

# Null model analysis of species associations using abundance data

WERNER ULRICH<sup>1,3</sup> AND NICHOLAS J. GOTELLI<sup>2</sup>

<sup>1</sup>*Department of Animal Ecology, Nicolaus Copernicus University, Gagarina 9, 87-100 Toruń, Poland*

<sup>2</sup>*Department of Biology, University of Vermont, Burlington, Vermont 05405 USA*

**Abstract.** The influence of negative species interactions has dominated much of the literature on community assembly rules. Patterns of negative covariation among species are typically documented through null model analyses of binary presence/absence matrices in which rows designate species, columns designate sites, and the matrix entries indicate the presence (1) or absence (0) of a particular species in a particular site. However, the outcome of species interactions ultimately depends on population-level processes. Therefore, patterns of species segregation and aggregation might be more clearly expressed in abundance matrices, in which the matrix entries indicate the abundance or density of a species in a particular site.

We conducted a series of benchmark tests to evaluate the performance of 14 candidate null model algorithms and six covariation metrics that can be used with abundance matrices. We first created a series of random test matrices by sampling a metacommunity from a lognormal species abundance distribution. We also created a series of structured matrices by altering the random matrices to incorporate patterns of pairwise species segregation and aggregation. We next screened each algorithm–index combination with the random and structured matrices to determine which tests had low Type I error rates and good power for detecting segregated and aggregated species distributions. In our benchmark tests, the best-performing null model does not constrain species richness, but assigns individuals to matrix cells proportional to the observed row and column marginal distributions until, for each row and column, total abundances are reached.

Using this null model algorithm with a set of four covariance metrics, we tested for patterns of species segregation and aggregation in a collection of 149 empirical abundance matrices and 36 interaction matrices collated from published papers and posted data sets. More than 80% of the matrices were significantly segregated, which reinforces a previous meta-analysis of presence/absence matrices. However, using two of the metrics we detected a significant pattern of aggregation for plants and for the interaction matrices (which include plant–pollinator data sets). These results suggest that abundance matrices, analyzed with an appropriate null model, may be a powerful tool for quantifying patterns of species segregation and aggregation.

**Key words:** *abundance matrix; biogeography; co-occurrence; covariation; null model; passive sampling; statistical test.*

## INTRODUCTION

A major research focus in ecology has been the elucidation of community assembly rules, a set of mechanisms that lead to nonrandom patterns in multispecies assemblages (Weiher and Keddy 1999). For example, Diamond (1975) hypothesized that pairs of species that are close competitors may never coexist in the same local assemblage, leading to a biogeographic “checkerboard distribution.” Patterson and Atmar (1986) hypothesized that orderly extinction sequences in fragmented habitats will lead to a pattern of species “nestedness” (Ulrich et al. 2009). More recent assembly rules have been based on patterns of phylogenetic clustering or overdispersion (Webb et al. 2002, Emerson and Gilgelsie 2008), patterns of species interaction

networks (Bascompte and Jordano 2007), and the distribution of species-level morphological or physiological traits (Bellwood et al. 2002, Lavorel and Garnier 2002).

Assembly rules are rarely tested experimentally (e.g., Fukami and Morin 2003, Irving and Connell 2006) and are controversial because different mechanisms, including stochastic processes, may lead to the same community pattern (Gotelli 2004). Null models provide a statistical test for whether an observed pattern is likely in the absence of a particular mechanism (Gotelli and Graves 1996), and they have always figured prominently in the assembly rules literature (Williams 1964, Harvey et al. 1983). The data for null model analyses are typically in the form of a binary presence/absence matrix (McCoy and Heck 1987): rows are species, columns are sites or samples, and the entries indicate the absence (0) or presence (1) of a species. In interaction or food web matrices, both rows and columns may represent species, and the entries represent the absence (0) or presence (1)

Manuscript received 20 November 2009; revised 25 March 2010; accepted 30 March 2010. Corresponding Editor: M. Fortin.

<sup>3</sup> E-mail: ulrichw@umk.pl

of an interaction link, such as a trophic or pollinator interaction (Jordano et al. 2003).

Although assembly rule patterns can be quantified from presence/absence data, colonization or extinction is usually preceded by changes in population size. Colonization and extinction thus represent special cases of change in abundance. An abundance matrix contains rows as species, columns as sites or samples, and entries representing the population size (which may be 0) of a particular species in a particular site. Presence/absence matrices can always be constructed from abundance matrices, but not vice versa. The pattern of abundances across replicated assemblages may potentially contain a more complex and subtle signal of community assembly rules than binary presence/absence matrices. Because many assembly rules are based on the premise of species interactions (Weiher and Keddy 1999), an appropriate null hypothesis for abundance data might be that the covariances between pairs of species equal zero (Schluter 1984). However, testing this null hypothesis with simple and partial correlation analysis is problematic because of constraints on the magnitude of correlations and covariances (Brown et al. 2004).

Null models and randomization methods may be preferable for the analysis of abundance matrices because they are not as restrictive in their assumptions as standard parametric statistics (Manly 1991). However, until recently, relatively little attention has been paid to assembly rule patterns in abundance matrices (Graves and Gotelli 1993, Hausdorf and Hennig 2007, Lester et al. 2009).

In this paper, we explore null model analyses of abundance matrices. We present six candidate indices that may be used to describe the structure of abundance matrices, and we test each index with 14 potential null model algorithms that randomize the pattern in an observed abundance matrix. These benchmark tests were applied to a series of artificial matrices that were simulated from distributions with specified patterns of randomness and structure. Finally, following the example of several meta-analyses of binary presence/absence matrices (Gotelli and McCabe 2002, Blüthgen et al. 2007, Ulrich and Gotelli 2007a), we assembled from the literature and the Internet 147 empirical abundance matrices and 36 mutualistic interaction matrices and analyzed them with a subset of null models that performed best in our initial benchmark tests.

#### MATERIALS AND METHODS

##### *Strategies for evaluation of null model algorithms*

Before a new randomization test is applied to empirical data, its performance needs to be evaluated with artificial data sets that have specified amounts of randomness and structure (Gotelli 2001). Two properties are desirable in a statistical test. First, when the test is confronted with “random” matrices, it should not reject the null hypothesis too frequently, and a traditional Type I error criterion of 5% is usually

employed. Second, when the test is confronted with “structured” matrices, it should not accept the null hypothesis too frequently. The statistical power of the test is the probability of correctly rejecting the null hypothesis when it is false. There is no convention for power levels, but a value of 0.8 (the null hypothesis is correctly rejected 80% of the time) has been suggested (Cohen 1992). However, power analyses are rarely conducted in ecological studies (Toft and Shea 1983), perhaps because they require specification of an alternative hypothesis and an effect size that can be detected by the test.

In the context of null model analysis, what constitutes a “random” or a “structured” matrix? The primary goal in much null model analysis has been testing for the effects of species interactions on community patterns (Gotelli and Graves 1996). So, a “random” matrix would consist of repeated samples from an assemblage that is not structured by species interactions. Unfortunately, we can only speculate what such assemblages would look like (Colwell and Winkler 1984).

Three approaches have been used to create “random” binary matrices for the purposes of benchmarking the performance of null models. First, Gotelli (2000) began with an empirical presence/absence matrix and created a mixture of random matrices by uniformly reshuffling elements within the matrix, sometimes constraining the reshufflings within each row or column of the original matrix. The disadvantages of this method are that it is somewhat arbitrary, that it may include some of the same null model algorithms that are being tested, and that the results may be conditional on a particular matrix size, dimension, or percentage fill. A second approach is to specify a mechanistic colonization model that does not include species interactions, such as the neutral model (Bell 2005), and then use that model to create random matrices that can be used to evaluate null model procedures (Ulrich 2004). The disadvantage of this method is that the test is narrowly optimized for one particular mechanistic model, and there is no logical reason that this model should have priority. Moreover, even the simplest mechanistic model may contain several parameters and specifying or estimating those parameters from real data is not easy (Gotelli and McGill 2006).

A third approach, which we have used here and in our other studies (Ulrich and Gotelli 2007a, b, Gotelli and Ulrich 2010), is to create unstructured matrices by randomly sampling from a specified statistical distribution. This distribution should have two properties: first, matrices that are created in this way should be similar to empirical matrices in “external” properties such as their marginal totals of species richness (number of species occurrences per site) and species occurrence (number of site occurrences per species), but should be random with respect to the “internal” structure of the matrix (co-occurrence and covariance in abundance). Second, sampling from this statistical distribution should mimic a generic process model of random, independent

colonization. However, like a nonparametric test, our random sampling algorithm does not specify the colonization and extinction details of a true process model. With this strategy, the hope is that the results will be more general and capture stochastic patterns that might arise from a variety of different process-based models.

