Methods in Ecology and Evolution

British Ecological Society

Methods in Ecology and Evolution 2015, **6,** 1239–1247 doi: 10.1111/2041-210X.12425

Should the Mantel test be used in spatial analysis?

Pierre Legendre[1]*, Marie-Josée Fortin[2] and Daniel Borcard[1]

[1]*Département de Sciences Biologiques, Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal, QC, Canada H3C 3J7; and* [2]*Department of Ecology & Evolutionary Biology, University of Toronto, Toronto, ON, Canada M5S 3G5*

Summary

1. The Mantel test is widely used in biology, including landscape ecology and genetics, to detect spatial structures in data or control for spatial correlation in the relationship between two data sets, for example community composition and environment. The study demonstrates that this is an incorrect use of that test.

2. The null hypothesis of the Mantel test differs from that of correlation analysis; the statistics computed in the two types of analyses differ. We examined the basic assumptions of the Mantel test in spatial analysis and showed that they are not verified in most studies. We showed the consequences, in terms of power, of the mismatch between these assumptions and the Mantel testing procedure.

3. The Mantel test H_0 is the absence of relationship between values in two dissimilarity matrices, not the independence between two random variables or data tables. The Mantel R^2 differs from the R^2 of correlation, regression and canonical analysis; these two statistics cannot be reduced to one another. Using simulated data, we show that in spatial analysis, the assumptions of linearity and homoscedasticity of the Mantel test (H_1: small values of $\mathbf{D}_1$ correspond to small values of $\mathbf{D}_2$ and large values of $\mathbf{D}_1$ to large values of $\mathbf{D}_2$) do not hold in most cases, except when spatial correlation extends over the whole study area. Using extensive simulations of spatially correlated data involving different representations of geographic relationships, we show that the power of the Mantel test is always lower than that of distance-based Moran's eigenvector map (dbMEM) analysis and that the Mantel R^2 is always smaller than in dbMEM analysis, and uninterpretable. These simulation results are novel contributions to the Mantel debate. We also show that regression on a geographic distance matrix does not remove the spatial structure from response data and does not produce spatially uncorrelated residuals.

4. Our main conclusion is that Mantel tests should be restricted to questions that, in the domain of application, only concern dissimilarity matrices, and are not derived from questions that can be formulated as the analysis of the vectors and matrices from which one can compute dissimilarity matrices.

Key-words: landscape ecology, landscape genetics, Mantel test, Moran's eigenvector maps (MEM), network analysis, numerical simulations, redundancy analysis, spatially structured data

Introduction

The Mantel test was originally designed for analysing disease clustering in epidemiological studies. In that procedure, Mantel (1967) related a matrix of spatial distances and a matrix of temporal distances in a generalized regression approach. The procedure was expanded by Mantel & Valand (1970) to a nonparametric form of analysis of the relationship between two dissimilarity matrices computed from two sets of multivariate data concerning the same *n* individuals or sampling units. Since that paper, 'the procedure, known as the Mantel test in the biological and environmental sciences, includes any analysis relating two distance matrices or, more generally, two resemblance or proximity matrices' (Legendre 2000).

In biology, Sokal (1979) was the first to use Mantel tests to study patterns of geographic variation in taxonomic data. In Sokal & Rohlf's (1995) *Biometry* book, the Mantel test is presented as a general procedure to test the relationship between multivariate data tables expressed as dissimilarity matrices in biological problems; for these authors, the usefulness of the Mantel test derived from the fact that 'in evolutionary biology and ecology, dissimilarity coefficients are frequently used to measure the degree of difference between individuals, populations, species, or communities' (Sokal & Rohlf 1995, p. 813). A further generalization was proposed by Anselin (1995) who showed that indices of spatial autocorrelation such as Moran's *I* and Geary's *c* may be considered to be special cases of the Mantel statistic.

The discussion and criticisms formulated in this study only concern the spatial analysis applications of the Mantel test in biology (ecology, genetics, evolutionary biology, landscape ecology and landscape genetics). They do not concern the original test developed by Mantel for epidemiological studies, where the question clearly involved the relationship between two types of distances (temporal and spatial) separating disease occurrences.

Applications to spatial analysis started when ecologists and geneticists discovered that a Mantel test offered an easy way of introducing spatial relationships, in the form of a geographic distance matrix, into a statistical framework for modelling

*Correspondence author. E-mail: Pierre.Legendre@umontreal.ca

multivariate data (Sokal 1979; Legendre & Troussellier 1988; Legendre & Fortin 1989; Cushman *et al.* 2006). The Mantel test quickly became a favourite statistical procedure for researchers interested in spatial [or temporal] processes. That was before more appropriate and powerful statistical procedures, such as dbMEM analysis, used in the simulations reported in this study, became available; see Legendre, Borcard & Peres-Neto (2005), Legendre & Fortin (2010) and Dray *et al.* (2012).

The main thesis of this study is that Mantel tests should be restricted to questions that, in the domain of application, only concern dissimilarity matrices, and are not derived from questions that can be formulated as the analysis of 'raw data tables', meaning the vectors and matrices from which one can compute dissimilarity or distance matrices. Matrices of geographic distances among sites derived from spatial coordinates are included in the cases where Mantel tests may be inappropriate. (i) We will show that the hypotheses of correlation tests of significance of raw data tables differ from the hypotheses that concern dissimilarity matrices; furthermore, the statistics involved in the two types of analyses differ and cannot be reduced to one another. (ii) We will refer to simulation papers that have shown that analyses in the world of *raw data* are consistently more powerful than in the world of *dissimilarities* when both approaches are possible. Appendix S1 (Supporting Information) retraces the history of the applications of the Mantel test to spatial data analysis and summarizes the most important simulation studies that have shown that the approach lacks statistical power by a broad margin. (iii) We will focus on the basic assumptions of linearity and homoscedasticity of the Mantel test in spatial analysis. Simulations involving spatially autocorrelated data will show that these assumptions are not verified in most studies. (iv) Finally, using again simulations of spatially autocorrelated data, we will show the consequences, in terms of power, of the mismatch between these assumptions and the Mantel procedure.

Formally, a *dissimilarity index* (or *coefficient*) is a function that measures the difference between two vectors. A *distance index* is a special type of dissimilarity that satisfies the metric properties (minimum value of 0, positiveness, symmetry and triangle's inequality); the Euclidean distance is the most widely used distance coefficient. In this study, the general term *dissimilarity* will be used except to designate a spatial or temporal distance.

