Ecology, 89(12), 2008, pp. 3400–3412

TESTING THE SPECIES TRAITS–ENVIRONMENT RELATIONSHIPS: THE FOURTH-CORNER PROBLEM REVISITED

Stéphane Dray[1,3] and Pierre Legendre[2]

[1]*Université de Lyon, Université Lyon 1, CNRS; UMR 5558, Laboratoire de Biométrie et Biologie Evolutive, 43 Boulevard du 11 Novembre 1918, Villeurbanne F-69622 France*
[2]*Département de Sciences Biologiques, Université de Montréal, C.P. 6128, Succursale Centre-ville, Montréal, Québec H3C 3J7 Canada*

Abstract. Functional ecology aims at determining the relationships between species traits and environmental variables in order to better understand biological processes in ecosystems. From a methodological point of view, this biological objective calls for a method linking three data matrix tables: a table **L** with abundance or presence–absence values for species at a series of sites, a table **R** with variables describing the environmental conditions of the sites, and a table **Q** containing traits (e.g., morphological or behavioral attributes) of the species. Ten years ago, the fourth-corner method was proposed to measure and test the relationships between species traits and environmental variables using tables **R**, **L**, and **Q** simultaneously. In practice, this method is rarely used. The major reasons for this lack of interest are the restriction of the original method and program to presence–absence data in **L** and to the analysis of a single trait and a single environmental variable at a time. Moreover, ecologists often have problems in choosing a permutation model among the four originally proposed. In this paper, we revisit the fourth-corner method and propose improvements to the original approach. First, we present an extension to measure the link between species traits and environmental variables when the ecological community is described by abundance data. A new multivariate fourth-corner statistic is also proposed. Then, using numerical simulations, we discuss and evaluate the existing testing procedures. A new two-step testing procedure is presented. We hope that these elements will help ecologists use the best possible methodology to analyze this type of ecological problem.

Key words: ecological community; fourth-corner statistic; functional ecology; permutational model; RLQ analysis; species traits.

Introduction

Southwood (1977) proposed the habitat templet theory, which assumes that "habitat provides the templet on which evolution forges characteristic life-history strategies" (Southwood 1988:3). Under this hypothesis, functional ecology has been developed with the objectives of "(1) constructing trait matrices through screening; (2) exploring empirical relationships among these traits; and (3) determining the relationships between traits and environments" (Keddy 1992*b*:621). Using species traits (morphological or behavioral attributes, for instance) as a surrogate to species taxonomy allows researchers to produce more general ecological models: "because the problem primarily involves traits and environment, answers should be generalizable to systems with very different taxonomic composition" (Keddy 1992*a*:157). This approach enables researchers to compare communities at a very broad spatial scale even when the lists of species are entirely distinct or even unknown (Blondel et al. 1984, Wiens 1991, Lamouroux et al. 2002). Species traits are useful in searching for functional types (combinations of attributes) in order to reduce the diversity of species to a diversity of functions. This approach is suitable to reduce the complexity of ecosystems in order to assess ecosystem sensitivity against changes in the environment (McIntyre et al. 1999). Field studies (e.g., Diaz et al. 1998) allow ecologists to select traits of interest, identify functional group, and study their responses to environmental variation. Results from these studies are essential for constructing models that predict the effects of environmental changes (e.g., disturbance, pollution, climatic change) on ecosystems (Campbell et al. 1999, Pausas 1999). This approach provides an efficient biomonitoring tool and could be used for conservation purposes (Dolédec et al. 1999).

From a methodological point of view, introducing a species trait matrix into an ecological analysis represents an exciting challenge. In most situations, considering species traits in a study implies that ecologists have to analyze three tables: a table **L** ($n \times p$) containing the abundances of p species at n sites, a second table **R** ($n \times m$) with the measurements of m environmental variables for the n sites, and a third table **Q** ($p \times s$) describing s traits for the p species. Some authors have simplified this situation by constructing a site × trait matrix ($n \times s$) in order to obtain two tables containing data about the

Manuscript received 19 February 2008; accepted 23 April 2008. Corresponding Editor: J. Franklin.
[3] E-mail: dray@biomserv.univ-lyon1.fr

same statistical units, the sites (Mabry et al. 2000). Other researchers use a two-step approach (indirect functional analysis) where tables **R** and **L** are first analyzed and then the results are interpreted using species trait data (Thuiller et al. 2004). Obviously, a direct functional analysis where species data, environmental data, and species-trait data are analyzed simultaneously represents a more optimal solution. The methodological question of linking species traits to environmental variables has been called the fourth-corner problem because of the nature of the matrix formulation that was used by Legendre et al. (1997) to solve it. These authors proposed four testing procedures to evaluate the link between an environmental variable and a species trait through a table **L** containing presence–absence data. One year before, Dolédec et al. (1996) developed RLQ analysis, a three-table ordination method. Although methods designed for the analysis of three tables have been available for a few years, there are few published applications (but see Barbaro et al. 2000, Charest et al. 2000, Ribera et al. 2001, Hausner et al. 2003, Hooper et al. 2004). Some reasons for this lack of interest can be found in the literature. Poff et al. (2006:731) argued that "despite much progress in recent years, the full potential of the functional traits-based approach is currently limited by several factors, both conceptual and methodological." They added on p. 732 that "the overarching need is to develop a more robust multivariate framework, in which the responses of multiple, independent traits can be related to multiple environmental gradients characteristic of most landscapes." They regretted that RLQ analysis and the fourth-corner method "do not explicitly account for evolutionary linkage of traits … and consider only a single trait at a time, and some can analyze only binary species data" and concluded that "a basic statistical challenge remains." Nygaard and Ejrnaes (2004:54) stated that "the fourth corner method has been developed for binary species by sample data sets, and an extension of the method to include species abundance data has not yet been developed." They incorrectly considered that "this statistical feature works only on three-table data sets where either matrix **B** (species × attributes matrix) or **C** (sample × environment matrix) is represented by binary variables. This restriction puts another severe limit to the applicability of the method in functional plant ecology." In short, ecologists regretted that the testing procedure and computer program of Legendre et al. (1997) were only devoted to species presence–absence data (this can be seen as a loss of information when abundance data are available) and considered only univariate measurements. Moreover, it appears that ecologists did not know how to choose among the four permutation methods proposed by Legendre et al. (1997), and tended to always use the first method, the one that had been used in the example contained in the original paper.

In this paper, we focus on testing procedures for direct functional analysis. First, we show how to measure the link between species traits and environmental variables and we propose an extension of the fourth-corner method of Legendre et al. (1997) to deal with abundance data. We propose also a multivariate statistic to consider several traits and environmental variables in a single analysis. Then, testing procedures are evaluated using simulated data representing various ecological situations. Last, we propose a two-step approach to test properly the relationships between species traits and environmental characteristics.

Measuring the Link between Species Traits and Environmental Characteristics

In this section, we show how to link environmental variables and species traits. We extend to abundance data the statistics proposed by Legendre et al. (1997) for presence–absence data. We first consider bivariate measurements, and then we develop a new multivariate statistic.

Inflating the original data tables

Table **L** ($n \times p$) contains the abundances of p species at n sites. Let $\mathbf{P} = [P_{ij}]$ be the table of relative frequencies (fractions of total abundance) with $P_{ij} = L_{ij}/L_{++}$ where L_{ij} is the abundance of the jth species in the ith site and $L_{++} = \Sigma_{i=1}^{n} \Sigma_{j=1}^{p} L_{ij}$ is the grand total computed from table **L**. The row and column weights derived from table **L** are denoted by

$$P_{i+} = \frac{L_{i+}}{L_{++}} = \sum_{j=1}^{p} \frac{L_{ij}}{L_{++}}$$

and

$$P_{+j} = \frac{L_{+j}}{L_{++}} = \sum_{i=1}^{n} \frac{L_{ij}}{L_{++}} \tag{1}$$

as in correspondence analysis. Let us consider the diagonal matrices of site and species weights defined by

$$\mathbf{D}_n = \mathrm{Diag}(P_{1+}, \ldots, P_{i+}, \ldots, P_{n+})$$

and

$$\mathbf{D}_p = \mathrm{Diag}(P_{+1}, \ldots, P_{+j}, \ldots, P_{+p}). \tag{2}$$

Table **R** ($n \times m$) describes the environment while species traits are contained in table **Q** ($p \times s$). The variables in **R** and **Q** can be quantitative or qualitative. Following Legendre et al. (1997), the information contained in tables **R**, **L**, and **Q** can be rewritten in the form of inflated tables. These tables correspond to a simple rewriting of the original tables where only nonempty cells are considered. Since we consider species abundance data, two procedures can be used for inflating the original data tables (Fig. 1).

The first way (Fig. 1b) considers the correspondences, which are the c nonempty cells of table **L** (Greenacre

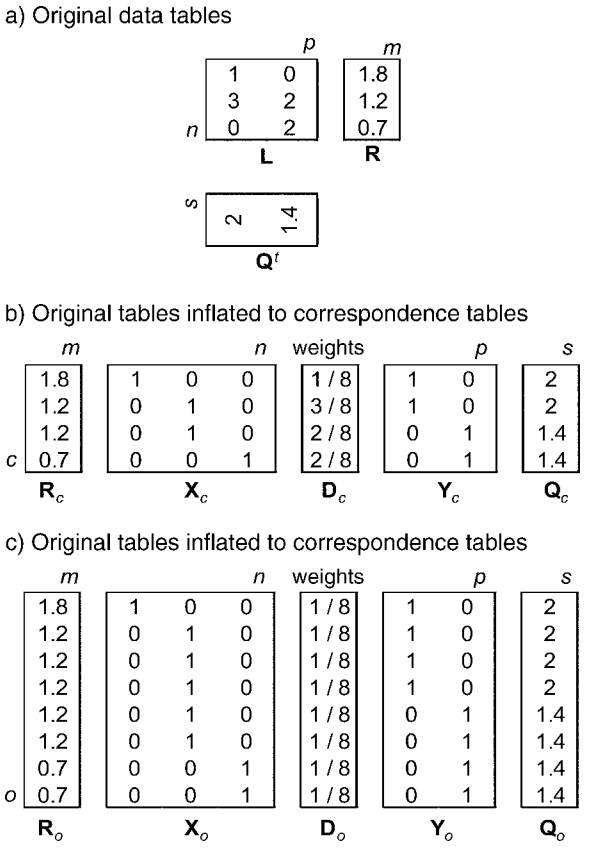

FIG. 1. A small example showing how to inflate the original tables (a) in the case of abundance data. When inflating to correspondence tables (b), the number of rows of the new tables is equal to the number of nonempty cells of table **L**, and we have $\mathbf{L} = \Sigma_{i=1}^{n}\Sigma_{j=1}^{p} L_{ij}\mathbf{X}_c^t\mathbf{D}_c\mathbf{Y}_c$. Weights proportional to the abundance values are assigned by matrix $\mathbf{D}_c$ to each correspondence. When inflating to occurrence tables (c), the number of rows of the new tables is equal to the number of individuals of table **L**, and we have $\mathbf{L} = \mathbf{X}_o^t\mathbf{Y}_o = \Sigma_{i=1}^{n}\Sigma_{j=1}^{p} L_{ij}\mathbf{X}_o^t\mathbf{D}_o\mathbf{Y}_o$. In this case, all occurrences have the same weight. For presence–absence data, the two modes of inflation are equivalent to the inflation proposed by Legendre et al. (1997). See *Measuring the link . . . : Inflating the original data tables* for details.

1984, Thioulouse and Chessel 1992). We can easily derive two tables $\mathbf{X}_c(c \times n)$ and $\mathbf{Y}_c$ $(c \times p)$ from table **L**. For the kth correspondence (i.e., kth row of $\mathbf{X}_c$ and $\mathbf{Y}_c$), we have

$$\mathbf{X}_c(k,i) = \begin{cases} 1 & \text{if the } k\text{th correspondence belongs to the } i\text{th site} \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbf{Y}_c(k,j) = \begin{cases} 1 & \text{if the } k\text{th correspondence belongs to the } j\text{th species} \\ 0 & \text{otherwise} \end{cases}$$

with $1 \le k \le c$, $1 \le i \le n$, and $1 \le j \le p$. This derivation can be performed in a column-wise or row-wise manner, for instance.