To meet these criteria (realistic marginal distributions and mimicry of a process model), we have created random matrices by drawing samples from a lognormal species abundance distribution. This distribution captures a property that has been observed in many real assemblages: a small number of species are very common, but most species are very rare (Preston 1962). When species are ranked according to their abundance or occurrence, this generates a characteristic right-skewed histogram that is approximated by sampling from a lognormal distribution. Whether the lognormal distribution itself is caused by species interactions or reflects neutral processes is still open to debate (May 1975, Sugihara 1980, McGill et al. 2006), but abundance and occurrence data collected for many taxa at widely different spatial scales often conform to an approximate lognormal distribution (McGill et al. 2007, Ulrich et al. 2010).

The distribution of species richness (species per site) is more problematic because it depends largely on the spatial grain and extent of sampling, which are often determined by the investigator. For island archipelagos, there may be a large amount of heterogeneity in the abundance and number of species per site, much of which is correlated with island area (Williamson 1981). In contrast, for small-scale samples of fixed area in homogeneous habitat, there may be relatively little variance in species richness per site. Without any a priori guide to modeling the species richness distribution, we used a random uniform distribution to determine the total abundance (number of individuals) per site and then sampled the individuals (and species) from the lognormal distribution. Although the formal definition of a Type I error is incorrect rejection of a true null hypothesis, we use an operational definition here of rejection of  $H_0$  on a set of appropriate test matrices created by random sampling from a lognormal distribution.

Just as it is challenging to specify community patterns in the absence of species interactions, it is equally challenging to specify the patterns that would be expected if communities were organized by strong species interactions. Formal mathematical theory is of little help here, because even simple models with appropriate parameter values can generate virtually any quantitative pattern of abundance and co-occurrence (Pielou 1981). However, the assembly rules literature (and much of the statistical analysis of empirical community structure) has emphasized that negative species interactions will lead to segregated patterns of species occurrence, including missing species

combinations and checkerboard distributions (Diamond 1975), species exclusion by strong predator effects (Morin 1983), and phylogenetic overdispersion of species that do co-occur (Emerson and Gillespie 2008). Aggregated patterns of co-occurrence can arise through species mutualisms and positive interactions, guild structuring, ordered extinction or differential colonization, and habitat filtering (see review in Ulrich et al. [2009]).

Therefore, a worthwhile strategy for evaluating null model algorithms is to measure their ability to detect these simple patterns. For example, Gotelli et al. (1997) began with a highly structured artificial matrix (first proposed by Diamond and Gilpin [1982]) that contained numerous segregated species pairs that never occurred together in the same site (checkerboard distributions). These kinds of extreme patterns are easily detected by most null model tests. Next, Gotelli et al. (1997) randomly swapped a few elements within some rows of the matrix, which has the effect of partially randomizing some of the species occurrences. The null model tests are applied to the new matrix, and the procedure is repeated, progressively adding more noise to the matrix. This procedure is analogous to taking a new (ordered) deck of playing cards, swapping two of the cards, and then asking a naive observer to inspect the deck after each swap and decide whether it has been shuffled or not. As more and more noise is progressively added to the deck through additional swaps, at some point it becomes impossible to recognize the “structure” that was present in the original ordering. Gotelli (2000) used this method and found that the fixed–fixed algorithm for presence/absence analysis could still detect significant patterns when ~50% of the original checkerboard matrix had been reshuffled.

In this study (and in Ulrich and Gotelli [2007a] and Gotelli and Ulrich [2010]), we have taken the opposite approach: we begin with a “random” matrix and then increase the numbers of segregated or aggregated species pairs within it and ask whether the null model algorithm can detect the change. This procedure directly addresses one of the early criticisms of null model analysis: matrices with a small number of strongly interacting species pairs might appear random because the interactions cannot be detected in a large matrix with many noninteracting species pairs (the “dilution effect”; Diamond and Gilpin 1982). Additional analysis of individual species pairs can also help to pinpoint which particular pairs are contributing to nonrandomness of the pattern for the entire matrix (Gotelli and Ulrich 2010).

It may be impossible to generate an optimal test because of the inevitable trade-off between Type I and Type II statistical errors. In our analyses, we have placed a greater priority on minimizing Type I errors (incorrect rejection of a true null hypothesis; Shrader-Frechette and McCoy 1992). There are two reasons for this. First, the entire null models controversy originat-

ed precisely over the question of Type I errors: did apparently unusual biogeographic patterns in species occurrence imply the existence of strong species interactions and assembly rules or might they have arisen by chance (Connor and Simberloff 1979)? Second, most of the data sets that have been used in null model analysis are “natural experiments” and are not based on controlled field manipulations. In such cases, the inference of mechanism from pattern is always weaker, so we prefer a more conservative approach that minimizes Type I errors. Therefore, we first confronted our candidate algorithms and metrics with the “random” matrices to eliminate tests with high Type I error rates. Then we evaluated a subset of algorithms and metrics for their performance on a set of “structured” matrices.

In spite of our attempt to test a broad array of null models and algorithms, these analyses are still optimized for their performance on the set of matrices that we created by random sampling from a lognormal distribution of species abundances. The tests are not fail-safe; it is certainly possible to generate matrices with a model of species interactions that would be incorrectly classified as random (Colwell and Winkler 1984) or, conversely, to generate matrices with a model of a stochastic process (Ulrich 2004) that would be incorrectly classified as nonrandom. But our benchmark analyses at least provide insight into how these analyses will perform with a set of artificial matrices that resemble real data in many respects and whose properties are known. In the future, perhaps it will be possible to tailor a particular test algorithm to a particular empirical matrix for maximum power. Recently, Ladau (2008) has proposed optimal null model tests based on formal parametric statistical theory. These alternative procedures are promising, although they are vulnerable to most of these same criticisms discussed here and are not as transparent as traditional null model analysis.

#### Matrix structures

We simulated two types of random abundance matrices (200 matrices each) to study the properties of 14 randomization algorithms and six measures of covariation (Fig. 1). Additionally we used 185 empirical matrices compiled from the literature that contained abundance data to apply the best performing randomization algorithms and measures and to infer the frequency of nonrandom species associations. Of these matrices, 149 were standard abundance matrices (rows = species, columns = sites) and 36 of these matrices were interaction matrices (rows, columns = species).

**Random matrices.**—We created 200 matrices ( $\mathbf{M}_R$ ) by assigning individuals randomly to matrix cells. The number of columns (= sites) in each matrix was determined by sampling from a random uniform distribution ( $5 \leq n \leq 50$  sites). To determine the number of rows (= species), we first set the total number of

species in the metacommunity  $S_T$  by sampling from a random uniform distribution ( $10 \leq S_T \leq 200$  species). However, not all of these species will necessarily be represented in the matrix because some rare species will be missing due to insufficient sampling. To mimic these sampling effects (the “veil line” of the lognormal distribution; Preston 1962), we first specified the total abundance  $N_i$  of species  $i$  by sampling from a lognormal distribution of abundances:

$$N_i = e^{x_i/2a} \quad (1)$$

where  $x_i \sim N(0, 1)$  and  $a$  is a shape-generating parameter for each matrix that is sampled from a continuous uniform distribution ( $0.1 \leq a \leq 1.0$ ). A recent meta-analysis (Ulrich et al. 2010) confirmed that the lognormal distribution most often provides the best fit to abundance data sampled from natural communities. The distribution of  $N_i$  is actually a scale mixture of a lognormal and a uniform distribution, with  $a = 0.5$  generating a standard lognormal. For large, well-sampled communities, the value of  $a$  is often  $\sim 0.2$  (Preston 1962, May 1975). To mimic the frequently observed lower truncation of the lognormal (the veil line), we sorted the  $S_T$  species according to decreasing abundance and used only the  $S_{\max}$  most abundant species in which the cutoff point “max” was sampled from a random uniform distribution ( $S_T/2 \leq S_{\max} \leq S_T$ ). This sampling procedure resulted in a matrix with  $S_{\max}$  rows and  $n$  columns.

To mimic different carrying capacities per site, the relative abundances  $A_j$  for each site  $j$  were also drawn from a random uniform distribution ( $0 < A_j \leq 1.0$ ). Having established the relative abundances for the columns and the rows of the matrix, we then assigned individuals randomly to each cell in the matrix, with the probability of choosing a particular row and column being set proportional to the row and column abundance totals (see Gotelli 2000). We placed individuals randomly in this way until all  $S_{\max}$  species were represented by at least one individual in at least one site. The relative abundance distributions generated in these matrices reflected a variety of patterns that can arise by sampling from a lognormal distribution and are similar to abundance distributions measured in nature (Magurran 2004, McGill et al. 2007, Ulrich et al. 2009).

