

RELATING BEHAVIOR TO HABITAT: SOLUTIONS TO THE FOURTH-CORNER PROBLEM

PIERRE LEGENDRE,¹ RENÉ GALZIN,^{2,4} AND MIREILLE L. HARMELIN-VIVIEN^{3,4}

¹*Département de Sciences biologiques, Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal, Québec, Canada H3C 3J7*

²*École Pratique des Hautes Études, Laboratoire d'Ichtyoécologie tropicale et méditerranéenne (Centre National de la Recherche Scientifique, Unité 1453), Université de Perpignan, F-66860 Perpignan Cedex, France*

³*Centre d'Océanologie de Marseille, Station marine d'Endoume (Centre National de la Recherche Scientifique, Unité 41), rue Batterie des Lions, F-13007 Marseille, France*

⁴*Centre de Recherches insulaires et Observatoire de l'Environnement, B.P. 1013, Moorea, Polynésie française*

Abstract. This paper addresses the following question: how does one relate the biological and behavioral characteristics of animals to habitat characteristics of the locations at which they are found? Ecologists often assemble data on species composition at different localities, habitat descriptions of these localities, and biological or behavioral traits of the species. These data tables are usually analyzed two by two: species composition against habitat characteristics, or against behavioral data, using such methods as canonical analysis. We propose a solution to the problem of estimating the parameters describing the relationship between habitat characteristics and biology or behavior, and of testing the statistical significance of these parameters; this problem is referred to as the fourth-corner problem, from its matrix formulation. In other words, the fourth-corner method offers a way of analyzing the relationships between the supplementary variables associated with the rows and columns of a binary (presence or absence) data table. The test case that motivated this study concerns a coral reef fish assemblage (280 species). Biological and behavioral characteristics of the species were used as supplementary variables for the rows, and characteristics of the environment for the columns. Parameters of the association between habitat characteristics (distance from beach, water depth, and substrate variables) and biological and behavioral traits of the species (feeding habits, ecological niche categories, size classes, egg types, activity rhythms) were estimated and tested for significance using permutations. Permutations can be performed in different ways, corresponding to different ecological hypotheses. Results were compared to predictions made independently by reef fish ecologists, in order to assess the method as well as the pertinence of the variables subjected to the analysis. The new method is shown to be applicable to a wide class of ecological problems.

Key words: *behavior; coral reef; fish; fourth-corner statistic; habitat; parameter estimation; permutation test; tropical ecology.*

INTRODUCTION

Niche theory tells us that species have ecological “preferences,” meaning that they are found at locations where they encounter appropriate living conditions. This statement is rooted in the observation that species have unimodal distributions along environmental variables, more individuals being likely to be found around some value that is “optimal” for the given species. This has been formalized by Hutchinson (1957) in his “fundamental niche” model, and used by ter Braak (1985) to justify correspondence analysis as a key method to study incidence or abundance data tables. Furthermore, Gause’s (1935) competitive exclusion principle suggests that in their evolution, species should have developed non-overlapping niches. These two principles together suggest that species should be

found to be roughly equally spaced in the n -dimensional space of resources. The influence of niche theory on competition theory (Bartlett 1960, Watanabe 1984) has raised some interesting questions recently (Tilman 1987). Niche theory remains limited, however, to answering questions such as: What is the ecological specificity of a species? How do different species apportion local resources among themselves? And, how does this mechanism control the associations of species found in nature?

It is understood that species have evolved biological and behavioral characteristics allowing them to exploit the given combination of resources represented in their niche. The following question, which also stems from niche theory, has been neglected, however, by lack of appropriate methods of analysis: *How do the biological and behavioral characteristics of species determine their relative locations in an ecosystem?* Observation of species in nature can lead ecologists to formulate

Manuscript received 5 June 1995; revised 22 April 1996; accepted 1 May 1996; final version received 5 June 1996.

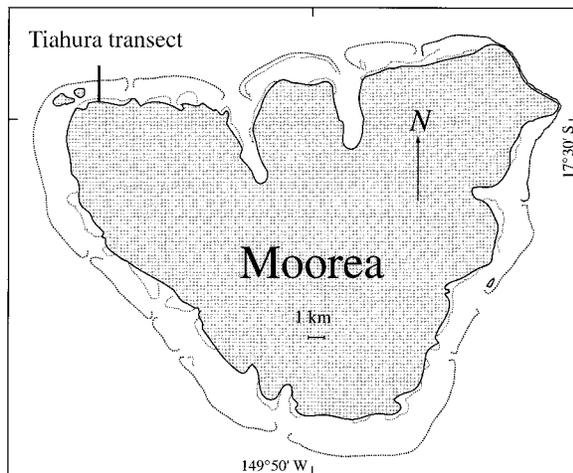


FIG. 1. Position of the Tiahura transect on Moorea Island, French Polynesia.

hypotheses in that respect. Testing such hypotheses would require (1) a way of detecting associations between species traits and habitat characteristics, and (2) a way of testing the significance of these associations. This paper presents such a method.

The question is difficult. Ecologists often assemble data on species composition at different localities, habitat descriptions of these localities, and/or biological or behavioral characteristics of the species. These data tables are usually analyzed by pairs: species composition against habitat characteristics, or against behavioral data, using such methods as canonical analysis. No standard statistical method can be used to directly analyze the relationship between the biological or behavioral characteristics of species and the characteristics of the habitat where they are found.

To fix ideas, imagine a first table **A** ($k \times m$) containing data on the presence or absence of k species at m locations (sampling stations, for instance). A second table **B** ($k \times n$) describes n biological or behavioral traits of the same k species. A third table **C** ($p \times m$) contains information about p environmental variables at the m locations. How does one go about relating the n biological and behavioral traits to the p environmental variables? In order to help find a solution, let us translate the problem into matrix algebra:

$$\begin{bmatrix} \mathbf{A} (k \times m) & \mathbf{B} (k \times n) \\ \mathbf{C} (p \times m) & \mathbf{D} (p \times n) \end{bmatrix}$$

The problem can now be stated as follows: How does one estimate the parameters in matrix **D** ($p \times n$) crossing the n biological and behavioral traits to the p environmental variables? Furthermore, are these associations significant in some sense, i.e., are they really different from 0 (no association), or from the value they could take in a randomly organized environment? Because of this matrix representation, the underlying

statistical problem can be referred to as that of estimating the parameters in the fourth-corner matrix **D**, or the *fourth-corner problem*. While the data in matrix **A** are necessarily of the presence/absence or frequency type, data in matrices **B** and **C** can be either quantitative or qualitative (nominal). Solutions have to be found that can accommodate any of these types of variables.

This paper presents statistical solutions to this problem, involving groups of qualitative or quantitative variables, or mixtures of quantitative and qualitative information. Our test case, which motivated the study, concerns a coral-reef fish assemblage of 280 species. Parameters of the association between habitat characteristics and biological and behavioral traits of the species will be estimated and tested for significance. They will then be compared to predictions made independently by reef-fish ecologists, in order to assess the method as well as the variables subjected to the analysis. This is the only way to validate the approach, in the absence of any other statistical method available at the present time to tackle this problem. A new array of biological questions are now open to statistical estimation. The method developed in this paper will be shown to be applicable to a wide class of ecological problems, and to other problems found in such fields as sociology, marketing, and political sciences.

MATERIALS AND METHODS

Coral reef fish data

Fish-assemblage data, used in this paper to illustrate the method, were collected by R. Galzin in 1982 and 1983, during daytime, along a transect extending from the beach to the outer reef slope, in the northwest part (called Tiahura) of the reef of the high volcanic island of Moorea in French Polynesia (Fig. 1). Moorea (149°50' W, 17°30' S) is a high volcanic tropical island of the Society archipelago, 25 km northwest of Tahiti. Water temperature varies from 25° to 29°C throughout the year. A coral reef, 800 m wide on average, and with 12 ocean passes, surrounds the island. Because it is narrow, this reef is ideal to study how marine organisms react to a coast-to-sea gradient. The transect was divided into 22 contiguous sampling stations from the beach (station 1) to the outer slope. On the reef, stations 1 to 16 were 50 m wide and all <8 m water depth, while station 17 was 40 m wide; stations 18 to 22, on the outer slope descending to the Pacific floor, were limited by depths of 3, 8, 15, 22, and 30 m (greater depths requiring diver safety stops). At each station, the presence/absence of 280 fish species was noted using two techniques: visual observations during 45 min inside the station (snorkeling or scuba diving), and limited rotenone collection around coral patches. Details about this transect and the fish counting method are found in Galzin and Pointier (1985), Galzin (1987a, b), and Galzin and Legendre (1987).

Two quantitative habitat variables were measured:

(1) distance (in meters) from the beach and (2) water depth (in centimeters). The other measured habitat variables were indices of percentage coverage of the reef bottom by different materials, based on 50 observation points. Along the 50-m rope, 50 observations were made at 1-m intervals. The following variables report what proportion of the 50 readings pertained to each category of substrate: (3) stone slab, (4) sand, (5) coral debris, (6) turf and dead coral, (7) live coral, (8) large algae, (9) calcareous algae, and (10) other substrate: large echinoderms (holothuroids and sea stars), sponges, anemones, or alcyonarians. Several of these categories represent biological material lying on top of, intermingled with, or attached to the mineral substrate. These variables are at an appropriate scale to be perceived by fish as important foraging characteristics of the habitat. When the 22 stations are considered globally, these eight categories of substrate respectively represented 2.5%, 31.1%, 11.4%, 14.3%, 13.9%, 18.7%, 7.5%, and 0.7% of the observed points.

