ROBUST STATISTICAL TESTS FOR DETECTING THE EFFECTS OF INTERSPECIFIC COMPETITION ON ECOLOGICAL COMMUNITIES

A Dissertation
Presented to the Faculty of the Graduate School
of Cornell University
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

by

Joshua Ladau

January 2006

ROBUST STATISTICAL TESTS FOR DETECTING THE EFFECTS OF INTERSPECIFIC COMPETITION ON ECOLOGICAL COMMUNITIES

Joshua Ladau, Ph.D.
Cornell University 2006

For seventy years ecologists have debated to what extent competition affects the composition of ecological communities. At one extreme, species have been proposed to assemble independently of each other, while at the other extreme, competition and other interspecific interactions have been proposed to account almost solely for the composition of communities. Although the debate bears broadly on applied and basic ecology, it has been challenging to resolve.

The most practical approach toward resolving the debate has been null model testing. The testing begins by assuming a null hypothesis that is reflective of an absence of competitive effects, which is then used to make a statistical prediction about the observed data. If observations are inconsistent with the prediction, then the null hypothesis is rejected, and effects of competition are inferred.

Unfortunately, as I show here, all existing null model tests are biased or non-robust. Although both qualities are problematic, the non-robustness is particularly troubling, because it means that when assumptions of the tests cannot be verified – as is usually the case – the tests will incorrectly indicate competitive effects unacceptably often. Thus, the tests are unreliable.

To fix the problem, I derive robust tests. Letting i and j denote the ith and jth colonists to arrive at a site, respectively, and W_{ij} the event that i and j belong to the

same “unit” (e.g., functional group, genus), I derive how partitions of colonists into units will be distributed if for all *i* and *j*, W_{ij} is conditionally independent of whether *i* and *j* share unit membership with the other colonists. Because the distribution can be derived without parametric assumptions, it can be used to test robustly for competitive effects.

I conclude by applying one of the tests to seven large data sets. In no cases does this test suggest effects of competition, although it does sometimes suggest effects of other interspecific interactions (e.g., facilitation). Overall, the predicted distribution accounts for over 95% of the variation in frequencies of partitions. Hence, the results suggest that although interspecific interactions may discernibly affect the composition of communities, those effects are generally minor.

BIOGRAPHICAL SKETCH

Joshua Ladau graduated from South Eugene High School in 1997, and received a B.S. with Distinction in Research in Entomology from Cornell University in 2000. He is the recipient of numerous awards and honors, including the Barry M. Goldwater and Westinghouse Science Talent Search Scholarships. He is originally from Oregon, and enjoys hiking, skiing, and swimming in his free time.

For Ron and Marge

ACKNOWLEDGEMENTS

The following people and organizations contributed greatly to my graduate research:

Roland Knapp: whose enthusiasm, advice, and hospitality were indispensable.

My Parents: who nurture my interests, and motivate and inspire me.

Thomas Eisner: whose support, wisdom, and patience made this work possible.

Steven J. Schwager: for a thorough critical reading of the dissertation, and for countless hours of invaluable discussion and assistance.

Committee Members: Thomas Eisner (Chair), Steven J. Schwager, Ronald R. Hoy, and H. Kern Reeve, for directing me through graduate school.

Daniel Dawson and the Sierra Nevada Aquatic Research Laboratory: for graciously providing the ideal working location.

Luana Maroja: for her sense of humor.

The Staff in the Department of Neurobiology and Behavior: Stacey Coil, Dawn Potter, Amy Bennet, Terri Natoli, and Janis Strope, who rescued me from administrative matters and deadlines.

National Science Foundation Graduate Research Fellowship Program: for copious funding.

National Park Service (Yosemite): for providing the location in which to test my theories, and to think up new ones.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER I

CHAPTER II

CHAPTER IV

LIST OF TABLES

CHAPTER I

CHAPTER II

CHAPTER IV

APPENDIX B

APPENDIX D

APPENDIX E

— CHAPTER I —

NULL MODEL TESTS OF SPECIES CO-OCCURRENCE: NULL HYPOTHESES AND BIOLOGICAL IMPLICATIONS

Abstract Null model tests have been extensively used to draw inferences from presence-absence data about effects of interspecific interactions on the composition of ecological communities. All tests published prior to 1996 test the same null hypothesis, that species occur independently. However, Gotelli and Graves (1996, *Null models in ecology*, Smithsonian Institution Press) and Gotelli (2000, *Ecology*, 81: 2606-2621) created several tests whose null hypothesis is less clear. Here I derive mathematical criteria to check whether these tests allow evaluation of the standard null hypothesis. I show that under many circumstances, they do not – their null distributions are inconsistent with the null hypothesis. However, the tests may allow evaluation of other null hypotheses relevant to understanding the effects of interspecific interactions on community composition.

INTRODUCTION

Do interspecific interactions affect large-scale patterns of species co-occurrence? The issue has proven contentious. Diamond (1975) suggested that they do, arguing that competition best explains the non-overlapping distributions of many bird species in the Bismarck Archipelago. However, Connor and Simberloff (1979) contended that such non-overlapping distributions can be expected by chance, and that interspecific interactions provide an unnecessarily complex explanation. Although the debate remains unresolved (e.g., Stone et al 2000, Brown et al 2002), Connor and Simberloff (1979) popularized an approach for addressing the matter, null model testing.

The data used in Connor and Simberloff's (1979) test consist of lists of species present at sets of sites. These "presence-absence data" are widely available (e.g., Patterson 1999), and can be obtained relatively easily. To summarize the data, a "presence-absence matrix" is constructed, in which rows and columns represent species and sites, respectively. If a particular species was observed at a particular site, a 1 is entered in the corresponding cell of the matrix; otherwise a 0 is entered. Connor and Simberloff's (1979) test begins by assuming that interspecific interactions did not affect the co-occurrence pattern. This assumption is translated into a mathematical statement, or null hypothesis, which, along with additional assumptions about colonization, comprises the "null model." The null model is then used to generate a "null distribution" of presence-absence matrices, which is consistent with an absence of interactive effects. If, as measured by a test statistic, the observed presence-absence matrix falls in a tail of this distribution, then the null hypothesis is rejected, and interspecific interactions are inferred to affect co-occurrence patterns.

Recently, Gotelli and Graves (1996) and Gotelli (2000) (hereafter referred to as "Gotelli and Graves") proposed four new null models ["SIM5" – "SIM8" in Gotelli (2000)] and corresponding tests. To simulate distributions of presence-absence matrices, the models sequentially place species according to specified probabilities, which vary depending on the placement of the preceding species (see below). Superficially, these models are plausible – species are placed "randomly."

Despite the plausibility of the models, their assumptions and null hypotheses have never been stated explicitly. This lack of explicitness presents two problems. First, like all statistical tests, Gotelli and Graves's tests are vulnerable to two types of errors: the null hypothesis can be rejected when it is true, or the null hypothesis can be accepted when it is false. In order to measure the rates of these errors and hence to assess the reliability of the tests, the assumptions and null hypothesis need to be defined explicitly [Bradley 1968; incorrect rates are reported in Gotelli (2000)]. Second, in null model testing, biological inferences are drawn from rejecting the null hypothesis. Thus, if the null hypothesis is unclear, any biological inferences will likewise be unclear.

Here, I investigate the null hypotheses of Gotelli and Graves's models. To do so, I begin by considering the null hypotheses of other null model tests, which are of two kinds. The first kind states that species are distributed independently of each other (Connor and Simberloff 1983, p. 463; Stone and Roberts 1990, p. 76). This null hypothesis is appropriate because interspecific interactions should cause species to occur non-independently – for instance, with competition, species should be less likely to occur when their competitors are present. The second null hypothesis states that species assort "randomly." In some cases, this null hypothesis has been employed to address a different biological question, whether or not colonization is stochastic (Connor & Simberloff 1978). In other cases, "randomness" and "independence" are used synonymously, although they have different meanings (e.g., Gotelli and McCabe 2002; see Stone & Roberts 1990). Thus, the null hypothesis of randomness is either inappropriate or equivalent to the null hypothesis of independence.

Herein I show that Gotelli and Graves's tests do not allow testing of the null hypothesis of independence. *A priori*, this finding is reasonable, because in the models, given that a particular species occurs at a site, subsequent arrivals have modified probabilities of belonging to the other species there. This finding clarifies the tests' biological implications and allows their error rates to be measured.

GOTELLI AND GRAVES'S MODELS

In the model SIM8 (Gotelli and Graves 1996, Gotelli 2000), probabilities are first assigned to each cell in the observed presence-absence matrix. The probabilities are chosen to sum to 1, and are proportional to the corresponding row and column totals (i.e., they are the product of the corresponding row and column totals, divided by the square of the grand total). Based on these probabilities, the first colonist's identity and location are selected. The probabilities are then renormalized to exclude the possibility of the second colonist belonging to the same species and occurring at the same site. That colonist is placed accordingly, and the process is repeated until the simulated and observed numbers of species-occurrences are equal. Using the same initial probabilities, many additional presence-absence matrices are then simulated, to create the null distribution. SIM7 and SIM6 follow the same filling algorithm, but have their initial probabilities set either (i) proportional to the row totals, but otherwise equal or (ii) proportional to the column totals, but otherwise equal, respectively (Gotelli 2000). In SIM5, columns (sites) are filled independently until the simulated and observed species richnesses match at each site, with the initial probabilities within each column set proportional to the corresponding row total (Gotelli 2000).

THE STANDARD NULL HYPOTHESIS

To check whether Gotelli and Grave's models can test the null hypothesis of independence, a key nuance of that null hypothesis requires explication. The nuance is illustrated by the null model of Connor and Simberloff (1979), as modified by

Wilson (1987). In this model, the null distribution contains just the presence-absence matrices having the observed row and column totals, each with equal probability measure. Hence, if the following presence-absence matrix was observed,

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \tag{1}$$

then the null distribution would consist of five presence-absence matrices, each equally likely:

$$P\left\{\begin{matrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{matrix}\right\} = P\left\{\begin{matrix} 0 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{matrix}\right\} = P\left\{\begin{matrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \end{matrix}\right\} = P\left\{\begin{matrix} 0 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \end{matrix}\right\} = P\left\{\begin{matrix} 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \end{matrix}\right\} = \frac{1}{5}. \tag{2}$$

Interestingly, examining this null distribution leads to an apparent contradiction: the null hypothesis states that species are distributed independently, suggesting that the first species should be equally likely to occur at the first site regardless of whether the second species occurs there. However, in the above distribution, the first species occurs with probability 0.5 when species 2 is absent, but probability 0 when it is present. One might wonder then, how can this model allow testing of the null hypothesis of independence?

The answer lies in the fact that the null distribution is conditioned on the observed row and column totals. For convenience, let O_1 and O_2 be the events that the first and second species occur at site 1, respectively, and let C be the event the observed row and column totals occur. Hence, letting O_2^c denote the complement of O_2, the model states that $P\{O_1 \mid O_2^c, C\} = 0.5$ and $P\{O_1 \mid O_2, C\} = 0$. Under conditional independence, these two terms would be equal. However, such independence differs from unconditional independence, which states for example that

$P\{O_1 \mid O_2^c\} = P\{O_1 \mid O_2\}$. Both conditional and unconditional independence are consistent with the verbal statement of the null hypothesis, and both could conceivably be used to test for interspecific interactions [Connor and Simberloff's (1979) model does allow testing for the latter]. The distinction between conditional and unconditional independence is important in evaluating Gotelli and Graves's models.

EVALUATING GOTELLI AND GRAVES'S MODELS

Like Connor and Simberloff's (1979) model, SIM5 – SIM8 condition on attributes of the sample. SIM6 – SIM8 condition on the total number of species-occurrences that were observed, and SIM5 conditions on the column totals of the observed presence-absence matrix. As per the preceding example, a cursory examination of null distributions generated by the models reveals that they cannot be generally implemented to test for conditional independence. However, as in Connor and Simberloff's (1979) model, they still may allow testing for unconditional independence.

In this section, I describe a method for checking this latter possibility, and in the next sections I apply that method. Theorem 1 (Appendices A) shows that for a model to be usable for testing for unconditional independence, its null distributions must possess certain properties. The properties can be illustrated by an example:

Suppose that a model gives a null distribution, in which the following presence-absence matrices occur with the following probabilities, conditional on attributes of the observed presence-absence matrix:

$$P\begin{Bmatrix}1 & 1 & 1\\ 1 & 1 & 0\\ 1 & 0 & 0\end{Bmatrix} = 0.05, \tag{3}$$

$$P\begin{Bmatrix}1 & 1 & 1\\ 1 & 1 & 0\\ 0 & 1 & 0\end{Bmatrix} = 0.04, \tag{4}$$

$$P\begin{Bmatrix}1 & 0 & 1\\ 1 & 1 & 1\\ 1 & 0 & 0\end{Bmatrix} = 0.01, \tag{5}$$

and

$$P\begin{Bmatrix}1 & 0 & 1\\ 1 & 1 & 1\\ 0 & 1 & 0\end{Bmatrix} = 0.09. \tag{6}$$

For convenience, denote these matrices $\mathbf{M}_1$ through $\mathbf{M}_4$, respectively. The set $\{\mathbf{M}_1, \mathbf{M}_2, \mathbf{M}_3, \mathbf{M}_4\}$ possesses two useful characteristics: (i) $\mathbf{M}_1$ and $\mathbf{M}_2$ share their first and second rows, as do $\mathbf{M}_3$ and $\mathbf{M}_4$, and (ii) $\mathbf{M}_1$ and $\mathbf{M}_3$ share their third rows, as do $\mathbf{M}_2$ and $\mathbf{M}_4$. Because of these characteristics, it can be shown that if species occur unconditionally independently, then in the model's null distribution:

$$\frac{P\{\mathbf{M}_1\}}{P\{\mathbf{M}_2\}} = \frac{P\{\mathbf{M}_3\}}{P\{\mathbf{M}_4\}} \tag{7}$$

(Appendices A and B). Hence, because in this example,

$$\frac{P\{\mathbf{M}_1\}}{P\{\mathbf{M}_2\}} = \frac{0.05}{0.04} \neq \frac{0.01}{0.09} = \frac{P\{\mathbf{M}_3\}}{P\{\mathbf{M}_4\}}, \tag{8}$$

it follows that the model cannot be used to test the null hypothesis of unconditional independence. Moreover, some (but not all) other quartets of presence-absence matrices will have the same property under unconditional independence.

METHODS

By checking null distributions generated by SIM5 – SIM8 for the properties predicted by Theorem 1 (Appendix A), I addressed two questions: First, are the null distributions consistent with unconditional independence? And second, when they are inconsistent, what is the magnitude of that inconsistency?

Presence-Absence Matrices Examined

SIM5 – SIM8 give different null distributions, dependent on the dimensions of the observed presence-absence matrix and the observed row and column totals. Hence, it was of interest to examine the consistency of the models with different observed presence-absence matrices. Figure 1 gives the matrices that were used.

Existence of Inconsistency

For presence-absence matrices 1-5, I computed empirical distribution functions (EDFs) using SIM6 - SIM8. Additionally, for presence-absence matrices 4-5, I computed EDFs using SIM5. For presence-absence matrices 1-3 and matrices 4-5, each EDF was based on 2×10^6 and 3×10^6 simulated matrices, respectively. All simulations were performed using EcoSim 7.70 Build 120404 (Gotelli & Entsminger 2001). This software is commonly used to implement Gotelli and Graves's models, and it simulates null distributions using Monte Carlo techniques.

To evaluate whether the EDFs were consistent with unconditional independence, I checked whether the quartets of presence-absence matrices listed in

Figure 1. Presence-absence matrices that were used to evaluate Gotelli and Graves's models. Matrices 6-11 were selected from the literature using the criteria of Gotelli and McCabe (2002) [respectively: Gotelli and Abele, Kodric-Brown and Brown 1993, Reed 1980, Bolger et al 1991 (matrices 9 and 10), Culver et al 1973]. All other matrices were created artificially. Matrices are listed in order of presentation in the text.

Matrix 1

0	0	1
0	1	0
1	0	1

Matrix 2

0	0	1
0	1	1
1	1	1

Matrix 3

0	0	1
1	1	1
1	1	1

Matrix 4

0	0	1
0	0	1
0	1	1
1	1	0

Matrix 5

0	0	1
0	0	1
1	1	0
1	1	0

Matrix 6

0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	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	0	0	0	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	1	1	1	1	1	1	1	0	0	1	1	1	1	0
0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
1	1	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	1	1	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	0	0	0	0	0	0
0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 1 (Continued)

Matrix 7

1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Matrix 8

1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1
1	0	0	0	0	1	0	0	0	0	0	0	1	0	0	1	0	1	0	0	1	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0
0	1	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	1	0	1	0	1	0	0	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
1	0	1	0	0	1	1	1	1	1	0	0	1	0	1	0	0	1	1	0	1	1	1	0	1
1	1	1	0	1	1	0	1	0	1	0	1	1	1	0	1	1	1	1	0	0	1	0	1	1

Matrix 9

1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 1 (Continued)

Matrix 10

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

Matrix 11

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

Matrix 12

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

Matrix 13

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

Matrix 14

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

Matrix 15

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

Matrix 16

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

Matrix 17

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

Matrix 18

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

Matrix 19

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

Matrix 20

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

Matrix 21

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

Matrix 22

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

Matrix 23

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

Appendix B had equal probability ratios, as predicted by Theorem 1. The quartets were chosen so that no presence-absence matrix occurred in more than one quartet. To check the ratios, I applied a generalized likelihood ratio test (GLRT). Let the total number of quartets under consideration be denoted by q, and let $\mathbf{M}_{i1}$ through $\mathbf{M}_{i4}$ denote the presence-absence matrices in quartet i so that by Theorem 1,

$$\frac{P\{\mathbf{M}_{i1}\}}{P\{\mathbf{M}_{i2}\}} = \frac{P\{\mathbf{M}_{i3}\}}{P\{\mathbf{M}_{i4}\}}. \tag{9}$$

Moreover, for $j = 1, 2, 3, 4$, let f_{ij} denote the simulated frequency of $\mathbf{M}_{ij}$, and let N be the total number of presence-absence matrices simulated. Then it can be shown that the generalized likelihood ratio statistic λ is

$$\lambda = \prod_i \left(\frac{\sum_{j=1}^{2} (f_{ij} + f_{i3}) \cdot (f_{ij} + f_{i4}) + \sum_{j=3}^{4} (f_{ij} + f_{i1}) \cdot (f_{ij} + f_{i2})}{N \cdot (f_{i1} + f_{i2} + f_{i3} + f_{i4})} \right). \tag{10}$$

Moreover, if the equalities of Theorem 1 hold, then $-2 \cdot \ln(\lambda)$ will be distributed approximately χ^2 with q degrees of freedom (Larsen & Marx 1986). For all tests, I set $\alpha = 0.003$ to give an overall significance level of 0.05. If the null hypothesis was rejected, I inferred that the corresponding EDF was inconsistent with unconditional independence.

For presence-absence matrices 6-11, the GLRT was unusable, because well over 10^6 presence-absence matrices needed to be simulated. However, for the null distribution of SIM5, numerical methods yielded exact probabilities (described in the next section). For each presence-absence matrix, I used these probabilities to check whether 100 quartets had equal probability ratios as predicted by Theorem 1.

Magnitude of Inconsistency

The magnitude of inconsistency can be quantified by the proportion of quartets violating the predictions of Theorem 1. One approach to measuring this proportion is to use EDFs generated by EcoSim. However, this approach suffers from two difficulties: Some quartets share presence-absence matrices, introducing non-independence, and very large sample sizes $(\gg 10^6)$ are needed to ensure adequate statistical power. An alternative approach, which I employ here, is to find exact null distributions. For SIM6 – SIM8, the drawback of this approach is that it can be applied only to small presence-absence matrices, because of the number of calculations that become necessary.

Exact null distributions can be found as follows. Suppose that the following presence-absence matrix is observed:

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}. \tag{11}$$

Then according to SIM8, the probabilities for placing the first colonist are

$$\begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} \equiv \begin{pmatrix} 2/9 & 1/9 \\ 4/9 & 2/9 \end{pmatrix}. \tag{12}$$

(That is, the first colonist has probability $p_{11} = 2/9$ of occurring at site 1 and belonging to species 1, etc.) By the definition of SIM8, it follows that the exact null distribution is as follows:

$$P\begin{Bmatrix} 0 & 1 \\ 1 & 1 \end{Bmatrix} = \sum_{\substack{i,j,k \in \{12,21,22\} \\ i,j,k \text{ distinct}}} p_i \cdot \frac{p_j}{1-p_i} \cdot \frac{p_k}{1-p_i-p_j} = 0.2135 \tag{13}$$

$$P\begin{Bmatrix} 1 & 0 \\ 1 & 1 \end{Bmatrix} = \sum_{\substack{i,j,k \in \{11,21,22\} \\ i,j,k \text{ distinct}}} p_i \cdot \frac{p_j}{1-p_i} \cdot \frac{p_k}{1-p_i-p_j} = 0.5079 \tag{14}$$

$$P\begin{Bmatrix}1 & 1\\ 0 & 1\end{Bmatrix} = \sum_{\substack{i,j,k\in\{11,12,22\}\\ i,j,k \text{ distinct}}} p_i \cdot \frac{p_j}{1-p_i}\cdot\frac{p_k}{1-p_i-p_j} = 0.0651 \tag{15}$$

$$P\begin{Bmatrix}1 & 1\\ 1 & 0\end{Bmatrix} = \sum_{\substack{i,j,k\in\{11,12,21\}\\ i,j,k \text{ distinct}}} p_i \cdot \frac{p_j}{1-p_i}\cdot\frac{p_k}{1-p_i-p_j} = 0.2135\,. \tag{16}$$

These arguments directly generalize to larger presence-absence matrices and other models.

For presence-absence matrices 1-5 and 12-23 (Figure 1), I calculated exact null distributions for SIM6 - SIM8, and for matrices 4-5 and 20-23, I additionally calculated exact null distributions using SIM5. For each matrix-model combination, I confirmed that the exact distributions agreed with EcoSim's distributions by simulating 25,000 presence-absence matrices, and performing goodness-of-fit testing. Next, from each exact distribution, I randomly chose a quartet of presence-absence matrices. If Theorem 1 made a prediction about the quartet, I checked the prediction by consulting the exact distribution; otherwise I disregarded the quartet. I repeated the process until 1000 prediction-making quartets had been considered for each matrix-model combination. I used the observed fractions of theorem-violating quartets to estimate the overall violation rates of the models, and calculated 95% confidence intervals.

All procedures were implemented using custom-written Visual Basic 6.0 software. For all measurements, degenerate presence-absence matrices were omitted (Gotelli 2000). However, it can be shown that if a model violates Theorem 1 with omission of degenerate matrices, then it will also violate it without omission (Ladau, unpublished results).

RESULTS

Existence of Inconsistency

For each model, at least one of the presence-absence matrices 1-5 gave an EDF inconsistent with independence (Table 1). For all matrices 6-11, SIM5 gave exact null distributions inconsistent with independence. For matrices 6 and 8, every quartet examined was inconsistent with independence, while for matrices 7, 9, 10, and 11; 84, 64, 30, and 99 quartets were inconsistent, respectively. (Non-random selection of quartets precludes inferences about overall rates of inconsistency here.)

Magnitude of Inconsistency

For presence-absence matrices 1-5 and 12-24, there were no significant differences between the exact distributions and the distributions given by EcoSim. For at least one matrix, each model generated a null distribution that was consistent with the property given by Theorem 1. However, for SIM6, SIM7, SIM8, and SIM5, respectively, on average 34.1%, 0.8%, 50.3%, and 12.9% of the quartets were violated per presence-absence matrix. Additionally, some presence-absence matrices gave violation rates as high as 75.2%, 5.7%, 95.7%, and 39.1% for each model, respectively (Table 2).

Table 1. P-values[1] for the existence of quartets violating Theorem 1 for Gotelli's models.

	Unfixed Rows, Unfixed Columns			Unfixed Rows, Fixed Columns
Matrix	Equiprobable Rows, Proportional Columns	Proportional Rows, Equiprobable Columns	Proportional Rows, Proportional Columns	Proportional Columns
1	0.7757	0.3945	0.0014	NA
2	6.14E-08	0.6049	3.20E-22	NA
3	0.0032	0.388	6.39E-11	NA
4	0.2862	0.0017	0.0575	0.1119
5	0.1096	0.0787	0.4613	3.39E-04

---

[1] Low p-values indicate the existence of a violation of Theorem 1. Shaded cells denote matrix-model combinations giving significant results after correction for multiple comparisons.

Table 2. Violation rates for Gotelli's models.

	Unfixed Rows, Unfixed Columns									Unfixed Rows, Fixed Columns		
	Equiprobable Rows, Proportional Columns			Proportional Rows, Equiprobable Columns			Proportional Rows, Proportional Columns			Proportional Columns		
Matrix	Agreement[1]	Point Estimate [2]	Confidence Interval[3]	Agreement	Point Estimate	Confidence Interval	Agreement	Point Estimate	Confidence Interval	Agreement	Point Estimate	Confidence Interval
1	0.2	0	(0,0)	0.2366	0	(0,0)	0.6735	0.402	(0.3714,0.4331)	NC	NA	NA
2	0.0646	0.743	(0.7147,0.7698)	0.7264	0	(0,0)	0.5925	0.957	(0.9425,0.9687)	NC	NA	NA
3	0.773	0.353	(0.3233,0.3835)	0.5653	0	(0,0)	0.357	0.511	(0.4795,0.5424)	NC	NA	NA
4	0.0154	0.714	(0.6849,0.7418)	0.3435	0.044	(0.0321,0.0586)	0.3199	0.843	(0.8189,0.865)	0.5391	0.391	(0.3606,0.422)
5	0.35	0	(0,0)	0.825	0.033	(0.0228,0.046)	0.8721	0.04	(0.0287,0.0541)	0.2088	0.383	(0.3528,0.4139)
12	0.6347	0.255	(0.2282,0.2832)	0.3075	0	(0,0)	0.8786	0.483	(0.4516,0.5145)	NC	NA	NA
13	0.8212	0.249	(0.2225,0.277)	0.2825	0	(0,0)	0.1735	0.519	(0.4875,0.5504)	NC	NA	NA
14	0.98	0.294	(0.2659,0.3233)	0.8639	0	(0,0)	0.3852	0.515	(0.4835,0.5464)	NC	NA	NA
15	0.0659	0.288	(0.2601,0.3172)	0.8346	0	(0,0)	0.07	0.477	(0.4456,0.5085)	NC	NA	NA
16	0.1583	0.752	(0.724,0.7785)	0.951	0	(0,0)	0.899	0.734	(0.7055,0.7612)	NC	NA	NA
17	0.6482	0	(0,0)	0.6902	0	(0,0)	0.7496	0	(0,0)	NC	NA	NA
18	0.5155	0.334	(0.3048,0.3642)	0.5299	0	(0,0)	0.7661	0.449	(0.4179,0.4804)	NC	NA	NA
19	0.1446	0.298	(0.2698,0.3274)	0.7889	0	(0,0)	0.9142	0.513	(0.4815,0.5444)	NC	NA	NA
20	0.1506	0.397	(0.3665,0.4281)	0.4761	0	(0,0)	0.6648	0.634	(0.6033,0.6639)	0.9378	0	(0,0)
21	0.5809	0.71	(0.6808,0.738)	0.0587	0	(0,0)	0.5965	0.853	(0.8295,0.8744)	0.4522	0	(0,0)
22	0.2566	0	(0,0)	0.6237	0	(0,0)	0.6878	0	(0,0)	0.1895	0	(0,0)
23	0.0977	0.409	(0.3783,0.4402)	0.6814	0.057	(0.0435,0.0732)	0.686	0.635	(0.6043,0.6649)	0.4351	0	(0,0)

[1] P-Value from goodness-of-fit test between exact and simulated distributions.

[2] Point estimate for fraction of violated equalites. Grey shading indicates matrix-model combinations with non-independence.

[3] 95% interval for fraction of violated equalities.

DISCUSSION

For many, but not all, presence-absence matrices, my results show that SIM5 – SIM8 generate null distributions inconsistent with the null hypothesis of unconditional independence. On theoretical grounds, such a result is expected. For instance, if all row and column totals are equal, the null distributions can be proven consistent with the null hypothesis. However, for most matrices, such consistency is not predicted, and as per the results, it is not observed. Because of the inconsistency, it appears that SIM5 – SIM8 cannot generally be applied to test the null hypothesis of independence.

Like other tests, the tests using SIM5 – SIM8 map the null distribution of presence-absence matrices onto a sampling distribution of statistics, from which the critical region is determined (Gotelli 2000). This mapping is not one-to-one; that is, different null distributions may result in the same sampling distributions. Hence, one might object that it is irrelevant to show that the null distributions are inconsistent with independence; what is really of concern is inconsistency of the sampling distributions with independence. However, checking the sampling distributions for consistency appears impractical – unlike the distributions of matrices, these distributions lack a "fingerprint" of independence, or its absence. Moreover, the mapping appears incapable of universally re-introducing independence (J. Ladau, unpublished data). Hence, SIM5 – SIM8 are indeed unusable for testing for independence.

Although rigorously testing for independence with SIM5 – SIM8 is impossible, perhaps approximate testing is still possible. However, precise tests of unconditional independence are readily implemented, so even if such testing is possible, it is unneeded.