What is the null hypothesis of the Mantel test?

Scientists who use Mantel tests when the analysis of raw data tables is possible are usually under the impression that the two types of methods are testing the same statistical hypothesis. For example, Guillot & Rousset (2013) wrote in the caption of their Fig. 2: 'The null hypothesis tested [in the Mantel test] is the independence between *x* and *y*' (in that part of their paper, *x* and *y* are two random variables, cf. their Fig. 1; they are not dissimilarity matrices). That description of the null hypothesis would be correct for the test of a correlation coefficient between two random variables. It is incorrect, however, for the Mantel test, which is a test of the absence of relationship *between the dissimilarities* in two dissimilarity matrices. A correct formulation of H_0 for the Mantel test is the following: 'H_0: *The distances among objects* in matrix $\mathbf{D_Y}$ are not (linearly or monotonically) related *to the corresponding distances* in $\mathbf{D_X}$' (Legendre & Legendre 2012, p. 600; italics added for empha-

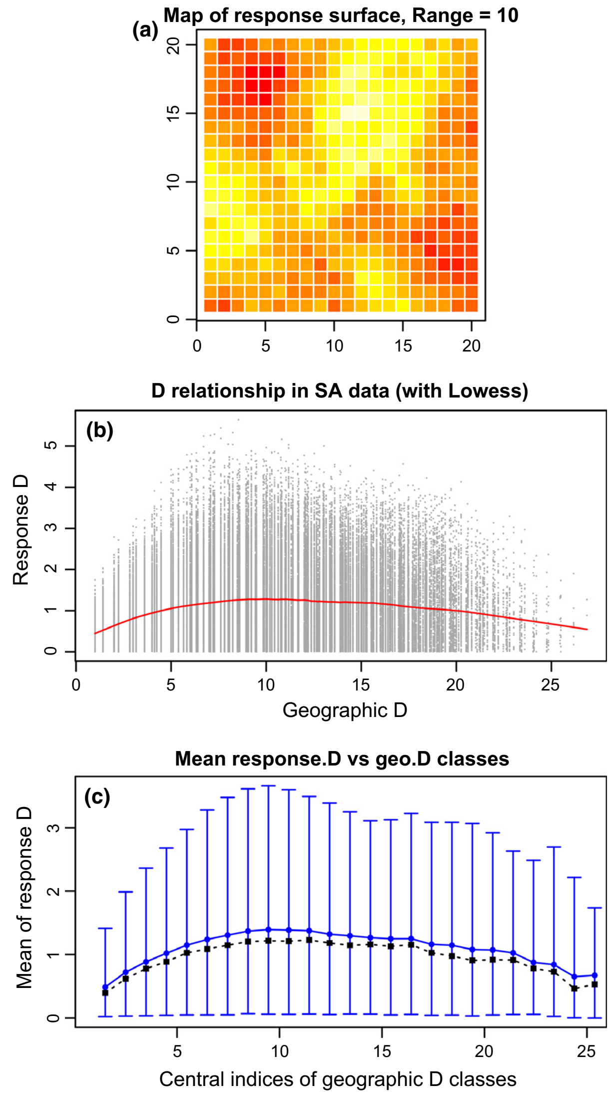

Fig. 1. (a) Map of a 20 × 20 pixel simulated autocorrelated surface. The variogram range controlling the autocorrelation structure was 10. Colour scale: from dark red (low) to pale yellow (high values). (b) Relationship between geographic distances (D) among pixels (abscissa) and dissimilarities (unsigned differences) computed from the simulated data (response D) whose values are represented by colours in panel a. This graph contains $(20^2(20^2 - 1)/2) = 79\ 800$ points (pairs of dissimilarities). Because of point superposition, the **D-D** relationship central tendency is not clear; a Lowess smoother (red line) was added to indicate the central tendency of the relationship across the plot. (c) The geographic distances in the abscissa of panel b are replaced by distance classes; the central tendencies of individual values of the response dissimilarities within classes are represented by their means (blue circles) and medians (black squares); variation is represented by empirical 95% coverage intervals, that is, intervals containing 95% of the response dissimilarities in the class.

sis). Similar formulations of the Mantel null hypothesis are found in Legendre (2000, p. 41): 'The simple Mantel test is a procedure to test the hypothesis that *the distances* among objects in a [distance] matrix **A** are linearly independent of *the distances* among the same objects in another [distance] matrix **B**' and in Legendre & Fortin (2010, p. 835). In partial Mantel test, mentioned in section 'Assumptions of the Mantel test', H_0 states that $\rho(\mathbf{AB.C}) = 0$, where **A**, **B** and **C** are dissimilarity matrices (Legendre 2000).

A complementary point is the demonstration by Legendre & Fortin (2010) (their eqs. 1, 2 and 9) that the statistic used in the Mantel test is unrelated to that used to test the R^2 statistic in [multiple] linear regression or redundancy analysis (RDA), or the simple correlation coefficient *r*. Here, we highlight the difference between the R^2 statistics tested in redundancy analysis (which is the multivariate form of multiple linear regression) and in a Mantel test.

In [multiple] linear regression and RDA, R^2 is the ratio of the sum of squared differences from the mean, or sum of squares (SS) for short, of the fitted values to the sum of squares of the data:

$$R^2 = \frac{\mathrm{SS}(\hat{\mathbf{Y}})}{\mathrm{SS}(\mathbf{Y})} \qquad \text{eqn 1}$$

(following usual notation), whose denominator is

$$\mathrm{SS}(\mathbf{Y}) = \sum\nolimits_{j=1:p} \sum\nolimits_{i=1:n} (y_{ij} - \bar{y}_j)^2 \qquad \text{eqn 2}$$

where *n* is the number of observations and *p* is the number of variables in matrix **Y**. This denominator can also be written as

$$\mathrm{SS}(\mathbf{Y}) = \left(\sum\nolimits_{i>h} {D_{ih}}^2\right)/n \qquad \text{eqn 3}$$

Proof of this equivalence if found in Appendix A1 of Legendre & Fortin (2010). R^2 represents the fraction of the total sum of squares of the response data **Y** that is explained by the explanatory variables **X**.

