Ecology, 99(1), 2018, pp. 103–115

Bi-dimensional null model analysis of presence-absence binary matrices

Giovanni Strona,[1,4] Werner Ulrich,[2] and Nicholas J. Gotelli[3]

[1]*Directorate D – Sustainable Resources, Joint Research Centre, European Commission, Via E. Fermi 2749, 21027 Ispra (VA) Italy*
[2]*Ecology and Biogeography, Nicolaus Copernicus University in Toruń, Lwowska 1, 87-100 Toruń Poland*
[3]*Department of Biology, University of Vermont, Burlington, Vermont 05405 USA*

Abstract. Comparing the structure of presence/absence (i.e., binary) matrices with those of randomized counterparts is a common practice in ecology. However, differences in the randomization procedures (null models) can affect the results of the comparisons, leading matrix structural patterns to appear either "random" or not. Subjectivity in the choice of one particular null model over another makes it often advisable to compare the results obtained using several different approaches. Yet, available algorithms to randomize binary matrices differ substantially in respect to the constraints they impose on the discrepancy between observed and randomized row and column marginal totals, which complicates the interpretation of contrasting patterns. This calls for new strategies both to explore intermediate scenarios of restrictiveness in-between extreme constraint assumptions, and to properly synthesize the resulting information. Here we introduce a new modeling framework based on a flexible matrix randomization algorithm (named the "Tuning Peg" algorithm) that addresses both issues. The algorithm consists of a modified swap procedure in which the discrepancy between the row and column marginal totals of the target matrix and those of its randomized counterpart can be "tuned" in a continuous way by two parameters (controlling, respectively, row and column discrepancy). We show how combining the Tuning Peg with a wise random walk procedure makes it possible to explore the complete null space embraced by existing algorithms. This exploration allows researchers to visualize matrix structural patterns in an innovative bi-dimensional landscape of significance/effect size. We demonstrate the rational and potential of our approach with a set of simulated and real matrices, showing how the simultaneous investigation of a comprehensive and continuous portion of the null space can be extremely informative, and possibly key to resolving longstanding debates in the analysis of ecological matrices.

Key words: *co-occurrence; C-score; ecological networks; effect size; nestedness; NODF;* P-*value.*

Introduction

The problem of identifying non-random structural patterns of species distribution compared to a background expectation of ecological noise has received much attention, leading to the development of statistical standards for pattern identification in binary presence-absence matrices (rows = species, columns = sites; Gotelli and Graves 1996, Gimenez et al. 2014). The most popular approach for this purpose is null model analysis (Gotelli 2000, Gotelli and Ulrich 2012), which typically consists of quantifying a structural pattern with an appropriate metric, and then comparing the observed measure with those obtained in a set of randomized (i.e., null) versions of the original matrix. The resulting effect size quantifies the strength of the observed pattern, and the simulation generates a frequentist estimate of statistical significance. This procedure, however, raises several issues at each step of analysis (Gotelli and Ulrich 2012): (1) the chosen metric has to quantify correctly the pattern of interest (Almeida-Neto et al. 2008, Baselga 2010, Tuomisto 2010); (2) the null model has to randomize the pattern of interest, leaving other confounding patterns unaffected (Gotelli and Graves 1996); (3) the null model should not be vulnerable to Type I errors, that is, it should not produce statistically significant results for a stochastic matrix generated by simple random processes; (4) the null model should not be vulnerable to Type II errors, and be therefore able to correctly reject the null hypothesis for a matrix in which structural patterns are caused by species interactions or other factors of ecological interest (Ulrich et al. 2017).

Null models for randomizing presence-absence matrices retain, at minimum, the observed number of rows (usually species) and columns (usually sample sites). Furthermore, most null models retain also the matrix fill (i.e., the ratio of the total number of occurrences to the number of matrix cells). The key difference among existing null models is therefore how they treat marginal sums of the matrix, which represent the occurrence frequencies of the different species (= row sums) and the number of species per site (= column sums).

Manuscript received 27 June 2017; revised 8 September 2017; accepted 4 October 2017. Corresponding Editor: Perry de Valpine.
[4] E-mail: giovanni.strona@ec.europa.eu

Marginal sums can typically be: (1) unconstrained, with the probability of a cell in the null matrix to be occupied completely independent from the corresponding row and/or column total of the original matrix; (2) perfectly constrained, with marginal totals of the null matrix that exactly match those of the original matrix; or (3) proportionally constrained, with marginal totals of the null matrix that match *on average* those of the original matrix, but differ slightly for each null matrix that is generated.

Because these three classes of marginal total constraints can be applied independently to rows and columns, a basic set of nine null model algorithms appears (Gotelli 2000). It is a well-known issue that these nine algorithms can sometimes generate contrasting results when applied to the same matrix (Gotelli and Ulrich 2012). Furthermore, they are also characterized by inevitable tradeoffs between Type I (rejecting a true null hypothesis) and Type II (failing to reject a false null hypothesis) statistical errors, with the equiprobable-equiprobable algorithm (which does not constrain neither row nor column marginal totals) being most vulnerable to Type I statistical error (Wright et al. 1997), and the fixed-fixed algorithm (which constrains both row and column marginal totals) being most vulnerable to Type II statistical error (Gotelli 2000).

The choice of one null model over another is usually made on the basis of ecological assumptions, and of benchmark performance testing on artificial matrices that contain specified amounts of structure and randomness (Gotelli 2000, Gotelli and Ulrich 2012). By both of these criteria, the fixed-fixed algorithm has proven most popular. It is fairly robust to Type I errors when applied to random matrices with heterogeneous row and column totals. The assumptions of the fixed-fixed algorithm also match the field ecologist's intuition that there are probably intrinsic differences between species' occurrence frequencies that do not reflect species interactions, and that sites can vary in total species richness because of factors that do not reflect resource limitation (Gotelli and Ulrich 2012). Yet, the fixed-fixed null model may be overly conservative and prone to type II errors in comparison to other null models that relax the assumptions on marginal totals (Ulrich and Gotelli 2012, Strona and Fattorini 2014).

The nine algorithms, however, represent only a subset of possible null models that could be ideally defined depending on how much they constrain marginal totals. To better illustrate this point, we can define a two-dimensional landscape with the dimensions corresponding, respectively, to the average discrepancy in row and column marginal totals between the original matrix and its randomized counterparts obtained using a particular null model. Ideally, the nine classical null models are regularly spaced on such a landscape (Fig. 1). Yet, as we will demonstrate and discuss extensively below, although suggested by intuition, the intermediate position of the proportional algorithms depicted in Fig. 1 does not always match what is observed in reality. Here we take the challenge of providing a method to exploring the whole landscape in a continued and controlled way. For this, we introduce the "Tuning-Peg" (TP) algorithm, which has two parameters that permit to generate null matrices with prescribed discrepancy in row and column marginal totals.

Thus, for a given data matrix and metric, the application of the TP algorithm with enough, varying combinations of the adjustment parameters makes it possible to generate visualizations of significance and effect size across the entire landscape. This approach has the potential to give deeper insights into the processes of community assembly, and may help to resolve long-standing issues in ecology, such as, for example, whether or not (and/or under which assumptions) nestedness, which measures the tendency for the community of a given locality to be a subsample (in terms of species composition) of richer communities (Atmar and Patterson 1993), as well as the tendency for species in an ecological networks to share interacting partners (Strona and Veech 2015), is ubiquitous in natural systems (Bascompte et al. 2003, Joppa et al. 2010, Strona and Fattorini 2014, Strona and Veech 2015).

