



A Note on the Analysis of Species Co-Occurrences

Author(s): Bryan F. J. Manly

Source: *Ecology*, Vol. 76, No. 4 (Jun., 1995), pp. 1109-1115

Published by: [Ecological Society of America](#)

Stable URL: <http://www.jstor.org/stable/1940919>

Accessed: 26/02/2014 16:42

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
<http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



Ecological Society of America is collaborating with JSTOR to digitize, preserve and extend access to *Ecology*.

<http://www.jstor.org>

A NOTE ON THE ANALYSIS OF SPECIES CO-OCCURRENCES¹

BRYAN F. J. MANLY

*Department of Mathematics and Statistics, University of Otago,
P.O. Box 56, Dunedin, New Zealand*

Abstract. The analysis of records of species occurrences on islands in an attempt to detect interactions between species has been an area of controversy in recent years in terms of the validity of some of the statistical methods used. In this note I make two contributions to the continuing debate. First, I advocate a generalized Monte Carlo testing procedure because this is easy to implement, is computationally efficient, and has guaranteed properties when the null hypothesis of no species interactions is correct. Second, I propose a test statistic that can be decomposed into a component for each individual species, and I demonstrate how this makes it possible to isolate species with unusual patterns of co-occurrence with other species, even after an allowance for multiple testing is made.

Key words: *community structure; computer-intensive methods; data analysis; generalized Monte Carlo test; randomization test; species co-occurrences; statistical methods.*

INTRODUCTION

There has in recent years been a considerable controversy about whether interactions between species can be detected in records of species occurrences on islands or other collections of discrete localities. For example, it might be asked whether the data shown in Table 1 for presences and absences of 20 lizard species on 25 islands in the Gulf of California indicate a tendency for some pairs of species to occur on the same islands and other pairs of species to occur on different islands.

One promising approach involves comparing statistics found for a real set of occurrences with the distributions of these statistics obtained by a randomization process that produces reconstructed sets of data for which certain specified constraints are the same as for the real data. For example, the number of times that species i occurs with species j might be calculated for the real data, for all possible pairs i and j . The standard deviation of these numbers of co-occurrences can then be found and compared with the distribution of the standard deviation that is obtained by generating a large number of random sets of data, for which certain characteristics are kept the same as for the real data. A significantly high observed standard deviation in comparison with the distribution for the random sets of data is then evidence that the distribution of the number of co-occurrences is more variable than can be expected by chance alone. This might then be interpreted in terms of some species avoiding each other (competitive exclusion) and some species tending to occur together (species aggregation).

The literature on randomization tests of this type up

to about 1988 has been reviewed by Manly (1991). More recent contributions have been concerned with the fitting and interpretation of a logistic model for the probability that species i is on island k (Ryti and Gilpin 1987) and the use of new test statistics designed to detect competitive exclusion and species aggregation (Wilson 1988, Roberts and Stone 1990, Stone and Roberts 1990, 1992).

One of the questions with these tests has been which constraints should be applied when generating random species occurrences. The point of view adopted here is that the number of occurrences of each species and the number of species on each island (i.e., the row and column totals indicated in Table 1) should be kept equal to the observed values in order to reflect the facts that some species are naturally more widespread than others and some islands are naturally able to accommodate more species than others. I recognize that the appropriateness of these constraints is debatable, and that it can be argued that the row and column sums may themselves be partly the outcome of species interactions (e.g., see Gilpin and Diamond 1984 and Connor and Simberloff 1984). However, other types of constraint are not considered in this note.

Two forms of algorithm have been used to construct random communities. One approach attempts to simulate random colonization by starting with empty islands and allocating species one by one to available "spaces" (Connor and Simberloff 1979, Wilson 1987, Stone and Roberts 1990, Manly, 1991). The other approach starts with the observed-occurrence matrix and changes this by a stepwise process of switching a randomly chosen pattern of the form A shown below to a pattern of the form B (Connor and Simberloff 1979, Roberts and Stone 1990).

¹ Manuscript received 23 July 1993; revised 7 July 1994; accepted 5 October 1994; final version received 3 November 1994.

