

Abstract Community assembly rules are often inferred from patterns in presence-absence matrices. A challenging problem in the analysis of presence-absence matrices has been to devise a null model algorithm to produce random matrices with fixed row and column sums. Previous studies by Roberts and Stone [(1990) *Oecologia* 83:560–567] and Manly [(1995) *Ecology* 76:1109–1115] used a “Sequential Swap” algorithm in which submatrices are repeatedly swapped to produce null matrices. Sanderson et al. [(1998) *Oecologia* 116:275–283]

duce similar results to Sanderson et al.'s (1998) algorithm, which samples nodes exhaustively and retreats sequentially. As with Sanderson et al.'s (1998) Knight's Tour, our Random Knight's Tour begins with an empty matrix and fills it by sequentially adding randomly chosen cells and backtracking when necessary until a solution is reached. In our Random Knight's Tour, the algorithm does not sample exhaustively at each node, but only examines a small number of randomly chosen cells for filling. If a cell cannot be filled, our algorithm retreats by randomly removing a filled cell from anywhere in the matrix, not necessarily the last cell filled. Although not mentioned by Sanderson et al. (1998), Stone and Roberts (1990) also reported on the results of a Random Knight's Tour algorithm (the "Milne method") that gave similar results to their swap algorithm (L. Stone, personal communication).

For small matrices, we sampled a single randomly chosen cell at each node. If a usable cell could not be found, the algorithm immediately retreated one step. For large matrices, this algorithm was inefficient because the program spent too much time backtracking. For the Vanuatu matrix, we analyzed a range of sampling intensities (Table 2). The lowest sampling intensity of 444 cells (~ 50% of the open cells) allowed us to create 1,000 random matrices in approximately 12 h of simulation time.

Exhaustive Knight's Tour

For this algorithm, we increased the sampling intensity to a very large number, so that the available cells at each node were exhaustively searched. For the Vanuatu matrix, we sampled 8,888 cells at each node, ensuring that all available cells were checked before the algorithm backtracked. Our Exhaustive Knight's Tour is identical to Sanderson et al.'s (1998) algorithm, except that we did not retreat sequentially when backtracking. As we show, this algorithm generates results that are qualitatively similar to those of Sanderson et al. (1998), although we did not observe the very large variances generated by their algorithm.

Our algorithms were programmed in Delphi Version 4.0, and implemented in EcoSim Version 5.0 (Gotelli and Entsminger 2000). The on-line version of EcoSim contains the Sequential Swap, Independent Swap, and Random Knight's Tour algorithms, and allows for the choice of the *C*-score, the number of checkerboards, the number of species combinations, and the variance ratio (Schluter 1984) as co-occurrence indices. The on-line version also contains data matrices for the West Indian finches and Virginia ants that are described in this paper. A special compiled version of EcoSim that analyzes the *S*

nificantly from that generated by the Knight's Tour algorithm. It is on the basis of this result that Sanderson et al. (1998) concluded that the S^2 metric is flawed and the randomization method of Manly (1995) is biased.

However, Sanderson et al.'s (1998) results are not directly comparable to Roberts and Stone's (1990), in part because Sanderson et al. (1998) used 5,000 simulations, whereas Roberts and Stone (1990) used 1,000. More important, Roberts and Stone (1990) imposed an additional constraint in their simulations that does not appear to have been incorporated by Sanderson et al. (1998). In their analysis, Roberts and Stone (1990) maintained the so-called "incidence function" of each species, so that the simulated range of island sizes occupied by each species matched that in the observed data set. This additional constraint means that the matrices created by Roberts and Stone (1990) constituted a subset of all random matrices that maintained row and column totals.

These different procedures do account for some, but not all, of the differences between the results of Sanderson et al. (1998) and Roberts and Stone (1990). When we analyzed the Vanuatu matrix with the Sequential Swap algorithm, we also obtained a significant result ($P=0.039$), though not as extreme as that reported by Roberts and Stone (1990; $P<0.001$). Our results from the Sequential Swap matched the results of the Independent Swap, in which each matrix is constructed from a series of Independent Swaps (Table 1).

