*J. Statist. Comput. Simul.*, 2000, Vol. 67, pp. 37 – 73

# COMPARISON OF PERMUTATION METHODS FOR THE PARTIAL CORRELATION AND PARTIAL MANTEL TESTS

PIERRE LEGENDRE*

*Département de sciences biologiques, Université de Montréal, C.P. 6128,*
*Succursale Centre-ville, Montréal, Québec H3C 3J7, Canada*

*(Received 16 July 1999; In final form 16 December 1999)*

This study compares empirical type I error and power of different permutation techniques that can be used for partial correlation analysis involving three data vectors and for partial Mantel tests. The partial Mantel test is a form of first-order partial correlation analysis involving three distance matrices which is widely used in such fields as population genetics, ecology, anthropology, psychometry and sociology. The methods compared are the following: (1) permute the objects in one of the vectors (or matrices); (2) permute the residuals of a null model; (3) correlate residualized vector 1 (or matrix **A**) to residualized vector 2 (or matrix **B**); permute one of the residualized vectors (or matrices); (4) permute the residuals of a full model. In the partial correlation study, the results were compared to those of the parametric *t*-test which provides a reference under normality. Simulations were carried out to measure the type I error and power of these permutation methods, using normal and non-normal data, without and with an outlier. There were 10000 simulations for each situation (100000 when $n = 5$); 999 permutations were produced per test where permutations were used. The recommended testing procedures are the following: (a) In partial correlation analysis, most methods can be used most of the time. The parametric *t*-test should not be used with highly skewed data. Permutation of the raw data should be avoided only when highly skewed data are combined with outliers in the covariable. Methods implying permutation of residuals, which are known to only have asymptotically exact significance levels, should not be used when highly skewed data are combined with small sample size. (b) In partial Mantel tests, method 2 can always be used, except when highly skewed data are combined with small sample size. (c) With small sample sizes, one should carefully examine the data before partial correlation or partial Mantel analysis. For highly skewed data, permutation of the raw data has correct type I error in the absence of outliers. When highly skewed data are combined with outliers in the covariable vector or matrix, it is still recommended to use the permutation of raw data. (d) Method 3 should never be used.

---

*Tel.: (514) 343-7591, Fax (514) 343-2293, e-mail: pierre.legendre@umontreal.ca

*Keywords:* Partial correlation, Mantel test, partial Mantel test, permutation test, power, simulation study, type I error.

## 1. INTRODUCTION

Multivariate tables of observations are often condensed into resemblance (or association) matrices among the observations or sampling units (objects), computed using similarity or distance (also called dissimilarity) coefficients. Without loss of generality, the following discussion will focus on distance matrices.

Several forms of data analysis, *e.g.*, clustering and ordination, are based upon distance matrices. Researchers may also wish to compare two or several distance matrices about the same objects in order to test a hypothesis concerning an assumed relationship between matrices. The most widely used method for assessing the relationship between two distance matrices is the Mantel (1967) test; the Mantel statistic is tested either by randomization or through an asymptotic normal approximation. The space-time clustering procedure of Mantel (1967) was originally designed to relate a matrix of spatial distance measures and a matrix of temporal distances in a generalized regression approach. Since Mantel and Valand (1970), 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. Indices of spatial autocorrelation such as Moran's *I* and Geary's *c* coefficients may be considered as special cases of the Mantel test (Anselin, 1995).

For three matrices, Smouse *et al.* (1986) have proposed an extension of the Mantel test to carry out partial correlation analysis in population genetics. The method was first applied to ecological data by Legendre and Troussellier (1988). In population genetics, the first matrix may reflect genetic distances among colonies, the other two matrices representing environmental and geographic distances. In ecology, one may be interested, for instance, in relating a matrix of faunal resemblance among sites, estimated from a species presence-absence or abundance data table, to a matrix of environmental resemblance, while controlling for geographic distances recorded in a third matrix. In partial Mantel analysis, randomization tests are used in all cases (Manly, 1997). The Mantel and partial Mantel tests are prime research tools in such fields as population genetics, ecology, anthropology, psychometry and sociology.

For more than three matrices, multiple regression on distance matrices has been suggested by several authors (Hubert and Golledge, 1981; Smouse *et al.*, 1986; Manly, 1986; Krackhardt, 1988). Legendre *et al.* (1994) have described appropriate randomization testing procedures for different types of dependent matrices (ordinary similarity or distance matrices, ultrametric matrices, and additive-tree matrices).

This paper reports the results of a simulation study undertaken to empirically compare type I error and power of different permutation methods that can be used for testing the significance of the Mantel correlation statistic in simple and partial Mantel tests. Some of these methods have been reported in the literature and represent accepted ways of testing the Mantel and partial Mantel statistics. The present research was motivated by the simulation study of Anderson and Legendre (1999) which highlighted problems encountered in various situations, in multiple regression analysis, when using these same permutation methods. Two other methods of performing partial Mantel tests are described here for the first time; they are derived from the literature on multiple regression; see Section 4. Simulations were carried out to measure the type I error and power of the various permutation methods, using normal and non-normal data, without and with an outlier. Conclusions are drawn as to the procedures which seem appropriate to each situation.

Some statistical questions had to be answered before launching the partial Mantel study. So, a study was first carried out to compare empirical type I error and power of the same permutation techniques, applied to first-order partial correlation coefficients. For normally-distributed data, the parametric normal-theory *t*-test can be used as reference to assess the results of the various permutation methods. The study of partial correlation coefficients will serve to (a) detect anomalies in the permutation techniques under study, in the well-understood context of partial correlations, and (b) explain some of the results that will also be found in the partial Mantel study.

Users of the Mantel and partial Mantel tests usually have one-tailed alternative hypotheses. For this reason, the simulations reported in this paper involve one-tailed tests, except where specified. The upper tail was used for convenience.

## 2. FIRST-ORDER PARTIAL CORRELATION

The well-known formula for a first-order partial correlation coefficient is:

$$r(\mathbf{x}_1\mathbf{x}_2.\mathbf{x}_3) = \frac{r(\mathbf{x}_1\mathbf{x}_2) - r(\mathbf{x}_1\mathbf{x}_3)\,r(\mathbf{x}_2\mathbf{x}_3)}{\sqrt{1 - r(\mathbf{x}_1\mathbf{x}_3)^2}\sqrt{1 - r(\mathbf{x}_2\mathbf{x}_3)^2}} \tag{1}$$

where $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$ are random variables (*i.e.*, data vectors). The value $r(\mathbf{x}_1\mathbf{x}_2.\mathbf{x}_3)$ measures the partial correlation between $\mathbf{x}_1$ and $\mathbf{x}_2$ when $\mathbf{x}_3$ is held constant. A parametric test of significance of partial correlation coefficients can be carried out using a *t*-statistic under the assumption of normality. If this condition is not met, a permutation method should be used to perform the test. The permutation methods compared in this study are:

1. Permute the raw data in vector $\mathbf{x}_1$.
2. Permute the residuals of a null model.
3. Correlate residualized $\mathbf{x}_1$to residualized $\mathbf{x}_2$; permute residualized $\mathbf{x}_1$.
4. Permute the residuals of a full model.

Details of these methods are given in the section on partial Mantel tests. In addition, parametric *t*-tests were computed in the partial correlation study. Under normality, one expects a permutation test to produce approximately the same results as the parametric *t*-test. So, the parametric *t*-test will be used as a reference to assess some important properties of the various permutation methods.

In correlation analysis, because the statistic $t = (r\sqrt{\nu})/\sqrt{1-r^2}$ is monotonic to *r* for any constant value of the number of observations *n*, it follows that *r* and *t* are equivalent statistics in permutation tests. The number of degrees of freedom, $\nu$, is equal to $(n-2)$ for a simple correlation coefficient (also called zero-order correlation; Sokal and Rohlf, 1995) and to $(n-3)$ for a first-order partial correlation coefficient. It could be taken to be $(Ndist-2)$ for a simple Mantel correlation or $(Ndist-3)$ for a first-order partial Mantel correlation, where

$Ndist = n(n - 1)/2$ is the number of distances in a half-matrix of distances involving $n$ objects. The number of degrees of freedom does not affect the probability obtained during a permutation test because this multiplicative constant is invariant under permutation; so, it can be left out of the equation. Permutation results reported in this paper for partial correlations and partial Mantel tests are identical using either the partial $r$ or partial $t$ statistics.

A further point of interest is that the $t$-statistic computed for testing the significance of a partial correlation coefficient $r(\mathbf{x}_1\mathbf{x}_2.\mathbf{x}_3)$ is identical to the $t$-statistic computed for testing the significance of the partial regression coefficient $b_{1.2}$ in the multiple linear regression equation $\hat{\mathbf{x}}_1 = b_0 + b_{1.2}\mathbf{x}_2 + b_{1.3}\mathbf{x}_3$. In partial regression analysis, permutation results obtained using a pivotal $t$-statistic differ from results obtained using a non-pivotal partial regression coefficient $b$. Only the partial $t$-statistic, which is pivotal, is expected to produce correct type I error in permutation tests in multiple regression.

## 3. SIMPLE MANTEL TEST

The simple Mantel test (Mantel, 1967; Mantel and Valand, 1970) is a procedure to test the hypothesis that the distances among objects in a matrix A are linearly independent of the distances among the same objects in another matrix B. The result of this test may be used as support for or against the hypothesis that the process that generated the first set of distances is independent of the process that generated the second set.

The unique feature of the Mantel test is the use of a linear statistic to assess the relationship between two distance matrices. Under the null hypothesis, the objects are the permutable units, not the distances which are not independent of one another; so, for the test of significance, randomization is obtained by permuting the $n$ objects of one of the distance matrices. Instead of recalculating the distances after permuting the objects, an algorithm of ‘matrix permutation’ is used; in this algorithm, 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, even this rewriting step can be avoided by indirect addressing of the matrix elements, using a vector of permuted object numbers. The testing procedure is the following:

1. Consider two symmetric resemblance matrices (similarities or distances) **A** and **B**, of size $(n \times n)$, whose rows and columns correspond to the same set of objects. Compute the Pearson correlation (or, alternatively, the Spearman correlation: Dietz, 1983) between the corresponding elements of the upper-triangular (or lower-triangular) portions of these matrices, obtaining the Mantel correlation (also called the standardized Mantel statistic) $r_M(\mathbf{AB})$, which will be used as the reference value in the test.
2. Permute at random the rows and corresponding columns of one of the matrices, say **A**, obtaining a permuted matrix **A***. This procedure is called 'matrix permutation'.
3. Compute the standardized Mantel statistic $r_M(\mathbf{A^*B})$ between matrices **A*** and **B**, obtaining a value $r^*_M$ of the test statistic under permutation.
4. Repeat steps 2 and 3 a large number of times to obtain the distribution of $r^*_M$ under permutation. Add the reference value $r_M(\mathbf{AB})$ to the distribution (Hope, 1968).
5. For a one-tailed test involving the upper tail (*i.e.*, $H_{1+}$: distances in matrices **A** and **B** are positively correlated), calculate the probability (p-value) as the proportion of values $r^*_M$ greater than or equal to $r_M(\mathbf{AB})$. For a test in the lower tail, the probability is the proportion of values $r^*_M$ smaller than or equal to $r_M(\mathbf{AB})$.

Manly (1997) and Legendre and Legendre (1998) have shown that the exact same p-value would be obtained by using the original statistic proposed by Mantel (1967), which is the cross-product of the two series of distances, instead of the now more generally used Mantel correlation coefficient $r_M$. This is true only for the two-matrix Mantel test and not for the partial Mantel test (below). For symmetric distance matrices, only the upper (or lower) triangular portions are used in the calculations. For non-symmetric matrices, the upper and lower triangular portions are included. The main diagonal elements need not be included in the calculation, but their inclusion does not change the p-value.

One expects the Mantel test to have correct type I error rate for any sample size *n*. Empirical illustration of this point will be provided. On the other hand, users of the Mantel test are often surprised to see that significant values of the simple Mantel statistics do not need to be as large as values of the Pearson correlation coefficient to reach significance. To illustrate this point, power curves will be produced for the simple Mantel test across values of the population correlation $\rho$, using multinormal data, and compared to power curves for the Pearson correlation coefficient, for sample sizes $n = 5$ to 50.

