

## Environmentally constrained null models: site suitability as occupancy criterion

Pedro R. Peres-Neto, Julian D. Olden and Donald A. Jackson

Peres-Neto, P. R., Olden, J. D. and Jackson, D. A. 2001. Environmentally constrained null models: site suitability as occupancy criterion. – *Oikos* 93: 110–120.

Null models have proven to be an important quantitative tool in the search for ecological processes driving local diversity and species distribution. However, there remains an important concern that different processes, such as environmental conditions and biotic interactions may produce similar patterns in species distributions. In this paper we present an analytical protocol for incorporating habitat suitability as an occupancy criterion in null models. Our approach involves modeling species presence or absence as a function of environmental conditions, and using the estimated site-specific probabilities of occurrence as the likelihood of species occupancy of a site during the generation of “null communities”. We validated this approach by showing that type I error is not affected by the use of probabilities as a site occupancy criterion and is robust against a variety of predictive performances of the species-environmental models. We describe the expected differences when contrasting classical and the environmentally constrained null models, and illustrate our approach with a data set of Dutch dune hunting spider assemblages. An environmentally constrained approach to null models will provide a more robust evaluation of species associations by facilitating the distinction between mutually exclusive processes that may shape species distributions and community assembly.

*P. R. Peres-Neto, J. D. Olden and D. A. Jackson, Dept of Zoology, Univ. of Toronto, Toronto, ON, Canada M5S 3G5 (pperes@zoo.utoronto.ca) (present address of JDO: Dept of Biology, Colorado State Univ., Fort Collins, CO 80523-1878, USA).*

Null models have been used widely to investigate patterns in species distributions and to identify possible mechanisms (e.g., competition, predation) or assembly rules (e.g., limiting similarity, species saturation) restricting local community membership to subsets of the regional pool of potential colonizers (e.g., Caswell 1976, Connor and Simberloff 1979, Jackson et al. 1992, Cook and Quinn 1995, Gotelli and Graves 1996, Weiher and Keddy 1999). Null models constitute a body of statistical methods to assess whether observed patterns in species distributions are distinct from arrangements of species taken at random from the regional pool. Conceptually, they are seen as quantitative tools for uncovering and testing conspicuous patterns in data rather than to elucidate the causal mechanisms responsible for such ecological patterns. For instance, nonrandom patterns in field data can serve as initial evidence for the

operation of particular ecological mechanisms shaping communities, which can subsequently be assessed experimentally (e.g., Werner 1984, Juliano and Lawton 1990). Given that experimental data alone cannot address the extent to which particular local mechanisms are influential at macroecological scales (Maurer 1999), null models and experimentation are regarded as complementary tools in the search for mechanisms structuring ecological communities.

Null models of species distributions generally involve data of species incidence (i.e., presence or absence) across a number of patches, sites, local communities or islands. Frequencies of species co-occurrence are compared to expectations based on random site occupation to determine whether species exhibit non-random patterns in their distribution (Gotelli 2000). However, evidence for non-random patterns in species distributions

Accepted 13 November 2000

Copyright © OIKOS 2001

ISSN 0030-1299

Printed in Ireland – all rights reserved

does not necessarily imply the role of biotic interactions, but could be equally related to other factors such as similarities or differences in dispersal abilities or environmental requirements of the species. In such cases, different processes like competition and environmental suitability could both lead to similar conclusions regarding patterns in species co-occurrences (i.e., rejection of the null hypothesis; see Schluter 1984, Bradley and Bradley 1985 for discussions). We argue that our treatment of the biotic and abiotic factors as independent components is a practical one in the sense that patterns in species distribution may be more conspicuous when these components are independent (but see Brown et al. 2000, Stone et al. 2000). To maximize the chance of distinguishing among competing hypotheses, the likelihood of species co-occurrences should be assessed again after the environment is factored out as a possible explanation for the patterns encountered. The idea that ecological requirements of species should be incorporated into null models has been long recognized, yet is rarely addressed in the literature (Zobel 1997). Approaches to control for species habitat requirements in null models have taken different forms. The simplest approach fixes the total number of species per site and species frequencies (see Gotelli 2000) when generating the “null communities” in order to account for differences in resource availability among sites and species-related characteristics. However, since species richness should vary with resource availability, this approach has been debated. Diamond and Marshall (1977) attempted to remove the effects of habitat diversity on species co-occurrences by using residuals from probit analysis to factor out differences in island sizes. Kelt et al. (1995) and Gotelli et al. (1997) incorporated geographic and habitat characteristics directly into null models during the generation of “null communities” by assigning species to sites where they should be able to disperse naturally and persist.

In general, attempts made to incorporate environmental constraints into null models have been based on environmental classifications where habitats are grouped into categories (e.g., forest versus woodland, low versus high altitude). The fact that habitats cannot always be easily classified into discrete units (e.g., Knight and Morris 1996, Dufrene and Legendre 1997), promoting subjective classifications, and that species are more likely to exhibit a much finer response to habitat conditions, emphasizes the need to apply more quantitative approaches for incorporating habitat suitability into null models. The objective of our study is to present a novel approach for incorporating environmental constraints into null models. This approach involves modeling species presence or absence as a function of a set of environmental factors, and using the estimated site-specific probabilities of occurrence as the likelihood of species occupancy of a site during the generation of “null communities”. After detailing the

protocol of the environmentally constrained null model approach, we conducted a simulation study to validate the model and described the expectations for the model. As an example, we used the approach with a data set of hunting spider assemblages from a Dutch dune area.