In light of these findings, it is appropriate to consider tests employing SIM5 – SIM8 not as tests of the standard null hypothesis of independence, but rather as tests of a biologically and mathematically different null hypothesis. If interspecific interactions affect species co-occurrence patterns, then one potential consequence is that at a given site, species occurrence probabilities will depend on which other species are present – the consequence investigable with the standard null hypothesis. However, interspecific interactions could have other consequences as well. For instance, interspecific interactions might cause non-independence not in the arrival of species *per se*, but in the arrival of *individuals*. The corresponding null hypothesis – that individuals colonize independently – differs quantitatively from the standard null hypothesis, and yields differing predictions. The null distributions generated by SIM5 – SIM8 can be shown to be consistent with this null hypothesis (Appendix C).

Testing the null hypothesis of individual independence could offer two benefits. First, because it and the standard null hypothesis reflect different effects of interspecific interactions, testing both null hypotheses could increase the likelihood of detecting effects of interspecific interactions. Second, the Type I and II error rates (i.e., reliability) of null model tests are intimately connected to the choice of null hypothesis. Testing the null hypothesis of individual independence could reduce these error rates, although this remains to be checked.

In sum, understanding what null hypothesis can be tested with SIM5 – SIM8 is a key step towards correctly and meaningfully applying these models. In null model testing, it is the null hypothesis that articulates the biological process of interest into the mathematics, and only with a detailed knowledge of that articulation can the biological implications of the tests be fully appreciated and utilized.

ACKNOWLEDGEMENTS

This research was supported by a National Science Foundation Graduate Research Fellowship. It benefited greatly from the computer resources of the Sierra Nevada Aquatic Research Laboratory, and the input of Steven J. Schwager and Roland Knapp.

LITERATURE CITED

Bolger, D. T., A. C. Alberts, and M. E. Soule. 1991. Occurrence patterns of bird species in habitat fragments: sampling, extinction, and nested species subsets. American Naturalist **137:**155-166.

Bradley, J. V. 1968. Distribution-Free Statistical Tests. Prentice-Hall, Englewood Cliffs, New Jersey.

Brown, J. H., D. A. Kelt, and B. J. Fox. 2002. Assembly Rules and Competition in Desert Rodents. American Natrualist **160:**815-818.

Connor, E. F. and D. Simberloff. 1978. Species number and compositional similarity of the Galapagos flora and avifauna. Ecological Monographs **48:**219-248.

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

Connor, E. F. and D. Simberloff. 1983. Interspecific competition and species co-occurrence patterns on islands: null models and the evaluation of evidence. Oikos **41:**455-465.

Culver, D., J. R. Holsinger, and R. Baroody. 1973. Toward a predictive cave biogeography: the Greenbriar Valley as a case study. Evolution **27:**689-695.

Diamond, J. M. 1975. Assembly of species communities. Pages 342-344 *in* M. L. Cody and J. M. Diamond, eds. Ecology and Evolution of Communities. Harvard University Press, Cambridge.

Gilpin, M. E. and J. M. Diamond. 1982. Factors Contributing to Non-Randomness in Species Co-Occurrences on Islands. Oecologia **52:**75-84.

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

Gotelli, N. J., and L. G. Abele. 1982. Statistical distributions of West Indian land bird families. Journal of Biogeography **9:**421-435.

Gotelli, N. J., and A. M. Ellison. 2002. Assembly rules for New England and assemblages. Oikos **99:**591-599.

Gotelli, N.J. and G.L. Entsminger. 2001. EcoSim: Null models software for ecology. Version 7.0. Acquired Intelligence Inc. & Kesey-Bear. http://homepages.together.net/~gentsmin /ecosim.htm.

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

Gotelli, N. J., and D. J. McCabe. 2002. Species co-occurrence: a meta-analysis of J. M. Diamond's assembly rules model. Ecology **83:**2091-2096.

Kodric-Brown, A. and J. H. Brown. 1993. Highly structured fish communities in Australian desert springs. Ecology **74:**1847-1855.

Larsen, R. J., and M. L. Marx. 1986. An Introduction to Mathematical Statistics and its Applications. Prentice-Hall, Englewood Cliffs, New Jersey.

Patterson, B. D. 1999. http://aics-research.com/nestedness/tempcalc.html.

Reed, T. M. 1980. Turnover frequency in island birds. Journal of Biogeography **7:**329-335.

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

Stone, L., T. Dayan, and D. Simberloff. 2000. On desert rodents, favored states, and unresolved issues: scaling up and down regional assemblages and local communities. American Naturalist **156:**322-328.

Wilson, J. B. 1987. Methods for detecting non-randomness in the specie co-occurrences: a ontrcibution. Oecologia **73:**579-582.

— CHAPTER II —

NULL MODEL TESTS OF SPECIES CO-OCCURRENCE: ROBUSTNESS AND POWER

Abstract An enduring question in ecology is whether interspecific competition affects co-occurrence patterns of species. Testing null models constitutes a popular approach to address the question, but until recently, all tests have proven highly controversial, as it was unclear how prone they were to falsely suggesting competition (a Type I error) or its absence (a Type II error). However, recently numerous investigators have begun citing procedures and error rates reported by Gotelli (2000, *Ecology*, 81: 2606-2621) to justify the application of certain tests. Here, I show that Gotelli's (2000) procedures have several shortcomings that result in underestimation of Type I error rates. I also show that even if they were correct, Gotelli's (2000) error rates would not have pertained to most of the testing situations in which they have been cited. I also examine the error rates of 76 tests in addition to those examined by Gotelli (2000). Overall, my results suggest that (1) without detailed biological information, no existing null model tests may be applicable, and (2) even with such information, tests perform inconsistently for different data sets, so error rates should be assessed on a case-by-case basis.

INTRODUCTION

For seventy years ecologists have debated to what extent competition structures ecological communities. At one extreme, the "Gleasonian" viewpoint posits that species assemble randomly, assorting by chance and abiotic factors (Gleason 1926). At the other extreme, putative "assembly rules" suggest that interspecific interactions are responsible for the assortment of species (Diamond 1975). Among the interactions, competition is often cited as key (Diamond 1975, Gotelli and Graves 1996).

Differentiating between the two viewpoints has been challenging. At small spatial scales, experiments can provide answers; for instance Connell (1961). However, at the scale of communities, experimentation is often impractical and unethical – extirpating or introducing species is difficult and disruptive to ecosystems. Moreover, interspecific interactions may take hundreds of years to have measurable effects (Connor and Simberloff 1986).

Testing null models constitutes a popular approach to compensate for the lack of experimental evidence. The testing utilizes *presence-absence data* – lists of species present at sets of sites. Such data are commonly summarized in a *presence-absence matrix*, wherein each row represents a species, while each column represents a site. If species *i* was observed at site *j*, the *i,j*th entry of the matrix is a 1; otherwise it is a 0. The tests begin by specifying a null model, which consists of assumptions about colonization, and a null hypothesis that is consistent with competition not affecting community composition. The null model is then used to generate a distribution of presence-absence matrices, from which the probability of seeing a statistic summarizing the observed presence-absence matrix or one more extreme is calculated. If the probability is sufficiently low, the null hypothesis is rejected, and an effect of competition is inferred. Within this framework, tests employing different null models and different statistics have been developed.

Like all statistical testing, null model testing is prone to two types of error. A Type I error occurs if the null hypothesis is falsely rejected, here resulting in competitive effects being inferred when they are absent. A Type II error occurs if the null hypothesis is falsely accepted, which will result here in competitive effects being dismissed when they are present.

It is the susceptibility of null model tests to Type I and II errors that has prolonged the debate over the effects of competition, as tests indicating competitive effects or their absence have been criticized as vulnerable to Type I and II errors, respectively (Gotelli 2000, Gotelli 2001). However, in a key work, Gotelli (2000) systematically measured error rates of 36 null model tests, and suggested that some could be applied generally, particularly a test employing the null model "SIM9" and statistic "*C* score" (see below).

Although the suggestions of Gotelli (2000) were circumspect, the error rates of Gotelli (2000) have been widely used to justify applying certain tests. For instance, Heino and Soininen (2005) state that they chose SIM9 to examine co-occurrence patterns of diatoms because the model "is relatively robust to Type I and Type II errors, especially when used with the C-score (Gotelli 2000)" (p. 571). Numerous other investigators similarly cite Gotelli (2000) to support their choice of tests (e.g., Mouillot et al 2005, pp. 450-451; Ribichich 2005, pp. 90-91; Chaves and Anez 2004, p. 220; Feeley 2003, p. 107). Gotelli also cites Gotelli (2000). For example, to analyze a myriad of different presence-absence matrices, Gotelli and McCabe (2002) write that SIM9 with *C* score was chosen because it "has good Type I error properties and does not reject the null hypothesis too frequently when tested with random matrices (Gotelli 2000)" (p. 2093). Similar citations can be found in Gotelli and Ellison (2002; p. 593), Sanders et al (2003; p. 2475), and Gotelli and Rohde (2002; p. 90).

Although the practice is widespread, it is unclear whether citing Gotelli's (2000) error rates is sound. Gotelli (2000) uses non-standard techniques for measuring Type I and II error rates, and effectively makes measurements for only one presence-absence matrix. Here, I show that these non-standard techniques give incorrect error rates. Moreover, even if the rates were correct, I show that they would not hold for other presence-absence matrices, as tests perform differently depending on the presence-absence matrix being analyzed. Overall, in contrast to the *de facto* conclusions of Gotelli (2000), I find that in many situations, no existing null model tests have adequate Type I and II error rates.

TESTS AND PROCEDURES

Description of the Tests Examined by Gotelli (2000)

To simulate the distributions of presence-absence matrices under the null hypothesis, each test examined by Gotelli (2000) uses one of nine null models, or "algorithms," denoted "SIM1"-"SIM9." In SIM1, probabilities are initially selected of the first colonist belonging to each species and occurring at each site. That colonist is placed accordingly, and then the probabilities are renormalized for the second colonist, conditional on it belonging to a new species or occurring at a new site. After it is placed, the procedure is iterated until the total simulated and total observed species-occurrences match. The entire procedure is repeated many times to generate the null distribution. SIM6, SIM7, and SIM8 follow identical procedures, but begin with different probabilities. In SIM3 and SIM5, colonists are added to one column at a time, rather than to the entire matrix, until the simulated and observed column totals match. In SIM2 and SIM4, colonists are added to each row of the matrix until the simulated and observed row totals are equal. Finally, for SIM9, both species-occurrences and species richnesses are held equal to the observed totals, with all corresponding presence-absence matrices assumed equally likely. Additional details of the models can be found in Gotelli (2000).

All of the models besides SIM9 can produce degenerate matrices. These matrices are defined as matrices in which an entire row or column lacks species; matrices in which at least one species occurs nowhere or one site lacks species (Connor and Simberloff 1983, Gotelli 2000). Degenerate matrices are omitted from null distributions in Gotelli (2000).

The different models entail different null hypotheses and assumptions. In SIM1, SIM3, SIM5, SIM6, SIM7, and SIM8, the null hypothesis states that at each site, individuals arrive independently of those that are already present there. For SIM2 and SIM4, the null hypothesis states that species occur independently of each other, conditional on the number of times each is observed to occur. Last, for SIM9, the null hypothesis asserts that species occur unconditionally independently. Although these null hypotheses differ mathematically and biologically, all permit inferences about competitive effects (Chapter I). As for assumptions, SIM1 – SIM8 assume a set of probabilities used to place the first colonist. SIM9 assumes that all species are equally likely to occur at all sites.

To implement its null model, each test maps the null distributions of presence-absence matrices onto a sampling distribution of one of four statistics – the checkerboard score, *C* score, number of unique species combinations, or *V* ratio (respectively: Diamond 1975; Stone and Roberts 1990; Pielou and Pielou 1968; Robson 1972, Schluter 1984). The statistics presumably respond to non-independent colonization, with competition causing large values for the checkerboard and *C* score, and small values for the number of species combinations and *V* ratio. The probability of seeing the observed matrix or one more extreme is determined from the sampling distribution of the statistic.

Standard Procedures for Measuring Type I and II Error Rates

Because the Type I and II error rates are of broad statistical interest, standard procedures have been developed for measuring them.

Measuring the Type I error rate is simple when assumptions are valid: the rate is equal to the significance level (α). On the other hand, invalid assumptions may cause the true Type I error rate to deviate from α, depending on the test being used, value of α, sample size, and nature of the assumption violations (Bradley 1968). To measure Type I error rates under the latter circumstances, a three-step process is usually followed: First, a sampling distribution, or "nominal distribution," is derived assuming the validity of the null hypothesis and model assumptions, and a critical region at the nominal significance level (α) is determined. Second, another sampling distribution is created, assuming the null hypothesis and violated assumptions. Taking the latter to be the true "state of nature," the true significance level is found by determining the proportion of statistics from the new distribution that fall in the old critical region. Last, because many potential true states of nature usually exist, additional "true distributions" are then often checked (Bradley 1968). If Type I error rates consistently remain near the nominal significance level, a test is said to be "robust" (Bradley 1968, Larsen and Marx 1986). Hence, robustness provides the appropriate criterion for distinguishing tests using their Type I error rates.

Type II errors can only occur when the null hypothesis is false. The Type II error rate depends on many factors, including the exact nature of the hypothesis violation, nominal significance level, and sample size. Under a false null hypothesis, the error rate also depends on the validity of the model assumptions, but measurements are usually nontrivial even with valid assumptions (Bradley 1968, Larsen and Marx 1986). To measure Type II error rates, initially a sampling distribution and critical region are initially derived, under the null hypothesis and model assumptions. A second distribution is then created, assuming an alternative hypothesis, and the proportion of the statistics that fall in the initial critical region is calculated. This gives the power, which equals 1 – Type II error rate. In most situations, it is also necessary to check error rates using other relevant alternative hypotheses (Bradley 1968).

Distinguishing tests based directly on their Type II error rates can be subjective because the rates depend on the set of alternative hypotheses examined. An objective criterion is provided by bias. A test is *unbiased* if it rejects alternative hypotheses at least as often as it rejects the null hypothesis when assumptions are true (Knight 1999). Unbiased tests are highly preferable to biased ones.

Gotelli's (2000) Procedures for Measuring Type I and II Error Rates

Gotelli's (2000) procedures for measuring Type I and II error rates differ from the standard procedures. To measure Type I error rates, Gotelli (2000) uses the following procedure:

1) Four sets of 100 "test matrices" are created, using SIM1, SIM2, SIM4, and an algorithm in which rows are fixed and columns are proportional to the are of each site. Figure 2 gives the marginal totals used in each case.

2) With the same marginal totals, four sets of 1000 matrices using each of SIM1 – SIM9 are created. One of the four aforementioned statistics are then applied to each of the sets, creating sampling distributions of each statistic for each algorithm.

	Cuba	Hispaniola	Jamaica	Puerto Rico	Guadeloupe	Martinique	Dominica	St. Lucia	Barbados	St. Vincent	Grenada	Antigua	St. Croix	Grand Cayman	St. Kitts	Barbuda	MontSt.errat	St. Martin	St. Thomas	**Row Sum**
Carduelis dominicensis	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**
Loxia leucoptera	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**
Volantinia jacarina	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	**1**
Sprophila nigricolis	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	**1**
Melopyrrha nigra	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	**2**
Loxigilla portoricensis	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**
Loxigilla violacea	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**2**
Loxigilla noxis	0	0	0	0	1	1	1	1	1	1	1	1	0	0	1	1	1	1	0	**12**
Melanospiza richardsoni	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	**1**
Tiara olivacea	1	1	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	**5**
Tiara bicolor	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	**17**
Tiara canora	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**
Loxipasser anoxanthus	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**
Saltator albicollis	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	**4**
Torreornis inexpectata	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**
Ammodramus	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**3**
Zonotrichia capensis	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**
Column Sum:	**4**	**7**	**5**	**4**	**3**	**3**	**3**	**4**	**2**	**2**	**4**	**2**	**1**	**2**	**2**	**2**	**2**	**2**	**1**	**55**

Figure 2**.** Presence-absence matrix showing distribution of West Indian Finches. From Gotelli and Abele (1982).

3) For each of the original test matrices, each of the four statistics are calculated.

4) Using the sampling distributions found in step 2, the probability of observing the statistic from each test matrix, or one more extreme is found. The relative frequency with which the statistics fall in the upper and lower 5% tails of the sampling distribution are recorded, as well as the mean values of the statistics.

5) Across all four sets of test matrices, the mean of the frequencies from step 4 are calculated.

6) If the mean relative frequency exceeds 0.1, the corresponding test is concluded "susceptible" to Type I errors.

To measure Type II error rates, Gotelli (2000) uses the following procedure:

1) A 20×20 presence-absence matrix is created (Figure 3).

2) In each row, two randomly selected entries are swapped, and the four statistics are then calculated.

3) Each algorithm is run using the new marginal totals where appropriate. The probability of observing the statistic from step 2, or a statistic more extreme is found – where "extreme" is taken to mean "larger" for the checkerboard and C scores, and "smaller" for the number of unique species combinations and V ratio.

4) Steps 2 and 3 are iterated nine times, always swapping new entries.

5) Steps 1-4 are repeated four more times.

6) For each statistic and algorithm, the probabilities from step 4 are plotted against the number of rows swapped ("noise level"). Means are used to summarize the probabilities from the five runs (i.e., repetitions from steps 4 and 5).

7) The means are used to infer Type II error rates: For a given noise level, larger means are taken to indicate greater susceptibility to errors.

	Sites																				Row Sum
Species	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	**10**
	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	**10**
	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	**10**
	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	**10**
	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	**10**
	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	**10**
	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	**10**
	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	**10**
	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	**10**
	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	**10**
	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	**10**
	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	**10**
	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	**10**
	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	**10**
	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	**10**
	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	**10**
	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	**10**
	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	**10**
	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	**10**
	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	**10**
Column Sum:	**10**	**10**	**10**	**10**	**10**	**10**	**10**	**10**	**10**	**10**	**10**	**10**	**10**	**10**	**10**	**10**	**10**	**10**	**10**	**10**	**200**

Figure 3. Hypothetical presence-absence matrix used by Gotelli to measure Type II error rates.

ERRORS IN MEASUREMENTS

Type I Error Rates

For the most part, Gotelli's (2000) procedures parallel standard practice. However, contrary to standard practice, Gotelli's (2000) procedures derive each pair of nominal and true distributions not by assuming the same null hypothesis, but by employing systematically chosen pairs of null models. The problem with this approach is that some pairs of models entail different null hypotheses. For instance, deriving nominal and true distributions using SIM1 and SIM4, respectively, means varying the validity of both the assumptions and null hypotheses, and hence measuring a probability unrelated to the Type I error rate. To measure Type I error rates, only pairs of models with the same null hypotheses should be used.

In addition to the varying null hypotheses, also problematic is Gotelli's (2000) averaging of Type I error rates across four possible sets of assumption violations. This averaging procedure is employed because "there is no a priori way to decide which of the four kinds of random matrices… are most valid." In other words, averaging is used because it is impossible to know beforehand how the assumptions will be violated.

However, the averaging approach conflicts with the notion of robustness. Because Type I errors are sufficiently serious, robustness is usually defined to reflect the worst-case behavior of a test, not the average (Bradley 1968). A similar situation arises in testing compound null hypotheses: Compound null hypotheses usually specify a range of parameter values, rather than a single value. Thus, it is often impossible to define a critical region resulting in a single probability of falsely rejecting the null hypothesis – the probability varies with the actual parameter value, and α cannot equal *the* probability of a Type I error. While one approach would be to define the critical region so that α gives the mean probability, the maximum is preferable (Knight 1999). The same reasoning applies here: For a given test, the plausible assumption violations can result in any number of true Type I error rates. Because it is the worst-case behavior that is of concern, it is the maximum error rate that provides the appropriate criterion for distinguishing tests.

Needless to say, an infinite number of possible assumption violations would need to be considered to find the maximum Type I error rates of most tests – a daunting task. The most reasonable alternative is to examine instead a representative or large sample of possible assumption violations. Towards that end, Gotelli (2000) examines up to three valid violations (not eight, due to the problem with the null hypotheses) per test. However, numerous qualitatively different violations can be constructed, which could increase the maximum error rates of the tests. In the empirical portion of this study, I show that the maximum Type I error rates do indeed increase when additional assumption violations are considered.

Type II Error Rates

Key to measuring Type II error rates are alternative distributions and alternative hypotheses. However, Gotelli (2000) never explicitly refers to either, instead substituting neologisms. Although a semantic matter, these neologisms must be interpreted to demonstrate a more substantive mistake.

For alternative distributions, Gotelli (2000) writes that Type II error rates should be estimated by "evaluating a nonrandom test matrix… A test that was vulnerable to Type II error would fail to detect nonrandom patterns in such a structured matrix." By "nonrandom test matrix" Gotelli (2000) appears to mean "a presence-absence matrix chosen from a 'nonrandom' distribution." But Gotelli (2000) takes "random" to mean "conforming to the null hypothesis," so "nonrandom test matrix" means "matrix from an alternative distribution." Gotelli (2000) then creates ten types of nonrandom test matrices, corresponding to noise levels zero through nine, so it is those nonrandom matrices' distributions that comprise the alternative distributions.

With regard to alternative hypotheses, Gotelli (2000) writes that to create each alternative distribution, "one strategy would be to build a specific model of species interactions." Later on, he implies that "a mathematical model" could be used to create alternative distributions. From the context, then, "models" denotes "alternative hypotheses."

That said, it is evident that Gotelli's (2000) procedures fail to measure any Type II error rates, or powers. To calculate power, one should find the proportion of statistics from each alternative distribution that falls in the critical region for each test – or equivalently, the proportion of statistics that result in p-values less than some preset α. However, Gotelli (2000) instead finds the mean p-value. The mean is undoubtedly an unreliable indicator of the Type II error rate – it is easy to construct examples where higher means entail lower rates.

METHODS

The empirical component of this investigation addressed three questions:

First, for the presence-absence matrix analyzed by Gotelli (2000; Figure 2), what are the correct Type I and II error rates of the tests? I addressed this question using the standard techniques for measuring error rates and a broad set of assumption violations. I also examined bias.

Second, do the Type I and II error rates of the tests depend on the presence-absence matrix being analyzed? As mentioned above, robustness and power are usually sensitive to differences in sample size (Bradley 1968), so one might *a priori* expect that they would be sensitive to the dimensions and marginal totals of the matrices. Finding dependency would indicate that even had Gotelli's (2000) error rates been correct, employing them to justify broadly applying analyses would still have been mistaken. To check for dependency, I examined error rates of Gotelli's (2000) tests for four additional presence-absence matrices.

Finally, what are the error rates of other tests? In addition to the 36 tests examined in Gotelli (2000), over 76 tests have been proposed or follow immediately from proposals, in the literature. I examined the error rates of these tests in an attempt to assess the overall utility of existing tests.

Computation

All simulations and statistical tests were performed using custom-written Visual Basic 6.0 software.

Presence-Absence Matrices

I considered the following published presence-absence matrices: plants in the Windward and Leeward Islands (Beard 1948), myxomycetes in the high latitudes (Stephenson et al. 2000), finches in the West Indies (Gotelli and Abele 1982), seabirds near Vancouver Island (Hay 1992), and flies in the Hawaiian Islands (Hardy 1965;

Figure 2 and Figure 4). Some matrices were abridged to allow simulations to be completed within a finite timeframe. For convenience, I labeled the matrices I-V, respectively.

Type I Error Rates of Gotelli's (2000) Tests

For SIM1-SIM8, I considered only violations of the placement probabilities for the first colonist. Appendix D gives the assumption violations that were considered. I grouped these violations into two sets: the "limited set," consisting of just the violations considered by Gotelli (2000), and the "full set," consisting of Gotelli's (2000) violations and violations constructed here. For SIM9, I considered only violations of the assumption that all species are equally likely to occur everywhere. I again considered "limited" and "full" sets of assumption violations (Appendix D).

Excluding degenerate matrices, I next generated nominal distributions of 3000 statistics (i.e., 3000 iterations). Each distribution was generated using a unique combination of test and presence-absence matrix, and almost all possible combinations were used. The exceptions were for SIM9 with matrices I-III, where despite extensive attempts, computer algorithms could not be developed to generate efficiently "true distributions." For all models except SIM9, I generated nominal distributions using EcoSim 7.0 (Gotelli and Entminger, 2001). For SIM9, I generated null distributions using custom-written software.

It was impossible to create usable critical regions for every nominal distribution. For instance, as Gotelli (2000) points out, V ratio cannot be used with SIM9, because the resulting distribution contains only one value, allowing for α to be set only at 1. Similar problems can arise with other distributions; for example, if a distribution consists of only two statistics in equal frequency, α can be set only at 0.5

Figure 4. Presence-absence matrices used to measure Type I and II error rates. (A) Windward and Leeward Island Vegetation (Matrix I), (B) High-latitude Myxomycetes (Matrix II), (C) Seabirds near Vancouver Island (Matrix IV), and (D) Hawaiian Flies (Matrix V). References listed in the text.

(A)

	Puerto Rico	St. Kitts-Nevis	Montserrat	Guadeloupe	Dominica	Martinique	St. Lucia	St. Vincent	Grenada	Trinidad and Tobago	Row Sum
Nectandra patens	1	0	0	1	0	1	0	1	0	0	**4**
Tapura antillana	1	0	0	1	1	1	1	0	0	0	**5**
Sloanea berteriana	1	1	1	1	1	0	0	0	0	0	**5**
Krugia ferruginea	0	0	0	1	0	1	0	1	0	1	**4**
Licania ternatensis	0	0	0	1	1	0	1	1	0	1	**5**
Ocotea martinicensis	0	0	0	1	0	1	0	1	1	0	**4**
Turpinia occidentalis	0	0	0	1	1	1	0	1	1	0	**5**
Column Sum:	**3**	**1**	**1**	**7**	**4**	**5**	**2**	**5**	**2**	**2**	**32**

(C)

	Zone 1	Zone 2	Zone 3	Zone 4	Zone 5	Zone 6	Zone 7	Zone 8	Row Sum
Gavia immer	0	0	0	0	0	0	1	1	**2**
Pterodroma inexpectat	0	0	0	0	1	0	0	0	**1**
Larus philadelphia	0	1	1	1	0	1	0	0	**4**
Cepphus columba	0	0	0	0	1	0	1	1	**3**
Cyclorrhynchus psittac	1	0	0	0	0	0	0	0	**1**
Column Sum:	**1**	**1**	**1**	**1**	**2**	**1**	**2**	**2**	**11**

(B)

	Iceland	Khibine Mountains	Polar Ural	Yamal Peninsula	Plateau Putorana	Taimyr Peninsula	Magadan Region	Chukchi Peninsula	Seward Peninsula	Central Alaska	Northern Alaska	Greenland	Row Sum
Arcyodes oncarnata	0	1	0	0	0	1	0	1	0	0	0	0	**3**
Comatricha typhoides	0	0	0	0	0	0	1	1	1	1	0	0	**4**
Trichia botrytis	1	1	1	0	0	1	0	1	0	0	0	0	**5**
Didymium difforme	0	1	0	1	0	1	0	1	0	0	1	1	**6**
D. melanospermum	0	0	1	0	0	1	1	1	1	1	0	0	**6**
Physarum bivalve	0	1	0	0	1	1	1	1	1	1	0	0	**7**
Column Sum:	**1**	**4**	**2**	**1**	**1**	**5**	**3**	**6**	**3**	**3**	**1**	**1**	**31**

(D)

	Hawaii	Maui	Molokai	Lanai	Oahu	Kauai	Row Sum
Gitonides perspicax	1	1	1	1	1	1	**6**
Antopocerus diamphidiopus	0	1	1	0	0	0	**6**
Chymomyza procnemis	1	1	1	1	1	1	**3**
Drosophila adiastola	1	1	0	0	1	0	**5**
Drosophila cilifemorata	1	1	1	0	0	0	**3**
Drosophila crassifemur	1	1	1	0	1	1	**2**
Column Sum:	**5**	**6**	**4**	**5**	**3**	**2**	**25**

and 1. Therefore, taking 0.05 to be maximum allowable nominal Type I error rate, I excluded tests from further consideration if they allowed α to be set only at > 0.05.

If usable critical regions were possible, for distributions of *V* ratio and the number of unique species combinations, I selected the statistic that gave an empirical distribution function (EDF) of 0.05, or, if none existed, the pair of statistics that gave the next smallest and largest values. Denoting these statistics $w_{0.05}$, $w_{<0.05}$, and $w_{>0.05}$, respectively, I defined the critical regions $[0, w_{0.05}]$, or $[0, w_{<0.05}]$ and $[0, w_{>0.05}]$, as appropriate. For the distributions of *C* score and checkerboard score, I found the statistic that gave an EDF of 0.95, or the two statistics that gave the closest bounds. Denoting these statistics $w_{0.95}$, $w_{<0.95}$, and $w_{>0.95}$, I defined the critical regions $[w_{0.95}, \infty)$, or $[w_{<0.05}, \infty)$ and $[w_{>0.95}, \infty)$, again as appropriate.

Again excluding degenerate presence-absence matrices, I next generated true distributions of 1000 statistics each (i.e., 1000 iterations) using each assumption violation given in Appendix D. To calculate true significance levels, I computed the proportion of statistics from each true distribution that fell in the corresponding $\alpha = 0.05$ critical region, or if no such region existed, the proportions that fell in the bounding regions. In the latter case, I used linear interpolation to infer the significance level at $\alpha = 0.05$ (Zar 1999).