The biological and behavioral characteristics of the 280 species (Table 1) are based on data from the literature (Hiatt and Strasburg 1960, Hobson 1974, Harmelin-Vivien 1979, 1989) and on personal observations by ichthyologists M. L. Harmelin-Vivien and R. Galzin. One variable is ordinal and four are nominal.

Statistical methods

Comparing two nominal variables.—The first situation considered implies two nominal variables, one from matrix **B** (behavior), the other from matrix **C** (habitat). Any nominal variable can be expanded into a series of binary variables (0, 1), one for each state. Assume that **B** and **C** each consist of a two-state nominal variable, as in test cases 1 and 2 (Table 2). To fix ideas, assume that the **B** variable describes two feeding-habit states (herbivorous, carnivorous) and that **C** is the nature of the substrate at two sampling stations (live coral, turf). We will use this example to describe the approach for nominal variables and to introduce a method of testing for statistical significance.

Matrices **A**, **B**, and **C** are all needed to obtain an estimate of the parameters in **D**. The simplest way to combine them is to multiply clockwise around the set of four matrices as follows, preserving matrix compatibility:

$$D = CA'B \tag{1}$$

or, which is equivalent counterclockwise:

$$D' = B'AC' \tag{2}$$

If the states of the nominal variables in **B** and **C** are 0 and 1 and mutually exclusive, or if they represent proportions, the sum of the values in **D** should be equal to the sum of the values in **A**, as it is the case in Table 2.

This equation has an equivalent in traditional statistics. The data in **A**, **B**, and **C** can be combined to form

TABLE 1. Biological and behavioral characteristics of the 280 fish species used as a test case in the proposed method for examining the relationships between habitat, biology, and behavior of organisms. Variable 2 is coded for diurnal behavior, since species can switch states from day to night. Frequencies of the states are given in parentheses.

Variable 1: Feeding habits (nominal)	
1	= herbivorous (<i>n</i> = 43)
2	= omnivorous (<i>n</i> = 43)
3	= diurnal grazer on sessile invertebrates: corals, alcyonarians, sponges, etc. (<i>n</i> = 25)
4	= carnivorous type 1 (diurnal): small crustaceans, molluscs, echinoderms, and polychaetes (<i>n</i> = 66)
5	= carnivorous type 2 (nocturnal): large crustaceans (crabs), cephalopods, fish (<i>n</i> = 58)
6	= fish eater (fish-only diet) (<i>n</i> = 16)
7	= zooplankton eater (copepods) (<i>n</i> = 29)
Variable 2: Ecological category (nominal)	
1	= hiding in holes, cavities, etc. (<i>n</i> = 58)
2	= living on the bottom (often poor swimmers) (<i>n</i> = 23)
3	= circling small territories around coral heads (<i>n</i> = 91)
4	= swimming above the coral heads; larger vital neighborhoods (<i>n</i> = 81)
5	= good swimmers, covering large distances on the reef (<i>n</i> = 8)
6	= sub-surface species (<i>n</i> = 11)
7	= pelagic species (<i>n</i> = 8)
Variable 3: Size class of adults (ordinal)	
1	= 0–15 cm (<i>n</i> = 81)
2	= 16–30 cm (<i>n</i> = 73)
3	= 31–60 cm (<i>n</i> = 83)
4	= 61–120 cm (<i>n</i> = 33)
5	= 121–240 cm (<i>n</i> = 9)
6	= larger than 240 cm (<i>n</i> = 1)
Variable 4: Egg type (nominal)	
1	= pelagic eggs (<i>n</i> = 192)
2	= benthic eggs (<i>n</i> = 86)
3	= viviparous species (<i>n</i> = 2)
Variable 5: Activity rhythm (nominal)	
1	= diurnal (<i>n</i> = 196)
2	= nocturnal (<i>n</i> = 38)
3	= indifferent (<i>n</i> = 46)

an “inflated data table” (Table 3a). Each row of that table corresponds to an occurrence in matrix **A**; the columns list the states of the nominal variables in **B** and **C** corresponding to that occurrence. Matrix **D**, defined by Eqs. 1 and 2, is exactly the result of crossing these two nominal columns; it is a contingency table containing frequencies (Table 3b). So, a solution for significance testing is to compute a χ^2 statistic, using either Pearson’s formula, or Wilk’s (also called the *G* statistic by Sokal and Rohlf 1995). The *G* statistic will be used in the present paper and forms the first type of fourth-corner statistic.

One cannot test these *G* statistics for significance in the usual manner, however, because in the general case, several species are observed at any one sampling station, and so the rows of the inflated table (Table 3) are not independent of one another; several rows of that matrix have resulted from observations at the same

TABLE 2. Test cases 1 and 2 for nominal variables. In each case, matrix **A** is (10 species × 2 stations), **B** is (10 species × 2 feeding habits), and **C** is (2 habitat types × 2 stations). So, **D** is (2 habitat types × 2 behavioral states).

Test case 1					Test case 2						
A)	Stn. 1	Stn. 2	B)	Herbiv.	Carniv.	A)	Stn. 1	Stn. 2	B)	Herbiv.	Carniv.
Sp. 1	0	1		0	1	Sp. 1	1	1		0	1
Sp. 2	0	1		0	1	Sp. 2	1	1		0	1
Sp. 3	1	0		0	1	Sp. 3	1	1		0	1
Sp. 4	1	0		0	1	Sp. 4	1	1		0	1
Sp. 5	1	0		0	1	Sp. 5	1	1		0	1
Sp. 6	0	1		1	0	Sp. 6	1	1		1	0
Sp. 7	0	1		1	0	Sp. 7	1	1		1	0
Sp. 8	0	1		1	0	Sp. 8	1	1		1	0
Sp. 9	0	1		1	0	Sp. 9	1	1		1	0
Sp. 10	0	1		1	0	Sp. 10	1	1		1	0
C)	Stn. 1	Stn. 2	D)	Herbiv.	Carniv.	C)	Stn. 1	Stn. 2	D)	Herbiv.	Carniv.
Live coral	1	0		0 -	3 +	Live coral	1	0		5	5
				<i>P</i> = 0.027	<i>P</i> = 0.494					<i>P</i> = 1.000	<i>P</i> = 1.000
				<i>E</i> = 0.03125	<i>E</i> = 0.500					<i>P</i> = 1.000	<i>E</i> = 1.000
Turf	0	1		5 +	2 -	Turf	0	1		5	5
				<i>P</i> = 0.027	<i>P</i> = 0.494					<i>P</i> = 1.000	<i>P</i> = 1.000
				<i>E</i> = 0.03125	<i>E</i> = 0.500					<i>E</i> = 1.000	<i>E</i> = 1.000
Contingency statistic: <i>G</i> = 5.4872; <i>P</i> (9999 perm.) = 0.058						Contingency statistic: <i>G</i> = 0.0000, <i>P</i> (9999 perm.) = 1.000					

Notes: Probabilities (*P*) are one-tailed, assuming that the sign of the relationship is stated in the hypothesis. The hypothesis is indicated by a sign in each cell of matrix **D**, + meaning that the actual value is expected to be in the upper tail (higher than the expected value), and - when it is expected to be in the lower tail. Probabilities were calculated after 9999 random permutations following model 1. *E* = exact probabilities.

sampling station. To solve the problem, we propose the following permutation (or randomization) test. A general introduction to randomization tests is found in Sokal and Rohlf (1995); more advanced texts are Edgington (1995) and Manly (1991).

1. Hypotheses.—

a) *H*₀: the species (reef fish in our examples) are distributed at random among the sampling stations.

b) *H*₁: the species are not distributed at random among the sampling stations. (See 5. *Permutation models*, below, for details on the ecological contents of the null and alternative hypotheses.)

2. Test statistic.—Compute a χ^2 statistic on the contingency table (matrix **D**) and use it as the reference value in the remainder of the test.

3. Distribution of the test statistic.—

a) Under *H*₀, the species found at any one station could have been observed at any other one. Where the species were actually observed is due to chance alone.

b) So, a realization of *H*₀ is obtained by permuting at random the values in matrix **A**, following one of the methods described below. After each permutation of matrix **A**, compute the χ^2 statistic on **D**.

c) Repeat this operation a large number of times (say, 999 or 9999 times). The different permutations produce a set of values of the χ^2 statistic, obtained under *H*₀.

d) Include in this set of values the reference χ^2 value computed for the unpermuted data matrix. This value is considered to be one that could be obtained under *H*₀ and, consequently, it should be added to the distribution (Hope 1968, Edgington 1995). The values form

TABLE 3. Inflated data table (a); there is one row in this table for each species “presence” (“1”) in matrix **A** of test case 1 (Table 2). From the inflated table, the contingency table (right) is constructed.

