



## Pattern detection in null model analysis

Werner Ulrich and Nicholas J. Gotelli

W. Ulrich (*ulrichw@umk.pl*), Nicolaus Copernicus Univ. in Toruń, Chair of Ecology and Biogeography, Gagarina 9, PL-87-100 Toruń, Poland. – N. J. Gotelli, Dept of Biology, Univ. of Vermont, Burlington, VT 05405, USA.

Synthesis

The identification of distinctive patterns in species x site presence-absence matrices is important for understanding meta-community organisation. We compared the performance of a suite of null models and metrics that have been proposed to measure patterns of segregation, aggregation, nestedness, coherence, and species turnover. We found that any matrix with segregated species pairs can be re-ordered to highlight aggregated pairs, indicating that these seemingly opposite patterns are closely related. Recently proposed classification schemes failed to correctly classify realistic matrices that included multiple co-occurrence structures. We propose using a combination of metrics and decomposing matrix-wide patterns into those of individual pairs of species and sites to pinpoint sources of non-randomness.

Null model analysis has been a popular tool for detecting pattern in binary presence-absence matrices, and previous tests have identified algorithms and metrics that have good statistical properties. However, the behavior of different metrics is often correlated, making it difficult to distinguish different patterns. We compared the performance of a suite of null models and metrics that have been proposed to measure patterns of segregation, aggregation, nestedness, coherence, and species turnover. We found that any matrix with segregated species pairs can be re-ordered to highlight aggregated pairs. As a consequence, the same null model can identify a single matrix as being simultaneously aggregated, segregated or nested. These results cast doubt on previous conclusions of matrix-wide species segregation based on the C-score and the fixed-fixed algorithm. Similarly, we found that recently proposed classification schemes based on patterns of coherence, nestedness, and segregation and aggregation cannot be uniquely distinguished using proposed metrics and null model algorithms. It may be necessary to use a combination of different metrics and to decompose matrix-wide patterns into those of individual pairs of species or pairs of sites to pinpoint the sources of non-randomness.

A major research effort in community ecology has been to infer mechanisms of community organization from patterns in a binary presence-absence matrix. In such a matrix, species are represented by rows, sites are represented by columns, and the entries are the presence or absence of a species in a particular site (McCoy and Heck 1987). Early research (and controversy) focused on the role of inter-specific competition in producing checkerboard patterns (Diamond 1975), species segregation (Stone and Roberts 1990), and missing species combinations (Pielou and Pielou 1968). A largely independent focus on species nestedness (Darlington 1957, Patterson and Atmar 1986) emphasized the role of selective, orderly extinction (and immigration) in producing species-poor assemblages whose composition is a perfectly nested subset of more species-rich assemblages.

Leibold and Mikkelsen (2002) significantly expanded this framework by proposing tests for additional patterns of Clementsian, Gleasonian, and evenly-spaced gradients.

Their method involves re-ordering the rows and columns of a presence-absence matrix by reciprocal averaging to maximize the coherence of species range sizes. A maximally coherent matrix state is one that contains the greatest number of uninterrupted sequences of species occurrences within rows of the re-ordered matrix. In such a re-ordered matrix, 'species turnover' is the replacement of one species by another, and 'boundary clumping' is the extent to which species boundaries are clumped or over-dispersed. Checkerboards, nestedness, Clementsian gradients, Gleasonian gradients, and evenly-spaced gradients are recognized by a combination of patterns of coherence, species turnover, and boundary clumping.

Presley et al. (2010) further expanded Leibold and Mikkelsen's (2002) framework, distinguishing among 12 different models based on the combination of boundary clumping and turnover patterns for matrices in which there is positive coherence. They distinguished three types of nested patterns, depending on whether boundary clumping was positive, random, or negative, and they recognized six additional 'quasi-structures' based on weakly positive or negative patterns of species turnover.

The review and decision to publish this paper has been taken by the above noted SE. The decision by the handling SE is shared by a second SE and the EiC.

Both Leibold and Mikkelsen (2002) and Presley et al. (2010) also recognized a pattern of ‘randomness’ which would imply purely stochastic forces in community assembly. This null model represents a ground state against which the other patterns can be meaningfully compared and statistically tested (Gotelli and Graves 1996). Mechanistically, such a null model can be thought of as a stochastic colonization model that excludes species interactions and abiotic forces that structure communities along gradients. Island biogeography models (MacArthur and Wilson 1963), neutral models (Hubbell 2001), and various meta-community models (Holyoak et al. 2005) can all be interpreted this way. However, estimating the parameters of these models from empirical data has been very challenging (Gotelli and McGill 2006). A more fruitful alternative has been to use a statistical null model, in which an observed presence–absence matrix is reshuffled or randomized in a way that would mimic stochastic assembly of a community without specifying all the details of speciation, colonization, dispersal and extinction (Gotelli and Ulrich 2012).

Null models are always used in conjunction with metrics that quantify the pattern in a presence–absence matrix, usually as a single number. There have been many such measures proposed for quantifying pattern in presence–absence matrices, including metrics of nestedness (Ulrich et al. 2009), species segregation (Stone and Roberts 1990), turnover and  $\beta$ -diversity (Leibold and Mikkelsen 2002, Baselga 2010, Tuomisto 2010), coherence (Leibold and Mikkelsen 2002), and boundary clumping (Clark and Evans 1954, Hoagland and Collins 1997). Similarly, many algorithms have been proposed for randomizing a presence–absence matrix to mimic a pattern of random community assembly (Gotelli 2000, Ulrich et al. 2009). These algorithms differ in whether the row and column sums of the randomized matrix are treated as fixed, equiprobable, or varying as a function of independently measured properties of sites or species.

As a consequence, there is now a plethora of seemingly reasonable ways to analyze such matrices. Some of these methods can lead to contradictory results when applied to the same data matrix (Gotelli and Ulrich 2012), and very different conclusions when applied to meta-analyses of sets of published data matrices (Wright et al. 1998, Ulrich and Gotelli 2010). However, not all combinations of metrics and null models perform well in practice. Before null models are used to analyze empirical data sets, the performance of candidate randomization algorithms and pattern metrics should be explored with benchmark testing (Gotelli 2001, Gotelli and Ulrich 2012). Such tests evaluate the behavior of null models and their associated metrics by applying them to sets of artificial stochastic matrices that have pre-defined mixtures of deterministic pattern and random associations. Benchmark analysis (Gotelli and Ulrich 2012) can help to identify combinations of null models and metrics that are most powerful for detecting certain kinds of patterns. Benchmark analysis can also steer ecologists away from certain algorithms and indices that lead to frequent rejection of the null hypothesis when applied to simple random matrices (a type I statistical error).

To date, there have been several benchmark analyses of null models for species nestedness (Ulrich and Gotelli

2007a) and species segregation (Gotelli 2001, Ulrich and Gotelli 2007b, 2010), but not for more recently developed metrics of species coherence, turnover, and boundary clumping. Both Leibold and Mikkelsen (2002) and Presley et al. (2010) proposed new metrics to be used in combination with existing null model algorithms. They analyzed patterns in artificial matrices exhibiting idealized (extreme) patterns, and they provided examples of different empirical matrices that fit different scenarios in their proposed frameworks. But in neither case have the metrics and null model algorithms been systematically tested against sets of artificial matrices created with varying mixtures of pattern and randomness. Such tests are important not only for finding combinations of metrics and algorithms that have good power for detecting pattern and avoiding type I statistical errors. In previous analyses, many presence–absence metrics for patterns of nestedness and co-occurrence have exhibited correlated responses when applied to both empirical and artificial data sets, and these correlations are sensitive to how the matrices were constructed as well as to their size, shape, and percent fill. The frameworks proposed by Leibold and Mikkelsen (2002), and Presley et al. (2010) implicitly assume that metrics of coherence, turnover, and boundary clumping describe orthogonal, independent properties of matrices. But if the metrics are strongly correlated, some of the proposed cells in their classification frameworks may be redundant or not achievable. Leibold and Mikkelsen (2002) recognized this problem and noted that they were able to identify empirical matrices that fit each of the five different scenarios they described.

In this paper, we systematically tested indices of coherence, turnover, and range boundaries proposed by Leibold and Mikkelsen (2002) and Presley et al. (2010), as well as some other candidate metrics that might also quantify these kinds of properties. We used these metrics in conjunction with several null model algorithms that have been proposed in previous benchmark tests of nestedness (Ulrich and Gotelli 2007a) and species co-occurrence (Ulrich and Gotelli 2007b). After identifying suitable combinations of metrics and null model algorithms, we tested for patterns of community organization in a meta-analysis of three sets of empirical presence–absence matrices compiled from the published literature and from unpublished sources (Ulrich and Gotelli 2010).

## Material and methods

### Artificial matrices

In order to understand the behavior of different null model algorithms and algorithms for detecting species turnover, we created two sets of artificial presence–absence matrices with specified amounts of randomness and structure. These matrices differed in their size (number of rows and columns), fraction of the matrix that was filled with species presences, and whether the cells were filled independently, or in a way that reflected specified patterns of species aggregation, segregation, or nestedness. Once these criteria were established, different matrices in each set were created stochastically, to generate variation among matrices of a particular type.

In the construction of these matrices, ‘uniform’ refers always to placement probabilities drawn from an equiprobable random distribution. ‘Exponential’ refers to placement probabilities drawn from an exponential random distribution, which was generated from the multiplication of two equiprobable random variables. Both the uniform and exponential distributions will generate variability in row or column totals, but the variability is more extreme with the exponential distribution. This variability reflects the intuition of ecologists that there is heterogeneity among species in their commonness or rarity (the row sums of the matrix), and there is heterogeneity among sites in the number of species they harbor (the column sums of the matrix). This kind of heterogeneity is probably present whether or not species interactions are important, so it is useful to incorporate it in a set of random null matrices for benchmark testing.

We constructed two types of random matrices ( $M_{\text{fixed size}}$  and  $M_{\text{fixed fill}}$ ). We generated these matrix types to span the range of empirical matrices that are typically generated by field ecologists who sample replicated assemblages at local and regional spatial scales. The  $M_{\text{fixed size}}$  matrices all had  $m = 50$  rows and  $n = 50$  columns. The proportion of cells in each matrix that was filled with 1 s was chosen from a random uniform distribution with limits of 0.1 and 0.9. In contrast, the  $M_{\text{fixed fill}}$  matrices were all filled with a fixed proportion of 0.4, but the number of rows and the number of columns for each matrix was determined by a draw of two random uniform numbers ( $m = 10$  to 100 rows and  $n = 10$  to 100 columns). For both matrix types, we generated 100 matrices each using nine different placement algorithms (Table 1; Supplementary material Appendix A1 contains examples of each matrix type):

1. *Uniform-uniform random matrices.* We filled the matrix cells randomly and independently until the desired matrix fill was reached (uniform – uniform random matrices). The probability of placing an occurrence in cell  $c_{ij}$  of the matrix was the product of two probabilities, each drawn from uniform random distribution, that specified placement in row  $i$ , column  $j$ . This matrix type represents random colonization in which all species are approximately equal in occurrence frequency, and all sites are approximately equal in suitability to be colonized.
2. *Exponential-exponential random matrices.* We filled the matrix cells randomly and independently until the

desired matrix fill was reached. The probability of placing an occurrence in cell  $c_{ij}$  of the matrix was the product of two probabilities, each drawn from a random exponential random distribution, that specified placement in row  $i$ , column  $j$ . This matrix type represents random colonization in which species differ greatly in occurrence frequency, and sites differ greatly in their suitability. The algorithm generates highly nested matrices.