Type II Error Rates of Gotelli's (2000) Tests

To measure Type II error rates, I used the same nominal distributions and critical regions as for Type I error rates. For SIM1, SIM3, and SIM 5 - SIM9, I examined the following alternative hypotheses:

Let R and C denote the observed numbers of species and sites, respectively. The first alternative hypothesis asserted that for all $i = 1, 2, 3, \ldots.$; $j < R/2$, j odd, and $k = 1, 2, \ldots, C$, for SIM1, SIM3, and SIM5,

$$\frac{P\{\text{arrival } i \text{ belong to species } j \text{ at site } k \mid \text{species } j-1 \text{ present at site } k\}}{P\{\text{arrival } i \text{ belong to species } j \text{ at site } k \mid \text{species } j-1 \text{ absent from site } k\}} = 0.9, \tag{1}$$

while for SIM9,

$$\frac{P\{\text{species } j \text{ occurs at site } k \mid \text{species } j-1 \text{ present at site } k\}}{P\{\text{species } j \text{ occurs at site } k \mid \text{species } j-1 \text{ absent from site } k\}} = 0.9. \tag{2}$$

The second and third alternative hypotheses were identical, except the ratios were set equal to 0.5 and 0.1. I also examined a second set of alternative hypotheses, wherein for all $i = 1, 2, 3, \ldots.$; $j < R$, j odd, and $k = 1, 2, \ldots, C$, the ratios were set equal to 0.9, 0.5, and 0.1. All of the alternative hypotheses reflected asymmetric competitive interactions between the species one and two, three and four, etc.

To measure Type II error rates, all assumptions of the models were maintained. For each test and alternative hypothesis, I simulated alternative distributions of 1000 statistics. To calculate Type II error rates, for each alternative distribution, I found the proportion of statistics falling outside the corresponding $\alpha = 0.05$ critical region, or if no such region existed, the proportion falling outside the corresponding bounding regions. In the latter case, I used linear interpolation to infer the Type II error rate at $\alpha = 0.05$ (Zar 1999). For each test and matrix, I assessed bias by determining whether any Type II error rate exceeded 0.95.

Assessment and Consistency of Critical Regions and Robustness

I used Cohen's Kappa Statistic (Siegel 1988) to check whether tests consistently gave usable critical regions for different presence-absence matrices. To check whether robustness was consistent in different testing situations, for each presence-absence matrix I began by classifying tests with respect to robustness. To make the classifications, I set the maximum tolerable Type I error rate at 0.1, higher than most investigators would allow (Sokal and Rohlf 1995, Zar 1999). For each matrix, the simplest approach would then have been to characterize tests as either "robust" or "non-robust," depending on whether their error rates ever exceeded 0.1.

However, such an approach would suffer from two problems: First, in this study, only a few assumption violations were investigated. Hence, universal robustness could not be inferred if rates failed to exceed 0.1, as unexamined violations could still have produced intolerably high rates. On the other hand, if the rates exceeded 0.1 for any violation, this would suffice to establish non-robustness. Hence, it was appropriate to classify tests not as "robust" or "non-robust," but as "indeterminate" or "non-robust." This distinction is important, because unlike other tests, where robustness can be predicted analytically, robustness here is not expected.

The second problem would arise from differences between the observed Type I error rates and the rates implicitly defined by the algorithms and assumption violations, for even though an observed rate may have exceeded the specified level, the pertinent, actual rate may still have fallen below it. Two sources of sampling error could have contributed to these differences: error in the creation of the critical regions, and error in creation of the true distributions. I assumed the first source of error negligible, because of the large sample sizes used to approximate the nominal distributions ($n = 3000$). As for the second source of error, some terminology is helpful. Let α_{ijk} denote the actual Type I error rate for test *i*, assumption violation *j*, and presence-absence matrix *k*. Likewise, let a_{ijk} denote the corresponding observed rate. The normal approximation to the binomial distribution implies that

$$P\{\alpha_{ijk} \leq 0.1\} \cong 1 - \Phi[\frac{a_{ijk} - 0.1}{\sqrt{a_{ijk} \cdot (1 - a_{ijk})/1000}}] \; \forall i, j, k, \qquad (3)$$

where Φ is the standard normal cumulative distribution function (Ross 1998). Hence, by the independence of the true distributions, the probability of test *i* being "indeterminate" for presence-absence matrix *k* and assumption violation set *A* (i.e., the "limited" or "full" set) is

$$P\{\max\{\alpha_{ijk}\}_{j \in A} \leq 0.1\} = \prod_{j \in A} P\{\alpha_{ijk} \leq 0.1\} \cong \prod_{j \in A} [1 - \Phi[\frac{a_{ijk} - 0.1}{\sqrt{a_{ijk} \cdot (1 - a_{ijk})/1000}}]]. \qquad (4)$$

It immediately follows that the probability of being found "non-robust" is:

$$1 - \prod_{j \in A} [1 - \Phi[\frac{a_{ijk} - 0.1}{\sqrt{a_{ijk} \cdot (1 - a_{ijk})/1000}}]]. \qquad (5)$$

Mindful of these issues, I classified each test-matrix-violation set combination as "non-robust" only if it could be guaranteed non-robust with at least 99% certainty. I used Cohen's Kappa Statistic (Siegel 1988) to check whether the same tests had consistent robustness for different presence-absence matrices and sets of assumption violations.

Assessment and Consistency of Bias

Because the nominal significance level was taken at 0.05, tests were considered biased if they had a power of less than 0.05 (Knight 1999). As above, because not all possible alternative hypotheses were considered, tests could only be concluded biased, not unbiased. I hence characterized tests as either “biased” or “indeterminate.” Also, as with the Type I error rates, the observed Type II error rates were estimates of actual rates and suffered from the same sources of sampling error. Here, I once again considered error from the nominal distributions negligible. To take into account error from the alternative distributions, I let β_{ijk} denote the actual Type II error rate for analysis *i*, alternative hypothesis *j*, and presence-absence matrix *k*. Likewise, I let b_{ijk} denote the corresponding estimate. Because power will be less than 0.05 only if the Type II error rate exceeds 0.95, by the arguments given above, the probability of test *i* being found “biased” with presence-absence matrix *k* and alternative hypothesis set *B* is

$$1-\prod_{j\in B}[1-\Phi[\frac{b_{ijk}-0.95}{\sqrt{b_{ijk}\cdot(1-b_{ijk})/1000}}]]. \qquad (6)$$

I classified each test-matrix combination as “biased” only if it could be guaranteed biased with at least 99% certainty. I used Cohen’s Kappa statistic (Siegel 1988) to check whether the same tests were biased for all presence-absence matrices.

Assessment of Utility

To assess whether any of Gotelli’s (2000) tests were usable for presence-absence matrices I-V, I checked whether there existed a test having both indeterminate robustness and bias under the full set of assumption violations

In addition to the 36 tests considered by Gotelli (2000), I also considered the robustness and bias of 76 additional tests. All of these tests followed the framework of the test proposed by Gilpin and Diamond (1982): Probabilities were first assigned to each cell of the observed presence-absence matrix to give the likelihood of each corresponding species occurring at each corresponding site. At least four methods for choosing the probabilities have been proposed or follow directly from the literature (Figure 5 A; Gilpin and Diamond 1982, Gotelli 2000). Next, using the probabilities, the null distribution of presence-absence matrices was simulated. The null distribution was conditioned on one of five attributes of the observed presence-absence matrix (Figure 5 B; Gilpin and Diamond 1982, Gotelli 2000, Connor and Simberloff 1979). Finally, the null distribution was transformed into a sampling distribution. One of the four aforementioned statistics was used to perform this transformation (Gotelli 2000). Thus, within Gilpin and Diamond's (1982) framework, $4 \times 5 \times 4 = 80$ tests were possible; four of these tests corresponded to Gotelli's (2000) tests employing SIM9. All tests checked the null hypothesis that species occur independently of each other.

To evaluate the robustness and bias of the 76 tests not examined in Gotelli (2000), I used procedures similar to those outlined above, with the following changes: different assumption violations were considered (Appendix E), and only the alternative hypotheses given for SIM9 were considered.

RESULTS

Gotelli's (2000) Tests: Error Rates and Consistency of Performance

For the matrix examined in Gotelli (2000; Matrix III), maximum Type I error rates increased upon implementing standard procedures and additional assumption violations. Six tests proved to be biased when standard procedures were used. Table 3 and Table 4 give maximum Type I error rates, minimum power values, critical region existence, and classifications of each test-matrix-assumption violation set

Figure 5. Tests examined, in addition to those of Gotelli (2000). (A) Probabilities assumed by the tests that the *i*th species occurs at the *j*th site. Each cell of the table gives a set of possible probabilities. With respect to the rows and columns, the probabilities can be set "equiprobable," so that they are constant, or "proportional," so that they are proportional to the row or column totals. In the "proportional rows-proportional columns" case, it is sometimes necessary to use an *ad hoc* procedure to ensure that all values are less than one (Gilpin and Diamond 1982). Variables are defined as follows: $N \equiv$ total number of species-occurrences observed; $R \equiv$ number of rows in the presence-absence matrix; $C \equiv$ number of columns; $S_i \equiv i$th row total; $T_j \equiv j$th column total (Gotelli 2000). (B) Marginal constraints imposed by the tests. The null distribution can be conditioned on the row or column totals. If it is conditioned on neither, then it may or may not be conditioned on the total number of species-occurrences.

(A)

		Columns	
		Equiprobable	Proportional
Rows	Equiprobable	N/RC	T_j/N
	Proportional	S_i/N	$\sim S_iT_j/N$

(B)

		Columns	
		Unfixed	Fixed
Rows	Unfixed	Total number of species-occurrences fixed / Total number of species-occurrences unfixed	Total number of species-occurrences fixed
	Fixed	Total number of species-occurrences fixed	Total number of species-occurrences fixed

Table 3. Error Rates and Critical Region Existence of Gotelli's Null Model Tests

Model; Statistic[2]	MatrixI[1]				MatrixII				MatrixIII				MatrixIV				MatrixV			
	Critical Region	Bias	Maximum Type I Error Rate: Lim. Set	Maximum Type I Error Rate: Full Set	Critical Region	Bias	Maximum Type I Error Rate: Lim. Set	Maximum Type I Error Rate:Full Set	Critical Region	Bias	Maximum Type I Error Rate: Lim. Set	Maximum Type I Error Rate: Full Set	Critical Region	Bias	Maximum Type I Error Rate: Lim. Set	Maximum Type I Error Rate: Full Set	Critical Region	Bias	Maximum Type I Error Rate: Lim. Set	Maximum Type I Error Rate: Full Set
Er, Ec; C		0.052	0.035	0.962		0.04	0.018	0.955		0.053	0.006	0.609		0.043	0.043	0.462		0.053	0.027	0.565
Er, Fc; C		0.039	0.054	0.27		0.054	0.033	0.373		0.053	0	0.678		0.051	0.025	0.177		0.039	0.028	0.187
Er, Pc; C		0.043	0.346	0.994		0.046	0.304	0.995		0.049	0.236	0.852		0.043	0.05	0.435		0.046	0.13	0.858
Fr, Ec; C		NC	6E-04	0.954		NC	0.003	0.931		NC	0.009	0.652		NC	0.056	0.269		NC	0.031	0.474
Fr, Fc; C		NC	NC	NC		NC	NC	NC		NC	NC	NC		0.017	0.038	0.826	x	NA	NA	NA

[1] An 'x' in the "Critical Region" column denotes only unusable critical regions. 'NC' means that a given rate was not calculated due to limitations of computer algorithms. Robust.: Lim. Set" and "Robust.: Full Set" denote robustness measurements for the limited and full sets of assumption violations, respectively. Values in the bias column are the smallest observed powers, while values in the robustness columns are the largest observed Type I error rates. Shading indicates classification as "biased" or "non-robust," respectively.

[2] "Fr," "Er," and "Pr" denote models with fixed, equiprobable, and proportional rows, respectively. Likewise, "Fc," "Ec," and "Pc" denote models with fixed, equiprobable, and proportional columns, respectively. "C," "Ch," "Co," and "V" indicate C score, checkerboard score, number of unique species combinations, and V ratio, respectively.

Table 3 (Continued)

Model; Statistic	MatrixI				MatrixII				MatrixIII				MatrixIV				MatrixV			
	Critical Region	Bias	Maximum Type I Error Rate: Lim. Set	Maximum Type I Error Rate: Full Set	Critical Region	Bias	Maximum Type I Error Rate: Lim. Set	Maximum Type I Error Rate:Full Set	Critical Region	Bias	Maximum Type I Error Rate: Lim. Set	Maximum Type I Error Rate: Full Set	Critical Region	Bias	Maximum Type I Error Rate: Lim. Set	Maximum Type I Error Rate: Full Set	Critical Region	Bias	Maximum Type I Error Rate: Lim. Set	Maximum Type I Error Rate: Full Set
Fr, Pc; C		NC	0.42	0.991		NC	0.403	0.992		NC	0.264	0.811		NC	0.057	0.305		NC	0.093	0.658
Pr, Ec; C		0.06	0.078	0.973		0.036	0.077	0.967		NC	0.998	1		0.033	0.119	0.564		0.05	0.134	0.863
Pr, Fc; C		0.045	0.063	0.312		0.035	0.095	0.527		0.041	1	1		0.031	0.09	0.283		0.029	0.095	0.46
Pr, Pc; C		0.043	0.325	0.996		0.042	0.399	0.996		NC	1	1		0.046	0.12	0.566		0.045	0.239	0.96
Er, Ec; Ch		0.041	0.053	0.78		0.048	0.085	0.795		0.047	0.798	0.96		0.045	0.059	0.214		0.057	0.116	0.691
Er, Fc; Ch		NA	NA	NA	x	NA	NA	NA	x	0.046	0.923	0.974		0.043	0.058	0.591		NA	NA	NA
Er, Pc; Ch		0.049	0.169	0.925		0.052	0.252	0.911		0.048	0.941	0.993		0.051	0.055	0.234		0.047	0.179	0.786
Fr, Ec; Ch		NC	0.013	0.943		NC	0.013	0.833		NC	0.013	0.194		NC	0.042	0.327		NA	NA	NA
Fr, Fc; Ch		NC	NC	NC		NC	NC	NC		NC	NC	NC		0.03	0.076	0.076	x	NA	NA	NA
Fr, Pc; Ch		NC	0.14	0.978		NC	0.121	0.872		NC	0.184	0.433		NC	0.046	0.325		NC	0.06	0.699
Pr, Ec; Ch		0.04	0.039	0.788		0.046	0.021	0.741		NC	0.008	0.272		0.042	0.048	0.169		0.044	0.028	0.585
Pr, Fc; Ch		NA	NA	NA	x	NA	NA	NA	x	0.04	0	0.195		0.05	0.044	0.586		NA	NA	NA
Pr, Pc; Ch		0.057	0.193	0.93		0.043	0.126	0.832		NC	0.172	0.471		0.041	0.05	0.187		0.036	0.088	0.636
Er, Ec; Co		0.048	0.069	0.865		0.047	0.097	0.848		0.04	0.511	0.972		0.044	0.079	0.188		0.047	0.114	0.741
Er, Fc; Co		0.047	0.05	0.341		0.042	0.057	0.342		0.05	0.341	0.889		0.045	0.069	0.246		NA	NA	NA
Er, Pc; Co		0.036	0.062	0.857		0.051	0.079	0.771		0.039	0.316	0.932		0.043	0.077	0.183		0.042	0.095	0.738

Table 3 (Continued)

Model; Statistic	MatrixI				MatrixII				MatrixIII				MatrixIV				MatrixV			
	Critical Region	Bias	Maximum Type I Error Rate: Lim. Set	Maximum Type I Error Rate: Full Set	Critical Region	Bias	Maximum Type I Error Rate: Lim. Set	Maximum Type I Error Rate:Full Set	Critical Region	Bias	Maximum Type I Error Rate: Lim. Set	Maximum Type I Error Rate: Full Set	Critical Region	Bias	Maximum Type I Error Rate: Lim. Set	Maximum Type I Error Rate: Full Set	Critical Region	Bias	Maximum Type I Error Rate: Lim. Set	Maximum Type I Error Rate: Full Set
Fr, Ec; Co		NC	0.067	0.847		NC	0.078	0.925		NC	0.116	0.839		NC	0.052	0.292		NC	0.044	0.361
Fr, Fc; Co		NC	NC	NC		NC	NC	NC		NC	NC	NC		0.023	0.038	0.471	x	NA	NA	NA
Fr, Pc; Co		NC	0.032	0.844		NC	0.035	0.876		NC	0.019	0.767		NC	0.04	0.267		NC	0.048	0.362
Pr, Ec; Co		0.052	0.062	0.872		0.051	0.069	0.771		NC	0.078	0.78		0.043	0.052	0.143		0.041	0.04	0.625
Pr, Fc; Co		0.043	0.043	0.31		0.044	0.033	0.307		0.035	0	0.511		0.045	0.041	0.206		NA	NA	NA
Pr, Pc; Co		0.045	0.057	0.84		0.043	0.032	0.711		NC	0.019	0.733		0.038	0.048	0.139		0.046	0.056	0.644
Er, Ec; V		0.044	0.061	0.457		0.051	0.041	0.603		0.045	0.038	0.277		0.052	0.055	0.123		0.056	0.057	0.637
Er, Fc; V		0.038	0.04	0.093		0.038	0.026	0.11		0.045	0	0.275		NA	NA	NA	x	NA	NA	NA
Er, Pc; V		0.043	0.456	0.948		0.052	0.4	0.968		0.055	0.363	0.829		0.036	0.035	0.131		0.047	0.116	0.662
Fr, Ec; V		NC	9E-04	0.483		NC	0.003	0.402		NC	1E-03	0.137		NA	NA	NA	x	NA	NA	NA
Fr, Fc; V	x	NA	NA	NA	x	NA	NA	NA	x	NA	NA	NA	x	NA	NA	NA	x	NA	NA	NA
Fr, Pc; V		NC	0.426	0.942		NC	0.416	0.908		NC	0.283	0.588		NA	NA	NA	x	NC	0.093	0.489
Pr, Ec; V		0.049	0.043	0.439		0.036	0.047	0.62		NC	0.06	0.366		0.039	0.087	0.161		0.051	0.049	0.621
Pr, Fc; V		0.025	0.04	0.093		0.035	0.098	0.201		0.05	1	1		0.04	0.101	0.139		0.01	0.125	0.274
Pr, Pc; V		0.054	0.469	0.948		0.057	0.456	0.981		NC	0.402	0.843		0.048	0.124	0.213		0.044	0.176	0.721

Table 4. Error Rates and Critical Region Existence of Null Model Tests Other than Gotelli's

Probabilities; Condition; Statistic^2	MatrixI[1]				MatrixII				MatrixIII				MatrixIV				MatrixV			
	Critical Region	Bias	Maximum Type I Error Rate: Lim. Set	Maximum Type I Error Rate: Full Set	Critical Region	Bias	Maximum Type I Error Rate: Lim. Set	Maximum Type I Error Rate:Full Set	Critical Region	Bias	Maximum Type I Error Rate: Lim. Set	Maximum Type I Error Rate: Full Set	Critical Region	Bias	Maximum Type I Error Rate: Lim. Set	Maximum Type I Error Rate: Full Set	Critical Region	Bias	Maximum Type I Error Rate: Lim. Set	Maximum Type I Error Rate: Full Set
Er, Ec; Fr, Uc; C		0.044	0.052	0.999		0.056	0.042	0.994		NC	NC	NC		0.038	0.043	0.308		0.059	0.06	0.573
Er, Ec; Ur, Fc; C		0.045	0.056	0.369		0.051	0.064	0.558		0.044	0.046	0.839		0.043	0.047	0.188		0.065	0.051	0.128
Er, Ec; Ur, Uc, Ft; C		0.049	0.034	0.998		0.049	0.02	0.999		0.045	0.002	0.779		0.055	0.056	0.66		0.074	0.003	0.208
Er, Ec; Ur, Uc, Ut; C		0.053	0.041	1		0.066	0.01	0.985		0.04	0.02	0.101		0.043	0.029	0.258		0.05	0	0.167
Er, Pc; Fr, Fc; C		NC	NC	NC	x	NA	NA	NA		NC	NC	NC		0.035	0.077	0.894	x	NA	NA	NA
Er, Pc; Fr, Uc; C		0.05	0.967	1		0.036	0.969	1		NC	NC	NC		0.045	0.065	0.352		0.033	0.624	0.959
Er, Pc; Ur, Fc; C		0.048	0.041	0.326		0.046	0.046	0.543		0.059	0.053	0.862		0.051	0.056	0.208		0.055	0.05	0.156

[1] An 'x' in the "Critical Region" column denotes only unusable critical regions. 'NC' means that a given rate was not calculated due to limitations of computer algorithms. Robust.: Lim. Set" and "Robust.: Full Set" denote robustness measurements for the limited and full sets of assumption violations, respectively. Values in the bias column are the smallest observed powers, while values in the robustness columns are the largest observed Type I error rates. Shading indicates classification as "biased" or "non-robust," respectively.

[2] "Er," "Pr," "Fr," and "Ur" denote equiprobable, proportional, fixed rows, and unfixed rows, respectively. Likewise, "Ec," "Pc," "Fc," and "Uc" denote equiprobable, proportional, fixed rows, and unfixed columns respectively. "Ut" denotes unfixed total.

Table 4 (Continued)

	MatrixI				**MatrixII**				**MatrixIII**				**MatrixIV**				**MatrixV**			
Probabilities; Condition; Statistic	**Critical Region**	**Bias**	**Maximum Type I Error Rate: Lim. Set**	**Maximum Type I Error Rate: Full Set**	**Critical Region**	**Bias**	**Maximum Type I Error Rate: Lim. Set**	**Maximum Type I Error Rate:Full Set**	**Critical Region**	**Bias**	**Maximum Type I Error Rate: Lim. Set**	**Maximum Type I Error Rate: Full Set**	**Critical Region**	**Bias**	**Maximum Type I Error Rate: Lim. Set**	**Maximum Type I Error Rate: Full Set**	**Critical Region**	**Bias**	**Maximum Type I Error Rate: Lim. Set**	**Maximum Type I Error Rate: Full Set**
Er, Pc; Ur, Uc, Ft; C		0.048	0.866	1		0.051	0.846	1		0.05	0.345	0.964		0.058	0.064	0.669		0.048	0.417	0.777
Er, Pc; Ur, Uc, Ut; C		0.063	0.856	1		0.062	0.838	1		0.039	0.117	0.17		0.04	0.055	0.297		0.054	0.359	0.622
Pr, Ec; Fr, Fc; C	x	NA	NA	NA		NC	0.081	0.081		NC	NC	NC		0.036	0.067	0.864	x	NA	NA	NA
Pr, Ec; Fr, Uc; C		0.07	0.046	0.999		0.062	0.041	0.988		0.069	0	0.626		0.035	0.059	0.328		0.05	0.047	0.584
Pr, Ec; Ur, Fc; C		0.044	0.071	0.402		0.046	0.154	0.724	x	NA	NA	NA		0.053	0.191	0.495		0.048	0.93	0.986
Pr, Ec; Ur, Uc, Ft; C		0.059	0.038	0.998		0.04	0.122	0.999		0.044	1	1		0.049	0.156	0.81		0.036	0.681	0.914
Pr, Ec; Ur, Uc, Ut; C		0.048	0.056	0.997		0.05	0.113	0.996		0.053	0.983	0.994		0.045	0.167	0.463		0.048	0.64	0.883
Pr, Pc; Fr, Fc; C		0.028	0	0		0.052	0.346	0.346		NC	NC	NC		0.05	0.121	0.956	x	NA	NA	NA
Pr, Pc; Fr, Uc; C		0.049	0.926	1		0.03	0.869	1		0.057	0.503	0.694		0.053	0.078	0.396		0.046	0.165	0.856
Pr, Pc; Ur, Fc; C		0.065	0.082	0.456		0.05	0.283	0.861		0.148	1	1		0.05	0.227	0.524		0.037	0.917	0.982
Pr, Pc; Ur, Uc, Ft; C		0.049	0.802	1		0.049	0.88	1		0.047	1	1		0.043	0.159	0.82		0.043	0.787	0.962
Pr, Pc; Ur, Uc, Ut; C		0.058	0.77	1		0.049	0.853	1		0.051	1	1		0.045	0.205	0.479		0.042	0.691	0.925
Er, Ec; Fr, Uc; Ch		0.059	0.058	0.999		0.056	0.049	0.975		NC	NC	NC		0.045	0.052	0.321	x	NA	NA	NA
Er, Ec; Ur, Fc; Ch	x	NA	NA	NA	x	NA	NA	NA		0.056	1	1		0.047	0.082	0.721	x	NA	NA	NA
Er, Ec; Ur, Uc, Ft; Ch		0.051	0.053	0.97		0.046	0.125	0.957		0.051	0.943	0.997		0.052	0.059	0.307		0.07	0.349	0.52

Table 4 (Continued)

Probabilities; Condition; Statistic	MatrixI				MatrixII				MatrixIII				MatrixIV				MatrixV			
	Critical Region	Bias	Maximum Type I Error Rate: Lim. Set	Maximum Type I Error Rate: Full Set	Critical Region	Bias	Maximum Type I Error Rate: Lim. Set	Maximum Type I Error Rate:Full Set	Critical Region	Bias	Maximum Type I Error Rate: Lim. Set	Maximum Type I Error Rate: Full Set	Critical Region	Bias	Maximum Type I Error Rate: Lim. Set	Maximum Type I Error Rate: Full Set	Critical Region	Bias	Maximum Type I Error Rate: Lim. Set	Maximum Type I Error Rate: Full Set
Er, Ec; Ur, Uc, Ut;		0.041	0.048	0.862		0.044	0.084	0.828		0.057	0.247	0.755		0.033	0.039	0.441		0.067	0.186	0.372
Er, Pc; Fr, Fc; Ch	x	NA	NA	NA	x	NA	NA	NA		NC	NC	NC		0.037	0.074	0.074	x	NA	NA	NA
Er, Pc; Fr, Uc; Ch	x	NA	NA	NA	x	NA	NA	NA		NC	NC	NC		0.04	0.061	0.35	x	NA	NA	NA
Er, Pc; Ur, Fc; Ch	x	NA	NA	NA	x	NA	NA	NA		0.056	1	1		0.049	0.095	0.708	x	NA	NA	NA
Er, Pc; Ur, Uc, Ft;	x	NA	NA	NA	x	NA	NA	NA		0.053	0.991	1		0.041	0.063	0.297	x	NA	NA	NA
Er, Pc; Ur, Uc, Ut;	x	NA	NA	NA	x	NA	NA	NA		0.039	0.712	0.957		0.051	0.066	0.508	x	NA	NA	NA
Pr, Ec; Fr, Fc; Ch		NC	NC	NC	x	NA	NA	NA		NC	NC	NC		0.034	0.078	0.078	x	NA	NA	NA
Pr, Ec; Fr, Uc; Ch		0.058	0.049	0.998		0.059	0.06	0.974		0.04	0.005	0.176		0.043	0.054	0.34	x	NA	NA	NA
Pr, Ec; Ur, Fc; Ch	x	NA	NA	NA	x	NA	NA	NA	x	NA	NA	NA		0.042	0.076	0.692	x	NA	NA	NA
Pr, Ec; Ur, Uc, Ft;		0.066	0.036	0.97		0.056	0.026	0.89		0.046	0.009	0.321		0.048	0.053	0.261		0.049	0.05	0.12
Pr, Ec; Ur, Uc, Ut;		0.044	0.043	0.873		0.06	0.023	0.797		0.053	0.015	0.474		0.053	0.059	0.489		0.033	0.034	0.11
Pr, Pc; Fr, Fc; Ch	x	NA	NA	NA	x	NA	NA	NA		NC	NC	NC		0.023	0.034	0.039	x	NA	NA	NA
Pr, Pc; Fr, Uc; Ch		0.043	0.418	1		0.055	0.179	0.987		0.05	0.211	0.211		0.038	0.048	0.317		0.041	0.08	0.781
Pr, Pc; Ur, Fc; Ch	x	NA	NA	NA	x	NA	NA	NA		0.016	0	0.164		0.037	0.035	0.679	x	NA	NA	NA
Pr, Pc; Ur, Uc, Ft;		0.054	0.42	0.998		0.029	0.098	0.948		0.037	0.206	0.559		0.044	0.051	0.265		0.039	0.099	0.195

Table 4 (Continued)

Probabilities; Condition; Statistic	MatrixI				MatrixII				MatrixIII				MatrixIV				MatrixV			
	Critical Region	Bias	Maximum Type I Error Rate: Lim. Set	Maximum Type I Error Rate: Full Set	Critical Region	Bias	Maximum Type I Error Rate: Lim. Set	Maximum Type I Error Rate:Full Set	Critical Region	Bias	Maximum Type I Error Rate: Lim. Set	Maximum Type I Error Rate: Full Set	Critical Region	Bias	Maximum Type I Error Rate: Lim. Set	Maximum Type I Error Rate: Full Set	Critical Region	Bias	Maximum Type I Error Rate: Lim. Set	Maximum Type I Error Rate: Full Set
Pr, Pc; Ur, Uc, Ut;		0.05	0.464	0.993		0.057	0.234	0.935		0.045	0.169	0.718		0.047	0.07	0.517		0.043	0.078	0.164
Er, Ec; Fr, Uc; Co		0.05	0.15	0.995		0.044	0.198	0.993		NC	NC	NC		0.032	0.066	0.337		0.039	0.085	0.428
Er, Ec; Ur, Fc; Co		0.052	0.069	0.613		0.041	0.095	0.432		0.052	0.918	0.961		0.042	0.075	0.255	x	NA	NA	NA
Er, Ec; Ur, Uc, Ft;		0.048	0.128	0.979		0.041	0.234	0.966		0.044	0.901	1		0.048	0.101	0.265		0.061	0.297	0.563
Er, Ec; Ur, Uc, Ut;		0.045	0.139	0.979		0.046	0.2	0.944		0.033	0.755	1		0.054	0.111	0.529		0.049	0.234	0.424
Er, Pc; Fr, Fc; Co		NC	NC	NC	x	NA	NA	NA		NC	NC	NC		0.041	0.063	0.518	x	NA	NA	NA
Er, Pc; Fr, Uc; Co		0.022	0.038	0.963		0.03	0.03	0.939		NC	NC	NC		0.031	0.047	0.337		0.006	0.029	0.346
Er, Pc; Ur, Fc; Co		0.037	0.052	0.58		0.047	0.095	0.436		0.044	0.918	1		0.05	0.08	0.262	x	NA	NA	NA
Er, Pc; Ur, Uc, Ft;		0.039	0.045	0.933		0.039	0.07	0.867		0.035	0.756	1		0.049	0.1	0.252		0.031	0.134	0.383
Er, Pc; Ur, Uc, Ut;		0.032	0.062	0.912		0.028	0.082	0.875		0.038	0.582	1		0.05	0.108	0.524		0.025	0.087	0.218
Pr, Ec; Fr, Fc; Co	x	NA	NA	NA	x	NA	NA	NA		NC	NC	NC		0.049	0.061	0.513	x	NA	NA	NA
Pr, Ec; Fr, Uc; Co		0.047	0.138	0.988		0.046	0.219	0.993		0.053	0.11	0.997		0.041	0.055	0.342		0.052	0.1	0.467
Pr, Ec; Ur, Fc; Co		0.046	0.049	0.562		0.033	0.064	0.312	x	NA	NA	NA		0.048	0.041	0.194	x	NA	NA	NA
Pr, Ec; Ur, Uc, Ft;		0.068	0.137	0.981		0.05	0.144	0.935		0.039	0.149	1		0.047	0.052	0.151		0.045	0.038	0.304
Pr, Ec; Ur, Uc, Ut;		0.066	0.155	0.97		0.05	0.113	0.917		0.045	0.094	1		0.045	0.059	0.396		0.044	0.04	0.175