(a) Inflated data table			(b) Contingency table (D)		
Occurrences in test case 1	Feeding habits from B	Habitat types from C		Herbiv.	Carniv.
Sp. 1 @ Stn. 2	Carnivorous	Turf	Live coral	0	3
Sp. 2 @ Stn. 2	Carnivorous	Turf			
Sp. 3 @ Stn. 1	Carnivorous	Live coral	Turf	5	2
Sp. 4 @ Stn. 1	Carnivorous	Live coral			
Sp. 5 @ Stn. 1	Carnivorous	Live coral			
Sp. 6 @ Stn. 2	Herbivorous	Turf			
Sp. 7 @ Stn. 2	Herbivorous	Turf			
Sp. 8 @ Stn. 2	Herbivorous	Turf			
Sp. 9 @ Stn. 2	Herbivorous	Turf			
Sp. 10 @ Stn. 2	Herbivorous	Turf			

an estimate of the sampling distribution of χ^2 under H_0 . (For small data sets, one could compute all possible permutations in a systematic way and thus obtain the exact, or complete randomization distribution of the statistic, to be used in the next step.)

4. *Statistical decision.*—The decision is made by comparing the reference value of the χ^2 statistic to the distribution obtained under H_0 . If the reference value of χ^2 is one likely to have been obtained under the null hypothesis (which states that there is no relationship between the rows and columns of matrix **D**), H_0 is not rejected. If it is too extreme (i.e., located out in a tail) to be considered a likely result under H_0 , the H_0 is rejected.

5. *Permutation models.*—Permutations can be accomplished in different ways, depending on the nature of the ecological hypotheses to be tested against observations. Technically, the fourth-corner statistical method presented above can accommodate any of the permutation models described below, as well as permutations constrained to accommodate spatial or temporal autocorrelation (e.g., Legendre et al. 1990, ter Braak 1990). With our data (see *Coral reef fish data*, above), the random component is clearly the species found at the various sampling locations; this is found in matrix **A**. It is thus matrix **A** that should be permuted (randomized) for the purpose of hypothesis testing. This can be accomplished in various ways (Fig. 2).

Model 1: Environmental control over individual species.—The environmental control model (Whittaker 1956, Bray and Curtis 1957, Hutchinson 1957) states that species are found at locations where they encounter appropriate living conditions. Species do that independently of one another, contrary to the assemblage model that has *species assemblages* randomly located (next). Realizations of this null hypothesis are generated by permuting at random the values within each row vector of table **A**, and this independently from row to row. In that model, species associations are not functional; they simply result from the co-occurrence of species at particular locations, driven by environmental control (Fager 1963, Legendre and Legendre 1978, 1983). The number of stations occupied by any given species in a row of matrix **A** (its ubiquity) is fixed because it is considered to reflect such characteristics of the species as abundance, intraspecific competition, and territoriality, as well as ecological plasticity. If all parts of the environment were equally suitable for all species, as stated by H_0 , they could eventually all be present at any given location; this permutation model would allow it, while the species assemblage model (next) would not. The alternative hypothesis is that *individual species* find optimal living conditions at the stations where they are actually found.

Model 2: Environmental control over species assemblages.—This permutation method would be appropriate to test a null ecological hypothesis that the *species composition* at any one location is as likely to have

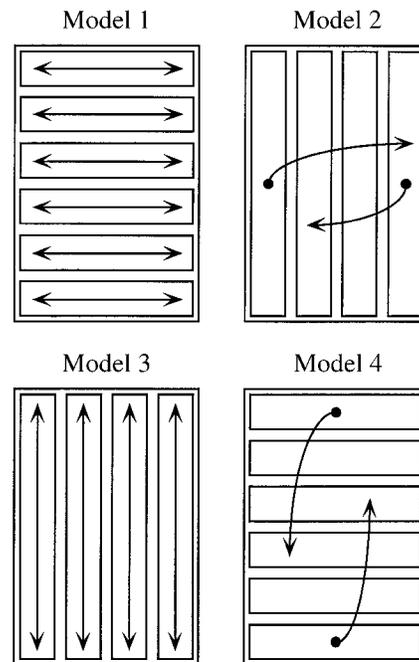


FIG. 2. Permutations of matrix **A** can be performed in four different ways, corresponding to different null ecological models. (1) The occurrence of a species on the reef is constant, but positions are random; permute at random within rows. (2) Positions of species assemblages are random; permute whole columns (assemblages). (3) Lottery hypothesis: the species that arrives first occupies a site; permute at random within columns. (4) Species have random attributes; permute whole rows.

occurred at any other location, against an alternative hypothesis that *species assemblages* are dependent upon the physical characteristics of the locations where they are actually found. In the context of the fourth-corner problem, this alternative hypothesis would be an extension of the environmental-control model (previous paragraph) to species assemblages, implying strong biotic ties among the species that are actually found together. This is the method of unrestricted permutations implemented in program CANOCO, for instance (ter Braak 1990). CANOCO is widely used to compute canonical correspondence analysis (ter Braak 1987), a technique allowing one to uncover relationships between a species presence-absence or abundance table on the one hand, and a table of environmental descriptors on the other. The hypothesis of no relationship can be tested by randomization, permuting the “site” vectors at random in one of the tables with respect to those in the other. The “site” vectors are the columns of our data table **A**.

Model 3: Lottery.—A third method consists in permuting values within columns of matrix **A**, and doing this independently from column to column. The null hypothesis says that there is a fixed number of niches at any one location, and that species invade them through some form of lottery, the identity (species) of

an individual settling at a station being a chance event. The lottery model has been advocated for coral reef fishes by Sale (1978), who argued that the main determinant of species composition at the various sites on coral reefs is chance, coupled with an overabundance of juveniles available for settlement. Instead of assuming the ubiquity of any one species to be fixed, it is the number of niches available for settlement that is assumed to be fixed. The alternative hypothesis is here that species have some competitive advantage over chance settlers in given habitats.

Model 4: Random species attributes.—A fourth method would be to permute whole rows at random with respect to one another. The corresponding null hypothesis is that species have random biological and behavioral attributes. This model is not appropriate for the data set analyzed in the present paper, because the relationships between species and their behavioral and biological characteristics are fixed. It may be appropriate to other types of problems, though.

6. *Remarks.*—

a) In all cases, some aspects of the data are considered fixed. In our data (see *Coral reef fish data*, above), the species are the same in matrices **A** and **B**. So, the rows of matrix **B** have to remain the same as the rows of matrix **A**; they cannot be permuted with respect to **A** (and consequently, permutation model 4 is deemed inappropriate). Furthermore, a given species has a fixed set of behavior states; the columns of **B** cannot be permuted with respect to one another. Likewise, the stations (column headings) are the same in matrices **A** and **C**; the habitat characteristics at any given sampling station cannot be permuted with those of other stations.

b) We have already shown in a previous paper (Galzin and Legendre 1987) that fish distributions along the Tiahura transect did not support the predictions of Sale's lottery model (permutation model 3). Because of the strong coast-to-sea gradient, it is the environmental control model that we consider the most appropriate. The number of times each species has been observed is considered a parameter of the ecological situation under description (and thus a parameter of the test), as explained in the environmental-control randomization model, and cannot be modified. To investigate possible associations between species traits and habitat characteristics, model 1 is certainly less restrictive than model 2, which would assume some kind of strong linkage among the species that are actually found together. So permutational model 1 will be used in the example analyzed in this paper.

c) Because the permutations are restricted to within the rows of matrix **A**, permutation model 1 is not equivalent to constructing the inflated data table (Table 3) and subjecting it to a standard χ^2 test. Results of the latter would be equivalent to carrying out unrestricted random permutations on one or the other of the right-hand columns of the inflated data table.

d) The observed values d in individual cells of matrix **D** can be tested for significance, in the same way as the general association between rows and columns of **D**, measured by the G statistic, is tested. In each cell, the values obtained during the permutations are compared to the reference value, and the number counted that are less than, equal to, and larger than the reference value. At the end of the permutation procedure, these values are used to compute probabilities of the relationship corresponding to each cell. In Table 2 (Test case 1), for instance, the association (correspondence) between turf and herbivores is found in the lower left-hand cell of matrix **D**. The probability shown in that cell is the one-tailed probability of the null hypothesis that herbivores are not positively associated with turf. If the actual value of the fourth-corner statistic d is lower than expected in that cell of the contingency table, the test is made in the left-hand tail of the distribution, and conversely. The mean of the d values found throughout the permutations is taken as an estimate of the expected value for any given cell; this value may differ markedly from the expected value of traditional contingency table analysis, which corresponds to a full randomization hypothesis.

e) The permutation testing procedure allows data in matrices **B** and **C** to be relative or absolute frequencies. This is the case with variables 3 to 10 in matrix **C** of the data on coral reef fishes. Each of these variables represents the proportion of the habitat covered by one of the eight categories of substrate. Using relative frequencies would have made these data unsuitable for traditional χ^2 testing. With the permutation procedure, however, the probabilities remain the same under any linear transformation of the frequency values, even though the value of the statistic is changed.