Two inflated tables $\mathbf{R}_c(c \times m)$ and $\mathbf{Q}_c(c \times s)$ are also constructed by duplicating the values of tables **R** and **Q**, respectively, according to the distribution of correspondences in tables $\mathbf{X}_c$ and $\mathbf{Y}_c$. If the kth correspondence belongs to the ith site and the jth species, then the kth row of $\mathbf{R}_c$ is equal to the ith row of **R**, and the kth row of $\mathbf{Q}_c$ is equal to the jth row of **Q**. Lastly, a weight must be associated to each correspondence, chosen to be a function of the abundance values in table **L**. A diagonal matrix $\mathbf{D}_c$ is therefore constructed where $\mathbf{D}_c(k, k) = L_{ij}/L_{++}$ if the kth correspondence belongs to the ith site and the jth species.

The second way to inflate the original data (Fig. 1c) considers the occurrences (i.e., the o individuals of table **L**). By definition, we have $o = L_{++}$. This inflating method is conceptually linked to the analysis of real occurrence data where the sampling sites are not considered (Pélissier et al. 2002, 2003, Gimaret-Carpentier et al. 2003). Two tables $\mathbf{X}_o(o \times n)$ and $\mathbf{Y}_o(o \times p)$ are derived from table **L**. For the kth occurrence (i.e., kth row of $\mathbf{X}_o$ and $\mathbf{Y}_o$) we have

$$\mathbf{X}_o(k,i) = \begin{cases} 1 & \text{if the } k\text{th occurrence belongs to the } i\text{th site} \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbf{Y}_o(k,j) = \begin{cases} 1 & \text{if the } k\text{th occurrence belongs to the } j\text{th species} \\ 0 & \text{otherwise.} \end{cases}$$

Inflated tables $\mathbf{R}_o(o \times m)$ and $\mathbf{Q}_o(o \times s)$ are also constructed by duplicating the values of tables **R** and **Q**, respectively. For this inflating approach, weights associated to occurrences are uniform ($\forall k$, $\mathbf{D}_o(k, k) = 1/o$). Note that while the inflation to correspondence tables can be performed for integer as well as continuous abundance measurements, the construction of the tables of occurrences is restricted to integer counts. However, the statistics proposed in the sequel to measure the link between species traits and environmental variables using the original data tables can also be used for abundances measured on a continuous scale (e.g., relative abundances). If table **L** contains only presence–absence data, which was the case considered by Legendre et al. (1997), the two mechanisms for inflating the original data tables are equivalent.

Linking two quantitative variables

Consider that **R** and **Q** each contain a single quantitative variable, as in the illustrative example proposed in Fig. 1. The link between **R** and **Q** can easily be measured from the pair of inflated tables $\mathbf{R}_o$ and $\mathbf{Q}_o$, which have the same number of rows. If the two variables in $\mathbf{R}_o$ and $\mathbf{Q}_o$ are standardized to means $\mathbf{Q}_o^t\mathbf{D}_o\mathbf{1}_o = \mathbf{R}_o^t\mathbf{D}_o\mathbf{1}_o = 0$ where $\mathbf{1}_o$ is a vector of 1 with o rows, and variances $\mathbf{Q}_o^t\mathbf{D}_o\mathbf{Q}_o = \mathbf{R}_o^t\mathbf{D}_o\mathbf{R}_o = 1$, then $\mathbf{Q}_o^t\mathbf{D}_o\mathbf{R}_o$ is a Pearson correlation coefficient r:

$$r = \mathbf{Q}_o^t\mathbf{D}_o\mathbf{R}_o = \mathbf{Q}_c^t\mathbf{D}_c\mathbf{R}_c = \mathbf{Q}^t\mathbf{P}^t\mathbf{R}. \quad (3)$$

(see Appendix C for details.)

Hence we can see that the link between a quantitative trait and a quantitative environmental variable can be computed in three different ways:

1) A correlation coefficient using the occurrence tables ($r = \mathbf{Q}_o^t\mathbf{D}_o\mathbf{R}_o$);

2) A weighted correlation coefficient using the correspondence tables ($r = \mathbf{Q}_c^t\mathbf{D}_c\mathbf{R}_c$). In that case, $\mathbf{Q}_c$ and $\mathbf{R}_c$ are standardized to means $\mathbf{Q}_c^t\mathbf{D}_c\mathbf{R}_c = \mathbf{R}_c^t\mathbf{D}_c\mathbf{1}_c = 0$ and variances $\mathbf{Q}_c^t\mathbf{D}_c\mathbf{Q}_c = \mathbf{R}_c^t\mathbf{D}_c\mathbf{R}_c = 1$ using the weights $\mathbf{D}_c$;

3) A weighted cross-correlation coefficient using the original tables ($r = \mathbf{Q}^t\mathbf{P}^t\mathbf{R}$). In that case, $\mathbf{Q}$ and $\mathbf{R}$ have to be first standardized to means $\mathbf{Q}^t\mathbf{D}_p\mathbf{1}_p = \mathbf{R}^t\mathbf{D}_n\mathbf{1}_n = 0$ and variances $\mathbf{Q}^t\mathbf{D}_p\mathbf{Q} = \mathbf{R}^t\mathbf{D}_n\mathbf{R} = 1$ using the weights $\mathbf{D}_p$ and $\mathbf{D}_n$, respectively.

The three calculation methods produce identical values of r. There is also another way to compute this value, which has a very clear ecological meaning. If we consider that species have unimodal responses to the environmental variable, niche centroids can be computed by weighted averaging (using species abundance values) of the environmental variable. The value of r is then equal to the slope of the linear model, weighted by total species abundances, with the niche centroids as the response variable and the species trait as the explanatory variable. Hence the fourth-corner statistic is a measure of the link between the characteristics (traits) of species and their positions along the environmental gradient. If table $\mathbf{L}$ contains presence–absence data, this statistic corresponds exactly to the Pearson r statistic proposed by Legendre et al. (1997).

Linking two qualitative variables

We consider now a qualitative environmental variable (k_r categories) and a qualitative species trait (k_q categories). Data are coded in $\mathbf{R}$ and $\mathbf{Q}$ by k_r and k_q dummy variables, respectively. For that case, one can create a $k_r \times k_q$ contingency table from tables $\mathbf{R}_o$ and $\mathbf{Q}_o$, and then compute a χ^2 statistic to measure the link between species trait and environment. The contingency table is obtained by the product $\mathbf{R}_o^t\mathbf{Q}_o$. From the table of proportions $\mathbf{R}_o^t\mathbf{D}_o\mathbf{Q}_o$, we derive diagonal matrices of row and column totals $\mathbf{D}_{k_r} = \mathbf{R}_o^t\mathbf{D}_o\mathbf{R}_o$ and $\mathbf{D}_{k_q} = \mathbf{Q}_o^t\mathbf{D}_o\mathbf{Q}_o$. The Pearson χ^2 statistic is given by

$$\chi^2 = o \cdot \text{trace}[\mathbf{D}_{k_q}^{-1}(\mathbf{R}_o^t\mathbf{D}_o\mathbf{Q}_o - \mathbf{D}_{k_r}\mathbf{1}_{k_r}\mathbf{1}_{k_q}^t\mathbf{D}_{k_q})^t \times \mathbf{D}_{k_r}^{-1}(\mathbf{R}_o^t\mathbf{D}_o\mathbf{Q}_o - \mathbf{D}_{k_r}\mathbf{1}_{k_r}\mathbf{1}_{k_q}^t\mathbf{D}_{k_q})] \quad (4)$$

where trace designates the sum of the diagonal elements of a matrix. After some manipulations (see Appendix C), we can rewrite the statistic as follows:

$$\chi^2 = o \cdot \text{trace}[(\mathbf{D}_{k_q}^{-1}\mathbf{Q}^t\mathbf{P}^t\mathbf{R}\mathbf{D}_{k_r}^{-1} - \mathbf{1}_{k_q}\mathbf{1}_{k_r}^t) \times \mathbf{D}_{k_r}(\mathbf{D}_{k_r}^{-1}\mathbf{R}^t\mathbf{P}\mathbf{Q}\mathbf{D}_{k_q}^{-1} - \mathbf{1}_{k_r}\mathbf{1}_{k_q}^t)\mathbf{D}_{k_q}]. \quad (5)$$

The components of the χ^2 statistic in individual cells of the contingency table, proposed by Legendre et al. (1997) to evaluate the link between categories of the two variables, can also be easily computed for abundance data.

Linking one qualitative variable and one quantitative variable

We consider now the case of a qualitative environmental variable (k_r categories) and a quantitative species trait. This choice is arbitrary, and the results presented here could easily be derived for the opposite and symmetric case of a quantitative environmental variable and a qualitative species trait. Data are coded in $\mathbf{R}$ by k_r dummy variables while $\mathbf{Q}$ contains a single quantitative variable. For that case, Legendre et al. (1997) proposed to compute an ANOVA-like pseudo-F statistic from tables $\mathbf{R}_o$ and $\mathbf{Q}_o$. One can also compute a correlation ratio by dividing the among-group sum of squares by the total sum of squares. If the variable in $\mathbf{Q}_o$ is standardized to mean 0 ($\mathbf{Q}_o^t\mathbf{D}_o\mathbf{1}_o = 0$) and variance 1 ($\mathbf{Q}_o^t\mathbf{D}_o\mathbf{Q}_o = 1$), the correlation ratio is given by

$$\eta^2 = \text{trace}\left[(\mathbf{R}_o\mathbf{D}_{k_r}^{-1}\mathbf{R}_o^t\mathbf{D}_o\mathbf{Q}_o)^t\mathbf{D}_o(\mathbf{R}_o\mathbf{D}_{k_r}^{-1}\mathbf{R}_o^t\mathbf{D}_o\mathbf{Q}_o)\right]. \quad (6)$$

This correlation ratio can be rewritten using the original data tables (see Appendix C):

$$\eta^2 = \text{trace}\left[(\mathbf{Q}^t\mathbf{P}^t\mathbf{R})\mathbf{D}_{k_r}^{-1}(\mathbf{Q}^t\mathbf{P}^t\mathbf{R})^t\right]. \quad (7)$$

In that case, $\mathbf{Q}$ is standardized to variance 1 ($\mathbf{Q}^t\mathbf{D}_p\mathbf{Q} = 1$) and mean 0 ($\mathbf{Q}^t\mathbf{D}_p\mathbf{1}_p = 0$) using the weights $\mathbf{D}_p$. The statistics proposed by Legendre et al. (1997) to evaluate the link between categories and a quantitative variable can also be extended to the case of abundance values.