## 4. PARTIAL MANTEL TESTS

A partial Mantel test is a first-order partial correlation analysis conducted on three distance matrices (Smouse *et al.*, 1986). Consider distance matrices **A**, **B**, and **C** computed for three univariate or multivariate data tables. The partial Mantel statistic (*sensu* Smouse *et al.*, 1986), $r_M(\mathbf{AB.C})$, estimating the correlation between matrices **A** and **B** while controlling for the effect of **C**, is computed in the same way as a partial correlation coefficient (eq. 1):

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

where $r_M(\mathbf{AB})$ is the simple Mantel statistic between matrices **A** and **B**.

There is no theory connecting permutation procedures to the partial Mantel test, where a linear model is assumed among sets of distances. So, the difficult aspect is to construct appropriate permutations to test the significance of the partial Mantel statistic. Simulations were performed to illustrate the properties of four testing procedures. The same procedures were used in the partial correlation study; it is easy to modify the descriptions that follow to conduct a partial correlation analysis of three data vectors $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$.

### 4.1. Method 1: Permute the Objects in Matrix A

The first method proposed by Smouse *et al.* (1986) is to perform ‘matrix permutation’ on matrix **A** (or, alternatively, on **B**). As in the simple Mantel test (above), the permutable units for the test are the $n$ objects. This procedure holds **B** and **C** constant (*i.e.*, unpermuted) with respect to each other, preserving their ancillarity (which means *relatedness*). Smouse *et al.* (1986) suggested that it implements a regression-type model where **B** and **C** are considered to be fixed predictors of **A**. Our results, using data generated under a correlation model described below, will show that a regression model does not have to be assumed. The testing procedure is as follows:

1. Compute the Mantel correlations $r_M(\mathbf{AB})$, $r_M(\mathbf{AC})$ and $r_M(\mathbf{BC})$. Calculate the reference value of the test statistic, $r_M(\mathbf{AB.C})$, using eq. 2.
2. Permute **A** at random using matrix permutation to obtain **A***.
3. 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. 2, to obtain a value $r_M^*$ of the partial correlation statistic under permutation.
4. Repeat steps 2 and 3 a large number of times to obtain the distribution of $r_M^*$ under permutation. Add the reference value $r_M(\mathbf{AB.C})$ to the distribution.
5. 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 $r_M$.

Only two of the three correlation coefficients have to be computed after each permutation of **A**, *i.e.*, $r_M(\mathbf{A^*B})$ and $r_M(\mathbf{A^*C})$. One does not have to recompute $r_M(\mathbf{BC})$ since neither **B** nor **C** were permuted.

Results obtained by Anderson and Legendre (1999) in the multiple regression context indicated that in most instances, permutation of the raw data had correct type I error and good power. When the covariable contained an extreme outlier, however, permutation of raw data resulted in unstable (often inflated) type I error. The simulations reported in this paper will examine the behavior of permutation method 1 in the context of the partial Mantel test.

### 4.2. Method 2: Permute the Residuals of a Null Model

There are other permutation methods in which the permutable units are the residuals of some model, linear or not. In multiple regression and canonical analysis, the null hypothesis is that of exchangeability of the residuals of the response variable(s) after fitting the explanatory variables using a linear regression model. Tests of significance based on permutation of residuals are not exact in the randomization sense but have asymptotically exact significance levels. Method 2 leads to reasonable results in multiple regression and canonical analysis when $n > 10$ (ter Braak and Smilauer, 1998; Anderson and Legendre, 1999).

Permuting the residuals of a null (or reduced) model was originally proposed by Freedman and Lane (1983) for testing the significance of partial regression coefficients. Consider the multiple regression model $\mathbf{y} = \beta_0 + \beta_{1.2}\mathbf{x} + \beta_{2.1}\mathbf{z} + \varepsilon_{\mathbf{x},\,\mathbf{z}}$ where **y** is the dependent variable, **x** is a covariable and **z** is the explanatory variable of interest; we are interested in testing the significance of parameter $\beta_{2.1}$. The null hypothesis is $H_0$: $\beta_{2.1} = 0$; this hypothesis is equivalent to $H_0$: $\rho(\mathbf{yz.x}) = 0$. Consider a ‘null model’ where $H_0$: $\beta_{2.1} = 0$ is true; the regression equation can be rewritten as $\mathbf{y} = b_0 + b_{1.2}\mathbf{x} + \mathbf{e_x}$. In this model, all the variation of **y** not explained by **x** is expressed by the vector of residuals $\mathbf{e_x}$, which represent the best estimates of the random errors $\varepsilon_{\mathbf{x}}$. Residuals are exchangeable among observations if they are independent and identically distributed (i.i.d.; note that this is not the case with residuals computed from distances in a distance matrix since distances are not independent of one another). The rationale is the following: after obtaining an estimate of the relationship between the dependent variable and the covariable(s), the null hypothesis states that there is no further variation in the dependent variable that can be explained by the explanatory variable being tested.

The procedure to test the significance of a partial regression coefficient is the following: (1) regress **y** on **x** and **z** to obtain an estimate $b_{2.1}$ of $\beta_{2.1}$ and a reference value $t_{\text{ref}}$ of the corresponding $t$-statistic for the real data. (2) Regress **y** on **x** alone to obtain fitted values $\hat{\mathbf{y}}_{\mathbf{x}}$ and regression residuals $\mathbf{e_x}$ under the ‘null model’. (3) Permute at random the vector of residuals $\mathbf{e_x}$ to obtain $\mathbf{e}^*_{\mathbf{x}}$. (4) Add the permuted residuals $\mathbf{e}^*_{\mathbf{x}}$ onto the (unpermuted) vector of fitted values $\hat{\mathbf{y}}_{\mathbf{x}}$, producing a new vector of permuted values $\mathbf{y}^*$. (5) Compute the multiple regression $\mathbf{y}^* = b^*_0 + b^*_{1.2}\mathbf{x} + b^*_{2.1}\mathbf{z} + \mathbf{e}^*$, producing an estimate $b^*_{2.1}$ of the regression coefficient $\beta^*_{2.1}$ and a value $t^*$ under permutation. (6) Repeat steps 3 to 5 a large number of times to obtain a null distribution using the $t^*$-statistics, which is used to (7) test the significance of $t_{\text{ref}}$. The procedure is described in more detail in Anderson and Legendre (1999), who noted that an equivalent model for the test would be, in step 3, to regress the permuted regression residuals $\mathbf{e}^*_{\mathbf{x}}$ alone on the original (*i.e.*, unpermuted) predictors **x** and **z**. This simplification is used in the following permutation method, derived from that of Freedman and Lane, which is now proposed for testing the significance of partial Mantel statistics:

1. Compute matrix $\mathbf{Res}_{\mathbf{A}|\mathbf{C}}$ containing the residuals of the simple linear regression of the distances in **A** over the distances in **C**.
2. Compute $r_{\mathrm{M}}(\mathbf{AB})$, $r_{\mathrm{M}}(\mathbf{AC})$ and $r_{\mathrm{M}}(\mathbf{BC})$. Calculate the reference value of the test statistic, $r_{\mathrm{M}}(\mathbf{AB.C})$, using eq. 2. The exact same result would be obtained by using $\mathbf{Res}_{\mathbf{A}|\mathbf{C}}$ instead of **A** in the calculation of $r_{\mathrm{M}}(\mathbf{AB})$ and $r_{\mathrm{M}}(\mathbf{AC})$.
3. Permute $\mathbf{Res}_{\mathbf{A}|\mathbf{C}}$ at random, using matrix permutation, to obtain $\mathbf{Res}^*_{\mathbf{A}|\mathbf{C}}$.
4. Compute $r_{\mathrm{M}}(\mathbf{Res}^*_{\mathbf{A}|\mathbf{C}}\mathbf{B})$ and $r_{\mathrm{M}}(\mathbf{Res}^*_{\mathbf{A}|\mathbf{C}}\mathbf{C})$. Using eq. 2, combine these values with $r_{\mathrm{M}}(\mathbf{BC})$ computed in step 2, obtaining a value $r^*_{\mathrm{M}}$ of the partial correlation statistic under permutation.
5. Repeat steps 3 and 4 a large number of times to obtain the distribution of $r^*_{\mathrm{M}}$ under permutation. Add the reference value $r_{\mathrm{M}}(\mathbf{AB.C})$ to the distribution.
6. For a one-tailed test involving the upper tail, calculate the probability as the proportion of values $r^*_{\mathrm{M}}$ greater than or equal to $r_{\mathrm{M}}$. For a test in the lower tail, the probability is the proportion of values $r^*_{\mathrm{M}}$ smaller than or equal to $r_{\mathrm{M}}$.

The exact same p-value would be obtained by using a permuted matrix **A*** in step 4 instead of matrix $\mathbf{Res}^*_{\mathbf{A}|\mathbf{C}}$, where $\mathbf{A}^* = \mathbf{Fit}_{\mathbf{A}|\mathbf{C}} + \mathbf{Res}^*_{\mathbf{A}|\mathbf{C}}$ and $\mathbf{Fit}_{\mathbf{A}|\mathbf{C}}$ is the matrix of fitted values of the regression in step 1. The procedure described above is computationally faster, however.

Results obtained by Anderson and Legendre (1999) in the multiple regression context indicated that permutation of the residuals of a null model had the most consistent results in terms of type I error and power, including the cases where an extreme outlier was present in the covariable. For small sample sizes in the presence of strongly non-normal (cubed exponential) error, this method became too conservative. As mentioned above, tests of significance involving permutation of residuals are only asymptotically exact; their error rates may be slightly off with small sample sizes. The simulations reported in this paper will show whether or not method 2 behaves as well for partial Mantel tests as the Freedman and Lane method does for tests of partial regression coefficients.

### 4.3. Method 3: Correlate Residualized A to Residualized B; Permute Residualized A

A partial correlation $r(\mathbf{x}_1\mathbf{x}_2.\mathbf{x}_3)$ may be obtained by computing a simple correlation between the vectors of residual values $\mathbf{Res}_{\mathbf{x}_1|\mathbf{x}_3}$ and $\mathbf{Res}_{\mathbf{x}_2|\mathbf{x}_3}$. This property is used in the alternative procedure designed by Smouse *et al.* (1986) to test the significance of partial Mantel statistics using permutations. They suggested that their procedure implements a correlation-type model where **B** and **C** are variable, instead of fixed, and are simply considered to be predictors of **A**.

1. Compute matrix $\mathbf{Res}_{\mathbf{A}|\mathbf{C}}$ containing the residuals of the simple linear regression of the distances in **A** over the distances in **C**.
2. Likewise, compute matrix $\mathbf{Res}_{\mathbf{B}|\mathbf{C}}$ containing the residuals of the simple linear regression of the distances in **B** over the distances in **C**.
3. Compute the standardized Mantel statistic between $\mathbf{Res}_{\mathbf{A}|\mathbf{C}}$ and $\mathbf{Res}_{\mathbf{B}|\mathbf{C}}$ to obtain the reference value of the test statistic, $r_{\mathrm{M}}(\mathbf{AB.C})$. The same reference value may be obtained by computing eq. 2 using the Mantel correlations $r_{\mathrm{M}}(\mathbf{AB})$, $r_{\mathrm{M}}(\mathbf{AC})$ and $r_{\mathrm{M}}(\mathbf{BC})$.
4. Using matrix permutation, permute $\mathbf{Res}_{\mathbf{A}|\mathbf{C}}$ at random to obtain a permuted residual matrix $\mathbf{Res}^*_{\mathbf{A}|\mathbf{C}}$. An equivalent method is to permute $\mathbf{Res}_{\mathbf{B}|\mathbf{C}}$ at random instead of $\mathbf{Res}_{\mathbf{A}|\mathbf{C}}$, obtaining the permuted matrix $\mathbf{Res}^*_{\mathbf{B}|\mathbf{C}}$.
5. Compute the standardized Mantel statistic between $\mathbf{Res}^*_{\mathbf{A}|\mathbf{C}}$ and $\mathbf{Res}_{\mathbf{B}|\mathbf{C}}$, to obtain a value $r^*_{\mathrm{M}}(\mathbf{AB.C})$ of the test statistic under permutation. An equivalent method is to compute the Mantel statistic between $\mathbf{Res}_{\mathbf{A}|\mathbf{C}}$ and $\mathbf{Res}^*_{\mathbf{B}|\mathbf{C}}$.
6. Repeat steps 4 and 5 a large number of times to obtain the distribution of $r^*_{\mathrm{M}}(\mathbf{AB.C})$ under permutation. Add the reference value $r_{\mathrm{M}}(\mathbf{AB.C})$ to the distribution.
7. For a one-tailed test involving the upper tail, calculate the probability as the proportion of values $r^*_{\mathrm{M}}$ greater than or equal to $r_{\mathrm{M}}$. For a test in the lower tail, the probability is the proportion of values $r^*_{\mathrm{M}}$ smaller than or equal to $r_{\mathrm{M}}$.