### **Unconstrained null models and co-occurrence test statistics**

A presence-absence matrix is the basis for the analysis of species distribution null models, where rows represent species and columns represent sites or samples. Each matrix cell is coded (1) for presence and (0) for absence. To test whether the matrix contains non-random patterns of species co-occurrences, a randomization test is used. The test begins by choosing a test statistic that reflects the question of interest and calculating the measure for the original data. Next the observed test statistic is contrasted against a null distribution that is generated by randomly allocating the incidence values (1/0) in the matrix a large number of times (i.e., generate “null communities”) and calculating the test statistic for each randomized set. Under the null hypothesis the observed test statistic is just one possible value from the null distribution and its likelihood can be evaluated as the proportion of randomized values that are more extreme than the observed.

Perhaps the main source of disagreement among ecologists is the protocol that is used to generate the “null communities” (see Gotelli and Graves 1996 for a review). Although no consensus has been reached on the most appropriate protocol, ecologists have commonly employed two algorithms for generating random matrices. The first randomizes the incidence values (1/0) fixing the sum of rows (i.e., species occurrences are maintained constant), whereas the second fixes the sum of both rows and columns (i.e., species occurrences and site richness are maintained constant). Depending on the algorithm, sites have different probabilities of being “randomly colonized”; however all sites are assumed to provide similar environmental conditions because any of them could be successfully colonized by any species under chance alone. We refer to these approaches as unconstrained null models. Gotelli (2000) found that both provide appropriate type I error rates and comparable statistical power.

There are many indices available to summarize patterns in species distribution (Jason and Vegelius 1981, Jackson et al. 1989) and we have chosen three that reflect different types of possible associations between species. (1) The C-score statistic (Stone and Roberts 1992, Gotelli 2000) calculates the number of checkerboard units for all species-pair combinations (i.e., the number of sites for which species A is present and species B is absent and vice versa). Increasing C-scores

indicate an increasing degree of mutual exclusivity between species, where the maximum is reached when half of the sites are occupied by one species and the other occupies the other half. (2) The T-score statistic (Stone and Roberts 1992) calculates the degree of togetherness by counting the number of sites that species A and B jointly are either present or absent. A high T-score indicates common occurrence between species, where the maximum is achieved when half of the sites are occupied by both species and the other half lacks both species. (3) The S-score (Stone and Roberts 1992) measures the number of shared sites occupied by both species A and B and is at its maximum when species occupy all sites. When conducting an analysis of the whole incidence matrix, C-, T- and S-scores are calculated as the average from all species pairs.

### Environmentally constrained null models

Figure 1 depicts the protocol for our environmentally constrained null model where two hypothetical species A and B were modeled according to two environmental

variables. The first step involves a single-species approach to estimate the probability of species presence or absence at each study site based on the environmental or habitat conditions. The resulting matrix contains site occurrence probabilities for each species at each site (i.e., site-specific probability matrix). A high predicted probability indicates that a site contains suitable environmental conditions for species occurrence, whereas a low probability suggests the lack of suitable environmental conditions. Predictions of site occurrence can be generated using a number of qualitative or quantitative approaches. For example, historical data could be used to assign probabilities of species colonization or information from experimental studies and field measures of habitat use could be used independently or combined to generate probabilities. An array of classification techniques such as discriminant analysis, logistic regression, classification trees, artificial neural networks or genetic algorithms is also available (see Hand 1997). In our hypothetical example (Fig. 1), we have used linear discriminant analysis to generate the species-specific probability matrix.