Our Random Knight's Tour generated results that were very similar to those of the Sequential and Independent Swap. For a small data set, we confirmed that the Random Knight's Tour does generate a non-biased sample of random matrices (Appendix). In contrast, our Exhaustive Knight's Tour generated results very different from the Random Knight's Tour and the two Swap algorithms (Table 1). As reported in Sanderson et al.'s (1998) study, our Exhaustive Knight's Tour led to a shift in the mean and an increase in the variance, though not as extreme as Sanderson et al. (1998) found with their algorithm.

The behavior of the Knight's Tour fill algorithms are very sensitive to the amount of sampling at each node. If exhaustive sampling is used, the results tend towards those reported by Sanderson et al. (1998). With less exhaustive sampling, the results converge towards those of the swap algorithms (Table 2).

Type I error analysis

Repeated swapping of submatrices will theoretically produce the full set of null matrices (Brualdi 1980), and we note that the Swap algorithms and the Random Knight's Tour produce consistent results. In their criticism of the Sequential Swap algorithms, Sanderson et al. (1998) caution that "it cannot be guaranteed that cyclic perturbations can be avoided. Thus, it is possible to perform two or more perturbations and end up with the same matrix".

However, cyclic perturbations are unlikely for most real matrices. For example, in the Vanuatu matrix, there are initially 14,676 different submatrices that can be swapped at the first step of the Sequential Swap procedure. Thus, the chance of reversing a forward step and generating the same matrix configuration is $1/14,676=6.8\times 10^{-5}$. We agree with Sanderson et al. (1998) that there are many possible matrix rearrangements that can be created for the Vanuatu matrix (and most real presence-absence matrices) that satisfy row and column totals. Therefore it is unlikely that cyclic perturbations are a serious problem for the Sequential Swap algorithm.

Sanderson et al. (1998) rejected the Sequential Swap algorithm because it gave different results for the Vanuatu presence-absence matrix than did their Knight's Tour algorithm. However, it is not logical to evaluate different algorithms by comparing their behavior with a real presence-absence matrix – such a matrix will contain unknown amounts of biological structure and randomness. Instead, we should compare the performance of algorithms on artificial data sets with known properties (e.g., Gotelli et al. 1997; Garvey et al. 1998; Shenk et al. 1998).

By claiming that Roberts and Stone's (1990) finding of a non-random S^2 metric was incorrect, Sanderson et al. (1998) invoked a Type I error, in which a true null hypothesis is incorrectly rejected. To attack this problem directly, we use artificial data sets to decide whether Roberts and Stone (1990) incorrectly rejected the null hypothesis for the Vanuatu matrix. We first created 100 random matrices in which we were sure there should be no pattern present. We wanted these matrices to resemble the Vanuatu matrix, but with no trace of biological structuring.

To create such matrices, we began with the original Vanuatu matrix and then reshuffled the cells within each row. Reshuffling the cells within each row eliminates any differences among sites due to area, resources, or

Table 2 Effects of sampling intensity on performance of the Knight's Tour algorithm. For the Vanuatu matrix, results are given for 1000 simulations. Sampling intensity indicates the number of cells the algorithm samples before backtracking. Other variables as in Table 1

Sampling intensity	\bar{x} Simulated	σ^2
--------------------	---------------------	------------

habitat availability. This is analogous to a colonization model in which all sites are equiprobable, and each species selects a site independently of the presence of any other species. We stress that the creation of test matrices in this way is not circular, because it is a completely different filling algorithm than either the swap or fill algorithms.

The resulting set of 100 matrices should not exhibit any non-randomness as a group, although we would expect that, by chance, approximately 5% of these matrices would generate significant results in a well-behaved statistical analysis. Running more than 100 matrices in this way would have been desirable, but the test is time-consuming, and the error frequency should not be appreciably different for 100 vs 1,000 matrices.

