

Appendices to:

Legendre P, Fortin MJ (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.

Appendix 1. Two ways of computing SS(Y)

Theorem. Matrix \mathbf{Y} contains data about n points in p -dimensional Euclidean space. The sum of squares of the data in \mathbf{Y} , $SS(\mathbf{Y})$, which is the sum of squares of the distances to the centroid for the group of points, is equal to the sum of squared Euclidean distances among the n points, in the half-matrix of distances \mathbf{D} (with elements D_{ih}), divided by n (Fig. A1.1). In other words,

$$SS(\mathbf{Y}) = \sum_{j=1}^p \sum_{i=1}^n (y_{ij} - \bar{y}_j)^2 = \left(\sum_{i \neq h} D_{ih}^2 \right) / n \quad (1a)$$

$$\text{or} \quad SS(\mathbf{Y}) = \sum_{j=1}^p \sum_{i=1}^n (y_{ij} - \bar{y}_j)^2 = \frac{1}{n} \sum_{i=1}^{n-1} \sum_{h=i+1}^n \sum_{j=1}^p (y_{ij} - y_{hj})^2 \quad (1b)$$

Proof. The data are p -dimensional. The proof is easy to follow in the case $p = 1$ which will be considered first. The left-hand portion of the equation is transformed into a form used for quick calculation of the variance. To simplify the proof, the right-hand part is transformed to incorporate all distances in the whole square distance matrix \mathbf{D} . The sum of squares of all distances in \mathbf{D} must be divided by 2 to be equal to the right-hand member of expression 1b; the distances along the diagonal of \mathbf{D} , which are 0, do not contribute anything to the sum.

$$\sum_{i=1}^n y_i^2 - \frac{1}{n} \left(\sum_{i=1}^n y_i \right)^2 = \frac{1}{2n} \sum_{i=1}^n \sum_{h=1}^n (y_i - y_h)^2 \quad (2)$$

$$\sum_{i=1}^n y_i^2 - \frac{1}{n} \left(\sum_{i=1}^n y_i \right)^2 = \frac{1}{2n} \left[\sum_{i=1}^n \sum_{h=1}^n y_i^2 + \sum_{i=1}^n \sum_{h=1}^n y_h^2 - 2 \sum_{i=1}^n \sum_{h=1}^n y_i y_h \right] \quad (3)$$

$$\sum_{i=1}^n y_i^2 - \frac{1}{n} \left(\sum_{i=1}^n y_i \right)^2 = \frac{1}{2n} \left[n \sum_{i=1}^n y_i^2 + n \sum_{i=1}^n y_i^2 - 2 \left(\sum_{i=1}^n y_i \right)^2 \right] \quad (4)$$

$$\sum_{i=1}^n y_i^2 - \frac{1}{n} \left(\sum_{i=1}^n y_i \right)^2 = \sum_{i=1}^n y_i^2 - \frac{1}{n} \left(\sum_{i=1}^n y_i \right)^2 \quad (5)$$

For the general case where $p > 1$, one only has to rewrite equations 2 to 5 with an additional summation over the p dimensions. ■

Details about the transformation of the last (rightmost) term in Eq. 3 to the corresponding term in Eq. 4: how to go from $\text{Res1} = \sum_{i=1}^n \sum_{h=1}^n y_i y_h$ in Eq. 3 to $\text{Res2} = \left(\sum_{i=1}^n y_i \right)^2$ in Eq. 4.

1. Res1 is the cross-product (scalar product) of two long vectors, each of size $k = n^2$, obtained by inflating vector \mathbf{y} . These vectors can be written as follows in the R language:

$\mathbf{x} = \text{rep}(\mathbf{y}, \text{each}=1, \text{times}=n)$ of length $k = n^2$

$\mathbf{z} = \text{rep}(\mathbf{y}, \text{each}=n, \text{times}=1)$ of length $k = n^2$

$$\text{so that } \text{Res1} = \sum_{i=1}^n \sum_{h=1}^n y_i y_h = \sum_{i=1}^k x_i z_i$$

or, in R: $\text{Res1} = \text{sum}(\mathbf{x} * \mathbf{z})$, or more simply: $\text{Res1} = \text{sum}(\mathbf{y} \%*\% \text{t}(\mathbf{y}))$

Note that the elements in \mathbf{x} and \mathbf{z} are the same; only their order differs.

2. Consider the formula for the covariance of \mathbf{x} and \mathbf{z} :

$$\text{cov}(\mathbf{x}, \mathbf{z}) = \frac{1}{k-1} \sum_{i=1}^k (x_i - \bar{x})(z_i - \bar{z}) = \frac{1}{k-1} \left[\sum_{i=1}^k x_i z_i - \frac{\sum_{i=1:k} x_i \sum_{i=1:k} z_i}{k} \right]$$

$$\sum_{i=1}^k (x_i - \bar{x})(z_i - \bar{z}) = \sum_{i=1}^k x_i z_i - \frac{\sum_{i=1:k} x_i \sum_{i=1:k} z_i}{k}$$

3. Because these two centred vectors $(x_i - \bar{x})$ and $(z_i - \bar{z})$ are orthogonal by construct, their cross-product is 0. We obtain:

$$\sum_{i=1}^k x_i z_i = \frac{\sum_{i=1:k} x_i \sum_{i=1:k} z_i}{n^2}$$

4. $\sum_{i=1:k} x_i = \sum_{i=1:k} z_i$ because these are the same vector with values in different orders. So,

$$\sum_{i=1}^k x_i z_i = \frac{\left(\sum_{i=1:k} x_i \right)^2}{n^2} = \left(\frac{1}{n} \sum_{i=1:k} x_i \right)^2$$

5. $\frac{1}{n} \sum_{i=1:k} x_i = \sum_{i=1:n} y_i$ because each value of \mathbf{y} is repeated n times in \mathbf{x} . So,

$$\sum_{i=1}^n \sum_{h=1}^n y_i y_h = \sum_{i=1}^k x_i z_i = \left(\frac{1}{n} \sum_{i=1:k} x_i \right)^2 = \left(\sum_{i=1:n} y_i \right)^2 = \left(\sum_{i=1}^n y_i \right)^2 = \text{Res2}. \blacksquare$$