A multivariate statistic

The statistics presented in Eqs. 3, 5, and 7 show how to measure the link between a single species trait and a single environmental variable using the original tables $\mathbf{R}$, $\mathbf{L}$, and $\mathbf{Q}$. According to the types of variables, one can compute a Pearson correlation coefficient (r), a χ^2 statistic, or a correlation ratio (η^2). In this part, we consider a general measure of the link between environment and species traits when matrices $\mathbf{R}$ and $\mathbf{Q}$ contain several quantitative as well as qualitative variables. Quantitative variables are standardized to mean 0 and variance 1 using weighting matrices $\mathbf{D}_p$ (for $\mathbf{Q}$) and $\mathbf{D}_n$ (for $\mathbf{R}$). Qualitative variables are coded using dummy variables. As in Legendre et al. (2002), the proposed statistic is the trace of the fourth-corner matrix. It is a generalization of Eqs. 3, 5, and 7, and is equal to

$$S_{\mathbf{RLQ}} = \text{trace}(\mathbf{Z}\mathbf{D}_r\mathbf{Z}^t\mathbf{D}_q) \quad (8)$$

with $\mathbf{Z} = (\mathbf{Q}\mathbf{D}_q^{-1})^t(\mathbf{P} - \mathbf{D}_n\mathbf{1}_n\mathbf{1}_p^t\mathbf{D}_p)^t(\mathbf{R}\mathbf{D}_r^{-1})$. Diagonal matrices $\mathbf{D}_r$ and $\mathbf{D}_q$ contain weights for columns of $\mathbf{R}$ and $\mathbf{Q}$. If the kth column of $\mathbf{R}$ ($\mathbf{Q}$, respectively)

corresponds to a quantitative variable, then $\mathbf{D}_r(k, k) = 1$ ($\mathbf{D}_q(k, k) = 1$, respectively). If the kth column of $\mathbf{R}$ denoted $\mathbf{R}^{(k)}$($\mathbf{Q}$ denoted $\mathbf{Q}^{(k)}$, respectively) corresponds to a dummy variable, then $\mathbf{D}_r(k, k) = \mathbf{R}^{(k)t}\mathbf{D}_n\mathbf{R}^{(k)}$ ($\mathbf{D}_q(k, k) = \mathbf{Q}^{(k)t}\mathbf{D}_p\mathbf{Q}^{(k)}$, respectively).

If we consider a single species trait and a single environmental variable, the proposed statistic $S_{\mathbf{RLQ}}$ is equal to r^2, χ^2/L_{++}, or η^2 depending on the types of variables. When considering both quantitative and qualitative variables in $\mathbf{R}$ and $\mathbf{Q}$, this statistic is simply a sum of r^2, χ^2/L_{++}, and η^2 for all combinations of species traits and environmental variables. It can be demonstrated that this quantity is also equal to the total inertia of an RLQ analysis (Dolédec et al. 1996).

Testing Ecological Hypotheses

Quantification of the link between species traits and environment, using the statistics proposed in the previous section, is a first step toward testing ecological hypotheses. The second step is to evaluate if the strength of the link may be attributed to chance alone, which is the most parsimonious hypothesis, or if it is likely to reflect ecological processes. In particular, we want to know if the structure of the community significantly denotes an association between the characteristics of the species and the environmental conditions. As shown in the previous section, the relationship between species traits and environmental variables can be measured using traditional statistics through inflated occurrence data tables. Classical testing procedures (e.g., t test for the Pearson correlation coefficient r) cannot be used, however, because the reference distribution of the fourth-corner statistic is unknown. The sampling unit of the analysis is the site; therefore the testing procedure must be applied to the original data tables, not the inflated tables. Permutation procedures can be implemented to solve this problem.

Permutation models

The principles of the randomization procedure are the following:

1) Compute a reference value of the statistic using the original data tables (e.g., χ^2 in the case of two qualitative variables).

2) Permute at random the values in $\mathbf{L}$ and recompute the statistic. This operation is repeated a number of times (e.g., 999) to obtain a set of values of the statistic under the null hypothesis H_0.

3) Compare the observed statistic to the distribution containing the values obtained by permutation as well as the reference value; compute the associated probability and take the appropriate statistical decision.

In the case of presence–absence species data, Legendre et al. (1997) proposed four methods of permuting table $\mathbf{L}$ to test the null hypothesis (H_0) that the species traits (table $\mathbf{Q}$) are unrelated to the characteristics of the sites (table $\mathbf{R}$), their relationships (links) being mediated by the species presence–absence data (table $\mathbf{L}$). The four permutation models are the following:

1) Model 1: Permute presence–absence values for each species independently (i.e., permute at random within each column of table $\mathbf{L}$). This not only destroys the link between $\mathbf{L}$ and $\mathbf{R}$, but also destroys the relationship between $\mathbf{L}$ and $\mathbf{Q}$, as shown in the last paragraph of Appendix A.

2) Model 2: Permute site vectors (i.e., permute entire rows of table $\mathbf{L}$). This is strictly equivalent to permuting the rows of table $\mathbf{R}$. This destroys the link between $\mathbf{L}$ and $\mathbf{R}$ but keeps $\mathbf{L}$ linked to $\mathbf{Q}$.

3) Model 3: Permute presence–absence values for each site independently (i.e., permute within each row of table $\mathbf{L}$). This not only destroys the link between $\mathbf{L}$ and $\mathbf{Q}$, but also destroys the relationship between $\mathbf{L}$ and $\mathbf{R}$, as shown in the last paragraph of Appendix A.

4) Model 4: Permute species vectors (i.e., permute entire columns of table $\mathbf{L}$). This is strictly equivalent to permuting the rows of table $\mathbf{Q}$ or the columns of table $\mathbf{Q}^t$. This destroys the link between $\mathbf{L}$ and $\mathbf{Q}$ but keeps $\mathbf{L}$ linked to $\mathbf{R}$.

In the present paper, we will also consider another randomization procedure:

5) Model 5: Permute the species values and, after (or before), permute the sites values (i.e., permute entire columns and, after (or before) that, entire rows of table $\mathbf{L}$), destroying the links between $\mathbf{L}$ and $\mathbf{Q}$ and between $\mathbf{L}$ and $\mathbf{R}$. This is strictly equivalent to permuting the rows of both tables $\mathbf{R}$ and $\mathbf{Q}$, as proposed by Dolédec et al. (1996).

These five models are easily extended to abundance data. For models 1 and 3, one could choose to permute either the correspondences (cell values) of the species, or the occurrences (individuals). The first alternative has been preferred because it keeps constant the number of sites occupied by a species (models 1 and 2) or the number of species per site (models 3 and 4). Permuting individuals in the case of abundance data would lead to tests with highly inflated rates of Type I error. Also, when permuting individuals with model 1, some sites would run the risk of becoming void, and some types of environment could cease to exist, while model 3 could make some species and their associated trait values extinct. These permutations (i.e., realizations under these models that remove species or sites) must not be considered for the computation of P values as they do not take into account the complete information about species and sites.

These methods represent five ways of testing different null hypotheses concerning the absence of a link between tables $\mathbf{Q}$ and $\mathbf{R}$ mediated by table $\mathbf{L}$ (Fig. 2). The link between three tables ($\mathbf{R}$-$\mathbf{L}$-$\mathbf{Q}$) can be decomposed by studying the links between two table pairs ($\mathbf{L}$-$\mathbf{R}$, $\mathbf{L}$-$\mathbf{Q}$). If tables $\mathbf{R}$ and $\mathbf{Q}$ are linked through $\mathbf{L}$ ($\mathbf{R} \leftrightarrow \mathbf{Q}$), both the conditions $\mathbf{L} \leftrightarrow \mathbf{R}$ and $\mathbf{L} \leftrightarrow \mathbf{Q}$ must be satisfied. On the contrary, the absence of a link between $\mathbf{R}$ and $\mathbf{Q}$ ($\mathbf{R} \nleftrightarrow \mathbf{Q}$) corresponds to $\mathbf{L} \nleftrightarrow \mathbf{R}$ and/or $\mathbf{L} \nleftrightarrow \mathbf{Q}$. In hypothesis

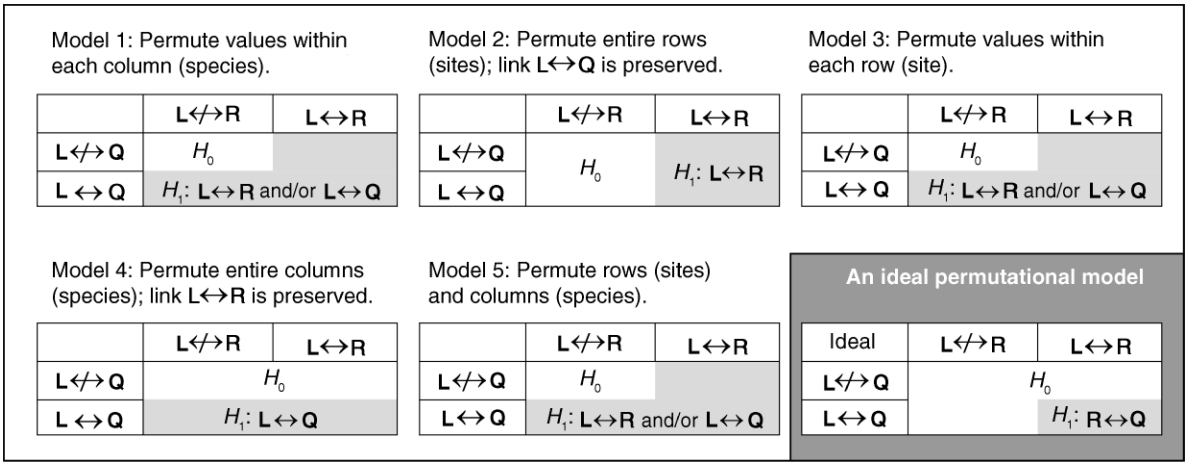

FIG. 2. Synoptic overview of the permutation models. For each model, cases corresponding to the null and the alternative hypotheses are given. See Appendix A for details.

testing, the null hypothesis is tested and either rejected in favor of an alternative hypothesis or not rejected, in which case the alternative hypothesis is not sustained. For the fourth-corner problem, an ideal permutational model would test the null hypothesis H_0: $\mathbf{R} \nleftrightarrow \mathbf{Q}$ against the alternative hypothesis H_1: $\mathbf{L} \leftrightarrow \mathbf{R}$ and $\mathbf{L} \leftrightarrow \mathbf{Q}$. The five models proposed in the literature are compared in Fig. 2 in terms of their null and alternative hypotheses; the figure also explains what action each permutation method has on the data, and what is preserved (i.e., not tested). The methods are spelled out in more detail in Appendix A. It appears that none of the five proposed models really corresponds to the ideal permutational model.

Simulation study

We simulated data to evaluate the five testing procedures for Type I error and power, according to different ecological scenarios. We considered both presence–absence and abundance data, and only the case of a single quantitative trait and a single quantitative environmental variable. The procedure to generate the data was the following:

1) Generate n values of an environmental variable x as a random sample from the uniform distribution of real numbers between 0 and 100 (table **R**).

2) Generate a vector $\boldsymbol{\mu}$ containing uniformly distributed random values between -1 and 101. The value μ_j represents the position of the optimum for species j. The optima (μ_j) of the p species form the species trait table (**Q**). Species optima are used as surrogates for species traits as it is assumed that species traits would be linked to species optima positions along the gradient if there is a link between the characteristics of species and their ecological preferences.

3) Generate a vector $\mathbf{h}$ containing values drawn at random from a uniform distribution from 0.5 to 1. Value h_j is the height of species j at its optimum.

4) Generate a vector $\boldsymbol{\sigma}$ containing normally distributed random values with mean μ_{tol} (defined in the last paragraph of the present section) and standard deviation 10. Values σ_j represent species tolerances, or niche breadths.

5) Generate a unimodal response curve for the jth species by

$$y_{ij} = h_j \exp\left[\frac{-(x_i - \mu_j)}{2\sigma_j^2}\right]$$

where x_i is the value of the environmental variable from table **R** at site i. For abundance data, table **L** was filled with y_{ij} values, whereas for presence–absence data, values were generated at random from a binomial distribution with probability y_{ij}.

In order to evaluate the five permutations models, six scenarios were considered.

1) Scenario 1: table **L** was structured and linked to tables **R** and **Q** as described in the previous paragraph.

2) Scenario 1N: table **L** was structured and linked to tables **R** and **Q**. The data were generated as in scenario 1. Normal random noise was added to tables **R** and **Q** (mean 5 and standard deviation 1) and to table **L** (mean 0 and standard deviation 2). This approach sometimes produced negative values in table **L**; they were replaced by 0.

3) Scenario 2: table **L** was structured and linked to table **R** only. This was obtained by simulating data as in scenario 1 and permuting at random the values in **Q**, creating realizations of H_0 for permutation model 4.