Consider now two dissimilarity matrices, $\mathbf{D_Y}$ and $\mathbf{D_X}$, computed from data vectors **y** and **x** or from matrices **Y** and **X**. String out the lower diagonal portions of these matrices as long vectors $\mathbf{d_Y}$ and $\mathbf{d_X}$, each of length $n(n-1)/2$. The Mantel correlation, r_M, is the correlation coefficient between these two vectors. The square of r_M is the coefficient of determination R^2_M of the linear regression of $\mathbf{d_Y}$ on $\mathbf{d_X}$:

$$R^2_M = \frac{\mathrm{SS}(\widehat{\mathbf{d_Y}})}{\mathrm{SS}(\mathbf{d_Y})} \qquad \text{eqn 4}$$

The denominator of that equation is

$$\mathrm{SS}(\mathbf{d_Y}) = \sum\nolimits_{i>h} (D_{ih\mathbf{Y}} - \bar{D}_{\mathbf{Y}})^2 = \sum\nolimits_{i>h} D^2_{ih\mathbf{Y}} - \frac{\left(\sum_{i>h} D_{ih\mathbf{Y}}\right)^2}{n(n-1)/2} \qquad \text{eqn 5}$$

This formula is written using dissimilarity values D_{ih} to make it comparable to Eqn 3. The important point here is that $\mathrm{SS}(\mathbf{d_Y})$ in Eqn 5 is not equal to, is not a simple function of, and cannot be reduced to $\mathrm{SS}(\mathbf{Y})$ in Eqn 3. They are different statistics, and so are R^2 and R^2_M.

The statistic used in each test reflects its null hypothesis and, because the null hypotheses differ, the statistics also differ and are not interchangeable. Hence, these two tests are not equivalent. This demonstration completes our proof that the Mantel test is inappropriate to test a hypothesis of correlation between two data vectors or matrices of raw data.

Assumptions of the Mantel test

The Mantel test makes two strong assumptions about the relationships between the two sets of dissimilarities, $\mathbf{D}_1$ and $\mathbf{D}_2$, under comparison.

The first assumption is that the relationship is linear, if a cross-product or a linear correlation coefficient is used as the Mantel statistic, or monotonic if the dissimilarities are replaced by their ranks (Mantel 1967) or if a Spearman or Kendall correlation coefficient is used to compute the Mantel statistic (Dietz 1983). The linearity or monotonicity assumption is linked to the choice of the statistic.

The second assumption, which is the basis for the alternative hypothesis (H_1) of the Mantel test, is that small values of $\mathbf{D}_1$ correspond to small values of $\mathbf{D}_2$ and large values of $\mathbf{D}_1$ to large values of $\mathbf{D}_2$. Mantel stated this assumption (alternative hypothesis H_1) as follows in his 1967 paper (p. 209) in the context of the disease clustering problem: 'if there is time-space clustering, cases in a cluster will be close both in time and space, while unrelated cases will tend to have a larger average separation in time and space'. In their *Biometry* textbook, Sokal & Rohlf (1995, pp. 814 and 816) formulated in similar terms the alternative hypothesis of the Mantel test for specific biological examples.

This assumption may hold for space-time clustering of epidemiological data, but does it hold for the various extensions of the Mantel test that are currently done by biologists? We will show in section 'Simulations involving spatially autocorrelated data: violation of the Mantel test assumptions' that for spatial analysis involving spatially autocorrelated data, that assumption, which refers to the homoscedasticity of the distribution of values in the distance–distance (**D**-**D**) plot, holds in a very limited number of situations; what is found in most cases is a hump-shaped or triangular distribution. This is a novel contribution to the Mantel debate.

Misuse of the Mantel test to analyse georeferenced data

In many applications, researchers incorrectly used the Mantel and partial Mantel tests to assess hypotheses of relationships between variables or data tables, *not* between dissimilarity matrices. A list of examples is found in Legendre, Borcard & Peres-Neto (2005, pp. 438–439). Based on the demonstration reproduced in section 'What is the null hypothesis of the Mantel test?' and on numerical simulations, Legendre & Fortin (2010) argued that Mantel and partial Mantel tests should only be used to test hypotheses that specifically concern dissimilarities, not those derived artificially from hypotheses about the raw data. In particular, to test the correlation between two spatially correlated vectors or matrices of raw data, one cannot use a partial Mantel test computed after transforming the raw data into dissimilarity matrices **A** and **B** and test H_0: ρ(**AB.C**) = 0, where **C** is some form of geographic distance or connexion matrix. There are alternative ways of testing the significance of the correlation between two raw data vectors or matrices while controlling for spatial structure, as shown in Peres-Neto & Legendre (2010) and in Legendre & Legendre (2012).

All simulation studies carried out to measure the capacity of the partial Mantel test to control for (auto)correlation in data have been done by generating raw data that were spatially correlated, for example Manly (1986), Oden & Sokal (1992), Legendre, Borcard & Peres-Neto (2005), Legendre & Fortin (2010), Guillot & Rousset (2013) and section 'Simulations involving spatially autocorrelated data: comparison of Mantel test and dbMEM analysis' of this study. Throughout, the Mantel test was consistently shown to have low power in these simulations, compared to analyses performed on the original data. Appendix S1 reviews some of the papers that showed, through simulations, important characteristics of tests of significance in the presence of spatial correlation, including Mantel and partial Mantel tests.

Simulations involving spatially autocorrelated data: violation of the Mantel test assumptions

Spatially autocorrelated surfaces of different sizes and degrees of autocorrelation were generated by Gaussian random field simulations, using function RFsimulate() of package **RandomFields** (Schlather *et al.* 2014) in R, implementing a spherical variogram model through function RMspheric(). Preliminary results, generated on a small surface (20 × 20 pixels), will be examined first.

Then, larger surfaces were generated in the same way and a subset of points was sampled: on each surface, we selected 100 points forming a square regular grid surrounded by 5-pixel-wide unsampled bands to reduce border effects in the sampled data. The points of the grid were spaced by 1–5 pixels; counting the border bands, the surfaces had {20, 29, 38, 47, 56} pixels in the horizontal and vertical directions, depending on the horizontal and vertical spacing {1, 2, 3, 4, 5} of the sampled points. Results for 5-pixel spacing will be examined. Similar (unreported) results were obtained for the smaller surfaces with horizontal and vertical spacing of 1–4 pixels. The results indicate the following about the assumptions of the Mantel test:

1. Linearity assumption of the **D**-**D** comparison – Let us examine first the response surface simulated on the (20 × 20 pixels) grid with spacing = 1 pixel and autocorrelation range = 10 units (Fig. 1a). The Lowess line in Fig. 1b and the response to distance classes in Fig. 1c show that the dissimilarities increased from geographic distance class 1 to 9 in this example; this is close to the range value (10) of the controlling variogram. The mean of the response dissimilarities decreased as geographic distance increased further. Hence, the **D**-**D** relationship was not linear or monotonic. Similar results are shown in Appendix S2 for larger (56 × 56 pixels) surfaces generated with different variogram range values. The only case where the **D**-**D** relationship was approximately linear was that with range = 70 (Fig. S1q–r), where the autocorrelation range was near the maximum distance between pixels on the surface (i.e. between the pixels in opposite corners, whose geographic distance was 79·2 units).