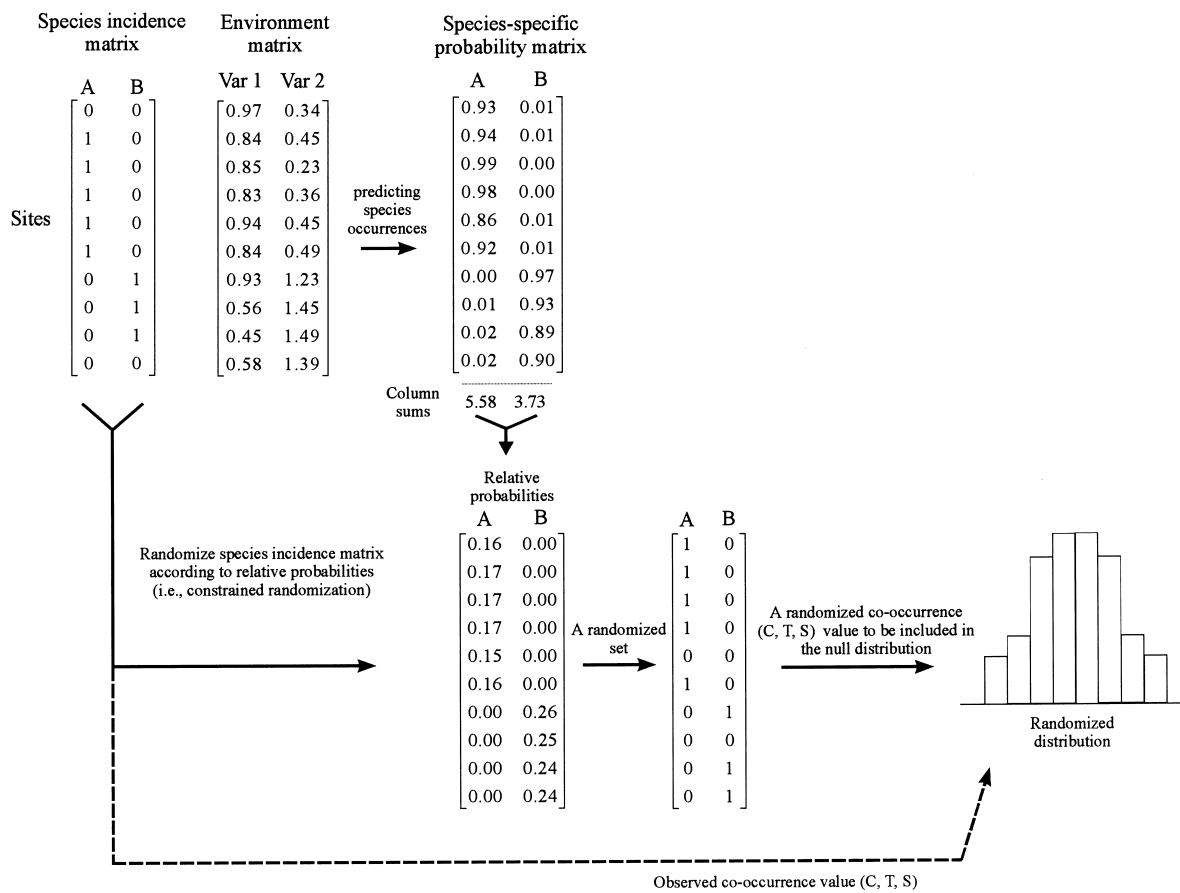


Fig. 1. Diagrammatic summary of the environmental constrained null model. Note that species are in columns and sites are in rows.

In the second step, probabilities of occurrence are transformed into relative probabilities, and species presences are reassigned to sites during the generation of “null communities” according to these relative values (Fig. 1). We refer to this algorithm as Ct-RA1. In addition we consider situations where the probability of species occurrence is low (i.e., smaller than 0.5), although the species is observed present (i.e., poor prediction by the model). Using these probabilities during the generation of “null communities” would identify these sites as unfavorable although the species is present. Therefore, we propose a second algorithm Ct-RA2 that assigns a probability of 1.0 to all sites in the species-specific probability matrix where the species is actually present in the observed incidence matrix. Then proceed as for Ct-RA1, where probabilities are transformed into relative probabilities (Fig. 1). Therefore, for Ct-RA2 only the probabilities from the species-habitat models for empty sites are used. In both protocols we maintain fixed species frequencies, so that the number of sites occupied by any species in each random matrix is the same as in the observed matrix (Fig. 1). The constrained randomization approach is then repeated a large number of times and at each time the index measuring species co-occurrence ( $C_{\text{rnd}}$ ,  $T_{\text{rnd}}$  or  $S_{\text{rnd}}$ ) is calculated and recorded. The observed value ( $C_{\text{obs}}$ ,  $T_{\text{obs}}$  or  $S_{\text{obs}}$ ) is then contrasted to the null distribution of random values and the probability of rejection is estimated (Fig. 1). The test is one-tailed and the probability is calculated as:  $(\text{number of } C_{\text{rnd}}, T_{\text{rnd}} \text{ or } S_{\text{rnd}} \text{ equal to or larger than } C_{\text{obs}}, T_{\text{obs}} \text{ or } S_{\text{obs}} + 1) / (\text{number of randomizations} + 1)$ , where 1 represents the observed value for the index being evaluated and is also included as a value of the randomized distribution. For all analyses we used 1999 random permutations.

### Validation of environmentally constrained null models

Although the above protocol is ecologically sound, there is a need to verify whether the use of site-specific probabilities in null models could inflate or deflate the expected type I error rate (Roxburgh and Matsuki 1999, Peres-Neto and Olden 2001). To address this question we designed a Monte Carlo experiment (Manly 1997, Peres-Neto and Marques 2000) to validate empirically our constrained randomization algorithms. A sample-based Monte Carlo approach was employed because our interest relates to null models for distribution matrices that represent samples from a larger universe rather than complete censuses of species in island archipelagos.

The experiment involved generating a large number of random species incidence matrices. Empirical estimations of type I error were calculated as the proportion

of random matrices that provided significant tests based on the null model. We began the experiment by generating an empty archipelago of 10 species and 20000 sites with 0s filling the whole matrix. Next, we generated a site-specific probability matrix (hereafter called SSP1) with the following data structure. For species 1, 90% of the sites received a high probability of occurrence based on a random value between 0.7 and 1.0 from a uniform distribution, and the remaining 10% of the sites received a low probability value ranging randomly between 0.0 and 0.3, again from a uniform distribution. For species 2, 80% of the sites received a probability value between 0.7 and 1.0, and the other 20% received a value between 0.0 and 0.3. The pattern continued with a 10% sequential decrease for all species until species 10, for which only 5% of the sites were assigned high probability values. Considering each species separately, the occupancy of any particular site in the empty archipelago was based on SSP1. This scenario mimics a situation where species distributions are related strongly to environmental suitability but their co-occurrences are random with respect with one another. Therefore, the outcome of any model based on species-environment relationship for this system would generate good performance, but the environmentally constrained null models should detect no association between species. To ensure that type I error rates were not influenced by the predictive success of the species-environmental models, we constructed three additional site-specific probability matrices based on modifications of SSP1. SSP2 randomly assigned new probabilities  $1 - P$  to 20% of the sites for each species where habitat suitability was small ( $< 0.30$ ) and the species was truly absent (i.e., good model prediction), where  $P$  is the original probability for the same site and species in SSP1. In addition, in SSP2, 20% of sites for each species where habitat suitability was high ( $> 0.70$ ) and the species was truly present (i.e., good model prediction) had their probabilities converted into  $1 - P$ . SSP3 randomly assigned new probabilities  $1 - P$  to 20% of the sites across all species independent of the classification success of the model. SSP4 randomly assigned new probabilities of 0.5 to 20% of the sites for each species.