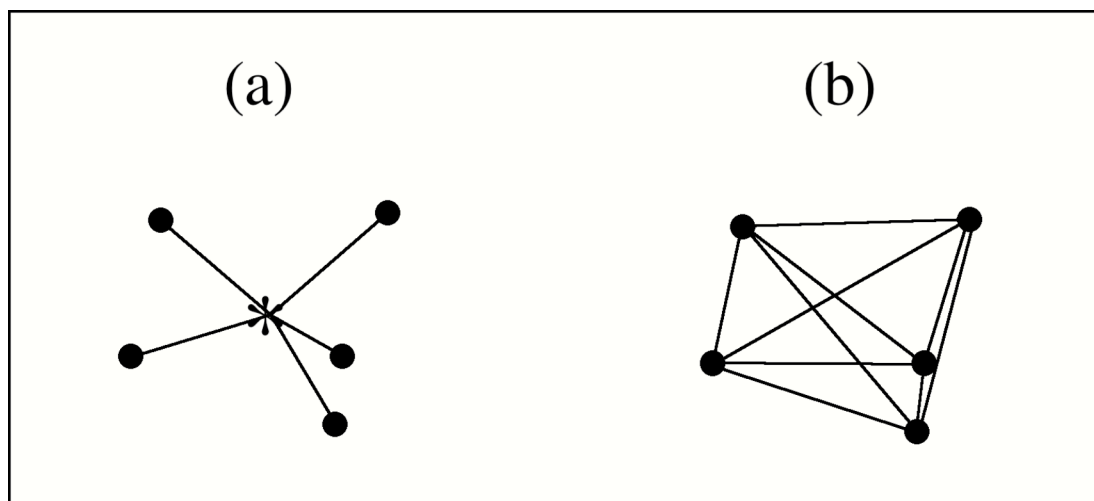


Fig. A1.1 The sum of squares of the distances to the centroid (*) of the group of points (a) is equal to the sum of the squared within-group distances divided by n , where n is the number of points in the group (b).

Appendix 2. Simulations involving multivariate, spatially structured data

The simulations reported here were carried out by Legendre *et al.* (2005) to compare modelling methods in the field of community ecology. The response variables were species presence-absence or abundance data; they are analogous to the gene frequency data analysed in spatial population genetics and landscape genetics. Multivariate response data were generated under the following model:

$$S_{ij} = \beta_{jk} \text{Env}_{ik} + \text{SA}_{Sij} + \epsilon_{Sij}$$

where S_{ij} is the value of species j at site i , Env_{ik} is the value of environmental variable k at site i , β_{jk} transfers the effect of environmental variable k to species j , SA_{Sij} is the added value given to species j by spatial autocorrelation at site i , and ϵ_{Sij} is the “innovation” value for species j at site i (normal error). In plain language, this model says that the spatial variation of a species is the sum of an environmental effect, plus the spatial effect of population processes generating spatial autocorrelation (SA) in the abundances of that species. If β_{jk} is 0, there is no effect of the environmental variables on the species. This model was translated into a simulation program that was first used to study the consequences of spatial structures for the design of ecological field surveys (Legendre *et al.* 2002) and field experiments (Legendre *et al.* 2004), and then in the simulations reported in Legendre *et al.* (2005) and here.

The program was used to generate “surfaces” (i.e., variables mapped in a geographic area) having some specified type of spatial structure. The full surface for data generation was a square grid containing $100 \times 100 = 10000$ points; the units of the grid are pixels of arbitrary size. After generating values on the whole grid, the program sampled it using a square regular grid design with $10 \times 10 = 100$ points with spacing of 10 units between neighbours. Ten species were generated on the surface (data table **Y**): 5 with spatial autocorrelation and random error only, and 5 with spatial autocorrelation, random error, and possibly an environmental effect, depending on the value of the parameter β . The environmental variables (**X**) had a deterministic spatial structure (a weak gradient), spatial autocorrelation, and random error. The table describing the spatial relationships was called **W**. Details of the simulation method are given in Legendre *et al.* (2005). A simulated data set consisted of three data tables: the species frequencies, the values of the environmental variables, and the spatial coordinates of the points on the grid.

Because the response data are multivariate (10 species), canonical redundancy analysis (RDA) was used for linear modelling; RDA is the multivariate equivalent of multiple regression, which can be used to compute linear models for single response variables. Three canonical analyses were used: one to obtain the fraction of the species variation (R^2) explained by the environmental variables (left-hand circle containing fraction [a+b] in Fig. A2.1), a second one for the fraction (R^2) explained by the spatial data (right-hand circle containing fraction [b+c]), and a third one for the fraction (R^2) explained by the environmental and spatial data jointly (union of the two circles, fraction [a+b+c]). The individual fractions [a], [b], [c], and [d] (residual variation) can easily be computed from these preliminary results, an operation called ‘variation partitioning’ described by Borcard *et al.* (1992) and Legendre & Legendre (1998). Peres-Neto *et al.* (2006) have shown that adjusted R^2 (R^2_{adj}) must be used to obtain unbiased estimates of the fractions [a] to [d] in variation partitioning. All these fractions of variation, except [b], can be tested for significance by RDA or partial RDA. The results of these tests of significance can be

used to assess the type I error rate and the power of different testing methods and types of representations of the spatial data.

For each data set, three types of representations of the spatial relationships were used as explanatory variables \mathbf{W} in RDA: (1) the X and Y geographic coordinates of the 100 points forming the sampling grid; (2) a polynomial function of order 3 of the X and Y geographic coordinates, as in Borcard *et al.* (1992); and (3) a set of spatial eigenfunctions (PCNM variables, which represent a spectral decomposition of the spatial relationships among the sites) computed from the X and Y geographic coordinates, as in Borcard & Legendre (2002). The PCNM eigenfunctions related to the response variables were selected by a forward selection procedure. PCNM variables have been shown to be able to identify spatial autocorrelation as well as other types of spatial structures in response data (Borcard & Legendre 2002). All tests of significance of the F and partial F statistics were permutational (999 random permutations).