3. *Exponential-uniform random matrices.* We filled the matrix cells randomly and independently until the desired matrix fill was reached. The probability of placing an occurrence in cell  $c_{ij}$  of the matrix was the product of two probabilities, one drawn from an exponential distribution that specified the probability of placement in row  $i$ , and one drawn from a uniform distribution that specified the probability of placement in column  $j$ . This matrix type represents random colonization in which species differ greatly in their occurrence frequency, but all sites are approximately equal in their suitability.
4. *Uniform-uniform segregated matrices.* We randomly selected two rows (uniform distribution) and two columns (uniform distribution) and filled the four cells with a  $2 \times 2$  checkerboard submatrix  $\{\{1,0\},\{0,1\}\}$ . We randomly selected additional pairs of rows and columns until the desired matrix fill was reached. This matrix type represents a pattern of species segregation in which all species are approximately equal in occurrence frequency, and all sites are approximately equal in suitability, but each species is part of a pair in which the pair members exclude one another in a classic checkerboard distribution. Note that, with this algorithm, each species forms a perfect checkerboard unit with one other species, but occurrence patterns should be random within the remaining  $S - 2$  species pairs of which it is a member.
5. *Uniform-uniform aggregated matrices.* We randomly selected two rows (uniform distribution) and two columns (uniform distribution) and filled the four cells with a  $2 \times 2$  aggregated submatrix  $\{\{1,1\},\{1,1\}\}$  until the desired matrix fill was reached. This matrix type represents a pattern of species aggregation in which all species are approximately equal in occurrence frequency, and all sites are approximately equal in suitability, but each species is a member of a pair that tends to occur in an aggregated distribution. Similar to algorithm no. 4, each species belongs to one such aggregated pair, but its occurrence should be random with respect to the remaining  $S - 2$  species pairs in which it is a member.
6. *Exponential-exponential segregated matrices.* We randomly selected two rows (exponential distribution) and two columns (exponential distribution) and filled the four cells with  $2 \times 2$  checkerboard submatrices  $\{\{1,0\},\{0,1\}\}$  until the desired matrix fill was reached. This matrix type represents a pattern of species segregation in which species differ greatly in their occurrence frequency, and sites differ greatly in their suitability, but species within individual pairs tend to exclude one another in a classic checkerboard distribution.
7. *Uniform-uniform compartmentalized matrices.* We sequentially filled the matrix with completely filled rectangular submatrices until the desired matrix fill was reached. At each step, the number of rows and columns

Table 1. Summary of nine different algorithms used to generate a suite of artificial matrices for testing null model algorithms and indices of community structure.

Algorithm	Heterogeneity among species	Heterogeneity among sites	Initially hypothesized association patterns
1	no	no	none
2	yes	yes	none
3	yes	no	none
4	no	no	segregation
5	no	no	aggregation
6	yes	yes	segregation
7	no	no	compartmentalized
8	yes	yes	nested
9	yes	yes	turnover

in the submatrix was determined by drawing a random uniform integer between 2 and  $m/2$  and  $n/2$ , respectively (compartmented matrices). This matrix type represents a pattern in which there are aggregated submatrices of species and sites in which occurrence is clustered.

8. *Proportional-proportional nested matrices.* As in algorithm no. 2, we filled the matrix cells randomly and independently until the desired matrix fill was reached. The probability of placing an occurrence in cell  $c_{ij}$  of the matrix was specified from two exponential random distributions. After sorting the matrix according to row and column totals, we randomly introduced a small number (5 to 10% of matrix fill) of unexpected absences in the upper left region of the matrix and a few unexpected presences in the lower right region of the matrix. Except for random placement, and the introduction of a small number of unexpected absences and presences, matrices generated this way should be strongly nested.
9. *Normal-uniform turnover matrices.* We filled the matrix until the desired fill was reached. The placement probabilities for columns were drawn from a uniform distribution. The placement probabilities for rows were drawn from a normal distribution with the mean centered at the corresponding diagonal cell of the matrix. This algorithm generates a pattern in which there is turnover of species across sites, expressed as a concentration of presences along a diagonal band when the matrix is ordered by reciprocal averaging (Leibold and Mikkelsen 2002).

In summary (Table 1), our set of artificial matrices include three algorithms in which species colonization is independent, but there are varying degrees of heterogeneity in species occurrences and site suitability (1, 2, 3), two algorithms which should generate segregated pairs of species (4, 6), and one algorithm each that should generate matrices that are aggregated (5), compartmentalized (7), nested (8), and exhibit turnover (9). Thus, our 1800 artificial matrices, with varying degrees of randomness and structure, were constructed to test the performance of different metrics and different null model algorithms. Examples of each matrix type are contained in Supplementary material Appendix A1.

## Empirical matrices

We analyzed three collections of empirical matrices that have previously been studied in species co-occurrence analysis (Wright et al. 1998, Ulrich and Gotelli 2007a, b, 2010). The first is the set of 288 biogeographic presence-absence matrices compiled by Atmar and Patterson (1995). The second is a set of 147 abundance matrices contained in Supplement A of Ulrich and Gotelli (2010). The third is a set of 36 interaction matrices (species  $\times$  species) that includes abundance data from the NCEAS database ([www.nceas.ucsb.edu/interactionweb/html/datasets.html#anemone\\_fish](http://www.nceas.ucsb.edu/interactionweb/html/datasets.html#anemone_fish)). In total, we analyzed 471 empirical species  $\times$  sites matrices. We classified these matrices as large scale (landscape to continental), small scale (local to landscape), and interaction matrices. Additionally, we sorted the matrices of

the first two sets into eight taxonomic groups (fishes, herpetiles [Amphibia and Reptilia], mammals, birds, insects, other invertebrates, plants, and protozoa).

## Metrics

We considered a set of 13 possible metrics (Table 2) that potentially can detect patterns of species segregation (4a, 4b, 4c, 12), species aggregation (4a, 5, 8, 9, 10, 12, 13), nestedness (1, 2, 3), turnover (6), range coherence (7), and clustering of species' range boundaries (11). Some metrics of species segregation, aggregation, and compositional dissimilarity (4, 5, 12, 13) do not depend on the ordering of the rows and columns of the matrix. Metrics of nestedness (1, 2, 3) are calculated from matrices that have been arranged in order of decreasing row and column sums. Metrics of turnover (6), range coherence (7), clustering (11), and some metrics of species aggregation (4b, 8, 9, 10), and species segregation (4c) are calculated from a matrix that is ordered on the basis of reciprocal averaging (correspondence analysis). Following the proposal of Leibold and Mikkelsen (2002), the first axis from a correspondence analysis orders the rows and columns of a presence-absence matrix in a way that maximizes uninterrupted strings of presences (coherence) in both the rows and columns of the matrix. Soerensen's and Simpson's indices have recently been proposed for use as metrics of species turnover and nestedness (Baselga 2010, Podani and Schmera 2011, but see Almeida-Neto et al. 2012). However, the indices are calculated using the row and column sums of a presence-absence matrix. We did not consider these indices, because we wanted to compare all indices with a fixed - fixed null model, which preserves row and column sums, and has performed well in previous analyses of species co-occurrence (Gotelli 2000) and nestedness (Ulrich and Gotelli 2007a).

## Metrics independent of matrix sorting

We estimated matrix wide species segregation by the *C-score* (Stone and Roberts 1990) that is a normalized count of the number of checkerboard submatrices ( $\{\{1,0\},\{0,1\}\}$  or  $\{\{0,1\},\{1,0\}\}$ ). The original normalization of the *C-score* was an average calculated over the number of species pairs (Stone and Roberts 1990). This normalization causes the raw *C-score* to be positively correlated with the number of sites (Ulrich and Gotelli 2007b). An additional scaling should therefore also account for the number of site combinations  $[n(n-1)/2]$ . We hence use a modified *C-score* of the form

$$C\text{-score} = \frac{4 \sum_{i,j} \begin{pmatrix} 1 & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & 1 \end{pmatrix} \vee \begin{pmatrix} 0 & \dots & 1 \\ \dots & \dots & \dots \\ 1 & \dots & 0 \end{pmatrix}}{mn(m-1)(n-1)} \quad (1)$$

$$= \frac{4 \sum_{i,j} (N_i - N_{ij})(N_j - N_{ij})}{mn(m-1)(n-1)}$$

where  $N_i$  and  $N_j$  denote total number of occurrences of species  $i$  and  $j$  and  $N_{ij}$  is the number of joint occurrences.

Table 2. Summary of 15 indices for co-occurrence analysis used in this paper.

Index	Name	Description	Matrix ordering	Pattern	Reference
1	<i>NODF</i>	normalized count of species overlap along the gradient of decreasing species richness of sites and species	row and column totals	nestedness	Almeida-Neto et al. 2008
2	<i>BR</i>	normalized count of shifts in occurrences to get a perfectly nested matrix	row and column totals	nestedness	Brualdi and Sanderson 1999
3	<i>BR<sub>turn</sub></i>	normalized count of shifts in occurrences to get a perfectly nested matrix	reciprocal averaging	nestedness	this work
4a	modified <i>C-score</i>	normalized number of 2 × 2 checkerboard units	unordered	aggregation/ segregation	Stone and Roberts 1990, this work
4b	<i>C<sub>turn</sub></i>	normalized number of 2 × 2 checkerboard units with occurrences in the turnover region of the sorted matrix	reciprocal averaging	segregation	this work
4c	<i>C<sub>seg</sub></i>	normalized number of 2 × 2 checkerboard units with occurrences not in the turnover region of the sorted matrix	reciprocal averaging	segregation	this work
5	<i>Clumping</i>	normalized number of clumped 2 × 2 submatrices	unordered	aggregation	this work
6	<i>R<sup>2</sup></i>	rank correlation of row and column position for occupied cells in an ordered matrix	reciprocal averaging	turnover	this work
7	<i>EmAbs</i>	number of embedded absences within rows	reciprocal averaging	range coherence	Presley et al. 2010
8	<i>NND</i>	average nearest neighbor distance of occupied cells	reciprocal averaging	aggregation (low values)	Clark and Evans 1954
9	<i>JoinOcc</i>	number of occupied adjacent grid cells	reciprocal averaging	aggregation	Moran 1948
10	<i>Block</i>	blocked quadrat variance	reciprocal averaging	aggregation	Hill 1973
11	<i>Morista</i>	relative abundance distribution of range boundaries among sites	reciprocal averaging	clustering of species range boundaries	Hoagland and Collins 1997
12	<i>Chao</i>	relative abundance distribution among sites	unordered	segregation	Chao et al. 2008
13	<i>Togetherness</i>	normalized number of paired 2 × 2 submatrices (togetherness score)	unordered	aggregation	Stone and Roberts 1992, this work

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

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

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

A simple metric for aggregation that is independent of matrix order is a count of the total number of 2 × 2 submatrices of the form {{1,1},{1,1}}. The *Clumping* metric is therefore the aggregation equivalent to the *C-score*. As in the case of the *C-score* we calculate *Clumping* in the normalized form divided by the total number of 2 × 2 submatrices  $[mn(m-1)(n-1)/4]$ .