Type I error rates for each simulated scenario (i.e., SSP1–SSP4) were estimated using 1000 random samples of 100 sites each from the archipelago. Each sample was constituted using a species incidence matrix and the corresponding (i.e., same sites) site-probability matrix based on to one of the four scenarios. Type I error rates were estimated for an unconstrained null model (fixed sum of species totals), Ct-RA1 and Ct-RA2 for C-, T- and S-scores, as the number of sample tests out of 1000 that were significant according to three alpha values (0.1, 0.05, 0.01). The results (Table 1) indicate that the environmental constrained null models are not prone to elevated or deflated type I error rates, even when the species-environment model exhibited poor

Table 1. Summary of type I error estimates for unconstrained and constrained null models. Each entry is the proportion of tests for which the null hypothesis was rejected according to a specific alpha level when tested against a sample from a random incidence matrix. See text for details on the simulation protocols and null models used.

	Unconstrained null models	Environmentally constrained null models							
		Cr-RA1				Cr-RA2			
		Species-specific site probability matrix				Species-specific site probability matrix			
Alpha = 0.10		1	2	3	4	1	2	3	4
C-score	0.108	0.115	0.112	0.109	0.127	0.093	0.086	0.075	0.107
T-score	0.106	0.135	0.110	0.108	0.132	0.100	0.089	0.085	0.108
S-score	0.093	0.126	0.099	0.090	0.128	0.091	0.085	0.096	0.098
Alpha = 0.05									
C-score	0.057	0.060	0.061	0.042	0.071	0.047	0.048	0.038	0.050
T-score	0.044	0.068	0.056	0.059	0.079	0.049	0.042	0.042	0.054
S-score	0.041	0.058	0.050	0.055	0.07	0.045	0.039	0.050	0.048
Alpha = 0.01									
C-score	0.017	0.012	0.011	0.003	0.015	0.009	0.010	0.010	0.013
T-score	0.009	0.012	0.009	0.015	0.020	0.009	0.010	0.006	0.012
S-score	0.007	0.012	0.008	0.017	0.019	0.007	0.008	0.007	0.010

correct classification success. We did not compare power since the unconstrained and constrained null models assess the likelihood of different alternative hypotheses. However, note that type I error comparisons are justifiable as the null hypothesis is the same in both cases (i.e., absence of species association).

### Environmentally constrained null models – expectations

We anticipate three qualitative differences in interpretations when contrasting the outcomes of unconstrained versus environmentally constrained null models. In Fig. 2 we provide a hypothetical example illustrating these expectations. Such expectations are due to the unconstrained null model having a greater proportion of sites occupied as a result of chance alone relative to the constrained versions.

*Expectation 1* – The unconstrained null model detects a significant negative association between species (i.e., high C-score), whereas the constrained null model is non-significant (Fig. 2). In this case, negative associations between species may be driven by the fact that they have different environmental requirements, so that biotic interactions between species are non-existent when species-environment relationships are taken into consideration.

*Expectation 2* – The unconstrained null model detects a significant positive association between species (i.e., high T- and S-scores), whereas the constrained null model is non-significant (Fig. 2). Here, positive associa-

tions between species can be explained by common species-environment affinities and not by biotic facilitation or similar dispersal capacities among species (Kelt et al. 1995). In these cases, species distributions overlapping in a relatively small fraction of sites should show the greatest differences between constrained and unconstrained null models. In contrast, species with greater frequency of occurrence in the data and exhibiting positive associations may show no difference when environment is taken into account because the site-specific probabilities will be closer to the actual percent of occurrence in the data set. Also, if one assumes that species with broad distributions are habitat generalists, environment should not be playing an important role in their distributions, so that in these cases both types of null models will likely provide similar outcomes.

*Expectation 3* – The unconstrained null model is non-significant, whereas the constrained null model detects a significant negative association between species (Fig. 2). In such cases, the potential for negative interactions exists only at sites where the environment conditions are suitable for the species involved, especially when they are closely related (e.g., guilds, congeneric species). Greater power of detecting patterns having negative associations can be achieved using the constrained approach because decreasing the number of suitable sites for species occupation increases the chance of finding checkerboard units under the permutation model. Similarly, Wilson and Gitay (1995) employing a null model where plants were re-assigned randomly only to quadrats near to their original location, found that decreasing the number of quadrats available increased the chances of detecting competitive interactions. Their

rationale was that nearby quadrats should offer more similar environments than more distant ones.

We do not expect cases where the unconstrained null model is non-significant, but the constrained null model detects a positive association among species. Under the constrained permutation, a smaller number of sites is available to species and they will tend to co-occur more often under chance alone, decreasing the probability of rejection of positive associations. We also do not expect the case where the unconstrained null model detects a positive association between species and the constrained null model detects a negative association between species, or vice versa (but see Schoener and Adler 1991), because in our constrained null model only the probability of the test statistic is affected (i.e., the observed test statistic is not changed). Cases where both constrained and unconstrained null models agree in their outcomes imply that species interactions cannot be explained by environment alone. In the case of positive interactions, species occupying a smaller fraction of the available habitats, but overlapping highly in their distribution, may suggest facilitation. As discussed above in expectation 2, ubiquitous species also may not exhibit differences between the two types of null models. There is also the case where species may remain positively associated when they only overlap in fewer sites that offer a unique combination of environmental conditions, but not others.