Methods

We introduce a procedure aimed at exploring, for a given focal empirical matrix, a bi-dimensional space whose x and y axes indicate, respectively, the degree of discrepancy between the marginal column and row

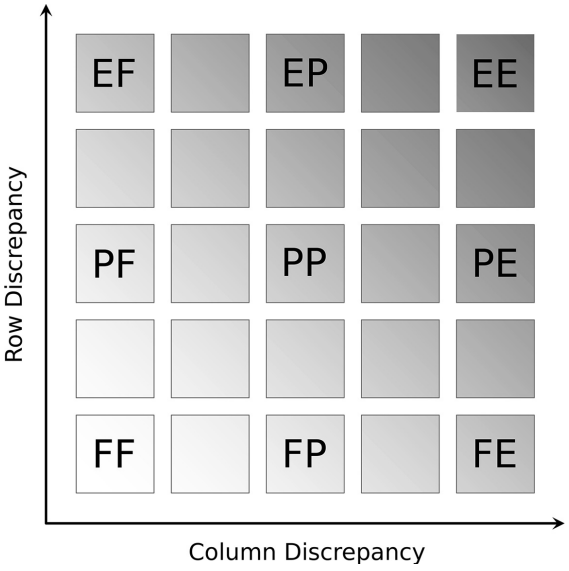

Fig. 1. Ideal representation of a 2-dimensional landscape filled up with hypothetical null models, with the two dimensions corresponding, respectively, to the discrepancy in row and column marginal totals between the original matrix and the randomized counterparts (obtained with the hypothetical null model). Superimposed across the landscape are the nine classical algorithms used for null model analysis of binary matrices.

totals of the focal matrix, and those of a target randomized counterpart. Thus, a given point having coordinates (x_i, y_i) identifies the exact position in the bi-dimensional null space of all possible null matrices whose column and row marginal totals have a discrepancy with respect to those of the focal matrix equal to x_i and y_i (see below for details on how discrepancy is computed).

The whole procedure is summarized in Fig. 2 and is based on combining a new matrix randomization algorithm named "Tuning Peg" (TP) with a random walk procedure aimed at reducing the computational demand of a thorough exploration of the null space. An area of interest is first identified in the null space, and then subdivided into a regular grid whose nodes correspond to target values of marginal total discrepancies between the original matrix, and the null matrices to be generated. The four cornering nodes are filled up using classical randomization algorithms (FF, EF, FE, EE). Then, the TP algorithm is used to generate a null matrix for each remaining node in the grid, by repeating the process of selecting a random node already hosting a null matrix as a starting point to generate a null matrix in one randomly extracted, empty neighboring node. The completion of a single random walk generates one null matrix for each node in the grid, ideally covering the whole null space of interest. Repeating the random walk several (e.g., 100–1,000) times for each node generates a set of null matrices large enough to estimate a *P* or SES value for that particular node, and hence to generate a bi-dimensional landscape of significance/effect size. A detailed explanation of these steps is provided in the following paragraphs.

Delimiting the null space

Given a focal presence-absence matrix with $i = 1$ to R rows (= species), and $j = 1$ to C columns (= sites), and a randomized (i.e., null) version of that matrix, both with rows and columns sorted in decreasing order of their respective marginal totals, we define the average normalized row and column mismatches between the two matrices as, respectively:

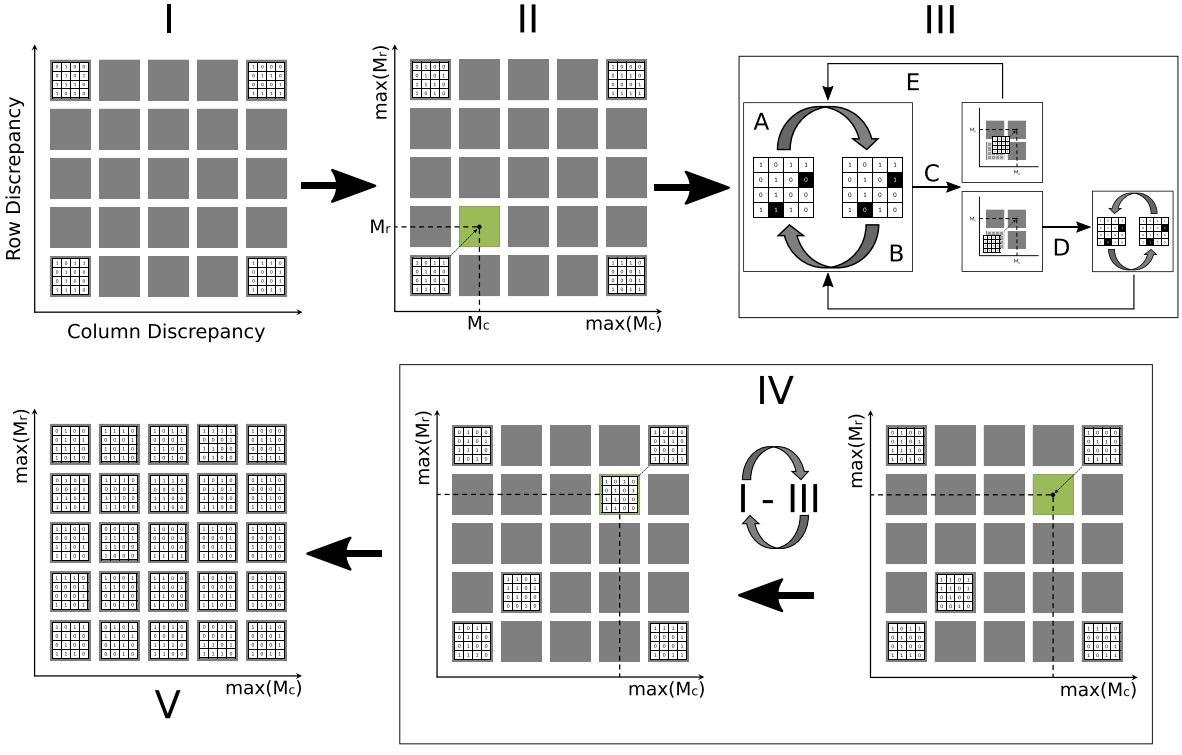

FIG. 2. Overview of the procedure to generate a landscape of significance. (I) The null space is subdivided in a regular grid whose node corresponds to evenly spaced discrepancy targets (i.e., combinations of M_r and M_c values). The nodes at the four corners of the null space are filled using the classical FF, FE, EF, and EE algorithms. (II) a node (node_0) in the grid already hosting a null matrix is extracted at random with uniform probability; if present, an empty node (node_e) neighboring node_0 is extracted at random with uniform probability. (III) the TP algorithm is applied to the matrix of node_0: two cells are selected at random (A), and swapped if different (B); if the swap increases the absolute difference between the target and the observed M_r and/or M_c values (C), the cells are swapped back to their previous position (D); otherwise, the modified matrix is retained (E). The swap procedure (A–E) is repeated until the absolute differences between the observed values of M_r and M_c and the target ones of node_e become smaller than specified thresholds t_r and t_c. (IV) steps I-III are repeated until all nodes in the grid have been filled up with a null matrix (V).

$$M_r = \frac{1}{R}\sum_{i=1}^{R}\frac{|r_i - sr_i|}{r_i} \quad (1)$$

$$M_c = \frac{1}{c}\sum_{j=1}^{c}\frac{|c_j - sc_j|}{c_j} \quad (2)$$

where r_i and c_j are the expected sums of the i-th row and j-th column of the original matrix, and sr_i and sc_j are the corresponding row and column sums of the null matrix.

For any focal matrix, we can then define a bi-dimensional space in which each point ideally corresponds to a set of null matrices having specific M_r and M_c values. Within this space, we identify a portion of interest delimited by the positions of classical null models. At one extreme, the most conservative algorithms (prone to type II errors) generate null matrices preserving exactly the marginal totals of the original matrix (FF). These conservative algorithms are located in the lower-left corner of the null space of interest. At the other extreme, the most liberal algorithms (prone to type I errors) generating matrices in which occurrences are placed in cells with equal probability (EE), whose average (over 1,000 replicates in our experiments) M_r and M_c values define the upper-right corner of the null space of interest (Figs. 1, 2-I). We denote those upper-right values as $\max(M_r)$ and $\max(M_c)$ because they define the upper bounds of the null space of interest. This, however, does not imply that null matrices having row and/or column mismatches higher than $\max(M_r)$ and $\max(M_c)$ are not possible for a given matrix.