The togetherness score (*Tog*) of Stone and Roberts (1992) is a complement of the *C-score* and defined from 2 × 2 submatrices of the form {{1,0},{1,0}} or {{0,1},{0,1}}. The metric *Tog* is a count of the total number of these submatrices. It is standardized with regard to matrix size by

$$Tog = \frac{4 \sum_{i,j} \begin{pmatrix} 1 & \dots & 0 \\ \dots & \dots & \dots \\ 1 & \dots & 0 \end{pmatrix} \vee \begin{pmatrix} 0 & \dots & 1 \\ \dots & \dots & \dots \\ 0 & \dots & 1 \end{pmatrix}}{mn(m-1)(n-1)} \quad (3)$$

$$= \frac{4 \sum_{i,j} N_{ij} (n - N_i - N_j + N_{ij})}{mn(m-1)(n-1)}$$

*Tog* can range from 0.0 to 1.0, with high values of *Tog* indicating positive species association. We note that *Tog* differs from the previous metrics in that it is column centered and scores of a matrix and its transpose differ.

### Metric depending of matrix sorting

With perfect spatial turnover, species occurrences should be concentrated along the diagonal of the matrix after ordination by reciprocal averaging (Leibold and Mikkelson 2002). Such a matrix can be subdivided into a nearly completely filled region (*AT*, area of turnover) and two empty regions (*AC*, areas of co-occurrence) (Fig. 1, 2: maximum species turnover). The area of *AT* equals the matrix fill defined as  $fill = w * S$ , where  $S$  is the number of species and  $w$  is the width of this sub-region of the matrix. For calculations, all cells that were intersected by the margins of the filled region were included in the *AT* region.

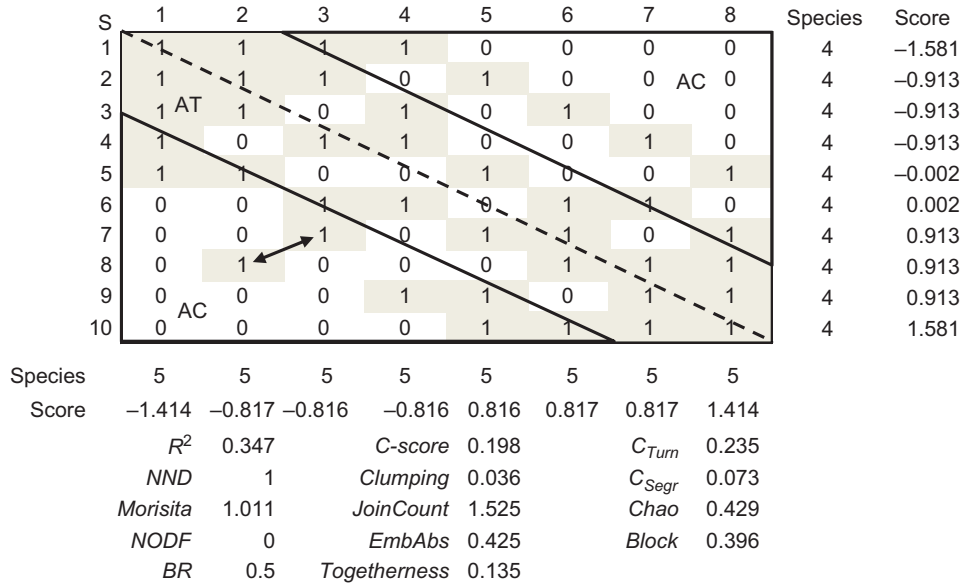


Figure 1. Ordinated example matrix with raw scores of the 14 metrics used in this study (Table 2). AT and AC: matrix areas used for calculating the  $C_{Turn}$  (AT) and the  $C_{Segr}$  (AC) scores, respectively.

Based on this spatial partitioning of the ordered matrix, the global  $C$ -score  $C_{total}$  can be decomposed into three components. The first component  $C_{Turn}$  is calculated for those checkerboard pairs for which both occurrences fall within AT.  $C_{Turn}$  measures the part of segregation that

stems from species turnover. The second component  $C_{Segr}$  is calculated only for those occurrences that fall outside AT (AC in Fig. 2).  $C_{Segr}$  measures the part of segregation that does not stem from species turnover. The third component  $C_{Mixed}$  is calculated for checkerboards for which one

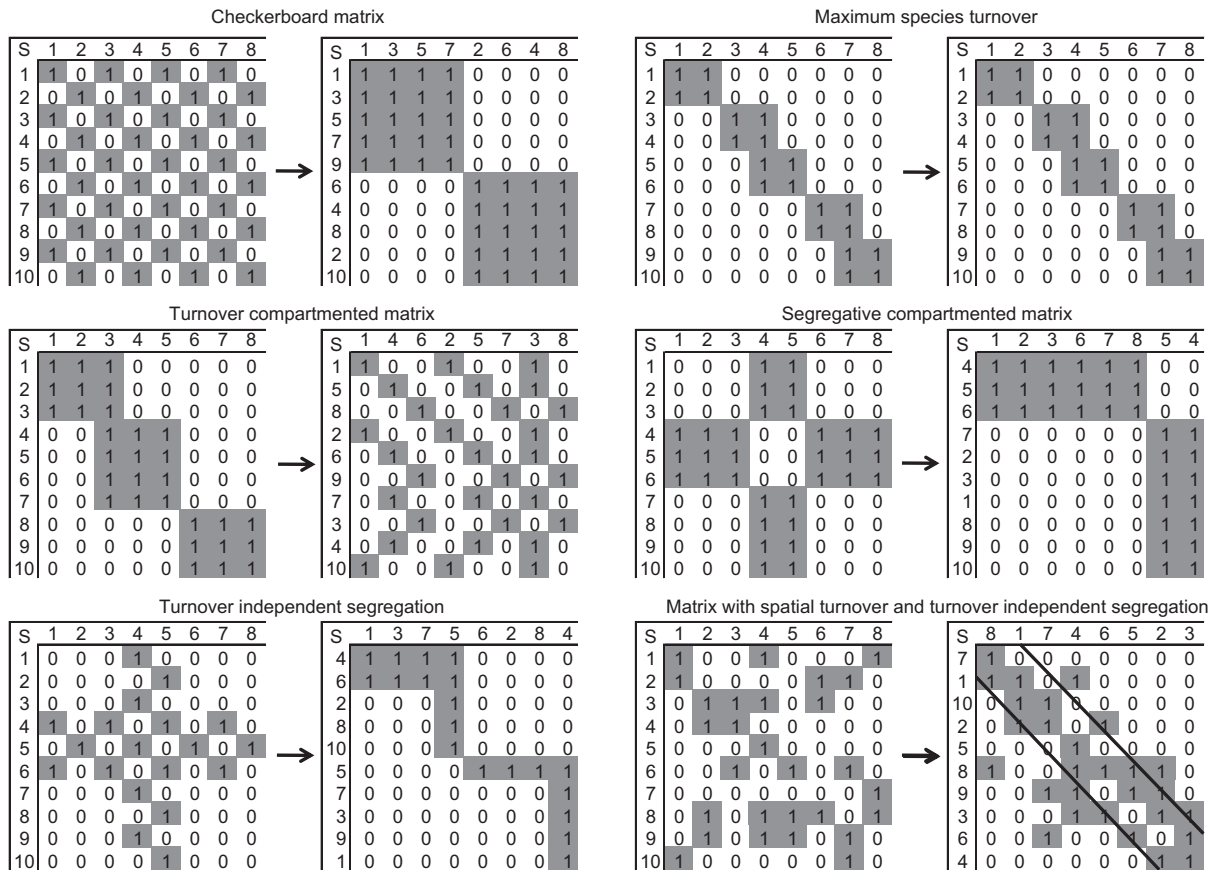


Figure 2. Six theoretical presence-absence matrices showing different degrees of co-occurrence. Left side: raw matrix; right side: ordinated matrix (order reversed for the compartmented matrix).

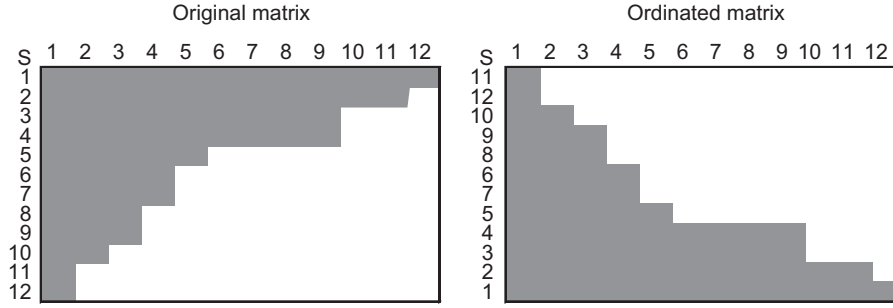


Figure 3. A perfectly nested matrix remains nested with reversed ordering of either species or sites after reciprocal averaging.

occurrences is within  $AT$  and the other is within  $AC$ .  $C_{Mixed}$  is calculated as the remaining fraction:

$$C_{Mixed} = C_{total} - C_{Turn} - C_{Segr} \quad (4)$$

In a matrix ordinated by reciprocal averaging, the position of the non-empty cells is specified by the row and column numbers. We calculated the coefficient of correlation  $r$  of the row and column numbers of non-empty cells. In this context,  $r^2$  is an index of how orderly the pattern is of the non-empty cells along the matrix diagonal. The index will be nearly 1.0 when the matrix has maximum coherence in combination with high turnover, and will be lower if the matrix has less structure. We compared the observed  $r^2$  with the  $r^2$  calculated for the randomized matrices, which were also ordinated by reciprocal averaging.

Presley et al. (2010) proposed a count of the number of embedded absences  $EmbAbs$  within rows in the ordinated matrix as a metric of coherence of species range sizes (Leibold and Mikkelsen 2002), with low numbers of embedded absences with respect to the null distribution indicating strong matrix-wide coherence (Fig. 3). We normalized  $EmbAbs$  by dividing it by the matrix fill. As in the case of  $Tog$ ,  $EmbAbs$  scores differ from those calculated after matrix transposition.

Hoagland and Collins (1997) proposed to use the Morisita index to assess clumping of species range boundaries in an ordinated matrix.  $Morisita$  is calculated as

$$Morisita = n \sum_{i=1}^n \frac{k_i}{N} \frac{k_i - 1}{N - 1} \quad (5)$$

where  $k_i$  denotes the number of starting and stopping boundaries at site  $i$ .  $N$  is the total number of starting and stopping boundaries and equals therefore two times the number of species. Large values of  $Morisita$  denote clustered boundaries of species occurrence.

To quantify patterns of species aggregation, we used a common member of the Clark and Evans (1954) nearest-neighbor class of metrics for spatial aggregation of the form

$$NND = \frac{\sum_{i=1}^m \sum_{j=1}^n d_{ij}^2}{2 \text{fill}} \quad (6)$$

where  $d_{ij}$  is the distance to the nearest neighbor (= occupied cell) in the ordinated matrix for each species occurrence

and  $\text{fill}$  is the total number of occurrences. Small values of NND indicate an aggregated pattern.