### Environmentally constrained null model in a community of hunting spiders

Data were taken from ter Braak (1986: Table 3) and comprise the abundance values of 12 species of hunting spiders and environmental data for six habitat variables

from 28 sites (Table 2), originally presented by Van der Aart and Smeek-Enserink (1975). We constructed species-habitat models by modeling species presence-absence as a function of six environmental characteristics using linear discriminant analysis. The results from the discriminant analyses produced a site-by-species matrix containing probability estimates for species occurrence of each site. A jackknife procedure (also called n-fold or leave-one-out cross validation) was used to validate the models because this approach provides a nearly unbiased estimate of prediction success (Olden and Jackson 2000). The jackknife approach excludes one site, constructs the models with the remaining 27 sites, and then predicts the probability of species occurrences for the excluded site using this model. This procedure is repeated 28 times so that each site, in turn, is excluded during the model construction and its response is predicted. Overall correct classification rate was calculated as the percentage of sites where the model correctly predicted the presence or absence of a species and Cohen's Kappa statistic was used to assess whether the model predictions differed from expectations based on chance alone (Titus et al. 1984). All single-species models were found to be statistically significant (Table 3). To acquire a basic description of species-environmental affinities, we performed a principal component analysis (PCA) on correlations of the sites-by-species probability matrix so that each species could be compared in the reduced ordination space (Fig. 3). Arrows represent correlations between environmental variables and principal components. Our results largely agree with those presented by ter Braak (1986: Fig. 1) based on a canonical correspondence analysis.

Based on the entire incidence matrix, the unconstrained null model and the two environmentally constrained null models showed significant positive associations between species (C-score = 30.2,  $P = 1.000$

Fig. 2. Expectations under the environmentally constrained null models compared to unconstrained null models. For the sake of illustration, values in the probability matrix were either 0 or 1. Species are in rows and sites are in columns. Reported values include C-, T- and S-scores and their associated probabilities. See text for explanations regarding differences between environmentally constrained and unconstrained null models.

Incidence matrix	Unconstrained null model	Probability matrix	Constrained null model
<p>Expectation 1</p> $\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$	<p>Negative</p> <p>C=25.0 (0.006) T=0.0 (1.000) S=0.0 (1.000)</p>	$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$	<p>Non-interactive</p> <p>C=25.0 (1.000) T=0.0 (1.000) S=0.0 (1.000)</p>
<p>Expectation 2</p> $\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$	<p>Positive</p> <p>C=0.0 (1.000) T=25.0 (0.003) S=5.0 (0.003)</p>	$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$	<p>Non-interactive</p> <p>C=0.0 (1.000) T=25.0 (1.000) S=5.0 (1.000)</p>
<p>Expectation 3</p> $\begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}$	<p>Non-interactive</p> <p>C=12.0 (0.137) T=0.0 (1.000) S=0.0 (1.000)</p>	$\begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \end{bmatrix}$	<p>Negative</p> <p>C=12.0 (0.034) T=0.0 (1.000) S=0.0 (1.000)</p>

Table 2. Incidence matrix (presence-absence) for the hunting spider data. Species are in rows, sites are in columns and environmental variables are categorized into 10 equal-sized classes (see ter Braak 1986: Table 3). Note that sites and species are ordered to maximize the correspondence between patterns in species incidence and environmental variables.

Species	Sites																														
	15	19	20	16	17	18	2	8	21	5	6	14	4	7	13	3	1	9	12	25	11	10	28	23	22	27	24	26			
<i>Arctosa lutetiana</i> (AL)	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0			
<i>Pardosa lugubris</i> (PL)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	0	0	0	1	0	0	0	0			
<i>Zora spinimana</i> (ZS)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0			
<i>Pardosa nigriceps</i> (PN)	0	1	0	1	0	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0			
<i>Pardosa pullata</i> (PP)	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	0	0	0	0	0	0	0			
<i>Autonia albimana</i> (AA)	0	0	0	0	0	0	1	1	0	1	1	1	1	1	1	1	1	0	0	1	1	0	0	0	0	0	0	0			
<i>Trochosa terricola</i> (TT)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			
<i>Alopecosa cuneata</i> (AC)	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0			
<i>Pardosa monticola</i> (PM)	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0			
<i>Alopecosa accentuata</i> (AAc)	0	0	0	0	0	0	0	0	0	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			
<i>Alopecosa fabrilis</i> (AF)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1	1	1	1	1	1	1	1	1			
<i>Arctosa pericola</i> (AP)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1			
Environmental variables																															
Water content	9	7	8	8	9	8	8	6	7	8	9	8	6	8	9	6	5	5	5	3	4	4	0	0	0	1	0	2	0		
Bare sand	0	0	0	0	0	0	0	0	0	0	5	0	0	0	3	0	0	0	0	7	0	8	7	6	7	5	7	9	0		
Cover moss	1	3	1	1	1	0	2	2	1	0	5	4	5	1	1	5	7	9	8	2	9	7	8	9	9	8	9	4	0		
Light reflection	1	0	0	0	2	2	3	1	0	5	1	2	6	5	7	8	8	7	8	5	8	8	8	9	8	8	9	9	0		
Fallen twigs	9	9	9	9	9	3	9	9	9	0	7	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
Cover herbs	5	2	0	0	5	5	9	6	2	9	6	9	9	9	9	9	6	8	8	8	7	5	6	6	0	6	5	2	0		

Table 3. Results from discriminant analysis for predicting species presence-absence of hunting spider species based on environmental factors (see Table 2). Reported values are percentage of species occurrence (%SO), percentage of sites that the species was correctly classified (%CC), Kappa statistic and the associated  $p$ -value ( $P$ ). Species codes follow Table 2.