The same data set were analysed by regression on distance matrices (Hubert & Golledge 1981; tests of significance described in Legendre *et al.* 1994), a generalisation of the Mantel analysis which is used by researchers to produce a form of variation partitioning involving R^2 statistics obtained by linear regression analysis on distances (Svenning 1999; Parris & McCarthy 1999; Duivenvoorden *et al.* 2002). The tests of fractions [a] and [c] were done by partial Mantel tests, a form of analysis that was introduced in population genetics by Smouse *et al.* (1986). For each data set, three types of representations of the spatial relationships were used, as found in the literature on Mantel tests and distance-based variation partitioning: (1) a geographic distance matrix $\mathbf{D}(XY)$ computed from the X and Y geographic coordinates, (2) a distance matrix $\mathbf{D}(\text{polyn.})$ obtained by computing Euclidean distances from the 3rd-order polynomial of the geographic coordinates described in the previous paragraph, and (3) a distance matrix $\ln(\mathbf{D}(XY))$ obtained by taking the natural logarithm of the values in $\mathbf{D}(XY)$. Most users of the Mantel test use $\mathbf{D}(XY)$ for analysing such data. The first and second analyses are comparable for the two methods: the first one uses the X and Y coordinates, raw or in the form of distances, while the second one uses the 3rd-order polynomial function of X and Y , raw or in the form of distances. For raw autocorrelated data, PCNM analysis is likely to account for more variation of the autocorrelated data than the two forms of trend-surface analysis, linear and polynomial. For distance matrices, a logarithmic relationship to geographic distance has been suggested as the most appropriate form for spatially autocorrelated data (Hubbell 2001, Fig. 7.9). One-tailed permutational tests of significance were used for the Mantel and partial Mantel tests (999 random permutations) since this is what most users of the Mantel test would do. Note that the use of one-tailed tests here created an artificial advantage for the Mantel test since the F tests used in RDA have less power than would be obtained if a one-tailed test could be used.

Type I error and power results from these simulations are reported in Table A2.1, which reproduces (with permission) a portion of Table 1 of the Legendre *et al.* (2005) paper. In section 1 of the Table, the species data were unrelated to the environmental variables (the parameter β was 0 for all species), so that the fractions of variation explained by the environmental variables (column [a+b]) reflect the type I error rates of the two methods of analysis and of the various representations of the spatial relationships. All rejection rates in that column are close to the significance level ($\alpha = 0.05$) used in all tests. All testing methods were valid since they had correct rates of type I error. Column [b+c] contains the rejection rates of the tests of significance for the spatial relationships that were present in the species data since they were autocorrelated. The rejection rates indicate that while there is variation among the methods of representation of

the spatial relationships, the tests based on linear models ('RDA' lines) have much higher power than the tests based on distance matrices ('Mantel' lines).

In section 2 of Table A2.1, the species data were forced to be related to the environmental variables (parameter β was 0.5 for 5 of the 10 simulated species). The main difference with section 1 is to be expected in the rejection rates displayed in column [a+b], which shows how the various methods detect the effect of the environmental variables. The results show that the linear analysis ('RDA' lines) detected the environmental influence on the species in about 97% of the simulations, whereas the distance-based analyses ('Mantel' lines) detected this effect in only about 50% of the cases.

References

- 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, Drapeau P (1992) Partialling out the spatial component of ecological variation. *Ecology*, 73, 1045-1055.
- Duivenvoorden JF, Svenning JC, Wright SJ (2002) Beta diversity in tropical forests. *Science*, 295, 636-637.
- Hubbell SP (2001) *The unified neutral theory of biodiversity and biogeography*. Princeton University Press, Princeton, NJ.
- Hubert LJ, Golledge RG (1981) A heuristic method for the comparison of related structures. *J. Math. Psychol.*, 23, 214-226.
- Legendre P, Borcard D, Peres-Neto PR (2005) Analyzing beta diversity: partitioning the spatial variation of community composition data. *Ecological Monographs*, 75, 435-450.
- Legendre P, Dale MRT, Fortin M-J, Casgrain P, Gurevitch J (2004) Effects of spatial structures on the results of field experiments. *Ecology*, 85, 3202-3214.
- Legendre P, Dale MRT, Fortin M-J, Gurevitch J, Hohn M, Myers D (2002) The consequences of spatial structure for the design and analysis of ecological field surveys. *Ecography*, 25, 601-615.
- Legendre P, Lapointe F-J, Casgrain P (1994) Modeling brain evolution from behavior: a permutational regression approach. *Evolution*, 48, 1487-1499.
- Legendre P, Legendre L (1998) *Numerical ecology*, 2nd English edition. Elsevier Science BV, Amsterdam.
- Parris KM, McCarthy MA (1999) What influences the structure of frog assemblages at forest streams? *Australian Journal of Ecology*, 24, 495-502.
- Peres-Neto PR, Legendre P, Dray S, Borcard D (2006) Variation partitioning of species data matrices: estimation and comparison of fractions. *Ecology*, 87, 2614-2625.
- Smouse PE, Long JC, Sokal RR (1986) Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Systematic Zoology*, 35, 627-632.
- Svenning J-C (1999) Microhabitat specialization in a species-rich palm community in Amazonian Ecuador. *Journal of Ecology*, 87, 55-65.

Table A2.1 Rates of rejection of the null hypothesis (H_0 : trace of the fraction is 0) at the $\alpha = 0.05$ level after 1000 simulations (excerpt from Table 1 of Legendre *et al.* 2005). S = species abundances; Env = environmental variables; Space = spatial relationships. XY = geographic coordinates of the sites. Autocorrelation in the species and environmental variables was controlled by a variogram with a range of 15 in a simulation field of size 100 x 100 (arbitrary units).