2. Assumption that small values of $\mathbf{D}_1$ correspond to small values of $\mathbf{D}_2$, and large values of $\mathbf{D}_1$ to large values of $\mathbf{D}_2$ – We will examine whether this assumption holds at least within the sections of the **D**-**D** plots within the range of the controlling variogram. This is the portion between geographic distances 1 and 9 or 10 in Fig. 1b,c. The graph shows that whereas small values of $\mathbf{D}_1$ (response) correspond to small values of $\mathbf{D}_2$ (geographic), an increasingly broad range of response values is associated with larger geographic distances, causing heteroscedasticity in the **D**-**D** distribution. The same absence of homoscedastic **D**-**D** relationships is found for the larger surfaces simulated with various range values (Fig. S2.1). The **D**-**D** relationship on the left of the geographic distance marking the end of the range of autocorrelation of the simulated surface is hump-shaped or triangular and, in any case, very far from homoscedasticity.

For spatially autocorrelated data, these two assumptions of the Mantel test are violated and that partly explains its lack of power. The violations are less important when autocorrelation is equal to or larger than the size of the study area; that is the case where the Mantel test performs best in terms of power, as we will see in the next section.

These two assumptions do not apply to the Mantel correlogram (Oden & Sokal 1986; Sokal 1986; Borcard & Legendre 2012) where the response dissimilarities $\mathbf{D}_1$ are analysed in separate tests against a set of binary model matrices, each representing a geographic distance class.

Simulations involving spatially autocorrelated data: comparison of Mantel test and dbMEM analysis

Despite several papers based on numerical simulations advising to the contrary (Appendix S1), the Mantel test is still widely used by ecologists and geneticists to carry out different forms of spatial analyses. That incentive led us to compare the power of the Mantel test to that of a test based on the original (non-dissimilarity) data, using extensive simulations carried out on the largest spatially autocorrelated surfaces of the previous section.

Spatially autocorrelated data were generated using function RFsimulate(), as in section 'Simulations involving spatially autocorrelated data: violation of the Mantel test assumptions'. The following statistical methods will be compared to study the relationship between the values associated to the points and their geographic positions: (i) the Mantel test between dissimilarity matrices (with one-tailed tests in the upper tail; mantel() function of the **vegan** package, Oksanen *et al.* 2013) and (ii) spatial eigenfunction analysis using the form known as distance-based Moran's eigenvector maps (dbMEM) (PCNM() function of the **PCNM** package, Legendre *et al.* 2012). That method is detailed in Legendre & Legendre (2012, Chapter 14) and in the original publications (Borcard & Legendre 2002;

Borcard *et al.* 2004; Dray, Legendre & Peres-Neto 2006) where it was called PCNM analysis. Spatial eigenfunctions can be used in linear models in the same way as any other set of explanatory variables. The analysis involves multiple linear regression when the response data are univariate (as in our simulation study) or redundancy analysis (RDA, Rao 1964) when it is multivariate. In both cases, R^2 and adjusted R^2 statistics (R^2_{adj}) can be computed and tested for significance using a parametric or permutational F-test (Legendre, Oksanen & ter Braak 2011). A permutational test based upon 999 random permutations of the response data will be used. No variable selection will be carried out in this study; the analyses will be based upon the whole set of eigenfunctions that model positive spatial correlation, that is, those with positive Moran's I coefficients.

In all simulations, 1000 random autocorrelated surfaces with 56×56 pixels were independently produced with variogram ranges of {0, 5, 10, 15, 20, 25, 30, 35, 40} grid units. These surfaces were sampled at 100 points forming a square regular grid with horizontal and vertical spacing of 5 units.

SERIES 1 SIMULATIONS INVOLVING ALL PAIRWISE GEOGRAPHIC DISTANCES

The simulated data sets were analysed with respect to geography using a dbMEM regression and a Mantel test. The truncation value for dbMEM generation was the point spacing, 5 grid units. Users of the Mantel test often square-root the geographic distances to increase the linearity of the relationships with the response dissimilarities, so we carried out our study using both the original and square-rooted geographic distances.

For each range value, the 1000 simulation results were summarized by tallying how many data sets produced significant dbMEM and Mantel results at the $\alpha = 0{\cdot}05$ significance level (one-tailed tests in the upper tail); these numbers were divided by 1000 to obtain rejection rates, which were plotted against the variogram range values (Fig. 2a). Confidence intervals, based on the binomial distribution, were also computed. They are not visible in the graph because they were smaller than the symbols representing the rejection rates.

Each dbMEM regression produced an R^2 and an R^2_{adj} statistic. The means of these R^2 and R^2_{adj} across 1000 simulations were computed for each variogram range value. The means were actually computed on R^2 transformed to $\sqrt{1-R^2}$ and $\sqrt{1-R^2_{adj}}$, which have symmetric distributions, and transformed back to R^2 and R^2_{adj}. Each Mantel test produced an r_{M} statistic, which was transformed to R^2_{M} by squaring it; with this transformation, the Mantel test is considered to be a form of regression analysis, following Mantel (1967). Many users of the Mantel test use that R^2_{M} statistic and erroneously interpret it as if it were equivalent to an R^2 computed by regression on the raw data. Note that there is no way of adjusting R^2_{M} to account for the number of explanatory variables in matrix $\mathbf{D}_2$. Means of the R^2_{M} values were computed as for the dbMEM R^2 and R^2_{adj}. The mean R^2 statistics were plotted against variogram range values (Fig. 2b) together with the mean R^2_{adj} statistics of dbMEM regression.