TABLE 1. Occurrences of 20 groups of lizard species on 25 islands in the Gulf of California (1 = presence, 0 = absence). From Case 1983a: Table and Appendix 1).

Species†	Species occurrences																				Total no.					
	Island*																									
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T		U	V	W	X	Y
1 <i>Coleonyx</i>	1	0	1	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0	1	0	1	0	7
2 <i>Phyllodactylus</i>	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	23
3 <i>Sceloporus orcutti</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	1	1	0	0	6
4 <i>S. magister</i>	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0	1	0	1	6
5 <i>S. clarki</i>	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	2
6 <i>Cnemidophorus tigris</i>	1	0	1	1	1	0	1	1	1	1	1	1	0	1	1	1	1	0	1	0	0	1	1	1	0	18
7 <i>C. hyperythrus</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	1	0	0	0	1	1	1	1	8
8 <i>Urosaurus</i>	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	0	8
9 <i>U. ornatus</i>	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
10 <i>Uta</i>	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	0	22
11 <i>Petrosaurus mearnsi</i>	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
12 <i>P. thalassinus</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	2
13 <i>Callisaurus draconoides</i>	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	1	1	1	1	9
14 <i>Crotaphytus</i>	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
15 <i>Gambelia wislizenii</i>	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
16 <i>Sauromalus</i>	1	1	1	1	0	0	0	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	0	18
17 <i>Dipsosaurus dorsalis</i>	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	1	0	0	1	0	1	1	9
18 <i>Ctenosaurus hemilopha</i>	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	3
19 <i>Sator</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	3
20 <i>Phrynosoma solare</i>	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Total	13	4	9	4	3	2	3	4	4	5	2	5	2	10	10	10	7	6	6	3	3	11	8	11	6	151

* Island name code: A = Tiburton, B = Mejia, C = Angel de la Guarda, D = Pond, E = Partida Norte, F = Raza, G = Salsipuedes, H = San Lorenzo Norte, I = San Lorenzo Sur, J = San Estaban, K = San Pedro Martir, L = San Pedro Nolasco, M = Tortuga, NM = San Marcos, O = Coronadoes, P = Carmen, Q = Danzante, R = Monserrate, S = Santa Catalina, T = Santa Cruz, U = San Diego, V = San José, W = San Francisco, X = Espirito Santo, and Y = Cerralvo.

† Species groupings are as used by Case (1983a), with the addition of the three species (*P. solare*, *S. clarki*, and *U. ornatus*) that are present only on one or both of Tiburton and San Pedro Nolasco.

Species	Pattern A island		Pattern B island	
	<i>i</i>	<i>j</i>	<i>i</i>	<i>j</i>
r	0	1	1	0
s	1	0	0	1

Switching species between islands in this manner maintains the total number of occurrences of each species and the total number of species on each island, and will eventually generate all possible occurrence matrices with these constraints (Brualdi 1980).

This note makes two contributions to the continuing debate about the analysis of species occurrences. First, it points out that Besag and Clifford's (1989) theory of generalized Monte Carlo significance tests is immediately applicable to the problem of deciding how to compare an observed-occurrence matrix with a series of occurrence matrices generated from it by the stepwise random switching process. This theory provides an exact test of significance even if only a relatively small number of switches are made, and, in particular, makes it unnecessary to run the switching process for some time in order to produce a "random" pattern. Second, the note addresses the question of how to isolate individual species with occurrence patterns that indicate they are tending to be associated positively or negatively with other species.

GENERALIZED MONTE CARLO TESTS

The principle involved in simple Monte Carlo tests was proposed by Barnard (1963). He pointed out that if an observed test statistic S_1 is compared with $n - 1$ other values S_2, S_3, \dots, S_n generated independently of each other using pseudo-random numbers on the basis of some null hypothesis, then (ignoring the possibility of ties) the probability of S_1 being amongst the k largest values is k/n . Therefore a one-sided test at the $100(k/n)\%$ level for whether the observed data were generated with the null hypothesis true consists of seeing whether S_1 is one of the top k in the set S_1, S_2, \dots, S_n . Two-sided tests and ties can be handled with obvious modifications of this procedure.