Partitioning methods ↓	Fractions of variation				
	[a+b+c] Env + Space	[a+b] Env	[b+c] Space	[a] Env Space	[c] Space Env
1. S unrelated to Env ($\beta = 0$), S autocorrelated, Env = $N(0,1)$. [a+b]: type I error. [b+c]: power					
RDA, XY	0.102	0.045	0.191	0.045	0.168
RDA, polynomial	0.261	0.047	0.344	0.045	0.324
RDA, PCNM	0.916	0.056	0.992	0.059	0.987
Mantel, D(XY)	0.106	0.067	0.133	0.066	0.133
Mantel, D(polyn.)	0.068	0.064	0.082	0.063	0.080
Mantel, ln(D(XY))	0.134	0.066	0.222	0.066	0.225
2. S related to Env ($\beta = 0.5$), S autocorrelated, Env autocorrelated. [a+b] and [b+c]: power					
RDA, XY	0.963	0.973	0.194	0.964	0.181
RDA, polynomial	0.944	0.971	0.327	0.950	0.304
RDA, PCNM	0.999	0.970	0.993	0.954	0.986
Mantel, D(XY)	0.501	0.487	0.124	0.478	0.121
Mantel, D(polyn.)	0.386	0.492	0.081	0.488	0.081
Mantel, ln(D(XY))	0.529	0.488	0.199	0.481	0.188

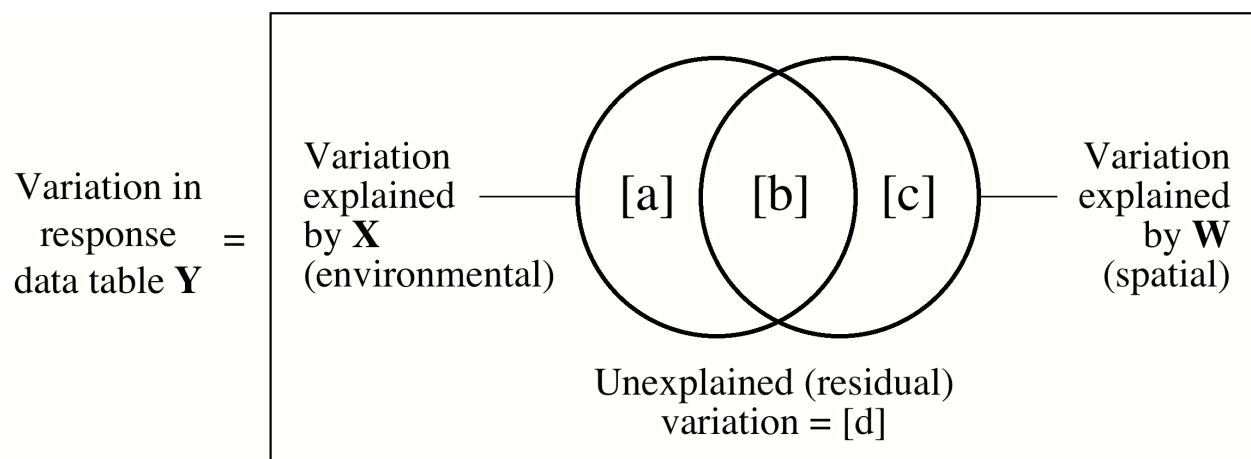


Fig. A2.1 Venn diagram representing the partition of the variation of a response matrix \mathbf{Y} between two sets of explanatory variables \mathbf{X} (environmental) and \mathbf{W} (spatial). The rectangle represents the total variation in \mathbf{Y} . Fraction [b] is the portion explained jointly linear models of \mathbf{X} and \mathbf{W} . Adapted from Legendre (1993).

Appendix 3 – Controversy about the validity of the partial Mantel test

Two papers appeared in *Evolution* in 2001-2002 about the use of partial Mantel tests in micro-evolutionary studies. The first one, by Raufaste & Rousset (2001), raised a valid point about the use of partial Mantel tests with general form of permutation to test for the correlation between two variables in the presence of spatial autocorrelation due to a special form of isolation by distance. The conclusion of that paper left the impression that partial Mantel tests, in general, used an inadequate testing procedure. Castellano & Balletto (2002) responded to that paper, using simulations based upon a more general model to show that the type I error of the partial Mantel test (Manly's 1991 procedure) was correct. Rousset (2002) questioned the appropriateness of the data simulation model used by Castellano & Balletto. The last two papers used inappropriate testing procedures, generating more confusion than they shed light.

These three papers ignored previously published work in which the properties of different forms of partial Mantel tests, with or without spatial autocorrelation, had been analyzed and clearly spelled out (Oden & Sokal 1992; Dutilleul *et al.* 2000; Legendre 2000). These papers, published in the statistical literature, were apparently overlooked by the above-mentioned authors. Using the findings of these papers, plus some new illustrative examples, we point out how partial Mantel statistics should be tested.

Raufaste & Rousset (2001) claimed that partial Mantel tests are not adequate in at least some situations encountered in micro-evolutionary studies. They used a counter-example to show that type I error of the partial Mantel test is inadequate. They positioned N populations of a species, at regular intervals, around a circle on a map, and assumed a non-directional migration flow between neighbouring populations. The response variable X was the abundance of the response species. They introduced an intervening variable Z , the temperature. They wanted to know if there was a relationship between X and Z , taking into account the autocorrelation due to migration around the circle of sites.

The study was concerned with isolation by distance. This is a situation where the null and alternative hypotheses are clearly formulated in terms of distances; so, in principle, the Mantel test is appropriate. The authors showed that random permutation of the values across the sites is an inadequate procedure in this particular case. They did not demonstrate the inadequacy of the Mantel test outside that case. Quite to the contrary: in their Conclusion section, they described the correct way of carrying out restricted permutations in this particular case (around the circle of sites).

The present Appendix provides an opportunity to reemphasize that a test of statistical significance contains three main components: a null hypothesis, a test statistic, and a reference distribution under the null hypothesis to assess the significance of the statistic with respect to the null hypothesis. The null hypothesis in a Mantel test is that the distance matrices are unrelated, in some way that determines the choice of the test statistic. At least three statistics have been proposed: the cross-product of the corresponding distances in the two matrices, a Pearson correlation coefficient, and a Spearman correlation coefficient. The statistic can be tested by some appropriate form of permutation, or, if the number of observations n is large, transformed into a statistic called t by Mantel (1967) and tested using a table of the standard normal distribution. Permutations can be done in different ways, depending on the nature of the data. For the partial Mantel test, Legendre (2000) studied four ways of permuting the distance matrices. For data forming a transect or a loop, as in the Raufaste & Rousset (2001) example, restricting

the permutations in such a way that the data are moved as a ring around the loop and repositioned at random, as proposed by Raufaste & Rousset (2001) in the last paragraph of their paper, is a standard method in data analysis; it has been used for years in the program CANOCO (ter Braak 1990), for example.