Table 4 (Continued)

Probabilities; Condition; Statistic	MatrixI				MatrixII				MatrixIII				MatrixIV				MatrixV			
	Critical Region	Bias	Maximum Type I Error Rate: Lim. Set	Maximum Type I Error Rate: Full Set	Critical Region	Bias	Maximum Type I Error Rate: Lim. Set	Maximum Type I Error Rate:Full Set	Critical Region	Bias	Maximum Type I Error Rate: Lim. Set	Maximum Type I Error Rate: Full Set	Critical Region	Bias	Maximum Type I Error Rate: Lim. Set	Maximum Type I Error Rate: Full Set	Critical Region	Bias	Maximum Type I Error Rate: Lim. Set	Maximum Type I Error Rate: Full Set
Pr, Pc; Fr, Fc; Co		0.008	0.009	0.009		0.05	0.122	0.122		NC	NC	NC		0.046	0.081	0.588	x	NA	NA	NA
Pr, Pc; Fr, Uc; Co		0.041	0.066	0.979		0.045	0.088	0.949		0.044	0.02	1		0.038	0.057	0.349		0.038	0.084	0.438
Pr, Pc; Ur, Fc; Co		0.054	0.057	0.574		0.048	0.041	0.324		0.084	0.012	1		0.045	0.056	0.222	x	NA	NA	NA
Pr, Pc; Ur, Uc, Ft;		0.042	0.056	0.937		0.041	0.033	0.821		0.037	0.008	1		0.036	0.046	0.136		0.053	0.063	0.308
Pr, Pc; Ur, Uc, Ut;		0.042	0.041	0.913		0.039	0.034	0.813		0.04	0.02	1		0.054	0.054	0.389		0.047	0.064	0.182
Er, Ec; Fr, Uc; V		0.066	0.051	0.809		0.057	0.052	0.565		NC	NC	NC	x	NA	NA	NA	x	NA	NA	NA
Er, Ec; Ur, Fc; V		0.046	0.047	0.125		0.044	0.05	0.095		0.044	0.044	0.384	x	NA	NA	NA	x	NA	NA	NA
Er, Ec; Ur, Uc, Ft; V		0.059	0.051	0.64		0.057	0.051	0.833		0.04	0.024	0.359		0.041	0.028	0.146		0.064	0.063	0.472
Er, Ec; Ur, Uc, Ut; V		0.062	0.059	0.625		0.053	0.034	0.753		0.039	0.027	0.432		0.053	0.034	0.133		0.065	0.039	0.322
Er, Pc; Fr, Fc; V	x	NA	NA	NA	x	NA	NA	NA	x	NA	NA	NA	x	NA	NA	NA	x	NA	NA	NA
Er, Pc; Fr, Uc; V		0.042	0.964	1		0.031	0.96	1		NC	NC	NC	x	NA	NA	NA		0.04	0.643	0.969
Er, Pc; Ur, Fc; V		0.051	0.045	0.152		0.041	0.054	0.098		0.042	0.041	1	x	NA	NA	NA	x	NA	NA	NA
Er, Pc; Ur, Uc, Ft; V		0.038	0.97	1		0.037	0.971	1		0.025	0.472	1		0.036	0.047	0.147		0.042	0.598	0.967
Er, Pc; Ur, Uc, Ut; V		0.055	0.961	1		0.051	0.946	1		0.037	0.498	1		0.056	0.088	0.295		0.032	0.563	0.955
Pr, Ec; Fr, Fc; V	x	NA	NA	NA	x	NA	NA	NA	x	NA	NA	NA	x	NA	NA	NA	x	NA	NA	NA

Table 4 (Continued)

Probabilities; Condition; Statistic	MatrixI				MatrixII				MatrixIII				MatrixIV				MatrixV			
	Critical Region	Bias	Maximum Type I Error Rate: Lim. Set	Maximum Type I Error Rate: Full Set	Critical Region	Bias	Maximum Type I Error Rate: Lim. Set	Maximum Type I Error Rate:Full Set	Critical Region	Bias	Maximum Type I Error Rate: Lim. Set	Maximum Type I Error Rate: Full Set	Critical Region	Bias	Maximum Type I Error Rate: Lim. Set	Maximum Type I Error Rate: Full Set	Critical Region	Bias	Maximum Type I Error Rate: Lim. Set	Maximum Type I Error Rate: Full Set
Pr, Ec; Fr, Uc; V		0.064	0.042	0.767		0.062	0.057	0.568		0.049	0.001	0.173	x	NA	NA	NA	x	NA	NA	NA
Pr, Ec; Ur, Fc; V		0.048	0.058	0.128		0.025	0.186	0.416	x	NA	NA	NA		0.023	0.173	0.226		0.04	0.911	0.966
Pr, Ec; Ur, Uc, Ft; V		0.067	0.053	0.657		0.048	0.071	0.84		0.059	0.089	1		0.038	0.15	0.357		0.046	0.039	0.417
Pr, Ec; Ur, Uc, Ut; V		0.07	0.051	0.628		0.06	0.063	0.804		0.052	0.106	1		0.056	0.073	0.259		0.045	0.067	0.345
Pr, Pc; Fr, Fc; V	x	NA	NA	NA	x	NA	NA	NA	x	NA	NA	NA	x	NA	NA	NA	x	NA	NA	NA
Pr, Pc; Fr, Uc; V		0.04	0.929	1		0.024	0.871	1		0.044	0.428	1	x	NA	NA	NA		0.048	0.143	0.627
Pr, Pc; Ur, Fc; V		0.037	0.082	0.154		0.044	0.273	0.56		0.131	1	1		0.05	0.212	0.295		0.047	0.946	0.984
Pr, Pc; Ur, Uc, Ft; V		0.042	0.923	1		0.036	0.886	1		0.044	0.626	1		0.044	0.196	0.524		0.039	0.276	0.814
Pr, Pc; Ur, Uc, Ut; V		0.045	0.899	1		0.052	0.887	1		0.055	0.632	1		0.061	0.152	0.591		0.043	0.135	0.531

combination. For Gotelli's (2000) tests, robustness was inconsistent between presence-absence matrices using both the limited and full sets of assumption violations ($\mathrm{K} = 0.234, -0.038$; $p = 0.101, 0.55$; both respectively). For the 76 other tests, robustness was consistent using the limited set of assumption violations, but not the full set ($\mathrm{K} = 0.425, -0.016$; $p = 7.5\times10^{-8}, 0.538$; both respectively). Bias was inconsistent between presence-absence matrices [Gotelli's (2000) tests: $\mathrm{K} = 0.133, p = 0.226$; other tests: $\mathrm{K} = 0.084, p = 0.215$], but the same tests consistently resulted in unusable critical regions [Gotelli's (2000) tests: $\mathrm{K} = 0.474, p = 0.006$; other tests: $\mathrm{K} = 0.474, p = 2.4\times10^{-6}$]. For all presence-absence matrices, robustness was inconsistent between the limited and full sets of assumption violations [for I-V respectively, Gotelli's (2000) tests: $\mathrm{K} = -0.34, -0.481, -0.306, -0.75, -0.613$; p = 0.998, 1, 0.997, 1, 1; other tests: $\mathrm{K} = -0.362, -0.178, -0.192, -0.581, -0.17$; p = 1, 0.985, 0.985, 1, 0.966].

Utility

Under the full set of assumption violations, for every presence-absence matrix where Type I and II error rates were calculated, all tests were either biased, non-robust, or without usable critical regions. The same did not hold under the limited set of assumption violations. Moreover, under the limited set, for every presence-absence matrix at least one test had a usable critical region and indeterminate robustness and bias.

DISCUSSION

Gotelli (2000) found that certain null model tests have low Type I and II error rates when applied to an artificial presence-absence matrix (Figure 3) and a presence-absence matrix summarizing the distribution of West Indian Finches. Based on these findings, Gotelli (2000) has been widely cited to justify the application of certain null model tests, particularly SIM9 with *C* score. However, my results show that citing Gotelli (2000) to support using certain null model tests is not justified. Gotelli (2000) employed incorrect procedures, causing underestimation of Type I error rates. Moreover, even if the rates in Gotelli (2000) were correct, they would not have pertained to other presence-absence matrices: my results show that robustness and bias vary depending on the matrix being analyzed.

Although simply citing other error rates does not justify applying analyses, such citation coupled with evidence limiting the set of potential assumption violations may be valid. Under the full set of assumption violations, none of the 112 tests considered here could be applied to any of the five presence-absence matrices. However, under the limited set of assumption violations, usable tests existed for every presence-absence matrix. Moreover, for the 76 tests not considered by Gotelli (2000), under the limited set, robustness was consistent between presence-absence matrices.

Therefore, to apply existing null model analyses, it appears necessary to restrict potential assumption violations. One tempting approach for making such restrictions is to exploit the marginal totals. For instance, Gilpin and Diamond (1982) and Gotelli and Graves (1996), set the placement probabilities proportional to the marginal totals, as per contingency table analysis. However, doing so implicitly assumes that for each individual, the probability of occurrence at each site is independent of species (Zar 1999), an assumption that *a priori* cannot be favored.

Also complicating matters, restrictions may be needed on the variability of the placement probabilities through time, and the values of those probabilities for unobserved species. The marginal totals appear uninformative in both regards.

What is evidently required for restrictions is supplemental information. The necessity of supplemental information is not a new idea: As has been abundantly argued, applying null model tests requires specific information on the dispersal abilities, habitat affinities, and persistence power of each species (e.g., Connor and Simberloff 1978, Simberloff and Connor 1981, Schoener and Adler 1991, Wilson 1995, Stone et al. 1996, Stone et al. 2000, Peres-Neto et al. 2001), as well as information on the habitats, environment, and isolation of each site (e.g., Stone et al. 1996, Gotelli et al. 1997, Stone et al. 2000, Peres-Neto et al. 2001). Also, although apparently unacknowledged, historical information is necessary to assess variation in the probabilities for each arrival.

Of course in many situations, obtaining such supplemental information may be difficult or impossible. Under such circumstances, the present results suggest that Type I and II error rates should be independently measured for every presence-absence matrix that is analyzed. Such measurements may indicate that all tests are unusable, as was the case here. Moreover, to ensure that Type I error rates always remain at sufficiently low levels, an infinite number of assumption violations will often need to be considered. Hence, rigorously inferring robustness may be difficult. Non-robustness can be inferred from observing high rates however, so measurements will often shed more light on which analyses to avoid, rather than which ones to use. The upshot is that correctly applying existing null model tests may often be impractical.

As an aside, Gotelli (e.g., Gotelli and McCabe 2002) has recently treated null model tests not as tests of the specific null hypothesis, but rather as tests of the entire null models. Under this rubric, if a significant result is found, the entire model is rejected, rather than the null hypothesis *per se*. To choose tests for this application, the Type I error rates of Gotelli (2000) have been cited. However, such citation is inappropriate, for if the entire null model is being tested, then the model comprises the *de facto* null hypothesis, and there are no additional assumptions that can be violated (all of the additional assumptions have been "pushed" into the null hypothesis). However, without additional assumptions, the Type I error rate will always be equal to α. Hence, within Gotelli's rubric, all tests will have the same Type I error rate – α – and it is incorrect to state that some tests are superior because they have low error rates: Robustness is only useful for distinguishing tests when tests require assumptions.

The present results leave at least three issues uninvestigated. First, due to limitations of computer algorithms, SIM9 could only be examined for one presence-absence matrix. Although this instance indicates that SIM9 is no panacea, how the algorithm behaves with other presence-absence matrices remains unclear. Second, I considered a null hypothesis that could imply competitive interactions only. Testing a double-sided null hypothesis – whose violation could also imply aggregative assembly – may result in differing bias, robustness, and usability. Finally, other null model tests have been developed besides the ones presented here (Gotelli and Graves 1996), and although they appear to be non-robust, they may yet yield insight into how to create improved techniques.

ACKNOWLEDGEMENTS

This work was funded by a NSF Graduate Research Fellowship. Computing facilities were kindly provided by Cornell Information Technologies and the Sierra Nevada Aquatic Research Laboratory (University of California, Santa Barbara). I thank S. J. Schwager and H. K. Reeve for helpful advice and comments. R. Knapp kindly provided extensive help and suggestions with writing.

LITERATURE CITED

Beard, J. S. 1948. The natural vegetation of the Windward and Leeward Islands. Oxford Forestry Memoirs **21:**1-192.

Bradley, J. V. 1968. Distribution-Free Statistical Tests. Prentice-Hall, Englewood Cliffs, New Jersey.

Chaves, L. F. and N. Anez. 2004. Species co-occurrence and feeding behavior in sand fly transmission of American cutaneous leishmaniasis in western Venezuela. Acta Tropica **92:**219-224.

Connell, J. H. 1961. The influence of interspecific competition and other factors on the distribution of the barnacle *Chthamalus stellatus*. Ecology **42:**710-723.

Connor, E. F., and D. Simberloff. 1978. Species number and compositional similarity of the Galapagos flora and avifauna. Ecological Monographs **48:**219-248.

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

Connor, E. F., and D. Simberloff. 1983. Interspecific competition and species co-occurrence patterns on islands: null models and the evaluation of evidence. Oikos **41:** 455-465.

Connor, E. F., and D. Simberloff. 1986. Competition, Scientific Method, and Null Models in Ecology. American Scientist **74:**155-162.

Diamond, J. M. 1975. Assembly of species communities. In Ecology and Evolution of Communities, ed. M. L. Cody and J. M. Diamond, 342-344. Harvard University Press, Cambridge.

Feeley, K. 2003. Analysis of avian communities in Lake Guri, Venezuela, using multiple assembly rule models. Oecologia **137:**104-113.

Gilpin, M. E., and J. M. Diamond. 1982. Factors contributing to nonrandomness in species co-occurrence on islands. Oecologia **52:**75-84

Gleason, H. A. 1926. The individualistic concept of plant association. Bulletin of the Torey Botanical Club **53:**7-26.

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

Gotelli, N. J. 2001. Research frontiers in null model analysis. Global Ecology and Biogeography **10:**337-343.

Gotelli, N. J., and L. G. Abele. 1982. Statistical distributions of West Indian land bird families. Journal of Biogeography **9:**421-435.

Gotelli, N. J., Buckley, N. J., and J. A. Wiens. 1997. Co-occurrence of Australian land birds: Diamond's assembly rules revisited. Oikos **80:**311-324.

Gotelli, N. J., and A. M. Ellison. 2002. Assembly rules for New England and assemblages. Oikos **99:**591-599.

Gotelli, N.J. and G.L. Entsminger. 2001. EcoSim: Null models software for ecology. Version 7.0. Acquired Intelligence Inc. & Kesey-Bear. http://homepages.together.net/~gentsmin/ecosim.htm.

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

Gotelli, N. J., and D. J. McCabe. 2002. Species co-occurrence: a meta-analysis of J. M. Diamond's assembly rules model. Ecology **83:**2091-2096.

Gotelli, N. J., and K. Rohde. 2002. Co-occurrence of ectoparasites of marine fishes: a null model analysis. Ecology Letters **5:**86-94.

Hardy, D. E. 1965. Diptera: Cyclorrhapha. In Insects of Hawaii, Volume 12, ed. E. C. Zimmerman, 1-10. University of Hawaii Press, Honolulu.

Heino, J. and J. Soininen. 2005. Assembly rules and community models for unicellular organisms: patterns in diatoms of boreal streams. Freshwater Biology **50:**567-577.

Hay, R. B. 1992. The Oceanic Habitats of Seabirds: Their Zonal Distribution off Vancouver Island, British Columbia, Canada. Journal of Biogeography **19:**67-85.

Knight, K. 1999. Mathematical Statistics. Chapman & Hall/CRC Press, Boca Raton, Florida.

Larsen, R. J., and M. L. Marx. 1986. An Introduction to Mathematical Statistics and its Applications. Prentice-Hall, Englewood Cliffs, New Jersey.

Mouillot, D., M. George-Nascimento, R. Poulin. 2005. Richness, structure and functioning in metazoan parasite communities. Oikos **109:**447-460.

Peres-Neto, P. R., J. D. Olden, and D. A. Jackson. 2001. Environmentally constrained null models: site suitability as occupancy criterion. Oikos **93:**110-120.

Pielou, D. P., and E. C. Pielou. 1968. Association among species of infrequent occurrence: the insect and spider fauna of *polyporus betulinus* (Bulliard) Fries. Journal of Theoretical Biology **21:**202-216.

Ribichich, A. M. 2005. From null community to non-randomly structured actual plant assemblages: parsimony analysis of species co-occurrences. Ecography **28:**88-98.

Robson, D. S. 1972. Appendix: statistical tests of significance. Journal of Theoretical Biology **34:**350-352.

Ross, S. 1998. A First Course in Probability: Fifth Edition. Prentice Hall, Upper Saddle River, New Jersey.

Sanders, N. J., N. J. Gotelli, N. E. Heller, and D. M. Gordon. 2003. Community disassembly by an invasive species. Proceedings of the National Academy of Science USA **100:**2474-2477.

Schluter, D. 1984. A variance test for detecting species associations, with some example applications. Ecology **65:**998-1005.

Schoener, T. W., and G. H. Adler. 1991. Greater resolution of distributional complementarities by controlling for habitat affinities: a study with Bahamian lizards and birds. The American Naturalist **137:**669-692.

Siegel, S. 1988. Nonparametric Statistics for the Behavioral Sciences: Second Edition. McGraw-Hill, New York.

Simberloff, D., and E. F. Connor. 1981. Missing species combinations. American Naturalist **118:**215-239.

Sokal, R. R., and F. J. Rohlf. 1995. Biometry: Third Edition. W. H. Freeman and Company, New York.

Stephenson, S. L., Y. K. Novozhilov, and M. Schnittler. 2000. Distribution and Ecology of Myxomycetes in High-Latitude Regions of the Northern Hemisphere. Journal of Biogeography **27:**741-754.

Stone, L., T. Dayan, and D. Simberloff. 1996. Community-wide assembly patterns unmasked: the importance of species' differing geographical ranges. The American Naturalist **148:**997-1015.

Stone, L., T. Dayan, and D. Simberloff. 2000. On Desert Rodents, Favored States, and Unresolved Issues: Scaling Up and Down Regional Assemblages and Local Communities. The American Naturalist **156:**322-328.

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

Wilson, J. B. 1995. Null models for assembly rules: The Jack Horner effect is more insidious than the Narcissus effect. Oikos **72:**139-144.

Zar, J. H. 1999. Biostatistical Analysis: Fourth Edition. Prentice Hall, Upper Saddle River, New Jersey.

— CHAPTER III —

NULL MODEL TESTS OF SPECIES CO-OCCURRENCE: ROBUST METHODS[1]

Abstract A contentious issue in ecology is to what extent competition affects the composition of ecological communities. To help resolve the matter, statistical tests that do not depend on parametric assumptions are needed. Here, in a step towards assumption-free tests, using random graph theoretic techniques, we derive tests that require one parametric assumption. Letting i and j denote the ith and jth colonists to arrive at a site, respectively, and W_{ij} the event that i and j belong to the same "unit" (e.g., functional group, genus), we show how colonists will be partitioned into units if for all i and j, W_{ij} is conditionally independent of whether i and j share unit membership with the other colonists. Our distribution of partitions is useful for inferring an absence of competitive effects, because they predict that for at least one i and j, $P(W_{ij})$ will be less when i and j share unit membership than when they do not.

INTRODUCTION

Interspecific competition is well documented, for instance between *Chthalamus* sp. and *Balanus* sp. barnacles in the intertidal zone (Connell 1961), *Plethodon* spp. salamanders in the Appalachian Mountains (Hairston 1980), and *Galium* spp. bedstraws in Britain (Tansley 1917, Begon et al 1996). However, it is

---

[1] Second author: Steven J. Schwager, Department of Computational Statistics and Computational Biology, Cornell University.

unclear whether competition affects large-scale spatial patterns of species co-existence. At these scales, experimentation is often impossible, and the only available information may be observational. In an influential work, Diamond (1975) argued that competitive effects can be inferred by determining which pairs of species never co-occur. However, Connor and Simberloff (1979) pointed out that such patterns might arise by chance, and that they may be due to interactions with the physical environment. Assessing the effects of competition has since proven challenging, and has motivated the creation of numerous statistical tests (Gotelli and Graves 1996, Gotelli 2000).

The most widely used tests employ *presence-absence data*, which consist of lists of species recorded at sets of sites. These data offer the advantages of being relatively easy to obtain, and widely available. Presence-absence data are usually summarized in a *presence-absence matrix*, wherein rows represent species, while columns represent sites. If species *i* is observed at site *j*, the *i,j*th entry is 1; otherwise it is 0. Tests using presence-absence data begin by assuming a null hypothesis that is inconsistent with an effect of competition – that species, or individuals of each species, occur independently. The tests also assume probabilities that species or individuals of each species occur at each site. Species are "placed" accordingly and independently until the simulated and observed species richnesses match at each site, or until the numbers of occurrences match for each species. The process is then repeated many times to produce an empirical sampling distribution for the presence-absence matrix. Because the null hypothesis is inconsistent with an effect of competition, if a test statistic from the observed matrix falls in the tail of its sampling distribution, a competitive effect can be inferred (Gotelli and Graves 1996, Gotelli 2000).

Although an effect of competition predicts non-independent occurrences, it makes no prediction about the probabilities of the occurrences *per se*. Hence, the probabilities constitute assumptions – assumptions that in biological terms correspond to the presumed effects of the physical environment on the viability of species or individuals (Ladau, in preparation). All existing unbiased tests require that the assumptions be true, because when they are not, the risk of falsely concluding competition may be intolerably high, up to 100% (Ladau, in preparation). Hence, existing tests are non-robust and provide unreliable conclusions.

As with other statistical tests, two options exist for mitigating the non-robustness (Bradley 1968). First, assumptions can be verified. Here, that means independently assessing the probability of each species occurring at each site, or the probability of each colonizing individual belonging to each species at each site (Ladau, in preparation). The second option is to develop tests that do not depend on parametric assumptions, similar to the alternatives that non-parametric tests provide to parametric tests. Here, the latter option appears most favorable, because assessing occurrence probabilities is often controversial and logistically difficult or impossible (Ladau, in preparation). In this paper we derive tests that rely on one extremely general parametric assumption, which will be useful for constructing tests that are completely free of parametric assumptions.

We begin by deducing a new null hypothesis. In null model testing, the null hypothesis provides the articulation between the biological process of interest and the mathematics of the model, and it consists of a mathematical prediction that will hold if competitive effects are absent. Our null hypothesis will meet this criterion, but, unlike existing null hypotheses, will not require parametric assumptions to be tested.

Competition can structure a community in four ways. At a given site, it can act intrinsically, affecting the biota by occurring within the site, or extrinsically, by occurring in the regions that supply immigrants (Simberloff 1970). It can also act evolutionarily, promoting the creation of ecologically dissimilar species, or ecologically, eliminating ecologically similar species (Connell 1980; Begon et al 1996, pp. 269-270). Hence, competition can act in a total of 2×2=4 ways, all of which can occur simultaneously. Regardless of how competition acts, it will reduce the co-occurrence of ecologically similar species (Connor and Simberloff 1983).

One way to characterize the ecological similarity of species is by grouping them: Pairs of species meeting a threshold for similarity are assigned to the same group, while those failing to meet it are assigned to different groups. Such grouping is equivalent to partitioning species functionally: Placing two species in the same functional group or guild is equivalent to saying that they are similar enough to merit the same classification (Root 2001). But such a notion can be generalized. “Subunits” of organisms – for instance, genera, families, or orders – can be classified either functionally or taxonomically into “units” – for instance, phyla – common membership in which implies ecological similarity.

So long as subunits within units are more ecologically similar than those from different units, competition will favor the co-existence of subunits from different units. Competition acting intrinsically and ecologically will cause subunits to be less likely to persist if they belong to already well-represented units, and all other actions will make arrivals unlikely to belong to the well-represented units (Fox 1987, Fox 1989, Wilson 1989). To formalize this idea, let *i* and *j* denote the *i*th and *j*th persistent arrivals at a site, W_{ij} the event that *i* and *j* belong to the same unit, and W_{ij}^c the event that *i* and *j* belong to different units. Thus, for example, if three subunits are observed, any action of competition predicts that

$$P(W_{13} \mid W_{12}) < P(W_{13} \mid W_{12}^c), \tag{1}$$

$$P(W_{23} \mid W_{12}) < P(W_{23} \mid W_{12}^c), \tag{2}$$

and

$$P(W_{13} \mid W_{23}) < P(W_{13} \mid W_{23}^c)\,. \tag{3}$$

When additional subunits are observed, competitive effects make additional predictions; for instance,

$$P(W_{34} \mid W_{23}) < P(W_{34} \mid W_{23}^c)\,, \tag{4}$$

$$P(W_{34} \mid W_{23} \bigcap W_{12}) < P(W_{34} \mid W_{23}^c \bigcap W_{12})\,, \tag{5}$$

and

$$P(W_{34} \mid W_{23} \bigcap W_{12}^c) < P(W_{34} \mid W_{23}^c \bigcap W_{12}^c)\,. \tag{6}$$

Thus, one possible null hypothesis states that for all *i* and *j*, W_{ij} is conditionally independent of the relationship between *i* and *j* and all other subunits. However, this null hypothesis overlooks an important issue: some relationships contain transitive information about others. For instance, W_{ij} and W_{jk} together imply W_{ik}. Hence, if the null hypothesis stated that

$$P(W_{ij} \mid W_{jk} \bigcap W_{ki}) = P(W_{ij} \mid W_{jk}^c \bigcap W_{ki}) \tag{7}$$

and that

$$P(W_{ij} \mid W_{jk} \bigcap W_{ki}^c) = P(W_{ij} \mid W_{jk}^c \bigcap W_{ki}^c) \quad , \tag{8}$$

it would imply that

$$P(W_{ij}) = 0\,. \tag{9}$$

Clearly, such an implication is inappropriate. It follows then that the null hypothesis should state conditional independence for W_{ij} only when no transitive information is conveyed.

Another potentially complicating issue is the consistency of the condition. Certain sets of events cannot occur; for instance, $W_{ij} \bigcap W_{jk} \bigcap W_{ik}^c$ implies both W_{ik}^c and W_{ik}, hence

$$P(W_{ij} \bigcap W_{jk} \bigcap W_{ik}^c) \equiv 0, \tag{10}$$

so $P(W_{kl} \mid W_{ij} \bigcap W_{jk} \bigcap W_{ik}^c)$ is undefined. It follows then that if the null hypothesis stated

$$P(W_{kl} \mid W_{ij} \bigcap W_{jk} \bigcap W_{ik}^c) = P(W_{kl} \mid W_{ij}^c \bigcap W_{jk} \bigcap W_{ik}^c), \tag{11}$$

it would imply the equality of an undefined and a defined term, a contradiction. Therefore, the null hypothesis should posit equality only when conditions are consistent.

Mindful of these issues, we present a null hypothesis in (12). Under this null hypothesis, assuming that for all *i* and *j*, $P(W_{ij}) > 0$, we derive how subunits will be distributed within units. We then show that under our assumption, this distribution entails the null hypothesis. We conclude by showing how our distribution can be used to test robustly for competition between subunits of the same unit.