The independence of the n test statistics is fundamental to the justification of this test. Therefore the test cannot be used as it stands for testing the observed value of a statistic designed to detect interaction between species occurrences on islands when alternative values of the statistic are obtained by making a series of stepwise changes to the observed-occurrence matrix of zeros and ones.

This problem has been appreciated by some of those carrying out tests based on making stepwise changes to the observed-occurrence matrix. For example, Roberts and Stone (1990) carried out 10^5 random changes on an observed matrix and only used the statistics

found from matrices obtained after $J_1, J_1 + J_2, J_1 + J_2 + J_3, \dots$ changes from then on, with J_i uniformly distributed from 1800 to 2200. They generated 1000 statistics in this way and made the assumption that they differed by enough changes to be effectively independent.

Roberts and Stone's procedure may well be valid, but it suffers from the disadvantage of requiring a large number of changes ($\approx 2.1 \times 10^6$) in order to generate only 1000 statistics. In contrast to this, an algorithm described by Besag and Clifford (1989) generates a test statistic with every stepwise change from an initial occurrence matrix. The importance of this difference in the context of the present paper is that the Besag and Clifford algorithm makes it possible to obtain very high levels of significance with a realistic number of changes. For example, with Roberts and Stone's (1990) procedure $\approx 2.1 \times 10^6$ changes can give a result that is significant at the 1/1000 (0.1%) level. On the other hand, Besag and Clifford's (1989) procedure with the same number of changes can give a significance level of $1/(2.1 \times 10^6)$ i.e., 0.00005%. As will be seen below, this difference makes the Besag and Clifford procedure more suitable for obtaining the high significance levels that are required to detect non-random patterns of occurrence for individual species.

The Besag and Clifford algorithm is as follows for testing whether an observed statistic is significantly large at the $100(k/n)\%$ level:

a) Generate a random integer m with a uniform distribution between 1 and n . This gives the position of the observed-occurrence matrix in a series of length n . The observed test statistic therefore becomes S_m .

b) If $m > 1$, carry out $m - 1$ stepwise changes of species occurrences in the observed-occurrence matrix as described in the previous section to obtain $m - 1$ occurrence matrices with their corresponding values $S_{m-1}, S_{m-2}, \dots, S_1$, i.e., one change gives S_{m-1} , two changes give S_{m-2} , etc. These are considered to be backwards changes.

c) If $m < n$, carry out $n - m$ stepwise changes of species occurrences in the observed-occurrence matrix to obtain $n - m$ occurrence matrices with their corresponding values $S_{m+1}, S_{m+2}, \dots, S_n$, i.e., one change gives S_{m+1} , two changes give S_{m+2} , etc. These are considered to be forwards changes.

d) If S_m is one of the top k values for the test statistics ordered by magnitude, then declare it significant at the $100(k/n)\%$ level.

It is shown in the Appendix that in the absence of ties this procedure is "exact" in the sense that the probability of obtaining a significant result is k/n when the observed-occurrence matrix is randomly chosen from the set of all possible occurrence matrices with the same row and column totals. If ties occur then these can be randomly broken. Alternatively, a conservative test can be constructed by placing the observed test statistic below any others with the same value in the

ordered list. The probability of a significant result will then $< k/n$.

Besag and Clifford (1989) called the algorithm (a)–(d) the "serial method." An alternative parallel method that they proposed is not described here because it requires many more changes of the occurrence matrix than the serial method to generate the same number of test statistics, and appears generally to be less efficient than the serial method (Manly 1993:333). Besag and Clifford discussed both methods in the context of testing the Rasch (1960) model for the responses ("1" for correct, "0" for incorrect) of a number of candidates to a series of test questions. Here the data have exactly the same form as Table 1 if questions replace islands and candidates replace questions. In fact, the Rasch model is similar to Rytty and Gilpin's (1987) logistic model for species occurrences.