A fundamental aspect to be accounted for in the computation of M_r and M_c (and hence of their maxima) is that different values of discrepancy can be reached by simply randomizing row and column order. This subtle issue is related to whether or not the identity of rows and columns is relevant to the specific analysis for which the null matrices are being generated. Ideally, if some particular hypothesis in which the actual order of rows and columns is meaningful, it would make sense to compute M_r and M_c using eqs. 1,2 with no particular adjustments. In all other cases for which row and column identity is irrelevant to the metric of choice (either because it is unaffected by the original row and column order, or because it requires a preliminary ordering of the focal matrix), M_r and M_c should be computed only after the original and the randomized matrices have been sorted according to the same criterion, which should also be applied in the computation of the target metric. Our choice of using the decreasing order of marginal totals simply reflects a common practice in ecological analysis, but is by no means the only possible criterion.

Identifying row and column mismatch targets

After $\max(M_r)$ and $\max(M_c)$ have been identified, one can designate a set of target M_r and M_c values within the intervals $[0, \max(M_r)]$ and $[0, \max(M_c)]$. The choice of the target values can be arbitrary, but because the goal of our procedure is a thorough and uniform exploration of the bi-dimensional null space defined by marginal constraint degrees, we will focus here on the case of equally spaced M_r and M_c target values. These correspond to the nodes of a square grid graph obtained by subdividing $[0, \max(M_r)]$ and $[0, \max(M_c)]$ into n equal intervals (Fig. 2-I), with the node in the lower left position corresponding to ($M_r = 0$, $M_c = 0$), and the node in the upper right position corresponding to (max (M_r) and $\max(M_c)$), consistent with our previous description of the bi-dimensional null space (Fig. 1).

Approaching the discrepancy targets

The procedure we describe here aims at generating efficiently, for each node in the grid, a random matrix with marginal row and column discrepancy (in respect to the original matrix) satisfying the corresponding target M_r and M_c values. The four extreme vertices of the grid can be filled using standard null model algorithms, and particularly, the lower left corner can be filled using a FF algorithm (here we used the Curveball algorithm by Strona et al. 2014), the lower right using a EF algorithm, the upper left using a FE algorithm, and the upper right using a EE algorithm (Fig. 2-I). The rest of the grid is filled using the following random walk procedure:

1. a node (node_0) in the grid already hosting a null matrix is chosen at random with uniform probability (Fig. 2-II);
2. if present, an empty node (node_e) neighboring node_0 is chosen at random with uniform probability (Fig. 2-II);
3. a swap procedure (called the Tuning-Peg algorithm and described in detail in the next section) is applied to the matrix of node_0 until the differences between its M_r and M_c values and those of node_e become smaller than pre-selected thresholds (Fig. 2-III; see next section for additional details on threshold values). If the algorithm fails, node_e is left empty, otherwise it is filled up with the randomized matrix. Prior to this, however, to enhance structural independence between matrices in neighboring nodes, the matrix of node_e is randomized using a standard FF algorithm.
4. steps 1–3 are repeated (Fig. 2-IV) until all nodes in the grid have been filled up with a null matrix (Fig. 2-V).

The Tuning-Peg (TP) algorithm

At each step of the algorithm, two cells are selected at random (Fig 2-III, A). If they are different, they are swapped (Fig 2-III, B). If the swap increases the absolute difference between the target and the observed M_r and/or M_c (Fig 2-III, C), the cells are swapped back to their previous position (Fig 2-III, D). Otherwise, the modified matrix is retained (Fig 2-III, E). The swap procedure (Fig 2-III, A–E) is repeated until the absolute differences between the target and the observed values of M_r and M_c are smaller than specified threshold t_r and t_c (see below). The

thresholds are needed because although the target M_r and M_c are ideally equally spaced within the intervals [0, max(M_r)] and [0, max(M_c)], there is a finite number of actual possible values for M_r and M_c, depending on matrix fill and structure (thus, in some cases, it could be impossible to reach a pre-selected target value of M_r and M_c).

Because the ultimate scope of our procedure is to explore the null space uniformly, in theory, the optimal t_r and t_c values are the smallest possible thresholds that permit the TP algorithm to complete its search, which would ensure that the null matrices are distributed as regularly as possible throughout the discrepancy landscape (ideally, perfectly achieving the grid nodes). However, small deviations from the pre-selected discrepancy targets do not constitute a major issue, considering that the actual discrepancy values are known (being computable by comparing a null matrix with the original one), and can be used to generate a truthful landscape of significance. Nevertheless, since it is important, in principle, to have a broad coverage of the null space, the thresholds should be at least small enough to prevent overlap between the different observations. For example, in case row and column discrepancy are subdivided into 10 intervals (as in Fig. 2), then the thresholds should be:

$$t_r \leq \frac{0.5 \times \max(M_r)}{10}$$

$$t_c \leq \frac{0.5 \times \max(M_c)}{10}$$

In all our analyses we conservatively set $t_r = \frac{0.25 \times \max(M_r)}{10}$, and $t_c = \frac{0.25 \times \max(M_c)}{10}$.

To prevent the algorithm from being trapped by a peculiar matrix structure, we allowed for a maximum of 50,000 consecutive "failed" swap attempts (i.e., not reducing the distance between the observed and the target M_r and/or M_c).

Exploring the bi-dimensional null space

Repeating n times the procedure described in the previous paragraphs (Fig. 2, I–V) generates, for each node in the discrepancy grid, a set of null matrices with given M_r and M_c. Those can then be used to obtain a set of corresponding P and/or SES values for a target structural pattern (such as nestedness), that can be plotted or interpolated across the bi-dimensional grid generating a landscape of significance and/or effect size.

Such landscapes provide visual information about the overall significance of a pattern under a wide set of null hypotheses. Besides being visually appealing and informative, the landscapes are also suitable for further analyses aimed at quantifying and characterizing the area of significance. Here we consider only a few of the most obvious measures, namely the fraction of nodes in the grid having significance (P) values lower than 0.05 or higher than 0.95, and the mean of all SES values computed across the whole landscape. We leave more complex analysis of the landscape patterns for future work.

It should be highlighted that, depending on the structure of the original matrix (size, fill, shape, and, most importantly, distribution of marginal totals), the explorable portion of the null space can vary. This applies in particular to matrices having uniform marginal totals, because those will not be much different from their null counterparts obtained using an EE algorithm, which would lead to very low max(M_r) and max(M_c) values. This dependency should be taken into account when interpreting the results of an analysis, especially when comparing results from different matrices having very different max(M_r) and/or max(M_c).

Tests on artificial matrices

As a proof of concept, and to test how Type I and II errors vary across the bi-dimensional null space, we applied the TP procedure to three sets of 100 artificial (i.e., randomly generated) matrices. The first set (**U**) included matrices filled completely at random (that is, by starting with a matrix filled with 0s, and then selecting cells with uniform random probability, until half of the cells have been filled up with 1s). The second set (**H**) included matrices having marginal totals drawn from an exponential distribution with exponents sampled from a uniform random distribution with values between one and two. Those matrices are expected to show a certain degree of nestedness emerging from passive sampling, and are thus expected to be more nested than most null matrices generated with weak constraints on marginal totals, but not more nested than null matrices having similar marginal totals (and hence having a structure equally affected by passive sampling). The third set (**N**) included nested matrices. Those were generated by starting from a perfectly nested matrix (having nestedness based on overlap and decreasing fill, NODF = 100, Almeida-Neto et al. 2008), that was then "perturbed" (as in Strona and Veech 2015) by randomly replacing presences with absences and vice versa in 25% of the total number of matrix cells (extracted, one cell at a time, with repetition). The size of each matrix was determined randomly, with row and column numbers sampled with uniform random probability between 15 and 100, and with fill varying with uniform random probability between 0.2 and 0.8.

We applied the TP algorithm to each matrix 100 times, hence generating 100 null matrices for each node in a grid of 121 equally spaced discrepancy targets, obtained by varying M_r and M_c from 0 to max(M_r) with step 0.1 × max(M_r), and from 0 to max(M_c) with step 0.1 × max (M_c). In total, we generated therefore 3.63 mol/L null matrices.