This method is shorter to compute than all other methods. Only one correlation coefficient has to be computed during each permutation, compared to two in the other methods.

Kennedy (1995) proposed this same method as a simple way of computing the permutation of residuals of a null model (method 2) in multiple linear regression. He demonstrated mathematically that, under permutation, method 3 produces the same estimate $b^*_{2.1}$ of a partial regression coefficient as method 2. Anderson and Legendre (1999) showed, however, that this equivalence does not hold when a $t$-statistic is used instead of the partial regression coefficient for permutation testing. They also showed that the Kennedy method of permutation, using a $t$-statistic, has inflated type I error, especially with small sample sizes. Using simulations, we will see whether or not method 3 for partial Mantel tests has the same drawback as the Kennedy method has for tests of partial regression coefficients.

## 4.4. Method 4: Permute the Residuals of a Full Model

Permutation of residuals of the full model has been used by ter Braak (1990, 1992) as a permutation analogue to the bootstrapping method proposed by Hall and Titterington (1989). In the regression context, (1) a full multiple regression model of $\mathbf{y}$ over all predictors (*e.g.*, $\mathbf{x}$ and $\mathbf{z}$) is computed, yielding fitted values $\hat{\mathbf{y}}_{\mathbf{x},\,\mathbf{z}}$, residual values $\mathbf{e}_{\mathbf{x},\,\mathbf{z}}$, as well as a reference value $b_{\text{ref}}$ for the parameter of interest and an associated pivotal statistic $t_{\text{ref}}$. The residuals $\mathbf{e}_{\mathbf{x},\,\mathbf{z}}$ are used as the permutable units for the test. (2) The vector of residuals $\mathbf{e}_{\mathbf{x},\,\mathbf{z}}$ is permuted at random to obtain $\mathbf{e}^*_{\mathbf{x},\,\mathbf{z}}$. (3) A permuted dependent variable $\mathbf{y}^*$ is computed by adding the fitted values $\hat{\mathbf{y}}_{\mathbf{x},\,\mathbf{z}}$ to the permuted residuals $\mathbf{e}^*_{\mathbf{x},\,\mathbf{z}}$. (4) A new multiple regression is computed between the permuted dependent variable $\mathbf{y}^*$ and all predictors, yielding a permuted value $b^*$ for the parameter of interest and a $t^*$-statistics computed for the differences $(b^* - b_{\text{ref}})$. (5) Steps 2 and 3 are repeated a large number of times. A null distribution is constructed using the $t^*$-statistics, which is used to (6) test the significance of $t_{\text{ref}}$. The two main differences with method 2 are: (1) the use of a full regression model and (2) the calculation of $t^*$-statistics for the differences $(b^* - b_{\text{ref}})$.

This procedure is described in more detail in Anderson and Legendre (1999). Manly (1997) as well as Anderson and Legendre (1999) noted that an equivalent model for the test consists in regressing the permuted residuals, $\mathbf{e}^*_{\mathbf{x},\mathbf{z}}$, on the unpermuted predictors **x** and **z**. This short-cut directly yields a value $b^*$ which is appropriate for computing the correct $t^*$-statistic for each permutation, instead of having to compute a $t^*$-statistic for the difference $(b^* - b_{\mathrm{ref}})$. This simplification is used in the following permutation method, derived from that of ter Braak, which is proposed for testing the significance of partial Mantel statistics:

1. Compute $r_{\mathrm{M}}(\mathbf{AB})$, $r_{\mathrm{M}}(\mathbf{AC})$ and $r_{\mathrm{M}}(\mathbf{BC})$. Calculate the reference value of the test statistic, $r_{\mathrm{M}}(\mathbf{AB.C})$, using eq. 2.
2. Compute matrix $\mathbf{Res}_{\mathbf{A|BC}}$ containing the residuals of the multiple linear regression of the distances in **A** over the distances in **B** and **C**.
3. Permute $\mathbf{Res}_{\mathbf{A|BC}}$ at random, using matrix permutation, to obtain $\mathbf{Res}^*_{\mathbf{A|BC}}$.
4. Compute $r_{\mathrm{M}}(\mathbf{Res}^*_{\mathbf{A|BC}}\mathbf{B})$ and $r_{\mathrm{M}}(\mathbf{Res}^*_{\mathbf{A|BC}}\mathbf{C})$. Using eq. 2, combine these values with $r_{\mathrm{M}}(\mathbf{BC})$ computed in step 1, obtaining a value $r^*_{\mathrm{M}}$ of the partial correlation test statistic under permutation.
5. Repeat steps 3 and 4 a large number of times to obtain the distribution of $r^*_{\mathrm{M}}$ under permutation. Add the reference value $r_{\mathrm{M}}(\mathbf{AB.C})$ to the distribution.
6. For a one-tailed test involving the upper tail, calculate the probability as the proportion of values $r^*_{\mathrm{M}}$ greater than or equal to $r_{\mathrm{M}}$. For a test in the lower tail, the probability is the proportion of values $r^*_{\mathrm{M}}$ smaller than or equal to $r_{\mathrm{M}}$.

In the multiple regression context, this method allows one to test the significance of several (or all possible) partial regression coefficients using a single series of permutations. With method 2, on the contrary, a different set of permutations has to be produced for each partial regression coefficient to be tested.

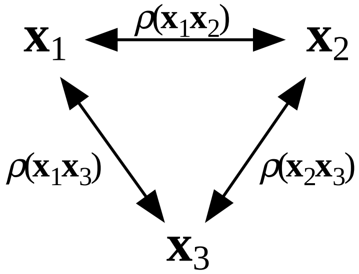
Results obtained by Anderson and Legendre (1999) in the multiple regression context indicated that permutation of the residuals of a full model produced results as good as those of Freedman and Lane in most cases, including when an extreme outlier was present in the covariable. We will see in Section 6.3 whether or not method 4 behaves as well for partial Mantel tests as the ter Braak method does for tests of partial regression coefficients.

## 5. SIMULATION METHODS

### 5.1. Correlation Model

In the partial correlation study, three vectors $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$ of length $n$ were created by random draw from a standard normal distribution ($\mu = 0$, $\sigma^2 = 1$), so that each one is a random component of the correlation model described in the next paragraph. Simulations were also run using highly skewed data obtained from standard exponential ($\mu = 1$, $\sigma^2 = 1$) or cubed exponential deviates, as in Manly (1997: 163-166) and Anderson and Legendre (1999). The three vectors were written to a matrix $\mathbf{X}$ of size $(n, 3)$ with elements $[x_{i,j}]$.

The deterministic component was introduced through the correlation model:



To implement this model, correlations were introduced between the random vectors $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$ through a correlation matrix $\mathbf{R}$ $(3 \times 3)$ reflecting the desired amounts of correlation $\rho(\mathbf{x}_1\mathbf{x}_2)$, $\rho(\mathbf{x}_1\mathbf{x}_3)$ and $\rho(\mathbf{x}_2\mathbf{x}_3)$ in the statistical population from which the points were drawn. Matrix $\mathbf{R}$ was decomposed using Cholesky factorization, $\mathbf{R} = \mathbf{L}'\mathbf{L}$, where $\mathbf{L}$ is a $(3 \times 3)$ upper triangular matrix with elements $[L_{\mathrm{j,k}}]$. Matrix $\mathbf{W}$ containing the correlated vectors was obtained by computing the linear equation $\mathbf{W} = \mathbf{XL}$. The model is thus:

$$\begin{aligned}
w_{i,1} &= L_{1,1}x_{i,1} \\
w_{i,2} &= L_{1,2}x_{i,1} + L_{2,2}x_{i,2} \\
w_{i,3} &= L_{1,3}x_{i,1} + L_{2,3}x_{i,2} + L_{3,3}x_{i,3}
\end{aligned}$$

The elements of the original data vectors, $x_{i,1}$, $x_{i,2}$ and $x_{i,3}$, are the random components of the model. The transformation $\mathbf{W} = \mathbf{XL}$ is such that, if the standard vectors in $\mathbf{X}$ are linearly independent, *i.e.*, if $[1/(n-1)]\mathbf{X}'\mathbf{X} = \mathbf{I}$, the covariances between the column vectors of **W** reflect the original correlations assigned to matrix **R**:

$$[1/(n-1)]\mathbf{W'W} = [1/(n-1)]\mathbf{L'X'XL} = \mathbf{L'IL} = \mathbf{L'L} = \mathbf{R}$$

Standard normal and standard exponential random deviates have variances of 1 and covariances of 0; so the vectors of random deviates did not require standardization prior to being correlated before using this procedure. With cubed exponential deviates, however, the variance of the parent distribution was not 1 (actually, it was close to 823); so, the columns of **X** were standardized before being correlated, in order for $[1/(n-1)]\mathbf{X'X} = \mathbf{I}$ to hold.

This is not the only way of introducing correlations among randomly generated data vectors. It may also be done using the square root of matrix **R**. The Cholesky matrix **L** is sometimes misleadingly called the square root of **R**. The square root of a square matrix **R** is a matrix $\mathbf{R}^{1/2} = \mathbf{M}$ such that $\mathbf{MM} = \mathbf{R}$; **M** is a square symmetric matrix, contrary to **L** which is a square (asymmetric) upper triangular matrix such that $\mathbf{L'L} = \mathbf{R}$; $\mathbf{R}^{1/2}$ is obtained as:

$$\mathbf{R}^{1/2} = \mathbf{U}\ [\lambda_i^{1/2}]\ \mathbf{U}^{-1}$$

where $[\lambda_i^{1/2}]$ is a matrix containing the square roots of the eigenvalues of **R** and **U** is the corresponding matrix of eigenvectors; $[\lambda_i^{1/2}]$ is also the matrix of singular values if **R** is subjected to singular value decomposition. The operations $\mathbf{V} = \mathbf{XM}$ and $\mathbf{W} = \mathbf{XL}$ represent two legitimate ways of introducing correlations among the vectors of **X**, but the matrices **V** and **W** they produce differ. The sample correlations among the vectors of **V** and **W** also differ sightly for a given data matrix **X**. Cholesky factorization was used for the simulations reported in this study because it is faster to compute. Some of the simulations were recomputed using the square root matrix; the results were consistent with those obtained using the Cholesky method.

Mantel tests are concerned with linear relationships between distances instead of between variables. So, in the Mantel and partial Mantel study, for each of the *n* objects, $p = 10$ vectors of data (*i.e.*, variables) were drawn randomly from either a standard normal distribution or a cubed exponential distribution. Euclidean distances were computed among the *n* objects. This was repeated for each of the two or three matrices involved in any given simulation, producing symmetric distance matrices **A**, **B** and **C** with zeros along the diagonal. These matrices were created independently; they were thus random with respect to one another (*i.e.*, they do not correspond to fixed factors). The upper triangular portions of matrices **A**, **B** and **C** were written as column vectors in a matrix **X** of size $[n(n-1)/2, 3]$ with elements $[x_{i,j}]$. The elements $x_{i,1}$, $x_{i,2}$ and $x_{i,3}$, which represent distances from the original matrices **A**, **B** and **C**, are the random components of the model, even though the elements in each single vector $\mathbf{x}_j$ are related to one another by the triangle inequality.

Correlations were introduced between the columns of matrix **X** in the same way as described above for the partial correlation study. Matrix **W** contained the correlated distance matrices. The vectors of correlated values were put back into square distance matrices. This way of generating correlations among distance matrices is appropriate to the study of simple and partial Mantel tests because the hypothesis one is interested in is that of a linear relationship between distances.

An alternative data generation method will be used in section 6.1. It consists of generating two random vectors and correlating them by a predetermined value. On the one hand, a simple correlation coefficient is computed between them; on the other hand, one computes a distance matrix for each vector and then a Mantel statistic between the two distance matrices. This method of data generation will be used to compare the power of the simple correlation to that of the simple Mantel tests for data for which the correlation coefficient is appropriate.