Let us examine how this procedure behaves when applied to two test data sets, each consisting of 10 hypothetical reef fish species and two sampling stations. The first test case of Table 2 was constructed to suggest that herbivores are found on turf while carnivores are more ubiquitously distributed. In matrix **D**, herbivores are clearly positively associated with turf and negatively with coral ($P = 0.0266$, computed by the random permutation procedure), while carnivores are not significantly associated with either live coral or turf ($P = 0.4940$, after 9999 random permutations following permutation model 1). These probability values are very close to the exact probabilities calculated for these data, which are the values obtained from a complete randomization procedure (E in Table 2). Values of exact probabilities E are computed as follows: consider all 2^{10} possible permutations that result from independently permuting the rows of matrix **A** in test case 1; count how many of these would produce values equal to, or more extreme than, the observed value in each given cell of matrix **D**. This value may differ slightly from the mean of the random experimental probabilities. Globally, the G statistic ($P = 0.0580$)

TABLE 4. Test cases 3 and 4 for quantitative or semi-quantitative (ordinal) variables. In each case, matrix **A** is (12 species \times 4 stations), **B** is (12 species \times 1 biological or behavioral variable), and **C** is (1 habitat variable \times 4 stations). So, **D** is (1 habitat variable \times 1 biological or behavioral variable).

Test case 3						Test case 4							
A)	Stn. 1	Stn. 2	Stn. 3	Stn. 4	B)	Size	A)	Stn. 1	Stn. 2	Stn. 3	Stn. 4	B)	Size
Sp. 1	1	0	0	0		1	Sp. 1	0	0	1	0		1
Sp. 2	1	0	0	0		2	Sp. 2	0	0	1	0		2
Sp. 3	1	0	0	0		3	Sp. 3	0	0	1	0		3
Sp. 4	0	1	0	0		4	Sp. 4	1	0	0	0		4
Sp. 5	0	1	0	0		5	Sp. 5	1	0	0	0		5
Sp. 6	0	1	0	0		6	Sp. 6	1	0	0	0		6
Sp. 7	0	0	1	0		7	Sp. 7	0	0	0	1		7
Sp. 8	0	0	1	0		8	Sp. 8	0	0	0	1		8
Sp. 9	0	0	1	0		9	Sp. 9	0	0	0	1		9
Sp. 10	0	0	0	1		10	Sp. 10	0	1	0	0		10
Sp. 11	0	0	0	1		11	Sp. 11	0	1	0	0		11
Sp. 12	0	0	0	1		12	Sp. 12	0	1	0	0		12
C)	Stn. 1	Stn. 2	Stn. 3	Stn. 4	D)	Size	C)	Stn. 1	Stn. 2	Stn. 3	Stn. 4	D)	Size
Depth	1	2	3	4		$r = 0.97163$ $P = 0.0001$	Depth	1	2	3	4		$r = 0.00000$ $P = 0.4960$

Notes: Probabilities (*P*) are one-tailed, and computed in the tail of the sign of the coefficient, assuming that the sign of the relationship is stated in the hypothesis. Probabilities were calculated after 9999 random permutations following model 1.

indicates a marginally significant association at significance level $\alpha = 0.10$ between behavioral states and types of habitat. Thus, the testing procedure for the association between behavior and habitat behaved as expected in this example, and the random permutation procedure produced values quite close to the exact probabilities.

The second test case illustrates a situation where the null hypothesis is true in all cases, matrix **A** indicating all 10 species to be present everywhere. Indeed, the testing procedure finds all permutation statistics to be equal to the unpermuted ones, so that the probability of the data under the null hypothesis is 1 everywhere. The procedure once more behaved correctly.

For large contingency tables **D**, relationships among descriptor states could be visualized from a correspondence-analysis ordination of the contingency table.

Comparing two quantitative or ordinal variables.—Quantitative data can be handled in almost the same way. Consider for the moment that **B** and **C** each contain a single quantitative variable, as in the test cases proposed in Table 4. An inflated data table can be constructed as above; there is one row in that table for each species “presence” (“1”) in matrix **A**. The inflated matrices corresponding to test cases 3 and 4 happen to have the same number of rows as the original matrix **A**; but in real-case studies, where several species may be present at each station, the number of rows of the inflated matrix should be higher, corresponding to the number of species occurrences in **A**. Two columns are written in the inflated matrix, the first one giving the values of the quantitative variable from **B**, the second the values of the variable from **C**. The cross-product of these two columns would give the fourth-corner statistic $d = \mathbf{CA}'\mathbf{B}$ for quantitative variables. If the two

variables in the inflated matrix are standardized to mean 0 and variance 1, and the cross-product is divided by (no. rows $- 1$), the fourth-corner statistic becomes a Pearson product-moment correlation coefficient *r*. For reasons given above, this correlation statistic should be tested using the permutation technique described in the previous section.

One should note that the general formula of correlation coefficients, applicable to ordinal as well as to quantitative data, is the same as that of the Pearson product-moment correlation coefficient (Kendall 1948). So **B** and **C** data tables containing either quantitative or ordinal variables can be handled in the same way, provided that the scale used for the ordinal variables is felt to be adequate. In case of doubt, one can calculate the inflated data table and draw a dispersion diagram, to verify that the scales used for the variables generate a linear dispersion of data points, when a relationship is present.

Test case 3, proposed in Table 4, is constructed in such a way that species of increasing sizes are found at increasing depths. The correlation coefficient of 0.972 indicates positive association between matrices **B** and **C**; the associated probability $P = 0.0001$ (one-tailed) shows that this association is strong. So, the permutation testing procedure behaved as expected with these test data. Test case 4, on the contrary, has been designed in such a way as to create no association between **B** and **C**. The correlation coefficient is 0 and the associated probability is 0.4960, which indicates a lack of association between **B** and **C**. Once again, the testing procedure performed as expected.

The unstandardized $d = \mathbf{CA}'\mathbf{B}$ statistic should not be used in place of the correlation statistic *r*, because the permutations of row contents of matrix **A**, inde-

TABLE 5. Test case 5 comparing a quantitative to a nominal variable. Matrix **A** is (12 species \times 4 stations), **B** is (12 species \times 1 quantitative biological or behavioral variable), and **C** contains a nominal variable (3 states \times 4 stations). So, **D** is (3 states \times 1 quantitative variable).

A) Stn. 1 Stn. 2 Stn. 3 Stn. 4					B) Size	
Sp. 1	1	0	0	0	1	
Sp. 2	1	0	0	0	1	
Sp. 3	1	0	0	0	1	
Sp. 4	0	1	0	0	4	
Sp. 5	0	0	1	0	5	
Sp. 6	0	0	0	1	6	
Sp. 7	0	1	0	0	7	
Sp. 8	0	0	1	0	8	
Sp. 9	0	0	0	1	9	
Sp. 10	0	1	0	0	10	
Sp. 11	0	0	1	0	11	
Sp. 12	0	0	0	1	12	

C) Stn. 1 Stn. 2 Stn. 3 Stn. 4					D) Size	
State 1	1	0	0	0	$d = 0.00000$ $P = 0.0141$	$r = -0.80472$ $P = 0.0022$
State 2	0	1	0	0	$d = 0.10573$ $P = 0.3420$	$r = 0.11496$ $P = 0.3715$
State 3	0	0	1	1	$d = 0.22026$ $P = 0.1516$	$r = 0.59735$ $P = 0.0233$

Notes: Probabilities (P) of the d statistics (matrix **D** left) are one-tailed and computed in the lower tail, to test for within-group homogeneity. Probabilities (P) of the correlation statistics r (matrix **D** right) are one-tailed, and computed in the tail of the sign of the coefficient, assuming that the sign of the relationship is stated in the hypothesis. Probabilities were calculated after 9999 random permutations following model 1.

pendently of one another, can produce cross-product values that are higher than the reference value, while the association between **B** and **C** is lower. In test case 3, for instance, the permutations might bring all "presences" in matrix **A** to station 4, except for the sixth one found in station 3. This would result in a cross-product statistic $d = 306$, higher than the actual value $d = 240$ for test case 3. There is no doubt, however, that the association between **B** and **C** would be much lower for that permutation of **A** than it is for test case 3, indicating inadequacy of the d statistic for this permutation test. On the contrary, correlation coefficients do behave correctly; in the case of that permutation, the correlation coefficient would be 0.044, much lower than the reference value of 0.972.

Comparing quantitative to nominal variables.—Comparing a quantitative or ordinal variable to a nominal variable recoded into dummy variables (binary states) can be done in two different ways.

a) Without standardization, the fourth-corner statistics $d = \mathbf{CA'B}$ are measures of within-group sums of squares, provided that the quantitative variable is centered within each group (= state) and the values are squared before summing. The within-group sums of squares are divided by the total sum of squares to provide normalized measures of within-group homogeneity taking values between 0 and 1 (see test case 5 in Table 5, matrix **D** left). No value is computed when a group contains a single, or no observation, since one cannot measure the homogeneity or heterogeneity of

such groups. Such cases, which can be produced by the permutation procedure but are likely to occur only in very small problems, are left out of the reference distribution during significance testing. The example has been constructed in such a way that only the first group, as defined by nominal variable **C**, is homogeneous. The permutation test allows one to reject the hypothesis of heterogeneity for the first group only (Table 5). The global statistic that can be derived, involving all groups of the nominal variable, is an analysis-of-variance statistic (F or Kruskal-Wallis; F will be used below). It can readily be computed from an inflated table, using the nominal variable as the classification criterion. It is tested by permutations for reasons given above.

b) Consider that the multistate nominal variable has been decomposed into a number of binary descriptors. If the quantitative variable and the binary descriptors are all standardized to mean 0 and variance 1, the fourth-corner statistics become correlation coefficients between the quantitative variable and the binary descriptors, as in *Statistical methods: Comparing two quantitative or ordinal variables*, above. An example is presented in Table 5 (matrix **D** right).