The join-count statistic (Moran 1948) is similar to the nearest neighbour method and uses sums of entries of all occupied cells in the vicinity of a given focal cell (Cliff and Ord 1973). We used the presence-absence version  $JoinOcc$  of this metric and counted the occurrences within a  $3 \times 3$  quadrature surrounding each occupied cell in the matrix.

$$JoinOcc = \frac{\sum_{i=2}^{m-1} \sum_{j=2}^{n-1} \sum_{i1=i-1}^{i+1} \sum_{j1=j-1}^{j+1} x_{ij} x_{i1j1}}{\sum_{i=1}^m \sum_{j=1}^n x_{ij}} \quad (7)$$

where  $x_{ij}$  and  $x_{i1j1}$  denote abundances (presence-absence) in cells  $(i,j)$  and  $(i1,j1)$  and the denominator is the total abundance in the matrix. High values of  $JoinOcc$  denote an aggregated pattern.

Blocked quadrature variances (Hill 1973, Dale 1999) are a diverse group of metrics that use the variances of contiguous submatrices of similar or differing size to infer non-random structure. We used a blocked quadrature variance index of the following form:

$$Block = \frac{\sum_{i=1}^{m-3} \sum_{j=1}^{n-3} [(x_{ij} - x_{i+2,j+2})^2 + (x_{ij} - x_{i,j+2})^2 + (x_{ij} - x_{i+2,j})^2]}{8L} \quad (8)$$

where  $x_{ij}$  denotes the sum of entries of the  $2 \times 2$  submatrices  $(x_{ij}, x_{i+1,j}, x_{i,j+1}, x_{i+1,j+1})$  and  $L = 3(m-3)(n-3)$  is the total number of summands. High values of  $Block$  indicate species aggregation.

We used NODF (nestedness based on overlap and decreasing fill, Almeida-Neto et al. 2008) and discrepancy BR (Brualdi and Sanderson 1999) to estimate the degree of nestedness. For this analysis, matrices were sorted according to row/columns totals. If a perfectly nested matrix is ordinated by reciprocal averaging, it retains its nested structure, but with a reversed ordering of either rows or columns (Fig. 3). Thus we tested whether the appropriately rescaled metric BR (termed  $BR_{norm}$ ) also works using ordinated matrices.

In all cases, random expectations were obtained from 1000 null matrices each that were either randomized using a null model that retains row and column totals (fixed - fixed null model, FF) or a null model that assigned occurrences equiprobably random to matrix cells (equiprobable -

equiprobable null model, EE). Randomization of the fixed – fixed null model was done with the independent swap algorithm (Gotelli and Entsminger 2003) using  $10*n*m$  swaps for each randomized matrix. Statistical significances came from the respective tail distributions at the two-sided 5% error level. Additionally we calculated Z-transformed scores ( $Z = (Obs - Exp) / StDev_{Exp}$ ; *Obs* and *Exp*: observed and expected scores,  $StDev_{Exp}$ : standard deviation of expectation). Z-scores should have values below  $-1.96$  and above  $+1.96$  at the two-sided 5% error level under the assumption that the respective null distribution is approximately normal.

## Results

### Metric and null model performance on random matrices

Regardless of the null model used, all of the metrics had a low frequency of rejection (type I error) when tested against random-uniform matrices (first row in Table 3 and 4). However, if the null model was *EE* (Table 3), all of the metrics had high frequencies of rejection when tested against random matrices with heterogeneous row and/or column totals (exponential-exponential random and exponential-uniform random). If the null model was *FF* (Table 4), type I errors were again low for exponential-exponential random matrices, but somewhat elevated for exponential-uniform random matrices. For this type of random matrix, the *BR*, *Morisita*, and *NND* indices performed best, and *Clumping* performed the worst (Table 4). When used with *FF* *C-score* and *Togetherness* identified all matrices in an identical manner. Regardless of whether the pattern was random, segregated, or aggregated both indices were consistently in the upper tail of the null distribution (Table 4).

Overall, *FF* controlled best for type I errors in purely random matrices with heterogeneous row and column totals. There was little difference in performance on matrices of fixed size (upper panel, Table 3 and 4) versus fixed fill (lower panel, Table 3 and 4). Because of the consistently poor performance of *EE* on random matrices, we restrict the description of results for structured matrices to tests with *FF*.

### Metric performance on segregated matrices

For uniformly and exponentially segregated matrices, three metrics that are independent of matrix sorting (*C-score*, *Togetherness* and *Clumping*) had the best power for detecting non-randomness regardless of the type of null model (Table 3, 4). However, *Clumping* detected at least 62% of the uniformly segregated and turnover matrices as being aggregated while *C-score* and *Togetherness* always performed identically. For the fixed-size matrices, *C-score* and *Togetherness* were less powerful at detecting non-randomness when the patterns were created by inserting segregated elements into a random matrix with exponential row and column totals (exponential segregated). In this case, the *Clumping* and *Chao* indices had better power. With the exception of *BR*,  $BR_{turn}$ , and *NND*, all metrics had very good power to detect turnover matrices

as being non-random. With these matrices,  $C_{Turn}$  and  $R^2$  performed equally well.

### Metric performance on aggregated matrices

*C-score*, *Togetherness*, *Clumping* and  $R^2$  detected at least 76% of the aggregated matrices as being non-random. (Table 4), whereas the other metrics performed much worse. For the nested matrices, the *C-score* identified all but one of the simulated matrices as non-random, but *Chao*,  $R^2$ , *EmbAbs* and *JoinOcc* also performed well for fixed size matrices. For fixed fill matrices, *C-score*, *Togetherness*, *Clumping*, *Chao*,  $C_{Seg}$ ,  $R^2$ , *BR*, *EmbAbs* and *JoinOcc* were all equally powerful in detecting pattern.  $BR_{turn}$ ,  $C_{Seg}$ , *Morisita*, *NND* and *Block* detected less than 50% of the aggregated matrices (Table 4, Supplementary material Appendix A1) as being non-random. *NODF* performed worse than *BR* in detecting a nested pattern. In the fixed fill matrices with an intermediate fill of 40%, *NODF* correctly detected 73% of the nested matrices (fixed – fixed null model), but detected only 35% of the nested fixed size matrices (Table 4). This difference shows the low power of *NODF* at high or low matrix fill. Compartmentalized matrices were detected most frequently by the *C-score*, *Togetherness*, *Clumping*,  $C_{Turn}$  and *NODF*, with minor differences between fixed fill and fixed size matrices.  $R^2$  and, to a lesser degree,  $C_{Turn}$  identified most aggregated matrices as having significant turnover (Table 4).

The exponential – exponential algorithm produced highly nested and therefore aggregated matrices. Such a matrix structure leaves a too low effective sample space for the *FF* null model and consequently *FF* was not able to detect this pattern (Table 4).

### Correlations among metrics

When tested against the set of all 1800 artificial matrices, many of the 15 indices were highly correlated in their performance (Z-scores), particularly for *EE* (Table 5, upper panel). The strongest positive correlations were between *BR* and *Clumping* ( $r = 0.98$ ) and between *BR* and *EmbAbs* ( $r = 0.86$ ). The strongest negative correlations were between *EmbAbs* and *JoinOcc* ( $r = -0.97$ ) and between *C-score* and *Clumping* ( $-0.96$ ). Of the 105 pairwise correlations among the indices, 28 had correlations  $> |0.7|$ .

For *FF*, there were only 13 such correlations (Table 5, lower panel). The largest positive correlations were between  $R^2$  and *JoinOcc* ( $r = 0.94$ ), and between *C-score* and *Togetherness* ( $r = 0.80$ ). The largest negative correlations were between *JoinOcc* and *EmbAbs* ( $r = -0.90$ ) and between  $R^2$  and *EmbAbs* ( $-0.89$ ). Although they should identify opposite patterns, there was a positive correlation ( $r = 0.55$ ) between *C-score* and *Clumping*. Z-scores of *Chao*, *NND* and *Morisita* were only weakly correlated with other metrics (Table 5), particularly with those designed to detect aggregation. These metrics and the also weak performing *Block* metric are not considered further.

### Performance of metrics on idealized matrices

Seven indices (*Clumping*, *C-score*,  $C_{Seg}$ ,  $C_{Turn}$ , *EmbAbs*,  $R^2$  and *NODF*) were used to test for patterns in the six



Table 3. Numbers of matrices (100 matrices each of nine types and two size/fill options as described in the methods section) that were detected by 15 different metrics as having scores significantly higher (score above the upper 95% confidence limit, UCL) or lower (score below the lower 95% confidence limit, LCL) than expected from the distribution of scores of the equiprobable – equiprobable null model.

Matrix pattern	Matrix type	Metric independent of row/column sorting					Ordination					Sorting according to row/column totals					Metric dependent on row/column sorting														
		C-score >UCL <LCL >UCL <LCL >UCL <LCL	Together >UCL <LCL >UCL <LCL >UCL <LCL	Clumping >UCL <LCL >UCL <LCL >UCL <LCL	Chao >UCL <LCL >UCL <LCL >UCL <LCL	C <sub>turn</sub> >UCL <LCL >UCL <LCL >UCL <LCL	C <sub>seg</sub> >UCL <LCL >UCL <LCL >UCL <LCL	BR <sub>turn</sub> >UCL <LCL >UCL <LCL >UCL <LCL	R <sup>2</sup> >UCL <LCL >UCL <LCL >UCL <LCL	NODF >UCL <LCL >UCL <LCL >UCL <LCL	BR >UCL <LCL >UCL <LCL >UCL <LCL	Block >UCL <LCL >UCL <LCL >UCL <LCL	EmbAbs >UCL <LCL >UCL <LCL >UCL <LCL	JoinOcc >UCL <LCL >UCL <LCL >UCL <LCL	MorSita >UCL <LCL >UCL <LCL >UCL <LCL	NND >UCL <LCL >UCL <LCL >UCL <LCL															
<b>Fixed size matrices</b>																															
Random	Uniform random	2	2	3	4	3	6	3	4	1	0	5	0	5	4	0	2	5	2	5	2	1	1	0	2	4	2	2	2	4	4
	Exponential-exponential random	0	100	0	100	100	0	100	0	1	94	0	100	62	34	0	86	93	5	0	100	11	59	0	100	88	2	51	0	23	7
	Exponential-uniform random	0	100	100	0	100	0	9	87	0	100	0	100	60	16	69	24	100	0	0	100	16	79	0	100	98	0	76	4	41	26
Segregated	Uniform segregated	69	0	1	1	0	43	0	2	20	0	23	0	0	55	0	0	68	51	0	4	4	3	2	2	2	5	2	2	3	
	Exponential segregated	0	89	29	2	79	0	1	79	0	42	0	55	9	10	0	37	56	0	0	56	14	1	0	55	46	1	12	5	0	10
	Turnover	0	94	0	74	100	0	87	0	81	18	0	100	73	13	100	0	87	0	0	95	0	99	0	100	100	0	0	95	0	30
Aggregated	Uniform aggregated	0	97	51	6	100	0	16	28	26	17	0	64	32	18	88	0	97	0	0	98	7	7	0	95	87	0	0	43	4	15
	Compartmented	0	95	66	17	94	0	16	76	18	60	0	89	20	58	71	8	94	1	0	92	19	33	0	84	87	0	17	29	28	14
	Nested	0	100	100	0	100	0	26	72	5	87	0	100	62	5	93	2	100	0	0	100	0	100	0	100	100	0	42	43	22	30
<b>Fixed fill matrices</b>																															
Random	Uniform random	4	6	2	4	4	3	2	0	5	2	3	6	4	1	0	5	2	4	3	4	1	2	4	1	2	3	0	3	1	1
	Exponential-exponential random	0	100	0	100	100	0	100	0	0	100	0	75	50	45	0	73	100	0	0	100	13	29	0	100	97	1	40	0	25	2
	Exponential-uniform random	0	100	99	0	100	0	2	93	0	100	0	87	54	2	79	3	100	0	0	100	0	96	0	100	99	0	100	0	37	11
Segregated	Uniform segregated	98	0	14	11	0	78	12	11	71	0	8	2	1	0	79	0	0	96	83	0	4	15	2	2	6	0	6	4	2	3
	Exponential segregated	0	94	30	4	91	0	0	85	0	88	0	57	32	5	3	27	92	0	0	93	8	2	0	76	67	0	13	8	4	7
	Turnover	0	94	0	74	100	0	87	0	81	18	0	100	64	17	100	0	87	0	0	93	0	99	0	100	100	0	0	92	0	29
Aggregated	Uniform aggregated	0	94	50	5	99	0	11	41	26	21	0	60	7	27	92	0	95	0	0	96	22	7	0	98	77	0	2	42	10	7
	Compartmented	0	95	71	16	95	0	12	81	10	77	0	77	7	73	76	10	94	0	0	95	24	44	1	92	94	0	40	23	30	23
	Nested	0	100	99	0	100	0	32	60	0	100	0	85	40	0	97	0	100	0	0	100	0	99	0	100	100	0	100	0	9	41