4) Scenario 3: table **L** was structured and linked to table **Q** only. This was obtained by simulating data as in scenario 1 and permuting at random the values in **R**, creating realizations of H_0 for permutation model 2.

5) Scenario 4: table **L** was structured but made unrelated to tables **R** and **Q**. This was obtained by simulating data as in scenario 1 and permuting at

TABLE 1. Results of the simulation study. Rejection rates of H_0 at the 5% significance level. The mean value of the link (r^2) is also given, except for scenario 5.

			Scenario 1 ($\mathbf{L} \leftrightarrow \mathbf{R}$, $\mathbf{L} \leftrightarrow \mathbf{Q}$)						Scenario 1N ($\mathbf{L} \leftrightarrow \mathbf{R}$, $\mathbf{L} \leftrightarrow \mathbf{Q}$)					
				Rejection rate for permutation model						Rejection rate for permutation model:				
μ_{tol}	No. sites	No. species	r^2	1	2	3	4	5	r^2	1	2	3	4	5
10	30	30	0.65855	1.000	1.000	1.000	1.000	1.000	0.00894	0.540	0.527	0.541	0.524	0.543
		50	0.66625	1.000	1.000	1.000	1.000	1.000	0.00815	0.722	0.702	0.724	0.713	0.722
		100	0.66797	1.000	1.000	1.000	1.000	1.000	0.00773	0.936	0.918	0.939	0.925	0.940
	50	30	0.66671	1.000	1.000	1.000	1.000	1.000	0.00824	0.705	0.703	0.711	0.675	0.713
		50	0.67258	1.000	1.000	1.000	1.000	1.000	0.00763	0.896	0.879	0.893	0.871	0.890
		100	0.67221	1.000	1.000	1.000	1.000	1.000	0.00716	0.992	0.991	0.993	0.990	0.991
	100	30	0.67367	1.000	1.000	1.000	1.000	1.000	0.00765	0.949	0.935	0.942	0.907	0.942
		50	0.67467	1.000	1.000	1.000	1.000	1.000	0.00735	0.991	0.990	0.991	0.986	0.991
		100	0.67655	1.000	1.000	1.000	1.000	1.000	0.00737	1.000	1.000	1.000	1.000	1.000
30	30	30	0.28416	1.000	1.000	1.000	1.000	1.000	0.01312	0.786	0.769	0.780	0.765	0.775
		50	0.28478	1.000	1.000	1.000	1.000	1.000	0.01305	0.944	0.931	0.943	0.937	0.947
		100	0.28414	1.000	1.000	1.000	1.000	1.000	0.01258	0.999	0.999	0.999	0.999	0.999
	50	30	0.28270	1.000	1.000	1.000	1.000	1.000	0.01302	0.936	0.929	0.935	0.916	0.934
		50	0.28708	1.000	1.000	1.000	1.000	1.000	0.01231	0.995	0.995	0.995	0.994	0.996
		100	0.28657	1.000	1.000	1.000	1.000	1.000	0.01231	1.000	1.000	1.000	1.000	1.000
	100	30	0.28766	1.000	1.000	1.000	1.000	1.000	0.01266	0.999	0.999	0.999	0.995	0.999
		50	0.28925	1.000	1.000	1.000	1.000	1.000	0.01248	1.000	1.000	1.000	1.000	1.000
		100	0.28925	1.000	1.000	1.000	1.000	1.000	0.01231	1.000	1.000	1.000	1.000	1.000
60	30	30	0.04532	1.000	1.000	1.000	1.000	1.000	0.00567	0.479	0.457	0.471	0.464	0.481
		50	0.04533	1.000	1.000	1.000	1.000	1.000	0.00514	0.680	0.652	0.670	0.662	0.683
		100	0.04596	1.000	1.000	1.000	1.000	1.000	0.00497	0.907	0.899	0.915	0.912	0.908
	50	30	0.04622	1.000	1.000	1.000	1.000	1.000	0.00518	0.668	0.675	0.670	0.662	0.679
		50	0.04654	1.000	1.000	1.000	1.000	1.000	0.00494	0.844	0.820	0.846	0.831	0.846
		100	0.04656	1.000	1.000	1.000	1.000	1.000	0.00455	0.983	0.986	0.980	0.982	0.983
	100	30	0.04592	1.000	1.000	1.000	1.000	1.000	0.00477	0.884	0.880	0.885	0.871	0.886
		50	0.04683	1.000	1.000	1.000	1.000	1.000	0.00458	0.986	0.985	0.986	0.986	0.986
		100	0.04714	1.000	1.000	1.000	1.000	1.000	0.00457	0.998	0.998	0.998	0.998	0.998

Notes: H_0 denotes the null hypothesis ($\mathbf{R} \nleftrightarrow \mathbf{Q}$) (see *Testing ecological hypotheses*); μ_{tol} is the average species tolerance. Permutation models are in defined in the text (see *Permutation models*). Scenarios are described in the *Simulation study* section.

random the values in **R** and **Q**, creating realizations of H_0 for all permutation models.

6) Scenario 5: table **L** was made to be unrelated to either **R**, or **Q**, or both. Data were simulated as in scenario 1 and the values in **L** were permuted at random according to the permutation model that we wished to evaluate. For model 1, individual values were permuted within each column of table **L**. For model 2, entire rows of **L** were permuted, leaving **L** linked to **Q** as in scenario 3. For model 3, individual values were permuted within each row of **L**. For model 4, entire columns of **L** were permuted, leaving **L** linked to **R** as in scenario 2. For model 5, the columns and then the rows of **L** were permuted at random, making **L** unrelated to tables **R** and **Q** as in scenario 4.

The first two scenarios (1 and 1N) correspond to a power study for all permutation models, while scenarios 2–5 allowed us to study the rates of Type I error. In scenarios 1–4, the covariance structure among the species in a simulated data set was the same for all permutation models. In scenario 5, the covariance structure among the species varied with the type of permutation of the **L** data. Scenario 2 is a study of the rate of Type I error for permutation model 4 and a power study for the other models. Likewise, scenario 3 is a study of the rate of Type I error for permutation model 2 and a power study for the other models. Scenarios 4 and 5 allowed us to study the rates of Type I error of all permutation models. A synthetic description of the simulation study is given in Appendix B.

For each scenario, three sample sizes were considered ($n = \{30, 50, 100\}$) as well as three sizes of the species pools ($p = \{30, 50, 100\}$). We also considered three values for the average species tolerances ($\mu_{tol} = \{10, 30, 60\}$). For each combination of parameters and scenarios, 1000 data sets were generated. For each data set, the five testing procedures were conducted with 999 random permutations.

Simulation results

For each data set, we tested the statistic $S_{\mathbf{RLQ}}$ (i.e., r^2 in this simulation study, which involved a single quantitative trait and a single quantitative environmental variable). For each combination of parameters and scenarios, we are reporting the rate of rejection of the null hypothesis at significance level $\alpha = 0.05$ and the mean value of r^2 over the 1000 generated data sets. In scenario 5, the values of r^2 varied with the type of permutation and are not reported. The results for abundance data are summarized in Table 1; the results

TABLE 1. Extended.

Scenario 2 ($\mathbf{L} \leftrightarrow \mathbf{R}$, $\mathbf{L} \nleftrightarrow \mathbf{Q}$)						Scenario 3 ($\mathbf{L} \leftrightarrow \mathbf{Q}$, $\mathbf{L} \nleftrightarrow \mathbf{R}$)					
	Rejection rate for permutation model:						Rejection rate for permutation model:				
r^2	1	2	3	4	5	r^2	1	2	3	4	5
0.02877	0.603	0.709	0.546	0.058	0.590	0.02105	0.556	0.046	0.547	0.689	0.573
0.01561	0.599	0.728	0.532	0.038	0.578	0.02288	0.640	0.066	0.623	0.760	0.651
0.00788	0.611	0.713	0.552	0.039	0.596	0.02250	0.751	0.053	0.739	0.844	0.758
0.02668	0.687	0.793	0.641	0.052	0.680	0.01211	0.529	0.040	0.515	0.663	0.538
0.01693	0.709	0.807	0.654	0.059	0.693	0.01294	0.649	0.041	0.630	0.778	0.664
0.00810	0.684	0.780	0.636	0.052	0.668	0.01326	0.762	0.047	0.748	0.855	0.770
0.02585	0.787	0.860	0.750	0.040	0.780	0.00681	0.554	0.057	0.533	0.695	0.567
0.01689	0.795	0.855	0.756	0.058	0.783	0.00601	0.620	0.035	0.597	0.744	0.632
0.00811	0.770	0.848	0.738	0.049	0.765	0.00615	0.752	0.045	0.731	0.830	0.749
0.00906	0.681	0.880	0.608	0.040	0.671	0.00858	0.601	0.043	0.612	0.865	0.642
0.00610	0.707	0.889	0.613	0.061	0.684	0.00952	0.710	0.057	0.723	0.916	0.749
0.00301	0.695	0.897	0.625	0.064	0.684	0.00932	0.792	0.053	0.797	0.930	0.810
0.00927	0.770	0.909	0.697	0.044	0.756	0.00497	0.650	0.048	0.657	0.883	0.685
0.00553	0.734	0.894	0.683	0.058	0.730	0.00518	0.694	0.044	0.700	0.887	0.723
0.00306	0.782	0.906	0.724	0.068	0.773	0.00538	0.808	0.044	0.812	0.933	0.830
0.00971	0.832	0.935	0.783	0.048	0.820	0.00235	0.611	0.036	0.615	0.860	0.649
0.00579	0.842	0.941	0.795	0.054	0.835	0.00276	0.712	0.048	0.718	0.919	0.740
0.00264	0.813	0.937	0.759	0.050	0.810	0.00277	0.805	0.047	0.804	0.931	0.820
0.00156	0.688	0.963	0.552	0.043	0.676	0.00151	0.642	0.059	0.617	0.955	0.695
0.00099	0.710	0.959	0.560	0.055	0.696	0.00147	0.686	0.049	0.672	0.967	0.723
0.00047	0.706	0.962	0.569	0.041	0.697	0.00149	0.813	0.052	0.799	0.984	0.850
0.00151	0.755	0.969	0.627	0.044	0.739	0.00091	0.618	0.054	0.604	0.959	0.669
0.00095	0.761	0.969	0.646	0.049	0.749	0.00087	0.697	0.042	0.675	0.974	0.736
0.00045	0.769	0.971	0.629	0.044	0.756	0.00088	0.766	0.052	0.758	0.977	0.807
0.00158	0.820	0.977	0.743	0.048	0.812	0.00047	0.644	0.058	0.605	0.954	0.692
0.00093	0.834	0.984	0.753	0.041	0.825	0.00045	0.715	0.051	0.692	0.964	0.772
0.00047	0.841	0.975	0.752	0.050	0.833	0.00041	0.769	0.034	0.753	0.982	0.804

for presence–absence data, which were very similar, are not presented.

When the three tables were linked (scenario 1), all permutational models identified the link with very high power. The strength of these results is due to the procedure of generation of the data, which created very strong links between the three tables. When the average species tolerance (μ_{tol}) increased, the number of correspondences per species increased and therefore the intensity of the link (r^2) decreased. In order to detect differences, we added random noise to tables **L**, **R**, and **Q** (scenario 1N): power increased with sample size and the five permutation models were quite equivalent.

Results of scenario 4 show that when neither **R** nor **Q** was linked to **L**, all permutational models had a rejection rate around 0.05. Hence they all had correct rates of Type I error (rate of rejection of H_0 when H_0 was true). In scenario 5, the simulated data were permuted at random according to the requirements of each permutation model; again, all permutation methods had correct rates of Type I error.