In permutation methods 2, 3 and 4, residuals are computed using a linear regression model; it is applied to distances in Mantel and partial Mantel tests. Residuals computed in this way do not have a simple relationship to the random components generated using the correlation model described above. Generation of the random components corresponds, however, to the hypothesis one is interested to test when using a correlation or partial correlation analysis (or a Mantel or partial Mantel test). This hypothesis may be stated as follows: given raw data vectors (or data tables) that are unrelated to each other, there is no linear relationship between the vectors (or derived distance matrices), above and beyond what may be expected by chance permutation of the original observations. The simulations reported in this paper will show empirically to what extent the various permutation methods behave correctly for data generated under the correlation model, (1) despite the lack of correspondence between the error generation procedure which follows from the correlation model and the model I regression method used for estimating residuals, and (2) despite the fact that in matrix permutation the residuals cannot be assumed to be i.i.d. This will be assessed under different conditions: different types of errors, and in the presence of outliers.

### 5.2. Regression Model I Data

In the simple and partial correlation study, and for purpose of comparison, data were generated, in some cases, with fixed values for variables $\mathbf{x}_2$ and $\mathbf{x}_3$, as in regression model I. This was done in either of two ways, depending on the situation. (a) In the comparison of properties of the simple correlation coefficient, variable $\mathbf{x}_2$ was generated with fixed values $= \{1, 2, \ldots, 10\}$ repeated 1 to 5 times, for a total of 10 to 50 data points; $\mathbf{x}_2$ was standardized. Variable $\mathbf{x}_1$ was then constructed as $\mathbf{x}_1 = \beta_2\mathbf{x}_2 + \varepsilon$ where $\varepsilon$ is a random deviate drawn from an error distribution; see below. (b) The second method of generating model I data, used in the study of partial correlation coefficients, is similar to that used by Manly (1997, pp. 162-166) and Anderson and Legendre (1999) to study the effect of outliers in multiple regression. Values of $\mathbf{x}_2$ and $\mathbf{x}_3$ were drawn at random from a uniform distribution in the interval [0, 3]. Variable $\mathbf{x}_1$ was then constructed as $\mathbf{x}_1 = \beta_2\mathbf{x}_2 + \beta_3\mathbf{x}_3 + \varepsilon$ where $\varepsilon$ is a random deviate drawn from an appropriate error distribution and added to the fixed effects of $\mathbf{x}_2$ and $\mathbf{x}_3$.

### 5.3. Normal and Highly Skewed Distributions

Simulations carried out using exponential ($\lambda = 1$, $\mu = 1$, $\sigma^2 = 1$) deviates or cubed exponential deviates corresponded to conditions known to be inappropriate for the parametric $t$-test. The permutation methods should produce better results than parametric tests under such conditions. We will see how well the various permutation methods do with this type of data.

In the partial Mantel study, data were first generated using standard normal deviates, producing multinormal distributions of objects in $p$-dimensional space. Simulation results obtained under these conditions should apply not only to distance matrices of such multinormal data, computed using the Euclidean distance, but also to any other triplets of distance matrices whose Euclidean representations are approximately multinormal. Highly non-multinormal data were also generated using cubed exponential deviates; they may be taken to also represent data distributions that may occur when using functions other than the Euclidean distance. The results in Section 6.3 will show which permutation method produces the most accurate results in that case.

### 5.4. Outliers

Outliers were introduced in vector $\mathbf{x}_3$ (or in matrix **C**), generated as described above, by including a value of 50 in variable $\mathbf{x}_3$ (in the partial correlation study) or in the first vector of the data matrix from which distance matrix **C** was computed (in the partial Mantel study). Large values for outliers have also been used by Manly (1997) and Anderson and Legendre (1999). When the null hypothesis of absence of partial relationship between $\mathbf{x}_1$ and $\mathbf{x}_2$ (or **A** and **B**) is true, an outlier in vector $\mathbf{x}_3$ (or matrix **C**) does not impinge on the null hypothesis; consequently, it should not affect type I error when using an appropriate testing procedure. Note that outliers may be high leverage points, affecting the values of parameter estimates.

### 5.5. Metric and Euclidean Properties

If the original distance matrices are Euclidean, the Cholesky-correlated matrices are also metric and Euclidean. This was empirically verified for a large number of distance matrices, using combinations of low, medium and high correlation values $\rho(\mathbf{AB})$, $\rho(\mathbf{AC})$ and $\rho(\mathbf{BC})$. For metricity, the property of positiveness of the resulting distances was checked, as well as the triangle's inequality property. For Euclideanarity, principal coordinate analysis was computed (Gower, 1966); non-Euclidean matrices would have been detected by their having negative eigenvalues (Gower and Legendre, 1986; Legendre and Legendre, 1998).

The matrices of residuals, however, are neither Euclidean nor metric. Since residuals are centered, these matrices are not even distance matrices, about half the values they contain being negative.

## 5.6. Computations

Two computer programs were written in Fortran 77 to carry out the simulations: one for the partial correlation study and one for the Mantel study. The same series of random simulated data were produced for all permutation methods in simulations involving the same set of parameters. Permutations of the simulated data were done using a uniform random generation algorithm *sensu* Furnas (1984). For very small samples ($n \leq 7$), where the number of possible permutations is small, complete enumeration of all possible $n!$ permutations was carried out using a procedure proposed by Edgington (1995, p. 207; for example, there are $5! = 120$ distinguishable permutations of $n = 5$ objects); in that case, 100000 simulations were produced instead of 10000 in order to keep the confidence intervals of the rejection rates narrow.

## 5.7. Type I Error

In the study of type I error, data were generated in such a way that the null hypothesis was true ($H_0$: $\rho(\mathbf{x}_1\mathbf{x}_2.\mathbf{x}_3) = 0$ or $\rho(\mathbf{AB.C}) = 0$). Empirical rates of type I error were studied for the four permutation methods and the parametric *t*-test with regard to the following factors:

1. The size of the samples: $n = \{5, 10, 20, 30, 40, 50\}$.
2. For simple correlations, the population correlation $\rho(\mathbf{x}_1\mathbf{x}_2)$ was set to 0 when generating the data. For partial correlations, there are various ways of varying the simple correlations between the three vectors to obtain $\rho(\mathbf{x}_1\mathbf{x}_2.\mathbf{x}_3) = 0$. The expected value of the partial correlation is zero if the numerator of eq. 1 is zero:

$$\rho(\mathbf{x}_1\mathbf{x}_2) - \rho(\mathbf{x}_1\mathbf{x}_3)\,\rho(\mathbf{x}_2\mathbf{x}_3) = 0 \qquad (3)$$

Hence the following combinations of simulation parameters produced realizations of the null hypothesis:

- $\rho(\mathbf{x}_1\mathbf{x}_3) = \{0.0, 0.5, 0.9\}$ while $\rho(\mathbf{x}_1\mathbf{x}_2)$ and $\rho(\mathbf{x}_2\mathbf{x}_3) = 0$; also $\rho(\mathbf{x}_2\mathbf{x}_3) = \{0.0, 0.5, 0.9\}$ while $\rho(\mathbf{x}_1\mathbf{x}_2)$ and $\rho(\mathbf{x}_1\mathbf{x}_3) = 0$;
- $\rho(\mathbf{x}_1\mathbf{x}_3) = \rho(\mathbf{x}_2\mathbf{x}_3) = \{0.0, 0.1, 0.2, \ldots, 0.9\}$ with corresponding values of $\rho(\mathbf{x}_1\mathbf{x}_2)$ set at $\{0.00, 0.01, 0.04, \ldots, 0.81\}$.

The same strategy was used in the partial Mantel study. Empirical type I error rates were calculated for each combination of factors. The significance level for the tests was set at $\alpha = 0.05$. For type I error to match the significance level, the number of significant p-values out of the 10000 (or 100000) simulations was expected to be 500 (or 5000) for each of the permutation methods. A 95% confidence interval was computed for each estimated error rate. In the partial correlation simulation study, parametric *t*-tests were also calculated in parallel with the various types of permutation tests.

### 5.8. Power

In the power study, data were simulated in such a way that the null hypothesis was false: $\rho(\mathbf{x}_1\mathbf{x}_2.\mathbf{x}_3) \neq 0$ or $\rho(\mathbf{AB}.\mathbf{C}) \neq 0$. Preliminary runs helped determine an appropriate value of $\rho(\mathbf{x}_1\mathbf{x}_2.\mathbf{x}_3)$ or $\rho(\mathbf{AB.C})$, for the chosen sample sizes, that would produce rejection rates always larger than 0 and smaller than 1. The selected value is 0.1. Power was studied with regard to the following factors:

1. The size of the samples: $n = \{5, 10, 30, 50\}$.
2. For the simple correlation study, $\rho(\mathbf{x}_1\mathbf{x}_2)$ was varied by steps of 0.1 from 0.0 to 0.9.
3. For the partial correlation study, eq. 1 indicates that there are various ways of varying the correlations between the three vectors while keeping the expected value of partial correlation constant at 0.1:

   - $\rho(\mathbf{x}_1\mathbf{x}_3) = 0$ while $\rho(\mathbf{x}_2\mathbf{x}_3) = \{0.0, 0.3, 0.5, 0.7, 0.9\}$. Equation (1) can be transformed to give the values of $\rho(\mathbf{x}_1\mathbf{x}_2)$ corresponding to different values of $\rho(\mathbf{x}_2\mathbf{x}_3)$ for constant $\rho(\mathbf{x}_1\mathbf{x}_2.\mathbf{x}_3)$:

     $$\rho(\mathbf{x}_1\mathbf{x}_2) = \rho(\mathbf{x}_1\mathbf{x}_2.\mathbf{x}_3)\sqrt{1-\rho(\mathbf{x}_2\mathbf{x}_3)^2} \tag{4}$$

     These values are: $\rho(\mathbf{x}_2\mathbf{x}_3) = 0.0$, $\rho(\mathbf{x}_1\mathbf{x}_2) = 0.1$; $\rho(\mathbf{x}_2\mathbf{x}_3) = 0.3$, $\rho(\mathbf{x}_1\mathbf{x}_2) = 0.0954$; $\rho(\mathbf{x}_2\mathbf{x}_3) = 0.5$, $\rho(\mathbf{x}_1\mathbf{x}_2) = 0.0866$; and $\rho(\mathbf{x}_2\mathbf{x}_3) = 0.9$, $\rho(\mathbf{x}_1\mathbf{x}_2) = 0.0436$. The same equation was used for the case where $\rho(\mathbf{x}_2\mathbf{x}_3) = 0$ while $\rho(\mathbf{x}_1\mathbf{x}_3) = \{0.0, 0.3, 0.5, 0.7, 0.9\}$.
   - $\rho(\mathbf{x}_1\mathbf{x}_3) = \rho(\mathbf{x}_2\mathbf{x}_3) = \{0.0, 0.1, \ldots 0.9\}$. The value $\rho(\mathbf{x}_1\mathbf{x}_2)$ was adjusted in such a way that $\rho(\mathbf{x}_1\mathbf{x}_2.\mathbf{x}_3)$ was kept constant at 0.1 in each plot. Equation (1) can be transformed to give the values of $\rho(\mathbf{x}_1\mathbf{x}_2)$

corresponding to different values of $\rho(\mathbf{x}_1\mathbf{x}_3) = \rho(\mathbf{x}_2\mathbf{x}_3)$ for constant $\rho(\mathbf{x}_1\mathbf{x}_2.\mathbf{x}_3)$:

$$\rho(\mathbf{x}_1\mathbf{x}_2) = \rho(\mathbf{x}_1\mathbf{x}_2.\mathbf{x}_3)\,(1 - \rho(\mathbf{x}_2\mathbf{x}_3)^2) + \rho(\mathbf{x}_2\mathbf{x}_3)^2 \qquad (5)$$

These values are: $\rho(\mathbf{x}_1\mathbf{x}_3) = \rho(\mathbf{x}_2\mathbf{x}_3) = 0.0$, $\rho(\mathbf{x}_1\mathbf{x}_2) = 0.1$; $\rho(\mathbf{x}_1\mathbf{x}_3) = \rho(\mathbf{x}_2\mathbf{x}_3) = 0.1$, $\rho(\mathbf{x}_1\mathbf{x}_2) = 0.109$; … $\rho(\mathbf{x}_1\mathbf{x}_3) = \rho(\mathbf{x}_2\mathbf{x}_3) = 0.9$, $\rho(\mathbf{x}_1\mathbf{x}_2) = 0.829$.