Table 4. Numbers of matrices (100 matrices each of nine types and two size/fill options as described in the methods section) that were detected by 15 different metrics as having scores significantly higher (score above the upper 95% confidence limit, UCL) or lower (score below the lower 95% confidence limit, LCL) than expected from the distribution of scores of the fixed – fixed null model.

Matrix pattern	Matrix type	Metric independent of row/column sorting			Ordination			Sorting according to row/column totals			Metric dependent on row/column sorting																				
		C-score	Together	Clumping	Chao	C <sub>turn</sub>	C <sub>segr</sub>	BR <sub>turn</sub>	R <sup>2</sup>	NODF	BR	Block	EmbAbs	JoinOcc	Morisita	NND															
		> UCL < LCL	> UCL < LCL	> UCL < LCL	> UCL < LCL	> UCL < LCL	> UCL < LCL	> UCL < LCL	> UCL < LCL	> UCL < LCL	> UCL < LCL	> UCL < LCL	> UCL < LCL	> UCL < LCL	> UCL < LCL	> UCL < LCL															
<b>Fixed size matrices</b>																															
Random	Uniform random	4	4	3	4	1	1	2	2	2	4	5	3	2	5	1	2	1	0	3	2	2	1	2	3	3	2	2	3	4	
	Exponential-exponential random	3	3	3	3	2	2	1	5	1	4	1	2	2	5	1	4	4	3	1	2	2	5	7	3	1	5	1	1	1	3
	Exponential-uniform random	22	12	22	12	81	0	0	34	8	37	15	20	0	18	0	66	0	51	1	6	11	1	47	0	2	14	6	8	6	0
Segregated	Uniform segregated	75	0	75	0	62	0	4	1	10	0	6	0	0	5	54	1	4	4	1	4	3	3	0	9	6	1	2	2	8	3
	Exponential segregated Turnover	38	0	38	0	0	68	55	0	2	3	4	3	19	1	1	1	40	1	4	1	1	3	3	3	1	3	9	0	4	4
Aggregated	Uniform aggregated	100	0	100	0	100	0	16	57	100	0	0	100	50	11	100	0	0	99	1	67	0	99	0	100	100	0	0	95	0	57
	Compartmented	89	0	89	0	76	0	5	7	67	0	13	5	5	7	92	0	2	11	4	2	4	10	0	62	29	0	1	8	2	8
	Nested	93	0	93	0	93	0	2	84	82	0	20	27	5	36	76	4	0	92	8	57	18	13	10	42	55	1	3	21	11	6
<b>Fixed fill matrices</b>																															
Random	Uniform random	1	2	1	1	2	4	1	5	4	5	3	6	3	2	0	4	3	4	1	1	1	2	3	0	2	2	1	3	4	3
	Exponential-exponential random	3	3	4	3	5	2	2	5	4	3	1	5	4	5	4	7	2	5	1	4	1	2	0	3	2	3	1	4	1	0
	Exponential-uniform random	22	1	22	1	86	0	0	37	1	35	0	33	2	16	0	78	0	53	2	3	23	0	52	0	2	20	1	8	4	1
Segregated	Uniform segregated	98	0	98	0	87	0	13	1	47	0	2	4	3	2	76	0	8	1	3	4	6	1	15	8	0	5	6	5	3	
	Exponential segregated Turnover	79	0	79	0	0	88	69	0	7	3	7	1	40	0	10	4	64	0	13	1	3	17	3	7	2	4	9	0	3	2
Aggregated	Uniform aggregated	91	0	91	0	93	0	26	61	90	0	0	62	46	14	87	1	0	84	2	65	13	68	1	83	82	0	7	41	5	23
	Compartmented	99	0	99	0	94	0	9	6	91	0	3	25	5	6	94	0	4	10	2	3	14	11	0	76	37	0	1	8	2	5
		95	1	95	1	95	0	4	78	91	1	11	44	3	30	80	6	0	93	4	69	11	15	8	43	73	2	4	22	20	3
		99	0	99	0	0	97	98	0	35	52	92	0	0	58	99	0	73	0	0	96	0	82	0	97	98	0	0	88	2	15

Table 5. Spearman's rank order correlation of Z-scores of 15 metrics applied to all 1800  $M_{\text{fixed-size}}$  and  $M_{\text{fixed-size}}$  matrices. Correlation coefficients  $> |0.7|$  are marked in bold.

Metrics	Block	BR	$BR_{\text{turn}}$	Chao	Clumping	C-score	$C_{\text{segr}}$	$C_{\text{turn}}$	EmbAbs	JoinOcc	Morisita	NND	Nodf	$R^2$	Together
Block															
BR	0.44	<b>0.72</b>	-0.47	0.20	-0.57	0.66	0.63	0.43	<b>0.88</b>	<b>-0.85</b>	-0.15	-0.14	-0.64	<b>-0.76</b>	-0.31
$BR_{\text{turn}}$	-0.10	0.31	-0.12	0.36	<b>-0.95</b>	<b>0.98</b>	<b>0.80</b>	0.69	<b>0.86</b>	<b>-0.90</b>	-0.32	-0.24	<b>-0.89</b>	-0.47	-0.42
Chao	-0.25	-0.20	-0.06	-0.02	-0.03	-0.08	-0.27	-0.11	-0.34	0.25	-0.11	0.08	-0.03	0.40	-0.03
Clumping	-0.23	-0.31	0.01	-0.07	-0.42	0.38	0.28	0.30	0.29	-0.34	-0.03	-0.34	-0.32	-0.14	<b>-0.73</b>
C-score	-0.54	-0.63	-0.14	0.10	0.55	<b>-0.96</b>	<b>0.79</b>	<b>0.72</b>	<b>0.80</b>	<b>-0.84</b>	0.33	0.24	<b>0.86</b>	0.36	0.49
$C_{\text{segr}}$	-0.16	-0.42	-0.32	0.39	-0.29	0.21	0.62	0.32	<b>0.83</b>	<b>-0.81</b>	-0.10	-0.34	-0.58	-0.36	-0.39
$C_{\text{turn}}$	-0.54	-0.44	0.10	-0.14	0.67	0.62	-0.29	0.42	0.42	-0.47	-0.32	-0.26	-0.67	-0.56	-0.22
EmbAbs	<b>0.88</b>	0.44	-0.12	-0.10	-0.40	-0.57	-0.02	<b>-0.70</b>	0.42	<b>-0.97</b>	-0.08	-0.17	<b>-0.70</b>	0.04	-0.23
JoinOcc	<b>-0.88</b>	-0.62	-0.04	0.23	0.36	0.69	0.26	0.61	<b>-0.90</b>	-0.43	0.12	0.20	<b>0.74</b>	<b>0.75</b>	-0.36
Morisita	0.31	0.47	0.30	-0.36	0.17	-0.30	-0.63	0.10	0.20	-0.43	0.08	-0.24	0.56	-0.17	0.44
NND	0.17	0.13	-0.05	-0.13	0.06	0.03	-0.06	-0.09	0.25	-0.18	0.08	-0.02	0.09	0.07	0.26
Nodf	0.15	0.40	0.03	0.44	<b>-0.79</b>	-0.41	0.30	<b>-0.67</b>	0.34	-0.30	-0.17	-0.02	0.30	0.30	0.04
$R^2$	<b>-0.87</b>	-0.62	-0.04	0.22	0.39	<b>0.73</b>	0.19	<b>0.72</b>	<b>-0.89</b>	<b>0.94</b>	-0.34	-0.14	-0.32	0.30	0.47
Together	-0.53	-0.61	-0.17	0.12	0.47	<b>0.80</b>	0.20	0.59	-0.57	0.67	-0.29	0.00	-0.38	<b>0.71</b>	0.29

Fixed – fixed null model

idealized matrices (checkerboard, maximum turnover, compartmented turnover, compartmented segregation, turnover-independent segregation, and a mixed matrix, Table 6, Fig. 2). In all comparisons, there were usually more significant results with the fixed – fixed null model than with the equiprobable – equiprobable null model, except for the mixed matrix, which did not differ from random for any of the indices with either null model (Table 6). For all comparisons, whenever an index was statistically significant with both null models, it deviated from the null hypothesis in the same direction, regardless of the matrix pattern.

## Empirical matrices

Of the 471 empirical matrices, 138 (29%) and 107 (23%), respectively, were detected by the *C-score* and the *Clumping* score as having a non-random structure when applied with FF (Table 7, 8). In 65 (14%) cases, both metrics jointly had scores above the upper two-sided 95% confidence limit, thus indicating simultaneous aggregation and segregation under a traditional interpretation of these scores (Table 8). Thirty-one percent of the matrices exhibited significant species turnover as measured by  $C_{\text{turn}}$  or  $R^2$ . Both of these metrics jointly classified 102 of 191 significant matrices as being spatially segregated (Table 8). The fact that only two matrices were classified differently (Table 8) shows that both metrics capture the same pattern. Of the 91 matrices that were jointly identified by the *C-score* and  $R^2$  as being segregated and having turnover, 49 were also identified by *Clumping* and 30 by *JoinCount* as being aggregated. These are matrices in which species occurrences were aggregated along the matrix diagonal after ordination. These patterns did not differ for matrices of different taxa or for matrices sampled at different spatial scales (Table 7).  $C_{\text{segr}}$  identified about 3% of the matrices as having turnover-independent segregation (Table 7). The *NODF* metric identified only seven matrices (1.5%) as being significantly nested (not shown). BR, in contrast, identified 98 matrices (21%) as being significantly nested (Table 7).