Landscapes of significance/effect size

For each matrix, we calculated nestedness using the NODF metric (Almeida-Neto et al. 2008), and the

average degree of segregation in species co-occurrences using the *C*-score (Stone and Roberts 1990). For each matrix, we compared the observed NODF and *C*-score with those measured on a set of 100 null matrices obtained at each of the 121 unique combinations of M_r and M_c. In addition, we compared the observed values of NODF and *C*-score with those measured for 100 null matrices generated separately for 6 "classic" null models that should generate an intermediate level of discrepancy between expected and observed row and/or column marginal totals. The 6 "classic" null models are: PP, proportional row and column totals; PE, proportional row totals, equiprobable column totals; EP, equiprobable row totals, proportional column totals; PF, proportional row totals, fixed column totals; FP, fixed row totals, proportional column totals (all of these were implemented with the procedures in Gotelli 2000); UG, unbiased proportional row and column totals (as implemented by Ulrich and Gotelli 2012).

All comparisons of observed vs. simulated results were based on a frequentist estimate of tail probability (*P*-value), computed as the fraction of null matrices having NODF and *C*-score higher than the respective values of the matrix under consideration. We also computed effect size as a standardized effect size (SES value), computed as $\text{SES} = (x-\mu)/\sigma$, with x being the observed metric score, μ being the mean of the scores of the null matrices, and σ being their standard deviation.

For each matrix, we created two landscapes of significance and effect size by plotting the *P* and SES values obtained using the TP algorithm in a bi-dimensional grid in which the coordinates (x and y) of each cell corresponded, respectively, to M_c and M_r values, that is, a grid where each cell corresponded to a unique setup of TP. We used different color gradients (magenta to blue, and light green to dark green) to discriminate between left and right tails of the null distribution (i.e., $P < 0.05$ and $P > 0.95$), and lighter colors (magenta to white, and light green to white) to indicate non-significant *P*-values. We use an analogous color-scale to map SES values, with values from 1.96 to the maximum recorded SES colored from magenta to blue, values from −1.96 to the minimum recorded SES colored from light to dark green, and lighter colors in between. Thus, for both NODF and *C*-score, a dark green, blue or light-colored cell at position $x = M_c$ and $y = M_r$ of the landscape indicates, respectively, that the observed metric is significantly smaller, significantly larger, or not significantly different than expected from the TP algorithm set-up at M_r and M_c.

For empirical matrices (see below), we plotted individual landscapes. Conversely, for each one of the three sets of artificial matrices (random with uniform marginal totals, random with exponential marginal totals, and nested) we averaged SES and *P*-values of all matrices in the set in a single landscape, where we also plotted six circles, representing the proportional null model algorithms (PP, PE, EP, PF, FP, UG). Each circle is placed in the landscape in a position with M_r and M_c respectively equal to the average row and column mismatches of the 100 matrices generated with the corresponding null model. In other words, the position of the circles indicates the TP setting generating matrices with the same (average) row/column mismatches of the six proportional null models. The fill color of each circle indicates the *P*-value obtained with the respective null model for the target metric. When these colors are the same as the background landscape, they indicate similar behavior of the proportional null model and the TP algorithm with comparable mismatch parameters. When the circle and background colors differ, the two algorithms (i.e., the proportional null model and the TP) are producing different results. When a circle falls outside the landscape, it means that the proportional algorithm produces matrices tending towards particular configurations having row and/or column marginal discrepancy higher than that expected under the assumption of equiprobable placement of occurrences (EE) we used to delimit the null space.

Error rates

We defined the frequency of Type I errors as the fraction of points in the null space (corresponding to a specific setting of the TP algorithm, i.e., to a unique combination of M_r and M_c) for which the null matrices were more or less structured than random expectation (i.e., having NODF or *C*-score *P*-values smaller than 0.05, or larger than 0.95). Similarly, we estimated Type II error rate, for each nested matrix, as the fraction of points in the null space for which the nested pattern was not detected (NODF *P*-value ≥ 0.05); an additional estimate of Type I error rate was obtained as the fraction of points in nested matrices for which a significant co-occurrence pattern (*C*-score $P < 0.05$) was detected.

Tests on Empirical matrices

We applied our approach to the well-known dataset by Atmar and Patterson (1995). As in the case of theoretical matrices, for each empirical matrix we generated both landscapes of effect size and of significance for NODF and *C*-score for 121 combinations of M_r and M_c. We provide the complete results as Supporting Information, while we focus our discussion below on five empirical matrices illustrating different situations of particular interest:

1. Ants in the Society Islands, from Wilson and Taylor (1967) (10 localities, 21 species, 57 occurrences; filename in Atmar and Patterson's dataset: "Sociantt.txt");
2. Terrestrial arthropods on 6 islets composed of pure *Spartina alterniflora*, in Northwest Florida, initially defaunated by fumigation, 24 weeks after the treatment, from Rey (1981) (6 localities, 41 species, 79 occurrences; filename in Atmar and Patterson's dataset: "Defau624.txt");

3. Reptiles on the California Islands, from Wilcox (1980) (15 localities, 27 species, 65 occurrences; filename in Atmar and Patterson's dataset: "Calirept.txt");
4. Birds of the Canary Islands, from Bacallado (1976) (7 localities, 78 species, 280 occurrences; filename in Atmar and Patterson's dataset: Canaboss.txt).
5. South Africa *Sciobius* weevil species distribution in arbitrary, connected quadrats, from Morrone (1994) (21 localities, 47 species, 124 occurrences; filename in Atmar and Patterson's dataset: Saweevil.txt);

Programming details

We implemented the algorithms and performed all of the analyses using the Python programming language (van Rossum and de Boer 1991). All the code necessary to generate null matrices using the TP algorithm, together with an implementation of the 10 classic null models and functions to compute NODF and *C*-score, is provided in Data S1 (all code has been developed and tested using Python version 2.7). In addition, we provide an *R* script including the functions necessary to perform the complete bi-dimensional null model analysis, and plot the landscape of significance/effect size for both NODF and *C*-score (Data S2).

Results

Artificial matrices

The landscape of significance/effect size for the three sets of artificial matrices were very consistent with the expectation (Figs. 3–4). In particular, the nearly totality of null matrices in all replicates and across the whole landscape had NODF and *C*-score values not different from those observed in the random artificial matrices with uniform marginal totals (U), while they resulted more nested (NODF) and with a smaller *C*-score than those observed in the artificial nested matrices (N). As predicted, the matrices with heterogeneous, exponential marginal totals (H) resulted more nested than expected across almost all the NODF landscape, with the exception of the lower left corner (corresponding to null matrices having marginal totals very similar or identical to those of the original matrix).

Type I error rates averaged over the whole landscape were 0.04 ± 0.08 (SD) for NODF and 0.05 ± 0.10 for *C*-score in the case of random matrices with uniform row and column marginal totals (U). Very few of the nested matrices (N) had a *C*-score higher than random expectation, resulting in a Type I error rate of 0.01 ± 0.00.