Species	%SO	%CC	Kappa	$P$
AL	25.0	92.9	0.810	0.001
PL	60.7	96.4	0.926	0.001
ZS	60.7	92.9	0.850	0.001
PN	53.6	82.1	0.643	0.001
PP	46.4	92.9	0.855	0.001
AA	42.9	89.3	0.779	0.001
TT	92.9	96.4	0.650	0.007
AC	67.8	92.9	0.826	0.001
PM	75.0	96.4	0.909	0.001
AAc	60.7	96.4	0.926	0.001
AF	39.3	92.9	0.845	0.001
AP	21.4	92.9	0.818	0.001

for all null models; T-score = 54.8,  $P < 0.0005$  for all null models; S-score = 8.9,  $P < 0.0005$  for all null models). Examining species-pair associations showed that the number of significant interactions (both positive and negative) was much smaller for the environmentally constrained null model compared to the unconstrained (Table 4). Although there were a large number of negative interactions (Tables 4, 5), the models based on the complete incidence matrix were not effective in detecting these associations, suggesting that when there are both positive and negative interactions between species, these indices may provide different power. Note that only three species had negative associations, whereas six species had positive associations (Table 5).

Table 4. Summary of the null models results for the hunting spider data. Number of rejections for species pair-wise comparisons for different significance levels is presented. For instance, for an alpha = 0.05, 13 species pairs were negatively associated (C-score) according to the unconstrained null model.

C-score	Unconstrained null models	Environmentally constrained null models	
		Cr-RA1	Cr-RA2
0.01	5	0	0
0.05	13	0	0
0.10	13	1	1
T-score			
0.01	15	1	1
0.05	18	4	4
0.10	20	10	11
S-score			
0.01	14	1	1
0.05	17	3	3
0.10	18	9	11

## Discussion

Our study describes a null-model protocol where species-environment associations can be accounted for when examining patterns in species incidence. The classical approach (i.e., unconstrained) assumes that the environment is homogeneous across the landscape and thus unimportant in shaping species distributions, whereas the constrained approach incorporates species-specific responses to the environment. Therefore, the constrained approach facilitates a better evaluation of

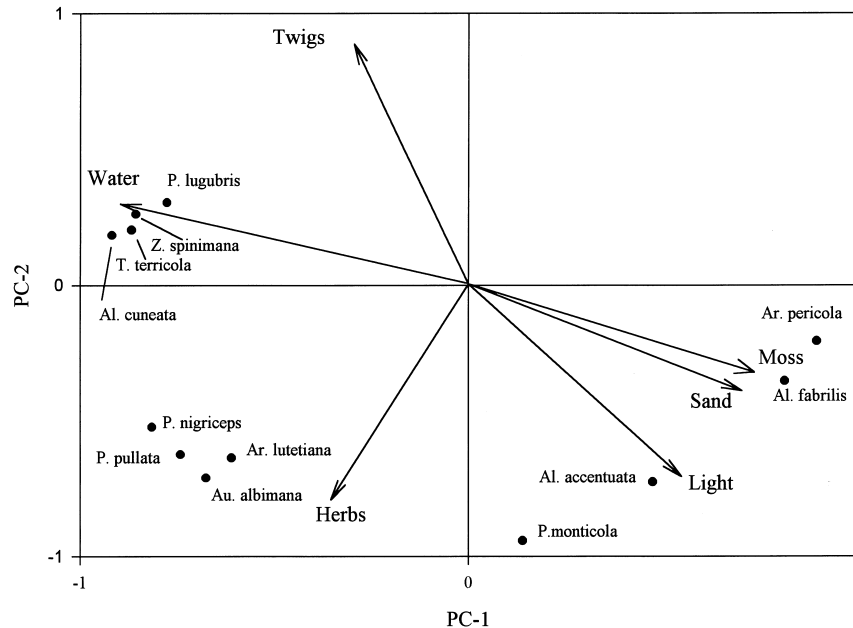


Fig. 3. Principal component plot of the site-by-species probability matrix resulting from linear discriminant functions for the hunting spider data (Table 2). Arrows represent the correlation between environmental variables and principal components.



the possible roles of biotic and abiotic factors shaping community structure. The comparison between classical null models and the constrained approach should provide the necessary contrast to judge which factor predominates. However, it is important to reiterate that the general value of null models based on distributional data is in verifying patterns related to similar or complementary distributions, rather than asserting that certain mechanisms are important or unimportant. For instance, because species that compete may be more likely to share similar habitats, positive associations are also expected under strong competition (Schluter 1984, Kelt and Brown 1999). The value of our analytical approach is not different in this regard; however, it addresses the question of whether or not associations between species may be simply ascribed to environmental affinities, rather than biotic interactions, and also tests whether species have more similar or distinct habitat requirements.