Results provided by scenarios 2 and 3 highlight the importance of choosing an appropriate permutational model, in accordance with the presuppositions of the study. When **L** was only linked to **R** (scenario 2), only permutation model 4 had a rejection rate near 0.05. The rejection rate for the other models varied between 0.532 and 0.984 even if the link (measured by r^2) between the three tables was very small. These rates are very high due to the strong link created between **L** and **R** in our simulated data; it would probably be lower with real data. Symmetrically, only model 2 (permute entire rows of table **L**, which is equivalent to permuting the rows of **R**) had a rejection rate of 0.05 for scenario 3 in which **L** was only linked to **Q**. All these results are in agreement with the expectations shown in Fig. 2.

Discussion of simulation results

In the original fourth-corner analysis, permutation models 1–4 were used in situations where the attributes of the species and those of the sites were considered fixed (i.e., not random). In many cases, the species characteristics are obtained from the literature and from past observations in similar systems, not from direct observation of the organisms making up table **L**. Likewise, observation of the characteristics of the sites can be made once and for all during a pilot study, before surveying the organisms to produce table **L**. The random component of the model is the observed presences or abundances of the species at the survey sites. Researchers are interested in finding links between the (fixed) traits of the species and the (fixed) characteristics of the sampling sites, the links being mediated by the observed data making up table **L** (Fig. 1a).

TABLE 1. Extended.

				Scenario 4 ($\mathbf{L} \nleftrightarrow \mathbf{R}$, $\mathbf{L} \nleftrightarrow \mathbf{Q}$)					Scenario 5 (depending on permutation model tested)				
				Rejection rate for permutation model:					Rejection rate for permutation model:				
μ_{tol}	No. sites	No. species	r^2	1	2	3	4	5	1	2	3	4	5
10	30	30	0.00176	0.059	0.051	0.029	0.053	0.052	0.048	0.045	0.049	0.044	0.059
		50	0.00103	0.056	0.052	0.035	0.043	0.046	0.041	0.043	0.059	0.050	0.057
		100	0.00048	0.047	0.039	0.029	0.043	0.044	0.053	0.050	0.051	0.058	0.052
	50	30	0.00120	0.065	0.055	0.042	0.058	0.056	0.047	0.052	0.041	0.055	0.051
		50	0.00060	0.052	0.044	0.026	0.045	0.043	0.045	0.047	0.045	0.046	0.047
		100	0.00029	0.049	0.047	0.032	0.048	0.049	0.042	0.049	0.044	0.048	0.048
	100	30	0.00050	0.044	0.046	0.035	0.033	0.043	0.045	0.059	0.051	0.043	0.039
		50	0.00030	0.043	0.042	0.024	0.043	0.036	0.052	0.043	0.052	0.042	0.037
		100	0.00017	0.065	0.052	0.050	0.054	0.057	0.050	0.044	0.049	0.046	0.045
30	30	30	0.00035	0.061	0.046	0.035	0.035	0.046	0.031	0.045	0.048	0.051	0.047
		50	0.00023	0.050	0.056	0.042	0.060	0.049	0.037	0.048	0.033	0.050	0.066
		100	0.00011	0.058	0.039	0.040	0.048	0.045	0.048	0.051	0.041	0.052	0.051
	50	30	0.00022	0.051	0.053	0.039	0.046	0.049	0.045	0.045	0.053	0.046	0.048
		50	0.00014	0.068	0.051	0.041	0.053	0.056	0.059	0.052	0.048	0.054	0.052
		100	0.00007	0.061	0.051	0.044	0.061	0.053	0.049	0.056	0.045	0.049	0.045
	100	30	0.00012	0.061	0.053	0.042	0.059	0.060	0.042	0.055	0.057	0.046	0.065
		50	0.00007	0.063	0.050	0.036	0.043	0.055	0.040	0.054	0.050	0.047	0.050
		100	0.00003	0.059	0.059	0.041	0.052	0.054	0.061	0.048	0.040	0.048	0.052
60	30	30	0.00005	0.059	0.062	0.021	0.052	0.055	0.054	0.049	0.049	0.045	0.046
		50	0.00003	0.044	0.046	0.026	0.038	0.040	0.062	0.047	0.045	0.037	0.054
		100	0.00001	0.056	0.034	0.020	0.044	0.049	0.041	0.053	0.050	0.046	0.043
	50	30	0.00003	0.052	0.051	0.024	0.043	0.044	0.058	0.063	0.049	0.055	0.038
		50	0.00002	0.044	0.043	0.023	0.048	0.039	0.031	0.046	0.056	0.049	0.057
		100	0.00001	0.055	0.052	0.020	0.048	0.053	0.056	0.044	0.042	0.058	0.042
	100	30	0.00002	0.062	0.057	0.024	0.053	0.054	0.046	0.051	0.044	0.044	0.044
		50	0.00001	0.052	0.050	0.017	0.041	0.043	0.051	0.056	0.050	0.050	0.056
		100	0.00001	0.066	0.063	0.030	0.046	0.061	0.033	0.046	0.041	0.057	0.052

In these circumstances, permutation model 1 can be used to test H_0 ($\mathbf{R} \nleftrightarrow \mathbf{Q}$) by distributing the presences or abundances of individual species (whose characteristics are fixed) at random through the sites whose characteristics are also fixed. Permutation model 2 tests the same null hypothesis but at the level of species assemblages. Permutation model 3 is quite different from models 1 and 2 in that the assumed random process is a lottery among individual settlers for space. Permutation model 4, where applicable, follows the same general approach as model 3.

The statistic $S_{\mathbf{RLQ}}$ measures the link between the species traits ($\mathbf{Q}$) and the environmental conditions ($\mathbf{R}$), mediated through $\mathbf{L}$. The tests of significance have a different purpose and meaning: they allow users to test hypotheses about the mechanisms that led to the measured link value. The simulation results (Table 1) showed that when $\mathbf{R}$ and $\mathbf{Q}$ were not linked to $\mathbf{L}$ (scenario 4), all permutation models produced correct rates of Type I error. The interesting question is: What are the results of the tests when only one matrix, $\mathbf{R}$ or $\mathbf{Q}$, is linked to $\mathbf{L}$? Simulation scenarios 2 and 3 showed that, unless one uses the appropriate permutation model, one can easily obtain false positives and incorrectly conclude that there is a link between $\mathbf{R}$ and $\mathbf{Q}$ through $\mathbf{L}$. On the contrary, when used appropriately (scenario 5), all permutation models had correct rates of Type I error, rejecting H_0 in about 5% of the simulations.

This simulation study showed that ecological conclusions drawn from the results of testing procedures developed for the fourth-corner statistic (models 1–4) or for RLQ analysis (model 5) must be made with real understanding of the tested hypothesis (Fig. 2). It must be noticed that the case where tables $\mathbf{R}$, $\mathbf{L}$, and $\mathbf{Q}$ are linked (i.e., $\mathbf{L} \leftrightarrow \mathbf{R}$ and $\mathbf{L} \leftrightarrow \mathbf{Q}$) is only one of the situations that correspond to the alternative hypothesis for these five models. These results show that in some circumstances, only conclusions about two tables can be drawn, even though the test statistic is measuring a relationship involving the three tables.

When the data in tables $\mathbf{R}$ and $\mathbf{Q}$ are considered random ($\mathbf{L}$ may be fixed or random), permutation method 5 may logically seem to be the most appropriate to test the whole link between $\mathbf{R}$ and $\mathbf{Q}$ through $\mathbf{L}$. The simulation results suggest another possibility, which consists of testing H_0: $\mathbf{R} \nleftrightarrow \mathbf{Q}$ against H_1: $\mathbf{L} \leftrightarrow \mathbf{R}$ and $\mathbf{L} \leftrightarrow \mathbf{Q}$, by combining the results of tests carried out separately using model 2 (H_1: $\mathbf{L} \leftrightarrow \mathbf{R}$) and model 4 ($H_1$: $\mathbf{L} \leftrightarrow \mathbf{Q}$) (Fig. 3):

1) Use model 2 to test H_0 against H_1, which states that tables $\mathbf{L}$ and $\mathbf{R}$ at least are linked. This test is performed at significance level α_1.

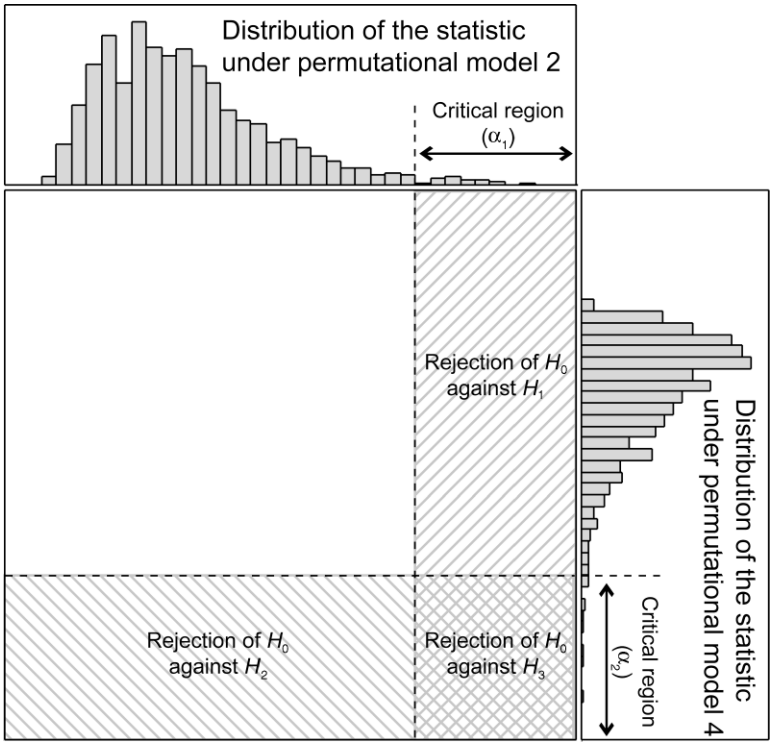

FIG. 3. An approach to combine results of two permutation models. Model 2 allows testing H_0 (no link between **R** and **Q**) against H_1 (tables **L** and **R** at least are linked). The distribution of the statistic under H_0 is obtained by permutation using model 2; the critical region corresponds to the proportion α_1 of the highest values. Model 4 allows testing H_0 against H_2 (tables **L** and **Q** at least are linked). The critical region is formed by the proportion α_2 of the highest values of the distribution of the statistic obtained by permutation using model 4. Combining the results of the two tests allows a test of H_0 against H_3 (tables **R**, **L**, and **Q** are linked) at significance level $\alpha_3 = \alpha_1\alpha_2$.

2) Use model 4 to test H_0 against H_2, which states that tables **L** and **Q** at least are linked. This test is performed at significance level α_2.

3) Test H_0 against H_3, which states that tables **R**, **L**, and **Q** are linked. This is achieved by combining the results of the two tests above: if H_0 has been rejected against H_1 at significance level α_1, and if H_0 has been rejected against H_2 at significance level α_2, then H_0 is rejected against H_3 at a significance level $\alpha_3 = \alpha_1\alpha_2$.

If we want to test H_0 against H_3 at significance level $\alpha_3 = 0.05$, we can choose for instance $\alpha_1 = \alpha_2 = \sqrt{0.05} = 0.2236$, or $\alpha_1 = 0.25$ and $\alpha_2 = 0.20$, if we are willing to use unequal significance levels for the two tests.

Results for the combined strategy with $\alpha_1 = \alpha_2 = \sqrt{0.05}$ are given in Table 2. This strategy seems to be slightly liberal (Type I error between 0.039 and 0.140 for scenario 4), especially when species tolerances μ_{tol} are very small, and induce loss of power in scenario 1N. However, as the simple procedures (models 2 and 4) are very powerful (scenario 1N), this loss of power is not too prejudicial. For scenarios 2 and 3, the Type I error varies between 0.198 and 0.258 because the simple tests (H_0 against H_1 and H_0 against H_2) were performed at significance levels $\alpha_1 = \alpha_2 = \sqrt{0.05} = 0.2236$.