Because the method of placing individuals in the matrix might potentially affect the assemblage patterns, we created 200 additional random matrices ( $\mathbf{M}_S$ ), using a slightly different algorithm. As before, the matrix dimensions were established by random uniform draws to determine the number of columns ( $5 \leq n \leq 50$ ) and the maximum number of species in the metacommunity ( $10 \leq S_T \leq 200$ ), again sampling from a truncated lognormal abundance distribution. However, for the  $\mathbf{M}_S$  matrices, we placed individuals sequentially (site by site) until for each site  $j$  the designated number of species  $S_j$

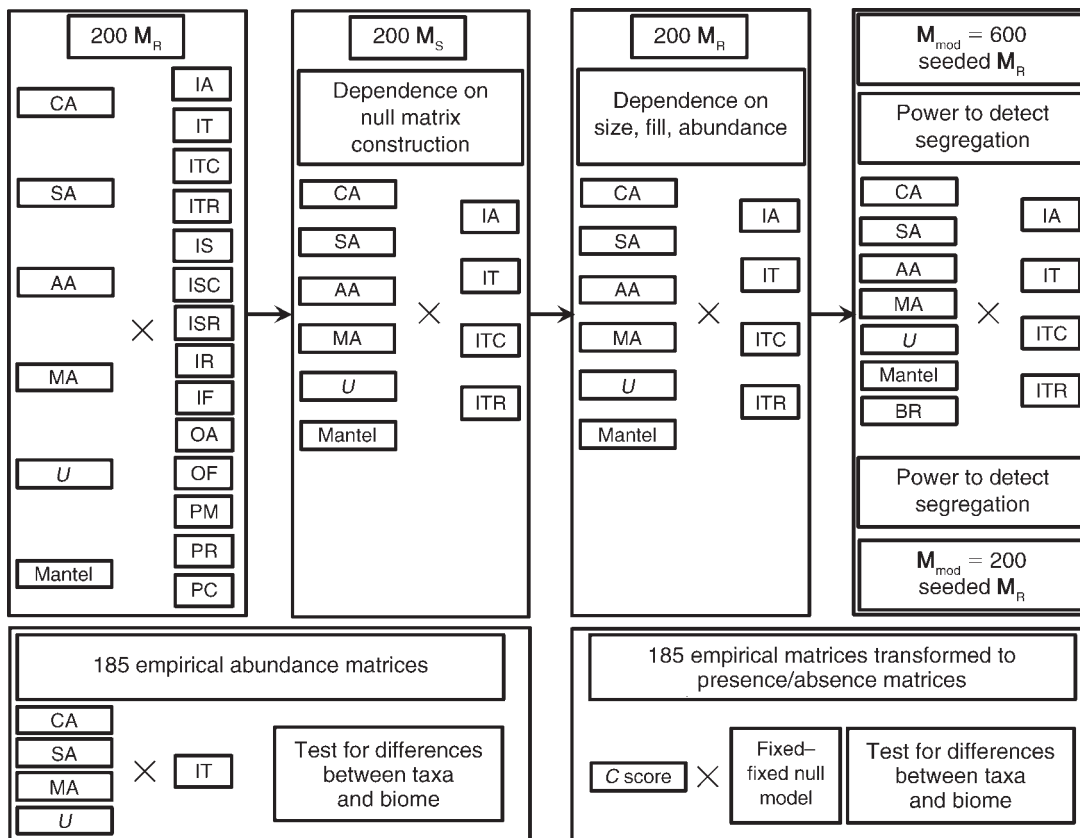


FIG. 1. Flowchart illustrating benchmark testing of covariation metrics (CA to Mantel) and null model algorithms (IA to PC) against two sets of random matrices (random,  $M_R$ , and sequential,  $M_S$ ) and two sets of seeded matrices ( $M_{mod}$ ), and testing of empirical matrices. See *Materials and methods: Strategies for evaluation of null model algorithms* and *Covariance metrics* for explanations of abbreviations.

was achieved.  $S_j$  was determined by random sampling from a uniform distribution ( $1 < S_j \leq S_T$ ).

*Empirical matrices.*—We also analyzed 149 species  $\times$  sites abundance matrices collected from the literature; all matrices used in our analyses are found in the Supplement. We classified matrices according to the taxon studied (mammals, birds, fish, arthropods, non-arthropod invertebrates, and plants), biome (terrestrial, aquatic), and, for terrestrial studies, habitat (mainland, island). We used only matrices that were based on quantitative sampling and provided integer counts of abundance (no biomass measures or ranked abundances). We analyzed separately 36 interaction matrices (species  $\times$  species) that included abundance data from the National Center for Ecological Analysis and Synthesis (NCEAS) database (*available online*<sup>4</sup>; see also the Supplement). In 53 of the data sets, the entries were density rather than abundance. We created abundance matrices in these cases by assuming the rarest species in the matrix was represented by a single individual.

<sup>4</sup> ([http://www.nceas.ucsb.edu/interactionweb/html/datasets.html#anemone\\_fish](http://www.nceas.ucsb.edu/interactionweb/html/datasets.html#anemone_fish))

Although multiple matrices were used from some studies, these almost always were for samples of different taxa or different sites. Twenty of the matrices were of the same taxa (studies on seasonal species turnover of spiders, ground, and rove beetles) sampled in different times. Therefore, we treated each empirical matrix as an independent observation in our analyses of summary patterns.

*Covariance metrics*

We developed and analyzed six indices that quantify the pattern of species aggregation and segregation in abundance matrices; some of these are directly analogous to co-occurrence metrics that are commonly used in the analysis of presence/absence matrices.

1) We used an abundance analog of “checkerboard” distributions (Diamond 1975). In presence/absence matrices, “checkerboard units” (Stone and Roberts 1990) represent submatrices of the following form:

$$\begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix}$$

(rows and columns of this submatrix form do not have

to be adjacent). The more checkerboard units there are in a matrix, the more segregated species are in their occurrence. We define an “abundance checkerboard” as a  $2 \times 2$  submatrix of the form

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad a > b \quad a > c \quad d > b \quad d > c$$

or

$$a < b \quad a < c \quad d < b \quad d < c \quad (2)$$

where  $a, b, c,$  and  $d$  represent the abundances of two species in two different sites. The metric  $CA$  is a count of the total number of abundance checkerboards in the matrix. This metric can be standardized with regard to matrix size ( $m$  rows,  $n$  columns) by

$$CA_{ST} = \frac{4CA}{m(m-1)n(n-1)}. \quad (3)$$

The standardized  $CA$  value can range from 0.0 to 1.0, with high values of  $CA$  indicating more negative covariation in abundances.

2) Similarly, we define the number of species abundance aggregations  $AA$  as a count of aggregated  $2 \times 2$  submatrices of the form

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad a > b \quad a > c \quad d < b \quad d < c$$

or

$$a < b \quad a < c \quad d > b \quad d > c. \quad (4)$$

Again a standardized metric has the following form:

$$AA_{ST} = \frac{4AA}{m(m-1)n(n-1)}. \quad (5)$$

The standardized  $AA$  value can range from 0.0 to 1.0, with high values of  $AA$  indicating positive covariation in abundance of species.  $AA$  and  $CA$  are correlated, but may differ in their ability to detect positive or negative covariation.

3) Rather than just counting abundance checkerboards as  $CA$  and  $AA$ , we can quantify the strength of covariance from the differences in abundance of the checkerboard elements:

$$A_{ST} = \frac{4 \sum \sqrt{\frac{(a-c)^2 + (b-d)^2}{a^2 + b^2 + c^2 + d^2}}}{m(m-1)n(n-1)}. \quad (6)$$

Large values of  $A_{ST}$  suggest strong positive covariance among species.

4) The variance test of Schluter (1984) is a metric of species covariance in abundance. This test compares the variance of row totals  $V$  with the sum of the column variances  $W$ . If the average covariance in abundance among all pairs of species  $\sim 0.0$ , the value  $U = V/W$  should be  $\chi^2$  distributed with  $n$  degrees of freedom. Low values of  $U$  indicate negative covariation in abundance.

5) Chao et al. (2008) extended the Morisita index of similarity for two communities to a matrix-wide metric for  $n$  communities of the following form:

$$MA = \frac{\sum_{i=1}^m \left[ \left( \sum_{j=1}^n p_{ij} \right)^2 - \sum_{j=1}^n (p_{ij})^2 \right]}{(n-1) \sum_{j=1}^n (p_{ij})^2} \quad (7)$$

where  $p_{ij}$  is the relative abundance of species  $i$  in site  $j$ . Low values of  $MA$  indicate dissimilarity of species relative abundance distribution among sites, which can be interpreted as a measure of negative covariance in relative abundances.