The same strategy was used in the partial Mantel study. Power is reported as the rate (fraction) of rejection of the null hypothesis after 10000 or 100000 simulations.

## 6. RESULTS

### 6.1. Simple Correlation and Simple Mantel Test

Figures 1a–c present empirical type I error rates at $\alpha = 0.05$ for the simple correlation coefficient $r(\mathbf{x}_1\mathbf{x}_2)$ (parametric $t$-test and permutation test) and the simple Mantel test $r(\mathbf{AB})$ (permutation test only), obtained using simulations where $\rho(\mathbf{x}_1\mathbf{x}_2)$ and $\rho(\mathbf{AB})$ were 0 in the statistical populations. The permutation tests always have correct type I error, whatever the shape of the distribution of errors. It is noticeable, however, that for non-normal data, the parametric $t$-test used in correlation analysis has inflated type I error for data generated under the correlation model (panel a), whereas it is too conservative for data generated under the regression model I (panel c), especially for cubed exponential error. Similarly, Anderson and Legendre (1999, Fig. 3) found that the normal-theory $t$-test is too conservative for regression model I data generated with cubed exponential error. Differences between the correlation and regression model I methods of data generation will also be found in partial correlation analysis (section 6.2). These results are summarized in Table I.

Figures 1d, e present power curves for the simple correlation coefficient and the simple Mantel test, for different sample sizes, across values of the population correlation $\rho$. Unsurprisingly, the powers of the simple correlation and simple Mantel tests increase with $n$. The power values

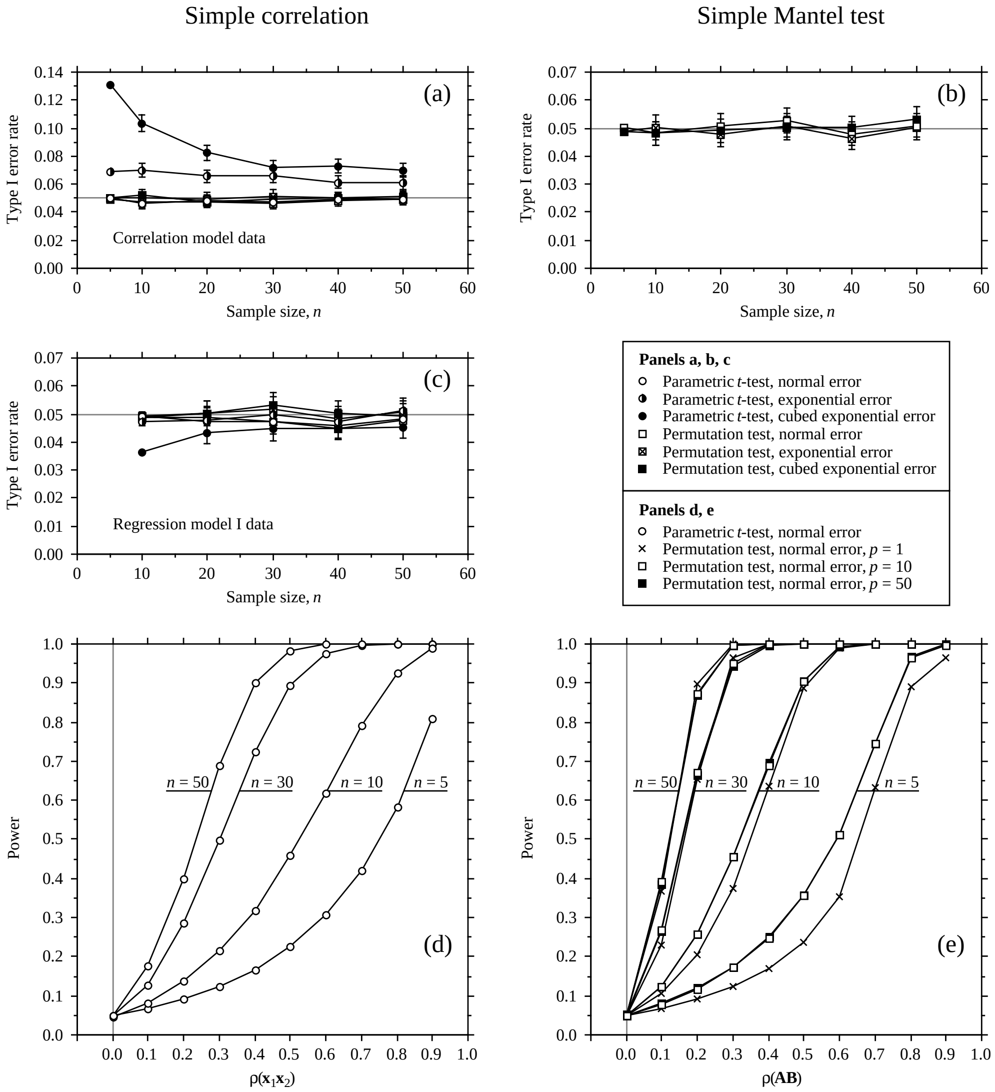


FIGURE 1 (a) Type I error and 95% confidence intervals (error bars) at $\alpha = 0.05$ with increasing sample size, for simple correlation coefficients (one-tailed tests; circles, parametric $t$-tests; squares, permutation tests), using data generated under the correlation model. (b) Same for the simple Mantel test ($p = 10$). (c) Same for simple correlations, using regression model I data (independent variable with fixed values {1, 2, …, 10} repeated 1 to 5 times, for a total of 10 to 50 data points). (d) Power curves for the simple correlation coefficient (parametric $t$-test) for different sample sizes $n$. (e) Same for the simple Mantel test, for different numbers of variables $p = \{1, 10, 50\}$. There were 10000 simulations in each case (100000 for $n = 5$) and 999 permutations per test where permutations were used. For $n = 5$, the confidence intervals are often so narrow that they are hidden by the symbols.

reported by Cohen (1988, Table 3.3.2) for one-tailed tests of the simple correlation coefficient are always inside the 95% confidence intervals obtained in our simulations (C.I. not plotted in Fig. 1d); this provides an

Table I Type I error: summary of the simulation findings. √ correct type I error; X incorrect type I error; # inflated type I error for small *n*. -----: not done.

| | *Normal error*<br>*Correlation model* | *Exponential error*<br>*Correlation model* | *Exponential*$^3$ *error*<br>*Correlation model* | *Normal + outlier*<br>*Regression model I* | *Normal + outlier*<br>*Correlation model* | *Exp*$^3$ *+ outlier*<br>*Correlation model* |
|---|---|---|---|---|---|---|
| **Simple correlation** | | | | | | |
| Parametric *t*-test | √ | X | X | | | |
| Permutation test | √ | √ | √ | | | |
| **Simple Mantel test** | | | | | | |
| Permutation test | √ | √ | √ | | | |
| **Partial correlation** | | | | | | |
| Parametric *t*-test | √ | ----- | X | √ | √ | X |
| 1. Permute raw data | √ | ----- | √ | X | √ | X |
| 2. Permute residuals of null model | √ | ----- | # | √ | √ | # |
| 3. Correlate residualized $\mathbf{x}_1$ to residualized $\mathbf{x}_2$ | X# | ----- | X# | ----- | ----- | ----- |
| 4. Permute residuals of full model | √ | ----- | # | √ | √ | # |
| **Partial Mantel test** | | | | | | |
| 1. Permute raw data | √ | ----- | √ | ----- | X | X |
| 2. Permute residuals of null model | √ | ----- | # | ----- | √ | # |
| 3. Correlate residualized **A** to residualized **B** | X# | ----- | X# | ----- | ----- | ----- |
| 4. Permute residuals of full model | # | ----- | # | ----- | # | # |

external validation of the simulation algorithm and computer code used to produce the simulations reported in the present paper. The power values were obtained in the same way as in the previous paragraph, using normal error and significance level $\alpha = 0.05$. The power of the simple Mantel test is the same for any number of variables, $p$, used in computing matrices **A** and **B**, except when $n$ is small and $p$ is very small.

The numerical value of the Mantel statistic required to reach significance, for any given $n$, is smaller than that of the Pearson correlation coefficient, a fact that users of the Mantel test often remark on. The results in Figures 1d and 1e are not entirely comparable, however, because the data have been generated in different ways; the correlation was added to the raw data vectors in the simple correlation study whereas it was added to the distances in the simple Mantel study.

Additional simulations carried out using data generated according to the method described in the penultimate paragraph of section 5.1, and reported in Table II, showed that the Mantel test is actually less powerful than the test of the Pearson correlation for analyzing the correlation between data vectors. Concerning the numerical values of the coefficients in the two cases, Dutilleul *et al.* (2000) have shown that in situations such as used in Table II, for any selected value of $\rho(\mathbf{x}_1\mathbf{x}_2)$ in the statistical population, the expected value of the simple correlation $r(\mathbf{x}_1\mathbf{x}_2)$ is equal to $\rho(\mathbf{x}_1\mathbf{x}_2)$ whereas that of the Mantel correlation $r(\mathbf{A}^{(2)}\mathbf{B}^{(2)})$ between matrices of squared Euclidean distances is $\rho^2(\mathbf{x}_1\mathbf{x}_2)$; it is thus smaller than $r(\mathbf{x}_1\mathbf{x}_2)$. The Mantel correlation $r(\mathbf{AB})$ between matrices of ordinary (not squared) Euclidean distances is always a bit smaller than $r(\mathbf{A}^{(2)}\mathbf{B}^{(2)})$.

Table II Comparison of power of the simple correlation and simple Mantel tests, using data vectors generated for power analysis of the simple correlation coefficient. The correlation imposed between the two vectors is $\rho(\mathbf{x}_1\mathbf{x}_2) = 0.5$, hence $H_0$ is false; error is normal. Power is the rejection rate of $H_0$ after 10000 simulations (or 100000 for $n = 5$); $\alpha = 0.05$, one-tailed tests.

| *n* | *Power of t-test of Pearson correlation* | *Power of simple Mantel test* |
|---|---|---|
| 5 | 0.2246 | 0.1244 |
| 10 | 0.4605 | 0.2872 |
| 30 | 0.8944 | 0.6268 |
| 50 | 0.9826 | 0.8067 |

## 6.2. Partial Correlation Analysis

### 6.2.1. *Type I Error*

Figures 2a, 2c and 2e present the type I error rates obtained for the various permutation methods, as well as for the parametric *t*-test, for increasing sample sizes, using data with normal error. When $\rho(\mathbf{x}_1\mathbf{x}_2) = \rho(\mathbf{x}_1\mathbf{x}_3) = \rho(\mathbf{x}_2\mathbf{x}_3) = 0$, all methods had correct type I error, except method 3