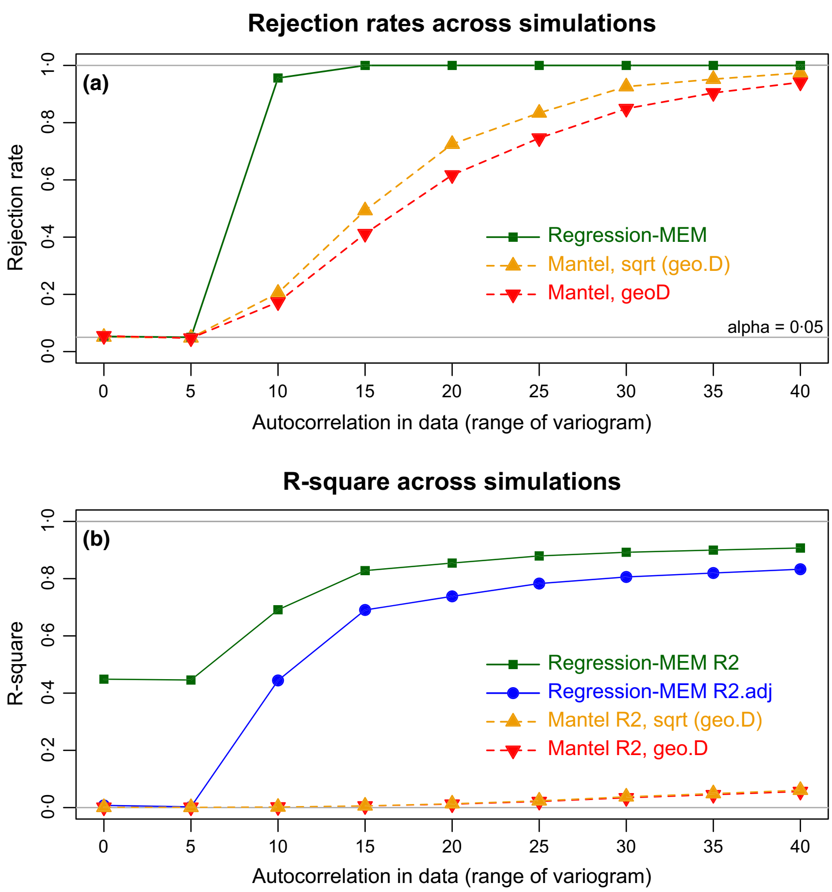

Fig. 2. (a) Rejection rates (i.e. number of rejections of H_0 at the 0·05 significance level divided by the number of simulations, 1000) of the regression–dbMEM and Mantel tests as a function of the variogram range in the simulated data. (b) Mean R-squares of the two methods of analysis. The mean adjusted R-square (R^2_{adj}) of the regression–MEM test, which is an unbiased estimate of the explained variation, is also shown. No R^2_{adj} statistic is available for Mantel tests.

The results (Fig. 2a) show first that the dbMEM analysis and Mantel test had correct levels of type I error; type I error was the rejection rate when there was no spatial autocorrelation in the data (range = 0) or when the range of the variogram used for generation of the data was not larger (range = 5) than the interval between the sampled grid points (here 5 units). This first result has been reported in other papers, for example Oden & Sokal (1992), Legendre & Fortin (2010) and Guillot & Rousset (2013).

When the range of the variogram controlling the autocorrelation in the data was larger than 5, dbMEM analysis was always far more powerful than the Mantel test (Fig. 2a). When the range of the autocorrelation process became very large and the patches nearly covered the whole surface (Fig. S2.1k), the Mantel test became usable although its power remained lower than that of dbMEM analysis. In all cases, the Mantel test based on square-rooted geographic distances was slightly more powerful than the Mantel test based on untransformed geographic distances.

In a regression context, R^2 is a useful measure of the variation of a response variable explained by explanatory data. Fig. 2b shows that the Mantel test R^2 (R^2_{M}) was much smaller than that of dbMEM regression. These two statistics are not comparable: in dbMEM analysis, R^2 measures how much of the variance *of the response data* is explained by geography. In the Mantel test, it measures the fraction of the variance *of the dissimilarities* $\mathbf{D}_1$ explained by the geographic *distances* $\mathbf{D}_2$. Hence, the Mantel R^2_{M} cannot be interpreted as an estimate of the R^2 produced by an analysis of the original data.

In our simulation functions, Mantel tests produced one-tailed tests in the upper tail. This is the normal output of **vegan**'s mantel() function and it was adequate for our study, where we wanted (H_1) to detect positive spatial autocorrelation (SA) in the simulated data when SA was present. We checked, however, what happened in the lower tail. In simulations with variogram ranges of {0, 5}, there was no SA in the data because the spacing between points on the sampled grid was 5; as expected, the rejection rates in the upper and lower tails were always near the significance level, 0·05. When there was SA in the simulated data, the rejection rate in the upper tail increased, as shown in Figs 2–3, while it decreased and became 0 in the lower tail (not shown in the figures). It never went above the significance level.

Legendre & Fortin (2010, their Fig. 4) showed complementary results. They simulated a univariate regular gradient crossing a square map diagonally and added error (noise) to the response data. As the amount of noise increased, power of the methods of analysis decreased, as expected. The comparison involved a linear regression of the response data on the geographic coordinates of the sampled points (i.e. a linear trend surface analysis) and a simple Mantel test. The Mantel test became non-significant after a small amount of noise was added, whereas the *F*-test of the linear regression remained significant for higher amounts of noise. So in that example again, linear regression had higher power than the Mantel test.