Next, we analyzed each random matrix with the Sequential Swap algorithm, completing 5,000 iterations and performing separate analyses for the S^2 and the C -score indices. Thus, although the column totals for each random matrix could vary freely as the matrix was being constructed, the resulting column and row totals were fixed in the Sequential Swap algorithm. For each matrix, we recorded the upper-tail probability that the observed index was more extreme than predicted by the null model. If Sanderson et al.'s (1998) claim of a Type I error is correct, we should find that many more than 5 of the 100 random matrices caused the null hypothesis to be rejected. On the other hand, if 5 or fewer of the 100 random matrices caused the null hypothesis to be rejected at the 0.05 significance level, there would be no indication of an excessive Type I error rate, and Roberts and Stone's (1990) original conclusion would be supported.

Table 3 gives the ordered upper-tail probabilities for the 15 most extreme random matrices. For the S^2 metric,

5 out of 100 random matrices rejected the null hypothesis at $P=0.05$, and 12 out of 100 matrices rejected the null hypothesis at $P=0.10$. The calculated P value for the actual Vanuatu matrix was 0.0042. This value was more extreme than 98 of the 100 random matrices, suggesting that the probability of a Type I error for the S^2 index of the Vanuatu matrix was less than 0.02.

For the C -score, 5 out of the 100 random matrices rejected the null hypothesis at $P=0.05$, and 12 out of 100 random matrices rejected the null hypothesis at $P=0.10$. The calculated P value for the actual Vanuatu matrix was $P=0.0006$. This value was more extreme than the tail probability for all 100 of the random matrices, suggesting that the probability of a Type I error for the C -score metric of the Vanuatu matrix was less than 0.01.

Thus, we find no evidence that Roberts and Stone (1990) made a Type I error when they rejected the null hypothesis for the S^2 metric of the Vanuatu matrix. When the Sequential Swap algorithm is used on a set of random matrices, it correctly generates a non-significant upper-tail pattern 95% of the time. In contrast, the S^2 and C -score indices for the actual Vanuatu matrix are significantly larger than expected by chance, and the pattern is more extreme than in most random matrices that have identical row totals. These analyses of random matrices support the original conclusions of Roberts and Stone (1990). The S^2 metric for the Vanuatu matrix is significantly larger than chance, and the Sequential Swap method does not have an excessive Type I error rate.

Discussion

Which algorithm should be used?

Our results suggest that the Knight's Tour algorithm of Sanderson et al. (1998) should not be used for null model analysis. This algorithm does not sample matrices equiprobably (Appendix), and the results it produces rest delicately on the assumption of exhaustive sampling at each node (Table 2). The Type I and Type II error properties of Sanderson et al.'s (1998) Knight's Tour are unknown, and the anomalous result it generates for the Vanuatu matrix (Table 1) is not a valid basis for claiming that "results from previous studies are generally flawed".

absence matrix they tested. G. Cobb (Department of Mathematics, Mt. Holyoke College) found that small runs of 1,000 replicates of the Sequential Swap were significantly heterogeneous and autocorrelated, due to similarity of sequential matrices generated by the algorithm. However, Cobb also found that a set of random matrices generated by the Sequential Swap from the presence-absence matrix of Darwin's finches had a uniform probability distribution, as it should for an unbiased test. Collectively, these results suggest the Sequential Swap performs well in empirical tests, and its behavior on the Vanuatu matrix converges with that of the Independent Swap, Zaman and Simberloff's (unpublished data) Weighted Swap, and our Random Knight's Tour.

How should indices be tested?