A scale-specific analysis (Table 7) revealed greater frequencies of nested, turnover and segregated/aggregated patterns for matrices assembled at the biogeographic scale versus the local scale. Compared to the standard presence-absence matrices (biogeographic, local), the interaction matrices had lower proportions of segregated and turnover patterns, but higher proportions of aggregated matrices. There were moderate differences among taxa in patterns of species co-occurrences (Table 7). Based on the  $C_{\text{turn}}$  and  $R^2$  scores, bird and plant matrices exhibited higher degrees of species turnover than herptile and invertebrate matrices. In contrast, fish, herptile, and mammal matrices appeared to have stronger patterns of checkerboard segregation than matrices for other taxa.

## Discussion

### The statistical performance of null models

Null models became popular in the late 1970s, when ecologists began to ask whether patterns of species co-occurrence

Table 6. Scores of important metrics of co-occurrence and lower and upper 95% confidence limits of the fixed – fixed and the equiprobable – equiprobable null models for the six idealized matrices of Figure 2. Statistically significant scores in bold face.

Matrix type	Metric	Observed value	Fixed – fixed null model		Equiprobable null model	
			Lower 95% confidence limit	Upper 95% confidence limit	Lower 95% confidence limit	Upper 95% confidence limit
Checkerboard	<i>Clumping</i>	0.095	0.031	<b>0.040</b>	0.044	<b>0.079</b>
	<i>C-score</i>	0.317	0.189	<b>0.208</b>	0.083	<b>0.167</b>
	<i>C<sub>Segr</sub></i>	0.050	<b>0.060</b>	0.150	0.006	0.096
	<i>C<sub>Turn</sub></i>	0.480	0.170	<b>0.270</b>	0.101	<b>0.235</b>
	<i>EmbAbs</i>	0.000	<b>0.325</b>	0.500	<b>0.100</b>	0.425
	<i>R<sup>2</sup></i>	0.577	0.240	<b>0.417</b>	0.200	<b>0.413</b>
	<i>NODF</i>	0.000	0.000	0.000	<b>0.382</b>	0.623
Maximum turnover	<i>Clumping</i>	0.004	0.000	<b>0.002</b>	0.001	0.007
	<i>C-score</i>	0.108	0.100	<b>0.105</b>	0.044	<b>0.090</b>
	<i>C<sub>Segr</sub></i>	0.000	0.000	0.040	0.000	0.040
	<i>C<sub>Turn</sub></i>	0.490	0.048	<b>0.450</b>	0.065	<b>0.305</b>
	<i>EmbAbs</i>	0.000	<b>0.100</b>	0.600	0.000	0.350
	<i>R<sup>2</sup></i>	0.907	0.590	<b>0.883</b>	0.405	<b>0.819</b>
	<i>NODF</i>	0.055	0.014	0.055	<b>0.148</b>	0.357
Turnover compartmented	<i>Clumping</i>	0.029	0.007	<b>0.013</b>	0.010	<b>0.028</b>
	<i>C-score</i>	0.188	0.145	<b>0.156</b>	0.075	<b>0.146</b>
	<i>C<sub>Segr</sub></i>	0.027	<b>0.032</b>	0.109	0.007	0.079
	<i>C<sub>Turn</sub></i>	0.527	0.151	<b>0.376</b>	0.104	<b>0.323</b>
	<i>EmbAbs</i>	0.000	<b>0.300</b>	0.600	0.000	0.500
	<i>R<sup>2</sup></i>	0.746	0.394	<b>0.645</b>	0.312	<b>0.612</b>
	<i>NODF</i>	0.048	<b>0.079</b>	0.109	<b>0.270</b>	0.502
Segregative compartmented	<i>Clumping</i>	0.052	0.025	<b>0.035</b>	0.015	<b>0.033</b>
	<i>C-score</i>	0.200	0.073	<b>0.098</b>	0.078	<b>0.151</b>
	<i>C<sub>Segr</sub></i>	0.169	0.000	<b>0.068</b>	0.000	<b>0.062</b>
	<i>C<sub>Turn</sub></i>	0.134	0.092	0.153	0.097	0.242
	<i>EmbAbs</i>	0.000	<b>0.125</b>	0.469	0.000	0.469
	<i>R<sup>2</sup></i>	0.515	0.242	<b>0.482</b>	0.278	0.557
	<i>NODF</i>	0.000	<b>0.279</b>	0.374	<b>0.299</b>	0.534
Turnover independent segregation	<i>Clumping</i>	0.005	0.001	0.006	0.000	0.006
	<i>C-score</i>	0.070	0.044	<b>0.058</b>	0.040	0.086
	<i>C<sub>Segr</sub></i>	0.037	0.000	0.049	0.000	<b>0.032</b>
	<i>C<sub>Turn</sub></i>	0.090	0.043	0.179	0.057	0.305
	<i>EmbAbs</i>	0.000	0.000	0.211	0.000	0.316
	<i>R<sup>2</sup></i>	0.718	0.490	0.770	0.368	0.849
	<i>NODF</i>	0.219	<b>0.267</b>	0.356	0.130	0.325
Mixed matrix	<i>Clumping</i>	0.009	0.006	0.013	0.007	0.021
	<i>C-score</i>	0.123	0.121	0.134	0.071	0.137
	<i>C<sub>Segr</sub></i>	0.025	0.012	0.077	0.007	0.074
	<i>C<sub>Turn</sub></i>	0.265	0.143	0.358	0.104	0.319
	<i>EmbAbs</i>	0.321	0.143	0.464	0.000	0.500
	<i>R<sup>2</sup></i>	0.503	0.392	0.679	0.326	0.639
	<i>NODF</i>	0.331	0.273	0.334	0.247	0.482

might be explained by chance, rather than by species interactions (Connor and Simberloff 1979). During the 1980s and the 1990s, ecologists began using null model methodology to test for other empirical community patterns, including body size ratios (Simberloff and Boecklen 1981), species nestedness (Patterson and Atmar 1986), favored states (Fox and Brown 1993), and, most recently, community phylogenetics (Webb et al. 2002). During the 2000s, some of these earlier analyses were re-assessed as ecologists began comparing the statistical performance of null models and metrics to artificial data sets with specified amounts of randomness and structure (Gotelli 2000, Ulrich and Gotelli 2007a, b, Ladau 2008). These tests used randomly constructed matrices to screen for algorithms with low type I errors and good power for detecting non-randomness (Gotelli 2001). This paper addresses an additional statistical issue that has only recently been considered: can null model tests uniquely

discriminate among different kinds of non-random patterns (Presley et al. 2010, Gotelli and Ulrich 2011)? Our results (Table 3, 4) are relevant to earlier analyses of species co-occurrence (Gotelli and McCabe 2002, Lehsten and Harmand 2006) and more recently proposed theoretical frameworks for analyzing patterns of coherence, species turnover, and boundary clumping (Leibold and Mikkelsen 2002, Presley et al. 2010).

### Species co-occurrence

The combination of Stone and Robert's (1990) *C-score* with the fixed-fixed randomization algorithm (Gotelli 2000) has become a standard method for testing for patterns of species co-occurrence. Several studies have shown that this combination rejects the null hypotheses infrequently when tested against random matrices with heterogeneous row

Table 7. Percentages of significant scores within each category (two sided 5% error level of the fixed – fixed null model distribution) of seven metrics applied to 471 empirical presence–absence matrices.

Scale	Pattern Metric	Co-occurrence C-score		Co-occurrence Clumping		Co-occurrence C <sub>Seg</sub>		Turnover C <sub>Turn</sub>		Turnover R <sup>2</sup>		Coherence EmbAbs		Nestedness BR		
		aggregated	segregated	aggregated	segregated	aggregated	segregated	aggregated	segregated	aggregated	segregated	positive	negative	aggregated	segregated	nested
Biogeographic	195	0.51	42.05	1.54	27.69	3.59	5.13	3.08	38.97	0.51	39.49	28.21	1.54	27.18	0.51	
Local	240	0.83	20.42	2.08	17.50	4.17	1.25	1.25	23.75	2.08	26.25	17.08	1.25	15.00	1.25	
Interaction	36	0.00	19.44	0.00	30.56	11.11	5.56	2.78	33.33	0.00	25.00	11.11	0.00	11.11	2.78	
Total	471	0.64	29.30	1.70	22.72	4.46	3.18	2.12	30.79	1.27	31.63	21.23	1.27	19.75	1.06	
<b>Taxon</b>																
Fish	12	0.00	25.00	0.00	25.00	8.33	8.33	0.00	33.33	0.00	33.33	33.33	0.00	33.33	0.00	
Amphiba/Reptiles	25	4.00	32.00	0.00	24.00	4.00	8.00	8.00	24.00	0.00	16.00	20.00	0.00	32.00	0.00	
Birds	52	0.00	51.92	0.00	36.54	3.85	1.92	0.00	46.15	0.00	40.38	26.92	3.85	34.62	0.00	
Mammals	65	0.00	43.08	1.54	24.62	7.69	7.69	1.54	20.00	0.00	38.46	30.77	0.00	29.23	0.00	
Insects	214	0.47	18.69	2.34	15.42	2.34	0.93	2.34	25.70	2.34	28.50	13.08	1.40	14.02	0.93	
Other invertebrates	26	0.00	26.92	7.69	23.08	11.54	3.85	3.85	30.77	0.00	34.62	34.62	0.00	11.54	3.85	
Plants	30	3.33	63.33	0.00	36.67	0.00	3.33	0.00	66.67	0.00	46.67	50.00	0.00	13.33	3.33	
Unicellular organisms	8	0.00	12.50	0.00	12.50	0.00	0.00	0.00	12.50	0.00	25.00	12.50	0.00	12.50	0.00	
Total	432	0.69	30.79	1.85	21.99	3.94	3.01	2.08	30.32	1.16	32.41	22.22	1.16	20.14	0.93	

and column totals, and has good power to detect checkerboard pairs embedded in noisy data matrices (Gotelli 2000, Ulrich and Gotelli 2007a). By this criterion, many empirical matrices exhibit significantly more species segregation than expected by chance (Gotelli and McCabe 2002).

However, Stone and Roberts (1990) warned that matrices with segregated species pairs must necessarily contain some species pairs that are aggregated. Indeed, in any presence–absence matrix positive and negative species associations are intimately linked. Consider three  $4 \times 4$  submatrices of the form

$$M = \begin{matrix} A \\ B \\ C \\ D \end{matrix} \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \end{pmatrix} \text{ and}$$

$$N = \begin{matrix} A \\ B \\ C \\ D \end{matrix} \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \end{pmatrix} \text{ and}$$

$$O = \begin{matrix} A \\ B \\ C \\ D \end{matrix} \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

where two species  $A$  and  $B$  form a checkerboard pair. In matrix  $M$ , any new species  $C$  that avoids either  $A$  or  $B$  will form a positive association with the complementary species of the form  $\{\{1,1\},\{1,1\}\}$  as does any new species  $D$  that is either linked with  $C$  or occurs everywhere. In matrix  $N$ , segregation of  $A$ ,  $B$ , and  $C$  is not perfect. Nevertheless, any new species  $D$ , which is at least as abundant as  $A$ ,  $B$ , or  $C$  will add at least one new positive association. Only new species with fewer occurrences than  $A$  and  $B$  might not add new positive associations as shown by matrix  $O$ . Hence, in equiprobably-filled matrices with a fill of less than 0.5, raw scores of the  $C$ -score and  $Clumping$  will always be positively correlated (Fig. 4A). Because the  $C$ -score has its theoretical maximum at 50% matrix fill matrices with higher fill will inevitably add more positive than negative associations and positive and negative associations will be negatively correlated (Fig. 4A). This correlation will be less apparent in matrices where species highly differ in the total numbers of occurrences (Fig. 4B). Nevertheless the  $C$ -score has always a tendency to be also a metric for positive species association (aggregation).