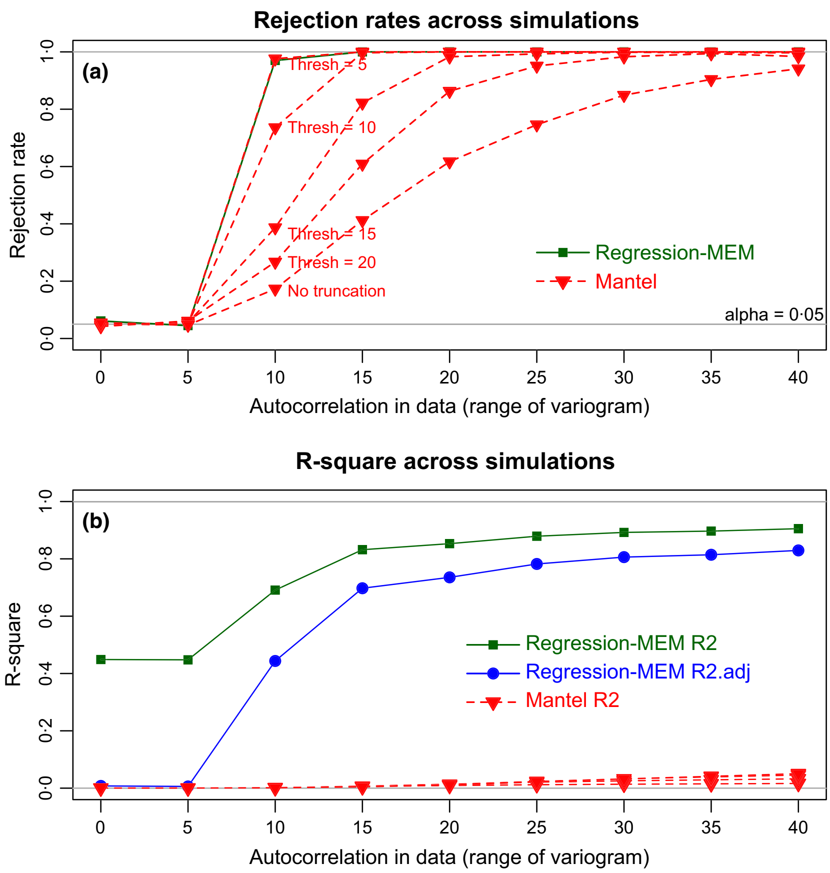

Fig. 3. (a) Rejection rates (i.e. number of rejections of H_0 at the 0·05 significance level divided by the number of simulations, 1000) of the regression–dbMEM and Mantel tests as a function of the variogram range in the simulated data. Mantel tests were computed with truncation levels (*thresh* in figure) of 5, 10, 15 and 20 grid units. (b) Mean *R*-squares of the two methods of analysis. The mean adjusted *R*-square (R^2_{adj}) of the regression–MEM test, which is an unbiased estimate of the explained variation, is also shown. No R^2_{adj} statistic is available for Mantel tests. For the Mantel test, the *R*-squares obtained in the truncated and untruncated simulations are nearly identical and superposed in the figure.

SERIES 2 SIMULATIONS INVOLVING TRUNCATED GEOGRAPHIC DISTANCE MATRICES

In these simulations, the matrix of geographic distances used in Mantel tests was truncated at different levels (thresholds, abbreviated *thresh* = {5, 10, 15, 20} grid units) and all distances larger than the truncation value were changed to the largest distance in the data set, which was the distance between the two opposite corners of the square grid (63·64 units). For each simulation condition (range and *thresh*), the analysis was repeated for 1000 independently generated surfaces. These simulations reproduced the method used by landscape ecologists and geneticists who apply Mantel tests to truncated distance matrices when they feel that the effect of the distance among sites can only be perceived up to a certain distance where contagion, dispersal of propagules in plants, or migration in animals, no longer creates spatial correlation among the sites (Dyer & Nason 2004; Fortuna *et al.* 2009; Murphy *et al.* 2010).

The truncated data, each with 100 observations, were analysed with respect to geography through a dbMEM regression using the full set of eigenfunctions modelling positive spatial correlation, as in series 1, and a Mantel test using the truncated geographic distance matrix (previous paragraph).

Rejection rates of the tests across the simulations are presented in Fig. 3a. For variogram ranges of 0 and 5, where there was no autocorrelation in the data, all tests had correct type I error as their rejection rates were close to the significance level. When the range was larger than 5, dbMEM analysis was always more powerful than the Mantel test for different truncation distance values (*thresh* in the figure), except when the truncation value was 5. The extreme case, with no truncation of geographic distances (or *thresh* larger than the largest distance in the data set), corresponds to the results in Fig. 2a. Hence, when more of the distances are kept (i.e. not truncated) in the geographic matrix, the Mantel test has less power to detect SA in the response data.

That the Mantel test with *thresh* = 5 had power identical to dbMEM analysis may seem surprising. This is because the geographic matrix only contained two different values in that case: $D = 5$ for points that were at that distance, and the largest distance in the data set, $D = 63{\cdot}63961$, for all other pairs of points. This was equivalent to the binary distance matrix used to test for autocorrelation in the first distance class of a Mantel correlogram. Our results thus show that the Mantel test used in this manner, with a single distance class, has the same power for detection of spatial autocorrelation as the dbMEM method of analysis. The simulation study of Borcard & Legendre (2012) had already shown that the test of significance in multivariate Mantel correlograms had high power. That is fine but it does not qualify the Mantel test as the equivalent of dbMEM analysis, which was developed to model the geographic distribution of univariate or multivariate data at different spatial scales, in addition to the production of a test for the presence of spatial correlation in data. In any case, when researchers use Mantel tests with truncated distance matrices, they have a specific ecological or genetic dispersion model in mind and they do not truncate to keep only the first distance class. More about this in the Discussion.

The R^2 results (Fig. 3b) tell the same story as reported in Fig. 2b: the square of the Mantel correlation (R^2_M) is always extremely low.

SERIES 3 SIMULATIONS INVOLVING DELAUNAY TRIANGULATIONS

In the interest of space, simulations involving Delaunay triangulations are described in Appendix S3. The results are essentially the same as those of the Series 2 simulations.

Does the Mantel test capture the spatial variation in response data?

Researchers who use Mantel tests in spatial analysis often assume that the Mantel correlation of a response **D** matrix on a geographic **D** matrix captures the spatial structure that may be present in the response data and, consequently, that regressing response **D** on geographic **D** removes to a large extent the spatial structure from the response **D**, producing residuals without spatial correlation. Appendix S4 shows that this is not the case through a proof-by-example based upon simulated data.

Discussion

This study has shown that there are more implicit assumptions behind the apparently simple decision to run a Mantel test in the context of spatial analysis than meets the eye.

We provided detailed reasons why the Mantel test is inappropriate to study spatial relationships in response data and supported them with numerical simulation results. The reasons invoked are as follows: (i) the hypothesis of correlation tests of significance that concern raw data differs from that concerning dissimilarity matrices; (ii) the statistics involved in the two types of analyses differ and cannot be reduced to one another; (iii) the Mantel test assumes linearity (or monotonicity) and homoscedasticity in the **D**-**D** comparison plots and that is not the case except in extreme cases where the range of spatial correlation is equal to or larger than the size of the study area.

Furthermore, our simulation results showed the following:

1. When the range of the variogram controlling the degree of spatial autocorrelation was larger than the interval between sampled grid points, dbMEM analysis was always far more powerful than the Mantel test (Fig. 2a).

2. The Mantel R^2_M cannot be interpreted as an estimate of the R^2 produced by an analysis of the original response data.