6) The Mantel test identifies nonrandom correlations between two matrices (Mantel 1967). To assess whether the  $\mathbf{M}_R$  matrices were nonrandom, we used the mean Mantel correlation between the  $\mathbf{M}_R$  matrix (using the Pearson correlation as distance metric) and the matrices generated by different null model algorithms. The expected correlation and confidence limits came from 100 randomly sampled null matrices. The Mantel test used in this way can identify nonrandomness, but it does not indicate whether an observed matrix is unusually aggregated or segregated.

To compare the performance of the above metrics to those that are commonly used for presence/absence analyses, we used the  $C$  score (CS; Stone and Roberts 1990) as a measure of species segregation and the discrepancy metric  $BR$  (Brualdi and Sanderson 1999) as a measure of species nestedness (Ulrich et al. 2009).

*Null model algorithms*

Compared to the analysis of binary presence/absence matrices (Gotelli 2000), there are many more possible algorithms and constraints that can be used to randomize an abundance matrix. Null model algorithms for abundance matrices can be divided into individual-based and population-based algorithms. Individual-based algorithms randomize the placement of individuals in the matrix. Population-based algorithms preserve population abundance values and randomize their occurrences among sites or species. Individual-based algorithms can be further divided into fixed-zero and floating-zero algorithms. Fixed-zero algorithms preserve species occurrences, so that the pattern of presences and absences (but not abundances) in the null matrices match those in the original matrix. Floating-zero algorithms allow for the placement of individuals in matrix cells that contained zeroes in the original matrix. We did not analyze any population-based floating-zero algorithms.

We used three population-based fixed-zero algorithms:

1)  $PM$  reshuffles populations equi-probably among the nonempty cells of the entire matrix. This model alters row and column abundance totals but preserves

species occurrences and the grand total of abundances for the entire matrix. Following the recommendation of Lehsten and Harmand (2006), we used  $100 \times n \times m$  such reshufflings to generate each null matrix for the PM, PC, and PR algorithms.

2) PC reshuffles populations equi-probably among the nonempty cells of each column. This model preserves species occurrences and total abundance per site, but alters abundances of each species.

3) PR reshuffles populations equi-probably among the nonempty cells of each row. This model preserves species occurrences and total abundance per species, but alters abundances of each site.

We used two individual-based fixed-zero algorithms:

1) OS first clears the matrix to be tested and then assigns individuals randomly only to cells that originally had nonzero values. The probability of placing an individual in a particular cell is proportional to the observed row and column abundance totals for that cell. Individuals are assigned sequentially to the matrix in this way until the total number of individuals in the original matrix is reached. OS preserves species occurrences, but does not preserve observed row and column total abundances.

2) OF first clears the matrix to be tested and then assigns individuals randomly only to cells that originally had nonzero values. The probability of placing an individual in a particular cell is proportional to the observed row and column abundance totals for that cell. Individuals are assigned sequentially to the matrix in this way until, for each row and column, total abundances are reached. This algorithm allows the abundance in each cell to vary, but preserves both species occurrences and row and column abundance totals of the original matrix. In a few cases, this algorithm stopped placing individuals before the total abundances were reached because the simultaneous constraints on row and column totals could not be met. However, the total number of individuals that could not be placed was always less than 10 (<0.1%) and should not affect the performance of the test.

We used nine individual-based floating-zero algorithms:

1) IR assigns individuals randomly to matrix cells with probabilities proportional to observed row and column abundance totals until total species richness is reached. In a few cases, this algorithm generated matrices with empty columns (sites), which were discarded prior to analysis.

2) IS assigns individuals randomly to matrix cells with probabilities proportional to observed row and column abundance totals until the total number of occurrences is reached for each row and column.

3) ISR sequentially (row by row) assigns individuals randomly to each row with probabilities proportional to observed column abundance totals until the respective number of row occurrences is reached.

4) ISC sequentially (column by column) assigns individuals randomly to each column with probabilities proportional to observed row abundance totals until the respective column total species richness is reached.

5) IT assigns individuals randomly to matrix cells with probabilities proportional to observed row and column abundance totals until, for each row and column, total abundances are reached.

6) ITR sequentially (row after row) assigns individuals randomly to each row with probabilities proportional to observed column abundance totals until the respective row total (the number of individuals) is reached.

7) ITC sequentially (column after column) assigns individuals randomly to each column with probabilities proportional to observed row abundance totals until the respective column total (the number of individuals) is reached.

8) IA reassigns all individuals randomly to matrix cells with probabilities proportional to observed row and column abundance totals until the matrix-wide total number of individuals is reached. In a few cases, this algorithm generated matrices with empty rows (species) or columns (sites), which were discarded prior to analysis.

9) IF is a two-step algorithm that preserves row and column abundances and species richness. In the first step, the algorithm converts the abundance matrix into a presence/absence matrix. Using a standard swap procedure (Gotelli 2000),  $2 \times 2$  submatrices of the form

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

to

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

are reshuffled, again using  $100 \times n \times m$  reshufflings. In the second step, the nonzero cells are cleared and then filled according to the OF algorithm.

All null models and indices were calculated with the software applications CoOccurrence and Matrix (see Supplement).

#### *Diagnostic tests*

We first determined, for each combination of null model (14 algorithms) and covariation index (six indices), whether the null model correctly identified most of the random matrices as being random. For each algorithm–index combination, we estimated the tail probabilities for the set of 200 random  $\mathbf{M}_R$  test matrices by simulating 1000 null assemblages for each random matrix. For the  $\mathbf{M}_S$  test matrices, we also simulated 1000 null assemblages, but used only the four most promising algorithm–index combinations (Fig. 1), based on their performance with the  $\mathbf{M}_R$  test matrices. If the analysis is not prone to Type I statistical errors, then approximate-

TABLE 1. Numbers of random matrices ( $\mathbf{M}_R$ ; from a total of 200) identified by six metrics of species covariation as being either segregated or aggregated (using the upper [UCL] and lower [LCL] 97.5% confidence limits).

Algorithm	CA		SA		AA		MA		U		Mantel	
	LCL	UCL	LCL	UCL	LCL	UCL	LCL	UCL	LCL	UCL	LCL	UCL
IA	0	39	0	40	6	10	5	8	3	2	0	0
IT	20	1	17	2	2	34	0	21	2	20	0	0
ITC	1	29	1	26	3	14	10	8	8	4	0	0
ITR	3	14	2	12	9	5	2	8	4	6	0	0
IS	43	25	38	19	1	89	0	39	0	98	0	2
ISC	1	34	1	30	42	0	10	15	1	80	4	0
ISR	90	68	7	174	189	9	131	66	94	74	189	0
IR	65	6	50	15	113	0	91	0	73	4	49	0
IF	8	11	10	12	54	0	25	12	3	45	0	0
OA	58	2	40	3	77	1	5	48	2	17	0	166
OF	0	72	0	88	12	4	8	24	20	56	0	164
PM	182	3	183	2	0	191	0	194	0	196	0	158
PR	171	1	173	0	0	174	1	145	1	181	15	142
PC	180	1	183	1	0	191	0	193	0	193	0	158

Note: See *Materials and methods: Strategies for evaluation of null model algorithms* and *Covariance metrics* for explanations of abbreviations.

ly five of the 200 test matrices should be statistically significant in the upper tail ( $P > 0.975$ ) and five should be significant in the lower tail ( $P < 0.025$ ) of the distribution.

After discarding a large number of model–index combinations that failed this test, we then studied the statistical power of the most promising combinations. We did this by modifying the 200  $\mathbf{M}_R$  matrices and generating from each of them four new matrices ( $\mathbf{M}_{\text{mod}}$ ). First, we generated 600 matrices for which 0.01% to maximally 33% randomly selected aggregated abundance checkerboards (as defined by Eq. 4) were rearranged as segregated checkerboards (as defined by Eq. 2). We then generated 200 matrices for which 0.01–33% randomly selected segregated abundance checkerboards (Eq. 2) of each  $\mathbf{M}_R$  matrix were changed into aggregated checkerboards. Because such changes potentially alter many other checkerboards in the matrix, the resulting number of segregated or aggregated checkerboards ranged from 0 to 7% of the total number of  $2 \times 2$  submatrices  $[nm(n-1)(m-1)/4]$ . We then calculated the fraction of these modified matrices that were correctly identified as statistically significantly segregated or aggregated ( $P < 0.05$  in either tail) by the different algorithms. From these analyses, we were able to identify combinations of algorithms and metrics that had the best power for detecting segregated and aggregated distributions from nonrandom matrices. Naturally, greater replication in benchmark tests would be desirable, but the tests are time-consuming, and the results were clear-cut with these samples of 200 and 600 test matrices.