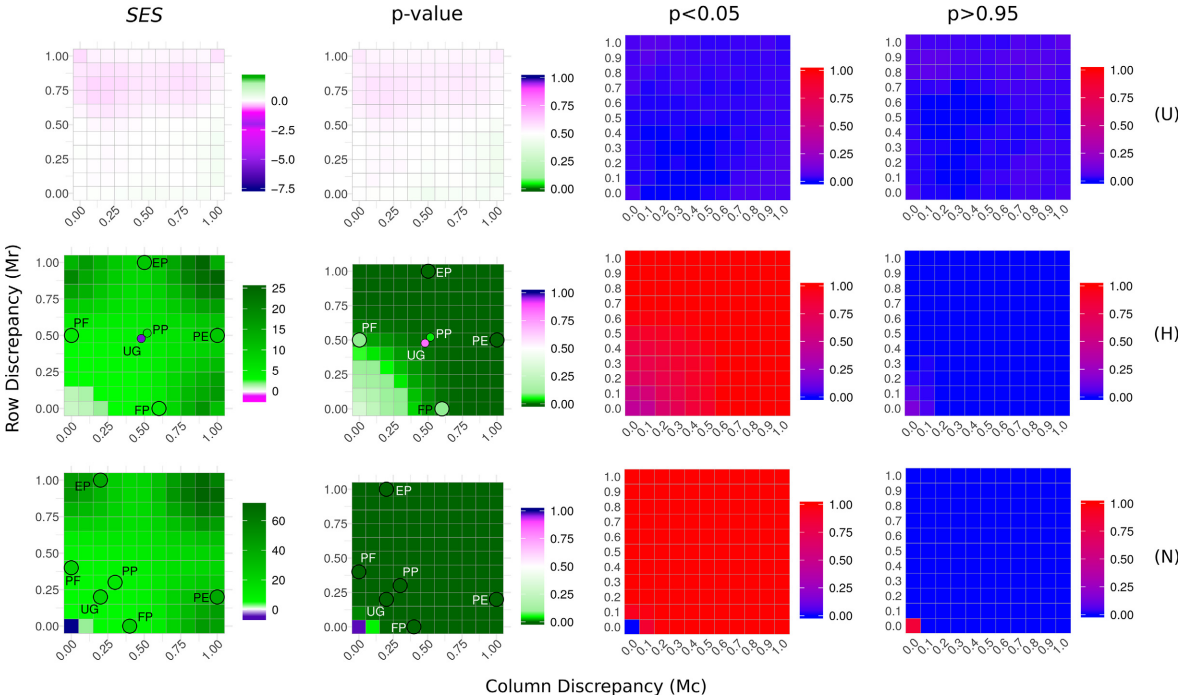

FIG. 3. The landscapes report the effect size (SES) and significance (*P*-value), as well as the fraction of matrices per cell in the landscape having *P*-value smaller than 0.05 or larger than 0.95 for NODF averaged over the sets of random matrices with uniform marginal totals (**U**); random matrices with heterogeneous marginal totals (**H**); and nested artificial matrices (**N**). We also plotted the dots summarizing the results obtained by using classical proportional randomization algorithms in a position corresponding to the average of the row/column mismatch of the null matrices. However, for the random matrices with uniform marginal totals (**U**), those fell out of the landscape boundaries, and are therefore shown in a separate picture (Fig. 5).

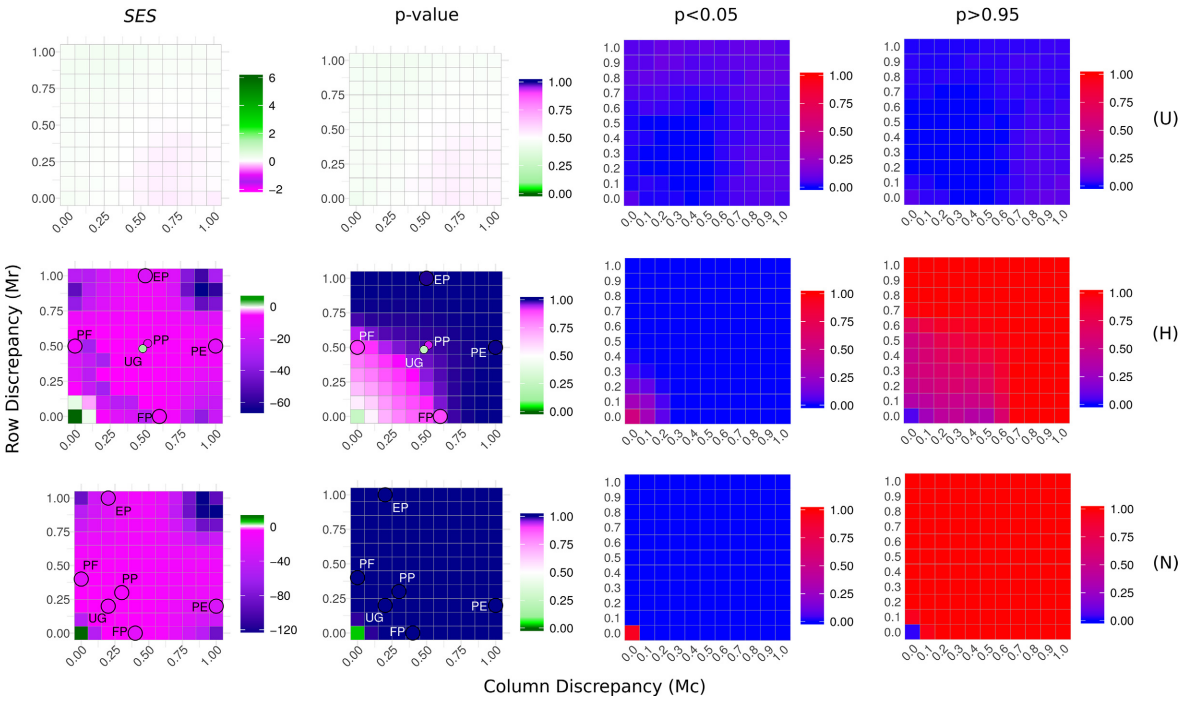

FIG. 4. The landscapes report the effect size (SES) and significance (*P*-value), as well as the fraction of matrices per cell in the landscape having *P*-value smaller than 0.05 or larger than 0.95 for *C*-score averaged over the sets of random matrices with uniform marginal totals (**U**); random matrices with heterogeneous marginal totals (**H**); and nested artificial matrices (**N**). We also plotted the dots summarizing the results obtained by using classical proportional randomization algorithms in a position corresponding to the average of the row/column mismatch of the null matrices. However, for the random matrices with uniform marginal totals (**U**), those fell out of the landscape boundaries, and are therefore shown in a separate picture (Fig. 5).

Conversely, a higher than expected nested structure was correctly identified nearly across the whole landscape in most of the nested matrices (N), leading to a Type II error rate (averaged over the landscape) of 0.01 ± 0.01 for NODF (Fig. 3). In fact, error rate for NODF was 0 in all cases, except when both M_r and M_c were close to 0. This seemingly strange pattern reflects the documented difficulties of detecting structure in highly nested matrices using NODF in combination with a FF null model, deriving from the reduced set of possible alternative configurations of the original matrices preserving marginal totals (see Strona and Fattorini 2014).

The random matrices with heterogeneous marginal totals (H) were correctly identified as "random" only when compared with null matrices with very similar marginal totals (i.e., with both M_r and M_c close to 0). For higher values of M_r and M_c, however, the matrices were identified as significantly nested (with an average type I error rate for NODF of 0.94 ± 0.09; similarly, the average of type I error rates measured as the fraction of points in the landscape where the measured *C*-score was lower than the expected one was 0.84 ± 0.14. In a small fraction of the landscape (0.01 ± 0.02) the observed *C*-score was significantly higher than the expectation.

In the case of nested matrices (N), and even more in that of random matrices with heterogeneous marginal totals (H), the placement of the proportional algorithms in the landscape according to their averaged M_r and M_c values indicated that these algorithms generated an intermediate degree of restrictiveness in between fully constrained or uniformly randomized marginal totals. By contrast, in the case of random artificial matrices with uniform marginal totals (U), most of the proportional algorithms produced null matrices with average M_r and M_c values higher than max(M_r) and max(M_c) (Fig. 5).