The power of the combined approach was evaluated using data generated under scenario 1N. In a first series of simulations, we evaluated the effect of the number of sites (n) and the number of species (p) on power. One hundred cases were considered (the number of species and the number of sites varied between 10 and 100 with regular intervals of 10; the average species tolerance was equal to 30); 1000 data sets were generated for each case. The results (Fig. 4a) show that power increased equivalently with the sampling size and the size of the species pool. In this simulation study, power varied between 0.29 and 1.0. It was always >0.8 when the number of cells of table **L** (number of species multiplied by the number of sites) was larger than 900. Power varied between 0.95 and 1.0 for tables with more than 1200 cells.

In a second series of simulations, the average species tolerances (μ_{tol}) was made to vary between 10 and 80 (by regular intervals of 10) for three sampling sizes ($n = \{30, 50, 100\}$) as well as three sizes of the species pool ($p = \{30, 50, 100\}$). Results concerning the influence of the average species tolerance are presented in Fig. 4b. Since the size of the gradient is constant, this figure can be interpreted in terms of alpha and beta diversities. As the average species tolerance increased, overlaps between

TABLE 2. Results of the simulation study for the combined approach. The table reports rejection rates of H_0 (tested against the alternative hypothesis H_3) at the 5% significance level.

			Combined approach				
			Rejection rate for scenario:				
μ_{tol}	No. sites	No. species	1	1N	2	3	4
10	30	30	1.000	0.802	0.232	0.217	0.130
		50	1.000	0.915	0.220	0.240	0.123
		100	1.000	0.992	0.228	0.237	0.116
	50	30	1.000	0.916	0.210	0.215	0.140
		50	1.000	0.976	0.219	0.213	0.122
		100	1.000	1.000	0.221	0.236	0.125
	100	30	1.000	0.989	0.212	0.246	0.108
		50	1.000	0.999	0.230	0.212	0.116
		100	1.000	1.000	0.224	0.229	0.126
30	30	30	1.000	0.938	0.213	0.204	0.078
		50	1.000	0.996	0.238	0.237	0.080
		100	1.000	1.000	0.258	0.243	0.077
	50	30	1.000	0.988	0.218	0.192	0.085
		50	1.000	0.999	0.219	0.215	0.091
		100	1.000	1.000	0.250	0.232	0.094
	100	30	1.000	1.000	0.213	0.199	0.083
		50	1.000	1.000	0.219	0.238	0.101
		100	1.000	1.000	0.198	0.230	0.082
60	30	30	1.000	0.781	0.224	0.217	0.068
		50	1.000	0.897	0.234	0.213	0.039
		100	1.000	0.989	0.218	0.218	0.052
	50	30	1.000	0.891	0.224	0.235	0.061
		50	1.000	0.958	0.217	0.219	0.057
		100	1.000	1.000	0.204	0.209	0.056
	100	30	1.000	0.974	0.211	0.250	0.057
		50	1.000	1.000	0.219	0.237	0.047
		100	1.000	1.000	0.223	0.212	0.066

species niches were more frequent and thus beta diversity decreased while alpha diversity increased. The results suggest that power is linked in a nonlinear manner to beta diversity. When beta diversity was low (high average species tolerance), power was also low (minimum is 0.414 for 30 species and 30 sites). This makes sense because, in this case, traits are linked to the average position on the gradient, but when species are very tolerant they can occupy a wide range of environmental conditions. More surprisingly, in the case of high beta diversity (low values of species tolerance), power was reduced (power was 0.667 for 30 species and 30 sites at an average species tolerance of 10). In our study, the best results were obtained for an average species tolerance of 30; power then varied between 0.875 and 1.0.

This combined approach seems at first glance equivalent to analyzing the two-tables links ($\mathbf{L} \leftrightarrow \mathbf{R}$ and $\mathbf{L} \leftrightarrow \mathbf{Q}$) using statistics and permutation procedures associated to co-inertia analysis (Dray et al. 2003), canonical correspondence analysis (ter Braak 1986), or redundancy analysis (Rao 1964), and then combining the results as in our combined approach. However, it must be remembered that our combined approach is based on a statistic that measures the link between the three tables and, therefore, it tests the link between the three tables. Results obtained by two-tables procedures could lead to false positive results by considering two significant two-tables links associated to different ecological processes.

The simulation studies reported here showed that all testing procedures were powerful (scenarios 1 and 1N) and had correct rates of Type I error when the three tables **R**, **L**, and **Q** were not linked (scenario 4) or were linked in the way corresponding to each permutation model (scenario 5). When only tables **L** and **R** were linked (scenario 2), model 4 had a correct level of Type I error while the other models correctly detected the $\mathbf{L} \leftrightarrow \mathbf{R}$ link. When tables **L** and **Q** only were linked (scenario 3), model 2 had a correct level of Type I error while the

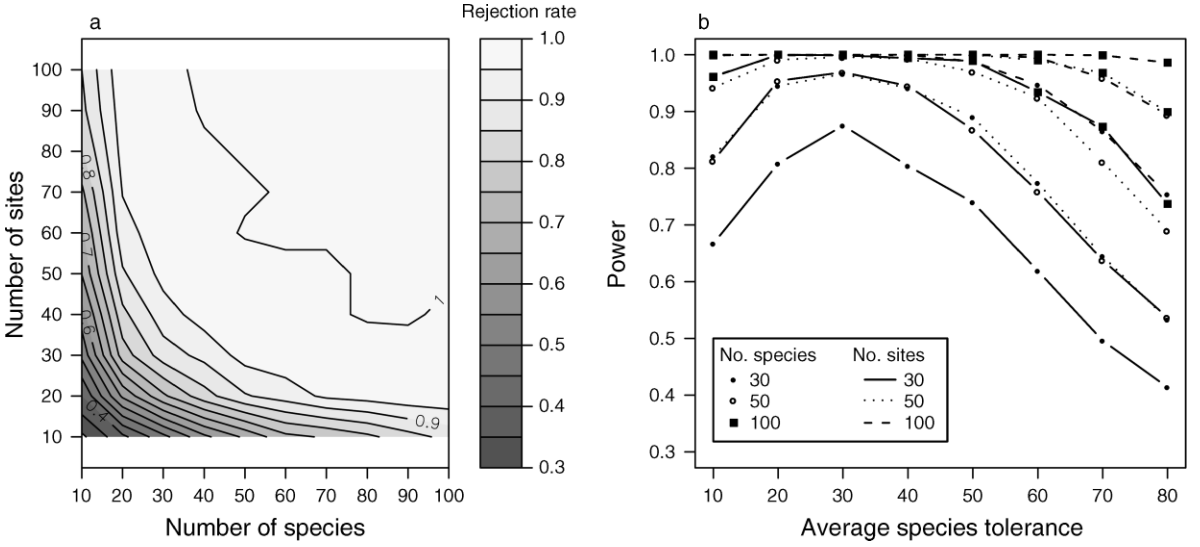

FIG. 4. Results of the power study for the combined approach. The influence on power of (a) the sample size and the number of species and (b) of the species tolerance. The isolines in panel (a) correspond to rejection rates of H_0 (tested against the alternative hypothesis H_3) at the 5% significance level after 1000 independent simulations.

other models correctly detected the $\mathbf{L} \leftrightarrow \mathbf{Q}$ link. Our proposition to combine the testing procedures based on models 2 and 4 is one way to solve these problems.

Conclusions

The fourth-corner analysis allows users to measure and test directly the link between the variations in species traits and the environmental structures through the link provided by community data, including situations where the data in **R** and **Q** are considered fixed or random. In this paper, we provide new tools to deal with abundance data and to test ecological hypotheses using several traits and environmental variables. The combined testing strategy seems to be the only way to test properly the whole link between species traits and environmental variables mediated by the abundances of species. Results of the power studies suggest that the procedure is more efficient when the sampling size and/or the number of species is fairly high. Hence sampling effort must be more important in ecological systems harboring few species in order to gain power. Power is also reduced when beta diversity is very low or very high. In such systems, increasing sample size is the best solution to increasing the power of the method, which is its ability to detect a relationship between species traits and environmental conditions when such a relationship exists.

In this paper, the methodological problem (simultaneous analysis of three tables) was directly related to the biological question of linking species traits and environmental variables. The methods developed for this particular biological issue could, of course, be used in other situations. For instance, Dray et al. (2002) adapted RLQ analysis to the problem of relating data sets from two different spatial samples. In that study, tables **R** and **Q** contained species composition for canopy and understory, respectively, while table **L** was a spatial neighborhood matrix. Using the principle of the fourth-corner statistic, Legendre et al. (2002) developed a method to test the significance of the global hypothesis of coevolution between parasites and their hosts. In that approach, tables **R** and **Q** contained principal coordinates describing the phylogenetic trees of the parasites and of the hosts, respectively, while a binary table **L** described the empirical associations between parasites and hosts. Tall et al. (2006) used the fourth-corner statistic and RLQ analysis in order to test if co-occurring grazers varying in size and taxonomy had different diet patterns and if these patterns were explained by ingestion of diatoms, which differed in size or spatial positions in the epiphyton mats. These examples show that the development of new methods for direct functional analysis can be very helpful in solving other ecological problems.

Introduction of species traits in ecological studies offers exciting perspectives for biologists. It allows tests of new and more general hypotheses to be performed in order to better understand the functioning of ecosystems. It represents a great challenge for numerical ecologists that must develop new methods in order to allow ecologists to test biological hypotheses with appropriate tools. Taking into account autocorrelation between sites (e.g., spatial structures) and between species (phylogenetic constraints) appears as an objective of prime interest for methodologists as well as ecologists. In order to provide efficient tools for conservation policy, development of a predictive functional ecology requiring predictive methods represents another objective. For this purpose, approaches developed in ecology, including a modeling step, (McIntyre and Lavorel 2001, Kleyer 2002, Nygaard and Ejrnaes 2004) could be used as a starting point in order to provide efficient predictive methods.

Software

R functions to compute the fourth-corner statistic for presence–absence or abundance data are available in the ade4 package (Dray and Dufour 2007). In addition, RLQ analysis is available in the ade4 package.

Acknowledgments

We are grateful to Anik Brind'Amour, Rosalie Léonard, and Marie-Hélène Ouellette, as well as Susan Wiser and one anonymous referee for useful comments on our manuscript. This research was supported by NSERC grant number 7738 to P. Legendre.

Literature Cited

Barbaro, L., E. Corcket, T. Dutoit, and J.-P. Peltier. 2000. Réponses fonctionnelles des communautés de pelouses calcicoles aux facteurs agro-écologiques dans les Préalpes françaises. Canadian Journal of Botany 78:1010–1020.

Blondel, J., F. Vuilleumier, L. F. Marcus, and E. Terouanne. 1984. Is there ecomorphological convergence among Mediterranean bird communities of Chile, California, and France? Pages 141–213 *in* M. K. Hecht, B. Wallace, and R. J. MacIntyre, editors. Evolutionary Biology. Plenum Press, New York, New York, USA.

Campbell, B. D., D. M. Stafford Smith, and A. J. Ash. 1999. A rule-based model for the functional analysis of vegetation change in Australasian grasslands. Journal of Vegetation Science 10:723–730.

Charest, R., L. Brouillet, A. Bouchard, and S. Hay. 2000. The vascular flora of Terra Nova National Park, Newfoundland, Canada: a biodiversity analysis from a biogeographical and life form perspective. Canadian Journal of Botany 78:629–645.

Diaz, S., M. Cabido, and F. Casanoves. 1998. Plant functional traits and environmental filters at a regional scale. Journal of Vegetation Science 9:113–121.

Dolédec, S., D. Chessel, C. J. F. ter Braak, and S. Champely. 1996. Matching species traits to environmental variables: a new three-table ordination method. Environmental and Ecological Statistics 3:143–166.

Dolédec, S., B. Statzner, and M. Bournaud. 1999. Species traits for future biomonitoring across ecoregions: patterns along a human-impacted river. Freshwater Biology 42:737–758.

Dray, S., D. Chessel, and J. Thioulouse. 2003. Co-inertia analysis and the linking of ecological data tables. Ecology 84: 3078–3089.