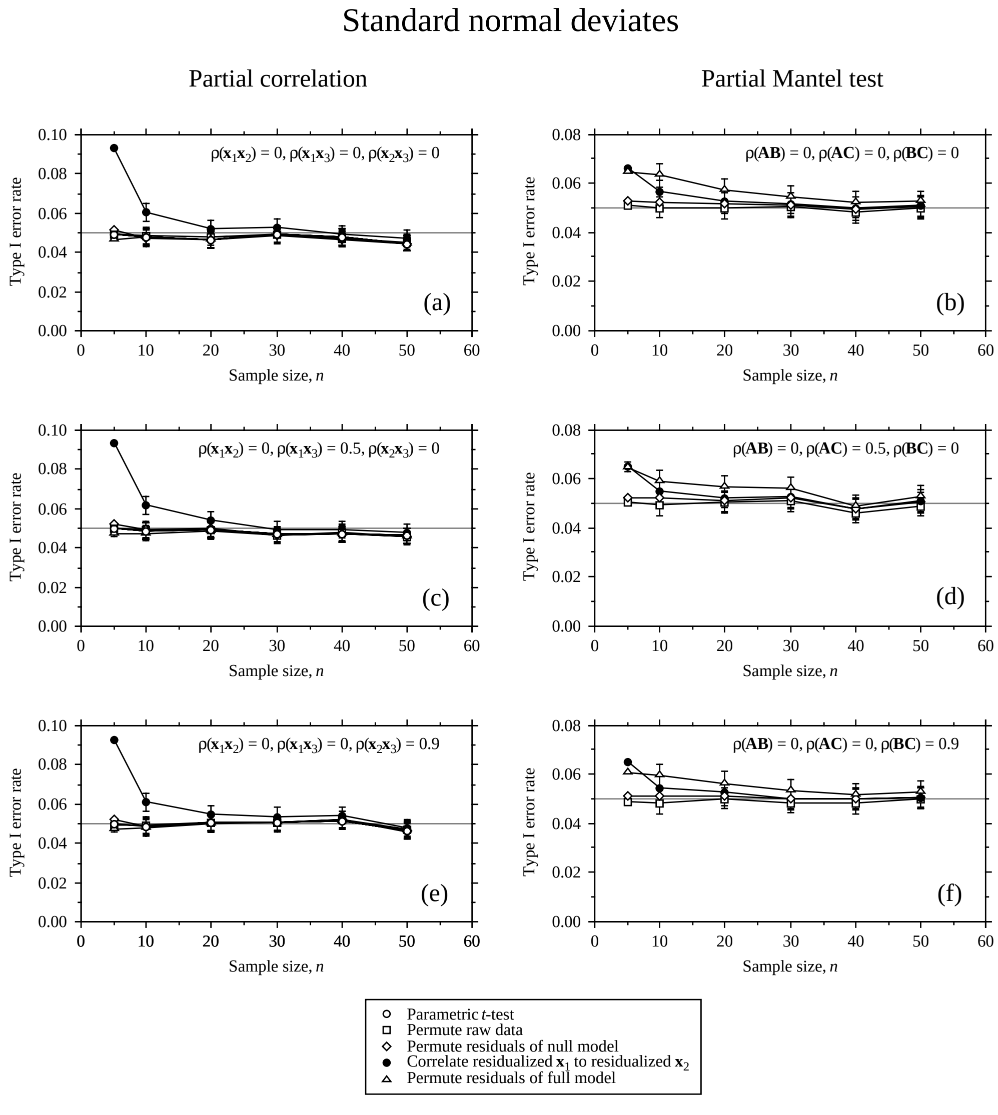


FIGURE 2 Mean and 95% confidence interval of the empirical probability of type I error ($\alpha = 0.05$) with increasing sample size. *Left:* comparison of the four permutation methods in partial correlation analysis, plus the parametric *t*-test. The population parameters were chosen in such a way that $H_0$ was true; the error was standard normal. *Right:* same for partial Mantel tests; parametric *t*-test could not be carried out here.

(correlate residualized $\mathbf{x}_1$ to residualized $\mathbf{x}_2$). This is also the case for increasing values of $\rho(\mathbf{x}_1\mathbf{x}_3)$ when $\rho(\mathbf{x}_1\mathbf{x}_2)$ and $\rho(\mathbf{x}_2\mathbf{x}_3)$ were set to zero (example in Fig. 2c), and for increasing values of $\rho(\mathbf{x}_2\mathbf{x}_3)$ when $\rho(\mathbf{x}_1\mathbf{x}_2)$ and $\rho(\mathbf{x}_1\mathbf{x}_3)$ were set to zero (example in Fig. 2e). Similar results were obtained for two-tailed tests (not shown).

Permutation method 3 (correlate residualized $\mathbf{x}_1$ to residualized $\mathbf{x}_2$) had inflated type I error with small sample sizes. This confirmed similar results obtained by Anderson and Legendre (1999) in a simulation study of permutation methods for testing the significance of partial regression coefficients in multiple linear regression.

Simulations of the null hypothesis were also conducted with $\rho(\mathbf{x}_1\mathbf{x}_3) = \rho(\mathbf{x}_2\mathbf{x}_3) = \{0.0, 0.1, 0.2, \ldots\ 0.9\}$. The results are the same as reported above: methods 1, 2 and 4 had error rates indistinguishable from that of the parametric *t*-test whereas method 3 had inflated type I error, significantly higher than that of the other methods for $n = 5$ and 10.

In simulations carried out using highly skewed data (*i.e.*, cubed exponential random deviates) in all three vectors, all permutation tests did much better than the normal-theory *t*-test in one-tailed or two-tailed tests (Figs. 3a, c); the normal-theory *t*-test had inflated type I error in such a case. This is opposite to the findings of Anderson and Legendre (1999) for data generated under regression model I (*i.e.*, explanatory variables controlled, error added), where the normal-theory *t*-test had reduced type I error when using cubed exponential deviates; this is also the case for simple correlation coefficients (Fig. 1c). No noticeable effect was found when cubed exponential deviates were used in $\mathbf{x}_1$ only while $\mathbf{x}_2$ and $\mathbf{x}_3$ contained random normal deviates (results not shown). In all cases, method 3 (correlate residualized $\mathbf{x}_1$ to residualized $\mathbf{x}_2$) had the same problem as it did with normal error (Fig. 2).

When cubed exponential deviates were used in all three vectors, the methods of permutation of residuals also had inflated type I error for very small sample sizes ($n = 5$ and 10) in one-tailed and two-tailed tests (Figs. 3a, c). This is not surprising since methods of permutation of residuals are known to only have asymptotically exact significance levels; they require $n$ to be relatively large. The permutation of residuals of a null model did better than the permutation of residuals of a full model. The

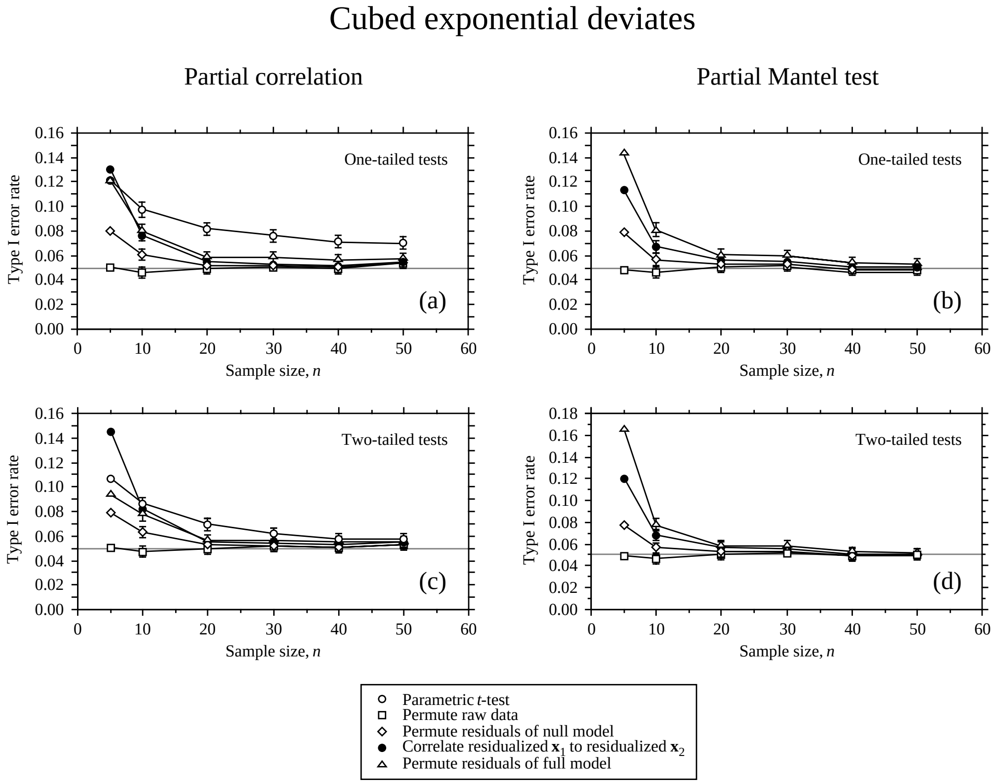


FIGURE 3 Type I error of one-tailed (a, b) and two-tailed tests (c, d) with increasing sample size for data generated under the correlation model with cubed exponential error in all variables. (a, c) Partial correlations; the population correlations $\rho(\mathbf{x}_1\mathbf{x}_2)$, $\rho(\mathbf{x}_1\mathbf{x}_3)$ and $\rho(\mathbf{x}_2\mathbf{x}_3)$ were 0. (b, d) Partial Mantel tests; the population correlations $\rho(\mathbf{AB})$, $\rho(\mathbf{AC})$ and $\rho(\mathbf{BC})$ were 0.

method of permutation of raw data was not affected by highly skewed data.

Anderson and Legendre (1999) noted that when the covariable of a multiple regression model ($\mathbf{x}_3$ in the present study) contained an extreme outlier, permutation of raw data resulted in unstable (often inflated) type I error. This effect is actually stronger in two-tailed than in one-tailed tests for regression model I data (Figs. 4a, c). To check the influence of outliers on permutation tests in the partial correlation context, data were generated under the correlation model, as described above, using standard normal deviates. An extreme outlier's value of 50 was included in vector $\mathbf{x}_3$ and the partial correlation $r(\mathbf{x}_1\mathbf{x}_2.\mathbf{x}_3)$ was tested for significance. No effect of the outlier was detected on type I error for any value of $\rho(\mathbf{x}_1\mathbf{x}_3)$ (Figs. 4e, g; simulations were run for $n$ = 10, 30, 50, 100) for any sample size, in one-tailed or two-tailed tests.

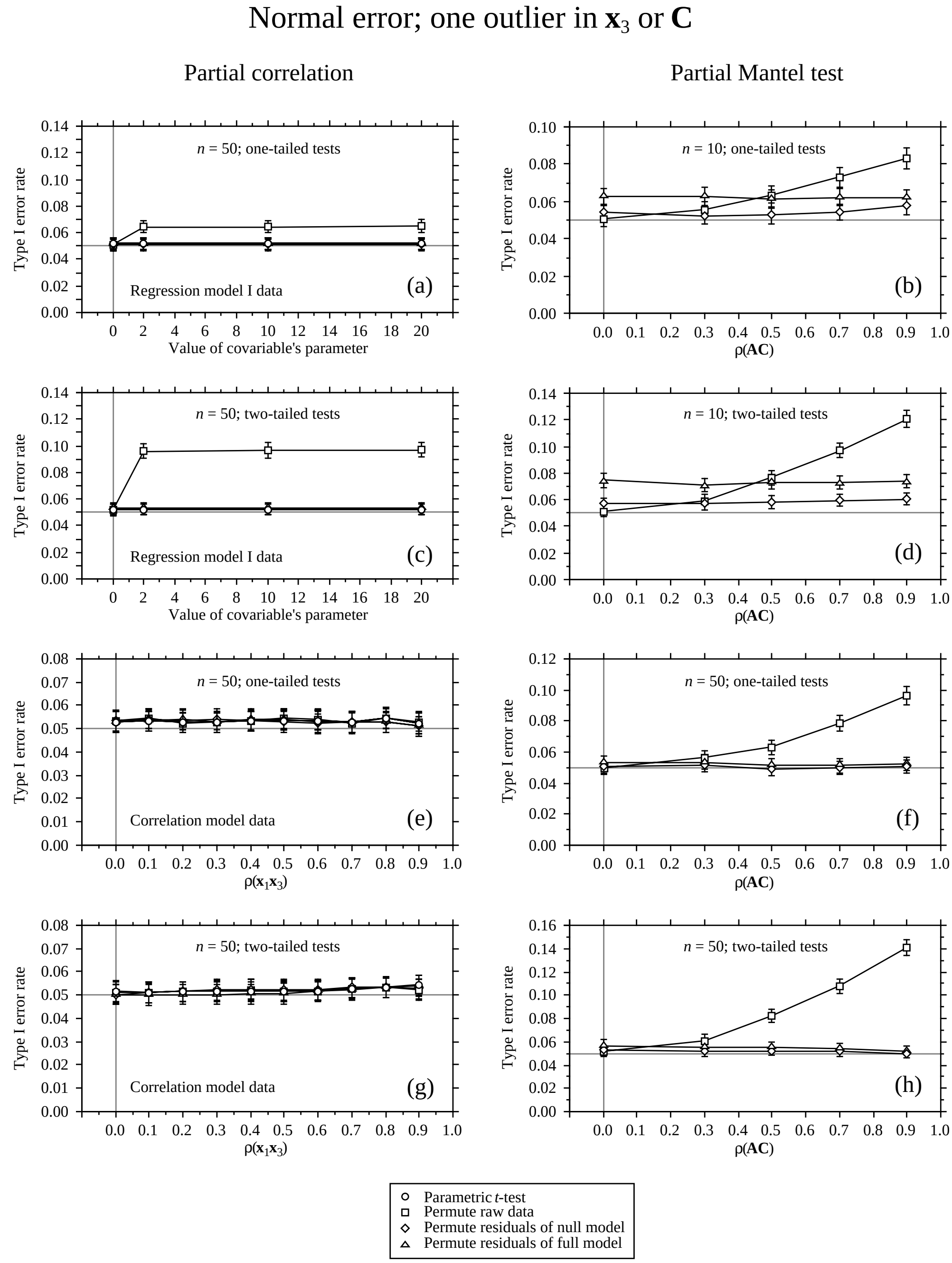


FIGURE 4 (a, c, e, g) Effect of an outlier in $\mathbf{x}_3$ on type I error of one-tailed (a, e) and two-tailed tests (c, g), for data generated under the regression model I (a, c: regression model with fixed predictors) or the correlation model (e, g) using normal error; $n = 50$. For the model I data, $\beta_0 = 0$, $\beta_1$ varied from 2 to 20 (abscissa of the graphs) and $\rho(\mathbf{x}_2\mathbf{x}_3) = 0$. For the correlation model data, $\rho(\mathbf{x}_1\mathbf{x}_2)$ and $\rho(\mathbf{x}_2\mathbf{x}_3)$ were 0; $\rho(\mathbf{x}_1\mathbf{x}_3)$ varied from 0.0 to 0.9 (abscissa of the graphs). (b, d) Same for partial Mantel tests (data generated under the correlation model) in one-tailed (b) and two-tailed tests (d); $n = 10$. (f, h) Same, $n = 50$.