DEFINITIONS AND NULL HYPOTHESIS

Let R be a set of observed subunits. Let the sample space be denoted *S*, i.e., $S \equiv \{\psi : \psi$ a partition of $R\}$. For any partition $\psi \in S$, let Eq_ψ denote the corresponding equivalence relation on R, and let $|\psi|$ denote the number of units in ψ. Let *k* be the total number of observed units. For any $r_1, r_2 \in R$, let $W_{r_1 r_2} \equiv \{\psi : \psi$ a partition of *R*, $(r_1, r_2) \in \mathrm{Eq}_\psi\}$. Hence, $W_{r_1 r_2}$ is the event that r_1 and r_2 share the same unit. Denote the complement of $W_{r_1r_2}$ by $W^c_{r_1r_2}$. Define $a \equiv P\{W_{12}\}$. For any $\psi \in S$, let $\{\psi\}$ be the event that ψ occurs. Let $\Omega \equiv \{\omega : \omega \subseteq R, \omega \neq \varnothing\}$. For any $\omega \in \Omega$, let ψ_ω denote a partition of ω, and $\mathrm{Eq}_{\psi_\omega}$ the corresponding equivalence relation. Let $\{\psi_\omega\} \equiv \{\psi : \psi \in S; \forall\, r_1, r_2 \in \omega, (r_1, r_2) \in \mathrm{Eq}_\psi$ if and only if $(r_1, r_2) \in \mathrm{Eq}_{\psi_\omega}\}$. Thus, $\{\psi_\omega\}$ is the event that ψ_ω occurs. Define the set $\mathrm{ran}(\mathbf{N}) \equiv \{(n_1, n_2, ..., n_{|R|}) : n_i$ a non-negative integer, $i = 1,.., |R|$; $\sum_{i=1}^{|R|} in_i = |R|\}$. Define the random vector $\mathbf{N} : S \to \mathrm{ran}(\mathbf{N})$ so that for any $\psi \in S$, $\mathbf{N}(\psi) = (n_1, n_2, ..., n_{|R|})$ if and only if ψ contains exactly n_i units with *i* subunits, $i = 1,2,...,|R|$. Define the random variable $U : S \to \{1,2,...,|R|\}$ so that for any $\psi \in S$, $U(\psi) = i$ if and only if ψ contains exactly *i* units, $i \in \{1,2,...,|R|\}$.

For any graph *G*, let the ordered pair $(V(G), E(G))$ represent the vertex set $V(G)$ and edge set $E(G)$ of *G*. Denote the complement of *G* by G^c. For any graph *G*, let $d(v)$ be the degree of vertex $v \in V(G)$. Let $\tau(G)$ be the number of trees in G. Let Q_G be a minimum vertex cover of *G*. For any Q_G, define $Q_G^+ \equiv \{v : v \in Q_G$ or $d(v) = 0\}$. For any $\omega \in \Omega$, let K_ω be a complete graph with vertex set ω. For any partition ψ_ω of ω with $|\psi_\omega| \geq 2$, let K_{ψ_ω} be the complete multipartite graph with partition ψ_ω. Given any $\psi \in S$ and spanning star forest *F* of K^c_ψ, for convenience let K^*_F be the complete graph with vertex set Q^+_F.

Null Hypothesis. H_0: For any distinct $r_1, r_2, r_3 \in R$ and $X, Y \subseteq E(K_{R-r_1})$, if $P[W_{r_2r_3} \bigcap (\bigcap_{\{i,j\}\in X} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in Y} W^c_{ij})] > 0$ and $P[W^c_{r_2r_3} \bigcap (\bigcap_{\{i,j\}\in X} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in Y} W^c_{ij})] > 0$, then

$$P[W_{r_1r_2} \mid W_{r_2r_3} \bigcap (\bigcap_{\{i,j\}\in X} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in Y} W^c_{ij})]$$

$$= P[W_{r_1r_2} \mid W_{r_2r_3}^c \bigcap (\bigcap_{\{i,j\}\in X} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in Y} W_{ij}^c)]. \qquad (12)$$

This states that for all *i* and *j*, W_{ij} is conditionally independent of the relationship between *i* and *j* and all other subunits, provided that there is consistency and that no transitive information is present.

IMPLICATIONS OF THE NULL HYPOTHESIS

Lemma 1. Under H_0, for any distinct $r_1, r_2, r_3 \in R$ and $X, Y \subseteq E(K_{R-r_1-r_2})$, if $P[(\bigcap_{\{i,j\}\in X} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in Y} W_{ij}^c)] > 0$, then

$$P[W_{r_1r_3} \bigcap W_{r_2r_3} \mid (\bigcap_{\{i,j\}\in X} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in Y} W_{ij}^c)]$$

$$= P[W_{r_1r_3} \mid (\bigcap_{\{i,j\}\in X} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in Y} W_{ij}^c)] \cdot P[W_{r_2r_3} \mid (\bigcap_{\{i,j\}\in X} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in Y} W_{ij}^c)].$$

Proof: Given distinct $r_1, r_2, r_3 \in R$ and $X, Y \subseteq E(K_{R-r_1-r_2})$, assume H_0 and that $P[(\bigcap_{\{i,j\}\in X} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in Y} W_{ij}^c)] > 0$. Because $P[(\bigcap_{\{i,j\}\in X} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in Y} W_{ij}^c)] > 0$, by the definition of conditional probability it will be sufficient to show that

$$P[W_{r_1r_3} \bigcap W_{r_2r_3} \bigcap (\bigcap_{\{i,j\}\in X} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in Y} W_{ij}^c)] \cdot P[(\bigcap_{\{i,j\}\in X} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in Y} W_{ij}^c)]$$

$$= P[W_{r_1r_3} \bigcap (\bigcap_{\{i,j\}\in X} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in Y} W_{ij}^c)] \cdot P[W_{r_2r_3} \bigcap (\bigcap_{\{i,j\}\in X} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in Y} W_{ij}^c)]. \qquad (13)$$

The proof will follow by a case analysis on whether $P[W_{r_2r_3} \bigcap (\bigcap_{\{i,j\}\in X} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in Y} W_{ij}^c)] > 0$ and $P[W_{r_2r_3}^c \bigcap (\bigcap_{\{i,j\}\in X} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in Y} W_{ij}^c)] > 0$.

Case 1: $P[W_{r_2r_3}\bigcap(\bigcap_{\{i,j\}\in X} W_{ij})\bigcap(\bigcap_{\{i,j\}\in Y} W_{ij}^c)] > 0$ and $P[W_{r_2r_3}^c\bigcap(\bigcap_{\{i,j\}\in X} W_{ij})\bigcap(\bigcap_{\{i,j\}\in Y} W_{ij}^c)] > 0$. By H_0,

$$P[W_{r_1r_3} \mid W_{r_2r_3}\bigcap(\bigcap_{\{i,j\}\in X} W_{ij})\bigcap(\bigcap_{\{i,j\}\in Y} W_{ij}^c)]$$

$$= P[W_{r_1r_3} \mid W_{r_2r_3}^c\bigcap(\bigcap_{\{i,j\}\in X} W_{ij})\bigcap(\bigcap_{\{i,j\}\in Y} W_{ij}^c)]. \qquad (14)$$

Applying the definition of conditional probability,

$$P[W_{r_1r_3}\bigcap W_{r_2r_3}\bigcap(\bigcap_{\{i,j\}\in X} W_{ij})\bigcap(\bigcap_{\{i,j\}\in Y} W_{ij}^c)]\cdot P[W_{r_2r_3}^c\bigcap(\bigcap_{\{i,j\}\in X} W_{ij})\bigcap(\bigcap_{\{i,j\}\in Y} W_{ij}^c)]$$

$$= P[W_{r_1r_3}\bigcap W_{r_2r_3}^c\bigcap(\bigcap_{\{i,j\}\in X} W_{ij})\bigcap(\bigcap_{\{i,j\}\in Y} W_{ij}^c)]$$

$$\cdot P[W_{r_2r_3}\bigcap(\bigcap_{\{i,j\}\in X} W_{ij})\bigcap(\bigcap_{\{i,j\}\in Y} W_{ij}^c)]. \qquad (15)$$

Also, the additivity axiom of probability gives

$$P[W_{r_2r_3}^c\bigcap(\bigcap_{\{i,j\}\in X} W_{ij})\bigcap(\bigcap_{\{i,j\}\in Y} W_{ij}^c)]$$

$$= P[(\bigcap_{\{i,j\}\in X} W_{ij})\bigcap(\bigcap_{\{i,j\}\in Y} W_{ij}^c)] - P[W_{r_2r_3}\bigcap(\bigcap_{\{i,j\}\in X} W_{ij})\bigcap(\bigcap_{\{i,j\}\in Y} W_{ij}^c)] \qquad (16)$$

and

$$P[W_{r_1r_3}\bigcap W_{r_2r_3}^c\bigcap(\bigcap_{\{i,j\}\in X} W_{ij})\bigcap(\bigcap_{\{i,j\}\in Y} W_{ij}^c)] = P[W_{r_1r_3}\bigcap(\bigcap_{\{i,j\}\in X} W_{ij})\bigcap(\bigcap_{\{i,j\}\in Y} W_{ij}^c)]$$

$$- P[W_{r_1r_3}\bigcap W_{r_2r_3}\bigcap(\bigcap_{\{i,j\}\in X} W_{ij})\bigcap(\bigcap_{\{i,j\}\in Y} W_{ij}^c)]. \qquad (17)$$

Hence, by substituting the results of (16) and (17) into (15), (13) follows.

Case 2: $P[W_{r_2r_3}\bigcap(\bigcap_{\{i,j\}\in X} W_{ij})\bigcap(\bigcap_{\{i,j\}\in Y} W_{ij}^c)] > 0$ and $P[W_{r_2r_3}^c\bigcap(\bigcap_{\{i,j\}\in X} W_{ij})\bigcap(\bigcap_{\{i,j\}\in Y} W_{ij}^c)] = 0$. Because $P[W_{r_2r_3}^c\bigcap(\bigcap_{\{i,j\}\in X} W_{ij})\bigcap(\bigcap_{\{i,j\}\in Y} W_{ij}^c)] = 0$, by the additivity axiom,

$$P[(\bigcap_{\{i,j\}\in X} W_{ij})\bigcap(\bigcap_{\{i,j\}\in Y} W_{ij}^{c})] = P[W_{r_2r_3}\bigcap(\bigcap_{\{i,j\}\in X} W_{ij})\bigcap(\bigcap_{\{i,j\}\in Y} W_{ij}^{c})]\,; \tag{18}$$

in addition,

$$P[W_{r_1r_3}\bigcap W_{r_2r_3}^{c}\bigcap(\bigcap_{\{i,j\}\in X} W_{ij})\bigcap(\bigcap_{\{i,j\}\in Y} W_{ij}^{c})] = 0\,; \tag{19}$$

and using (19) and the additivity axiom,

$$P[W_{r_1r_3}\bigcap(\bigcap_{\{i,j\}\in X} W_{ij})\bigcap(\bigcap_{\{i,j\}\in Y} W_{ij}^{c})] = P[W_{r_1r_3}\bigcap W_{r_2r_3}\bigcap(\bigcap_{\{i,j\}\in X} W_{ij})\bigcap(\bigcap_{\{i,j\}\in Y} W_{ij}^{c})]\,. \tag{20}$$

Then (13) follows by multiplying (18) and (20).

Case 3: $P[W_{r_2r_3}\bigcap(\bigcap_{\{i,j\}\in X} W_{ij})\bigcap(\bigcap_{\{i,j\}\in Y} W_{ij}^{c})] = 0$ and $P[W_{r_2r_3}^{c}\bigcap(\bigcap_{\{i,j\}\in X} W_{ij})\bigcap(\bigcap_{\{i,j\}\in Y} W_{ij}^{c})] > 0$. $P[W_{r_2r_3}\bigcap(\bigcap_{\{i,j\}\in X} W_{ij})\bigcap(\bigcap_{\{i,j\}\in Y} W_{ij}^{c})] = 0$ implies that $P[W_{r_1r_3}\bigcap W_{r_2r_3}\bigcap(\bigcap_{\{i,j\}\in X} W_{ij})\bigcap(\bigcap_{\{i,j\}\in Y} W_{ij}^{c})] = 0$. Thus,

$$\begin{aligned} &P[W_{r_1r_3}\bigcap W_{r_2r_3}\bigcap(\bigcap_{\{i,j\}\in X} W_{ij})\bigcap(\bigcap_{\{i,j\}\in Y} W_{ij}^{c})]\cdot P[(\bigcap_{\{i,j\}\in X} W_{ij})\bigcap(\bigcap_{\{i,j\}\in Y} W_{ij}^{c})] = 0 \\ &= P[W_{r_1r_3}\bigcap(\bigcap_{\{i,j\}\in X} W_{ij})\bigcap(\bigcap_{\{i,j\}\in Y} W_{ij}^{c})]\cdot P[W_{r_2r_3}\bigcap(\bigcap_{\{i,j\}\in X} W_{ij})\bigcap(\bigcap_{\{i,j\}\in Y} W_{ij}^{c})]\,. \end{aligned} \tag{21}$$

Case 4: $P[W_{r_2r_3}\bigcap(\bigcap_{\{i,j\}\in X} W_{ij})\bigcap(\bigcap_{\{i,j\}\in Y} W_{ij}^{c})] = 0$ and $P[W_{r_2r_3}^{c}\bigcap(\bigcap_{\{i,j\}\in X} W_{ij})\bigcap(\bigcap_{\{i,j\}\in Y} W_{ij}^{c})] = 0$. (13) follows by the arguments given in Case 3.

■

Lemma 2. Under H_0, for any distinct $r_1, r_2, r_3 \in R$ and $X, Y \subseteq E(K_{R-r_1-r_2})$, if $P[W_{r_2r_3}\bigcap(\bigcap_{\{i,j\}\in X} W_{ij})\bigcap(\bigcap_{\{i,j\}\in Y} W_{ij}^{c})] > 0$, then

$$P[W_{r_1r_3}\,|\,W_{r_2r_3}\bigcap(\bigcap_{\{i,j\}\in X} W_{ij})\bigcap(\bigcap_{\{i,j\}\in Y} W_{ij}^{c})] = P[W_{r_1r_3}\,|\,(\bigcap_{\{i,j\}\in X} W_{ij})\bigcap(\bigcap_{\{i,j\}\in Y} W_{ij}^{c})]\,.$$

Proof: Given distinct $r_1, r_2, r_3 \in R$ and $X, Y \subseteq E(K_{R-r_1-r_2})$, assume H_0 and that $P[W_{r_2r_3} \bigcap (\bigcap_{\{i,j\}\in X} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in Y} W_{ij}^c)] > 0$. The latter assumption and the definition of conditional probability imply that

$$P[W_{r_1r_3} \mid W_{r_2r_3} \bigcap (\bigcap_{\{i,j\}\in X} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in Y} W_{ij}^c)] = \frac{P[W_{r_1r_3} \bigcap W_{r_2r_3} \bigcap (\bigcap_{\{i,j\}\in X} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in Y} W_{ij}^c)]}{P[W_{r_2r_3} \bigcap (\bigcap_{\{i,j\}\in X} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in Y} W_{ij}^c)]}$$

$$= \frac{P[W_{r_1r_3} \bigcap W_{r_2r_3} \mid (\bigcap_{\{i,j\}\in X} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in Y} W_{ij}^c)]}{P[W_{r_2r_3} \mid (\bigcap_{\{i,j\}\in X} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in Y} W_{ij}^c)]}. \tag{22}$$

It follows from $P[W_{r_2r_3} \bigcap (\bigcap_{\{i,j\}\in X} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in Y} W_{ij}^c)] > 0$ that

$P[(\bigcap_{\{i,j\}\in X} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in Y} W_{ij}^c)] > 0$, and so by Lemma 1,

$$\frac{P[W_{r_1r_3} \bigcap W_{r_2r_3} \mid (\bigcap_{\{i,j\}\in X} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in Y} W_{ij}^c)]}{P[W_{r_2r_3} \mid (\bigcap_{\{i,j\}\in X} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in Y} W_{ij}^c)]}$$

$$= \frac{P[W_{r_1r_3} \mid (\bigcap_{\{i,j\}\in X} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in Y} W_{ij}^c)] \cdot P[W_{r_2r_3} \mid (\bigcap_{\{i,j\}\in X} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in Y} W_{ij}^c)]}{P[W_{r_2r_3} \mid (\bigcap_{\{i,j\}\in X} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in Y} W_{ij}^c)]}$$

$$= P[W_{r_1r_3} \mid (\bigcap_{\{i,j\}\in X} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in Y} W_{ij}^c)]. \qquad \blacksquare$$

Lemma 3. For any $X, Y \in E(K_R)$, and $r_1, r_2 \in R, r_1 \neq r_2$, if $\exists\ r_3 \in R$ such that $\{r_1, r_3\}, \{r_2, r_3\} \in X$, then

$$W_{r_1r_2} \bigcap (\bigcap_{\{i,j\}\in X} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in Y} W_{ij}^c) = (\bigcap_{\{i,j\}\in X} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in Y} W_{ij}^c).$$

Proof: Given $X, Y \in E(K_R)$, and $r_1, r_2 \in R, r_1 \neq r_2$, assume $\exists \; r_3 \in R$ such that $\{r_1, r_3\}, \{r_2, r_3\} \in X$. Fix r_3. Given $\psi \in S$, assume that $\psi \in W_{r_1 r_2} \bigcap (\bigcap_{\{i,j\} \in X} W_{ij}) \bigcap (\bigcap_{\{i,j\} \in Y} W_{ij}^c)$. Because by definition

$$W_{r_1 r_2} \bigcap (\bigcap_{\{i,j\} \in X} W_{ij}) \bigcap (\bigcap_{\{i,j\} \in Y} W_{ij}^c) \subseteq (\bigcap_{\{i,j\} \in X} W_{ij}) \bigcap (\bigcap_{\{i,j\} \in Y} W_{ij}^c), \tag{23}$$

$\psi \in (\bigcap_{\{i,j\} \in X} W_{ij}) \bigcap (\bigcap_{\{i,j\} \in Y} W_{ij}^c)$. Now assume that $\psi \in (\bigcap_{\{i,j\} \in X} W_{ij}) \bigcap (\bigcap_{\{i,j\} \in Y} W_{ij}^c)$. Thus, $(r_1, r_3), (r_2, r_3) \in \mathrm{Eq}_\psi$, and transitively, $(r_1, r_2) \in \mathrm{Eq}_\psi$. It follows that $\psi \in W_{r_1 r_2}$, and that $\psi \in W_{r_1 r_2} \bigcap (\bigcap_{\{i,j\} \in X} W_{ij}) \bigcap (\bigcap_{\{i,j\} \in Y} W_{ij}^c)$. Hence, $\psi \in W_{r_1 r_2} \bigcap (\bigcap_{\{i,j\} \in X} W_{ij}) \bigcap (\bigcap_{\{i,j\} \in Y} W_{ij}^c)$ if and only if $\psi \in (\bigcap_{\{i,j\} \in X} W_{ij}) \bigcap (\bigcap_{\{i,j\} \in Y} W_{ij}^c)$. ■

Lemma 4. Under H_0, $P(W_{r_1 r_2}) = a$ for all distinct $r_1, r_2 \in R$.

Proof: Assume H_0. Given distinct $r_1, r_2 \in R$, fix $r_3 \in R, r_3 \neq r_1, r_2$. By Lemma 3 (with $X = \{\{r_1, r_3\}, \{r_2, r_3\}\}, Y = \varnothing$)

$$W_{r_1 r_2} \bigcap W_{r_1 r_3} \bigcap W_{r_2 r_3} = W_{r_1 r_3} \bigcap W_{r_2 r_3}, \tag{24}$$

implying by Lemma 1 (with $X, Y = \varnothing$) that

$$P(W_{r_1 r_2} \bigcap W_{r_1 r_3} \bigcap W_{r_2 r_3}) = P(W_{r_1 r_3}) P(W_{r_2 r_3}). \tag{25}$$

By similar arguments,

$$P(W_{r_1 r_2} \bigcap W_{r_1 r_3} \bigcap W_{r_2 r_3}) = P(W_{r_1 r_2}) P(W_{r_1 r_3}) \tag{26}$$

and

$$P(W_{r_1 r_2} \bigcap W_{r_1 r_3} \bigcap W_{r_2 r_3}) = P(W_{r_1 r_2}) P(W_{r_2 r_3}). \tag{27}$$

(25), (26), and (27) comprise a system of three independent equations with four unknowns. By assumption, $P(W_{r_1 r_2}), P(W_{r_1 r_3}), P(W_{r_2 r_3}) > 0$, implying that $P(W_{r_1 r_2}) = P(W_{r_1 r_3}) = P(W_{r_2 r_3})$. Therefore, in the case $r_3 = 2$,

$$P(W_{r_1 r_2}) = P(W_{r_1 2}) = P(W_{r_2 2}), \tag{28}$$

while in the case $r_1 = 1, r_3 = 2$,

$$P(W_{1 r_2}) = P(W_{12}) = P(W_{r_2 2}). \tag{29}$$

Hence $P(W_{r_1 r_2}) = P(W_{12}) = a$. ■

Lemma 5. For any $X, Y \in E(K_R)$, and distinct $r_1, r_2 \in R$, if $\exists \; r_3 \in R$ such that $\{r_1, r_3\} \in X$ and $\{r_2, r_3\} \in Y$, then

$$W^c_{r_1 r_2} \bigcap \Big(\bigcap_{\{i,j\} \in X} W_{ij} \Big) \bigcap \Big(\bigcap_{\{i,j\} \in Y} W^c_{ij} \Big) = \Big(\bigcap_{\{i,j\} \in X} W_{ij} \Big) \bigcap \Big(\bigcap_{\{i,j\} \in Y} W^c_{ij} \Big).$$

Proof: Given $X, Y \in E(K_R)$, and distinct $r_1, r_2 \in R$, assume $\exists \; r_3 \in R$ such that $\{r_1, r_3\} \in X$ and $\{r_2, r_3\} \in Y$. Fix r_3. Given $\psi \in S$, because by definition

$$W^c_{r_1 r_2} \bigcap \Big(\bigcap_{\{i,j\} \in X} W_{ij} \Big) \bigcap \Big(\bigcap_{\{i,j\} \in Y} W^c_{ij} \Big) \subseteq \Big(\bigcap_{\{i,j\} \in X} W_{ij} \Big) \bigcap \Big(\bigcap_{\{i,j\} \in Y} W^c_{ij} \Big), \tag{30}$$

if $\psi \in W^c_{r_1 r_2} \bigcap (\bigcap_{\{i,j\} \in X} W_{ij}) \bigcap (\bigcap_{\{i,j\} \in Y} W^c_{ij})$, then $\psi \in (\bigcap_{\{i,j\} \in X} W_{ij}) \bigcap (\bigcap_{\{i,j\} \in Y} W^c_{ij})$. Now assume that $\psi \in (\bigcap_{\{i,j\} \in X} W_{ij}) \bigcap (\bigcap_{\{i,j\} \in Y} W^c_{ij})$, implying that $(r_1, r_3) \in \mathrm{Eq}_\psi$ and $(r_2, r_3) \notin \mathrm{Eq}_\psi$. It follows that $(r_1, r_2) \notin \mathrm{Eq}_\psi$. Thus, $\psi \in W^c_{r_1 r_2} \bigcap (\bigcap_{\{i,j\} \in X} W_{ij}) \bigcap (\bigcap_{\{i,j\} \in Y} W^c_{ij})$. Therefore, for any $\psi \in S$, $\psi \in W^c_{r_1 r_2} \bigcap (\bigcap_{\{i,j\} \in X} W_{ij}) \bigcap (\bigcap_{\{i,j\} \in Y} W^c_{ij})$ if and only if

$\psi \in (\bigcap_{\{i,j\} \in X} W_{ij}) \bigcap (\bigcap_{\{i,j\} \in Y} W^c_{ij})$. ■

Lemma 6. For any $\omega \in \Omega$ with $|\omega| \geq 2$, integer *t* with $2 \leq t \leq |\omega|$, and enumeration $r_1, r_2, r_3, \ldots, r_{|\omega|}$ of ω,

$$W_{r_1 r_2} \bigcap (\bigcap_{\{i,j\} \in E(K_{\omega - r_1})} W_{ij}^c) = W_{r_1 r_2} \bigcap (\bigcap_{\{i,j\} \in E(K_{\omega - r_1}) \cup E(K_{\{\{r_1\},\{r_2,\ldots,r_t\}\}}) - \{r_1, r_2\}} W_{ij}^c) .$$

Proof: Given $\omega \in \Omega$ with $|\omega| \geq 2$, integer *t* with $2 \leq t \leq |\omega|$, enumeration $r_1, r_2, r_3, \ldots, r_{|\omega|}$ of ω, and $\psi \in R$, assume that $\psi \in W_{r_1 r_2} \bigcap (\bigcap_{\{i,j\} \in E(K_{\omega - r_1}) \cup E(K_{\{\{r_1\},\{r_2,\ldots,r_{t-1}\}\}}) - \{r_1, r_2\}} W_{ij}^c)$. Then

$$W_{r_1 r_2} \bigcap (\bigcap_{\{i,j\} \in E(K_{\omega - r_1}) \cup E(K_{\{\{r_1\},\{r_2,\ldots,r_t\}\}}) - \{r_1, r_2\}} W_{ij}^c) \subseteq W_{r_1 r_2} \bigcap (\bigcap_{\{i,j\} \in E(K_{\omega - r_1})} W_{ij}^c) , \tag{31}$$

so $\psi \in W_{r_1 r_2} \bigcap (\bigcap_{\{i,j\} \in E(K_{\omega - r_1})} W_{ij}^c)$. For the opposite direction, assume that $\psi \in W_{r_1 r_2} \bigcap (\bigcap_{\{i,j\} \in E(K_{\omega - r_1})} W_{ij}^c)$. By Lemma 5, for any $\{r_1, r_i\} \in E(K_{\{\{r_1\},\{r_2,\ldots,r_t\}\}}) - \{r_1, r_2\}$,

$$W_{r_1 r_2} \bigcap (\bigcap_{\{i,j\} \in E(K_{\omega - r_1})} W_{ij}^c) = W_{r_1 r_i} \bigcap W_{r_1 r_2} \bigcap (\bigcap_{\{i,j\} \in E(K_{\omega - r_1})} W_{ij}^c) . \tag{32}$$

Thus, $\psi \in W_{r_1 r_i} \bigcap W_{r_1 r_2} \bigcap (\bigcap_{\{i,j\} \in E(K_{\omega - r_1})} W_{ij}^c)$. It follows that

$$\begin{aligned} \psi \in (\bigcap_{\{i,j\} \in E(K_{\{\{r_1\},\{r_2,\ldots,r_t\}\}}) - \{r_1, r_2\}} W_{ij}^c) \bigcap W_{r_1 r_2} \bigcap (\bigcap_{\{i,j\} \in E(K_{\omega - r_1})} W_{ij}^c) \\ = W_{r_1 r_2} \bigcap (\bigcap_{\{i,j\} \in E(K_{\omega - r_1}) \cup E(K_{\{\{r_1\},\{r_2,\ldots,r_t\}\}}) - \{r_1, r_2\}} W_{ij}^c) . \end{aligned} \tag{33}$$

Thus, $\psi \in W_{r_1 r_2} \bigcap (\bigcap_{\{i,j\} \in E(K_{\omega - r_1})} W_{ij}^c)$ if and only if

$\psi \in W_{r_1 r_2} \bigcap (\bigcap_{\{i,j\} \in E(K_{\omega - r_1}) \cup E(K_{\{\{r_1\},\{r_2,\ldots,r_{t-1}\}\}}) - \{r_1, r_2\}} W_{ij}^c)$. ■

Lemma 7. Under H_0, for all $\omega \in \Omega$ with $|\omega| \neq R$, $r_1 \notin \omega$, and $r_2 \in \omega$, if $P(\bigcap_{\{i,j\} \in E(K_\omega)} W_{ij}^c) > 0$, then

$$P(W_{r_1 r_2} \mid \bigcap_{\{i,j\} \in E(K_\omega)} W_{ij}^c) = a .$$

Proof: Assume H_0. Given $\omega \in \Omega$ with $|\omega| \neq R$, $r_1 \notin \omega$, and $r_2 \in \omega$, assume that $P(\bigcap_{\{i,j\} \in E(K_\omega)} W_{ij}^c) > 0$. The proof will follow by induction on $|\omega|$. If $|\omega| = 1$, then $E(K_\omega) = \varnothing$, and by Lemma 4,

$$P(W_{r_1 r_2} \mid \bigcap_{\{i,j\} \in E(K_\omega)} W_{ij}^c) = P(W_{r_1 r_2}) = a . \tag{34}$$

Now assume that for some integer $1 \leq t$, the result holds for $|\omega| \leq t$. Take $|\omega| = t+1$. Fix $r_3 \in \omega, r_3 \neq r_2$. By assumption $P(\bigcap_{\{i,j\} \in E(K_\omega)} W_{ij}^c) > 0$, so $P(\bigcap_{E(K_{\omega - r_3})} W_{ij}^c) > 0$. Thus, by the induction hypothesis,

$$P(W_{r_2 r_3} \mid \bigcap_{E(K_{\omega - r_3})} W_{ij}^c) = a \tag{35}$$

and

$$P(W_{r_1 r_2} \mid \bigcap_{E(K_{\omega - r_3})} W_{ij}^c) = a . \tag{36}$$

From (35) and the multiplication rule for conditional probability,

$$P[W_{r_2 r_3} \bigcap (\bigcap_{E(K_{\omega - r_3})} W_{ij}^c)] = P(W_{r_2 r_3} \mid \bigcap_{E(K_{\omega - r_3})} W_{ij}^c) \cdot P(\bigcap_{E(K_{\omega - r_3})} W_{ij}^c) = a \cdot P(\bigcap_{E(K_{\omega - r_3})} W_{ij}^c) > 0 . \tag{37}$$

Thus, it follows from Lemma 2 that

$$P[W_{r_1 r_2} \mid W_{r_2 r_3} \bigcap (\bigcap_{E(K_{\omega - r_3})} W_{ij}^c)] = a . \tag{38}$$

By Lemma 6,

$$P[W_{r_1 r_2} \mid W_{r_2 r_3} \bigcap (\bigcap_{E(K_{\omega - r_3})} W_{ij}^c)] = a = P[W_{r_1 r_2} \mid W_{r_2 r_3} \bigcap (\bigcap_{E(K_\omega) - \{r_2, r_3\}} W_{ij}^c)] . \tag{39}$$