In the case of nested matrices (N), the resulting *P* and SES values obtained using the proportional null models were in close accordance with corresponding results (i.e., at the same position in the null space) obtained by using the TP algorithm (Figs. 3–4). The same applies to matrices with heterogeneous marginal totals (H) with the exception of the proportional algorithm UG (Ulrich and Gotelli 2012), which correctly indicated randomness both for NODF ($P = 0.83$) and for the *C*-score ($P = 0.23$). The placements of the proportional null models outside the boundaries of the bi-dimensional null space explored by the TP in the case of random matrices corresponded to a discrepancy in the SES and *P*-values. In particular, all proportional algorithms, and particularly the UG generated null matrices tending towards a more nested structure (thus having higher NODF and

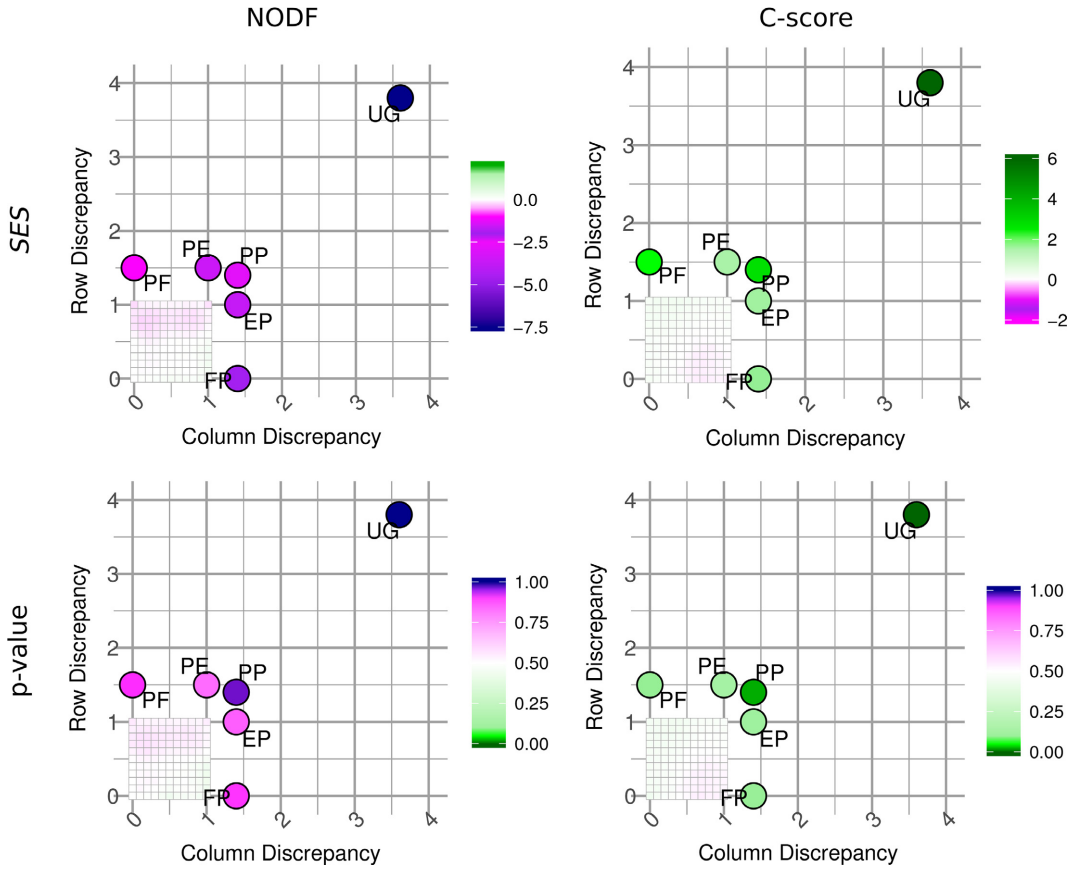

FIG. 5. Relative position in the null space of the classical proportional randomization algorithms in respect to the landscapes of effect size (SES) and significance (*P*-value) of the set of theoretical random matrices with uniform marginal totals, for both NODF and *C*-score. Dots are placed in a position corresponding to the average of the row/column mismatch of the null matrices, and are colored according to the resulting *P*-value (with the same color-scale used for the landscape).

smaller *C*-score) than that of the focal random matrices generated with uniform marginal totals (U).

Effect of matrix properties

In random matrices with uniform marginal totals (U), matrix properties (number of rows and columns, size, fill and shape) were not correlated neither with the averaged effect sizes (SES), nor with the fraction of M_r and M_c combinations identifying significant patterns for both NODF and *C*-score (Table 1). Conversely, in the case of random matrices with heterogeneous marginal totals (H), artificial nested matrices (N), and empirical matrices from the Atmar and Patterson's (1995) set, matrix properties were moderately correlated with both the averaged effect size and with the fraction of M_r and M_c combinations for which *P*-values were significant (Table 1).

Empirical matrices

The landscapes of NODF and *C*-score effect size and significance for the whole Atmar and Patterson's (1995) dataset are reported in Appendix S1. Among those, we identified five particular cases showing different situations (Fig. 6). First, the matrix of the Society Islands ants was consistently identified as being strongly (and significantly) nested across the whole landscape (Fig. 6). Consequently, *C*-score was lower than expected (Fig. 6A). Conversely, the second matrix (terrestrial arthropods on six defaunated islets in Northwest Florida composed of pure *Spartina alterniflora*) appeared to be random across the whole landscape (Fig. 6B).

The other three matrices were characterized by heterogeneous landscapes of significance/effect size that highlight the need for further investigation, and emphasize how focusing on a single null model could lead to deceiving results. Specifically, the matrix of reptiles on the California (Channel) Islands was characterized by a diagonal gradient of structure across the landscape, where nestedness became evident only when relaxing the constraints on marginal totals (i.e., moving towards high M_r and M_c values) (Fig. 6C). The matrix of birds of the Canary Islands showed a horizontal gradient, in which strong constraints on column totals (i.e., low values of M_c) led

TABLE 1. Effect of matrix properties on the landscapes of significance/effect size.

	NODF (Z)	C-score (SES)	NODF (*P*)	C-score (*P*)
Random (H)				
R	0.48*	−0.47*	0.22*	0.33*
C	0.3*	−0.37*	−0.3*	0.38*
Size	0.62*	−0.66*	0.01	0.56*
Shape	0.09	−0.04	0.29*	−0.05
Fill	−0.17	−0.71*	−0.15	0.48*
Random (U)				
R	−0.13	0.14	0.10	0.15
C	−0.01	−0.08	−0.12	−0.12
Size	−0.10	0.06	0.00	0.04
Shape	−0.07	0.12	0.16	0.14
Fill	−0.04	0.02	−0.01	0.11
Nested				
R	0.63*	−0.64*	0.57*	0.15
C	0.57*	−0.7*	−0.12	0.39*
Size	0.83*	−0.94*	0.33*	0.38*
Shape	−0.03	0.14	0.45*	−0.27*
Empirical				
R	0.68*	−0.62*	0.48*	0.28*
C	0.23*	−0.37*	−0.34*	0.15*
Size	0.61*	−0.72*	0.00	0.25*
Shape	0.23*	−0.09	0.55*	0.05
Fill	−0.12*	0.03	−0.41*	0.03

Notes: We compared: SES values averaged over the whole landscape of significance, and the fraction of combinations of M_r and M_c where a significant pattern was detected ($P < 0.05$) with: number of rows (**R**) and columns (**C**); matrix size, computed as **R** × **C**; matrix shape, computed as **R/C**; matrix fill, computed as the ratio between the number of occurrences in the matrix, and matrix size. Pairwise relationships between variables (expressed as Spearman's rank correlation coefficient) are reported separately for random matrices with both heterogeneous (**H**) and uniform (**U**) marginal totals, nested artificial matrices, and empirical matrices from Atmar and Patterson's (1995) dataset. Significant relationships ($P < 0.05$) are denoted by *. We do not report fill for nested matrices since those were all generated with the same fill (0.5) in order to obtain the desired degree of nestedness (see *Methods*).

to the identification of less than expected nestedness regardless of M_r values, while relaxing the constraints on M_c led to the identification of significant nestedness (Fig. 6D). The fifth matrix (South Africa *Sciobius* weevil species) showed an analogous situation, but with a vertical gradient of effect size/significance driven by the degree of constraints in row marginal totals (Fig. 6E).