Dray, S., and A. B. Dufour. 2007. The ade4 package: implementing the duality diagram for ecologists. Journal of Statistical Software 22(4):1–20.

Dray, S., N. Pettorelli, and D. Chessel. 2002. Matching data sets from two different spatial samples. Journal of Vegetation Science 13:867–874.

Gimaret-Carpentier, C., S. Dray, and J.-P. Pascal. 2003. Broad-scale biodiversity pattern of the endemic tree flora of the Western Ghats (India) using canonical correlation analysis of herbarium records. Ecography 26:429–444.

Greenacre, M. J. 1984. Theory and applications of correspondence analysis. Academic Press, London, UK.

Hausner, V. H., N. G. Yoccoz, and R. A. Ims. 2003. Selecting indicator traits for monitoring land use impacts: birds in northern coastal birch forests. Ecological Applications 13: 999–1012.

Hooper, E. R., P. Legendre, and R. Condit. 2004. Factors affecting community composition of forest regeneration in deforested, abandoned land in Panama. Ecology 85:3313–3326.

Keddy, P. A. 1992*a*. Assembly and response rules: two goals for predictive community ecology. Journal of Vegetation Science 3:157–164.

Keddy, P. A. 1992*b*. A pragmatic approach to functional ecology. Functional Ecology 6:621–626.

Kleyer, M. 2002. Validation of plant functional types across two contrasting landscapes. Journal of Vegetation Science 13: 167–178.

Lamouroux, N., N. L. Poff, and P. L. Angermeier. 2002. Intercontinental convergence of stream fish community traits along geomorphic and hydraulic gradients. Ecology 83:1792–1807.

Legendre, P., Y. Desdevises, and E. Bazin. 2002. A statistical test for host–parasite coevolution. Systematic Biology 5:217–234.

Legendre, P., R. Galzin, and M. L. Harmelin-Vivien. 1997. Relating behavior to habitat: solutions to the fourth-corner problem. Ecology 78:547–562.

Mabry, C., D. Ackerly, and F. Gerhardt. 2000. Landscape and species-level distribution of morphological and life history traits in a temperate woodland flora. Journal of Vegetation Science 11:213–224.

McIntyre, S., S. Diaz, S. Lavorel, and W. Cramer. 1999. Plant functional types and disturbance dynamics: introduction. Journal of Vegetation Science 10:604–608.

McIntyre, S., and S. Lavorel. 2001. Livestock grazing in subtropical pastures: steps in the analysis of attribute response and plant functional types. Journal of Ecology 89: 209–226.

Nygaard, B., and R. Ejrnaes. 2004. A new approach to functional interpretation of vegetation data. Journal of Vegetation Science 15:49–56.

Pausas, J. G. 1999. Response of plant functional types to changes in the fire regime in Mediterranean ecosystems: a simulation approach. Journal of Vegetation Science 10:717–722.

Pélissier, R., P. Couteron, S. Dray, and D. Sabatier. 2003. Consistency between ordination techniques and diversity measurements: two strategies for species occurrence data. Ecology 84:242–251.

Pélissier, R., S. Dray, and D. Sabatier. 2002. Within-plot relationships between tree species occurrences and hydrological soil constraints: an example in French Guiana investigated through canonical correlation analysis. Plant Ecology 162:143–156.

Poff, N., J. Olden, N. Vieira, D. Finn, M. Simmons, and B. Kondratieff. 2006. Functional trait niches of North American lotic insects: traits-based ecological applications in light of phylogenetic relationships. Journal of the North American Benthological Society 25:730–755.

Rao, C. R. 1964. The use and interpretation of principal component analysis in applied research. Sankhya A 26:329–359.

Ribera, I., S. Dolédec, I. S. Downie, and G. N. Foster. 2001. Effect of land disturbance and stress on species traits of ground beetle assemblages. Ecology 82:1112–1129.

Southwood, T. R. E. 1977. Habitat, the templet for ecological strategies? Journal of Animal Ecology 46:337–365.

Southwood, T. R. E. 1988. Tactics, strategies and templets. Oikos 52:3–18.

Tall, L., A. Cattaneo, L. Cloutier, S. Dray, and P. Legendre. 2006. Resource partitioning in a grazer guild feeding on a multilayer diatom mat. Journal of the North American Benthological Society 25:800–810.

ter Braak, C. J. F. 1986. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. Ecology 67:1167–1179.

Thioulouse, J., and D. Chessel. 1992. A method for reciprocal scaling of species tolerance and sample diversity. Ecology 73: 670–680.

Thuiller, W., S. Lavorel, G. Midgley, S. Lavergne, and T. Rebelo. 2004. Relating plant traits and species distributions along bioclimatic gradients for 88 *Leucadendron* taxa. Ecology 85:1688–1699.

Wiens, J. A. 1991. Ecological similarity of shrub-desert avifaunas of Australia and North America. Ecology 72: 479–495.

APPENDIX A

Full description of the five permutation methods (*Ecological Archives* E089-195-A1).

APPENDIX B

Synthetic description of the simulation study (*Ecological Archives* E089-195-A2).

APPENDIX C

Mathematical proofs (*Ecological Archives* E089-195-A3).

Electronic Appendices

APPENDIX A

FULL DESCRIPTION OF THE FIVE PERMUTATION METHODS

This Appendix presents a detailed description of the permutation methods compared in Table 1 of the main paper.

Model #1.—Permute presence-absence values for each species independently (i.e., permute within each column of table **L**). The null hypothesis (H_0) states that individuals of a species are randomly distributed with respect to site characteristics. The corresponding alternative hypothesis (H_1) states that individuals of a species are distributed according to their preferences for site conditions, or "*individual species find optimal living conditions at the stations where they are actually found*" (Legendre et al. 1997, p. 551). The link between the species and their traits is also modified by this permutation procedure and so is tested. Under this permutation model, the number of sites occupied by a given species (i.e., niche breadth) is kept constant.

Model #2.—Permute site vectors (i.e., permute entire rows of table **L**). This is strictly equivalent to permuting the rows of table **R**. H_0 states that species assemblages are randomly attributed to sites, irrespective of the site characteristics. The corresponding alternative hypothesis (H_1) states that "*species assemblages are dependent upon the physical characteristics of the locations where they are actually found*" (Legendre et al. 1997, p. 551). The link between the species and their traits is not questioned nor tested. Under this permutation model, the number of sites occupied by a given species (i.e., niche breadth) is kept constant. This is the only permutation method that preserves the covariances among the species and the link between tables **L** and **Q**.

Model #3.— Permute presence-absence values for each site independently (i.e., permute within each row of table **L**). H_0 states that the distribution of the presences of various species at a site is the result of a random allocation process (lottery for space among the species); it is not due to the adaptation of that species' traits to the sites. The alternative hypothesis (H_1) states that due to their traits, "*species have some competitive advantages over chance settlers in given habitats*" (Legendre et al. 1997, p. 552). The link between the species composition of sites and the environmental conditions is also modified by this permutation procedure and so is tested. Under this permutation model, the number of species present in a given site (i.e., species richness) is kept constant.

Model #4.— Permute species vectors (i.e., permute entire columns of table **L**). This is strictly equivalent to permuting the rows of table **Q** (or the columns of table $\mathbf{Q}^t$). H_0 states that species are distributed according to their preferences for site conditions, but irrespective of their traits or other characteristics included in table **Q**. The alternative hypothesis (H_1) states that the distributions of the species among the sites, which are related to their preferences for site conditions, depend on the adaptations (traits) of the species. The preferences of the species for site conditions is taken for granted; it is not questioned nor tested. Under this permutation model, the number of species present in a given site (i.e., species richness) is kept constant. This is the only permutation method that preserves the link between **L** and **R**.

Model #5.— Permute species vectors, then permute sites vectors, or in the reverse order (i.e., permute entire columns, then (or before) permute entire rows of table **L**). This is strictly equivalent to permuting simultaneously the rows of tables **R** and **Q**, as proposed by Dolédec et al. (1996). The null model (H_0) is that species distributions among the sites are not related to the site conditions nor to the traits of the species. The alternative hypothesis (H_1) states that the species

distributions across the sites are related to species traits *and/or* that species assemblages are dependent upon the environmental conditions.

The statements that the **L↔Q** link is broken in permutation method #1, and the **L↔R** link is broken in permutation method #3, need support. We can envision the **L-R** link as a sum of squared covariances between the species in **L** and the environmental variables in **R** (i.e., sum of eigenvalues of the co-inertia analysis between **L** and **R**), and the **L-Q** link as a sum of squared covariances between the sites in **L** and the traits in **Q**. In permutation models #1 and #2, the species vectors in **L** are modified, but not in the same way. Model #2 does not change the **L-Q** relation, whereas by permuting the values within each column of **L**, and thus modifying the covariances between the sites in **L** and the traits in **Q**, model #1 changes that relation. The same phenomena happen, *mutatis mutandis*, for permutation method #3. As an additional illustration, consider canonical redundancy analysis (RDA), which is familiar to community ecologists: permutation method #3, which permutes values within each row of **L** independently of the other rows, would change the regressions of individual species (columns of **L**) on the site variables in **R**, and thus the canonical relationship.

LITERATURE CITED

Dolédec, S., D. Chessel, C. J. F. ter Braak, and S. Champely. 1996. Matching species traits to environmental variables: a new three-table ordination method. Environmental and Ecological Statistics **3**:143-166.

Legendre, P., R. Galzin, and M. L. Harmelin-Vivien. 1997. Relating behavior to habitat: solutions to the fourth-corner problem. Ecology **78**:547-562.

APPENDIX B

SYNTHETIC DESCRIPTION OF THE SIMULATION STUDY

This table presents a short summary of the different scenarios used in the simulation study.

Scenario	Structure simulated	Modification of the data (compared to scenario 1)	Power study	Type I error study
1	**L**↔**R**, **L**↔**Q**	—	For all models	—
1N	**L**↔**R**, **L**↔**Q**	Add noise $\mathcal{N}(5,1)$ to **R** and **Q** and $\mathcal{N}(0,2)$ to **L**	For all models	—
2	**L**↔**R**, **L**↮**Q**	Permute **Q**	For all models except #4	For model #4
3	**L**↮**R**, **L**↔**Q**	Permute **R**	For all models except #2	For model #2
4	**L**↮**R**, **L**↮**Q**	Permute **R** and **Q**	—	For all models
5	Vary with model used	Vary with model used	—	For all models

APPENDIX C

MATHEMATICAL PROOFS

Table **L** ($n \times p$) contains the abundances of p species at n sites. Let $\mathbf{P} = \left[P_{ij} \right]$ be the table of relative frequencies with $P_{ij} = L_{ij} / L_{++}$ where L_{ij} is the abundance of the j-th species in the i-th site and $L_{++} = \sum_{i=1}^{n} \sum_{j=1}^{p} L_{ij}$ is the grand total computed from table **L**. The row and column weights derived from table **L** are respectively denoted by $P_{i+} = \frac{L_{i+}}{L_{++}} = \sum_{j=1}^{p} \frac{L_{ij}}{L_{++}}$ and $P_{+j} = \frac{L_{+j}}{L_{++}} = \sum_{i=1}^{n} \frac{L_{ij}}{L_{++}}$. Let us consider the diagonal matrices of site and species weights defined respectively by $\mathbf{D}_n = Diag(P_{1+},...,P_{i+},...,P_{n+})$ and $\mathbf{D}_p = Diag(P_{+1},...,P_{+j},...,P_{+p})$. Table **R** ($n \times m$) describes the environment while species traits are contained in table **Q** ($p \times s$). The variables in **R** and **Q** can be quantitative or qualitative.