In a reply to Raufaste & Rousset (2001), Castellano and Balleto (2002) noted that the case considered by Raufaste & Rousset (2001) for their simulations was very restrictive. They produced a short series of simulations, based upon a more general model, showing that the partial Mantel test had a correct rate of type I error. This had already been shown by the more extensive simulations of Legendre (2000). The simulations reported by Castellano & Balleto (2002) had two pitfalls, however.

(1) For their simulations, they created two rectangular data matrices \mathbf{Y} and \mathbf{X} that were both related to a third rectangular matrix \mathbf{H} . They then computed distance matrices derived from \mathbf{Y} , \mathbf{X} and \mathbf{H} and used these matrices in simple and partial Mantel tests. They noted that the null hypothesis of the Mantel test is the *absence of correlation between the distances* in two \mathbf{D} matrices, but they simulated a *correlation between the original data tables*. These two hypotheses are not equivalent; see next section. So, these authors' simulation procedure does not correspond to the null hypothesis of the Mantel test.

(2) In their simulations, Castellano & Balleto (2002) used two procedures: Manly's (1991) procedure, which they called "the partial Mantel test", and the second Smouse *et al.* (1986) testing procedure, called method 3 in Legendre (2000). (2a) Manly's (1991) procedure for permutation of raw data is incorrect because it uses the partial regression coefficient as the test statistic, instead of the t -statistic associated with the partial regression or partial correlation coefficient. The importance of using, in permutation tests, either the partial correlation coefficient, or the t -statistic associated with the partial correlation or partial regression coefficient, which are pivotal statistics, is discussed in Anderson and Legendre (1999; section 2.1) where references are given. Permutation probabilities obtained using a pivotal r - or t -statistic differ from results obtained from a non-pivotal (partial) regression coefficient b (which was Manly's 1991 procedure). Only the partial r - or t -statistic, which are pivotal, are expected to produce correct type I error in permutation tests in multiple regression. Manly modified his point of view and his procedure in the 1997 edition of his book (p. 180). (2b) For the second Smouse *et al.* (1986) testing procedure (Legendre method 3), Castellano & Balleto (2002) found that it had slightly inflated type I error rate for $n = 20$ objects, a fact that had already been shown by Legendre (2000) who had produced simulations for $n = 5$ to 50 and concluded that this method should not routinely be used. The inflation of type I error for that method asymptotically disappears for large n , so that results published in the past are correct for $n \geq 50$, or when H_0 was not rejected.

Castellano & Balleto (2002) produced additional simulations indicating that the power of the partial Mantel test was affected by correlations between the independent matrices. This situation had been studied in more detail by Legendre (2000). In his Figure 2f, Legendre found no deviation of rate of type I error from the expected value, in the presence of strong correlation between the two independent matrices, for permutation of raw data and permutation of residuals of the null model. In his Figure 6, Legendre found no significant differences in power among the various permutation procedures available for the partial Mantel test (Legendre method 3 was not included in these simulations), for increasing correlation values between the two independent

matrices. That Castellano and Balleto found such an effect is probably due to their inappropriate simulation protocol and/or use of the inappropriate statistic of Manly (1991).

Rousset (2002) questioned the appropriateness of the data simulation model used by Castellano and Balleto, on different grounds than the argument presented above. Rousset (2002) indicates that he also used the incorrect testing procedure of Manly (1991) in the new simulations reported in this paper.

Abusive uses of the Mantel test and derived forms should be questioned, as we did in the main paper, but that questioning should be based on sound statistical arguments. The arguments developed by the authors of the three papers published in *Evolution* did not, in our opinion, offer a sound basis for criticizing Mantel tests. A description of the various permutation methods proposed for partial Mantel tests is given in Appendix 4, together with recommendations about the use of three of these procedures.

References

- Anderson MJ, Legendre P (1999) An empirical comparison of permutation methods for tests of partial regression coefficients in a linear model. *Journal of Statistical Computation and Simulation*, 62, 271-303.
- Castellano S, Balleto E (2002) Is the partial Mantel test inadequate? *Evolution*, 56, 1871–1873.
- Dutilleul P, Stockwell JD, Frigon D, Legendre P (2000) The Mantel test versus Pearson's correlation analysis: assessment of the differences for biological and environmental studies. *Journal of Agricultural, Biological and Environmental Statistics*, 5, 131-150.
- Legendre P (2000) Comparison of permutation methods for the partial correlation and partial Mantel tests. *Journal of Statistical Computation and Simulation*, 67, 37-73.
- Manly BFJ (1991) *Randomization and Monte Carlo methods in biology*. Chapman and Hall, London.
- Manly BJF (1997) *Randomization, bootstrap and Monte Carlo methods in biology*. 2nd edition. Chapman and Hall, London.
- Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research*, 27, 209-220.
- Oden NL, Sokal RR (1992) An investigation of three-matrix permutation tests. *Journal of Classification*, 9, 275-290.
- Raufaste N, Rousset F (2001) Are partial Mantel tests adequate? *Evolution*, 55, 1703-1705.
- Rousset F (2002) Partial Mantel tests: reply to Castellano and Balleto. *Evolution*, 56, 1874-1875.
- Smouse PE, Long JC, Sokal RR (1986) Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Systematic Zoology*, 35, 627-632.
- ter Braak CJF (1990) *Update notes: CANOCO version 3.10*. Agricultural Mathematics Group, Wageningen, The Netherlands.

Appendix 4 – Tests of significance for partial Mantel tests

Proposed by Smouse *et al.* (1986), the partial Mantel test, also called the S-L-S test (Smouse-Long-Sokal), is a form of first-order partial correlation analysis involving three dissimilarity or distance matrices. It is widely used in such fields as evolution, population genetics, ecology, anthropology, psychometry, and sociology. The work of Smouse *et al.* (1986) was illustrated by a study involving matrices of genetic, linguistic, and geographic distances among Yanomama tribes in southern Venezuela and northern Brazil, testing a null hypothesis of genetic isolation by distance against an alternative hypothesis of influence of the linguistic similarity on the genetic similarity among the tribes.