Lemma 6 and (37) also imply that

$$P[W_{r_2r_3}\bigcap(\bigcap_{E(K_\omega)-\{r_2,r_3\}} W_{ij}^c)]>0. \tag{40}$$

Moreover, by definition

$$P[W_{r_1r_2}\mid W_{r_2r_3}^c\bigcap(\bigcap_{E(K_\omega)-\{r_2,r_3\}} W_{ij}^c)]=P(W_{r_1r_2}\mid \bigcap_{\{i,j\}\in E(K_\omega)} W_{ij}^c) \tag{41}$$

and by assumption $P(\bigcap_{\{i,j\}\in E(K_\omega)} W_{ij}^c)>0$, so H_0 gives

$$P[W_{r_1r_2}\mid W_{r_2r_3}\bigcap(\bigcap_{E(K_\omega)-\{r_2,r_3\}} W_{ij}^c)]=P[W_{r_1r_2}\mid W_{r_2r_3}^c\bigcap(\bigcap_{E(K_\omega)-\{r_2,r_3\}} W_{ij}^c)]. \tag{42}$$

Therefore, by (39)

$$P[W_{r_1r_2}\mid W_{r_2r_3}^c\bigcap(\bigcap_{E(K_\omega)-\{r_2,r_3\}} W_{ij}^c)=a. \qquad \blacksquare$$

Lemma 8. Under H_0, for any $\omega\in\Omega$ with $|\omega|\geq 2$, enumeration $r_1,r_2,r_3,\ldots,r_{|\omega|}$ of ω, and integer *t* with $2\leq t\leq|\omega|$, if $P(\bigcap_{\{i,j\}\in E(K_{\omega-r_1})} W_{ij}^c)>0$, then

$$P(W_{r_1r_t}\mid \bigcap_{\{i,j\}\in E(K_{\omega-r_1})\cup E(K_{\{\{r_1\},\{r_2,\ldots,r_t\}\}})-\{r_1,r_t\}} W_{ij}^c)=\frac{a}{1-a(t-2)}.$$

Proof: Given $\omega\in\Omega$ with $|\omega|\geq 2$, fix an enumeration of ω, $r_1,r_2,r_3,\ldots,r_{|\omega|}$. Assume H_0 and that $P(\bigcap_{\{i,j\}\in E(K_{\omega-r_1})} W_{ij}^c)>0$. The proof will follow by induction on *t*. For the case $t=2$, because $E(K_{\{\{r_1\},\{r_2,\ldots,r_t\}\}})-\{r_1,r_t\}=\varnothing$, Lemma 7 can be applied to obtain

$$P(W_{r_1r_t}\mid \bigcap_{\{i,j\}\in E(K_{\omega-r_1})\cup E(K_{\{\{r_1\},\{r_2,\ldots,r_t\}\}})-\{r_1,r_t\}} W_{ij}^c)=P(W_{r_1r_2}\mid \bigcap_{\{i,j\}\in E(K_{\omega-r_1})} W_{ij}^c)=a. \tag{43}$$

For $t>2$, assume the result for $t-1$. The multiplication rule for conditional probability and $P(\bigcap_{\{i,j\}\in E(K_{\omega-r_1})} W_{ij}^c)>0$ imply that

$$P[W_{r_1r_t}\bigcap(\bigcap_{\{i,j\}\in E(K_{\omega-r_1})} W_{ij}^c)]=P(W_{r_1r_t}\mid \bigcap_{\{i,j\}\in E(K_{\omega-r_1})} W_{ij}^c)\cdot P(\bigcap_{\{i,j\}\in E(K_{\omega-r_1})} W_{ij}^c). \tag{44}$$

By Lemma 7,

$$P(W_{r_1r_t} \mid \bigcap_{\{i,j\}\in E(K_{\omega-r_1})} W_{ij}^c)\cdot P(\bigcap_{\{i,j\}\in E(K_{\omega-r_1})} W_{ij}^c) = a\cdot P(\bigcap_{\{i,j\}\in E(K_{\omega-r_1})} W_{ij}^c) > 0\,. \tag{45}$$

Thus,

$$P[W_{r_1r_t}\bigcap(\bigcap_{\{i,j\}\in E(K_{\omega-r_1})} W_{ij}^c)] > 0\,. \tag{46}$$

Moreover, by Lemma 6,

$$P[W_{r_1r_t}\bigcap(\bigcap_{\{i,j\}\in E(K_{\omega-r_1})} W_{ij}^c)] = P[W_{r_1r_t}\bigcap(\bigcap_{\{i,j\}\in E(K_{\omega-r_1})\cup E(K_{\{\{r_1\},\{r_2,\ldots,r_{t-1}\}\}})} W_{ij}^c)]\,, \tag{47}$$

so

$$P[W_{r_1r_t}\bigcap(\bigcap_{\{i,j\}\in E(K_{\omega-r_1})\cup E(K_{\{\{r_1\},\{r_2,\ldots,r_{t-1}\}\}})} W_{ij}^c)] > 0\,. \tag{48}$$

Thus, by the multiplication rule for conditional probability:

$$\begin{aligned}
&P[W_{r_1r_t}\bigcap(\bigcap_{\{i,j\}\in E(K_{\omega-r_1})\cup E(K_{\{\{r_1\},\{r_2,\ldots,r_{t-1}\}\}})} W_{ij}^c)]\\
&= P(\bigcap_{\{i,j\}\in E(K_{\omega-r_1})\cup E(K_{\{\{r_1\},\{r_2,\ldots,r_{t-1}\}\}})-\{r_1,r_{t-1}\}} W_{ij}^c)\cdot P(W_{r_1r_{t-1}}^c \mid \bigcap_{\{i,j\}\in E(K_{\omega-r_1})\cup E(K_{\{\{r_1\},\{r_2,\ldots,r_{t-1}\}\}})-\{r_1,r_{t-1}\}} W_{ij}^c)\\
&\qquad \cdot P(W_{r_1r_t} \mid \bigcap_{\{i,j\}\in E(K_{\omega-r_1})\cup E(K_{\{\{r_1\},\{r_2,\ldots,r_{t-1}\}\}})} W_{ij}^c)
\end{aligned} \tag{49}$$

and

$$\begin{aligned}
&P[W_{r_1r_t}\bigcap(\bigcap_{\{i,j\}\in E(K_{\omega-r_1})\cup E(K_{\{\{r_1\},\{r_2,\ldots,r_{t-1}\}\}})} W_{ij}^c)]\\
&= P(\bigcap_{\{i,j\}\in E(K_{\omega-r_1})\cup E(K_{\{\{r_1\},\{r_2,\ldots,r_{t-1}\}\}})-\{r_1,r_{t-1}\}} W_{ij}^c)\cdot P(W_{r_1r_t} \mid \bigcap_{\{i,j\}\in E(K_{\omega-r_1})\cup E(K_{\{\{r_1\},\{r_2,\ldots,r_{t-1}\}\}})-\{r_1,r_{t-1}\}} W_{ij}^c)\\
&\qquad \cdot P[W_{r_1r_{t-1}}^c \mid W_{r_1r_t}\bigcap(\bigcap_{\{i,j\}\in E(K_{\omega-r_1})\cup E(K_{\{\{r_1\},\{r_2,\ldots,r_{t-1}\}\}})-\{r_1,r_{t-1}\}} W_{ij}^c)]\,.
\end{aligned} \tag{50}$$

It follows from (48) that all terms in both (49) and (50) are defined and nonzero. Examining the second term on the right side in (49), by the induction hypothesis,

$$P(W_{r_1r_{t-1}}^c \mid \bigcap_{\{i,j\}\in E(K_{\omega-r_1})\cup E(K_{\{\{r_1\},\{r_2,\ldots,r_{t-1}\}\}})-\{r_1,r_{t-1}\}} W_{ij}^c) = 1-\frac{a}{1-a(t-3)} = \frac{1-a(t-2)}{1-a(t-3)}\,. \tag{51}$$

Examining the second term on the right side of (50), by the induction hypothesis,

$$P(W_{r_1 r_t} \mid \bigcap_{\{i,j\}\in E(K_{\omega-r_1})\cup E(K_{\{\{r_1\},\{r_2,\ldots,r_{t-1}\}\}})-\{r_1,r_{t-1}\}} W_{ij}^c) = \frac{a}{1-a(t-3)}. \tag{52}$$

For the final term in (50), from Lemma 5,

$$W_{r_1 r_{t-1}}^c \bigcap W_{r_1 r_t} \bigcap (\bigcap_{\{i,j\}\in E(K_{\omega-r_1})\cup E(K_{\{\{r_1\},\{r_2,\ldots,r_{t-1}\}\}})-\{r_1,r_{t-1}\}} W_{ij}^c)$$

$$= W_{r_1 r_t} \bigcap (\bigcap_{\{i,j\}\in E(K_{\omega-r_1})\cup E(K_{\{\{r_1\},\{r_2,\ldots,r_{t-1}\}\}})-\{r_1,r_{t-1}\}} W_{ij}^c), \tag{53}$$

which implies that

$$P[W_{r_1 r_{t-1}}^c \mid W_{r_1 r_t} \bigcap (\bigcap_{\{i,j\}\in E(K_{\omega-r_1})\cup E(K_{\{\{r_1\},\{r_2,\ldots,r_{t-1}\}\}})-\{r_1,r_{t-1}\}} W_{ij}^c)] = 1. \tag{54}$$

Hence, equating (49) and (50) gives

$$P(\bigcap_{\{i,j\}\in E(K_{\omega-r_1})\cup E(K_{\{\{r_1\},\{r_2,\ldots,r_{t-1}\}\}})-\{r_1,r_{t-1}\}} W_{ij}^c) \cdot \frac{1-a(t-2)}{1-a(t-3)}$$

$$\cdot P(W_{r_1 r_t} \mid \bigcap_{\{i,j\}\in E(K_{\omega-r_1})\cup E(K_{\{\{r_1\},\{r_2,\ldots,r_{t-1}\}\}})} W_{ij}^c)$$

$$= P(\bigcap_{\{i,j\}\in E(K_{\omega-r_1})\cup E(K_{\{\{r_1\},\{r_2,\ldots,r_{t-1}\}\}})-\{r_1,r_{t-1}\}} W_{ij}^c) \cdot \frac{a}{1-a(t-3)} \cdot 1, \tag{55}$$

and thus,

$$P(W_{r_1 r_t} \mid \bigcap_{\{i,j\}\in E(K_{\omega-r_1})\cup E(K_{\{\{r_1\},\{r_2,\ldots,r_t\}\}})-\{r_1,r_t\}} W_{ij}^c).$$

$$= P(W_{r_1 r_t} \mid \bigcap_{\{i,j\}\in E(K_{\omega-r_1})\cup E(K_{\{\{r_1\},\{r_2,\ldots,r_{t-1}\}\}})} W_{ij}^c) = \frac{a}{1-a(t-2)} \qquad \blacksquare$$

Lemma 9. Under H_0, for any $\omega \in \Omega$ with $|\omega| \geq 2$ and $r_1 \in \omega$, if

$P(\bigcap_{\{i,j\}\in E(K_{\omega-r_1})} W_{ij}^c) > 0$, then

$$P\Big(\bigcap_{\{i,j\}\in E(K_{\omega})} W_{ij}^{c}\Big)=\prod_{i=0}^{|\omega|-1}(1-ia).$$

Proof: Given $\omega \in \Omega$ with $|\omega| \geq 2$ and $r_1 \in \omega$, assume H_0 and that $P(\bigcap_{\{i,j\}\in E(K_{\omega-r_1})} W_{ij}^{c}) > 0$. The proof will follow by induction on $|\omega|$. For the case $|\omega| = 2$, fix $r_2 \in \omega, r_2 \neq r_1$. By Lemma 4,

$$P\Big(\bigcap_{\{i,j\}\in E(K_{\omega})} W_{ij}^{c}\Big)=P(W_{r_1r_2}^{c})=1-a. \tag{56}$$

Assume now that for some integer $2 \leq t < |R|$, the result holds if $2 \leq |\omega| \leq t$. Assume that $|\omega| = t+1$ and fix an enumeration of $\omega \equiv r_1, r_2, \ldots, r_{t+1}$. By assumption, $P(\bigcap_{\{i,j\}\in E(K_{\omega-r_1})} W_{ij}^{c}) > 0$, so by the multiplication rule for conditional probability

$$P\Big(\bigcap_{\{i,j\}\in E(K_{\omega})} W_{ij}^{c}\Big)=P\Big(\bigcap_{\{i,j\}\in E(K_{\omega-r_1})} W_{ij}^{c}\Big)\cdot P\Big(\bigcap_{\{i,j\}\in E(K_{\{\{r_1\},\omega-r_1\}})} W_{ij}^{c} \mid \bigcap_{\{i,j\}\in E(K_{\omega-r_1})} W_{ij}^{c}\Big). \tag{57}$$

Examining the first term on the right side in (57), because $|\omega - r_1| = t$ and $P(\bigcap_{\{i,j\}\in E(K_{\omega-r_1})} W_{ij}^{c}) > 0$, the induction hypothesis can be applied. Thus,

$$P\Big(\bigcap_{\{i,j\}\in E(K_{\omega-r_1})} W_{ij}^{c}\Big)=\prod_{i=0}^{t-1}(1-ia). \tag{58}$$

As for the second term on the right side in (57), by the multiplication rule for conditional probability and Lemma 8,

$$\begin{aligned} &P\Big(\bigcap_{\{i,j\}\in E(K_{\{\{r_1\},\omega-r_1\}})} W_{ij}^{c} \mid \bigcap_{\{i,j\}\in E(K_{\omega-r_1})} W_{ij}^{c}\Big) \\ &\qquad=\prod_{l=2}^{t+1} P\Big(W_{r_1r_l}^{c} \mid \bigcap_{\{i,j\}\in E(K_{\omega-r_1})\cup E(K_{\{\{r_1\},\{r_2,\ldots,r_l\}\}})-\{r_1,r_l\}} W_{ij}^{c}\Big) \\ &\qquad\qquad=\prod_{l=2}^{t+1}\Big[1-\frac{a}{1-a(l-2)}\Big]=1-ta. \end{aligned} \tag{59}$$

Hence,

$$P(\bigcap_{\{i,j\}\in E(K_\omega)} W_{ij}^c) = [\prod_{i=0}^{t-1}(1-ia)]\cdot(1-ta) = \prod_{i=0}^{t}(1-ia) \ . \tag{60}$$

As $|\omega|-1=t$, the proof is complete. ■

Lemma 10. Under H_0, for any $\omega\in\Omega$ with $|\omega|\geq 2$ and $r_1\in\omega$, if $P(\bigcap_{\{i,j\}\in E(K_{\omega-r_1})} W_{ij}^c)>0$, then $a\leq\frac{1}{|\omega|-1}$.

Proof: Given $\omega\in\Omega$ with $|\omega|\geq 2$ and $r_1\in\omega$, assume H_0 and that $P(\bigcap_{\{i,j\}\in E(K_{\omega-r_1})} W_{ij}^c)>0$. Fix an enumeration $r_1,r_2,r_3,\ldots,r_{|\omega|}$ of ω. Lemma 8 applies with $t=|\omega|$ to give

$$P(W_{r_1r_{|\omega|}} \mid \bigcap_{\{i,j\}\in E(K_{\omega-r_1})\cup E(K_{\{\{r_1\},\{r_2,\ldots,r_{|\omega|}\}\}})-\{r_1,r_{|\omega|}\}} W_{ij}^c) = \frac{a}{1-a(|\omega|-2)}. \tag{61}$$

The expression on the right side of (61) is greater than 1 when $\frac{1}{|\omega|-1}<a<\frac{1}{|\omega|-2}$, undefined when $a=\frac{1}{|\omega|-2}$, and less than 0 when $a>\frac{1}{|\omega|-2}$. Because this expression is a probability by (61), it follows that $a\leq\frac{1}{|\omega|-1}$. ■

Lemma 11. Under H_0, for any $\omega\in\Omega$ with $|\omega|\geq 2$, if $P(\bigcap_{\{i,j\}\in E(K_\omega)} W_{ij}^c)=0$, then $a\in\{\frac{1}{|\omega|-1},\frac{1}{|\omega|-2},\ldots,\frac{1}{2},1\}$.

Proof: Given $\omega \in \Omega$ with $|\omega| \geq 2$, assume H_0 and that $P(\bigcap_{\{i,j\} \in E(K_\omega)} W_{ij}^c) = 0$. The proof will follow by a case analysis on whether $\exists\, \omega' \subset \omega$ with $|\omega'| \geq 2$ so that $P(\bigcap_{\{i,j\} \in E(K_{\omega'})} W_{ij}^c) > 0$.

Case 1: $\exists\, \omega' \subset \omega$ with $|\omega'| \geq 2$ so that $P(\bigcap_{\{i,j\} \in E(K_{\omega'})} W_{ij}^c) > 0$. Fix ω' as the largest proper subset of ω so that $P(\bigcap_{\{i,j\} \in E(K_{\omega'})} W_{ij}^c) > 0$, and fix $r_1 \in \omega - \omega'$. By Lemma 9,

$$P(\bigcap_{\{i,j\} \in E(K_{\omega' \cup r_1})} W_{ij}^c) = \prod_{i=0}^{|\omega'|} (1 - ia) = 0\,. \tag{62}$$

This will hold only if $a \in \{\frac{1}{|\omega'|}, \frac{1}{|\omega'| - 1}, \ldots, \frac{1}{2}, 1\}$. Because $|\omega'| < |\omega|$, the result follows.

Case 2: $\neg\exists\, \omega' \subset \omega$ with $|\omega'| \geq 2$ so that $P(\bigcap_{\{i,j\} \in E(K_{\omega'})} W_{ij}^c) > 0$. Fix $r_1, r_2 \in \omega$, implying that $P(W_{r_1 r_2}^c) = 0$. By Lemma 4, $P(W_{r_1 r_2}^c) = 1 - a$, so

$$a = 1 \in \{\frac{1}{|\omega| - 1}, \frac{1}{|\omega| - 2}, \ldots, \frac{1}{2}, 1\}\,. \qquad \blacksquare$$

Lemma 12. Under H_0, if for all $\omega \in \Omega$ with $|\omega| \geq 2$, $P(\bigcap_{\{i,j\} \in E(K_\omega)} W_{ij}^c) > 0$, then $a \in (0, \frac{1}{|R| - 1})$.

Proof: Assume H_0, and that for all $\omega \in \Omega$ with $|\omega| \geq 2$, $P(\bigcap_{\{i,j\} \in E(K_\omega)} W_{ij}^c) > 0$. Then taking $\omega = R$ and fixing $r_1 \in R$,

$$P(\bigcap_{\{i,j\} \in E(K_{R - r_1})} W_{ij}^c) > 0\,. \tag{63}$$

Hence, by Lemma 10, $a \leq \frac{1}{|R|-1}$. Moreover, by Lemma 9,

$$P(\bigcap_{\{i,j\}\in E(K_R)} W_{ij}^c) = \prod_{i=0}^{|R|-1}(1-ia). \tag{64}$$

Thus, if $a = \frac{1}{|R|-1}$, then $P(\bigcap_{\{i,j\}\in E(K_R)} W_{ij}^c) = 0$, contradicting the assumption that for all $\omega \in \Omega$ with $|\omega| \geq 2$, $P(\bigcap_{\{i,j\}\in E(K_\omega)} W_{ij}^c) > 0$. Therefore $a < \frac{1}{|R|-1}$. ■

Theorem 1. Under H_0, $a \in (0, \frac{1}{|R|-1}) \cup \{\frac{1}{|R|-1}, \frac{1}{|R|-2}, ..., \frac{1}{2}, 1\}$.

Proof: Assume H_0. By definition, either (i) $\exists$ $\omega \in \Omega$ with $|\omega| \geq 2$ so that $P(\bigcap_{\{i,j\}\in E(K_\omega)} W_{ij}^c) = 0$, or (ii) for all $\omega \in \Omega$ with $|\omega| \geq 2$, $P(\bigcap_{\{i,j\}\in E(K_\omega)} W_{ij}^c) > 0$. If (i) is true, fix ω. By Lemma 11,

$$a \in \{\frac{1}{|\omega|-1}, \frac{1}{|\omega|-2}, ..., \frac{1}{2}, 1\} \subseteq \{\frac{1}{|R|-1}, \frac{1}{|R|-2}, ..., \frac{1}{2}, 1\}. \tag{65}$$

If (ii) is true, then by Lemma 12,

$$a \in (0, \frac{1}{|R|-1}). \tag{66}$$

Thus, $a \in (0, \frac{1}{|R|-1}) \cup \{\frac{1}{|R|-1}, \frac{1}{|R|-2}, ..., \frac{1}{2}, 1\}$. ■

Lemma 13. Under H_0, for any (i) ladder graph G with vertex sets $\omega_1, \omega_2 \in \Omega \cup \varnothing$, (ii) $\omega_3 \in \Omega$ with ω_3 and ω_1 disjoint and $\omega_3 \supset \omega_2$, (iii) $r_1 \in (\omega_1 \cup \omega_3)^c$, and (iv) $r_2 \in \omega_3 - \omega_2$: if $P(\bigcap_{\{i,j\}\in E(K_{\omega_3})} W_{ij}^c) > 0$, then

$$P[W_{r_1r_2} \mid (\bigcap_{\{i,j\}\in E(G)} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in E(K_{\omega_3})} W_{ij}^c)] = a .$$

Proof: Given (i) to (iv), assume H_0 and that $P(\bigcap_{\{i,j\}\in E(K_{\omega_3})} W_{ij}^c) > 0$. The proof will follow by induction on $|\omega_1|$. Assume that $|\omega_1| = 0$. Hence, $E(G) = \varnothing$ and by Lemma 7,

$$P[W_{r_1r_2} \mid (\bigcap_{\{i,j\}\in E(G)} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in E(K_{\omega_3})} W_{ij}^c)] = P(W_{r_1r_2} \mid \bigcap_{\{i,j\}\in E(K_{\omega_3})} W_{ij}^c) = a . \tag{67}$$

Now assume that for some integer $0 \le t$, if $|\omega_1| \le t$, then the result holds.

Assume

$|\omega_1| = t+1$ and fix an edge $\{r_3, r_4\} \in E(G)$ with $r_3 \in \omega_2$. Because $P(\bigcap_{\{i,j\}\in E(K_{\omega_3})} W_{ij}^c) > 0$,

$P(\bigcap_{\{i,j\}\in E(K_{\omega_3 - r_3})} W_{ij}^c) > 0$. Thus, by the induction hypothesis,

$$P[W_{r_2r_3} \mid (\bigcap_{\{i,j\}\in E(G)-\{r_3,r_4\}} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in E(K_{\omega_3 - r_3})} W_{ij}^c)] = a , \tag{68}$$

and

$$P[W_{r_2r_4} \mid (\bigcap_{\{i,j\}\in E(G)-\{r_3,r_4\}} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in E(K_{\omega_3 - r_3})} W_{ij}^c)] = a . \tag{69}$$

From (68), it follows that

$$P[W_{r_2r_3} \bigcap (\bigcap_{\{i,j\}\in E(G)-\{r_3,r_4\}} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in E(K_{\omega_3 - r_3})} W_{ij}^c)] > 0 .$$

Hence, by Lemma 2, (68) and (69) imply that

$$P[W_{r_2r_4} \mid W_{r_2r_3} \bigcap (\bigcap_{\{i,j\}\in E(G)-\{r_3,r_4\}} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in E(K_{\omega_3 - r_3})} W_{ij}^c)] = a . \tag{70}$$

Likewise,

$$P[W_{r_1r_2} \mid W_{r_2r_3} \bigcap (\bigcap_{\{i,j\}\in E(G)-\{r_3,r_4\}} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in E(K_{\omega_3 - r_3})} W_{ij}^c)] = a . \tag{71}$$

(70) implies that

$$P[W_{r_2r_3} \bigcap W_{r_2r_4} \bigcap (\bigcap_{\{i,j\} \in E(G) - \{r_3, r_2\}} W_{ij}) \bigcap (\bigcap_{\{i,j\} \in E(K_{\omega_3 - r_3})} W_{ij}^c)] > 0$$

so by Lemma 2 again, (70) and (71) imply that

$$P[W_{r_1r_2} \mid W_{r_2r_3} \bigcap W_{r_2r_4} \bigcap (\bigcap_{\{i,j\} \in E(G) - \{r_3, r_2\}} W_{ij}) \bigcap (\bigcap_{\{i,j\} \in E(K_{\omega_3 - r_3})} W_{ij}^c)] = a \,. \tag{72}$$

By Lemma 3,

$$W_{r_2r_3} \bigcap W_{r_2r_4} \bigcap (\bigcap_{\{i,j\} \in E(G) - \{r_3, r_4\}} W_{ij}) \bigcap (\bigcap_{\{i,j\} \in E(K_{\omega_3 - r_3})} W_{ij}^c)$$

$$= W_{r_2r_3} \bigcap W_{r_2r_4} \bigcap W_{r_3r_4} \bigcap (\bigcap_{\{i,j\} \in E(G) - \{r_3, r_4\}} W_{ij}) \bigcap (\bigcap_{\{i,j\} \in E(K_{\omega_3 - r_3})} W_{ij}^c)$$

$$= W_{r_2r_3} \bigcap W_{r_3r_4} \bigcap (\bigcap_{\{i,j\} \in E(G) - \{r_3, r_4\}} W_{ij}) \bigcap (\bigcap_{\{i,j\} \in E(K_{\omega_3 - r_3})} W_{ij}^c) \,, \tag{73}$$

and by definition,

$$W_{r_2r_3} \bigcap W_{r_3r_4} \bigcap (\bigcap_{\{i,j\} \in E(G) - \{r_3, r_4\}} W_{ij}) \bigcap (\bigcap_{\{i,j\} \in E(K_{\omega_3 - r_3})} W_{ij}^c)$$

$$= W_{r_2r_3} \bigcap (\bigcap_{\{i,j\} \in E(G)} W_{ij}) \bigcap (\bigcap_{\{i,j\} \in E(K_{\omega_3 - r_3})} W_{ij}^c) \,. \tag{74}$$

Moreover, by Lemma 6,

$$W_{r_2r_3} \bigcap (\bigcap_{\{i,j\} \in E(G)} W_{ij}) \bigcap (\bigcap_{\{i,j\} \in E(K_{\omega_3 - r_3})} W_{ij}^c) = W_{r_2r_3} \bigcap (\bigcap_{\{i,j\} \in E(G)} W_{ij}) \bigcap (\bigcap_{\{i,j\} \in E(K_{\omega_3}) - \{r_2, r_3\}} W_{ij}^c) \,. \tag{75}$$

Thus, (72), (74), and (75) imply that

$$P[W_{r_1r_2} \mid W_{r_2r_3} \bigcap (\bigcap_{\{i,j\} \in E(G)} W_{ij}) \bigcap (\bigcap_{\{i,j\} \in E(K_{\omega_3}) - \{r_2, r_3\}} W_{ij}^c)] = a. \tag{76}$$

Finally, because $P\{ \bigcap_{\{i,j\} \in E(K_{\omega_3})} W_{ij}^c \} > 0$, by the multiplication rule for conditional probability

$$P[(\bigcap_{\{i,j\} \in E(G)} W_{ij}) \bigcap (\bigcap_{\{i,j\} \in E(K_{\omega_3})} W_{ij}^c)] = P(\bigcap_{\{i,j\} \in E(G)} W_{ij} \mid \bigcap_{\{i,j\} \in E(K_{\omega_3})} W_{ij}^c) \cdot P(\bigcap_{\{i,j\} \in E(K_{\omega_3})} W_{ij}^c) \,. \tag{77}$$

Moreover, it follows from the induction hypothesis that

$$P(\bigcap_{\{i,j\}\in E(G)} W_{ij} \mid \bigcap_{\{i,j\}\in E(K_{\omega_3})} W_{ij}^c) > 0 . \tag{78}$$

Hence,

$$P[(\bigcap_{\{i,j\}\in E(G)} W_{ij})\bigcap(\bigcap_{\{i,j\}\in E(K_{\omega_3})} W_{ij}^c)] > 0 , \tag{79}$$

and H_0 applies to yield

$$P[W_{r_1r_2} \mid W_{r_2r_3}^c \bigcap(\bigcap_{\{i,j\}\in E(G)} W_{ij})\bigcap(\bigcap_{\{i,j\}\in E(K_{\omega_3})-\{r_2,r_3\}} W_{ij}^c)]$$

$$= P[W_{r_1r_2} \mid W_{r_2r_3} \bigcap(\bigcap_{\{i,j\}\in E(G)} W_{ij})\bigcap(\bigcap_{\{i,j\}\in E(K_{\omega_3})-\{r_2,r_3\}} W_{ij}^c)] . \tag{80}$$

The result follows directly from (76). ■

Lemma 14. Under H_0, for any partition ψ of *R*, spanning star forest *F* of K_ψ^c, proper edge-induced subgraph F' of F, and $\{r_1,r_2\} \in E(F')^c \bigcap E(F)$ with $r_2 \in V(F')$, if $P(\bigcap_{\{i,j\}\in E(K_F^*)} W_{ij}^c) > 0$, then

$$P[W_{r_1r_2} \mid (\bigcap_{\{i,j\}\in E(F')} W_{ij})\bigcap(\bigcap_{\{i,j\}\in E(K_F^*)} W_{ij}^c)] = a .$$

Proof: Given partition ψ, a spanning star forest *F* of K_ψ^c, a proper edge-induced subgraph F' of F, and $\{r_1,r_2\} \in E(F')^c \bigcap E(F)$ with $r_2 \in V(F')$, assume H_0 and that $P(\bigcap_{\{i,j\}\in E(K_F^*)} W_{ij}^c) > 0$. It follows that $\exists$ an edge in $E(F')$ containing r_2. Fix this edge as $\{r_2,r_3\}$. The proof will follow by induction on $|E(F')| - \tau(F')$.