We also calculated a standardized effect size (SES) as a  $Z$ -transformed score  $[Z = (x - \mu)/\sigma]$ , where  $x$  is the observed index for the  $\mathbf{M}_R$  or  $\mathbf{M}_S$  test matrix,  $\mu$  is the mean of the 1000 simulated indices for each of the null model algorithms, and  $\sigma$  is the standard deviation of the 1000 simulated indices. The use of SES is based on the assumption of an approximately normal error distribution. This was indeed the case: the mean skewness of all

null model distributions was only 0.004 with a standard deviation of 0.37. For a random sample of scores that follows a normal distribution, ~95% of the SES values should be  $<|2.0|$ . We used the SES to test whether null model results were sensitive to basic matrix properties (size, fill, and mean species abundances).

Based on the results of these tests, we used one of the algorithms (IT) in combination with four metrics (MA, U, CA, SA) to test for patterns of segregation and aggregation in the empirical matrices.

## RESULTS

### *Performance of null model algorithms with random test matrices*

Null models that reshuffled whole populations (PM, PR, PC) performed poorly (Table 1). Irrespective of metric, they identified more than 70% of the random  $\mathbf{M}_R$  matrices as being segregated. The two null models that resampled observed occurrences (OA, OF) identified between 10% (OA with the variance test) and 83% (OA with the Mantel test) of the test matrices as nonrandom. Individual-based null models that were conditioned on row, column, or total abundances (IA, IT, ITC, ITR) performed better than those that were constrained to match observed species richness (IS, ISC, ISR, IR). In particular, the individual-based null models IT, ITC, and ITR identified only 0–15% of the  $\mathbf{M}_R$  matrices as being not random (Table 1). The performance of the two-step IF null model was a bit worse: 0–27% of random  $\mathbf{M}_R$  matrices were statistically significant. For the sequentially filled random  $\mathbf{M}_S$  matrices, IA, ITC, and ITR performed worse and incorrectly classified 0–79% as not random (Table 2). Overall, the best performing null model was IT, which identified a maximum of 17% of the  $\mathbf{M}_R$  matrices (with the AA metric) and 13% of the  $\mathbf{M}_S$  matrices (with the Mantel test) as being not random.



TABLE 2. Numbers of sequential matrices ( $M_S$ ; from a total of 200) identified by six metrics of species covariation as being either segregated or aggregated (using the upper [UCL] and lower [LCL] 97.5% confidence limits), using only the most promising null models of Table 1.

Algorithm	CA		SA		AA		MA		U		Mantel	
	LCL	UCL	LCL	UCL	LCL	UCL	LCL	UCL	LCL	UCL	LCL	UCL
IA	0	133	0	133	94	0	20	2	64	25	0	0
IT	3	9	1	11	5	2	0	11	0	13	26	1
ITC	0	69	0	72	8	2	61	3	69	3	0	0
ITR	0	84	0	83	82	0	2	67	1	37	0	158

Note: See *Materials and methods: Strategies for evaluation of null model algorithms* and *Covariance metrics* for explanations of abbreviations.

*Performance of covariance metrics with random test matrices*

For the best performing null models IA, IT, ITC, and ITR, many of the six metrics tested had satisfactory Type I error rates and correctly classified >77% of the  $M_R$  matrices as being random (Table 1). For the  $M_S$  matrices, all metrics except the Mantel test performed well with the IT null model, but largely failed with IA, ITR, and ITC (Table 2). Results of null model tests using the metrics MA, AA, and U were least dependent on matrix size, fill, and mean abundance (Table 3). The Z-transformed values of CA, SA, and, particularly, the Mantel test

TABLE 3. Pearson correlation coefficients between matrix size, fill, and mean abundance (= mean number of individuals per cell), and the Z-standardized covariation metric for the four best-performing null models (Table 1) of the random matrices ( $M_R$ ).

Matrix and properties	IA	IT	ITC	ITR
CA				
Size	0.00	<b>-0.16</b>	0.02	-0.09
Fill	<b>0.44</b>	<b>-0.26</b>	<b>-0.30</b>	<b>-0.31</b>
Mean abundance	<b>0.31</b>	-0.14	<b>-0.23</b>	<b>-0.15</b>
SA				
Size	-0.02	<b>-0.16</b>	0.01	0.09
Fill	<b>0.43</b>	<b>-0.26</b>	<b>-0.29</b>	<b>-0.31</b>
Mean abundance	<b>-0.29</b>	-0.12	<b>-0.20</b>	-0.13
AA				
Size	0.00	<b>0.17</b>	0.00	-0.03
Fill	<b>0.23</b>	0.11	0.10	0.11
Mean abundance	0.10	0.09	0.02	0.07
MA				
Size	-0.03	0.01	-0.02	-0.09
Fill	0.14	0.01	<b>0.18</b>	0.04
Mean abundance	0.09	-0.01	<b>0.15</b>	0.06
U				
Size	0.10	<b>0.17</b>	0.11	0.09
Fill	<b>0.18</b>	0.00	<b>0.18</b>	-0.07
Mean abundance	0.12	0.05	<b>0.19</b>	-0.01
Mantel				
Size	-0.04	0.05	-0.06	0.08
Fill	<b>0.54</b>	<b>0.27</b>	<b>0.36</b>	<b>0.41</b>
Mean abundance	<b>0.27</b>	<b>0.21</b>	0.13	<b>0.21</b>

Note: Significant correlations ( $P < 0.05$ ) appear in boldface. See *Materials and methods: Strategies for evaluation of null model algorithms* and *Covariance metrics* for explanations of abbreviations.

TABLE 4. Proportions of modified matrices ( $M_{mod}$ ; total of 800) that were detected as being either aggregated (-1% to -10% checkerboards, given as a percentage of the total number of  $2 \times 2$  submatrices) or segregated (0.01% to 10%).

Change in number of checkerboards	IA	IT	ITC	ITR
CA				
-1% to -10%	0.05	0.45	0.06	0.27
0.01% to 0.1%	0.25	0.02	0.17	0.10
0.1% to 1%	0.53	0.11	0.43	0.35
1% to 10%	0.94	0.81	0.94	0.89
SA				
-1% to -10%	0.06	0.46	0.08	0.28
0.01% to 0.1%	0.27	0.01	0.16	0.11
0.1% to 1%	0.50	0.10	0.40	0.30
1% to 10%	0.93	0.76	0.94	0.87
AA				
-1% to -10%	0.09	0.13	0.06	0.06
0.01% to 0.1%	0.07	0.03	0.16	0.09
0.1% to 1%	0.25	0.13	0.43	0.33
1% to 10%	0.80	0.72	0.91	0.84
CS				
-1% to -10%	0.08	0.44	0.10	0.30
0.01% to 0.1%	0.26	0.01	0.16	0.07
0.1% to 1%	0.39	0.06	0.30	0.20
1% to 10%	0.91	0.69	0.90	0.79
MA				
-1% to -10%	0.03	0.15	0.02	0.06
0.01% to 0.1%	0.09	0.12	0.12	0.09
0.1% to 1%	0.36	0.36	0.46	0.33
1% to 10%	0.90	0.91	0.94	0.84
U				
-1% to -10%	0.02	0.14	0.02	0.13
0.01% to 0.1%	0.03	0.06	0.08	0.05
0.1% to 1%	0.25	0.27	0.35	0.28
1% to 10%	0.87	0.89	0.90	0.87
Mantel				
-1% to -10%	0.00	0.00	0.00	0.00
0.01% to 0.1%	0.02	0.01	0.02	0.01
0.1% to 1%	0.14	0.05	0.12	0.09
1% to 10%	0.54	0.42	0.51	0.41
BR				
-1% to -10%	0.05	0.29	0.05	0.25
0.01% to 0.1%	0.20	0.02	0.14	0.07
0.1% to 1%	0.19	0.05	0.13	0.11
1% to 10%	0.76	0.58	0.74	0.70