GLOBAL AND CONDITIONAL P VALUES

Suppose that the serial algorithm is applied and a P value is calculated for the observed test statistic, where this is equal to the proportion of the n values in the series that are greater than or equal to that observed. Then this P value is an unbiased estimate of the proportion of test statistics as large as or larger than the observed in all series of length n that include the observed value. It is therefore a conditional P value that is not an unbiased estimate of the proportion of *all* statistics as large or larger than the observed for the complete collection of possible occurrence matrices. This difference between the P value from the serial test using a series of length n and the global P value will reduce as n increases, but it is not clear how large n has to be before the difference becomes negligible.

In fact, it can be expected that the P value from a serial test will be higher than the global P value in the cases of interest where the observed test statistic S is itself quite large. This is because it is the nature of the serial test to produce test statistics that are similar to S in magnitude for occurrence matrices that differ by only a few stepwise changes from the observed-occurrence matrix. Therefore, the probability of large values of S will tend to be overestimated to some extent in this test.

The implication of this is that it should not be surprising to find situations where the P value estimated by comparing an observed test statistic with $n - 1$ independently generated random values is smaller (and therefore more significant) than the P value estimated from the serial test with a series of length n . However, this difference should disappear as the series length n increases.

The alternative situation, where the serial test gives a smaller P value than a test using the same number of independently generated statistics, would be harder to explain (assuming that it is not just a question of the sampling errors in estimating P values). It would appear to require that the observed-occurrence matrix

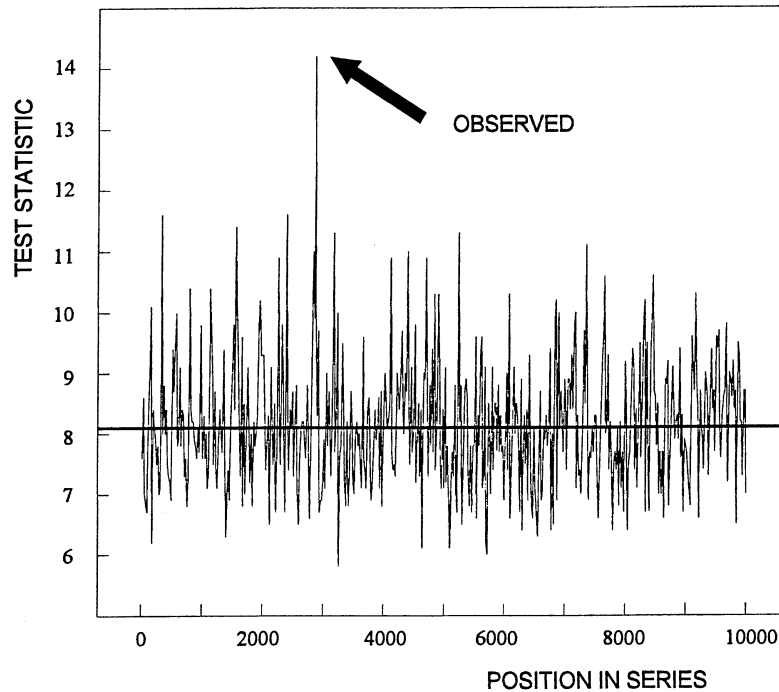


FIG. 1. Test statistics for the first of 100 replications of the serial algorithm with a series of length 10^4 . The overall mean for the entire series (8.09) is indicated by a horizontal line.

has a large value of S but the stepwise changes to this matrix then tend to produce smaller values of the statistic than are obtained by generating values independently. This seems unlikely to occur except with an unusual series of matrices, and it can be guarded against by replicating the serial algorithm as described in the next section.

If necessary, the global P value can be estimated by generating a large number of independent species occurrence matrices, either using one of the methods for simulating random colonization that are referenced above, or a method such as that of Roberts and Stone (1990), which assumes that an independent matrix is achieved after making one or two thousand changes to a starting matrix.

It might be thought that because the P value from the serial test is not an unbiased estimate of the global

P value this implies that the null hypothesis being tested is not the same as the one that is tested by generating independent occurrence matrices with their corresponding test statistics. It must therefore be emphasized that the null hypothesis with the serial test is that the test statistic for the observed-occurrence matrix is a randomly chosen value from the distribution obtained from all possible occurrence matrices with the same row and column totals as the observed matrix (see Appendix). This is the same null hypothesis that is tested by comparing the observed test statistic with independently generated occurrence matrices. The principle used for testing is just different.