Discussion

The fact that different constraints in null models can have a fundamental effect on the detection of ecological patterns has been at the center of the debate for decades (e.g., Connor and Simberloff 1979, Ulrich and Gotelli 2013). In this paper, we provide a new perspective on the issue, by showing how different null models can be reconciled on a bi-dimensional continuum of restrictiveness, whose dimensions are defined by the marginal mismatch in row and column totals between the target matrix and its randomized counterparts. In doing this, we introduced a new null modeling framework which provides an immediate and valuable visualization of how the effect size (or significance) of a given pattern varies across the null space.

Adding a second dimension to null model analysis, that is, presenting the results in a landscape of significance/effect size is preferable to providing a single SES score or *P*-value resulting from the choice of a particular null model for several reasons. Justifying the choice of one null model over another on the basis of ecological/biological assumptions is not always straightforward. Moreover, biases can derive from the restrictiveness (or lack of) of the chosen null model. Showing how the effect size changes under different combinations of row and column constraints offers a broader perspective, permitting to distinguish between cases where a pattern is always significant regardless of null model restrictiveness, from cases where the significance is limited to specific regions of the null space. Moreover, investigating how the effect size varies in the null space defined by row and column restrictiveness can provide interesting insights on the determinants of matrix structure.

The FF algorithm, being the most conservative among the typical matrix randomization procedures, has been often recommended due to its ability to reduce Type I errors in most situations. However, it is also known to be prone to Type II errors in particular situations. One of those is that of highly nested matrices, especially when using the FF algorithm in combination with the popular NODF nestedness metric. This problem is well highlighted in Fig. 3: the only point in the landscape where the analysis failed to identify the non-random pattern in the set of the artificial nested matrices (N) is at the FF algorithm (i.e., at zero M_r and M_c). However, relatively small changes in row and column restrictiveness lead to an abrupt transition to the correct identification of significant nestedness, that is also preserved across the rest of the landscape. It is noteworthy that the region of the landscape adjacent to the lower left corner (zero M_r and M_c) relaxes only slightly the constraints of the FF algorithm, hence generating null matrices whose marginal totals are still very close to those of the original ones. Therefore, although satisfying the need to rule out the effect of marginal totals (e.g., local species richness and species prevalence in typical species-area incidence matrix) on the pattern of interest, those quasi-FF matrices proved themselves more reliable than the typical approach (FF) in detecting structure of highly nested matrices. Furthermore, they correctly identified as not-nested the random matrices with exponential marginal totals (Fig. 3). Conversely, higher values of M_r and M_c lead to the incorrect detection of nestedness due to passive sampling.

In the case of random artificial matrices with uniform marginal totals (U), most of the proportional algorithms produced null matrices with average M_r and/or M_c

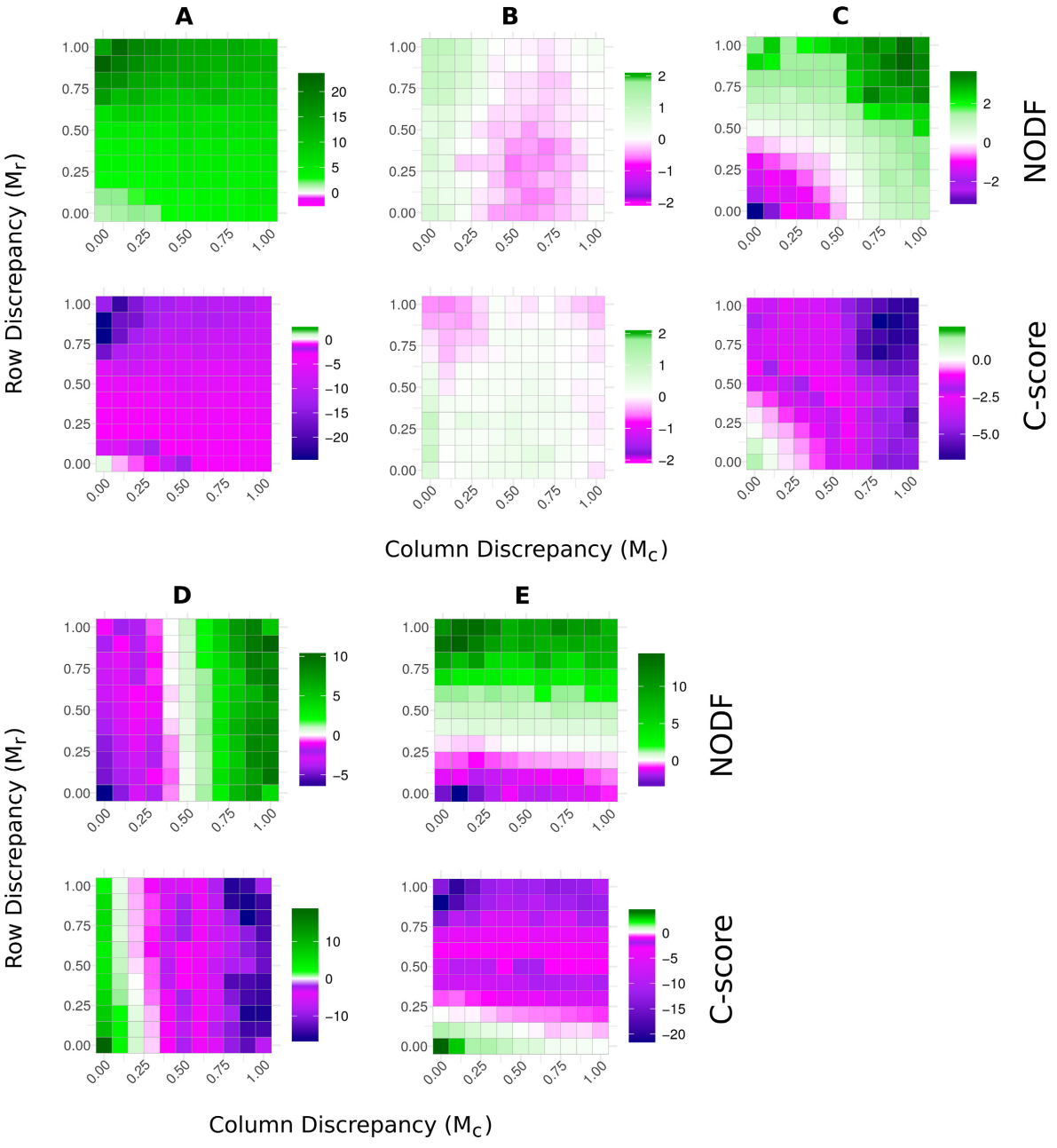

FIG. 6. Landscapes of effect size (SES) of five empirical matrices selected from Atmar and Patterson's (1995) dataset. (A) Ants in the Society Islands; (B) terrestrial arthropods on 6 islets composed of pure *Spartina alterniflora*, in Northwest Florida, initially defaunated by fumigation, 24 weeks after the treatment; (C) reptiles on the California Islands; (D) birds in the Canary Islands; (E) South Africa *Sciobius* weevil species distribution in arbitrary, connected quadrats.

values higher than max(M_r) and/or max(M_c) (Fig. 5). Apparently these algorithms, and specifically the PP and UG algorithms, when used in combination with matrices with uniform marginal totals, produce null matrices with a tendency towards more nestedness (low *C*-score and high NODF) that fall outside the null space of interest. This is in contrast with the random pattern correctly identified in the rest of the investigated null space.

In the case of random matrices with heterogeneous marginal totals (H), all the classical algorithms generate matrices with lower nestedness than that observed in the focal matrices, with the exception of FF (which of course, preserving the heterogeneous distribution of marginal totals permits to get rid of the potential confounding effect of passive sampling), and UG. As in the case of random matrices with uniform marginal totals

(U), the UG algorithm tends to generate matrices with higher NODF and lower *C*-score than those observed in random matrices with heterogeneous marginal totals (H), hence leading to the identification of less than expected nestedness, and more than expected segregation. However, due to the increased level of nestedness due to passive sampling in the random matrices with heterogeneous marginal totals (H), the differences in NODF and *C*-score compared to the UG null matrices were not large enough to be identified as significant.