These two procedures serve different purposes and together provide three types of information. (1) First, the F statistic is a global measure of association between the two variables. In Table 5, the F test ($F = 9.30405$, $P = 0.0138$) tells us that at least one (and perhaps more) of the states of **C** differs from the others,

in terms of the values of **B** to which it is associated. A posteriori testing can be done by repeating the *F* test for pairs of groups (with appropriate correction for multiple testing). (2) Secondly, the fourth-corner statistics *d* answer the question of within-group homogeneity. The first group (state 1) is significantly more homogeneous than the hypothesis of random allocation of fish to sampling stations would suggest (Table 5). (3) Finally, the correlation coefficients *r* indicate the strength of association of the 1s of each state of the nominal variable to small or large values of the quantitative variable. Consider the first group in Table 5, the only one found above to be significantly homogeneous; the significant negative correlation coefficient indicates that the presence of environmental state 1 is associated with small values of size. By way of consequence, environmental state 3 is associated with large values of size. Interpretation of the associated probabilities is not without problems, though, because the states of the nominal variable are mutually exclusive; therefore, testing the relationship of the quantitative variable with each state of the nominal descriptor, through simple correlation coefficients, makes little sense. A multiple-regression approach, which is the equivalent of an analysis of variance, would be more appropriate; in test case 5, a multiple regression of the **B** variable on any two of the three binary state variables provides an R^2 of 0.67401 ($P = 0.0146$). The problem of interpreting tables of correlation coefficients is discussed further below.

When the nominal variable consists of a series of percentage coverage indices, as it is for our variables describing substrates (see description of the data, above), an analysis-of-variance approach cannot be used because states are not mutually exclusive (fuzzy classification). We will be using the correlation fourth-corner statistic, although a multiple-correlation approach also would be appropriate to obtain a global statistic linking the quantitative variable to all percentage coverage indices.

When analyzing real data, one has to correct individual tests to accommodate the increased probability of committing a Type I error in the case of multiple simultaneous tests. This can be done either by adjusting individual probability values (*P* values), or by adjusting the significance level α . A traditional way is to use the Bonferroni correction, where the significance level, say $\alpha = 0.05$, is replaced by an adjusted level $\alpha' = \alpha/k$ where *k* is the number of simultaneous tests. This is equivalent to adjusting individual *P* values P_i to kP_i and comparing P'_i to the unadjusted significance level α . While the Bonferroni method is appropriate to test that the null hypothesis is true for the whole set of simultaneous hypotheses (i.e., reject H_0 for the whole set of *k* hypotheses if the smallest unadjusted *P* value in the set is $\leq \alpha/k$), the Bonferroni method is overly conservative. Several alternatives have been proposed in the literature; see Wright (1992) for a review. In this

paper, we will use Holm's (1979) procedure for adjusting individual probabilities. It is nearly as simple to carry out as the Bonferroni adjustment and it is much more powerful, leading to rejecting the null hypothesis more often. Other solutions, such as Hochberg's (1988) and Hommel's (1988) procedures, are even more powerful, but they are known to have the desired experimentwise error rate α only for independent tests (Wright 1992). This condition is not met in some of our tables of results, where for instance "Other substrate" is what remains after considering all the other substrates in the analysis.

Holm's (1979) procedure, which remains valid in this type of situation, is computed as follows. (1) Order the *P* values from left to right so that $P_1 \leq P_2 \leq \dots \leq P_k$. (2) Compute adjusted probability values $P'_i = P_i(k - i + 1)$; adjusted probabilities can be > 1 . (3) Proceeding from left to right, if an adjusted *P* value in the ordered series is smaller than the one occurring at its left, make the smaller one equal to the larger one. (4) Compare each adjusted P'_i to the unadjusted α significance level and take the statistical decision. Because the adjusted probabilities form a nondecreasing series, this procedure presents the property that an hypothesis in the ordered series cannot be rejected unless all previous hypotheses in the series have also been rejected, and, equal *P* values receive equal adjusted *P* values.

Validation

In an attempt to validate the method, we compared the fourth-corner results to predictions made independently by reef fish ecologists. The difficulty was that there exists no other statistical method, at the present time, to address the problem and provide results with which the fourth-corner results could be compared; hence the use of traditional ecological analysis by experts. Description of how this was done is saved for the *Discussion*.

RESULTS

Tables 6c and 7c present relationships of the reef bottom materials to feeding habits and ecological categories. Since each of these tables implies 56 simultaneous tests, Holm's procedure did not allow detection at the 5% significance level after only 999 permutations (the smallest attainable probability is 0.001, which, when multiplied by 56 simultaneous tests, gives $P = 0.056$, a value > 0.05). Therefore, 9999 permutations were used in these comparisons, as well as Table 8c where the same problem existed to a certain extent (24 simultaneous tests for each of the two biological or behavioral variables). In any case, considering the well-known instability of randomization probabilities computed from a small number of permutations (Jackson and Somers 1989), as many permutations as practically possible should be used during permutation tests.

TABLE 6. Feeding habits are compared to all habitat variables. The table reports probabilities adjusted using Holm's procedure within each habitat variable: Parts a and b, 999 permutations; part c, 9999 permutations, following model 1. Sign in part c indicates whether a statistic is above (+) or below (−) the estimated expected value. Results of the global tests of significance (F , G) are also given.

	Feeding habits						
	Herbivorous	Omnivorous	Sessile inverteb.	Carniv. 1 diurnal	Carniv. 2 nocturnal	Fish only	Copepod eater
a) Distance from the beach (d = homogeneity, r = correlation fourth-corner statistics)							
$d(i,j)$	0.20360	0.16772	0.11741	0.24177	0.12146	0.02898	0.09053
Pr(d)	0.195	2.061	0.195	0.048	0.007	2.061	2.061
$r(i,j)$	−0.01124	−0.09030	−0.05195	−0.01669	0.14451	0.00259	0.03357
Pr(r)	0.621	0.007	0.015	0.621	0.007	0.621	0.288
$F = 6.661$, P (999 perm.) = 0.001							
b) Water depth (d = homogeneity, r = correlation fourth-corner statistics)							
$d(i,j)$	0.20849	0.11549	0.09352	0.22850	0.18997	0.02654	0.12210
Pr(d)	1.580	0.450	0.168	0.360	1.964	1.580	1.964
$r(i,j)$	0.00650	−0.04320	−0.04506	−0.04694	0.07819	0.00081	0.07469
Pr(r)	0.744	0.081	0.056	0.040	0.014	0.744	0.014
$F = 3.544$, P (999 perm.) = 0.001							
c) Materials covering reef bottom ($G = 15.426$, P (9999 perm.) = 0.0001)							
Stone slab	6.20−	5.84+	3.72−	8.42−	5.18+	0.96+	2.40−
P	0.429	0.232	1.535	2.650	2.650	2.650	2.650
Sand	81.22−	54.26−	43.34−	94.38−	35.90−	8.94−	26.26−
P	0.039	0.799	0.006	0.006	0.006	0.799	0.039
Coral debris	34.96+	20.22−	24.32+	46.74+	25.60+	4.48+	12.08−
P	1.976	1.976	0.006	0.009	0.645	2.650	2.650
Turf, dead coral	45.46+	27.88+	28.28+	57.58+	33.58+	6.20+	15.76+
P	0.207	2.650	0.081	0.013	0.029	1.976	2.650
Live coral	49.86+	28.50+	29.20+	58.28+	40.82+	6.22+	21.06+
P	0.006	1.976	0.006	0.006	0.006	1.976	0.006
Large algae	44.66−	37.50+	28.12−	59.68−	32.26−	6.34−	19.20−
P	0.006	2.650	0.105	0.048	0.140	2.650	2.650
Calcar. algae	29.12+	16.32+	16.08+	31.00+	26.02+	4.50+	11.32+
P	0.006	1.030	0.079	0.122	0.006	0.207	0.036
Other substrate	2.52+	1.48+	1.94+	2.92+	1.64+	0.36+	0.92+
P	0.105	2.650	0.006	0.795	1.734	1.976	1.976

Comparing two nominal variables

In Table 6c, the relationship between reef bottom materials and feeding habits is globally significant ($P_{G \text{ statistic}} = 0.0001$), and 20 of the fourth-corner statistics d are marked as significant. According to these significant d values, fish are under-represented on sand and large algae, and are unrelated to stone slab. In addition, herbivores are overrepresented on live coral and calcareous algae. Grazers of sessile invertebrates as well as carnivores types 1 and 2 are overrepresented on coral debris, turf and dead coral, live coral, calcareous algae, and "other substrate" (large echinoderms, sponges, anemones, alcyonarians); this includes all areas where herbivores were found. Copepod eaters are overrepresented on live coral and calcareous algae. Omnivores and specialist piscivores (fish-only diet) do not exhibit significant associations with substrate.