Sanderson et al. (1998) followed the lead of Connor and Simberloff (1979) and plotted the number of species pairs sharing 0,1,2,...*n* islands. They tested for deviations of observed and expected values in each category by constructing a parametric 99% confidence interval for a Student's *t*-test. Observations that fell outside of this interval were classified as significant. This procedure is very similar to the original chi-square test that Connor and Simberloff (1979) used to evaluate the same distribution of data. As Roberts and Stone (1990) thoroughly explained, a parametric chi-square test is inappropriate in this case because the set of shared island numbers are not independent observations. Sanderson et al. (1998) revived this error by using a *t*-test and evaluating all of the species pair combinations. No appeals to parametric theory can get around the fact that these numbers are not independent. Indeed, one of the major reasons for using randomization methods in the first place is to avoid problems of exactly this sort.

To test the significance of patterns in null model analysis, we prefer the methods that are used in classic randomization tests (Edgington 1987; Manly 1991): describe the pattern in the matrix with a single index, and compare the observed value of that index directly to the distribution of index values from the simulated matrices. Extreme values of the observed index in the tails of the distribution indicate statistically significant patterns.

Which co-occurrence index should be used?

The utility of species co-occurrence indices should be based not only on their statistical properties, but on their relationship to ecological theory. Sanderson et al.'s (1998) used the individual deviations of species-pair classes as their index. Thus, an assembly rule derived from their analysis of the Vanuatu matrix (their Fig. 2) would be: "There are fewer avian species pairs occupying 2 islands and 9 islands, and more species pairs occupying 10 islands, than expected by chance in the Vanuatu Archipelago". We do not see how meaningful biological

interpretations could be attached to such an assembly rule.

In contrast, we prefer three different indices that address Diamond's (1975) assembly rules model: the number of species combinations (Pielou and Pielou 1968), the number of species pairs forming perfect checkerboard (Diamond 1975), and Stone and Robert's (1990) *C*-score. A fourth metric, Schuller's (1984) variance ratio, is also a useful measure of species covariance. However, the variance ratio is determined solely by the row and column sums of the matrix, so it cannot be tested using the randomization algorithms described in this paper, all of which preserve row and column sums (Gotelli 2000).

Pielou and Pielou (1968) first introduced the number of species combinations as an index of community structure. This index is directly related to Diamond's (1975) first and second assembly rules:

- "1. If one considers all the combinations that can be formed from a group of related species, only certain ones of these combinations exist in nature."
- "2. These permissible combinations resist invaders that would transform them into a forbidden combination."

If assembly rules 1 and 2 are met, a set of islands or sites should harbor significantly fewer species combinations than expected by chance.

A second useful index is the number of species pairs that never co-occur, forming "checkerboard" distributions. This index describes Diamond's (1975) fifth assembly rule:

- "5. Some pairs of species never coexist, either by themselves or as part of a larger combination."

ture.Oes c ux isails tions

randomness in these indices (Weiher and Keddy 1999). However, the null model analysis is the important first step towards at least establishing whether the patterns are random or not.

Although the *C*-score, number of checkerboards, and number of species combinations are closely related, they measure different patterns in a matrix and have different statistical properties. Gotelli (2000) has carried out simulation tests with these indices and compared their performance on random matrices and on structured matrices that have differing levels of random “noise” added to the co-occurrence patterns.

The *C*-score, when used in combination with the Sequential Swap algorithm, has good statistical power for detecting pattern in structured matrices that have a considerable amount of random noise added to them. At the same time, it has good Type I error properties and does not lead to an excessive rejection of the null hypothesis when tested with random matrices (Gotelli 2000). A low frequency of Type I errors for the *C*-score and the Sequential Swap algorithm are confirmed by our additional analyses of the Vanuatu matrix in this paper (Table 3).

Empirical examples

To illustrate utility of the methods we propose, we analyzed three presence-absence matrices with the Sequential Swap algorithm and present patterns based on the *C*-score, the number of checkerboard pairs, and the number of species combinations. We used 5,000 iterations in each analysis. Results were similar when these matrices were analyzed with the Random Knight’s Tour.

For the Vanuatu avifauna, the *C*-score was significantly greater than expected by chance (Fig. 1a), but the number of checkerboard pairs (Fig. 1b) and the number of species combinations (Fig. 1c) did not differ from random. For the number of species combinations, we note that each of the 28 islands in the Vanuatu matrix supported a different species combination. Similarly, in all 5,000 simulated communities, species combinations were never repeated for any pair of islands, so 28 combinations were observed in every case.

