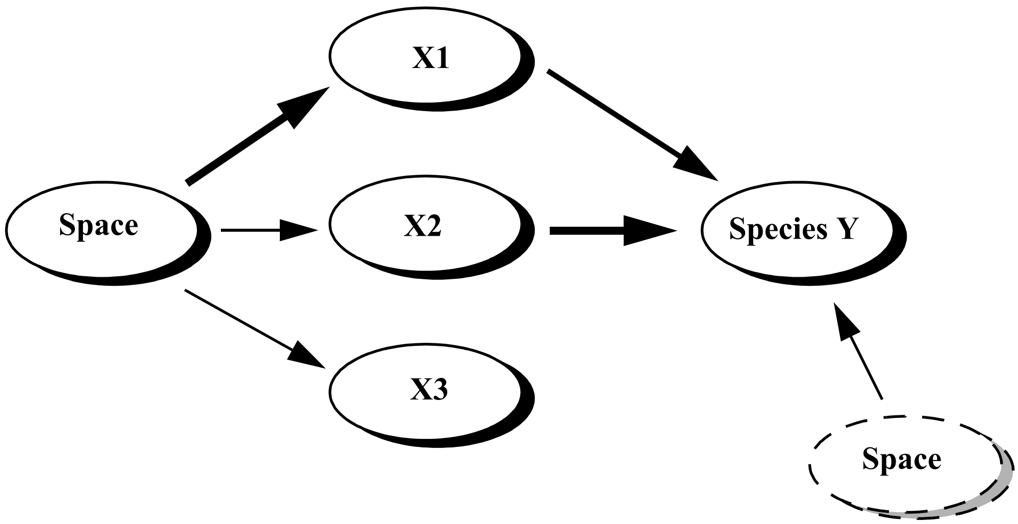


Figure 2. Schematic of a partial Mantel's test in the framework of a path diagram. Thicker lines denote stronger correlations (missing lines are nonsignificant). Lines from Space to the X variables denote spatial structure in the predictors; lines to species Y denote direct effects on the species. A line from Space to the species denotes residual spatial variation in the species not accounted by the X's; this "Space" is ghosted because we cannot know whether the cause is a species process (dispersal, intraspecific competition) or another unmeasured environmental variable.

Path analysis is, of course, a matter of interpretation and it has been used and abused in ecological applications (Petraitis *et al.* 1996). Because path analysis is based on correlation, finding a significant correlation between two variables actually cannot indicate cause; yet the converse *is* true: failing to find a correlation between two variables certainly argues against a causal relationship. Thus, conservatively interpreted, path analysis provides a useful framework for the interpretation of partial regression such as in Mantel's test.

## Suggested Readings

While Mantel (1967) is the original authority on the test, the Legendre lab is responsible to a large extent for a recent popularization of Mantel tests (Legendre and Fortin 1989, Borcard *et al.* 1992, Fortin and Gurevitch 1993, Legendre and Troussellier 1988, Leduc *et al.* 1992; see especially Legendre and Legendre 1998). Manly (1986, 1991, 1997) has also played a strong role. Both the Legendre lab (the "R" package) and Manly ("RT") provide code for Mantel tests, as well—a significant consideration as the procedure is not widely available in common statistical packages. Sokal and Rohlf cover Mantel tests in their (1995) textbook. Partial Mantel tests are typically performed as defined by Smouse *et al.* (1986), and Oden and Sokal (1986) introduced Mantel correlograms (see also Legendre and Fortin 1989). The literature of ecological applications of these tests is rapidly expanding.