In Table 7c, the relationship between reef bottom materials and ecological categories is globally significant ($P_{G \text{ statistic}} = 0.0004$), and 17 of the fourth-corner statistics d are marked as significant. Materials avoidance is the same as in Table 6. In addition, fish hiding

in holes and cavities are overrepresented on live coral and calcareous algae. Fish circling small territories around coral heads, as well as those with larger vital neighborhoods above coral heads, are mostly found on live coral, turf and dead coral, coral debris, calcareous algae, and other substrate. Good swimmers, as well as pelagic species, were found everywhere. Subsurface species are often associated with the stone slab—the only group overrepresented in this habitat.

In Table 8c, the relationships are significant between reef bottom materials, on the one hand, and egg types ($P_{G \text{ statistic}} = 0.0029$) and activity rhythms ($P_{G \text{ statistic}} = 0.0001$) on the other. Except for the general underrepresentation on sand and large algae mentioned above and the lack of association with stone slab, species producing pelagic eggs are significantly associated with all other substrate types; species producing benthic eggs are positively associated only with live coral and calcareous algae. Not much can be concluded about the two viviparous species in the study. On the other hand, typically diurnal species are significantly associated with all substrate types except those generally

TABLE 7. Ecological categories are compared to all habitat variables. The table reports probabilities adjusted using Holm's procedure within each habitat variable. Parts a and b, 999 permutations; part c, 9999 permutations, following model 1. Sign in part c indicates whether a statistic is above (+) or below (-) the estimated expected value. Results of the global tests of significance (*F*, *G*) are also given.

	Hiding in holes	Living on bottom	Around cor. heads	Above cor. heads	Good swimmers	Sub-surface species	Pelagic species
a) Distance from the beach (<i>d</i> = homogeneity, <i>r</i> = correlation fourth-corner statistics)							
<i>d</i> (<i>i,j</i>)	0.14457	0.05515	0.38554	0.36568	0.01692	0.01564	0.01204
Pr(<i>d</i>)	3.795	3.795	3.795	0.007	0.186	3.795	3.795
<i>r</i> (<i>i,j</i>)	0.02452	-0.05784	0.00146	0.00162	0.00040	-0.00467	0.02764
Pr(<i>r</i>)	0.830	0.091	1.776	1.776	1.776	1.776	0.810
<i>F</i> = 1.020, <i>P</i> (999 perm.) = 0.281							
b) Water depth (<i>d</i> = homogeneity, <i>r</i> = correlation fourth-corner statistics)							
<i>d</i> (<i>i,j</i>)	0.16225	0.04764	0.36999	0.38412	0.01577	0.00045	0.01460
Pr(<i>d</i>)	1.872	1.872	1.872	1.670	1.212	0.280	1.872
<i>r</i> (<i>i,j</i>)	0.00503	-0.03082	0.02306	-0.00311	-0.01586	-0.05335	0.02923
Pr(<i>r</i>)	0.836	0.744	0.744	0.836	0.777	0.007	0.744
<i>F</i> = 1.180, <i>P</i> (999 perm.) = 0.178							
c) Materials covering reef bottom (<i>G</i> = 11.686, <i>P</i> (9999 perm.) = 0.0004)							
Stone slab	5.42+	1.82+	11.22-	11.78-	0.90+	1.12+	0.46+
<i>P</i>	1.778	2.918	0.946	0.456	2.918	0.006	2.918
Sand	44.26-	20.70-	124.14-	142.56-	8.14-	1.66-	2.84-
<i>P</i>	0.006	2.918	0.006	0.006	1.930	0.044	0.946
Coral debris	23.70+	6.34-	60.38+	69.34+	5.36+	1.22-	2.06+
<i>P</i>	1.925	1.778	2.039	0.009	0.099	2.902	2.902
Turf, dead coral	29.52+	10.06+	80.74+	85.12+	4.24-	2.84+	2.22+
<i>P</i>	1.778	2.918	0.021	0.006	2.918	0.946	2.918
Live coral	35.04+	9.42+	88.02+	90.60+	5.70+	1.94+	3.22+
<i>P</i>	0.006	2.918	0.006	0.006	0.946	2.918	0.456
Large algae	33.50-	14.06+	86.20-	83.98-	4.20-	3.94+	1.88-
<i>P</i>	2.183	2.183	0.062	0.006	0.602	0.946	1.833
Calcareous algae	18.98+	4.18-	53.62+	51.02+	3.08+	1.28+	2.20+
<i>P</i>	0.009	2.343	0.006	0.006	2.039	2.918	0.230
Other substrate	1.58+	0.42-	4.68+	4.60+	0.38+	0.00-	0.12+
<i>P</i>	1.980	2.902	0.006	0.006	0.368	0.126	2.918

avoided by fish. During daytime, nocturnal species are overrepresented around live coral and coral debris; rhythm-indifferent species are significantly associated with live coral, turf and dead coral, and calcareous algae.

Comparing quantitative to ordinal variables

The two quantitative habitat variables, distance from the beach and water depth, were compared to the size classes of adult fishes (ordinal) using the correlation fourth-corner statistic (Table 9a). Results indicate that adult fish size is slightly positively related to distance from the beach (*r* = 0.05043, *P* = 0.011), but not to water depth (*r* = 0.02271, *P* = 0.143). Size of adult fish significantly increases at stations with more coral debris and fewer large algae (Table 9b).

Comparing quantitative to nominal variables

The relationship between feeding habits and distance from the beach is globally significant (Table 6a), as is the relationship of feeding habits with water depth. Omnivores and grazers on sessile invertebrates are significantly associated with sites closer to the beach, while nocturnal carnivores and copepod eaters are sig-

nificantly associated with deeper stations and/or sites located farther offshore. Diurnal carnivores avoid the very shallow stations near the beach. Nocturnal carnivores are the only group that is strongly homogeneous in terms of distance from the beach; analysis of the inflated data table shows that most sightings were made at the nine farthest stations, farther than 650 m from the beach.

Ecological categories, on the other hand, are not significantly related to distance from the beach or to water depth (Table 7a, b). Egg types are weakly associated with distance from the beach, and not at all to depth (Table 8a, b). Species with benthic eggs are found at shallower stations, and those with pelagic eggs at deeper locations; species with pelagic eggs are significantly homogeneous in terms of distance from the beach; analysis of the inflated data table shows that the frequency of sightings increases with distance from the beach. Activity rhythm is strongly associated with distance from the beach, and weakly with water depth. Diurnal species are found closer to the beach, while rhythm-indifferent species are significantly found farther away and in deeper waters.

TABLE 8. Egg types (left) and activity rhythms (right) are compared to all habitat variables. The table reports probabilities (Pr) adjusted using Holm's procedure within each habitat variable. Parts a and b, 999 permutations; part c, 9999 permutations, following model 1. Sign in part c indicates whether a statistic is above (+) or below (-) the estimated expected value. Results of the global tests of significance (F , G) are also given.

	Egg type			Activity rhythm		
	Pelagic	Benthic	Viviparous	Diurnal	Nocturnal	Indifferent
a) Distance from the beach (d = homogeneity, r = correlation fourth-corner statistics)						
$d(i,j)$	0.72426	0.26842	0.00281	0.78613	0.11920	0.08132
Pr(d)	0.003	1.000	0.590	0.832	0.832	0.018
$r(i,j)$	0.06217	-0.06533	0.01810	-0.07745	-0.00634	0.11528
Pr(r)	0.006	0.006	0.244	0.004	0.399	0.003
$F = 3.094$, $P = 0.018$				$F = 9.235$, $P = 0.001$		
b) Water depth (d = homogeneity, r = correlation fourth-corner statistics)						
$d(i,j)$	0.77843	0.21856	0.00146	0.73333	0.13655	0.12566
Pr(d)	0.873	0.441	0.844	0.069	1.622	1.622
$r(i,j)$	0.03742	-0.03509	-0.01657	-0.03984	-0.00999	0.06673
Pr(r)	0.141	0.141	0.355	0.096	0.311	0.015
$F = 1.061$, $P = 0.218$				$F = 3.055$, $P = 0.022$		
c) Materials covering reef bottom						
$G = 4.759$, P (9999 perm.) = 0.0029				$G = 7.712$, P (9999 perm.) = 0.0001		
Stone slab	24.16-	8.46+	0.10-	24.92-	4.78+	3.02-
P	1.055	1.527	1.527	0.431	0.431	0.767
Sand	251.70-	91.60-	1.00-	278.98-	42.26-	23.06-
P	0.002	0.021	0.853	0.002	0.004	0.002
Coral debris	133.50+	34.08-	0.82+	130.08+	22.76+	15.56+
P	0.002	0.216	1.527	0.017	0.016	0.767
Turf, dead coral	163.04+	50.34+	1.36+	166.70+	24.82+	23.22+
P	0.002	0.726	0.506	0.002	0.754	0.002
Live coral	180.28+	52.58+	1.08+	177.50+	29.48+	26.96+
P	0.002	0.018	1.527	0.002	0.002	0.002
Large algae	164.76-	62.18-	0.82-	179.90-	26.42-	21.44-
P	0.002	1.527	1.527	0.002	0.072	0.270
Calcareous algae	102.44+	31.12+	0.80+	102.58+	14.08+	17.70+
P	0.002	0.003	0.899	0.002	0.431	0.002
Other substrate	9.12+	2.64+	0.02-	9.34+	1.40+	1.04+
P	0.002	1.034	1.527	0.002	0.431	0.767

Validation

Before the results of these analyses were known, reef fish ecologists (M. L. Harmelin-Vivien and R. Galzin) made the predictions reported in Table 10. They concern feeding habits and ecological categories, compared to reef bottom materials. Admittedly, these predictions are not independent of the data that were used in the analyses presented here, since the data were collected by R. Galzin on Tiahura in the early 1980s, but the two analyses were carried out independently from one another.