Assume $|E(F')| - \tau(F') = 0$. Then $\exists$ a ladder graph with edge set $E(F') - \{r_2,r_3\}$, and so by Lemma 13,

$$P[W_{r_1 r_2} \mid (\bigcap_{\{i,j\}\in E(F')-\{r_2,r_3\}} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in E(K_F^*)} W_{ij}^c)] = a \tag{81}$$

and

$$P[W_{r_2 r_3} \mid (\bigcap_{\{i,j\}\in E(F')-\{r_2,r_3\}} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in E(K_F^*)} W_{ij}^c)] = a\,. \tag{82}$$

From (82),

$$P[W_{r_2 r_3} \bigcap (\bigcap_{\{i,j\}\in E(F')-\{r_2,r_3\}} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in E(K_F^*)} W_{ij}^c)] > 0\,, \tag{83}$$

and so by Lemma 2,

$$P[W_{r_1 r_2} \mid (\bigcap_{\{i,j\}\in E(F')} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in E(K_F^*)} W_{ij}^c)] = a\,. \tag{84}$$

Assume now that for some integer $t \geq 0$, the result holds if $|E(F')| - \tau(F') \leq t$. Assume that $|E(F')| - \tau(F') = t+1$. A case analysis is required on whether $\{r_2, r_3\}$ is the only edge in $E(F')$ containing r_2.

Case 1: $\{r_2, r_3\}$ is the only edge in $E(F')$ containing r_2.

Because $|E(F')| - \tau(F') \geq 1$, $\exists$ at least one $v \in V(F')$ with degree of at least 2. Fix this vertex as r_4 and an edge containing this vertex as $\{r_4, r_5\}$. By the case assumption, r_2, r_3, r_4, r_5 must be distinct. Thus, for the edge-induced subgraph $(V(F') - r_5, E(F') - \{r_4, r_5\})$ of F, the difference between the number of edges and trees is

$$|E(F') - \{r_4, r_5\}| - \tau(V(F') - r_5, E(F') - \{r_4, r_5\}) = |E(F')| - 1 - \tau(F') = t\,. \tag{85}$$

Likewise, for the edge-induced subgraph $(V(F') - r_2 - r_3 - r_5, E(F') - \{r_4, r_5\} - \{r_2, r_3\})$ of F, the difference is

$$|E(F') - \{r_4, r_5\} - \{r_2, r_3\}| - \tau(V(F') - r_2 - r_3 - r_5, E(F') - \{r_4, r_5\} - \{r_2, r_3\})$$
$$= |E(F')| - 2 - [\tau(F') - 1] = t\,. \tag{86}$$

Therefore, by the induction hypothesis,

$$P[W_{r_4r_5} \mid (\bigcap_{\{i,j\}\in E(F')-\{r_4,r_5\}} W_{ij})\bigcap(\bigcap_{\{i,j\}\in E(K_F^*)} W_{ij}^c)] = a\,, \tag{87}$$

$$P[W_{r_2r_3} \mid (\bigcap_{\{i,j\}\in E(F')-\{r_4,r_5\}-\{r_2,r_3\}} W_{ij})\bigcap(\bigcap_{\{i,j\}\in E(K_F^*)} W_{ij}^c)] = a\,, \tag{88}$$

and

$$P[W_{r_4r_5} \mid (\bigcap_{\{i,j\}\in E(F')-\{r_4,r_5\}-\{r_2,r_3\}} W_{ij})\bigcap(\bigcap_{\{i,j\}\in E(K_F^*)} W_{ij}^c)] = a \tag{89}$$

It follows then that

$$a = \frac{P[W_{r_4r_5} \mid (\bigcap_{\{i,j\}\in E(F')-\{r_4,r_5\}} W_{ij})\bigcap(\bigcap_{\{i,j\}\in E(K_F^*)} W_{ij}^c)]}{P[W_{r_4r_5} \mid (\bigcap_{\{i,j\}\in E(F')-\{r_4,r_5\}-\{r_2,r_3\}} W_{ij})\bigcap(\bigcap_{\{i,j\}\in E(K_F^*)} W_{ij}^c)]}$$

$$\cdot P[W_{r_2r_3} \mid (\bigcap_{\{i,j\}\in E(F')-\{r_4,r_5\}-\{r_2,r_3\}} W_{ij})\bigcap(\bigcap_{\{i,j\}\in E(K_F^*)} W_{ij}^c)]$$

$$= \frac{P[W_{r_2r_3}\bigcap(\bigcap_{\{i,j\}\in E(F')-\{r_2,r_3\}} W_{ij})\bigcap(\bigcap_{\{i,j\}\in E(K_F^*)} W_{ij}^c)]}{P[(\bigcap_{\{i,j\}\in E(F')-\{r_2,r_3\}} W_{ij})\bigcap(\bigcap_{\{i,j\}\in E(K_F^*)} W_{ij}^c)]}$$

$$= P[W_{r_2r_3} \mid (\bigcap_{\{i,j\}\in E(F')-\{r_2,r_3\}} W_{ij})\bigcap(\bigcap_{\{i,j\}\in E(K_F^*)} W_{ij}^c)]\,. \tag{90}$$

Turning to the edge induced subgraphs

$(V(F')\cup r_1 - r_3 - r_5, E(F')\cup\{r_1,r_2\}-\{r_2,r_3\}-\{r_4,r_5\})$ and

$(V(F') - r_2 - r_3 - r_5, E(F')-\{r_4,r_5\}-\{r_2,r_3\})$ of F, by similar arguments,

$$P[W_{r_1r_2} \mid (\bigcap_{\{i,j\}\in E(F')-\{r_2,r_3\}} W_{ij})\bigcap(\bigcap_{\{i,j\}\in E(K_F^*)} W_{ij}^c)] = a\,. \tag{91}$$

By the definition of conditional probability, (90) implies that

$$P[W_{r_2r_3}\bigcap(\bigcap_{\{i,j\}\in E(F')-\{r_2,r_3\}} W_{ij})\bigcap(\bigcap_{\{i,j\}\in E(K_F^*)} W_{ij}^c)] > 0\,. \tag{92}$$

Hence, applying Lemma 2 to (90) and (91) gives the result.

Case 2: $\{r_2, r_3\}$ is not the only edge in $E(F')$ containing r_2. For the edge-induced subgraph $(V(F') - r_3, E(F') - \{r_2, r_3\})$ of F,

$$\left| E(F') - \{r_2, r_3\} \right| - \tau(V(F') - r_3, E(F') - \{r_2, r_3\}) = t. \tag{93}$$

Thus, by the induction hypothesis,

$$P[W_{r_1 r_2} \mid (\bigcap_{\{i,j\} \in E(F') - \{r_2, r_3\}} W_{ij}) \bigcap (\bigcap_{\{i,j\} \in E(K_F^*)} W_{ij}^c)] = a \tag{94}$$

and

$$P[W_{r_2 r_3} \mid (\bigcap_{\{i,j\} \in E(F') - \{r_2, r_3\}} W_{ij}) \bigcap (\bigcap_{\{i,j\} \in E(K_F^*)} W_{ij}^c)] = a. \tag{95}$$

(94) implies that

$$P[W_{r_2 r_3} \bigcap (\bigcap_{\{i,j\} \in E(F') - \{r_2, r_3\}} W_{ij}) \bigcap (\bigcap_{\{i,j\} \in E(K_F^*)} W_{ij}^c)] > 0, \tag{96}$$

so again the result follows by Lemma 2. ■

Lemma 15. For any star *G*, $|Q_G| = 1$.

Proof: Given a star *G*, by definition $\exists$ a vertex $v \in V(G)$ adjacent to all others. Let v be the covering vertex for the star, creating a vertex cover of size 1. Because a vertex cover cannot be empty, the constructed cover must also be minimal and hence $|Q_G| = 1$.

■

Lemma 16. For any $X, Y \in E(K_R)$, and $r_1, r_2 \in R$ with $r_1 \neq r_2$, if $\exists$ $r_3, r_4 \in R$ such that $\{r_1, r_3\}, \{r_2, r_4\} \in X$ and $\{r_3, r_4\} \in Y$, then

$$W_{r_1 r_2}^c \bigcap (\bigcap_{\{i,j\} \in X} W_{ij}) \bigcap (\bigcap_{\{i,j\} \in Y} W_{ij}^c) = (\bigcap_{\{i,j\} \in X} W_{ij}) \bigcap (\bigcap_{\{i,j\} \in Y} W_{ij}^c).$$

Proof: Given $X, Y \in E(K_R)$, and $r_1, r_2 \in R$ with $r_1 \neq r_2$, assume $\exists\ r_3, r_4 \in R$ such that $\{r_1, r_3\}, \{r_2, r_4\} \in X$ and $\{r_3, r_4\} \in Y$. Fix r_3 and r_4. Given $\psi \in S$, assume that $\psi \in W^c_{r_1r_2} \bigcap (\bigcap_{\{i,j\}\in X} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in Y} W^c_{ij})$. Because by definition,

$$W^c_{r_1r_2} \bigcap (\bigcap_{\{i,j\}\in X} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in Y} W^c_{ij}) \subseteq (\bigcap_{\{i,j\}\in X} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in Y} W^c_{ij}), \tag{97}$$

it follows that $\psi \in (\bigcap_{\{i,j\}\in X} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in Y} W^c_{ij})$. Now assume that $\psi \in (\bigcap_{\{i,j\}\in X} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in Y} W^c_{ij})$. Thus, $(r_1, r_3), (r_2, r_4) \in \mathrm{Eq}_\psi$ and $(r_3, r_4) \notin \mathrm{Eq}_\psi$. It follows that $(r_1, r_4) \notin \mathrm{Eq}_\psi$, because otherwise, transitively $(r_3, r_4) \in \mathrm{Eq}_\psi$, a contradiction. $(r_1, r_4) \notin \mathrm{Eq}_\psi$ and $(r_2, r_4) \in \mathrm{Eq}_\psi$ imply by similar arguments that $(r_1, r_2) \notin \mathrm{Eq}_\psi$. Thus, $\psi \in W^c_{r_1r_2}$, and $\psi \in W^c_{r_1r_2} \bigcap (\bigcap_{\{i,j\}\in X} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in Y} W^c_{ij})$. Therefore, $\psi \in W^c_{r_1r_2} \bigcap (\bigcap_{\{i,j\}\in X} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in Y} W^c_{ij})$ if and only if $\psi \in (\bigcap_{\{i,j\}\in X} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in Y} W^c_{ij})$. ■

Lemma 17. For any partition ψ of R, $X \subseteq E(K^c_\psi), Y \subseteq E(K_\psi)$, $\{r_1, r_2\} \in E(K_\psi)$, and spanning star forest F of K^c_ψ,

$$W^c_{r_1r_2} \bigcap (\bigcap_{\{i,j\}\in E(F)\cup X} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in E(K^*_F)\cup Y} W^c_{ij}) = (\bigcap_{\{i,j\}\in E(F)\cup X} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in E(K^*_F)\cup Y} W^c_{ij}).$$

Proof: Given partition ψ, $X \subseteq E(K^c_\psi), Y \subseteq E(K_\psi)$, $\{r_1, r_2\} \in E(K_\psi)$, and a spanning star forest F of K^c_ψ, the proof will follow from a case analysis on whether $r_1, r_2 \in Q^+_F$.

Case 1: $r_1, r_2 \in Q^+_F$. By the definition of K^*_F, $\{r_1, r_2\} \in E(K^*_F)$. Hence, $\{r_1, r_2\} \in E(K^*_F) \cup Y$, and the claim follows immediately.

Case 2: $r_1 \notin Q_F^+$ and $r_2 \in Q_F^+$. By Lemma 15, $\exists$ a vertex cover of size 1 for r_1's star. Fix the vertex in this cover as r_3. Clearly, because $r_1 \notin Q_F^+$, $r_1 \neq r_3$. Moreover, by definition r_1 and r_3 are adjacent in r_1's star, and hence they are adjacent in F. Thus, we have that $\{r_1,r_3\} \in E(F) \subseteq E(F) \cup X \subseteq E(K_\psi^c)$. Additionally, because $r_2,r_3 \in Q_F^+$, $\{r_2,r_3\} \in E(K_F^*) \subseteq E(K_F^*) \cup Y \subseteq E(K_\psi)$. The hypotheses of Lemma 5 are hence satisfied, and the claim follows.

Case 3: $r_1 \in Q_F^+$ and $r_2 \notin Q_F^+$. By symmetry, this case follows from the same arguments as case 2.

Case 4: $r_1,r_2 \notin Q_F^+$. Because by definition $\{r_1,r_2\} \in E(K_\psi)$, r_1 and r_2 must be elements of different stars in F. By Lemma 15, $\exists$ vertex covers each with a single element for both r_1 and r_2's stars. Fix the covering vertices as r_3 and r_4, respectively. As in case 2, $\{r_1,r_3\},\{r_2,r_4\} \in E(F) \subseteq E(F) \cup X \subseteq E(K_\psi^c)$. Moreover, because r_3 and r_4 are both covering vertices, $\{r_3,r_4\} \in E(K_F^*) \subseteq E(K_F^*) \cup Y \subseteq E(K_\psi)$, as in case 1. Hence, the hypotheses of Lemma 16 are satisfied and the claim follows.

■

Lemma 18. For any partition ψ of R, $X \subseteq E(K_\psi^c), Y \subseteq E(K_\psi)$, $\{r_1,r_2\} \in E(K_\psi^c)$, and spanning star forest F of K_ψ^c,

$$W_{r_1r_2} \bigcap \Big(\bigcap_{\{i,j\} \in E(F) \cup X} W_{ij} \Big) \bigcap \Big(\bigcap_{\{i,j\} \in E(K_F^*) \cup Y} W_{ij}^c \Big) = \Big(\bigcap_{\{i,j\} \in E(F) \cup X} W_{ij} \Big) \bigcap \Big(\bigcap_{\{i,j\} \in E(K_F^*) \cup Y} W_{ij}^c \Big).$$

Proof: Given partition ψ, $X \subseteq E(K_\psi^c), Y \subseteq E(K_\psi)$, $\{r_1,r_2\} \in E(K_\psi^c)$, and a spanning star forest F of K_ψ^c, the proof will follow by a case analysis on whether $r_1,r_2 \in Q_F^+$.

Case 1: $r_1, r_2 \in Q_F^+$. Note that because $\{r_1, r_2\} \in E(K_\psi^c)$, r_1 and r_2 must be distinct vertices in the same star of F with $d(r_1) \geq 1$ and $d(r_2) \geq 1$. But by Lemma 15 $\exists$ only one element of Q_F per star, implying a contradiction because Q_F^+ contains only of vertices of degree zero and elements of Q_F. Therefore, r_1 and r_2 cannot both be elements of Q_F^+.

Case 2: $r_1 \notin Q_F^+$ and $r_2 \in Q_F^+$. By definition, r_1 and r_2 must be distinct vertices in the same star. Also, by Lemma 15, $\exists$ exactly one element of Q_F in r_1 and r_2's star. Because $d(r_2) \neq 0$ and $r_2 \in Q_F^+$, this vertex must be r_2, implying that r_1 and r_2 are adjacent in F. Thus, we have that $\{r_1, r_2\} \in E(F) \cup X$, implying the claim.

Case 3: $r_1 \in Q_F^+$ and $r_2 \notin Q_F^+$. By symmetry, the arguments from case 2 apply.

Case 4: $r_1, r_2 \notin Q_F^+$. By Lemma 15, $\exists$ exactly one element of Q_F in r_1 and r_2's star. Fix this vertex as r_3. Because $r_1, r_2 \notin Q_F^+$, $r_1 \neq r_3$ and $r_2 \neq r_3$. However, by definition, both r_1 and r_2 must be adjacent to r_3, implying that $\{r_1, r_3\}, \{r_2, r_3\} \in E(F) \subseteq E(F) \cup X \subseteq E(K_\psi^c)$. Hence, the hypotheses of Lemma 3 are satisfied and the claim follows. ■

Lemma 19. For any partition ψ of R, $X \subseteq E(K_\psi^c)$, $Y \subseteq E(K_\psi)$, and spanning star forest F of K_ψ^c,

$$\Big(\bigcap_{\{i,j\} \in E(K_\psi^c)} W_{ij}\Big) \bigcap \Big(\bigcap_{\{i,j\} \in E(K_\psi)} W_{ij}^c\Big) = \Big(\bigcap_{\{i,j\} \in E(F)} W_{ij}\Big) \bigcap \Big(\bigcap_{\{i,j\} \in E(K_F^*)} W_{ij}^c\Big).$$

Proof: Given a partitions ψ and ψ' of R, $X \subseteq E(K_\psi^c)$, $Y \subseteq E(K_\psi)$, and a spanning star forest F of K_ψ^c, assume that $\psi' \in \Big(\bigcap_{\{i,j\} \in E(K_\psi^c)} W_{ij}\Big) \bigcap \Big(\bigcap_{\{i,j\} \in E(K_\psi)} W_{ij}^c\Big)$. By definition,

$$\Big(\bigcap_{\{i,j\}\in E(K_\psi^c)} W_{ij}\Big)\bigcap\Big(\bigcap_{\{i,j\}\in E(K_\psi)} W_{ij}^c\Big)\subseteq\Big(\bigcap_{\{i,j\}\in E(F)} W_{ij}\Big)\bigcap\Big(\bigcap_{\{i,j\}\in E(K_F^*)} W_{ij}^c\Big), \tag{98}$$

so $\psi'\in\Big(\bigcap_{\{i,j\}\in E(F)} W_{ij}\Big)\bigcap\Big(\bigcap_{\{i,j\}\in E(K_F^*)} W_{ij}^c\Big)$. Now assume that

$\psi'\in\Big(\bigcap_{\{i,j\}\in E(F)} W_{ij}\Big)\bigcap\Big(\bigcap_{\{i,j\}\in E(K_F^*)} W_{ij}^c\Big)$. Given $\{r_1,r_2\}\in E(K_\psi^c)-E(F)$, by Lemma 18,

$$W_{r_1r_2}\bigcap\Big(\bigcap_{\{i,j\}\in E(F)} W_{ij}\Big)\bigcap\Big(\bigcap_{\{i,j\}\in E(K_F^*)} W_{ij}^c\Big)=\Big(\bigcap_{\{i,j\}\in E(F)} W_{ij}\Big)\bigcap\Big(\bigcap_{\{i,j\}\in E(K_F^*)} W_{ij}^c\Big), \tag{99}$$

implying that

$$\psi'\in W_{r_1r_2}\bigcap\Big(\bigcap_{\{i,j\}\in E(F)} W_{ij}\Big)\bigcap\Big(\bigcap_{\{i,j\}\in E(K_F^*)} W_{ij}^c\Big). \tag{100}$$

Because (100) holds for any $\{r_1,r_2\}\in E(K_\psi^c)-E(F)$, it follows that

$$\psi'\in\Big(\bigcap_{\{i,j\}\in E(K_\psi^c)} W_{ij}\Big)\bigcap\Big(\bigcap_{\{i,j\}\in E(K_F^*)} W_{ij}^c\Big). \tag{101}$$

Likewise, for any $\{r_1,r_2\}\in E(K_\psi)-E(K_F^*)$, by Lemma 17,

$$\psi'\in W_{r_1r_2}^c\bigcap\Big(\bigcap_{\{i,j\}\in E(K_\psi^c)} W_{ij}\Big)\bigcap\Big(\bigcap_{\{i,j\}\in E(K_F^*)} W_{ij}^c\Big), \tag{102}$$

implying that

$$\psi'\in\Big(\bigcap_{\{i,j\}\in E(K_\psi^c)} W_{ij}\Big)\bigcap\Big(\bigcap_{\{i,j\}\in E(K_\psi)} W_{ij}^c\Big). \tag{103}$$

Therefore, $\psi'\in\Big(\bigcap_{\{i,j\}\in E(K_\psi^c)} W_{ij}\Big)\bigcap\Big(\bigcap_{\{i,j\}\in E(K_\psi)} W_{ij}^c\Big)$ if and only if

$\psi'\in\Big(\bigcap_{\{i,j\}\in E(F)} W_{ij}\Big)\bigcap\Big(\bigcap_{\{i,j\}\in E(K_F^*)} W_{ij}^c\Big)$. ■

Lemma 20. For any partition ψ of *R*,

$$\{\psi\}=\Big(\bigcap_{\{i,j\}\in E(K_\psi^c)} W_{ij}\Big)\bigcap\Big(\bigcap_{\{i,j\}\in E(K_\psi)} W_{ij}^c\Big).$$

Proof: Given partitions ψ and ψ' of *R*, assume first that $\psi' \in \{\psi\}$. For any $r_1, r_2 \in R$, if $\{r_1, r_2\} \in E(K_\psi^c)$, then $(r_1, r_2) \in \mathrm{Eq}_\psi$. Hence,

$$\psi \in \bigcap_{\{i,j\} \in E(K_\psi^c)} W_{ij} . \tag{104}$$

Moreover, by definition, if $\{r_1, r_2\} \in E(K_\psi)$, then $(r_1, r_2) \notin \mathrm{Eq}_\psi$, so

$$\psi \in \bigcap_{\{i,j\} \in E(K_\psi)} W_{ij}^c . \tag{105}$$

By assumption, $\psi' \in \{\psi\}$, so $\psi' = \psi$, and by (104) and (105)

$$\psi' \in (\bigcap_{\{i,j\} \in E(K_\psi^c)} W_{ij}) \bigcap (\bigcap_{\{i,j\} \in E(K_\psi)} W_{ij}^c) . \tag{106}$$

Now assume that $\psi' \in (\bigcap_{\{i,j\} \in E(K_\psi^c)} W_{ij}) \bigcap (\bigcap_{\{i,j\} \in E(K_\psi)} W_{ij}^c)$. Thus, given distinct $r_1, r_2 \in R$, $\psi' \in W_{r_1 r_2}$ if $\{r_1, r_2\} \in E(K_\psi^c)$, and $\psi' \notin W_{r_1 r_2}$ if $\{r_1, r_2\} \in E(K_\psi)$. Because either $\{r_1, r_2\} \in E(K_\psi^c)$ or $\{r_1, r_2\} \in E(K_\psi)$, it follows that $\psi' \in W_{r_1 r_2}$ if and only if $\{r_1, r_2\} \in E(K_\psi^c)$. Moreover, $\psi' \in W_{r_1 r_2}$ if and only if $(r_1, r_2) \in \mathrm{Eq}_{\psi'}$. Thus, $(r_1, r_2) \in \mathrm{Eq}_{\psi'}$ if and only if $\{r_1, r_2\} \in E(K_\psi^c)$. Finally, $\{r_1, r_2\} \in E(K_\psi^c)$ if and only if $(r_1, r_2) \in \mathrm{Eq}_\psi$, so $(r_1, r_2) \in \mathrm{Eq}_{\psi'}$ if and only if $(r_1, r_2) \in \mathrm{Eq}_\psi$. Equivalence relations "define" partitions and vice versa (Hrbacek and Jech, Chapter 2), so $\psi' = \psi$, and hence $\psi' \in \{\psi\}$. Therefore, $\psi' \in \{\psi\}$ if and only if

$$\psi' \in (\bigcap_{\{i,j\} \in E(K_\psi^c)} W_{ij}) \bigcap (\bigcap_{\{i,j\} \in E(K_\psi)} W_{ij}^c) . \qquad \blacksquare$$

Lemma 21. For any partition ψ of *R* and spanning star forest *F* of K_ψ^c, if $|V(K_F^*)| = 1$, then

$$P[(\bigcap_{\{i,j\}\in E(F)} W_{ij})\bigcap(\bigcap_{\{i,j\}\in E(K_F^*)} W_{ij}^c)] = a^{|E(F)|}.$$

Proof: Given a partition ψ of R and spanning star forest F of K_ψ^c, assume that $|V(K_F^*)| = 1$. Thus, $|E(K_F^*)| = 0$, and

$$P[(\bigcap_{\{i,j\}\in E(F)} W_{ij})\bigcap(\bigcap_{\{i,j\}\in E(K_F^*)} W_{ij}^c)] = P(\bigcap_{\{i,j\}\in E(F)} W_{ij}). \tag{107}$$

Fix an enumeration $r_1, r_2, ..., r_{|V(F)|}$ of $V(F)$, with $\{r_1\} = V(K_F^*)$. The remainder of the proof will follow by induction on $|V(F)|$. $|V(K_F^*)| = 1$ implies by definition that F contains exactly 1 tree, whose vertices are all of the subunits in R. Hence, $|V(F)| \geq 3$. Assume that $|V(F)| = 3$. By Lemma 4, $P(W_{r_1r_2}), P(W_{r_1r_3}) = a$, so by Lemma 2,

$$P(W_{r_1r_2} \mid W_{r_1r_3}) = P(W_{r_1r_2}), \tag{108}$$

implying that

$$P(W_{r_1r_2} \bigcap W_{r_1r_3}) = a^2. \tag{109}$$

Because any tree must have 1 fewer edges than vertices (Bogart 1983), $|E(F)| = 2$ and the result follows.

Now assume that for the integer $3 \leq t$, if $|V(F)| \leq t$, then the result holds. Assume that $|V(F)| = t+1$. By the induction hypothesis,

$$P(\bigcap_{\{i,j\}\in E(F)-\{r_1,r_t\}} W_{ij}) = a^t, \tag{110}$$

$$P(\bigcap_{\{i,j\}\in E(F)-\{r_1,r_{t-1}\}} W_{ij}) = a^t, \tag{111}$$

and

$$P(\bigcap_{\{i,j\}\in E(F)-\{r_1,r_t\}-\{r_1,r_{t-1}\}} W_{ij}) = a^{t-1}. \tag{112}$$

Therefore,

$$P(W_{r_1 r_t} \mid \bigcap_{\{i,j\}\in E(F)-\{r_1,r_t\}-\{r_1,r_{t-1}\}} W_{ij}) = \frac{P(\bigcap_{\{i,j\}\in E(F)-\{r_1,r_{t-1}\}} W_{ij})}{P(\bigcap_{\{i,j\}\in E(F)-\{r_1,r_t\}-\{r_1,r_{t-1}\}} W_{ij})} = a \tag{113}$$

and

$$P(W_{r_1 r_{t-1}} \mid \bigcap_{\{i,j\}\in E(F)-\{r_1,r_t\}-\{r_1,r_{t-1}\}} W_{ij}) = \frac{P(\bigcap_{\{i,j\}\in E(F)-\{r_1,r_t\}} W_{ij})}{P(\bigcap_{\{i,j\}\in E(F)-\{r_1,r_t\}-\{r_1,r_{t-1}\}} W_{ij})} = a\,. \tag{114}$$

Thus, by Lemma 1, (112), (113), and (114) imply that

$$P(\bigcap_{\{i,j\}\in E(F)} W_{ij}) = P(W_{r_1 r_t} \bigcap W_{r_1 r_{t-1}} \mid \bigcap_{\{i,j\}\in E(F)-\{r_1,r_t\}-\{r_1,r_{t-1}\}} W_{ij})$$

$$\cdot P(\bigcap_{\{i,j\}\in E(F)-\{r_1,r_t\}-\{r_1,r_{t-1}\}} W_{ij}) = a^2 \cdot a^{t-1} = a^{t+1}\,. \qquad \blacksquare$$

Theorem 2. Under H_0, for any partition ψ of R,

$$P\{\psi\} = a^{|R|-|\psi|} \cdot \prod_{i=0}^{|\psi|-1} (1-ia)\,.$$

Proof: Given a partition ψ of R and a spanning star forest F of K_ψ^c, assume H_0. By Lemma 19 and Lemma 20,

$$P\{\psi\} = P[(\bigcap_{\{i,j\}\in E(F)} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in E(K_F^*)} W_{ij}^c)]\,. \tag{115}$$

The remainder of the proof will follow by a case analysis on whether $|V(K_F^*)| \geq 2$ and $P\{\bigcap_{\{i,j\}\in E(K_F^*)} W_{ij}^c\} = 0$.

Case 1: $|V(K_F^*)| \geq 2$ and $P(\bigcap_{\{i,j\}\in E(K_F^*)} W_{ij}^c) = 0$. It follows that $P[(\bigcap_{\{i,j\}\in E(F)} W_{ij}) \bigcap (\bigcap_{\{i,j\}\in E(K_F^*)} W_{ij}^c)] = 0$ and hence by (115) that $P\{\psi\} = 0$. Also, by definition, $V(K_F^*) \in \Omega$, so because $\left|V(K_F^*)\right| \geq 2$, by Lemma 11 $a \in \{\frac{1}{\left|V(K_F^*)\right|-1}, \frac{1}{\left|V(K_F^*)\right|-2}, \ldots, \frac{1}{2}, 1\}$. Moreover, by definition,

$$\left|V(K_F^*)\right| = |\psi|, \tag{116}$$

so the result follows by substitution.