A simulation study for permutation methods in partial Mantel tests

Using simulations, Legendre (2000) investigated the general statistical properties of four testing procedures for partial correlations and partial Mantel tests. The findings of that paper, which appeared in the statistical literature, seem to have been overlooked by evolutionary biologists; see Appendix 3. Thus the main findings are summarized here. Two permutation procedures had originally been proposed by Smouse *et al.* (1986) for the partial Mantel test; they are referred to as methods 1 and 3 in the Legendre (2000) paper and below. Method 3 has been used by many authors who published partial Mantel test results during the past 23 years because it is computationally faster. Methods 2 and 4 were adapted from the literature on canonical analysis. These methods had not been thoroughly validated to detect flaws in type I error or to compare their power. Legendre & Legendre (1998) had, however, raised questions, in particular about method 3, after considering simulation results obtained about this permutation method in the framework of multiple regression analysis (Anderson & Legendre 1999; Anderson & Robinson 2001). The four permutation methods investigated in the Legendre (2000) study will now be described.

Permutation testing procedures used in the partial Mantel simulation study

The permutation methods for partial Mantel tests compared in the Legendre (2000) simulation paper were the following: (1) permute the response distance matrix **A** (application to partial Mantel tests: Smouse *et al.* 1986); (2) permute the residuals of a null regression model (application to multiple regression: Freedman & Lane 1983; to canonical analysis: ter Braak 1990; to partial Mantel tests: Legendre 2000); (3) correlate residualized matrix **A** (**Res_{A|C}**) to residualized matrix **B** (**Res_{B|C}**); permute one of the residualized matrices (application to partial Mantel tests: Smouse *et al.* 1986; to partial regression: Kennedy 1995); (4) permute the residuals of a full regression model (application to partial regression and canonical analysis: ter Braak 1990; to partial Mantel tests: Legendre 2000).

In all cases, the reference value of the partial Mantel statistic $r_M(\mathbf{AB.C})$, estimating the correlation between matrices **A** and **B** while controlling for the effect of **C**, is computed in the same way as a first-order partial correlation coefficient:

$$r_M(\mathbf{AB.C}) = \frac{r_M(\mathbf{AB}) - r_M(\mathbf{AC}) \cdot r_M(\mathbf{BC})}{\sqrt{1 - r_M(\mathbf{AC})^2} \sqrt{1 - r_M(\mathbf{BC})^2}} \quad (1)$$

where $r_M(\mathbf{AB})$ is the simple Mantel correlation between matrices **A** and **B**. The four testing methods are the following. All methods include step 1 which is only described once:

1. Compute the three simple Mantel correlations (also called standardized Mantel statistics) $r_M(\mathbf{AB})$, $r_M(\mathbf{AC})$ and $r_M(\mathbf{BC})$. Combine these values using Eq. 1, obtaining the reference value of the test statistic, $r_M(\mathbf{AB.C})$.

Method 1: permute the response distance matrix **A**

2. [This step is void in method 1, which does not involve residuals.]

3. Permute **A** at random using matrix permutation (described below), obtaining **A***.

4. Compute $r_M(\mathbf{A^*B})$ and $r_M(\mathbf{A^*C})$. Using the value $r_M(\mathbf{BC})$ calculated in step 1, compute $r_M(\mathbf{A^*B.C})$ using Eq. 1, obtaining a value r_M^* of the partial correlation statistic under permutation.

5. Repeat steps 3 and 4 a large number of times to obtain the distribution of r_M^* under permutation. Add the reference value $r_M(\mathbf{AB.C})$ to that distribution.

Method 2: permute the residuals of a null regression model

2. Compute matrix **Res_{A|C}** containing the residuals of the simple linear regression of the distances in **A** over the distances in **C**.

3. Permute **Res_{A|C}** at random using matrix permutation, obtaining **Res*_{A|C}**.

4. Compute $r_M(\mathbf{Res^*_{A|C}B})$ and $r_M(\mathbf{Res^*_{A|C}C})$. Using Eq. 1, combine these values with $r_M(\mathbf{BC})$ computed in step 2, obtaining a value r_M^* of the partial correlation statistic under permutation.

5. Repeat steps 3 and 4 a large number of times to obtain the distribution of r_M^* under permutation. Add the reference value $r_M(\mathbf{AB.C})$ to that distribution.

Method 3: correlate residualized matrix **A** to residualized matrix **B**

2. Compute matrix **Res_{A|C}** containing the residuals of the simple linear regression of the distances in **A** over the distances in **C**. Likewise, compute matrix **Res_{B|C}** containing the residuals of the simple linear regression of the distances in **B** over the distances in **C**.

3. Using matrix permutation, permute **Res_{A|C}** at random to obtain a permuted residual matrix **Res*_{A|C}**. An equivalent method is to permute **Res_{B|C}** at random instead of **Res_{A|C}**, obtaining the permuted matrix **Res*_{B|C}**.

4. Compute the standardized Mantel statistic between **Res*_{A|C}** and **Res_{B|C}**, to obtain a value $r_M^*(\mathbf{AB.C})$ of the test statistic under permutation. An equivalent method is to compute the Mantel statistic between **Res_{A|C}** and **Res*_{B|C}**.

5. Repeat steps 3 and 4 a large number of times to obtain the distribution of $r_M^*(\mathbf{AB.C})$ under permutation. Add the reference value $r_M(\mathbf{AB.C})$ to that distribution.

Method 4: permute the residuals of a full regression model

2. Compute matrix **Res_{A|BC}** containing the residuals of the multiple linear regression of the distances in **A** over the distances in **B** and **C**.

3. Permute **Res_{A|BC}** at random using matrix permutation, obtaining **Res*_{A|BC}**.

4. Compute $r_M(\mathbf{Res}^*_{A|BC}\mathbf{B})$ and $r_M(\mathbf{Res}^*_{A|BC}\mathbf{C})$. Using Eq. 1, combine these values with $r_M(\mathbf{BC})$ computed in step 1, obtaining a value r_M^* of the partial correlation test statistic under permutation.

5. Repeat steps 3 and 4 a large number of times to obtain the distribution of r_M^* under permutation. Add the reference value $r_M(\mathbf{AB.C})$ to that distribution.

All methods include step 6 which is only described once:

6. For a one-tailed test involving the upper tail, calculate the probability as the proportion of values r_M^* greater than or equal to r_M . In the lower tail, the probability is the proportion of values r_M^* smaller than or equal to r_M .