REPLICATING THE SERIAL TEST

It can be argued that the dependence of the serial method on a single choice of the position in the se-

TABLE 2. Estimated significance levels (P , in %) for the test statistics v_i for 20 individual species, with estimated standard errors (SE). The estimated significance levels are mean values from 100 replicates of a serial test with series lengths of $n = 10^4$. Species code as in Table 1.

	Species, number, i											
	1	2	3	4	5	6	7	8	9	10	11	12
P (%)	8.18	46.04	9.20	62.25	9.53	4.11	1.33	0.69	1.64	0.64	28.13	87.01
SE	0.16	0.46	0.14	0.38	0.21	0.08	0.03	0.02	0.06	0.02	0.05	0.02

† Significant at the 0.5% level.

‡ Significant at the 0.25% level.

quence where the observed test statistic will be situated is unsatisfactory. This limitation can be overcome by repeating the algorithm M times in order to obtain independent estimates p_1, p_2, \dots, p_M of the proportion of test statistics as large as or larger than that observed in a series of n statistics that includes the observed one. Then the mean of these estimates $\bar{p} = \Sigma p_i/M$ gives a combined estimate of the P value for the observed test statistic, which can be declared to be significant at the $100\alpha\%$ level if it is $< \alpha$. This replicated serial method has the advantage of providing information about the accuracy of the estimated P value because a standard error for \bar{p} can be estimated in the usual way by $\sqrt{\{\text{Var}(p_i)/M\}}$.

IDENTIFYING INTERACTING SPECIES

Simply deciding that an observed-occurrence matrix appears to display interaction between species is not satisfactory. In addition it is desirable to identify the nature of at least the main interactions. An immediate problem that arises is the large number of possible interactions that can occur. For example, in Table 1 data are given for 20 species. There are therefore $20 \times 19/2 = 190$ possible pairwise interactions. In other examples there might be as many as 50 species with 1225 possible pairwise interactions. To put it mildly, this generally rules out the possibility of estimating interaction parameters in some type of parametric model. However, progress is possible by asking which species exhibit non-random interaction with other species.

To this end, let e_{ij} denote the expected number of times that species i and j occur together on the same island if species occur independently of each other, with the row and column totals of the occurrence matrix fixed at those obtained for an observed matrix. Then the overall deviation from expected numbers for all species from an occurrence matrix can be measured by the mean square of deviations

$$S = \sum_{i=1}^R \sum_{j=1}^R (o_{ij} - e_{ij})^2/R^2, \quad (1)$$

where o_{ij} is the observed number of co-occurrences, and R is the number of species. This can also be written as

$$S = \sum_{i=1}^R v_i/R,$$

where

$$v_i = \sum_{j=1}^R (o_{ij} - e_{ij})^2/R \quad (2)$$

is a measure of the extent to which numbers of co-occurrences of other species with species i differ from expectation. Because $o_{ii} = e_{ii}$ (every species always occurs with itself) it matters little whether or not cases with $i = j$ are included or excluded from the sums in Eqs. 1 and 2. For simplicity of expression they are included here.

The following testing scheme is proposed based on these equations. First, S is tested to see whether it is significantly large and therefore gives overall evidence of deviations from expectations. This can be done either by using one application of the serial method or by repeating the serial algorithm M times and combining the results as discussed at the end of the last section. Secondly, all the v_i values are individually tested using the serial method, with or without replication, with the same generated-occurrence matrices as for the test on S . But, to make an allowance for the multiple testing of R species, v_i is required to be significantly large at the $100(\alpha/R)\%$ level before it is considered to provide evidence of interactions. This procedure for testing the v_i value for each species is based on the Bonferroni inequality, which says that if R tests are each carried out at the $100(\alpha/R)\%$ level then the probability of declaring any of them significant by chance is α or less (Alt 1982).