Tables of correspondences

We consider the two tables of correspondences $\mathbf{X}_c$ ($c \times n$) and $\mathbf{Y}_c$ ($c \times p$) (see the main paper for details). The two inflated tables $\mathbf{R}_c$ ($c \times m$) and $\mathbf{Q}_c$ ($c \times s$) are also constructed by duplicating the values of tables **R** and **Q** respectively, according to the distribution of correspondences in tables $\mathbf{X}_c$ and $\mathbf{Y}_c$. If the k-th correspondence belongs to the i-th site and the j-th species, then the k-th row of $\mathbf{R}_c$ is equal to the i-th row of **R** and the k-th row of $\mathbf{Q}_c$ is equal to the j-th row of **Q**. A diagonal matrix of weights $\mathbf{D}_c$ is constructed where $\mathbf{D}_c(k,k) = L_{ij} / L_{++}$ if the k-th correspondence belongs to the i-th site and the j-th species.

We note the following relationships (Dray 2003 p. 25):

$$\mathbf{Q}_c = \mathbf{Y}_c\mathbf{Q},\ \mathbf{Q} = \mathbf{D}_p{}^{-1}\mathbf{Y}_c{}^t\mathbf{D}_c\mathbf{Q}_c;\ \mathbf{R}_c = \mathbf{X}_c\mathbf{R},\ \mathbf{R} = \mathbf{D}_n{}^{-1}\mathbf{X}_c{}^t\mathbf{D}_c\mathbf{R}_c \quad \text{(C.1)}$$

and

$$\mathbf{P} = \mathbf{X}_c{}^t\mathbf{D}_c\mathbf{Y}_c,\ \mathbf{D}_n = \mathbf{X}_c{}^t\mathbf{D}_c\mathbf{X}_c,\ \mathbf{D}_p = \mathbf{Y}_c{}^t\mathbf{D}_c\mathbf{Y}_c \quad \text{(C.2)}$$

Tables of occurrences

The second way to inflate the original data considers the occurrences, i.e. the o individuals of table **L**. By definition, we have $o = L_{++}$. Two tables $\mathbf{X}_o$ ($o \times n$) and $\mathbf{Y}_o$ ($o \times p$) are derived from table **L**. Inflated tables $\mathbf{R}_o$ ($o \times m$) and $\mathbf{Q}_o$ ($o \times s$) are also constructed by duplicating the values of tables **R** and **Q** respectively. For this inflating approach, weights associated to occurrences are uniform ($\forall k,\ \mathbf{D}_o(k,k) = 1/o$).

We note the following relationships:

$$\mathbf{Q}_o = \mathbf{Y}_o\mathbf{Q},\ \mathbf{Q} = \mathbf{D}_p{}^{-1}\mathbf{Y}_o{}^t\mathbf{D}_o\mathbf{Q}_o;\ \mathbf{R} = \mathbf{D}_n{}^{-1}\mathbf{X}_o{}^t\mathbf{D}_o\mathbf{R}_o,\ \mathbf{R}_o = \mathbf{X}_o\mathbf{R} \quad \text{(C.3)}$$

and

$$\mathbf{P} = \mathbf{X}_o{}^t\mathbf{D}_o\mathbf{Y}_o,\ \mathbf{L} = \mathbf{X}_o{}^t\mathbf{Y}_o,\ \mathbf{D}_n = \mathbf{X}_o{}^t\mathbf{D}_o\mathbf{X}_o,\ \mathbf{D}_p = \mathbf{Y}_o{}^t\mathbf{D}_o\mathbf{Y}_o \quad \text{(C.4)}$$

Linking two quantitative variables

Consider that **R** and **Q** each contain a single quantitative variable. If the two variables in $\mathbf{R}_o$ and $\mathbf{Q}_o$ are standardized to means $\mathbf{Q}_o{}^t\mathbf{D}_o\mathbf{1}_o = \mathbf{R}_o{}^t\mathbf{D}_o\mathbf{1}_o = 0$ where $\mathbf{1}_o$ is a vector of 1 with o rows and variances $\mathbf{Q}_o{}^t\mathbf{D}_o\mathbf{Q}_o = \mathbf{R}_o{}^t\mathbf{D}_o\mathbf{R}_o = 1$, then $\mathbf{Q}_o{}^t\mathbf{D}_o\mathbf{R}_o$ is a Pearson correlation coefficient r:

$$r = \mathbf{Q}_o{}^t\mathbf{D}_o\mathbf{R}_o \quad \text{(C.5)}$$

Using (C.3) and (C.4), we demonstrate that (C.5) is equivalent to compute a cross-correlation coefficient using the original tables:

$$r = \mathbf{Q}_o{}^t \mathbf{D}_o \mathbf{R}_o = \mathbf{Q}^t \mathbf{Y}_o{}^t \mathbf{D}_o \mathbf{X}_o \mathbf{R} = \mathbf{Q}^t \mathbf{P}^t \mathbf{R} \qquad \text{(C.6)}$$

Using (C.1) and (C.2), we demonstrate that (C.6) is equivalent to compute a weighted correlation coefficient using the correspondence tables:

$$\begin{aligned} r &= \mathbf{Q}_c{}^t \mathbf{D}_c \mathbf{Y}_c \mathbf{D}_p{}^{-1} \mathbf{Y}_c{}^t \mathbf{D}_c \mathbf{X}_c \mathbf{D}_n{}^{-1} \mathbf{X}_c{}^t \mathbf{D}_c \mathbf{R}_c \\ &= \mathbf{Q}_c{}^t \mathbf{D}_c \mathbf{Y}_c (\mathbf{Y}_c{}^t \mathbf{D}_c \mathbf{Y}_c)^{-1} \mathbf{Y}_c{}^t \mathbf{D}_c \mathbf{X}_c (\mathbf{X}_c{}^t \mathbf{D}_c \mathbf{X}_c)^{-1} \mathbf{X}_c{}^t \mathbf{D}_c \mathbf{R}_c \\ &= \mathbf{Q}_c{}^t \mathbf{D}_c \mathbf{R}_c \end{aligned} \qquad \text{(C.7)}$$

Linking two qualitative variables

We consider now a qualitative environmental variable (k_r categories) and a qualitative species trait (k_q categories). Data are coded in $\mathbf{R}$ and $\mathbf{Q}$ by respectively k_r and k_q dummy variables. For that case, one can create a k_r-by-k_q contingency table from tables $\mathbf{R}_o$ and $\mathbf{Q}_o$ and then compute a χ^2 statistic to measure the link between species trait and environment. The contingency table is obtained by the product $\mathbf{R}_o{}^t \mathbf{Q}_o$. From the table of proportions $\mathbf{R}_o{}^t \mathbf{D}_o \mathbf{Q}_o$, we derive diagonal matrices of row and column totals $\mathbf{D}_{k_r} = \mathbf{R}_o{}^t \mathbf{D}_o \mathbf{R}_o$ and $\mathbf{D}_{k_q} = \mathbf{Q}_o{}^t \mathbf{D}_o \mathbf{Q}_o$. The Pearson χ^2 statistic is given by:

$$\chi^2 = o.trace(\mathbf{D}_{k_q}{}^{-1} (\mathbf{R}_o{}^t \mathbf{D}_o \mathbf{Q}_o - \mathbf{D}_{k_r} \mathbf{1}_{k_r} \mathbf{1}_{k_q}{}^t \mathbf{D}_{k_q})^t \mathbf{D}_{k_r}{}^{-1} (\mathbf{R}_o{}^t \mathbf{D}_o \mathbf{Q}_o - \mathbf{D}_{k_r} \mathbf{1}_{k_r} \mathbf{1}_{k_q}{}^t \mathbf{D}_{k_q})) \qquad \text{(C.8)}$$

Using (C.3) and (C.4), we demonstrate that (C.8) can also be expressed using the original tables:

$$\begin{aligned} \chi^2 &= o.trace(\mathbf{D}_{k_q}{}^{-1} (\mathbf{R}^t \mathbf{P} \mathbf{Q} - \mathbf{D}_{k_r} \mathbf{1}_{k_r} \mathbf{1}_{k_q}{}^t \mathbf{D}_{k_q})^t \mathbf{D}_{k_r}{}^{-1} (\mathbf{R}^t \mathbf{P} \mathbf{Q} - \mathbf{D}_{k_r} \mathbf{1}_{k_r} \mathbf{1}_{k_q}{}^t \mathbf{D}_{k_q})) \\ &= o.trace(\mathbf{D}_{k_q}{}^{-1} (\mathbf{Q}^t \mathbf{P}^t \mathbf{R} - \mathbf{D}_{k_q} \mathbf{1}_{k_q} \mathbf{1}_{k_r}{}^t \mathbf{D}_{k_r}) \mathbf{D}_{k_r}{}^{-1} (\mathbf{R}^t \mathbf{P} \mathbf{Q} - \mathbf{D}_{k_r} \mathbf{1}_{k_r} \mathbf{1}_{k_q}{}^t \mathbf{D}_{k_q})) \\ &= o.trace((\mathbf{D}_{k_q}{}^{-1} \mathbf{Q}^t \mathbf{P}^t \mathbf{R} \mathbf{D}_{k_r}{}^{-1} - \mathbf{1}_{k_q} \mathbf{1}_{k_r}{}^t) \mathbf{D}_{k_r} (\mathbf{D}_{k_r}{}^{-1} \mathbf{R}^t \mathbf{P} \mathbf{Q} \mathbf{D}_{k_q}{}^{-1} - \mathbf{1}_{k_r} \mathbf{1}_{k_q}{}^t) \mathbf{D}_{k_q}) \end{aligned} \qquad \text{(C.9)}$$

Linking one qualitative variable and one quantitative

We finally consider the case of a qualitative environmental variable (k_r categories) and a quantitative species trait. Data are coded in **R** by k_r dummy variables while **Q** contains a single quantitative variable. If the variable in $\mathbf{Q}_o$ is standardized to mean 0 ($\mathbf{Q}_o{}^t\mathbf{D}_o\mathbf{1}_o = 0$) and variance 1 ($\mathbf{Q}_o{}^t\mathbf{D}_o\mathbf{Q}_o = 1$), the correlation ratio is given by:

$$\eta^2 = trace((\mathbf{R}_o\mathbf{D}_{k_r}{}^{-1}\mathbf{R}_o{}^t\mathbf{D}_o\mathbf{Q}_o)^t\mathbf{D}_o(\mathbf{R}_o\mathbf{D}_{k_r}{}^{-1}\mathbf{R}_o{}^t\mathbf{D}_o\mathbf{Q}_o)) \qquad \text{(C.10)}$$

Using (C.3) and (C.4), we demonstrate that (C.10) can also be expressed using the original tables:

$$\begin{aligned}\eta^2 &= trace(\mathbf{Q}_o{}^t\mathbf{D}_o\mathbf{R}_o\mathbf{D}_{k_r}{}^{-1}\mathbf{R}_o{}^t\mathbf{D}_o\mathbf{R}_o\mathbf{D}_{k_r}{}^{-1}\mathbf{R}_o{}^t\mathbf{D}_o\mathbf{Q}_o)\\ &= trace(\mathbf{Q}^t\mathbf{Y}_o{}^t\mathbf{D}_o\mathbf{X}_o\mathbf{R}\mathbf{D}_{k_r}{}^{-1}\mathbf{D}_{k_r}\mathbf{D}_{k_r}{}^{-1}\mathbf{R}^t\mathbf{X}_o{}^t\mathbf{D}_o\mathbf{Y}_o\mathbf{Q})\\ &= trace(\mathbf{Q}^t\mathbf{P}^t\mathbf{R}\mathbf{D}_{k_r}{}^{-1}\mathbf{R}^t\mathbf{P}\mathbf{Q})\\ &= trace((\mathbf{Q}^t\mathbf{P}^t\mathbf{R})\mathbf{D}_{k_r}{}^{-1}(\mathbf{Q}^t\mathbf{P}^t\mathbf{R})^t)\end{aligned} \qquad \text{(C.11)}$$

LITERATURE CITED

Dray, S. 2003. Eléments d'interface entre analyses multivariées, systèmes d'information géographique et observations écologiques. Thèse de doctorat. Université Lyon I, Lyon.