In Mantel tests, the objects are the permutable units under the null hypothesis, not the individual distances. ‘Matrix permutation’ is an algorithm in which the rows and corresponding columns of the matrix are rewritten as if the objects had been permuted in the original rectangular data matrix and the distances recomputed. In computer programs, this rewriting step can be avoided by indirect addressing of the matrix elements, using a vector of permuted object numbers.

Computation note – In method 2, in order to obtain the correct reference estimate of the partial regression coefficient $b_{2.1}$, one must use the original matrix \mathbf{A} as the dependent variable in the regression equation, as described by Anderson & Legendre (1999) in the case of multiple regression on data vectors. However, the correct reference estimate of the partial correlation coefficient is obtained using either the original matrix \mathbf{A} or the residuals of the regression of \mathbf{A} on \mathbf{C} . After permutation of the residuals, the correct estimates of the partial correlation and regression coefficients are obtained using either the permuted residuals alone as the dependent variable, or the permuted residuals added on to the unpermuted fitted values.

Recommendations

The recommended testing procedures are the following (Table A4.1, which was not included in the original publication): (a) In partial Mantel tests, permutation method 2 can always be used, except when highly skewed data are combined with small sample size. Skewness of the distributions can be assessed by examining ordination diagrams of the distance matrices (principal coordinate analysis, Gower 1966). (b) With small sample sizes, one should carefully examine the data before partial Mantel analysis. For highly skewed data, permutation of the response distance matrix \mathbf{A} has correct type I error in the absence of outliers. Outliers can, again, be detected in ordination diagrams. When highly skewed data are combined with outliers in matrix \mathbf{C} (covariable), it is recommended to use permutation of distance matrix \mathbf{A} (method 1). (c) Permutation method 3 should not be routinely used. (d) In general, one should avoid using partial Mantel studies when $n < 20$. Reasonable values are $n \geq 40$ for multinormal data and $n \geq 50$ for highly skewed data. As in all tests of significance, the power of partial Mantel tests increases with the value of n .

These recommendations are based on simulations carried out by Legendre (2000) to measure the type I error and power of these permutation methods, using normal and non-normal data, without and with outlier; an outlier was generated by introducing a high value, 50, in the vector representing the covariable before computing matrix \mathbf{C} . For each situation, 10000 simulations were run (100 000 when $n = 5$); 999 permutations were generated for each test. The

simulations used unrestricted, random, equally-likely (EL) permutations of the data. In some cases, restricted permutations are in order (Manly 1997). Sokal *et al.* (1987) used restricted permutations in Mantel tests. The method of restricted permutations is routinely used in some computer programs, e.g., CANOCO (ter Braak & Smilauer 2002), to control for the sampling design (e.g., transect or regular grid of points in space) or the effect of blocking in field experiments. A non-equally-likely “phylogenetic permutation” (PP) procedure has been described by Lapointe & Garland (2001) to correct for inflated type I error when analysing phylogenetically nonindependent observations using conventional statistical methods. This method could be incorporated into Mantel testing procedures.

What is wrong with permutation method 3?

As mentioned above, permutation method 3 has been used by many authors who published partial Mantel test results during the past 23 years because it is computationally faster. Smouse *et al.* (1986) showed that, after residualizing vector \mathbf{d}_A on \mathbf{d}_C and vector \mathbf{d}_B on \mathbf{d}_C , correlating residualized \mathbf{d}_A to residualized \mathbf{d}_B produces a correct estimate of the partial correlation coefficient $r(\mathbf{d}_A \mathbf{d}_B \cdot \mathbf{d}_C)^1$. Kennedy (1995) presented this method as a simpler way of computing the permutation of residuals of a reduced model (method 2) in multiple linear regression. He demonstrated mathematically that, under permutation, method 3 produces the same estimate of a partial regression coefficient (b) as method 2.

What can go wrong, then, when using this permutation method to test the significance of a partial correlation or regression coefficient, or a partial Mantel statistic? Under permutation, the estimate of the partial regression coefficient is indeed the same in methods 2 and 3, but the corresponding t statistics and the partial correlation coefficients, which are pivotal statistics, are affected differently.

This is illustrated using three simple vectors \mathbf{x}_1 , \mathbf{x}_2 and \mathbf{x}_3 as an example (Table A4.2) instead of distance matrices. Data were generated according to the correlation model, where correlations are imposed onto vectors of random data: the three vectors \mathbf{x}_1 , \mathbf{x}_2 and \mathbf{x}_3 contain random deviates $N(0,1)$, $n = 50$; $r(\mathbf{x}_1 \mathbf{x}_2) = 0.0$, $r(\mathbf{x}_1 \mathbf{x}_3) = 0.5$, $r(\mathbf{x}_2 \mathbf{x}_3) = 0.5$. A permutation \mathbf{x}_1^* was selected where $r(\mathbf{x}_1^* \mathbf{x}_3)$ was fairly large; the value was $r(\mathbf{x}_1^* \mathbf{x}_3) = 0.29318$. When $r(\mathbf{x}_1^* \mathbf{x}_3)$ is small, the difference between methods is hard to detect. Table A4.2 illustrates the following points:

- For the original data, the same (correct) estimates of $r(\mathbf{x}_1 \mathbf{x}_2 \cdot \mathbf{x}_3)$, b_2 and t are obtained using methods 2 and 3.
- For permuted data, method 3 obtains the same estimate of b_2 as method 2. This has been shown by Kennedy (1995) on theoretical grounds.
- For permuted data, methods 2 and 3 do not produce the same estimates of the statistics $r(\mathbf{x}_1 \mathbf{x}_2 \cdot \mathbf{x}_3)$ and t . r or t can both be used in permutation tests because they are both pivotal.

Anderson & Robinson (2001) showed that although the difference between methods 2 and 3 disappears asymptotically, r^2 for method 3 is consistently smaller than or equal to r^2 for method 2 under permutation, so that the observed values r^2 appear more extreme more often for method 3

¹ For symmetric matrices \mathbf{A} , \mathbf{B} and \mathbf{C} , vectors \mathbf{d}_A , \mathbf{d}_B and \mathbf{d}_C contain the upper-diagonal portions of distance matrices \mathbf{A} , \mathbf{B} and \mathbf{C} strung out as long vectors.

than for method 2 when compared to the permuted values, resulting in probabilities which are too small, and thus inflating the rate of type I error.