With many species this procedure requires a high level of significance before the result for an individual species gives evidence of interaction, which will only be possible with a large number of computer-generated test statistics. For example, with 50 species and a 5% level of significance v_i is required to be significantly large at the $5\%/50 = 0.1\%$ level before the result for species i is considered significant, which is just possible with 999 computer-generated statistics. However, the good thing about the serial method of testing is that even with a large occurrence matrix the computer time required to generate long series of statistics is not unreasonable.

In order to calculate statistics S and v_i of Eqs. 1 and 2 it is necessary to determine the expected numbers of co-occurrences, e_{ij} . This does not appear to be simple to do analytically, and therefore the approach proposed here is to use the mean values obtained from the n sets of data generated with each application of the serial algorithm. This results in the deviations from expectation being generally somewhat less than would be obtained by using the mean values for all possible occurrence matrices. However, because all of the n statistics generated from one application of the serial al-

TABLE 2. Continued.

Species number, i							
13	14	15	16	17	18	19	20
2.22	13.29	1.59	40.46	10.84	0.50†	0.06‡	1.58
0.06	0.23	0.06	0.33	0.18	0.02	0.01	0.06

gorithm will be affected in the same way, it seems that this should not result in much loss of power in the test.

EXAMPLE

The serial algorithm was repeated 100 times on the data in Table 1 with series of length 10^4 . That is to say, for each of the replications a random position between 1 and 10^4 was chosen for the observed-occurrence matrix and the remaining 9999 occurrence matrices were determined by random stepwise backwards and forwards changes as discussed above (see *Generalized Monte Carlo tests*). All random numbers used were generated using Wichmann and Hill's (1982) method.

For the first run of the serial algorithm the random starting position was 2868, and the test statistic for the observed data was $S_{2868} = 14.2$. This was the largest statistic in the whole series and therefore had a significance level of $(10^2/10^4)\% = 0.01\%$. From Fig. 1 it can be seen that the observed statistic is far from typical, and the series returned to more usual values of ≈ 5 –12 after ≈ 10 steps away from the observed-occurrence matrix in either direction.

From the 100 replications of the serial algorithm it was found that the average significance level was 0.050%, with an estimated standard error of 0.004%. There is therefore very strong evidence that the observed-occurrence matrix has a non-random pattern.

The significance levels for the v_i values for 20 individual species are shown in Table 2. With a Bonferroni adjustment for multiple testing significance levels of $10\%/20 = 0.5\%$ and $5\%/20 = 0.25\%$ are required in order to ensure that probabilities of declaring any result significant by chance are no more than 0.1 and 0.05, respectively. These levels are of course arbitrary to some extent, but still it is reasonable to regard a result for a species as indicating possible interactions if the significance level is $>0.25\%$ and $\leq 0.5\%$, and to provide clear evidence for interactions if the significance level is $\leq 0.25\%$. On this basis there is some evidence of interactions associated with species 18 (*Ctenosaurus hemilopha*), and clear evidence of interactions with species 19 (*Sator*).

Inspection of the occurrence matrix in Table 1 and differences between observed and expected numbers of co-occurrences with other species indicates that:

a) *Sator* occurs with other species 12 times, although the expected total number of co-occurrences is 29.1. The deficit of co-occurrences is found with all except species 2 (*Phyllodactylus*) and species 18 (*Ctenosaurus hemilopha*), but is particularly marked with species 6 (*Cnemidophorus tigris*) and species 10 (*Uta*). About 2.8 co-occurrences are expected with each of the species 6 and 10, but none are found.

b) *Ctenosaurus hemilopha* occurs with other species 16 times, although the expected number of occurrences is 29.1. In this case a comparison of observed and expected co-occurrences with other species does not indicate any particular interactions as being important.

It is more a question of a general tendency for co-occurrences to be too low.

The lack of co-occurrences of *Sator* with other species has been noted by Case (1983b), who points out that these lizards are fiercely aggressive in the field against smaller species. This might explain the absence of co-occurrences with smaller species, but not the absences with larger species. He concluded that the restricted distribution of *Sator* may just be due to a failure to disperse to some areas. Case also mentions in passing that the distribution of *Ctenosaurus hemilopha* may have been altered by man. This then offers one explanation for the apparently non-random occurrences of this species.