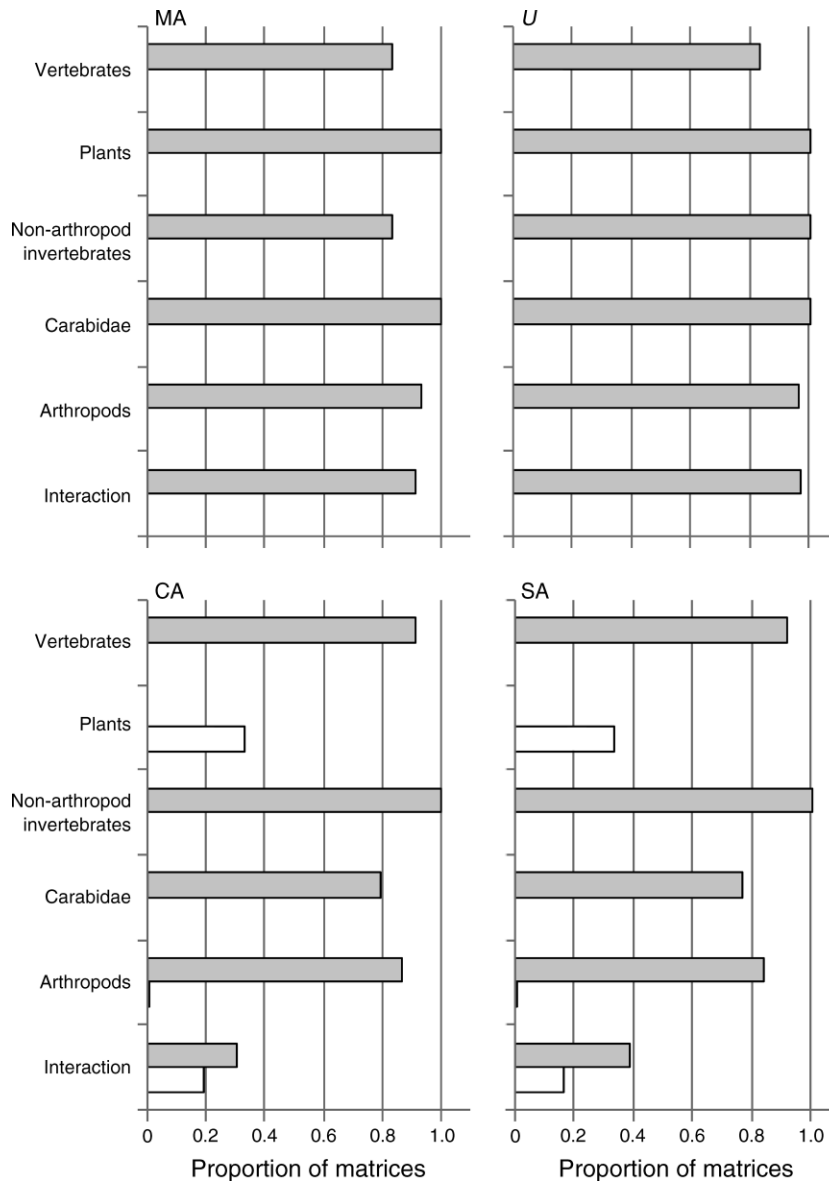


FIG. 2. Fraction of 185 empirical abundance matrices that were significantly segregated (gray bars) or aggregated (white bars) according to the metrics MA,  $U$ , CA, and SA under the IT null model: vertebrates ( $N = 14$ ), plants ( $N = 3$ ), non-arthropod invertebrates ( $N = 6$ ), Carabidae ( $N = 39$ ), all arthropods ( $N = 126$ ), and interaction matrices ( $N = 36$ ). See *Materials and methods: Strategies for evaluation of null model algorithms* and *Covariance metrics* for explanations of abbreviations.

tended to correlate with matrix fill. The largest correlation was between matrix fill and the Mantel value for the IA algorithm (Pearson's  $r = 0.54$ ,  $P < 0.01$ ).

*Performance of null model algorithms and covariance metrics with segregated and aggregated test matrices*

The diagnostic tests with matrices in which we increased or decreased the numbers of abundance checkerboards reveal the power of the null models to correctly identify nonrandom patterns. These tests suggest that the IT algorithm was less powerful than IA, ITC, and ITR for detecting species segregation, but

was more powerful for detecting species aggregation (Table 3). CA and SA in combination with IT had the best power to detect species aggregation and correctly identified 45% and 46%, respectively, of the manipulated matrices as aggregated (Table 4). The power of both tests was comparable to the power of the  $C$  score to detect aggregations when the same matrices were analyzed in the form of presence/absence data. With the IT algorithm, the abundance based metrics CA, MA, and  $U$  correctly identified at least 81% of the segregated matrices (1–10% increase in checkerboards) as nonrandom. The presence/absence indices CS ( $C$

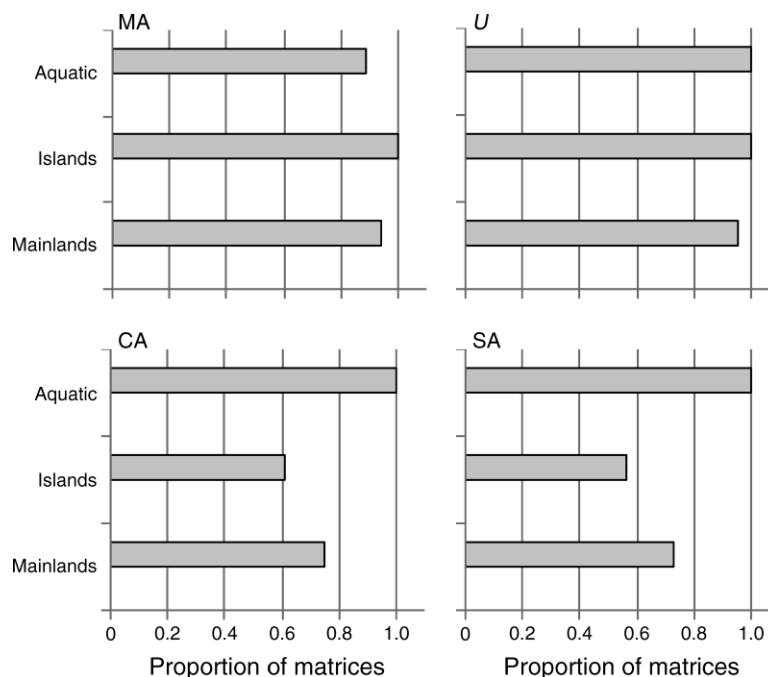


FIG. 3. Fraction of 149 empirical abundance matrices that were significantly segregated (gray bars) or aggregated (white bars) according to the metrics MA, *U*, CA, and SA under the IT null model: aquatic ( $N = 9$ ), mainlands ( $N = 139$ ), and islands ( $N = 23$ ). See *Materials and methods: Strategies for evaluation of null model algorithms* and *Covariance metrics* for explanations of abbreviations.

score) and BR (nestedness) were more conservative and identified only 69% and 58%, respectively, of these matrices as segregated (Table 4). With the IT algorithm, the MA performed best (91% of segregated matrices correctly identified), but the Mantel test was too conservative (only 42% of the matrices identified). Based on analyses of both random matrices (Tables 1–3) and structured matrices (Table 4) the IT algorithm is best for testing empirical matrices. The CA and SA metrics are more conservative, and the MA and *U* metrics are more liberal in the detection of nonrandom patterns.

#### *Meta-analysis of empirical abundance matrices*

MA and *U* identified >80% of the 185 real abundance matrices as being significantly segregated (Fig. 2). CA and SA gave similar results, except that most plant and interaction matrices were classified as random. However, CA and SA also identified aggregated patterns in the plant and interaction matrices that were not detected with MA and *U*. All four metrics identified >50% of the aquatic, mainland, and island data set as being significantly segregated (Fig. 3). Again, CA and SA appeared to be more conservative than MA and *U*. The *C* score, which is based only on presence/absence data, was more conservative than the abundance-based metrics and identified at most 39% of the animal and interaction matrices as being segregated. However, this score classified 33–81% of the matrices as being either random or even aggregated (Fig. 4A) and the results did

not differ greatly among island/aquatic/mainland matrices (Fig. 4B).

#### DISCUSSION

Careful benchmark testing of potential randomization algorithms and community metrics is essential for valid null model analyses (Gotelli 2001). In this case, the vast majority of algorithms and metrics that we evaluated had unacceptably high Type I error rates when tested with a series of random matrices that were created from random sampling of a lognormal species abundance distribution (Table 1). In fact, the only metric that met the strict criterion of rejecting  $H_0$  for <5% of the null model matrices was the Mantel metric. However, this index does not indicate whether a nonrandom matrix is predominantly aggregated or segregated, and it had poor power for detecting significant aggregation or segregation (Table 4). Most of the null models tests showed some correlations with measures of matrix size or fill (Table 3). However, most of the correlations are fairly weak, and they reflect the universal property that, as sample size becomes very large, the null hypothesis will inevitably be rejected because the randomization algorithm is not identical to the (lognormal) sampling model that was used to generate the test matrices.