Another way to look at the problem is to say that method 3 explicitly removes the effect of \mathbf{x}_3 by initial regressions of \mathbf{x}_1 and \mathbf{x}_2 on \mathbf{x}_3 . The relationship between \mathbf{x}_3 and each of the residualized variables is (wrongly) assumed to remain zero during the permutations. This would only be strictly correct for infinite n . Method 2 (permute residuals of a null model), on the contrary, recognizes that a small non-zero correlation may appear between permuted residualized \mathbf{x}_1 and residualized \mathbf{x}_3 , due to the finite sample size, and it takes it into account.

So, method 3 is not an appropriate substitute for method 2 for tests of significance of partial correlation or regression coefficients through a pivotal statistic t or r^2 . And, since method 3 has an inflated rate of type I error in partial correlation and partial regression analysis (Legendre 2000), as well as in partial Mantel tests, it should not be used. Anderson & Robinson (2001) have shown, however, that the inflation of type I error asymptotically disappears for large n , so that results published in the past using method 3 for large sample sizes ($n \geq 50$), or when H_0 was not rejected, are correct.

References

- Anderson MJ, Legendre P (1999) An empirical comparison of permutation methods for tests of partial regression coefficients in a linear model. *Journal of Statistical Computation and Simulation*, 62, 271-303.
- Anderson MJ, Robinson J (2001) Permutation tests for linear models. *Australian and New Zealand Journal of Statistics*, 43, 75-88.
- Freedman D, Lane D (1983) A nonstochastic interpretation of reported significance levels. *Journal of Business & Economic Statistics*, 1, 292-298.
- Gower JC (1966) Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, 53, 325-338.
- Kennedy PE (1995) Randomization tests in econometrics. *Journal of Business & Economic Statistics*, 13, 85-94.
- Lapointe F-J, Garland T Jr (2001) A generalized permutation model for the analysis of cross-species data. *Journal of Classification*, 18, 109-127.
- 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, Legendre L (1998) *Numerical ecology*, 2nd English edition. Elsevier Science BV, Amsterdam.
- Manly BJF (1997) *Randomization, bootstrap and Monte Carlo methods in biology*. 2nd edition. Chapman and Hall, London.
- Smouse PE, Long JC, Sokal RR (1986) Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Systematic Zoology*, 35, 627-632.

Sokal RR., Lengyel IA, Derish PA, Wooten MC, Oden NL (1987) Spatial autocorrelation of ABO serotypes in mediaeval cemeteries as an indicator of ethnic and familial structure. *Journal of Archaeological Science*, 14, 615–633.

ter Braak CJF (1990) *Update notes: CANOCO version 3.10*. Agricultural Mathematics Group, Wageningen, The Netherlands.

ter Braak CJF, Smilauer P (2002) *CANOCO reference manual and CanoDraw for Windows user's guide: software for canonical community ordination (version 4.5)*. Ithaca, NY: Microcomputer Power.

Table A4.1 Recommended testing procedures (✓) in partial Mantel studies. The permutation methods are: (1) permute the response distance matrix **A**, (2) permute the residuals of a null regression model, (4) permute the residuals of a full regression model. Permutation method (3), correlate residualized matrix **A** to residualized matrix **B**, should not be used routinely because it has an inflated rate of type I error for $n < 50$.

n	Error distribution in the data	Outlier present in C	Permutation methods		
			1	2	4
≥ 40	normal	no	✓	✓	✓
< 40	normal	no	✓	✓	
≥ 30	normal	yes		✓	✓
< 30	normal	yes		✓	
≥ 20	highly skewed	no	✓	✓	✓
< 20	highly skewed	no	✓		
≥ 50	highly skewed	yes	✓*	✓	✓
< 50	highly skewed	yes	✓*		

* One-tailed tests using permutation of the response distance matrix **A** are too conservative in this situation. Two-tailed tests using permutation of the response distance matrix **A** have inflated type I error when $n < 50$, but less so than the other testing methods. To alleviate the problem, use a conservative significance level.

Table A4.2 Results of permutation methods 2 and 3 for a permutation, selected for illustration, obtained from three simple random normal vectors \mathbf{x}_1 , \mathbf{x}_2 and \mathbf{x}_3 with correlations $r(\mathbf{x}_1\mathbf{x}_2) = 0.0$, $r(\mathbf{x}_1\mathbf{x}_3) = 0.5$, and $r(\mathbf{x}_2\mathbf{x}_3) = 0.5$; $n = 50$.

Method 2: Permute residuals of reduced model

r	\mathbf{x}_2	\mathbf{x}_3
\mathbf{x}_2	1	0.56825
\mathbf{x}_3	0.56825	1
$\text{Res}(\mathbf{x}_1 \mathbf{x}_3)$	-0.29424	0
$\text{Res}^*(\mathbf{x}_1 \mathbf{x}_3)$	0.28701	0.29318

Original data: $r(\text{Res}(\mathbf{x}_1|\mathbf{x}_3) \mathbf{x}_2.\mathbf{x}_3) = -0.35759$ $b_2 = -0.39494$ $t = -2.62506$

Permutation: $r(\text{Res}^*(\mathbf{x}_1|\mathbf{x}_3) \mathbf{x}_2.\mathbf{x}_3) = 0.15305$ $b_2 = 0.16161$ $t = 1.06177$

Method 3: Correlate residualized \mathbf{x}_1 to residualized \mathbf{x}_2 ; permute residualized \mathbf{x}_1

r	$\text{Res}(\mathbf{x}_2 \mathbf{x}_3)$
$\text{Res}(\mathbf{x}_1 \mathbf{x}_3)$	-0.35759
$\text{Res}^*(\mathbf{x}_1 \mathbf{x}_3)$	0.14633

Original data: $r(\text{Res}(\mathbf{x}_1|\mathbf{x}_3) \text{Res}(\mathbf{x}_2|\mathbf{x}_3)) = -0.35759$ $b_2 = -0.39494$ $t = -2.62506$

Permutation: $r(\text{Res}^*(\mathbf{x}_1|\mathbf{x}_3) \text{Res}(\mathbf{x}_2|\mathbf{x}_3)) = 0.14633$ $b_2 = 0.16161$ $t = 1.01407$