The landscapes of significance/effect size obtained with our procedure show very well how, for random matrices with heterogeneous marginal totals (H), only a comparison with null matrices with row and column marginal totals close to the target matrix (and hence retaining the heterogeneity in marginal totals and accounting for the consequent passive sampling bias towards nestedness) correctly avoid Type I statistical errors and false significance. Using less restrictive null models (or setting the tuning peg to higher M_r and/or M_c target values) leads to false positives.

In our analysis of artificial matrices, we knew a priori that the placement of presences in the heterogeneous matrices was random, and that therefore the degree of structure was simply a product of passive sampling driven by marginal totals. For empirical matrices, it may be hard to say whether (and to what extent) the marginals are a cause or a consequence of the marginal distributions. In the absence of any further information, the parsimonious approach is to preserve the structure of the marginals, and test for non-random patterns above and beyond this constraint. Our landscapes of significance/effect size provide better insights into this issue, because they reveal more clearly the extent to which the marginal constraints are (or are not) responsible for the outcome of the null model analysis.

The experiments on the empirical matrices showed further why examining a broader portion of the null space than that usually explored in ecological studies is fundamental to a proper investigation of matrix structure. In the first two cases (Fig. 6A-B), the observed structural patterns were consistent across the whole landscape. Furthermore, they were reasonable from an ecological perspective, with the distribution of ant species in the Society Islands showing the typical nested structure (Fig. 6A) expected in relatively stable insular systems (Atmar and Patterson 1993); and the latter, representing the arthropod communities of six islets composed of *Spartina alterniflora* a few weeks after having been defaunated by fumigation, showing a random structure consistent with the initial disturbance, and the relatively small time for re-colonization (Fig. 6B).

The other three examples demonstrated that matrix patterns can be often too complex to be assessed by single metric-null model combinations, and that their interpretation is strictly dependent on the ecological meaning of null model assumptions (Ulrich and Gotelli 2013). For instance, the matrix of California (Channel) Islands reptiles showed a tendency towards less than expected nestedness for low values of M_r and M_c, and the opposite tendency (more than expected nestedness) for higher values (Fig. 6C). In an over-simplification, if we assume that the observed number of reptile species per island, as well as the number of islands where a given species occurs are not really representative of any important ecological mechanism, we would identify the matrix as nested. This, however, would correspond to the assumption that the islands in the dataset do not differ in their likelihood of being colonized by a reptile species, as well as in their habitat and resource availability, and that all species have similar colonization abilities and ecological requirements. Conversely, if we relax this assumption, then we have to identify the matrix as not nested. Similar reasoning, applied for only column or only row totals, can be applied for the fourth and fifth examples (birds of the Canary Islands and *Sciobius* weevils of South Africa; Fig. 6E–F).

In a typical analysis, one would have probably focused on just one null model, either identifying the matrices as nested or not, and then building the conclusions on only one of these contrasting results. The landscape of significance, conversely, provides, at glance, a broader spectrum of results, hence representing a more cautious, and definitely more informative, approach to null model analysis of binary matrices.

Similar situations highlight how trying to synthesize the structure of an ecological matrix with a single number may hide more complex scenarios that are worth a more in-depth investigation. For this, our new approach offers a valuable alternative to classical null model analysis, which has relied on a handful of algorithms that, in fact, represent specific settings of the more general TP algorithm. We are convinced that focusing on the landscape of significance/effect size, which can reveal how the strength of an ecological pattern varies according to a broad, continuous range of possible assumptions, could be much more rewarding than trying to justify interpretations based on a single algorithm. Considering the long history of species-area matrix analysis, we ecologists have often focused on one or another single dot of a Georges Seurat pointilist painting. We hope our method could help us taking some steps back, and get a better look at the big picture.

Acknowledgments

WU acknowledges funding by the Polish National Science Centre (grant NCN 2014/13/B/NZ8/04681). NJG was supported by U. S. NSF DEB 1257625, NSF DEB 1144055, and NSF DEB 1136644. The views expressed are purely those of the writers and may not under any circumstances be regarded as stating an official position of the European Commission.

Literature Cited

Almeida-Neto, M., P. Guimaraes, P. R. Guimarães, R. D. Loyola, and W. Ulrich. 2008. A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. Oikos 117:1227–1239.

Atmar, W., and B. D. Patterson. 1993. The measure of order and disorder in the distribution of species in fragmented habitat. Oecologia 96:373–382.

Atmar, W., and B. D. Patterson. 1995. The nestedness temperature calculator: a visual basic program, including 294 presence-absence matrices. AICS Research Incorporate and The Field Museum.

Bacallado, J. J. 1976. Notas sobre la distribucion y evolucion de la avifauna Canaria. Pages 413–431 *in* G. Kunkel, editor. Biogeography and ecology in the Canary Islands. Dr. W. Junk, The Hague, The Netherlands.

Bascompte, J., P. Jordano, C. J. Melián, and J. M. Olesen. 2003. The nested assembly of plant–animal mutualistic networks. Proceedings of the National Academy of Sciences of the United States of America 100:9383–9387.

Baselga, A. 2010. Partitioning the turnover and nestedness components of beta diversity. Global Ecology and Biogeography 19:134–143.

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

Gimenez, O., et al. 2014. Statistical ecology comes of age. Biology Letters 10:20140698.

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

Gotelli, N. J., and G. R. Graves. 1996. Null models in ecology. Smithsonian Institution Press, Washington, D.C.

Gotelli, N. J., and W. Ulrich. 2012. Statistical challenges in null model analysis. Oikos 121:171–180.

Joppa, L. N., J. M. Montoya, R. Solé, J. Sanderson, and S. L. Pimm. 2010. On nestedness in ecological networks. Evolutionary Ecology Research 12:35–46.

Morrone, J. J. 1994. On the identification of areas of endemism. Systematic Biology 43:438–444.

Rey, J. R. 1981. Ecological biogeography of arthropods on Spartina islands in Northwest Florida. Ecological Monographs 51:237–265.

Stone, L., and A. Roberts. 1990. The checkerboard score and species distributions. Oecologia 85:74–79.

Strona, G., and S. Fattorini. 2014. On the methods to assess significance in nestedness analyses. Theory in Biosciences 133:179–186.

Strona, G., and J. A. Veech. 2015. A new measure of ecological network structure based on node overlap and segregation. Methods in Ecology and Evolution 6:907–915.

Strona, G., D. Nappo, F. Boccacci, S. Fattorini, and J. San Miguel-Ayanz. 2014. A fast and unbiased procedure to randomize ecological binary matrices with fixed row and column totals. Nature Communications 5:4114.

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 N. J. Gotelli. 2012. A null model algorithm for presence–absence matrices based on proportional resampling. Ecological Modelling 244:20–27.

Ulrich, W., and N. J. Gotelli. 2013. Pattern detection in null model analysis. Oikos 122:2–18.

Ulrich, W., A. Baselga, B. Kusumoto, T. Shiono, H. Tuomisto, and Y. Kubota. 2017. The tangled link between β-and γ-diversity: a Narcissus effect weakens statistical inferences in null model analyses of diversity patterns. Global Ecology and Biogeography 26:1–5.

van Rossum, G., and J. de Boer. 1991. Interactively testing remote servers using the Python programming language. CWI Quarterly 4:283–303.

Wilcox, B. A. 1980. Species number, stability, and equilibrium status of reptile faunas on the California Islands. Pages 551–564 *in* D. M. Power, editor. The California Islands: Proceedings of a multi- disciplinary symposium. Santa Barbara Museum of Natural History, Santa Barbara, California, USA.

Wilson, E. O., and R. W. Taylor. 1967. The ants of Polynesia. Pacific Insects Monographs 14:1–109.

Wright, D. H., B. D. Patterson, G. M. Mikkelson, A. Cutler, and W. Atmar. 1997. A comparative analysis of nested subset patterns of species composition. Oecologia 113:1–20.

Supporting Information

Additional supporting information may be found in the online version of this article at http://onlinelibrary.wiley.com/doi/10.1002/ecy.2043/suppinfo