The results show that accounting for environmental suitability of the dune sites greatly influences the interpretation of interactions between hunting spider species as we found that differences or similarities in species environmental requirements largely described the patterns of association between species (Tables 4, 5). Consequently, this removed the need to invoke facilitation or competition as plausible mechanisms determining most species distributions. For instance, *A. pericola* and *P. lugrabis* are negatively associated based on the results from the unconstrained null model; however, they have very different environmental affinities for the amount of bare sand, cover moss and water content (Fig. 3), thus resulting in no significant association according to the constrained null model. In contrast, *A. albimana* and *A. lutetiana* are positively associated according to the unconstrained null model, but are randomly associated in the environmentally constrained null models because these species were found at sites with very similar environmental conditions (Fig. 3). Interestingly, *A. cuneata* and *P. nigriceps* remained

positively associated after the environment was incorporated (Table 5). Although these two species exhibit somewhat different environmental preferences (Fig. 3), they frequently overlap at sites having the combination of few fallen twigs but high light reflectivity (Table 2), suggesting an interaction of these two variables in facilitating coexistence. Associations that remain significant after accounting for species-environment relationships may be related to three aspects: (1) species associations are truly related to biotic interactions; (2) interactions between some environmental factors at particular sites might facilitate coexistence; and (3) important environmental variables not used in the species-environment models may contribute to their joint or disjoint distribution. Regardless, our null model will be effective in revealing those associations that can be explained by the measured environmental variables from those that cannot be explained.

A number of conditions may decouple species-habitat relationships and introduce errors into species-environmental models. For instance, some sites may provide environmental conditions suitable for persistence; however, dispersal barriers may impede immigration into the site (e.g., Lonzarich et al. 1998). Competitive interactions may also displace species from their optimum into sub-optimal habitats in a number of sites. An interesting extension to the constrained approach would be to sample site-specific probabilities within confidence intervals for site probabilities (see Taylor 1991) at each randomization so that the degree of error would be also incorporated into the null model. In the worst-case scenario, where species models present extremely high degrees of errors, the unconstrained and constrained null models would provide equivalent outcomes (Peres-Neto unpubl.).

By fixing species frequencies in our null models, we assumed that the probability of random colonization of sites is a function of the fraction of sites occupied. However, species frequencies are also an important outcome of competitive interactions, and perhaps may

Table 5. Pair-wise associations between hunting spiders. Positive associations (+) were assessed by the significance of the T-score (note: S-score provided similar results). Negative associations (-) were judged by the significance of the C-score. All results based on alpha = 0.05. The upper diagonal contains the results based on the unconstrained null model, whereas the lower diagonal has the results for Cr-RA1 (note: Cr-RA2 provided similar results). Species codes follow Table 2.

	AL	PL	ZS	PN	PP	AA	TT	AC	PM	AAc	AF	AP
AL			+	+	+	+		+				
PL								+		-	-	-
ZS				+	+	+		+		-	-	-
PN					+	+		+			-	-
PP						+		+	+			-
AA								+	+			-
TT												-
AC				+	+	+					-	-
PM												
AAc												
AF												
AP												

also reflect positive associations (Gotelli 2000 and references therein). As a consequence, fixing species frequencies may incorporate the outcomes from previous species interactions into the “null communities”, lowering the power of detecting true associations. Relaxing this assumption may provide important insights. For instance, some components of metapopulation or biogeographic models may be adapted and incorporated into constrained null models for predicting site occupancy, as well as generating random site occupancy via simple stochastic processes (e.g., Haydon et al. 1993, Hanski 1994). The use of more complex species-environmental models for predicting species incidence may be modified to incorporate aspects related to spatial location of sites (Roxburgh and Matsuki 1999) such as isolation, connectivity and corridor quality and species dispersal abilities into null models.

Our study has detailed methods for incorporating site-specific probabilities into null models so that species selected to compose “null communities” could occupy the environment offered by randomly selected sites under normal conditions rather than occurrences simply being mediated by species interactions. Given the vast number of data sets containing data on species distributions and site environmental conditions, the method presented here should facilitate a more robust evaluation of the factors contributing to species associations and community organization.

*Acknowledgements* – We would like to thank Nicholas Gotelli for stimulating discussions regarding null models and helpful comments on this paper and Bryan Manly for his insights regarding Monte Carlo simulations for type I error estimation. Funding for this project was provided by a CNPq Fellowship to PRP-N, a Natural Sciences and Engineering Research Council of Canada (NSERC) Graduate Scholarship to JDO and an NSERC Research grant to DAJ. A computer program to carry out the null model protocols presented here is available on < <http://www.zoo.utoronto.ca/jackson/software/> > .

## References

Bradley, R. A. and Bradley, D. W. 1985. Do non-random patterns of species in niche space imply competition? – *Oikos* 45: 443–446.

Brown, J. H., Fox, B. J. and Kelt, D. A. 2000. Assembly rules: desert rodent communities are structured at scales from local to continental. – *Am. Nat.* 156: 314–321.

Caswell, H. 1976. Community structure: a neutral model analysis. – *Ecol. Monogr.* 46: 327–354.

Connor, E. F. and Simberloff, D. 1979. The assembly of species communities: chance or competition? – *Ecology* 60: 1132–1140.

Cook, R. and Quinn, J. F. 1995. The importance of colonization in nested species subsets. – *Oecologia* 102: 413–424.

Diamond, J. M. and Marshall, A. G. 1977. Distributional ecology of New Hebridian birds: a species keleidoscope. – *J. Anim. Ecol.* 46: 703–727.

Dufrène, M. and Legendre, P. 1997. Species assemblages and indicator species: the need for a flexible asymmetrical approach. – *Ecol. Monogr.* 67: 345–366.

Gotelli, N. J. 2000. Null model analysis of species co-occurrence patterns. – *Ecology* 81: 2606–2621.

Gotelli, N. J. and Graves, G. R. 1996. Null models in ecology. – Smithsonian Inst. Press.