DISCUSSION

Dolédec et al. (1996) have recently proposed a different but complementary approach to the problem of jointly analyzing **A**, **B**, and **C**. Their approach consists in an ordination of table **A** under a double constraint consisting of **B** and **C**. The two approaches differ in that we are primarily interested here in relating the variables in **B** and **C** and testing the significance of these relationships, while the purpose of Dolédec et al. (1996) is primarily to obtain an ordination of sites and

species in table **A** under constraints, and to project the variables in **B** and **C** onto the ordination axes for interpretation.

Comparison of Tables 8 and 9 indicates that the fish ecologist predictions generally agreed with the results of the fourth-corner statistics (Table 10). For feeding habits, there are three opposite results over 29 predictions, for an error rate of 10%; 15 predictions were not verified by the data (52%), possibly because insufficient observations were available (poor statistical power). For ecological categories, the fish ecologists did even better; a single case is found where the fourth-corner statistics contradict ecologists' predictions (5% error rate); a further eight predictions were not verified by the data (40%). All four discrepancies in Table 10 (i.e., opposite predictions and statistics) concern carnivores 1 and 2.

Confronted with the fourth-corner statistic results, the fish ecologists (M. L. Harmelin-Vivien and R. Galzin) checked their predictions and confirmed that they did not think they had made any gross ecological mistake about the habitat and behavior of these fish. Com-

parison of the two sets of results is informative because it reveals conceptual differences between the two approaches.

a) The fish families involved in the relationship between sand and diurnal carnivores (discrepancy 1) are essentially the Mullidae (goatfishes), Labridae (wrasses), Balistidae (triggerfishes), Tetraodontidae (puffers), Lethrinidae (emperors), and Gobiidae (gobies). The goatfishes, emperors, and gobies find their prey preferably inside the sandy part of the reef during daytime (Harmelin-Vivien 1979), and therefore a positive relationship was predicted.

b) The same families are involved in the relationship between large algae (discrepancy 2) and diurnal carnivores. From the results of Naim (1988) for the Tiahura transect, we know that macroalgae provide shelter for small motile invertebrates that are preyed upon by wrasses, triggerfishes, and puffers, suggesting a positive relationship, which is what was predicted.

c) The nocturnal carnivores involved in the relationship with sand (discrepancy 3) belong essentially to families Lutjanidae (snappers), Holocentridae (squirrelfishes and soldierfishes), Scorpaenidae (scorpionfishes), Serranidae (groupers), Diodontidae (porcupinefishes), Muraenidae (morays), and Cirrhitidae (hawkfishes). Ecologists predicted a positive relationship essentially by considering the snappers, which feed at night on the invertebrates living in the sand. No such relationship is known for the other nocturnal carnivores. Considering the scarcity of snappers on the Tiahura transect, we could understand an absence of relationship between nocturnal carnivores and sand, but not a negative one.

d) The prediction of a positive association between fish living above coral heads and large algae (discrepancy 4) is based on the observation that large algae are largely found attached to hard substrates (Payri 1987).

The fourth-corner statistics produced a negative relationship between large algae and diurnal carnivores (discrepancy 2) as well as position above the coral heads (discrepancy 4). The reason is the following. On Tiahura, large algae are most abundant in the first stations of the transect, near the beach. Since few species are present there, all fish types appear to avoid large algae as well as sand or, at best, to be indifferent to these substrates (full-transect analysis reported in Tables 8 and 9). At these very shallow stations, carnivores, in particular, are less abundant than in the rest of the transect. The fourth-corner method thus has found a negative relationship, which is witness of an indirect effect. Fish ecologists, on the other hand, had first in mind the feeding habits of the fish. On the Tiahura transect, all carnivores find their prey essentially in or around macroalgae (Naim 1988) or inside hard substrates (Harmelin-Vivien 1979). So, considering feeding habits, the situation is that of a positive relationship between carnivores and large algae. To reach such a conclusion, fish ecologists did not take

TABLE 9. Size of adult fish is compared to all habitat variables, using correlation fourth-corner statistics. Permutations are computed following model 1. In part b, probabilities (999 permutations) were adjusted using Holm's procedure.

Habitat variables	Size of adult fish
a) Quantitative variables	
Distance from the beach	0.05043
Pr(<i>r</i>)	0.011
Water depth	0.02271
Pr(<i>r</i>)	0.143
b) Materials covering reef bottom	
Stone slab	-0.02236
Pr(<i>r</i>)	0.381
Sand	-0.03798
Pr(<i>r</i>)	0.220
Coral debris	0.06079
Pr(<i>r</i>)	0.035
Turf, dead coral	0.02548
Pr(<i>r</i>)	0.381
Live coral	0.03476
Pr(<i>r</i>)	0.220
Large algae	-0.06740
Pr(<i>r</i>)	0.016
Calcareous algae	0.05174
Pr(<i>r</i>)	0.066
Other substrate	0.02385
Pr(<i>r</i>)	0.381

into account the spatial distribution of the algae and of the fish on the transect, and the dominance of macroalgae in shallow areas; sand (discrepancies 1 and 3) and large algae (discrepancies 2 and 4) are dominant at stations near the beach, where all fish types are not abundant because of the absence of hard substrate and of very shallow water. The fourth-corner method, on the other hand, found negative relationships because it takes into account all stations in the study area; it would probably have found the predicted positive relationships had the study been conducted at a smaller spatial scale. This is an illustration that correlation does not mean causation.

The results presented in Tables 9 to 12, establishing the linkage of ecological and behavioral characteristics of fish species to the distance from the beach, statistically confirm and refine the results obtained by Harmelin-Vivien (1989), who compared and contrasted the characteristics of fish assemblages on the Tiahura reef flat and reef slope. (1) Diurnal fish were found to be more abundant on the reef flat, and nocturnal fish on the reef slope; the fourth-corner statistics indicate a similar significant negative relationship between diurnal fish and distance from the beach (Table 8). (2) Fish laying benthic eggs dominated on the reef flat, and fish with pelagic eggs on the reef slope; the fourth-corner method confirms these results, with a positive relationship between distance from the beach and pelagic eggs, and a negative relationship for benthic eggs (Table 8). (3) In the same way, smaller fish were ob-

TABLE 10. Relations between variables C3–C10 (reef bottom, rows) and variables B1 (feeding habits, part a) and B2 (ecological category, part b). Predictions made by coral reef ecologists are in parentheses, followed by the results from Tables 6 and 7. Blanks indicate no relation, (+) or + a positive relation, and (–) or – a negative relation. Discrepancies are numbered to facilitate reference in the text.

a)		Feeding habits					
Reef bottom	Herbivor.	Omnivorous	Sessile inverteb.	Carniv. 1 diurnal	Carniv. 2 nocturnal	Fish only	Copepod eater
Stone slab			(–)–	(+)– ¹	(+)– ³	(+)	(–)
Sand	–		+	(+)– ¹	(+)– ³	(+)	(–)–
Coral debris	(+)	(+)	+	(+)– ¹	(+)– ³	(+)	(–)–
Turf, dead coral	(+)	(+)		(+)– ¹	(+)– ³		
Live coral	(+)–	(+)	(+)–	(+)– ¹	(+)– ³	(+)	(+)–
Large algae	–	(+)	(–)	(+)– ²	(+)– ²	(+)	(+)–
Calcareous algae	+			(+)– ²	(+)– ²		+
Other substrate	(–)	(–)	(+)–				

b)		Ecological category					
Reef bottom	Hiding in holes	Living on bottom	Around coral heads	Above coral heads	Good swimmers	Sub-surface species	Pelagic species
Stone slab			(–)	(–)		+	
Sand	(–)–	(+)	(–)–	(–)–		–	
Coral debris			(+)–	(+)–			
Turf, dead coral		(–)	(+)–	(+)–			
Live coral	(+)–	(–)	(+)–	(+)–			
Large algae		(–)	(+)–	(+)– ⁴			
Calcareous algae	(+)–	(–)	+	+			
Other substrate		(–)	(+)–	+			

served on the reef flat; a positive correlation between size of fish and distance from shore is obtained. (4) Omnivores and sessile invertebrate feeders were found on reef slopes. These relationships are confirmed by the fourth-corner statistics and their tests of significance: a negative correlation is found in Table 6 between distance from the beach and omnivores as well as sessile invertebrate feeders, and a positive correlation for carnivores 2.