The generation of 100 series of 10^4 occurrence matrices (i.e., 10^6 matrices in all) by stepwise changes took ≈ 4 h on a 50 megahertz 486 computer. This may seem excessive given the fact that the serial test is valid even for quite short series. However it needs to be kept in mind that: (a) enough replications are needed to ensure a good spread of the starting points used and to provide good estimates of the standard errors of estimated P values; 100 replications is a realistic minimum from this point of view, and (b) each replication must consist of a series that is long enough so that the influence of the observed-occurrence matrix on the series is quite small, and high levels of significance can be obtained for individual species. Because significance levels of the order of 0.1% are required for evidence of interaction with individual species, series of length 10^3 are absolutely minimal. The choice of a length of 10^4 is therefore reasonable in the sense of being much more than the minimum required. In situations where high levels of significance are not necessarily needed, the series lengths could be much less.

Finally, a comment can be made about the number of random changes that were required in this example in order to change the observed-occurrence matrix into one with the appearance of a typical matrix from the sequences of 10^4 . This question was examined by comparing the distributions obtained for the overall test statistic S from the replicated series after 10, 20, 40, 80, 160, 320, 640, 1280, and 2560 changes. Without going into details, it can be said that the distribution of S was found to become more or less constant after 80 changes. It appears, therefore, that for this example this smaller number of steps is sufficient to remove most traces of the original matrix.

ACKNOWLEDGMENTS

I thank J. Bastow Wilson and two referees for constructive comments that have helped me to improve earlier versions of this paper.

LITERATURE CITED

- Alt, F. B. 1982. Bonferroni inequalities and intervals. *Encyclopedia of statistical sciences* 1:294–300.
- Barnard, G. A. 1963. Discussion on a paper by M. S. Bartlett. *Journal of the Royal Statistical Society* B25:294.

- Besag, J., and P. Clifford. 1989. Generalized Monte Carlo significance tests. *Biometrika* **76**:633–642.
- Brualdi, R. A. 1980. Matrices of zeros and ones with fixed row and column sum vectors. *Linear Algebra and its Applications* **33**:159–231.
- Case, T. J. 1983a. Niche overlap and the assembly of island lizard communities. *Oikos* **41**:427–433.
- . 1983b. The reptiles: ecology. Pages 159–209 in T. J. Case and M. L. Cody, editors. *Island biogeography in the Sea of Cortéz*. University of California Press, Berkeley, California, USA.
- Connor, E. E., and D. Simberloff. 1979. The assembly of species communities: chance or competition? *Ecology* **60**:1132–1140.
- Connor, E. E., and D. Simberloff. 1984. Neutral models of species' co-occurrence patterns. Pages 316–331 in D. R. Strong, D. Simberloff, L. G. Abele, and A. B. Thistle, editors. *Ecological communities: conceptual issues and the evidence*. Princeton University Press, Princeton, New Jersey, USA.
- Gilpin, M. E., and J. M. Diamond. 1984. Are species co-occurrence on islands non-random, and are null hypotheses useful in community ecology? Pages 297–315 in D. R. Strong, D. Simberloff, L. G. Abele, and A. B. Thistle, editors. *Ecological communities: conceptual issues and the evidence*. Princeton University Press, Princeton, New Jersey, USA.
- Manly, B. F. J. 1991. *Randomization and Monte Carlo methods in biology*. Chapman & Hall, London, England.
- . 1993. A review of computer intensive multivariate methods in ecology. Pages 307–346 in G. P. Patil and C. R. Rao, editors. *Multivariate environmental statistics*. North-Holland, Amsterdam, The Netherlands.
- Rasch, G. 1960. Probabilistic models for some intelligence and attainment tests. Danish Educational Research Institute, Copenhagen, Denmark.
- Roberts, A., and L. Stone. 1990. Island-sharing by archipelago species. *Oecologia* **83**:560–567.
- Ryti, R. T., and M. E. Gilpin. 1987. The comparative analysis of species occurrence patterns on archipelagos. *Oecologia* **73**:282–287.
- Stone, L., and A. Roberts. 1990. The checkerboard score and species distributions. *Oecologia* **85**:74–79.
- Stone, L., and A. Roberts. 1992. Competitive exclusion, or species aggregation: an aid in deciding. *Oecologia* **91**:419–424.
- Wichmann, B. A., and I. D. Hill. 1982. Algorithm AS2: an efficient and portable pseudo-random number generator. *Applied Statistics* **31**:188–190. [Correction in 1984: *Applied Statistics* **33**:123.]
- Wilson, J. B. 1987. Methods for detecting non-randomness in species co-occurrences: a contribution. *Oecologia* **73**:579–582.
- . 1988. Community structure in the flora of islands in Lake Manapouri, New Zealand. *Journal of Ecology* **76**:1030–1042.