However, combining highly skewed data with an outlier, permutation of the raw data produced a reduction in type I error in one-tailed tests as $\rho(\mathbf{x}_1\mathbf{x}_3)$ increased, and an inflation of type I error in two-tailed tests (Fig. 5). The normal-theory *t*-test had inflated type I error in most instances and mimicked the behavior of the permutation of raw data. In these extreme conditions, permutation of the residuals of a null or full model had mostly correct type I error for $n = 50$ or more (Figs. 5e, g; simulations were run for $n = 10$, 50 and 100).

These results are summarized in Table I. In the Table, method 3 (correlate residualized $\mathbf{x}_1$ to residualized $\mathbf{x}_2$) is marked as having incorrect type I error, not only because the simulations have shown it to be so for small samples, but also because it is intrinsically flawed, as explained in the Discussion. This is the reason why simulations have not been run for this method in the presence of outliers, nor in the power study.

#### *6.2.2. Power*

The power curves obtained with all testing methods, except method 3, are presented in Figures 6(a, c, e, g) for increasing values of $\rho(\mathbf{x}_2\mathbf{x}_3)$ and different sample sizes when $\rho(\mathbf{x}_1\mathbf{x}_2.\mathbf{x}_3) = 0.1$ (*i.e.*, the null hypothesis is false); $\rho(\mathbf{x}_1\mathbf{x}_3)$ was 0 in these simulations. Identical results were obtained when $\rho(\mathbf{x}_2\mathbf{x}_3)$ was 0 and $\rho(\mathbf{x}_1\mathbf{x}_3)$ was made to vary. Results of methods 1, 2 and 4 are indistinguishable from those of the parametric *t*-test. As expected, power increases with sample size. Similar results again were obtained by varying both $\rho(\mathbf{x}_1\mathbf{x}_3)$ and $\rho(\mathbf{x}_2\mathbf{x}_3)$ at the same time, as described in the Methods (results not shown).

### 6.3. Partial Mantel tests

#### *6.3.1. Type I Error*

Among the methods for partial Mantel analysis (Figs. 2b, d, f), only methods 1 (random permutation of the objects in matrix **A**) and 2 (permute residuals of null model) had correct type I error for all sample sizes, in simulations using standard normal deviates without outlier. Methods 3 and 4 systematically produced inflated type I error with small sample

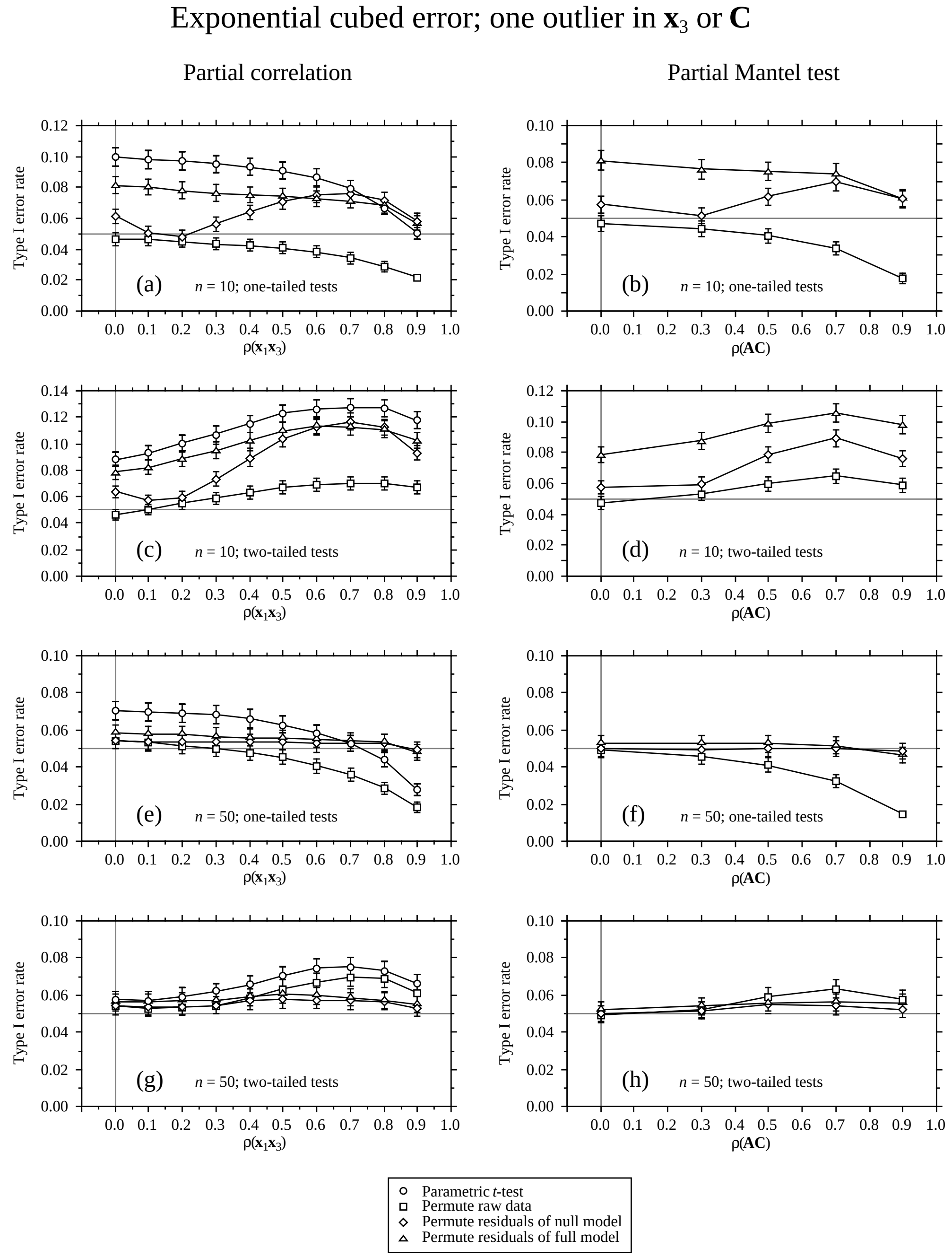


FIGURE 5 (a, c, e, g) Effect of an outlier in $\mathbf{x}_3$ (value = 50) on type I error of one-tailed (a, e) and two-tailed tests (c, g), for data generated under the correlation model with cubed exponential error; $n = 10$ and 50. The population correlations $\rho(\mathbf{x}_1\mathbf{x}_2)$ and $\rho(\mathbf{x}_2\mathbf{x}_3)$ were 0; $\rho(\mathbf{x}_1\mathbf{x}_3)$ varied from 0.0 to 0.9 (abscissa of the graphs). There were 10000 simulations per data point. (b, d, f, g) Same for partial Mantel tests in one-tailed (b) and two-tailed tests (d); $n = 10$. (f, h) Same, $n = 50$.

sizes. The error decreased with sample size, but it always remained. Section 6.2 has shown the problem inherent with method 3 (correlate residualized **A** to residualized **B**). Method 4 (permutation of residuals of a

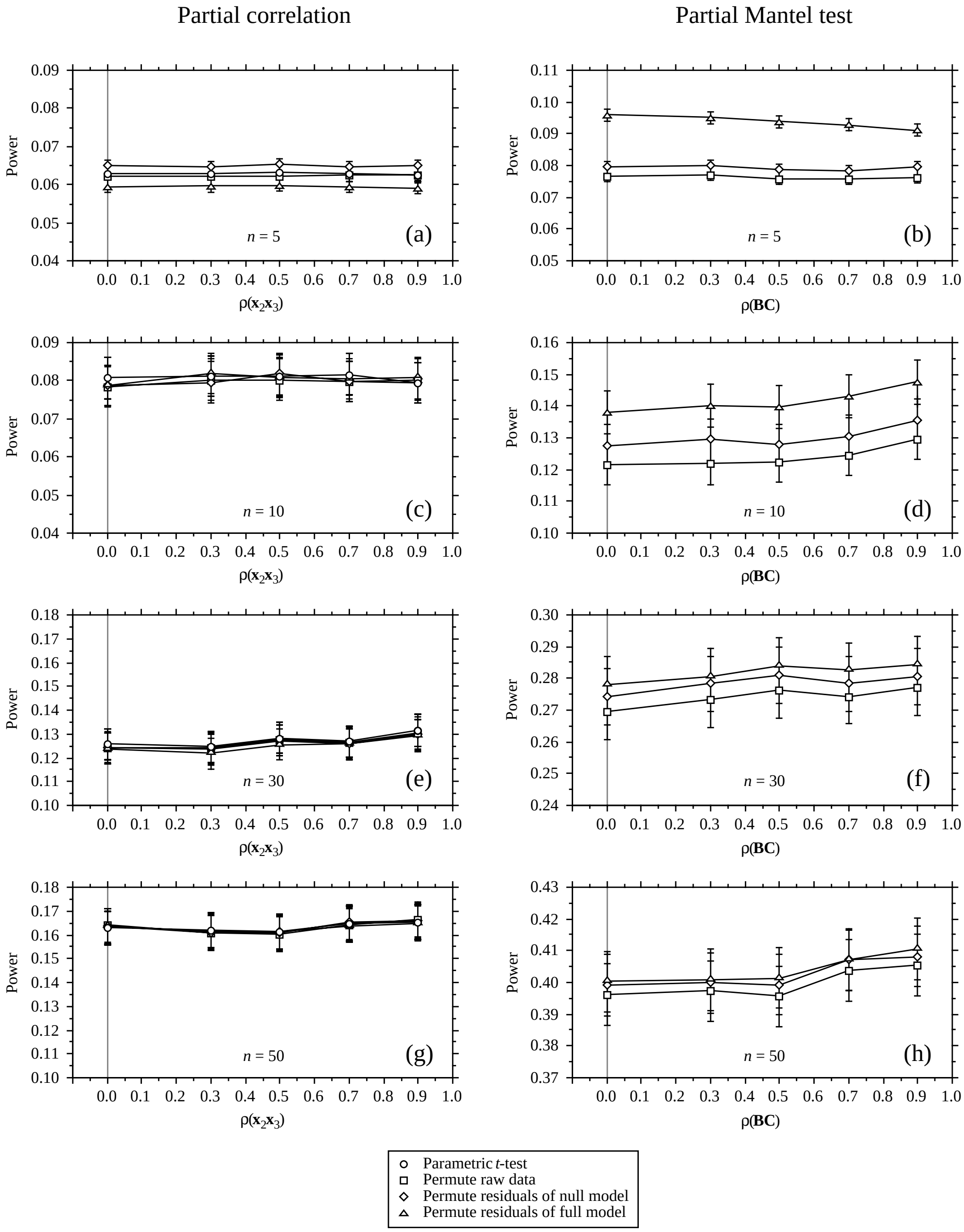


FIGURE 6 Mean and 95% confidence interval of power (at $\alpha = 0.05$), using different testing methods and sample sizes. *Left:* partial correlation analysis; $\rho(\mathbf{x}_2\mathbf{x}_3)$ varies from 0.0 to 0.9 while $\rho(\mathbf{x}_1\mathbf{x}_3) = 0$; $\rho(\mathbf{x}_1\mathbf{x}_2.\mathbf{x}_3) = 0.1$. *Right:* partial Mantel tests; $\rho(\mathbf{BC})$ varies from 0.0 to 0.9 while $\rho(\mathbf{AC}) = 0$; $\rho(\mathbf{AB.C}) = 0.1$.

full model) had inflated type I error up to $n = 40$; this problem was not present in partial correlation analysis. This is due, most likely, to the lack of correspondence between the error-generation process and the calculation of residuals, described in section 5.1, and to the fact that in matrix permutation, the residuals are not i.i.d. Method 2, which is also a method of permutation of residuals, does not present this problem.