Gotelli, N. J., Buckley, N. J. and Wiens, J. A. 1997. Co-occurrence of Australian birds: Diamond’s assembly rules revisited. – *Oikos* 80: 311–324.

Hand, D. J. 1997. Construction and assessment of classification rules. – John Wiley and Sons.

Hanski, I. 1994. A practical model of metapopulation dynamics. – *J. Anim. Ecol.* 63: 151–162.

Haydon, D., Radtkey, R. R. and Pianka, E. R. 1993. Experimental biogeography: interactions between stochastic, historical, and ecological processes in a model archipelago. – In: Ricklefs, R. E. and Schluter, D. (eds), Species diversity in ecological communities – historical and geographical perspectives. Univ. of Chicago Press, pp. 117–130.

Jackson, D. A., Somers, K. M. and Harvey, H. H. 1989. Similarity coefficients: measures of co-occurrence and association or simply measures of occurrence. – *Am. Nat.* 133: 436–453.

Jackson, D. A., Somers, K. M. and Harvey, H. H. 1992. Null models and fish communities: evidence of nonrandom patterns. – *Am. Nat.* 139: 930–951.

Jason, S. and Vegelius, J. 1981. Measures of ecological association. – *Oecologia* 49: 371–376.

Juliano, S. A. and Lawton, J. H. 1990. The relationship between competition and morphology. II. Experiments on co-occurring dystiscid beetles. – *J. Anim. Ecol.* 59: 831–848.

Kelt, D. A. and Brown, J. H. 1999. Community structure and assembly rules: confronting conceptual and statistical issues with data on desert rodents. – In: Weiher, E. and Keddy, P. (eds), Ecological assembly rules: perspectives, advances, retreats. Cambridge Univ. Press, pp. 75–107.

Kelt, D. A., Taper, M. L. and Meserve, P. L. 1995. Assessing the impact of competition on community assembly: a case study using small mammals. – *Ecology* 76: 1283–1296.

Knight, T. W. and Morris, D. W. 1996. How many habitats do landscapes contain? – *Ecology* 77: 1756–1764.

Lonzarich, D. G., Warren, M. L. and Conzarich, M. E. 1998. Effects of habitat isolation on the recovery of fish assemblages in experimentally defaunated stream pools in Arkansas. – *Can. J. Fish. Aquat. Sci.* 55: 2141–2149.

Manly, B. J. F. 1997. Randomization, bootstrap and Monte Carlo methods in biology. – Chapman and Hall.

Maurer, B. A. 1999. Untangling ecological complexity: the macroscopic perspective. – Univ. of Chicago Press.

Olden, J. D. and Jackson, D. A. 2000. Torturing data for the sake of generality: how valid are our regression models. – *Écoscience* 7 (in press).

Peres-Neto, P. R. and Marques, F. 2000. When are random data not random, or is the PTP test useful? – *Cladistics* (in press).

Peres-Neto, P. R. and Olden, J. D. 2001. Assessing the robustness of randomization tests: examples from behavioural studies. – *Anim. Behav.* (in press).

Roxburgh, S. H. and Matsuki, M. 1999. The statistical validation of null models used in spatial association analysis. – *Oikos* 85: 68–78.

Schluter, D. A. 1984. A variance test for detecting species associations, with some example applications. – *Ecology* 65: 998–1005.

Schoener, T. W. and Adler, G. H. 1991. Greater resolution of distributional complementarities by controlling for habitat affinities: a study with Bahamian lizards and birds. – *Am. Nat.* 137: 669–692.

Stone, L. and Roberts, A. 1992. Competitive exclusion, or species aggregation? An aid in deciding. – *Oecologia* 91: 419–424.

Stone, L., Dayan, T. and Simberloff, D. 2000. On desert rodents, favored states, and unresolved issues: scaling up and down regional assemblages and local communities. – *Am. Nat.* 136: 322–328.

- Taylor, B. 1991. Investigating species incidence over habitat fragmentation of different areas – a look at error estimation. – In: Gilpin, M. and Hanski, I. (eds), *Metapopulation dynamics: empirical and theoretical investigations*. Academic Press, pp. 177–191.
- ter Braak, C. J. F. 1986. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. – *Ecology* 67: 1167–1179.
- Titus, K., Mosher, J. A. and Williams, B. K. 1984. Chance-corrected classification for use in discriminant analysis: ecological applications. – *Am. Midl. Nat.* 111: 1–7.
- Van der Aart, P. J. and Smeek-Enserink, N. 1975. Correlation between distributions of hunting spiders (Lycosidae, Ctenidae) and environmental characteristics in a dune area. – *Neth. J. Zool.* 25: 1–45.
- Weihner, E. and Keddy, P. 1999. Introduction: the scope and goals of research on assembly rules. – In: Weihner, E. and Keddy, P. (eds), *Ecological assembly rules: perspectives, advances, retreats*. Cambridge Univ. Press, pp. 1–20.
- Werner, E. E. 1984. The mechanisms of species interactions and community organization in fish. – In: Strong, D. R., Jr., Simberloff, D., Abele, L. G. and Thistle, A. B. (eds), *Ecological communities: conceptual issues and evidence*. Princeton Univ. Press, pp. 360–382.
- Wilson, J. B. and Gitay, H. 1995. Limitations to species coexistence: evidence for competition from field observations, using a patch model. – *J. Veg. Sci.* 6: 369–376.
- Zobel, M. 1997. The relative role of species pools in determining plant species richness: an alternative explanation of species coexistence? – *Trends Ecol. Evol.* 12: 266–269.