For the West Indies fishes, the *C*-score was also larger than expected by chance (Fig. 2a). Only 10 unique species combinations occurred on the 19 islands, which was significantly less than expected by chance (Fig. 2b), confirming Diamond’s (1975) 1st and 2nd assembly rule patterns. The number of checkerboard pairs did not differ from random (Fig. 2c).

For the Virginia ants, none of the three co-occurrence indices were different from random (Fig. 3a–c). Because the ant matrix was created from a grid of pitfall traps that sampled species at a very small spatial scale (5×5 m), we would not have expected to see much structure in comparison to the avifauna of island archipelagos.

In conclusion, we are in agreement with Sanderson et al. (1998) on several major issues. First, we agree that

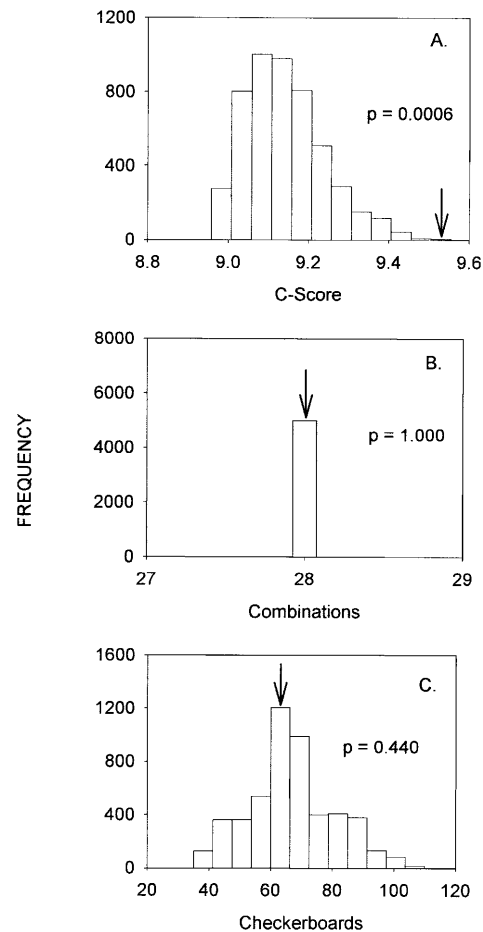


Fig. 1a–c Null model analyses of the Vanuatu avifauna. Each histogram shows the range of values from 5,000 simulated assemblages. The arrow indicated the observed value, and *p* is the one-tailed probability. **a** *C*-score; **b** number of species combinations; **c** number of species pairs forming perfect checkerboards

the entire issue of testing for non-randomness in species co-occurrence patterns is important, especially with the continued interest in community assembly rules (Weiher and Keddy 1999) and computer simulation models (Hilborn and Mangel 1997; Cipra 2000). Second, we agree that an algorithm that maintains observed row and column sums is useful for testing community patterns, although this is by no means the only valid null model (Gotelli 2000). We also agree that, for most real matrices, there are many different matrix rearrangements that can be constructed that satisfy the row and column constraints.

However, there are also some major areas of disagreement. Our results suggest that Sanderson et al.’s (1998) Knight’s Tour is a biased algorithm that should not be used for null model analysis. We think it is inappropriate to use a *t*-test on non-independent data, and that deviations from the shared-species distribution cannot be interpreted biologically. We prefer a simple randomization test for assessing statistical significance. We recommend the *C*-score, number of species combinations, and number of

checkerboard pairs as three indices with good statistical properties that are directly relevant to Diamond's (1975) assembly rules model. Finally, an error analysis of random matrices supports Roberts and Stone's (1990) original conclusions: the S^2 metric for the Vanuatu matrix is larger than expected by chance. There is no evidence that this pattern is caused by a Type I error because random matrices tested with the Sequential Swap algorithm rejected the null hypothesis 5% of the time (Table 3).