CONCLUSION

The fourth-corner method offers a way of analyzing the relationships between the supplementary variables associated with the rows and columns of a binary (presence-absence) data table. The parameters describing these relationships can be estimated and tested for statistical significance. In the example analyzed in the present paper, biological and behavioral characteristics of species were used as supplementary variables for the rows, and characteristics of the environment for the columns. Other ecological situations are considered below.

The fourth-corner method allows the detection of a new family of associations or correlations, between species traits and habitat characteristics, in a traditional data-analysis framework. The matrix operation leading to estimating the parameters in the fourth-corner matrix, $\mathbf{D} = \mathbf{CA}^t\mathbf{B}$, is very general and can accommodate any type of variable (quantitative, semi-quantitative, or qualitative). The method is not a modelling technique

since it does not take into account the dynamics of individual species or the spatial and temporal structure of environmental variation. However, the fourth-corner statistics can be incorporated into causal ecological models describing the mechanisms determining the observed associations or correlations. On the one hand, the sign and indication of strength (value, significance) of fourth-corner statistics can be incorporated as parameters in deterministic models; on the other hand, the fourth-corner correlation coefficients can be used directly in path analysis or other related forms of statistical modelling.

Analysis of the Tiahura fish data set (280 species), which motivated the development of the method, illustrated that fourth-corner statistics adequately identify positive or negative associations between the biological or behavioral traits of animals, and the habitat characteristics of the locations at which they are found, much in the same way as correlation coefficients do in traditional data analysis. In many instances, the statistical results confirmed the intuitive analysis of field ecologists. The statistical analysis also revealed a set of relationships that had not been predicted by the fish ecologists, and which can now be embodied in our understanding of the ecological determinants of these animals.

Currently, this method does not take into account indirect relationships that may exist among variables, as can be done with path analysis for instance. A multiple-interaction form of fourth-corner statistics re-

mains to be developed (log-linear modelling, multiple regression, path analysis, etc.) to test more sophisticated hypotheses. Another interesting development would be to extend the fourth-corner method to a table **A** of species abundance data; this extension is not straightforward.

Other ecological problems could be studied using this method. In the study of feeding behavior for instance, consider a matrix **A** with rows that are individuals and columns corresponding to locations. The prey ingested by each individual are found in matrix **B** (in columns). Matrix **C** may contain either microhabitat environmental variables, or prey availability variables. One could use fourth-corner statistics to identify relationships between prey availability and consumption, or between prey consumption and environmental variables. Problems of the same type are found in such fields as sociology, marketing, political science, and the like.

ACKNOWLEDGMENTS

Ottar Bjørnstad, Charlotte K. Hemelrijk, and François-Joseph Lapointe commented on the permutation models. The work of P. Legendre in French Polynesia and in Perpignan in 1994 has been supported by a CNRS (France) Associate Research position in URA 1453, Université de Perpignan. This research was also supported by NSERC grant No. OGP0007738 to P. Legendre. A FORTRAN program (source code and compiled versions for Macintosh and DOS) carrying out the computations of the fourth-corner statistics, with documentation, is available from P. Legendre (http://alize.ere.umontreal.ca/~casgrain/R/4th_corner/ or ftp://ftp.umontreal.ca/pub/casgrain/R/4th_corner/). The program can easily be adapted to other operating systems for which a FORTRAN compiler is available. The above-mentioned World Wide Web site also makes available the data set upon which Table 1 is based.

LITERATURE CITED

- Bartlett, M. S. 1960. Stochastic population models in ecology and epidemiology. Methuen, London, UK.
- Bray, R. J., and J. T. Curtis. 1957. An ordination of the upland forest communities of southern Wisconsin. *Ecological Monographs* **27**:325–349.
- 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.
- Edgington, E. S. 1995. Randomization tests. Third edition. Marcel Dekker, New York, New York, USA.
- Fager, E. W. 1963. Communities of organisms. Pages 415–437 in M. N. Hill, editor. *The sea*. Volume Two. Interscience, New York, New York, USA.
- Galzin, R. 1987a. Structure of fish communities of French Polynesian coral reefs. I. Spatial scales. *Marine Ecology Progress Series* **41**:129–136.
- . 1987b. Structure of fish communities of French Polynesian coral reefs. II. Temporal scales. *Marine Ecology Progress Series* **41**:137–145.
- Galzin, R., and P. Legendre. 1987. The fish communities of a coral reef transect. *Pacific Science* **41**:158–165.
- Galzin, R., and J.-P. Pointier. 1985. Moorea island, Society Archipelago. Pages 73–102 in *Proceedings of the Fifth International Coral Reef Congress, Tahiti*. Volume One: French Polynesian Coral Reefs.
- Gause, G. F. 1935. Vérification expérimentale de la théorie mathématique de la lutte pour la vie. *Actualités Scientifiques et Industrielles* Number 277. Hermann, Paris, France.
- Harmelin-Vivien, M. L. 1979. Ichtyofaune des récifs coralliens de Tuléar (Madagascar): écologie et relations trophiques. Thèse de Doctorat des Sciences, Université Aix-Marseille II, France.
- . 1989. Reef fish community structure: an Indo-Pacific comparison. Pages 21–60 in M. L. Harmelin-Vivien and F. Bourlière, editors. *Vertebrates in complex tropical systems*. Ecological Studies 69, Springer-Verlag, Berlin, Germany.
- Hiatt, R. W., and D. W. Strasburg. 1960. Ecological relationships of the fish fauna on coral reefs of the Marshall Islands. *Ecological Monographs* **30**:65–127.
- Hobson, E. S. 1974. Feeding relationships of teleostean fishes on coral reefs in Kona, Hawaii. *Fisheries Bulletin* **72**:915–1031.
- Hochberg, Y. 1988. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* **75**:800–802.
- Holm, S. 1979. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* **6**:65–70.
- Hommel, G. 1988. A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika* **75**:383–386.
- Hope, A. C. A. 1968. A simplified Monte Carlo significance test procedure. *Journal of the Royal Statistical Society Series B* **30**:582–598.
- Hutchinson, G. E. 1957. Concluding remarks. Cold Spring Harbor Symposium of Quantitative Biology **22**:415–427.
- Jackson, D. A., and K. M. Somers. 1989. Are probability estimates from the permutation model of Mantel's test stable? *Canadian Journal of Zoology* **67**:766–769.
- Kendall, M. G. 1948. Rank correlation methods. Charles Griffin, London, England.
- Legendre, L., and P. Legendre. 1978. Associations. Pages 261–272 in *Phytoplankton manual*. Monographs on oceanographic methodology, Number Six. UNESCO, Paris.
- Legendre, L., and P. Legendre. 1983. Numerical ecology. Elsevier Scientific, Amsterdam, The Netherlands.
- Legendre, P., N. L. Oden, R. R. Sokal, A. Vaudor, and J. Kim. 1990. Approximate analysis of variance of spatially autocorrelated regional data. *Journal of Classification* **7**:53–75.
- Manly, B. F. J. 1991. Randomization and Monte Carlo methods in biology. Chapman and Hall, New York, New York, USA.
- Naim, O. 1988. Distributional patterns of mobile fauna associated with *Halimeda* in the Tiahura coral-reef complex (Moorea, French Polynesia). *Coral Reefs* **6**:237–250.
- Payri, C. E. 1987. Zonation and seasonal variation of the commonest algae on Tiahura Reef (Moorea island, French Polynesia). *Botanica Marina* **30**:141–149.
- Sale, P. F. 1978. Coexistence of coral reef fishes—a lottery for living space. *Environmental Biology of Fishes* **3**:85–102.
- Sokal, R. R., and F. J. Rohlf. 1995. Biometry, Third edition. W. H. Freeman, New York, New York, USA.
- ter Braak, C. J. F. 1985. Correspondence analysis of incidence and abundance data: properties in terms of a unimodal response model. *Biometrics* **41**:859–873.
- . 1987. Ordination. Pages 91–173 in R. H. G. Jongman, C. J. F. ter Braak, and O. F. R. van Tongeren, editors. *Data analysis in community and landscape ecology*. Center for Publishing and Documentation (PUDOC), Wageningen, The Netherlands.
- . 1990. Update notes: CANOCO version 3.10. Agricultural Mathematics Group, Wageningen, The Netherlands.
- Tilman, D. 1987. The importance of the mechanisms of interspecific competition. *American Naturalist* **129**:769–774.

- Watanabe, J. M. 1984. The influence of recruitment, competition and benthic predation on spatial distributions of three species of kelp forest gastropods (Trochidae: *Tegula*). *Ecology* **65**:920–936.
- Whittaker, R. H. 1956. Vegetation of the Great Smoky Mountains. *Ecological Monographs* **26**:1–80.
- Wright, S. P. 1992. Adjusted *P*-values for simultaneous inference. *Biometrics* **48**:1005–1013.