Overall, the IT algorithm performed best, with fairly low Type I error rates (Tables 1 and 2) but good power for detecting aggregated or segregated distributions with a variety of metrics (Table 4). It is interesting to note

that the IT algorithm assigns individuals to matrix cells proportional to observed row and column totals until, for each row and column, total abundances are reached. A similar null distribution underlies contingency table analysis (Everitt 1980), in which the null hypothesis of no species  $\times$  site interaction is tested by assuming independent marginal probabilities for each cell in the matrix (cf. Diamond and Gilpin 1982). However, Monte Carlo simulations usually do not produce identical results to parametric tests of the same data (Gotelli and Ellison 2004). In this case, a parametric chi-square test (without corrections for small sample size or sparse matrices; Gotelli and Ellison 2004) of the 200 random  $M_R$  matrices identified 33 (16.5%) as being not random at the 5% error level, a value within the range observed for the IT null model (Table 1). Alternative algorithms are available that do not fix row and column abundance totals, but allow them to vary among different simulated matrices (see Gotelli and Graves [1996] and Gotelli [2000] for a discussion of this algorithm for presence/absence analysis). However, such tests are potentially prone to greater Type I error, because the null hypothesis might be rejected due to differences in row and column sums per se, rather than because of aggregated or segregated abundance distributions.

When we applied the IT test with several metrics to the empirical abundance matrices, nearly all of them showed strongly segregated patterns (Fig. 2), although a few aggregated distributions were also detected for plant and interaction matrices. Although differences among taxa were strong (Fig. 2), differences among habitat type were not (Fig. 2), which is similar to the findings of Gotelli and McCabe (2002) for presence/absence matrices. The frequency of segregated distributions in these real data sets (Fig. 2) is far greater than would be expected from the frequencies expected in our null model tests (Tables 1 and 2). When these same matrices were converted to a presence/absence form and analyzed with the standard fixed-fixed null model (Gotelli 2001), segregated patterns still dominated, although the frequency of nonrandom matrices was much lower (Fig. 4). These results suggest that null model analysis of abundance matrices may potentially be more powerful than null model analysis of presence/absence matrices (Hausdorf and Hennig 2007), although the latter are more common in the literature and are easier for field biologists to generate. Detection errors and imprecise counts are certainly present in both abundance and presence/absence matrices, and they potentially affect the power of the tests. However, these factors have only recently been incorporated into statistical tests for species interactions (Royle and Dorazio 2008, Waddle et al. 2010).

Although ecologists routinely test for pairwise correlations of species abundances (Brown et al. 2004), there has been relatively little use of null models with abundance data. Some null model tests have been applied to the analysis of relative abundance distributions (McGill et al. 2007), although these tests often use

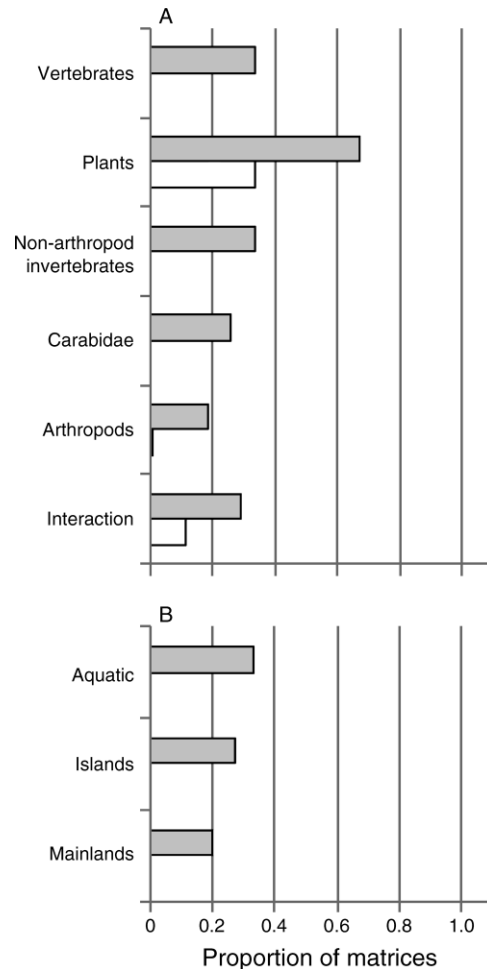


FIG. 4. Fraction of empirical abundance matrices (converted to presence/absence) with a significant  $C$  score (5% error level) under the fixed-fixed null model classified (A) according to taxon and type and (B) according to biome. In panel (B) only the 147 species  $\times$  site matrices are included. Gray bars indicate species segregation; white bars indicate aggregation.

just the row sums of the species abundance matrix. Null model tests have been used with species abundance matrices in which the columns represent sampling periods rather than sites. In the 1980s, ecologists used null model tests of these data to determine whether species ranks remained concordant through time (Grossman 1982, Ebeling et al. 1990), which is an important measure of community, persistence, and stability (Pimm 1984). More recently, Houlahan et al. (2007) calculated Schluter's (1984) variance ratio (the  $U$  metric in our analyses) for a large number of published species temporal matrices. Most of these indices indicated a pattern of species aggregation: abundances tended to covary positively through time, suggesting that compensatory dynamics were not important. Schluter (1984) found a similar pattern when he analyzed species  $\times$  site abundance matrices. However, the tests by Houlahan et al. (2007) and Schluter (1984) assume that the columns

of the matrix (sites or times) are equivalent and do not affect the probability of occurrence of individuals of different species (Gotelli 2001). In contrast, the IT algorithm that we used preserves the column totals for abundance, taking into account differences in suitability or conditions among sites. With this null model, the  $U$  metric reveals mostly segregated patterns in species abundance matrices (Fig. 2).

In summary, null model analysis that was first developed for binary presence/absence matrices can be effectively extended to abundance matrices and may be a more powerful test for segregated, aggregated, and random patterns of abundance. The statistical performance of null models against mechanistic colonization and extinction scenarios is largely unknown, but some insight can be gained by comparing the performance of null models to simple sampling distributions, as we have done here for the lognormal species abundance distribution. Like nearly all previous null models, our algorithms do not use information on the location of the samples or take into account spatial autocorrelation (Lichstein et al. 2002). Most existing tests (including ours) also treat abundance and presence/absence data as error-free and do not address the problem of undetected species (Dorazio 2007). Species detection and spatial autocorrelation are promising, but largely unexplored, avenues for future null model analyses.

#### ACKNOWLEDGMENTS

Comments by Lewi Stone, Robert Dorazio, and three referees greatly improved an earlier version of this paper. We thank Adam Barcikowski, Izabela Hajdamowicz, Piotr Jastrzebski, Bartłomiej Pacuk, Marzena Stańska, January Weiner, and Marcin Żalewski for the permission to include unpublished data sets in the present analysis. N. J. Gotelli acknowledges support from grants NSF DEB-0541936 and DOE DE-FG02-08ER64510. W. Ulrich was in part supported by grants from the Polish Science Committee (KBN 3 P04F 034 22, and KBN 2 P04F 039 29).