This paper represents the latest contribution to a controversy over the statistical analysis of the Vanuatu presence-absence matrix that has lasted for over 20 years (e.g., Connor and Simberloff 1979, 1983; Diamond and Gilpin 1982; Wilson 1987; Roberts and Stone 1990; Sanderson et al. 1998; Zaman and Simberloff, unpublished data). No wonder some ecologists are uneasy about the use of null models! Future exchanges can be minimized by carefully examining the statistical properties of null models, and systematically testing their behavior with artificial data sets of known structure. Such analyses are tedious and not very exciting, but they are necessary if null models are to serve as reliable statistical tools for ecologists and biogeographers.

Acknowledgements We thank A. Ellison, G. Cobb, P. Peres-Neto, J. Sanderson, D. Simberloff, L. Stone, and the University of Vermont Ecology Discussion Group for comments on the manuscript. We thank Inprise Corporation/Borland for Delphi development

Suppose that we wish to construct random matrices with four cells filled. The matrix constraints we use are row totals of (1, 2, 1) and column totals of (1, 2, 1). There are

126 unique ways to fill a 3×3 matrix with four 1s. Of these 126 matrices, exactly 5 matrices satisfy the specified row and column constraints. We label these matrices A-E:

Matrix A

Matrix B

Matrix C

Matrix D

Matrix E

An unbiased null model algorithm should generate all five matrices with frequency $f=0.20$.

Exhaustive Knight's Tour

The probabilities for the Exhaustive Knight's Tour can be calculated by hand for this small example. We illustrate the method with matrix A, which is the easiest to compute. Matrix A has entries in cells 2,4,6, and 8. There are $4! = (4)(3)(2)(1) = 24$ sequences by which this

matrix can be reached filling sequentially with the exhaustive knight's tour.

We calculate the probabilities of reaching each particular sequence through the Exhaustive Knight's Tour, and then sum those probabilities to obtain the total frequency with which the matrix is reached. Suppose we wish to calculate the probability that Matrix A is reached by filling in sequence the cells 2, 4, 6, and 8. Thus, we seek to calculate $P(\text{Matrix A} | 2468)$.

For the initial placement of cell 2 in an empty matrix, the probability is $(1/9)$.

Once cell 2 has been selected, matrices C and D are eliminated from the solution space. For the next cell selection, cell 2 cannot be chosen again because it is already filled. Moreover, cells 1 and 3 are also not possible because these would violate row or column constraints. Thus, on the next step of the exhaustive knight's tour,

$P(\text{Matrix B})=0.21296$. Matrices B–E are oversampled by the Exhaustive Knight's Tour because there are more starting positions from which these matrices can be reached. Note that the patterns in Matrices B–E can all be produced by reflection or transposition, and the probabilities of reaching these matrices turn out to be identical. We also confirm that $0.14815+(4)(0.21296)=1.0000$, as it should.

Sequential Swap

To calculate the expected frequencies for the Sequential Swap, we used a different approach. For each matrix A–E, a single swap will produce one of the other matri-

- Gotelli NJ, Graves GR (1996) Null models in ecology. Smithsonian Institution, Washington, D.C.
- Gotelli NJ, Buckley NJ, Wiens JA (1997) Co-occurrence of Australian land birds: Diamond's assembly rules revisited. *Oikos* 80:311–324
- Hilborn R, Mangel M (1997) The ecological detective: confronting models with data. Princeton University Press, Princeton
- Manly BFJ (1991) Randomization and Monte Carlo methods in biology. Chapman and Hall, London
- Manly BFJ (1995) A note on the analysis of species co-occurrences. *Ecology* 76:1109–1115
- Pielou DP, Pielou EC (1968) Association among species of infrequent occurrence: the insect and spider fauna of *Polyporus betulinus* (Bulliard) Fries. *J Theor Biol* 21:202–216