3. In simulations involving truncated distance matrices and Delaunay graph distance matrices, dbMEM analysis was always more powerful than the Mantel test for different values of the truncation distance, except when the truncation value was equal to the interval between the sampled grid points, which created a single distance class with value different from the largest distance in the data set. When more of the distances were kept (i.e. not truncated) in the geographic matrix, the Mantel test had less power to detect SA in the response data.

4. Simulations with a truncation value of 5 were equivalent to a test of the first distance class in a Mantel correlogram; it simply indicated the presence of significant SA in the first distance class. However, when researchers use the Mantel test with truncated distance matrices, they have a specific ecological or genetic dispersion model in mind and they do not truncate to keep only the first distance class.
5. Previous simulations in Legendre, Borcard & Peres-Neto (2005) had shown that spatial variation was, at best, weakly captured by direct regression of a response dissimilarity matrix on a geographic distance matrix. In the present study, we went further and showed that regression on a geographic distance matrix does not control for the spatial structure from response data, and does not produce residuals without spatial correlation.

SHOULD THE MANTEL TEST BE USED IN SPATIAL ANALYSIS?

Our conclusions and recommendations to users for spatial analysis of ecological and genetic data are the following.
1. dbMEM analysis by regression or RDA is a more powerful and informative method of spatial analysis than Mantel tests conducted with distance matrices (truncated or not) or Delaunay triangulations. For one, the tests of significance in dbMEM analysis have much greater power to detect SA in data than Mantel tests. Secondly, dbMEM analysis is a method for modelling the spatial structure in univariate or multivariate response data at different scales; the fitted values of the regression or RDA models can be mapped, providing a visual representation of the structure at different spatial scales; the R^2_{adj} of univariate or multivariate models are unbiased estimates of the portion of the information of the response data explained by the eigenfunctions (Peres-Neto *et al.* 2006). Groups of eigenfunctions representing the variation at different spatial scales can be used in variation partitioning together with other matrices of explanatory variables.
2. Series 2 and 3 simulations showed that when the geographic distance matrix or the Delaunay triangulation are truncated and become binary, the Mantel test becomes identical to a test of the first distance class in a Mantel correlogram and that test has the same power as the test of significance in dbMEM analysis to detect spatial autocorrelation (SA) in response data. The simulation study of Borcard & Legendre (2012) had shown that the Mantel test, used in the context of the Mantel correlogram, had good power to detect SA in data. By opposition, the present series of simulations showed that the ordinary Mantel test has little power to detect SA in data, except in the particular case where a single distance class is studied.

In studies of empirical data, scientists do not know the range of action of SA in the response data. They can use Mantel correlogram analysis to discover it.
3. If ecologists want to use statistical tests to identify SA in field data whose spatial relationships are represented by a Delaunay triangulations or some other connection network, they should check the theoretical framework of their study and decide whether they expect positive or negative SA to be present, and this for each graph distance. If negative SA is expected for some graph distance, they should use dbMEM or Mantel tests accordingly: for dbMEM, they should use only the eigenfunctions that model negative SA, whereas for Mantel analysis, they should look for significance in the lower tail; these *p*-values are equal to or larger than 0·95 in the output of **vegan**'s mantel() function.

To summarize, the Mantel test does not answer the same question and assess the same hypothesis as its raw-data counterparts. When the question concerns the spatial structure of univariate or multivariate data, the lack of concordance of the null hypothesis of the Mantel test with the question produces a test that has low power. In statistics, when several tests of significance are available, one should choose the one that has the highest power, that is, the highest capacity to detect an effect when one is present. The low power of the Mantel test is a symptom of its inadequacy. One should prefer a method with high power, such as dbMEM analysis, to detect spatial structures in data.

PARTIAL MANTEL TEST

In ecology and genetics, many papers used partial Mantel tests to control for spatial structures in the analysis of the relationships between response and environmental data, using a geographic distance matrix as covariable. Oden & Sokal (1992) were the first to demonstrate that partial Mantel tests had inflated type I error rates in analyses of dissimilarity matrices computed from independently autocorrelated data. Guillot & Rousset (2013) repeated the Oden & Sokal study in a more extensive way and came to the same conclusion (Appendix S1). This is likely due to the fact that the partial Mantel test suffers from the same problems as the simple test in the context of spatial analysis: inadequate statistic (Eqn 5), lack of linearity of the relationship, and triangular distribution of the distances.

Users of partial Mantel tests should know that when the question of interest is stated in the world of raw data, the analysis should be performed by partial regression or partial canonical analysis and that these linear forms of partial analysis offer greater power than partial Mantel tests. This is especially true in spatial analysis, where simulation studies have shown that the partial Mantel test is less powerful than partial canonical analysis (Legendre, Borcard & Peres-Neto 2005) and can lead to erroneous conclusions (Oden & Sokal 1992).

SHOULD THE MANTEL TEST BE USED AT ALL?

Mantel tests are valid and useful when applied to the study of relationships among dissimilarities in dissimilarity matrices. Such questions are rarely encountered in ecology and genetics, but they exist; one example is found in Le Boulengé *et al.* (1996). Mantel tests should simply not be used to test hypotheses that concern the raw data from which dissimilarity matrices can be computed or to control for spatial structures in tests of relationships between two autocorrelated data sets.

In population genetics, researchers often use the Mantel method to test hypotheses of isolation by distance (IBD). What is the most appropriate and powerful method to test this hypothesis should be the subject of a separate study. It seems clear, however, that a Mantel correlogram or a multivariate variogram would provide more complete and interesting results than a Mantel test because these analyses would indicate what is the range of the autocorrelation in the data. On the other hand, a dbMEM analysis could be conducted to detect and model the spatial correlation in the genetic data. This is done by computing principal coordinates from the genetic distance matrix and using them as response data in a dbMEM analysis by RDA. After running these analyses, researchers could decide what sets of results are the most useful to answer their landscape genetic question.

Acknowledgements

The authors dedicate this study to the memory of Professor Robert R. Sokal who passed away on 9 April 2012. PL and MJF learned about Mantel and partial Mantel tests in Professor Sokal's laboratory. They also want to thank two anonymous reviewers for stimulating comments. This research was supported by Natural Sciences and Engineering Research Council of Canada (NSERC) research grants to P. Legendre and M.-J. Fortin.