APPENDIX

JUSTIFICATION FOR THE SERIAL TEST

The proposed significance test with the serial algorithm is justified by the fact that the probability of a significant result for a test at the $100\alpha\%$ level is α when the occurrence matrix being tested is randomly chosen from the set of all occurrence matrices with the same row and column totals.

To establish that this is the case, imagine an infinite sequence of occurrence matrices with the same row and column totals, where each of these matrices differs from the two neighboring matrices by random switches of species occurrences as described in the Introduction to this paper. Then the relative frequencies of all possible occurrence matrices with the given row and column totals will be equal in the sequence, and the overall distribution of a test statistic S calculated for each matrix in the sequence will correspond to the distribution generated by giving each of these matrices an equal probability of arising (Besag and Clifford 1989).

The null hypothesis to be tested is that the value of S for an observed-occurrence matrix is randomly chosen from the distribution in the full sequence, against the alternative hypothesis that the method by which the occurrence matrix was obtained tended to give large values of S . A one-sided test is therefore required, with large values of S casting doubt on the null hypothesis.

Suppose that an observed-occurrence matrix provides a test statistic S . Following the serial test procedure this value of S is then made to be in the i th position in a sequence of n such statistics, with the other values $S_1, S_2, \dots, S_{i-1}, S_{i+1}, \dots, S_n$ generated by making stepwise changes in the observed matrix, and i chosen as a random integer between 1 and n . In this way a fragment of length n from the infinite sequence of occurrence matrices is constructed such that all sequences of this length that include the observed matrix with its corresponding S value are equally likely to occur.

If the null hypothesis is true, so that S_i is a random choice from the values in the infinite sequence, then the fragment of length n will be a random choice from all possible fragments of this length. Thus S_1 to S_n will all have the same

distribution, although they will be correlated. Assuming that this is the case, let $P(r|i)$ be the probability that S_j is the r th in the list of values that is obtained when S_1 to S_n are placed in order from the smallest to the largest, with any tied values in a part of the list being randomly allocated an order within that part. Then, because i is randomly chosen from the integers 1 to n , the probability that S_i will be in the r th position in the ordered list is

$$\begin{aligned} P(r) &= \sum_{i=1}^n P(r|i)(1/n) \\ &= (1/n) \sum_{i=1}^n P(r|i) \\ &= 1/n \end{aligned}$$

because one, and only one, of the statistics is in the r th position in the ordered list.

It follows from this argument that the probability of an observed test statistic S being one of the k largest values in the generated sequence of n values is k/n if S is a random choice from the null hypothesis distribution. Therefore, an exact test at the $100(k/n)\%$ level is obtained by declaring S to be significantly large when it is one of the k largest values.

Note that it can be expected that $P(r|i)$ will vary with i . For example, the first statistic in a series is more likely to be the largest out of the n statistics than is a statistic in the middle of the series. This is because the similarity between neighboring statistics means that all of the largest values will tend to be close together in series. As a result, if S_1 is large then it will be competing with the first few values that follow it for the status of being the largest in the whole series as these will also necessarily be of a similar size. However, a large statistic in the middle of the series will be competing with statistics that occur both immediately before and immediately after it for the status of being the largest in the series. A middle statistic will therefore have more alternative statistics to compete with.