Specifically, any sufficiently filled perfect checkerboard matrix can be reordered by reciprocal averaging to reveal a perfectly aggregated pattern (Stone and Roberts 1990). Such a matrix has also a high degree of species turnover (Table 8). Irrespective of null model, the  $C$ -score identifies such a matrix as having far more checkerboard pairs than expected by chance while the  $Clumping$  score,  $C_{Seg}$  and  $EmbAbs$  identify the same matrix as being aggregated. The significant  $R^2$  value points to turnover (Table 8, Fig. 2). Hence

Table 8. Pair wise comparisons of seven co-occurrence and nestedness scores to identify patterns in 471 empirical presence–absence matrices. Note that for the *EmbAbs* score values above the upper two-sided 95% confidence limit indicate scatter of range. For  $R^2$  and  $C_{turn}$  values above upper 95% confidence limits indicate species turnover.

Metric pair	No. of significant scores			No. of significant scores			
	<i>C-score</i>	<i>Clumping</i>	joint indication of segregation/ aggregation	<i>C-score</i>	<i>Clumping</i>	joint indication of aggregation/ segregation	Contrasting results
<i>C-score</i> - <i>Clumping</i>	138	107	65	3	8	0	5
$R^2$ - $C_{turn}$			joint indication of turnover	$R^2$	$C_{turn}$	joint indication of no turnover	Contrasting results
	149	145	102	6	10	2	2
$R^2$ - <i>C-score</i>			joint indication of turnover	$R^2$	<i>C-score</i>	joint indication of no turnover	Contrasting results
	149	138	91	6	3	0	1
<i>EmbAbs</i> - <i>Clumping</i>			joint indication of aggregation	<i>EmbAbs</i>	<i>Clumping</i>	joint indication of segregation	Contrasting results
	100	107	58	6	8	0	3
<i>EmbAbs</i> - <i>C-score</i>			joint indication of coherence/ segregation	<i>EmbAbs</i>	<i>C-score</i>	joint indication of scatter/ aggregation	Contrasting results
	100	138	64	6	3	0	2
<i>EmbAbs</i> - $R^2$			joint indication of coherence/ turnover	<i>EmbAbs</i>	$R^2$	joint indication of scatter	Contrasting results
	100	149	75	6	6	1	0
<i>NODF</i> - <i>BR</i>			joint indication of nestedness	<i>NODF</i>	<i>BR</i>	joint indication of antinestedness	Contrasting results
	7	93	1	98	5	1	40

such a matrix can be viewed as combining three seemingly exclusive patterns (segregation, aggregation, turnover).

While this link is apparent for raw scores it holds also when comparing scores with those obtained from a null distribution that constraints row and column totals. The fixed – fixed null model does not eliminate the intimate link between the numbers of aggregated and checkerboard  $2 \times 2$  submatrices. As a result, the significance levels of the *C-score* and *Clumping* will still be positively correlated (Fig. 5). With the exception of the nested and proportionally segregated matrices (with their limited null spaces) high *Z*-scores of *Clumping* were in all cases combined with respective high *C-scores* irrespective of matrix size and fill.

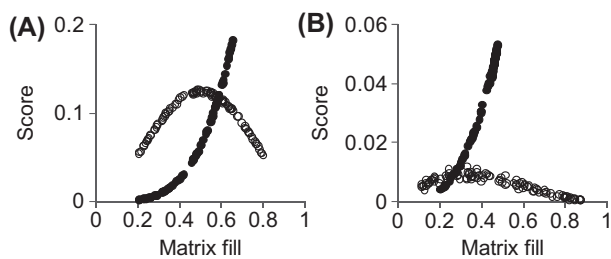


Figure 4. (A) In uniform – uniform matrices raw scores of the *C-score* (open dots) and the clumping score (black dots) are positively linked up to a matrix fill of 50%. Above 50% fill the clumping score increases with fill while the *C-score* decreases. (B) In proportional – proportional matrices the region of fill where both scores are positively linked is shifted towards low fill. Results from 100  $M_{fixed\ size}$  matrices.

The same holds for the relationship between the *C-score* and *Togetherness*. The latter was designed to measure joint occurrences and therefore niche similarity (Stone and Roberts 1992), and the raw scores of both metrics do behave properly. Dormann et al. (2009) showed that these metrics capture the antagonistic patterns of negative and positive association in ecological interaction matrices. However, when applied in combination with the fixed – fixed null model (Table 4) both metrics give identical results that do not allow for an unequivocal matrix classification.

Perhaps the most useful step forward is to recognize that the *C-score* represents an average calculated over all possible species pairs. Analysis of which individual species pairs show aggregation or segregation provides insight into the performance of the index. Although the interpretation of the *C-score* is complicated when it is applied to all the species pairs in a matrix, it is straightforward when applied to a single pair of species. In a meta-analysis of the Patterson and Atmar empirical matrices, Gotelli and Ulrich (2010) found that consistent non-randomness (as measured by the *C-score*) was caused by a preponderance of negatively associated species pairs. Even using the most conservative criterion (based on an empirical Bayes approach) for classifying species pairs as statistically significant, there were approximately four times as many species pairs exhibiting segregation as aggregation. At least for these empirical matrices, the frequencies of aggregation and segregation measured among species pairs were consistent with the overall *C-score*. However, future analyses should dissect the *C-score* pattern of non-randomness to identify particular pairs of species with positive and negative associations.

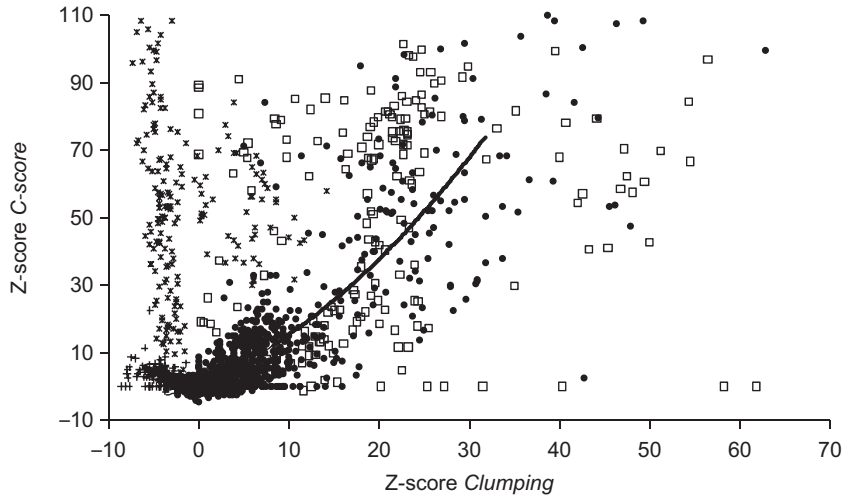


Figure 5. Relationship between Z-scores (fixed – fixed null model) of the Clumping score and the C-score ( $M_{\text{fixed size}}$  and  $M_{\text{fixed fill}}$  combined). Open squares: Turnover matrices, Stars: nested matrices, plus signs: proportionally segregated matrices, black dots: all other matrices. Quadratic regression for black dots:  $R^2 = 0.79$ .

### Coherence, species turnover and boundary clumping

Our analysis in general demonstrated that metrics that use marginal totals for calculation such as *NODF*, *Chao* or *Morisita* perform poorly when used with null models that constrain these totals (Table 4). Recently, these metrics (Almeida-Neto et al. 2008, Chao et al. 2008, Leibold and Mikkelsen 2002, Presley et al. 2010) and multispecies extensions of the Soerensen and Jaccard indices of species overlap (Baselga et al. 2007, Baselga 2010) have been proposed to quantify patterns of nestedness or species co-occurrence. These authors showed that the metrics behave well as raw scores and with null models that assume equiprobable occurrence probabilities. However, these metrics have not been tested rigorously with more constrained null models and should be used with caution (Almeida-Neto et al. 2012) in combination with the fixed – fixed null model.

We also showed that metrics based on nearest neighbor distances (*NND*) and local variances differences (*JoinCount*, *Block*) appeared also to be of little value. In matrices of higher fill (> 50%) most occurrences have at least one direct neighbor. As a consequence, metrics of the *NND* family (Clark and Evans 1954) achieve their minimum value and therefore lose any statistical power. These nearest-neighbor metrics appear to be of value only in very large matrices of relatively low fill. At least in presence–absence matrices, metrics of the *JoinCount* and *Block* family suffer from the same restrictions.

Our analysis did not point to any simple classification schemes of ecological presence–absence matrices. In comparison to the fixed – fixed null models, *NODF* detected 241 of the 1800  $M_{\text{fixed fill}}$  and  $M_{\text{fixed size}}$  matrices as being significantly nested. Of these matrices, 133 were also detected by  $R^2$  as having significant species turnover and 208 by the *C-score* as being segregated. Only 14 of these 241 significantly nested matrices were detected by *Clumping* as being aggregated. In turn, 205 of these 241 nested matrices were identified as being significantly not aggregated. Further,  $C_{\text{Segr}}$  detected 238 matrices as having

turnover independent (checkerboard) segregation and  $R^2$  identified 877 as having significant turnover. As many as 194 of these matrices were jointly detected by  $C_{\text{Segr}}$  and  $R^2$  as having significant turnover and turnover independent segregation. These were mainly matrices of a high degree of nestedness (not shown). This mix of classifications in combination with the apparent differences in the power of certain metrics makes any clear-cut classification scheme for ecological matrices challenging.

Previous attempts to classify ecological presence–absence matrices and to link this classification with ecological processes (Leibold and Mikkelsen 2002, Presley et al. 2010, Podani and Schmera 2011, but see Gotelli and Ulrich 2011, 2012) were largely based on idealized artificial matrices. These approaches particularly focused on the gradient between nestedness and species turnover and aggregation and segregation. Our results highlight the intimate link between an aggregated and a segregated pattern. A matrix might be highly aggregated and segregated (Fig. 2) as had been demonstrated previously for a perfect checkerboard matrix (Stone and Roberts 1990). Because a nested subset pattern is a special form of species aggregation, nestedness may not be really distinct from species segregation. Almeida-Neto et al. (2007) and Podani and Schmera (2011) focused on the gradient from nestedness to antinestedness and equated antinestedness with species segregation. We showed that species segregation can be divided in true species turnover and turnover independent (diffuse) segregation. Thus classification schemes that focus on the aggregation–segregation dichotomy might miss important aspects of matrix structure.