Simulations carried out using highly skewed data in all three matrices led to the same results as in the partial correlation study (Figs. 3b, d). Permutation of the raw data was not affected by highly skewed data. The three methods of permutation of residuals had correct type I error for $n \geq 20$.

The effect of an outlier in matrix **C** differed from the findings of the partial correlation study. With normal error (Figs. 4b, d, f, g), the method of permutation of raw data was affected by having increasingly inflated type I error as $\rho(\mathbf{AC})$ increased. There was no difference between one-tailed and two-tailed tests. Inflation of type I error, already noted for method 4 (permutation of residuals of a full model) with small *n*, was also found here. The two methods of permutation of residuals behaved well for $n \geq 30$ (simulations were run for $n = 10$, 30 and 50). With highly skewed data and an outlier (Figs. 5b, d, f, g), permutation of the raw data had either inflated or reduced type I error as $\rho(\mathbf{AC})$ increased, depending on the test being one-tailed or two-tailed, as in the partial correlation study. The two methods of permutation of residuals maintained correct type I error for $n > 30$ (simulations were run for $n = 10$, 30 and 50).

These results are summarized in Table I. For the reason given in the results of the partial correlation study (Section 6.2), simulations have not been run for method 3 (correlate residualized **A** to residualized **B**) in the presence of outliers, nor in the power study.

### *6.3.2. Power*

The power curves obtained with all permutation methods, except method 3, are presented in Figures 6(b, d, f, h) for increasing values of $\rho(\mathbf{BC})$ and different sample sizes when $\rho(\mathbf{AB.C}) = 0.1$ (*i.e.*, the null hypothesis is false); $\rho(\mathbf{AC})$ was 0 in these simulations. Identical results were obtained when $\rho(\mathbf{BC})$ was 0 and $\rho(\mathbf{AC})$ was made to vary. The power curves of the three methods converge as *n* increases, although method 2 (permute residuals of null model) seems to maintain a slight (though not significant) advantage over the permutation of raw data. The apparent advantage of method 4 (permute residuals of full model) at small *n* simply results from it having inflated type I error; it should not be construed as a reason to use this method in partial Mantel tests. As expected, power increases with sample size.

## 7. DISCUSSION

The theory of permutation methods for testing partial regression coefficients for a univariate response is now well understood (Anderson and Legendre, 1999; Anderson and Robinson, in press), but using the same approach in the partial correlation case and for Mantel tests involving distance matrices presents problems. For partial correlations, the main problem is the lack of correspondence, in the methods of permutation of residuals, between the error generation procedure which follows from the correlation model and the estimation of residuals using model I regression. For distance matrices, the main problem is that the residuals of a linear regression model computed for distance matrices cannot be assumed to be i.i.d.; this remains a problem even though the residuals are permuted using ‘matrix permutation’, which corresponds to permuting the original observations. This paper used simulations to show empirically to what extent the various permutation methods described in section 4 behaved correctly for data generated under the correlation model, using different types of errors, and in the absence or presence of outliers.

The simulation results summarized in Table I show that, when the null hypothesis is true, permutation testing always leads to correct type I error in simple correlation analysis and simple Mantel tests for normal, skewed (exponential), or highly skewed data (cubed exponential). The parametric *t*-test used in correlation studies, however, has correct type I error only for normally distributed data.

In the partial correlation study, the parametric *t*-test had incorrect type I error in the same situations where it experienced problems in the simple correlation study, *i.e.*, in the presence of non-normal data. It did well, however, in the presence of outliers, a situation that caused problems to the method of permutation of raw data. The two methods of permutation of residuals (methods 2 and 4) produced equivalent results—correct type I error except for $n < 20$ in the presence of highly skewed data, and for $n < 50$ when cubed exponential deviates were combined with an outlier.

Although method 3 (correlate residualized $\mathbf{x}_1$ to residualized $\mathbf{x}_2$) had originally been proposed, in the context of multiple regression, as equivalent to method 2 but computationally simpler (Kennedy 1995), it actually has inflated type I error, especially for small sample sizes. According to Anderson and Legendre (1999), the reason for this anomaly, which will also be found in the partial Mantel study (below), is 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 (erroneously) assumed to remain zero during the permutations; this would only be correct if $n$ were infinite. Method 2, on the contrary, recognizes that a small non-zero correlation may appear between permuted residualized $\mathbf{x}_1$ and $\mathbf{x}_3$, due to the finite sample size, and it takes it into account during estimation. Anderson and Robinson (in press) 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 than for method 2 when compared to the permuted values, resulting in probabilities which are too small and thus inflating type I error. A numerical example is presented by Legendre (in preparation). It follows that method 3 should never be used.

The conclusions of the partial Mantel study are essentially the same, except for the fact that permutation of the residuals of a full model (method 4) had slightly inflated type I error for $n < 40$, even with normal error, a case where the permutation of residuals of a null model (method 2) had correct type I error.

The study leads to the following recommendations to users concerning testing procedures that seem appropriate for partial correlation analysis and partial Mantel tests:

- In partial correlation analysis (original variables), all testing procedures except method 3 can be used most of the time. The parametric $t$-test should not be used with highly skewed data. Permutation of the raw data should be avoided only when highly skewed data are combined with outliers in $\mathbf{x}_3$. The methods of permutation of residuals, which are known to only have asymptotically exact significance levels, should not be used when highly skewed data are combined with small sample size ($n < 20$); *very* highly skewed data (cubed exponential deviates) were used in the simulations. Highly skewed data can be detected by looking at frequency histograms, examining the skewness parameter of the distribution, and/or performing tests of normality on the three data vectors.

- In partial Mantel tests (distance matrices), method 2 (permute residuals of null model) can always be used, except when highly skewed data are combined with small sample size ($n < 20$, or $n < 50$ in the presence of outliers). For distance matrices computed using functions other than the Euclidean distance, highly skewed data can be detected by examining ordination diagrams obtained for each distance matrix, using the method of principal coordinate analysis (Gower, 1966); examining the distribution and skewness parameter of each individual ordination axis; and/or performing tests of multivariate normality on principal coordinates computed for the three distance matrices. Permutation of the raw data can be used with normal or highly skewed data in the absence of extreme outliers. Method 4 (permute residuals of full model) can only be used safely with medium to large sample sizes ($n > 40$).
- With small sample sizes, one should carefully examine the data before partial correlation or partial Mantel analysis. For highly skewed data, permutation of the raw data has correct type I error in the absence of outliers (Fig. 3). When highly skewed data are combined with outliers in $\mathbf{x}_3$ or $\mathbf{C}$ (Fig. 5), it is recommended to use the permutation of raw data in partial correlation analysis or partial Mantel tests, for the following reasons: (1) with one-tailed tests, the results are conservative; so, one is less likely to incorrectly reject $H_0$; (2) with two-tailed tests, permutation of the raw data has the least amount of inflation of type I error of all the testing procedures investigated in this study; this effect can be compensated by using a conservative significance level, *e.g.*, $\alpha = 0.01$ instead of 0.05. Outliers can be detected by examining frequency distributions or multivariate ordination diagrams, or by cluster analysis.
- Method 3 (correlate residualized $\mathbf{x}_1$ or $\mathbf{A}$ to residualized $\mathbf{x}_2$ or $\mathbf{B}$) should never be used.

The above recommendations depend to a large extent on the arbitrary decisions that were made about the simulation programme (*e.g.*, normal and highly skewed data, with or without an outlier). Simulations carried out using other types of data (*e.g.*, with many tied values) could lead to complementary recommendations.

## *Acknowledgments*

I am most thankful to M. J. Anderson for comments on the penultimate version of the manuscript, and to an anonymous reviewer for pertinent suggestions. This work was supported by NSERC grants OGP0007738 and EQP0196089 to P. Legendre.

## *References*

Anderson, M. J. and Legendre, P. (1999) An empirical comparison of permutation methods for tests of partial regression coefficients in a linear model, *J. Statist. Comput. Simulation*, **62**, 271 – 303.

Anderson, M. J. and Robinson, J., Permutation tests for linear models, *Aust. N. Z. J. Stat.* (in press).

Anselin, L. (1995) Local indicators of spatial association — LISA, *Geogr. Anal.*, **27**, 93 – 115.

Cohen, J. (1988) *Statistical Power Analysis for the Behavioral Sciences* (Second Edition). Hillsdale, New Jersey: Lawrence Erlbaum Assoc.

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

Dutilleul, P., Stockwell, J. D., Frigon, D. and Legendre, P. (2000) The Mantel-Pearson paradox: statistical considerations and ecological implications, *J. Agric. Biol. Environ. Stat.*, **5**(2), 131 – 150.

Edgington, E. S. (1995) *Randomization Tests* (Third Edition). New York: Marcel Dekker.

Freedman, D. and Lane, D. (1983) A nonstochastic interpretation of reported significance levels, *J. Bus. Econ. Statist.*, **1**, 292 – 298.

Furnas, G. W. (1984) The generation of random, binary unordered trees, *J. Classif.*, **1**, 187 – 233.

Gower, J. C. (1966) Some distance properties of latent root and vector methods used in multivariate analysis, *Biometrika*, **53**, 325 – 338.

Gower, J. C. and Legendre, P. (1986) Metric and Euclidean properties of dissimilarity coefficients, *J. Classif.*, **3**, 5 – 48.

Hall, P. and Titterington, D. M. (1989) The effect of simulation order on level accuracy and power of Monte Carlo tests, *J. R. Statist. Soc. B*, **51**, 459 – 467.

Hope, A. C. A. (1968) A simplified Monte Carlo test procedure, *J. R. Statist. Soc. B*, **50**, 35 – 45.

Hubert, L. J. and Golledge, R. G. (1981) A heuristic method for the comparison of related structures, *J. Math. Psychol.*, **23**, 214 – 226.

Kennedy, P. E. (1995) Randomization tests in econometrics, *J. Bus. Econ. Statist.*, **13**, 85 – 94.

Krackhardt, D. (1988) Predicting with networks: nonparametric multiple regression analysis of dyadic data, *Soc. Networks*, **10**, 359 – 381.

Legendre, P. Permutation methods for partial Mantel tests: practical aspects (in preparation).

Legendre, P., Lapointe, F.-J. and Casgrain, P. (1994) Modeling brain evolution from behavior: a permutational regression approach, *Evolution*, **48**, 1487 – 1499.

Legendre, P. and Legendre, L. (1998) *Numerical Ecology* (Second English Edition). Amsterdam: Elsevier Science B.V.

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

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

Manly, B. J. F. (1997) *Randomization, Bootstrap and Monte Carlo Methods in Biology* (Second Edition). London: Chapman and Hall.

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

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

Smouse, P. E., Long, J. C. and Sokal, R. R. (1986) Multiple regression and correlation extensions of the Mantel test of matrix correspondence, *Syst. Zool.*, **35**, 627–632.

Sokal, R. R. and Rohlf, F. J. (1995) *Biometry – The Principles and Practice of Statistics in Biological Research* (Third Edition). New York: W. H. Freeman.

ter Braak, C. J. F. (1990) *Update Notes: CANOCO version 3.10.* Wageningen, The Netherlands: Agricultural Mathematics Group.

ter Braak, C. J. F. (1992) Permutation versus bootstrap significance tests in multiple regression and ANOVA. In: *Bootstrapping and Related Techniques* (Eds., Jöckel, K.-H., Rothe, G. and Sendler, W.), pp. 79 – 86, Berlin: Springer-Verlag.

ter Braak, C. J. F. and Smilauer, P. (1998) *CANOCO Reference Manual and User's Guide to Canoco for Windows: Software for Canonical Community Ordination (version 4).* Ithaca, New York: Microcomputer Power.

Due to the small size of the journal and the poor quality of printing, the figures were not clearly presented in the Journal and in the reprints received from the Publisher. The present facsimile was prepared, using Frame Maker, to address these problems.