Data accessibility

This study uses simulated spatially autocorrelated data. The R software used for the simulations presented in section 'Simulations involving spatially autocorrelated data: comparison of Mantel test and dbMEM analysis' of the article is shown in Appendix S5.

References

Anselin, L. (1995) Local Indicators of Spatial Association–LISA. *Geographical Analysis*, **27**, 93–115.

Borcard, D. & Legendre, P. (2002) All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling*, **153**, 51–68.

Borcard, D. & Legendre, P. (2012) Is the Mantel correlogram powerful enough to be useful in ecological analysis? A simulation study. *Ecology*, **93**, 1473–1481.

Borcard, D., Legendre, P., Avois-Jacquet, C. & Tuomisto, H. (2004) Dissecting the spatial structure of ecological data at multiple scales. *Ecology*, **85**, 1826–1832.

Cushman, S.A., McKelvey, K.S., Hayden, J. & Schwartz, M.K. (2006) Gene flow in complex landscapes: testing multiple hypotheses with causal modeling. *American Naturalist*, **168**, 486–499.

Dietz, E.J. (1983) Permutation tests for association between two distance matrices. *Systematic Zoology*, **32**, 21–26.

Dray, S., Legendre, P. & Peres-Neto, P.R. (2006) Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling*, **196**, 483–493.

Dray, S., Pélissier, R., Couteron, P., Fortin, M.-J., Legendre, P., Peres-Neto, P.R. *et al.* (2012) Community ecology in the age of multivariate multiscale spatial analysis. *Ecological Monographs*, **82**, 257–275.

Dyer, R.J. & Nason, J.D. (2004) Population graphs: the graph theoretic shape of genetic structure. *Molecular Ecology*, **13**, 1713–1727.

Fortuna, M.A., Albaladejo, R.G., Fernández, L., Aparicio, A. & Bascompte, J. (2009) Networks of spatial genetic variation across species. *Proceedings of the National Academy of Sciences USA*, **106**, 19044–19049.

Guillot, G. & Rousset, F. (2013) Dismantling the Mantel tests. *Methods in Ecology and Evolution*, **4**, 336–344.

Le Boulengé, É., Legendre, P., de le Court, C., Le Boulengé-Nguyen, P. & Languy, M. (1996) Microgeographic morphological differentiation in muskrats. *Journal of Mammalogy*, **77**, 684–701.

Legendre, P. (2000) Comparison of permutation methods for the partial correlation and partial Mantel tests. *Journal of Statistical Computation and Simulation*, **67**, 37–73.

Legendre, P., Borcard, D. & Peres-Neto, P.R. (2005) Analyzing beta diversity: partitioning the spatial variation of community composition data. *Ecological Monographs*, **75**, 435–450.

Legendre, P. & Fortin, M.-J. (1989) Spatial pattern and ecological analysis. *Vegetatio*, **80**, 107–138.

Legendre, P. & Fortin, M.-J. (2010) Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Molecular Ecology Resources*, **10**, 831–844.

Legendre, P. & Legendre, L. (2012) *Numerical Ecology*, 3rd English edn. Elsevier Science BV, Amsterdam.

Legendre, P., Oksanen, J. & ter Braak, C.J.F. (2011) Testing the significance of canonical axes in redundancy analysis. *Methods in Ecology and Evolution*, **2**, 269–277.

Legendre, P. & Troussellier, M. (1988) Aquatic heterotrophic bacteria: modeling in the presence of spatial autocorrelation. *Limnology and Oceanography*, **33**, 1055–1067.

Legendre, P., Borcard, D., Blanchet, F.G. & Dray, S. 2012. *PCNM: MEM spatial eigenfunction and principal coordinate analyses*. R package version 2.1-2. https://r-forge.r-project.org/R/?group_id = 195.

Manly, B.F.J. (1986) Randomization and regression methods for testing for associations with geographical, environmental and biological distances between populations. *Researches on Population Ecology*, **28**, 201–218.

Mantel, N. (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research*, **27**, 209–220.

Mantel, N. & Valand, R.S. (1970) A technique of nonparametric multivariate analysis. *Biometrics*, **26**, 547–558.

Murphy, M.A., Dezzani, R., Pilliod, D.S. & Storfer, A. (2010) Landscape genetics of high mountain frog populations. *Molecular Ecology*, **19**, 3634–3649.

Oden, N.L. & Sokal, R.R. (1986) Directional autocorrelation: an extension of spatial correlograms to two dimensions. *Systematic Zoology*, **35**, 608–617.

Oden, N.L. & Sokal, R.R. (1992) An investigation of three-matrix permutation tests. *Journal of Classification*, **9**, 275–290.

Oksanen, J., Blanchet, G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B. *et al.* (2013) *vegan: Community ecology package*. R package version 2.0-10. http://cran.r-project.org/package = vegan.

Peres-Neto, P.R. & Legendre, P. (2010) Estimating and controlling for spatial structure in the study of ecological communities. *Global Ecology and Biogeography*, **19**, 174–184.

Peres-Neto, P.R., Legendre, P., Dray, S. & Borcard, D. (2006) Variation partitioning of species data matrices: estimation and comparison of fractions. *Ecology*, **87**, 2614–2625.

Rao, C.R. (1964) The use and interpretation of principal component analysis in applied research. *Sankhyaá, Series A*, **26**, 329–358.

Schlather, M., Malinowski, A., Oesting, M., Boecker, D., Strokorb, K., Engelke, S. *et al.* (2014) *RandomFields: Simulation and analysis of random fields*. R package version 3.0.10. http://cran.r -project.org/package = RandomFields.

Sokal, R.R. (1979) Testing statistical significance of geographic variation patterns. *Systematic Zoology*, **28**, 227–232.

Sokal, R.R. (1986) Spatial data analysis and historical processes. *Data analysis and Informatics, IV* (eds E. Diday *et al.*), pp. 29–43. North-Holland, Amsterdam.

Sokal, R.R. & Rohlf, F.J. (1995) *Biometry – The Principles and Practice of Statistics in Biological Research*, 3rd edn. W.H. Freeman and Co., New York.

Received 24 February 2015; accepted 8 June 2015
Handling Editor: Pedro Peres-Neto

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Appendix S1. Analysis of spatially correlated data and Mantel test: who has shown what?.

Appendix S2. Analysis of simulated random autocorrelated surfaces.

Appendix S3. Series 3 simulations involving Delaunay triangulations.

Appendix S4. Regression on a geographic distance matrix does not control for SA in data.

Appendix S5. Software used in the simulations.