#### LITERATURE CITED

- Bascompte, J., and P. Jordano. 2007. Plant–animal mutualistic networks: the architecture of biodiversity. *Annual Review of Ecology, Evolution, and Systematics* 38:567–593.
- Bell, G. 2005. The co-distribution of species in relation to the neutral theory of community ecology. *Ecology* 86:1757–1770.
- Bellwood, D. R., P. C. Wainwright, C. J. Fulton, and A. Hoey. 2002. Assembly rules and functional groups at global biogeographical scales. *Functional Ecology* 16:557–562.
- Blüthgen, N., F. Menzel, T. Hovestadt, B. Fiala, and N. Blüthgen. 2007. Specialization, constraints, and conflicting interests in mutualistic networks. *Current Biology* 17:341–346.
- Brown, J. H., E. J. Bedrick, S. K. M. Ernst, J.-L. E. Carton, and J. F. Kelly. 2004. Constraints on negative relationships: mathematical causes and ecological consequences. Pages 298–323 in M. L. Taper and S. R. Lele, editors. *The nature of scientific evidence: statistical, philosophical, and empirical considerations*. University of Chicago Press, Chicago, Illinois, USA.
- Brualdi, R. A., and J. G. Sanderson. 1999. Nested species subsets, gaps, and discrepancy. *Oecologia* 119:256–264.
- Chao, A., L. Jost, S. C. Chiang, Y.-H. Jiang, and R. L. Chazdon. 2008. A two stage probabilistic approach to multiple-community similarity indices. *Biometrics* 64:1178–1186.
- Cohen, J. 1992. A power primer. *Psychological Bulletin* 112:155–159.
- Colwell, R. K., and D. W. Winkler. 1984. A null model for null models in biogeography. Pages 344–359 in D. R. Strong, Jr., D. Simberloff, L. G. Abele, and A. B. Thistle, editors. *Ecological communities: conceptual issues and the evidence*. Princeton University Press, Princeton, New Jersey, USA.
- Connor, E. F., and D. Simberloff. 1979. The assembly of species communities: Chance or competition? *Ecology* 60:1132–1140.
- Diamond, J. M. 1975. Assembly of species communities. Pages 342–444 in M. L. Cody and J. M. Diamond, editors. *Ecology and evolution of communities*. Harvard University Press, Cambridge, Massachusetts, USA.
- Diamond, J. M., and M. E. Gilpin. 1982. Examination of the “null” model of Connor and Simberloff for species co-occurrences on islands. *Oecologia* 52:64–74.
- Dorazio, R. M. 2007. On the choice of statistical models for estimating occurrence and extinction from animal surveys. *Ecology* 88:2773–2782.
- Ebeling, A. W., S. J. Holbrook, and R. J. Schmitt. 1990. Temporally concordant structure of a fish assemblage: Bound or determined? *American Naturalist* 135:63–73.
- Emerson, B. C., and R. G. Gillespie. 2008. Phylogenetic analysis of community assembly and structure over space and time. *Trends in Ecology and Evolution* 23:619–630.
- Everitt, B. S. 1980. *The analysis of contingency tables*. Second edition. Monographs on Statistics and Applied Probability 45. CRC Press, Boca Raton, Florida, USA.
- Fukami, T., and P. J. Morin. 2003. Productivity–biodiversity relationships depend on the history of community assembly. *Nature* 424:423–426.
- Gotelli, N. J. 2000. Null model analysis of species co-occurrence patterns. *Ecology* 81:2606–2621.
- Gotelli, N. J. 2001. Research frontiers in null model analysis. *Global Ecology and Biogeography* 10:337–343.
- Gotelli, N. J. 2004. Assembly rules. Pages 1027–1035 in M. V. Lomolino, D. F. Sax, and J. H. Brown, editors. *Foundations of biogeography: classic papers with commentaries*. University of Chicago Press, Chicago, Illinois, USA.
- Gotelli, N. J., N. J. Buckley, and J. A. Wiens. 1997. Co-occurrence of Australian land birds: Diamond’s assembly rules revisited. *Oikos* 80:311–324.
- Gotelli, N. J., and A. M. Ellison. 2004. *A primer of ecological statistics*. Sinauer, Sunderland, Massachusetts, USA.
- Gotelli, N. J., and G. R. Graves. 1996. *Null models in ecology*. Smithsonian Institution Press, Washington, D.C., USA.
- Gotelli, N. J., and D. J. McCabe. 2002. Species co-occurrence: a meta-analysis of J. M. Diamond’s assembly rules model. *Ecology* 83:2091–2096.
- Gotelli, N. J., and B. J. McGill. 2006. Null versus neutral models: What’s the difference? *Ecography* 29:793–800.
- Gotelli, N. J., and W. Ulrich. 2010. The empirical Bayes approach as a tool to identify non-random species associations. *Oecologia* 162:463–477.
- Graves, G. R., and N. J. Gotelli. 1993. Assembly of avian mixed-species flocks in Amazonia. *Proceedings of the National Academy of Sciences USA* 90:1388–1391.
- Grossman, G. D. 1982. Dynamics and organization of a rocky intertidal fish assemblage: the persistence and resilience of taxocene structure. *American Naturalist* 119:611–637.
- Harvey, P. H., R. K. Colwell, J. W. Silvertown, and R. M. May. 1983. Null models in ecology. *Annual Review of Ecology and Systematics* 14:189–211.
- Hausdorf, B., and C. Hennig. 2007. Null model tests of clustering of species, negative co-occurrence patterns and nestedness in meta-communities. *Oikos* 116:818–828.

- Houlahan, J. E., et al. 2007. Compensatory dynamics are rare in natural ecological communities. *Proceedings of the National Academy of Sciences USA* 104:3273–3277.
- Irving, A. D., and C. D. Connell. 2006. Predicting understorey structure from the presence and composition of canopies: an assembly rule for marine algae. *Oecologia* 148:491–502.
- Jordano, P., J. Bascompte, and J. M. Olesen. 2003. Invariant properties in coevolutionary networks of plant–animal interactions. *Ecology Letters* 6:69–81.
- Ladau, J. 2008. Validation of null model tests using Neyman–Pearson hypothesis testing theory. *Theoretical Ecology* 1: 241–248.
- Lavorel, S., and E. Garnier. 2002. Predicting changes in community composition and ecosystem functioning from plant traits: revisiting the Holy Grail. *Functional Ecology* 16: 545–556.
- Lehsten, V., and P. Harmand. 2006. Null models for species co-occurrence patterns: assessing bias and minimum iteration number for the sequential swap. *Ecography* 29:786–792.
- Lester, P. J., K. L. Abbott, M. Sarty, and K. C. Burns. 2009. Competitive assembly of South Pacific invasive ant communities. *BMC Ecology* 9:3.
- Lichstein, J. W., T. R. Simons, S. A. Shriener, and K. E. Franzreb. 2002. Spatial autocorrelation and autoregressive models in ecology. *Ecological Monographs* 72:445–463.
- Magurran, A. E. 2004. *Measuring biological diversity*. Blackwell, Oxford, UK.
- Manly, B. F. J. 1991. *Randomization and Monte Carlo methods in biology*. Chapman and Hall, London, UK.
- Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Research* 27:209–220.
- May, R. M. 1975. Patterns of species abundance and diversity. Pages 81–120 in M. L. Cody and J. M. Diamond, editors. *Ecology and evolution of communities*. Belknap, Cambridge, Massachusetts, USA.
- McCoy, E. D., and K. L. Heck, Jr. 1987. Some observations on the use of taxonomic similarity in large-scale biogeography. *Journal of Biogeography* 14:79–87.
- McGill, B. J., B. A. Maurer, and M. D. Weiser. 2006. Empirical evaluation of neutral theory. *Ecology* 87:1411–1423.
- McGill, B. J., et al. 2007. Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecology Letters* 10:995–1015.
- Morin, P. J. 1983. Predation, competition, and the composition of larval anuran guilds. *Ecological Monographs* 53:119–138.
- Patterson, B. D., and W. Atmar. 1986. Nested subsets and the structure of insular mammalian faunas and archipelagos. *Biological Journal of the Linnean Society* 28:65–82.
- Pielou, E. C. 1981. The usefulness of ecological models: a stock-taking. *Quarterly Review of Biology* 56:17–31.
- Pimm, S. L. 1984. The complexity and stability of ecosystems. *Nature* 307:321–326.
- Preston, F. W. 1962. The canonical distribution of commonness and rarity: parts I and 2. *Ecology* 43:185–215, 410–432.
- Royle, J. A., and R. M. Dorazio. 2008. *Hierarchical modeling and inference in ecology*. Academic Press, Amsterdam, The Netherlands.
- Schluter, D. 1984. A variance test for detecting species associations, with some example application. *Ecology* 65: 998–1005.
- Shrader-Frechette, K. S., and E. D. McCoy. 1992. Statistics, costs and rationality in ecological inference. *Trends in Ecology and Evolution* 7:96–99.
- Stone, L., and A. Roberts. 1990. The checkerboard score and species distributions. *Oecologia* 85:74–79.
- Sugihara, G. 1980. Minimal community structure: an explanation of species abundance patterns. *American Naturalist* 116: 770–787.
- Toft, C. A., and P. J. Shea. 1983. Detecting community-wide patterns: estimating power strengthens statistical inference. *American Naturalist* 122:618–625.
- Ulrich, W. 2004. Species co-occurrences and neutral models: reassessing J. M. Diamond's assembly rules. *Oikos* 107:603–609.
- Ulrich, W., M. Almeida-Neto, and N. J. Gotelli. 2009. A consumer's guide to nestedness analysis. *Oikos* 118:3–17.
- Ulrich, W., and N. J. Gotelli. 2007a. Null model analysis of species nestedness patterns. *Ecology* 88:1824–1831.
- Ulrich, W., and N. J. Gotelli. 2007b. Disentangling community patterns of nestedness and species co-occurrence. *Oikos* 116: 2053–2061.
- Ulrich, W., M. Ollik, and K. I. Ugland. 2010. A meta-analysis of species–abundance distributions. *Oikos* 119:1149–1155.
- Waddle, J. H., R. M. Dorazio, S. C. Walls, K. G. Rice, J. Beauchamp, M. J. Schuman, and F. J. Mazzotti. 2010. A new parameterization for estimating co-occurrence of interacting species. *Ecological Applications*, *in press*.
- Webb, C. O., D. D. Ackerly, M. A. McPeck, and M. J. Donoghue. 2002. Phylogenies and community ecology. *Annual Review of Ecology and Systematics* 33:475–505.
- Weiher, E., and P. Keddy, editors. 1999. *Ecological assembly rules, perspectives, advances, retreats*. Cambridge University Press, Cambridge, UK.
- Williams, C. B. 1964. *Patterns in the balance of nature*. Academic Press, New York, New York, USA.
- Williamson, M. 1981. *Island populations*. Oxford University Press, Oxford, UK.

## SUPPLEMENT

Data matrices and source codes of the software used in the paper (*Ecological Archives* E091-240-S1).