Species turnover is intimately linked to  $\beta$ -diversity (Tuomisto 2010) and refers to the replacement in species composition across sites. Temporal turnover is an explicit component of the theory of island biogeography (MacArthur and Wilson 1963, Simberloff 1976) but spatial turnover (among sites) is more often linked to environmental gradients, and to vicariant speciation triggered by environmental barriers. For these reasons, spatial turnover is a common pattern in biogeography (Tuomisto 2010). The second,

turnover-independent, part of species segregation has gained much less attention, although it is the classical pattern on which most discussions on the assembly rule concept focused (Diamond 1975, Weiher and Keddy 1999). Within the competition framework, we might speculatively think of  $C_{Segr}$  as capturing the effect of actual competition as a force towards species segregation and of  $C_{Turn}$  as quantifying 'the ghost of competition past' (Connell 1980). Our analysis of empirical presence-absence matrices clearly points to turnover as the dominant pattern of species segregation (Table 6).

Irrespective of spatial scale (local or regional) and type of matrix (biogeographic – interaction) less than 6% of matrices were significantly diffusely segregated but more than 20% showed significant turnover. In accordance with expectation the degree of turnover was highest for matrices at regional to continental scales.

Our results point to a gradient from a nested subset pattern to a turnover pattern (Table 5) as has been previously hypothesized (Gaston and Blackburn 2000, Baselga 2010). Nestedness was originally interpreted as an ordered sequence of species extinction (Atmar and Patterson 1993, Ulrich et al. 2009) whereas turnover has been interpreted as an ordered sequence of species replacement across an environmental gradient (Tuomisto 2010). In combination with the fixed – fixed null model  $C_{Segr}$  identified at least 58% of the nested matrices as being also checkerboarded (Table 5).  $C_{Segr}$  is not strongly correlated to *BR* and *NODF* (Table 6). Thus, a nestedness metric in combination with the *C-score* that does not differentiate between the two types of segregation might give an incorrect classification.

## Summary

Our analysis showed that metrics designed to quantify different and sometimes exclusive matrix properties are often correlated (Table 5). Undesired correlations, even if statistically not significant, might strongly influence the interpretation of pattern. Such undesired correlations are particularly obvious using the fixed – fixed null model. For instance *Clumping* correlated positively with the *C-score*, *Tog*,  $R^2$  and negatively with *NODF*, while *EmbAbs* was negatively correlated with the *C-score* and *Clumping*. This seemingly contradictory behavior casts further doubts on the utility of simple classification schemes (Leibold and Mikkelsen 2002, Presley et al. 2010). The use of multiple metrics – perhaps dissected for pairs of individual species or sites – may be necessary for valid pattern identification.

*Acknowledgements* – WU was in part supported by grants from the Polish Science Ministry (KBN, 3 P04F 034 22, KBN 2 P04F 039 29). NJG was supported by grants from the US National Science Foundation (DEB-0541936) and the Dept of Energy (022821).

## References

Almeida-Neto, M. et al. 2007. On nestedness analyses: rethinking matrix temperature and anti-nestedness. – *Oikos* 116: 716–722.

- Almeida-Neto, M. et al. 2008. A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. – *Oikos* 117: 1227–1239.
- Almeida-Neto, M. et al. 2012. Rethinking the relationship between nestedness and beta diversity: a comment on Baselga (2010). – *Global Ecol. Biogeogr.* 21: 772–777.
- Atmar, W. and Patterson, B. D. 1993. The measure of order and disorder in the distribution of species in fragmented habitat. – *Oecologia* 96: 373–382.
- Atmar, W. and Patterson, B. D. 1995. The nestedness temperature calculator: a Visual Basic program, including 294 presence-absence matrices. – AICS Research, Univ. Park, NM and Field Museum, Chicago <<http://aics-research.com/nestedness/tempcalc.html>>.
- Baselga, A. 2010. Partitioning the turnover and nestedness components of beta diversity. – *Global Ecol. Biogeogr.* 19: 134–143.
- Baselga, A. et al. 2007. A multiple-site similarity measure independent of richness. – *Biol. Lett.* 3: 642–645.
- Brualdi, R. A. and Sanderson, J. G. 1999. Nested species subsets, gaps, and discrepancy. – *Oecologia* 119: 256–264.
- Chao, A. et al. 2008. A two stage probabilistic approach to multiple-community similarity indices. – *Biometrics* 64: 1178–1186.
- Clark, P. J. and Evans, F. C. 1954. Distance to nearest neighbour as a measure of spatial relationships in populations. – *Ecology* 35: 445–453.
- Cliff, A. D. and Ord, J. K. 1973. *Spatial autocorrelation*. – Pion, London.
- Connell, J. H. 1980. Diversity and the coevolution of competitors, or the ghost of competition past. – *Oikos* 35: 131–138.
- Connor, E. F. and Simberloff, D. 1979. The assembly of species communities: chance or competition? – *Ecology* 60: 1132–1140.
- Dale, M. R. T. 1999. *Spatial pattern analysis in plant ecology*. – Cambridge Univ. Press.
- Darlington, P. J. 1957. *Zoogeography: the geographical distribution of animals*. – Wiley.
- Diamond, J. M. 1975. Assembly of species communities. – In: Cody, M. L. and Diamond, J. M. (eds), *Ecology and evolution of communities*. Harvard Univ. Press, pp. 342–444.
- Dormann, C. F. et al. 2009. Indices, graphs and null models: analysing bipartite ecological networks. – *Open Ecol. J.* 2: 7–24.
- Fox, B. J. and Brown, J. H. 1993. Assembly rules for functional groups in North American desert rodent communities. – *Oikos* 67: 358–370.
- Gaston, K. J. and Blackburn, T. M. 2000. Pattern and process in macroecology. – Blackwell.
- 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 Ecol. Biogeogr.* 10: 337–343.
- Gotelli, N. J. and Entsminger, G. L. 2003. Swap algorithms in null model analysis. – *Ecology* 84: 532–535.
- Gotelli, N. J. and Graves, G. R. 1996. *Null models in ecology*. – Smith. Inst. Press, Washington D.C.
- Gotelli, N. J. and McCabe, D. J. 2002. Species co-occurrence: a meta-analysis of J. M. Diamond's assembly rules model. – *Ecology* 83: 2091–2096.
- Gotelli, N. J. and McGill, B. J. 2006. Null versus neutral models: what's the difference? – *Ecography* 29: 793–800.
- Gotelli, N. J. and Ulrich, W. 2010. The empirical Bayes distribution as a tool to identify non-random species associations. – *Oecologia* 162: 463–477.
- Gotelli, N. J. and Ulrich, W. 2011. Over-reporting bias in null model analysis: a response to Fayle and Manica (2010). – *Ecol. Modell.* 222: 1337–1339.
- Gotelli, N. J. and Ulrich, W. 2012. Statistical challenges in null model analysis. – *Oikos* 21: 171–180.



- Hill, M. O. 1973. The intensity of spatial pattern plant communities. – *J. Ecol.* 61: 225–235.
- Hoagland, B. W. and Collins, S. L. 1997. Gradient models, gradient structure and hierarchical structure in plant communities. – *Oikos* 78: 23–30.
- Holyoak, M. et al. (eds) 2005. *Metacommunities: spatial dynamics and ecological communities*. – Univ. of Chicago Press.
- Hubbell, S. P. 2001. *The unified theory of biogeography and biodiversity*. – Princeton Univ. Press.
- Ladau, J. 2008. Validation of null model tests using Neyman–Pearson hypothesis testing theory. – *Theor. Ecol.* 1: 241–248.
- Lehsten, V. and Harmand, P. 2006. Null models for species co-occurrence patterns: assessing bias and minimum iteration number for the sequential swap. – *Ecography* 29: 786–792.
- Leibold, M. A. and Mikkelsen, G. M. 2002. Coherence, species turnover, and boundary clumping: elements of meta-community structure. – *Oikos* 97: 237–250.
- MacArthur, R. H. and Wilson, E. O. 1963. An equilibrium theory of insular zoogeography. – *Evolution* 17: 373–387.
- McCoy, E. D. and Heck Jr., K. L. 1987. Some observations on the use of taxonomic similarity in large-scale biogeography. – *J. Biogeogr.* 14: 79–87.
- Moran, P. A. P. 1948. The interpretation of statistical maps. – *J. R. Stat. Soc. B* 10: 243–251.
- Patterson, B. D. and Atmar, W. 1986. Nested subsets and the structure of insular mammalian faunas and archipelagos. – *Biol. J. Linn. Soc.* 28: 65–82.
- Pielou, D. P. and Pielou, E. C. 1968. Association among species of infrequent occurrence: the insect and spider fauna of *Polyporus betulinus* (Bulliard) Fries. – *J. Theor. Biol.* 21: 202–216.
- Podani, J. and Schmera, D. 2011. A new conceptual and methodological framework for exploring and explaining pattern in presence–absence data. – *Oikos* 120: 1625–1638.
- Presley, S. J. et al. 2010. A comprehensive framework for the evaluation of metacommunity structure. – *Oikos* 119: 908–917.
- Simberloff, D. 1976. Species turnover and equilibrium island biogeography. – *Science* 194: 572–578.
- Simberloff, D. S. and Boecklen, W. 1981. Santa Rosalia reconsidered: size ratios and competition. – *Evolution* 35: 1206–1228.
- Stone, L. and Roberts, A. 1990. The checkerboard score and species distributions. – *Oecologia* 85: 74–79.
- Stone, L. and Roberts, A. 1992. Competitive exclusion, or species aggregation? An aid in deciding. – *Oecologia* 91: 419–424.
- Tuomisto, H. 2010. A diversity of beta diversities: straightening up a concept gone awry. Part 1. Defining beta diversity as a function of alpha and gamma diversity. – *Ecography* 33: 2–22.
- Ulrich, W. and Gotelli, N. J. 2007a. Null model analysis of species nestedness patterns. – *Ecology* 88: 1824–1831.
- Ulrich, W. and Gotelli, N. J. 2007b. Disentangling community patterns of nestedness and species co-occurrence. – *Oikos* 116: 2053–2061.
- Ulrich, W. and Gotelli, N. J. 2010. Null model analysis of species associations using abundance data. – *Ecology* 91: 3384–3397.
- Ulrich, W. et al. 2009. A consumer's guide to nestedness analysis. – *Oikos* 118: 3–17.
- Webb, C. O. et al. 2002. Phylogenies and community ecology. – *Annu. Rev. Ecol. Syst.* 33: 475–505.
- Weihert, E. and Keddy, P. A. (eds) 1999. *Ecological assembly rules: perspectives, advances, retreats*. – Cambridge Univ. Press.
- Wright, D. H. et al. 1998. A comparative analysis of nested subset patterns of species composition. – *Oecologia* 113: 1–20.

Supplementary material (available online as Appendix O20325 at <[www.oikosoffice.lu.se/appendix](http://www.oikosoffice.lu.se/appendix)>). Appendix A1 (Figure A1.docx: Typical presence–absence matrices used for testing the performance of metric – null model combination. Figure A2.docx: Relationships between the  $R^2$  metric and nine metrics of matrix structure (raw scores and Z-transforms of two null models) for the fixed size matrices. Figure A3.docx: Relationships between the  $R^2$  metric and nine metrics of matrix structure (raw scores and Z-transforms of two null models) for the fixed fill matrices. Figure A4.docx: Relationships between the  $R^2$  metric and nine metrics of matrix structure (raw scores and Z-transforms of two null models) for all 900 artificial fixed fill and 900 fixed size matrices.