Case 1: $\left|V(K_F^*)\right| \geq 2$ and $P(\bigcap_{\{i,j\}\in E(K_F^*)} W_{ij}^c) > 0$. Because $V(K_F^*) \in \Omega$ and $\left|V(K_F^*)\right| \geq 2$, Lemma 9 implies that

$$P\{\bigcap_{\{i,j\}\in E(K_F^*)} W_{ij}^c\} = \prod_{i=0}^{\left|V(K_F^*)\right|-1} (1-ia). \tag{117}$$

Fix an enumeration η_F of $E(F)$, with the first $\tau(F)$ elements being in different trees, and denote the *l*th element in η_F as $\{i_l, j_l\}$. It follows from Lemma 13 and Lemma 14 that for all $1 \leq l \leq \left|E(F_\psi^*)\right|$,

$$P\{W_{i_l j_l} \mid (\bigcap_{\{i,j\}\in E(K_F^*)} W_{ij}^c) \bigcap (\bigcap_{m=1}^{l} W_{i_m j_m}) - W_{i_l j_l}\} = a. \tag{118}$$

Thus, by the product rule,

$$P\{\bigcap_{\{i,j\}\in E(F)} W_{ij} \mid \bigcap_{\{i,j\}\in E(K_F^*)} W_{ij}^c\} = \prod_{l=1}^{|E(F)|} P\{W_{i_l j_l} \mid (\bigcap_{\{i,j\}\in E(K_F^*)} W_{ij}^c) \bigcap (\bigcap_{m=1}^{l} W_{i_m j_m}) - W_{i_l j_l}\}. \tag{119}$$

Also, any forest G contains $|V(G)| - \tau(G)$ edges (Bogart 1983). Hence, $|E(F)| = |V(F)| - \tau(F)$. But by definition, $|V(F)| = |R|$ and $\tau(F) = \left|V(K_F^*)\right|$, implying that

$$|E(F)| = |R| - \left|V(K_F^*)\right|. \tag{120}$$

Therefore, (118), (119), and (120) imply that

$$P\{\bigcap_{\{i,j\}\in E(F)} W_{ij} \mid \bigcap_{\{i,j\}\in E(K_F^*)} W_{ij}^c\} = a^{|R|-|V(K_F^*)|}. \tag{121}$$

It follows by (117), (121), and the product rule that

$$P[(\bigcap_{\{i,j\}\in E(K_\psi^c)} W_{ij})\bigcap(\bigcap_{\{i,j\}\in E(K_\psi)} W_{ij}^c)] = P\{\bigcap_{\{i,j\}\in E(K_F^*)} W_{ij}^c\}\cdot P\{\bigcap_{\{i,j\}\in E(F)} W_{ij} \mid \bigcap_{\{i,j\}\in E(K_F^*)} W_{ij}^c\}$$

$$= a^{|R|-|V(K_F^*)|}\cdot\prod_{i=0}^{|V(K_F^*)|-1}(1-ia), \tag{122}$$

and by (115) that

$$P\{\psi\} = a^{|R|-|V(K_F^*)|}\cdot\prod_{i=0}^{|V(K_F^*)|-1}(1-ia). \tag{123}$$

The result follows from (116).

Case 3: $|V(K_F^*)| = 1$. By Lemma 21,

$$P[(\bigcap_{\{i,j\}\in E(F)} W_{ij})\bigcap(\bigcap_{\{i,j\}\in E(K_F^*)} W_{ij}^c)] = a^{|E(F)|}. \tag{124}$$

By (120) and the case assumption,

$$a^{|E(F)|} = a^{|R|-|V(K_F^*)|}\cdot\prod_{i=0}^{|V(K_F^*)|-1}(1-ia). \tag{125}$$

The result follows from (116). ■

Theorem 3. Under H_0 for any $(n_1, n_2, ..., n_{|R|}) \in \operatorname{ran}\mathbf{N}$,

$$P[\mathbf{N} = (n_1, n_2, ..., n_{|R|})] = \frac{|R|!}{\prod_{i=1}^{|R|}(i!)^{n_i}\cdot n_i!}\cdot a^{|R|-|(n_1,n_2,...,n_{|R|})|_1}\cdot\prod_{i=0}^{|(n_1,n_2,...,n_{|R|})|_1-1}(1-ia),$$

where

$$|(n_1, n_2, ..., n_{|R|})|_1 = \sum_{i=1}^{|R|} n_i\,.$$

Proof: Given $(n_1, n_2, ..., n_{|R|}) \in \operatorname{ran} \mathbf{N}$, assume H_0. Because partitions of R are mutually exclusive,

$$P[\mathbf{N} = (n_1, n_2, ..., n_{|R|})] = P[\psi \in \mathbf{N}^{-1}(n_1, n_2, ..., n_{|R|})] = \sum_{\psi \in \mathbf{N}^{-1}(n_1, n_2, ..., n_{|R|})} P\{\psi\}. \qquad (126)$$

By definition, for any $\psi \in S$, $\psi \in \mathbf{N}^{-1}(n_1, n_2, ..., n_{|R|})$ if and only if ψ contains exactly n_i blocks of size i, $i = 1, 2, ..., |R|$. Thus,

$$\left|\mathbf{N}^{-1}(n_1, n_2, ..., n_{|R|})\right| = \frac{|R|!}{\prod_{i=1}^{|R|} (i!)^{n_i} \cdot n_i!}. \qquad (127)$$

[$\frac{|R|!}{\prod_{i=1}^{|R|} (i!)^{n_i} \cdot n_i!}$ gives the number of unique partitions of $|R|$ objects into n_i groups of size $i \in \{1, 2, ..., |R|\}$ (Bogart 1983, p. 46)]. Moreover, for any $\psi \in S$, if $\psi \in \mathbf{N}^{-1}(n_1, n_2, ..., n_{|R|})$, then ψ has exactly $\left|(n_1, n_2, ..., n_{|R|})\right|_1$ units. Thus, by Theorem 3, for any $\psi \in \mathbf{N}^{-1}(n_1, n_2, ..., n_{|R|})$,

$$P\{\psi\} = a^{|R| - \left|(n_1, n_2, ..., n_{|R|})\right|_1} \cdot \prod_{i=0}^{\left|(n_1, n_2, ..., n_{|R|})\right|_1 - 1} (1 - ia). \qquad (128)$$

Therefore, (126) and (127) imply that

$$P[\mathbf{N} = (n_1, n_2, ..., n_{|R|})] = \frac{|R|!}{\prod_{i=1}^{|R|} (i!)^{n_i} \cdot n_i!} \cdot a^{|R| - \left|(n_1, n_2, ..., n_{|R|})\right|_1} \cdot \prod_{i=0}^{\left|(n_1, n_2, ..., n_{|R|})\right|_1 - 1} (1 - ia). \blacksquare$$

Theorem 4. Under H_0 for any $(n_1, n_2, ..., n_{|R|}) \in \operatorname{ran} \mathbf{N}$, if $\left|(n_1, n_2, ..., n_{|R|})\right|_1 = k$ and $P(U = k) > 0$, then,

$$P[\mathbf{N}=(n_1,n_2,...,n_{|R|})\,|\,U=k]=\frac{|R|!}{[\prod_{i=1}^{|R|}(i!)^{n_i}\cdot n_i!]\cdot S(|R|,k)},$$

where $S(|R|,k)$ is a Stirling number of the second kind, given by

$$\frac{1}{k!}\cdot\sum_{i=1}^{k}(-1)^i\cdot\binom{k}{i}\cdot(k-i)^{|R|}.$$

Proof: Given $(n_1,n_2,...,n_{|R|})\in \text{ran}\,\mathbf{N}$, assume H_0, that $\left|(n_1,n_2,...,n_{|R|})\right|_1=k$, and that $P(U=k)>0$. By the definition of conditional probability,

$$P[\mathbf{N}=(n_1,n_2,...,n_{|R|})\,|\,U=k]=\frac{P[\mathbf{N}=(n_1,n_2,...,n_{|R|})\bigcap U=k]}{P(U=k)}. \tag{129}$$

Looking at the numerator of (129), because partitions of R are mutually exclusive,

$$P[\mathbf{N}=(n_1,n_2,...,n_{|R|})\bigcap U=k]=P[\psi\in\mathbf{N}^{-1}(n_1,n_2,...,n_{|R|})\bigcap U^{-1}(k)]$$

$$=\sum_{\psi\in\mathbf{N}^{-1}(n_1,n_2,...,n_{|R|})\bigcap U^{-1}(k)}P\{\psi\}. \tag{130}$$

By definition, for any $\psi\in S$, $\psi\in\mathbf{N}^{-1}(n_1,n_2,...,n_{|R|})$ if and only if ψ contains exactly n_i blocks of size i, $i=1,2,...,|R|$. Also, $\psi\in U^{-1}(k)$ if and only if ψ has exactly k units. Thus, because $\left|(n_1,n_2,...,n_{|R|})\right|_1=k$, for any $\psi\in S$, if $\psi\in\mathbf{N}^{-1}(n_1,n_2,...,n_{|R|})$, then $\psi\in U^{-1}(k)$. It follows that

$$\mathbf{N}^{-1}(n_1,n_2,...,n_{|R|})\subseteq U^{-1}(k), \tag{131}$$

and that

$$\mathbf{N}^{-1}(n_1,n_2,...,n_{|R|})\bigcap U^{-1}(k)=\mathbf{N}^{-1}(n_1,n_2,...,n_{|R|}). \tag{132}$$

Hence, by (130), Theorem 3 implies that

$$P[\mathbf{N}=(n_1,n_2,...,n_{|R|})\bigcap U=k]=\frac{|R|!}{\prod_{i=1}^{|R|}(i!)^{n_i}\cdot n_i!}\cdot a^{|R|-k}\cdot\prod_{i=0}^{k-1}(1-ia)\,. \tag{133}$$

Turning now to the denominator of (129), again because partitions of *R* are mutually exclusive,

$$P(U=k)=P[\psi\in U^{-1}(k)]=\sum_{\psi\in U^{-1}(k)}P\{\psi\}\,. \tag{134}$$

$\left|U^{-1}(k)\right|$ will be equal to the number of partitions of $|R|$ objects into *k* groups – given by a Stirling number of the second kind (Bogart 1983,, p. 48):

$$S(|R|,k)=\frac{1}{k!}\cdot\sum_{i=1}^{k}(-1)^i\cdot\binom{k}{i}\cdot(k-i)^{|R|}\,. \tag{135}$$

Moreover, every $\psi\in U^{-1}(k)$ must by definition have *k* units, so by Theorem 2,

$$P\{\psi\}=a^{|R|-k}\cdot\prod_{i=0}^{k-1}(1-ia)\,. \tag{136}$$

Thus, by (134),

$$P(U=k)=S(|R|,k)\cdot a^{|R|-k}\cdot\prod_{i=0}^{k-1}(1-ia)\,. \tag{137}$$

As by assumption $P(U=k)>0$, (137) implies that

$$a^{|R|-k}\cdot\prod_{i=0}^{k-1}(1-ia)>0\,. \tag{138}$$

Therefore, (129), (133), and (137) imply that

$$P[\mathbf{N}=(n_1,n_2,...,n_{|R|})\,|\,U=k]=\frac{\dfrac{|R|!}{\prod_{i=1}^{|R|}(i!)^{n_i}\cdot n_i!}\cdot a^{|R|-k}\cdot\prod_{i=0}^{k-1}(1-ia)}{S(|R|,k)\cdot a^{|R|-k}\cdot\prod_{i=0}^{k-1}(1-ia)}$$

$$= \frac{|R|!}{[\prod_{i=1}^{|R|} (i!)^{n_i} \cdot n_i!] \cdot S(|R|,k)}. \qquad \blacksquare$$

IMPLICATIONS OF THE DISTRIBUTION OF PARTITIONS

Lemma 22. For any $\omega \in \Omega$ with $|\omega| \geq 2$; partition ψ_ω of ω; $r_1, r_2 \in \omega$; and $\psi \in S$: if $\psi \in \{\psi_\omega\}$, then $(r_1, r_2) \in \text{Eq}_{\psi_\omega}$ if and only if $(r_1, r_2) \in \text{Eq}_\psi$.

Proof: Given any $\omega \in \Omega$ with $|\omega| \geq 2$; partition ψ_ω of ω; $r_1, r_2 \in \omega$; and $\psi \in S$, assume that $\psi \in \{\psi_\omega\}$. Because $r_1, r_2 \in \omega$, by the definition of $\{\psi_\omega\}$, $(r_1, r_2) \in \text{Eq}_{\psi_\omega}$ if and only if $(r_1, r_2) \in \text{Eq}_\psi$. Thus, the result follows. $\blacksquare$

Lemma 23. For any $\omega \in \Omega$ with $|\omega| \geq 2$; $X, Y \subseteq E(K_\omega)$; partition ψ_ω of ω; and $\psi, \psi' \in S$: if $\psi, \psi' \in \{\psi_\omega\}$ and $\psi \in (\bigcap_{\{i,j\} \in X} W_{ij}) \bigcap (\bigcap_{\{i,j\} \in Y} W_{ij}^c)$, then $\psi' \in (\bigcap_{\{i,j\} \in X} W_{ij}) \bigcap (\bigcap_{\{i,j\} \in Y} W_{ij}^c)$.

Proof: Given $\omega \in \Omega$ with $|\omega| \geq 2$; $X, Y \subseteq E(K_\omega)$; partition ψ_ω of ω; and $\psi, \psi' \in S$: assume that $\psi, \psi' \in \{\psi_\omega\}$ and $\psi \in (\bigcap_{\{i,j\} \in X} W_{ij}) \bigcap (\bigcap_{\{i,j\} \in Y} W_{ij}^c)$. The proof will follow by a case analysis on whether $X = \varnothing$ and $Y = \varnothing$.

Case 1: $X \neq \varnothing$ and $Y \neq \varnothing$. Given $\{r_1, r_2\} \in X$ and $\{r_3, r_4\} \in Y$, because $\psi \in (\bigcap_{\{i,j\} \in X} W_{ij}) \bigcap (\bigcap_{\{i,j\} \in Y} W_{ij}^c)$, $(r_1, r_2) \in \text{Eq}_\psi$ and $(r_3, r_4) \notin \text{Eq}_\psi$. Also, because $X, Y \subseteq E(K_\omega)$, $r_1, r_2, r_3, r_4 \in \omega$. Hence, as $\psi \in \{\psi_\omega\}$, by Lemma 22, $(r_1, r_2) \in \text{Eq}_{\psi_\omega}$ and

$(r_3, r_4) \notin \mathrm{Eq}_{\psi_\omega}$. Moreover, again by Lemma 22, because $\psi' \in \{\psi_\omega\}$, $(r_1, r_2) \in \mathrm{Eq}_{\psi'}$ and $(r_3, r_4) \notin \mathrm{Eq}_{\psi'}$. It follows that, $\psi' \in \bigcap_{\{i,j\} \in X} W_{ij}$ and $\psi' \in \bigcap_{\{i,j\} \in Y} W_{ij}^c$.

Case 2: $X \neq \varnothing$ and $Y = \varnothing$. Given $\{r_1, r_2\} \in X$, by the arguments in Case 1, it follows that $\psi' \in \bigcap_{\{i,j\} \in X} W_{ij}$. Hence, $\psi' \in (\bigcap_{\{i,j\} \in X} W_{ij}) \bigcap (\bigcap_{\{i,j\} \in Y} W_{ij}^c)$.

Case 3: $X = \varnothing$ and $Y \neq \varnothing$. Given $\{r_3, r_4\} \in Y$, by the arguments given in case 1, $\psi' \in \bigcap_{\{i,j\} \in Y} W_{ij}^c$. Hence, the result follows.

Case 4: $X = \varnothing$ and $Y = \varnothing$. The result follows immediately. ■

Lemma 24. For any $\omega_1, \omega_2 \in \Omega$ with $\omega_1 \subseteq \omega_2$ and $|\omega_1| \geq 2$; partitions ψ_{ω_1} and ψ_{ω_2} of ω_1 and ω_2, respectively; and $\psi, \psi' \in S$: if $\psi \in \{\psi_{\omega_1}\}$ and $\psi, \psi' \in \{\psi_{\omega_2}\}$, then $\psi' \in \{\psi_{\omega_1}\}$.

Proof: Given $\omega_1, \omega_2 \in \Omega$ with $\omega_1 \subseteq \omega_2$ and $|\omega_1| \geq 2$; partitions ψ_{ω_1} and ψ_{ω_2} of ω_1 and ω_2, respectively; and $\psi, \psi' \in S$; assume that $\psi \in \{\psi_{\omega_1}\}$ and $\psi, \psi' \in \{\psi_{\omega_2}\}$.

Given distinct $r_1, r_2 \in \omega_1$, assume that $(r_1, r_2) \in \mathrm{Eq}_{\psi_{\omega_1}}$. Because $\psi \in \{\psi_{\omega_1}\}$, it follows from Lemma 22 that $(r_1, r_2) \in \mathrm{Eq}_{\psi}$. Moreover, $\omega_1 \subseteq \omega_2$ implies that $r_1, r_2 \in \omega_2$. Thus, because $\psi \in \{\psi_{\omega_2}\}$, by Lemma 22, $(r_1, r_2) \in \mathrm{Eq}_{\psi_{\omega_2}}$. Applying Lemma 22 again, because $\psi' \in \{\psi_{\omega_2}\}$, $(r_1, r_2) \in \mathrm{Eq}_{\psi'}$. Thus, for any $r_1, r_2 \in \omega_1$, if $(r_1, r_2) \in \mathrm{Eq}_{\psi_{\omega_1}}$, then $(r_1, r_2) \in \mathrm{Eq}_{\psi'}$.

Now assume that $(r_1, r_2) \in \mathrm{Eq}_{\psi'}$. As above, $r_1, r_2 \in \omega_2$. Because $\psi' \in \{\psi_{\omega_2}\}$, by Lemma 22, $(r_1, r_2) \in \mathrm{Eq}_{\psi_{\omega_2}}$. Moreover, because $\psi \in \{\psi_{\omega_2}\}$, by Lemma 22, $(r_1, r_2) \in \mathrm{Eq}_{\psi}$. Finally, because $\psi \in \{\psi_{\omega_1}\}$, by Lemma 22, $(r_1, r_2) \in \mathrm{Eq}_{\psi_{\omega_1}}$. Thus, for any $r_1, r_2 \in \omega_1$, if $(r_1, r_2) \in \mathrm{Eq}_{\psi'}$, then $(r_1, r_2) \in \mathrm{Eq}_{\psi_{\omega_1}}$.

Therefore, for any $r_1, r_2 \in \omega_1$, $(r_1, r_2) \in \mathrm{Eq}_{\psi'}$ if and only if $(r_1, r_2) \in \mathrm{Eq}_{\psi_{\omega_1}}$. The result follows directly from the definition of $\{\psi_{\omega_1}\}$. ■

Lemma 25. For any $\omega_1, \omega_2 \in \Omega$ with $|\omega_1| \geq 2$ and $|\omega_2| - |\omega_1| = 1, r_1 = \omega_2 - \omega_1$, partition ψ_{ω_1} of ω_1, and $Z = \{\psi_{\omega_2} : \psi_{\omega_2}$ a partition of ω_2, $\{\psi_{\omega_2}\} \subseteq \{\psi_{\omega_1}\} \bigcap (\bigcup_{i \in \omega_1} W_{r_1 i})\}$,

$$\bigcup_{i \in \omega_1} (\{\psi_{\omega_1}\} \bigcap W_{r_1 i}) = \bigcup_{\psi_{\omega_2} \subseteq Z} \{\psi_{\omega_2}\}.$$

Proof: Given $\omega_1, \omega_2 \in \Omega$ with $|\omega_1| \geq 2$ and $|\omega_2| - |\omega_1| = 1, r_1 = \omega_2 - \omega_1$, partition ψ_{ω_1} of ω_1, and $Z = \{\psi_{\omega_2} : \{\psi_{\omega_2}\} \subseteq \{\psi_{\omega_1}\} \bigcap (\bigcup_{i \in \omega_1} W_{r_1 i})\}$, the proof will follow by showing that for any $\psi \in S$, $\psi \in \bigcup_{i \in \omega_1} (\{\psi_{\omega_1}\} \bigcap W_{r_1 i})$ if and only if $\psi \in \bigcup_{\psi_{\omega_2} \subseteq Z} \{\psi_{\omega_2}\}$. Given $\psi \in S$, assume first that $\psi \in \bigcup_{i \in \omega_1} (\{\psi_{\omega_1}\} \bigcap W_{r_1 i})$. Thus, $\exists\, r_2 \in \omega_1$ so that $\psi \in W_{r_1 r_2}$. Fix r_2. Also, by definition, $\exists$ a partition $\psi^*_{\omega_2}$ of ω_2 so that $\psi \in \{\psi^*_{\omega_2}\}$. Fix $\psi^*_{\omega_2}$. Because $\psi \in W_{r_1 r_2}$, $\psi \in \{\psi^*_{\omega_2}\}$, $r_1 \in \omega_2$, and $r_2 \in \omega_2$, Lemma 23 (with $X = \{\{r_1, r_2\}\}$ and $Y = \varnothing$) implies that for any $\psi' \in S$, if $\psi' \in \{\psi^*_{\omega_2}\}$, then $\psi' \in W_{r_1 r_2}$. It follows that

$$\{\psi^*_{\omega_2}\} \subseteq W_{r_1 r_2}. \tag{139}$$

Moreover, because $\psi \in \bigcup_{i \in \omega_1} (\{\psi_{\omega_1}\} \bigcap W_{r_1 i})$, $\psi \in \{\psi_{\omega_1}\}$. As $\psi \in \{\psi^*_{\omega_2}\}$ and $\omega_1 \subseteq \omega_2$, by Lemma 24, for any $\psi' \in S$, if $\psi' \in \{\psi^*_{\omega_2}\}$, then $\psi' \in \{\psi_{\omega_1}\}$. Thus,

$$\{\psi^*_{\omega_2}\} \subseteq \{\psi_{\omega_1}\}. \tag{140}$$

(139) and (140) imply that $\{\psi^*_{\omega_2}\} \subseteq \{\psi_{\omega_1}\} \bigcap (\bigcup_{i \in \omega_1} W_{r_1 i})$, and hence that $\psi^*_{\omega_2} \subseteq Z$.

Because by definition $\psi \in \{\psi^*_{\omega_2}\}$, $\psi \in \bigcup_{\psi_{\omega_2} \subseteq Z} \{\psi_{\omega_2}\}$.

Now assume that $\psi \in \bigcup_{\psi_{\omega_2} \subseteq Z} \{\psi_{\omega_2}\}$. Thus, $\exists$ a partition $\psi^*_{\omega_2}$ of ω_2 so that $\psi \in \{\psi^*_{\omega_2}\}$ and $\{\psi^*_{\omega_2}\} \subseteq \{\psi_{\omega_1}\} \bigcap (\bigcup_{i \in \omega_1} W_{r_1 i})$. This implies that $\psi \in \{\psi_{\omega_1}\}$ and $\psi \in \bigcup_{i \in \omega_1} W_{r_1 i}$, so $\psi \in \bigcup_{i \in \omega_1} (\{\psi_{\omega_1}\} \bigcap W_{r_1 i})$. ■

Lemma 26. For any $\omega_1, \omega_2 \in \Omega$ with $|\omega_1| \geq 2$ and $|\omega_2| - |\omega_1| = 1, r_1 = \omega_2 - \omega_1$, partition ψ_{ω_1} of ω_1, and $Z = \{\psi_{\omega_2} : \psi_{\omega_2}$ a partition of ω_2, $\{\psi_{\omega_2}\} \subseteq \{\psi_{\omega_1}\} \bigcap (\bigcap_{i \in \omega_1} W^c_{r_1 i})\}$

$$\{\psi_{\omega_1}\} \bigcap (\bigcap_{i \in \omega_1} W^c_{r_1 i}) = \bigcup_{\psi_{\omega_2} \subseteq Z} \{\psi_{\omega_2}\}.$$

Proof: Given $\omega_1, \omega_2 \in \Omega$ with $|\omega_1| \geq 2$ and $|\omega_2| - |\omega_1| = 1, r_1 = \omega_2 - \omega_1$, partition ψ_{ω_1} of ω_1, and $Z = \{\psi_{\omega_2} : \{\psi_{\omega_2}\} \subseteq \{\psi_{\omega_1}\} \bigcap (\bigcap_{i \in \omega_1} W^c_{r_1 i})\}$, the proof will follow similarly to the proof of Lemma 25. Given $\psi \in S$, assume first that $\psi \in \{\psi_{\omega_1}\} \bigcap (\bigcap_{i \in \omega_1} W^c_{r_1 i})$. Thus, $\psi \in \bigcap_{i \in \omega_1} W^c_{r_1 i}$. Also clearly, $\exists$ a partition $\psi^*_{\omega_2}$ of ω_2 so that $\psi \in \{\psi^*_{\omega_2}\}$. Fix $\psi^*_{\omega_2}$. Therefore, by Lemma 23, for any $\psi' \in S$, if $\psi' \in \{\psi^*_{\omega_2}\}$, then $\psi' \in \bigcap_{i \in \omega_1} W^c_{r_1 i}$. Hence,

$$\{\psi^*_{\omega_2}\} \subseteq \bigcap_{i \in \omega_1} W^c_{r_1 i}. \tag{141}$$

Moreover, by definition, $\psi \in \{\psi^*_{\omega_2}\}$, and $\omega_1 \subseteq \omega_2$. Hence, as in Lemma 25,

$$\{\psi^*_{\omega_2}\} \subseteq \{\psi_{\omega_1}\}. \tag{142}$$

(141) and (142) imply that $\{\psi^*_{\omega_2}\} \subseteq \{\psi_{\omega_1}\} \bigcap (\bigcap_{i \in \omega_1} W^c_{r_1 i})$, and hence that $\psi^*_{\omega_2} \subseteq Z$.

Because by definition $\psi \in \{\psi^*_{\omega_2}\}$, $\psi \in \bigcup_{\psi_{\omega_2} \subseteq Z} \{\psi_{\omega_2}\}$.

Now assume that $\psi \in \bigcup_{\psi_{\omega_2} \subseteq Z} \{\psi_{\omega_2}\}$. Thus, $\exists$ a partition $\psi^*_{\omega_2}$ of ω_2 so that $\psi \in \{\psi^*_{\omega_2}\}$ and $\{\psi^*_{\omega_2}\} \subseteq \{\psi_{\omega_1}\} \bigcap (\bigcap_{i \in \omega_1} W^c_{r_1 i})$. This implies that $\psi \in \{\psi_{\omega_1}\}$ and $\psi \in \bigcap_{i \in \omega_1} W^c_{r_1 i}$, so $\psi \in \{\psi_{\omega_1}\} \bigcap (\bigcap_{i \in \omega_1} W^c_{r_1 i})$. ■

Lemma 27. If $a \in (0, \frac{1}{|R|-1}) \cup \{\frac{1}{|R|-1}, \frac{1}{|R|-2}, \ldots, \frac{1}{2}, 1\}$ and for all $\psi \in S$, $P\{\psi\} = a^{|R|-|\psi|} \cdot \prod_{i=0}^{|\psi|-1} (1 - ia)$, then for any $\omega \in \Omega$ with $|\omega| \geq 2$ and partition ψ_ω of ω,

$$P\{\psi_\omega\} = a^{|\omega|-|\psi_\omega|} \cdot \prod_{i=0}^{|\psi_\omega|-1} (1 - ia).$$

Proof: Given $\psi \in S$, $\omega \in \Omega$ with $|\omega| \geq 2$, and a partition ψ_ω of ω, assume that $a \in (0, \frac{1}{|R|-1}) \cup \{\frac{1}{|R|-1}, \frac{1}{|R|-2}, \ldots, \frac{1}{2}, 1\}$ and that $P\{\psi\} = a^{|R|-|\psi|} \cdot \prod_{i=0}^{|\psi|-1} (1 - ia)$. The proof will follow by induction on $|R| - |\omega|$. Assume that $|R| - |\omega| = 0$. Thus, $\omega = R$, and the result follows.

Now assume that for some integer $0 \leq t < |R| - 2$, if $|R| - |\omega| \leq t$, then the result holds. Assume that $|R| - |\omega| = t + 1$. Fix $r_1 \in R - \omega$. For any partition $\psi \in S$, either $\exists\ r_2 \in \omega$ so that $(r_1, r_2) \in \text{Eq}_\psi$, or $\neg\exists$ such an r_2. Therefore,

$$(\bigcup_{i\in\omega} W_{r_1 i}) \bigcup (\bigcap_{i\in\omega} W^c_{r_1 i}) = S, \tag{143}$$

implying that

$$\{\psi_\omega\} = [\bigcup_{i\in\omega}(\{\psi_\omega\}\bigcap W_{r_1 i})] \bigcup [\{\psi_\omega\}\bigcap_{i\in\omega} W^c_{r_1 i}]. \tag{144}$$

Looking at the first term on the right side of (144), let $\omega' = \omega \cup r_1$. By Lemma 25,

$$\bigcup_{i\in\omega}(\{\psi_\omega\}\bigcap W_{r_1 i}) = \bigcup_{\psi_{\omega'}\subseteq Z}\{\psi_{\omega'}\}, \tag{145}$$

where $Z \equiv \{\psi_{\omega'} : \psi_{\omega'}$ a partition of ω', $\{\psi_{\omega'}\} \subseteq \{\psi_\omega\}\bigcap(\bigcup_{i\in\omega} W_{r_1 i})\}$. Clearly, Z contains $|\psi_\omega|$ unique values of $\psi_{\omega'}$, each with $|\psi_\omega|$ units. Moreover, $R - \omega' = t$, so by the induction hypothesis and fact that partitions of ω' are mutually exclusive,

$$P(\bigcup_{i\in\omega}(\{\psi_\omega\}\bigcap W_{r_1 i})) = \sum_{\psi_{\omega'}\subseteq Z} P\{\psi_{\omega'}\} = |\psi_\omega| \cdot a^{|R|-|\psi_\omega|} \cdot \prod_{i=0}^{|\psi_\omega|-1}(1-ia). \tag{146}$$

Turning to the last term of (144), by Lemma 26,

$$\{\psi_{\omega_1}\}\bigcap(\bigcap_{i\in\omega_1} W^c_{r_1 i}) = \bigcup_{\psi_{\omega'}\subseteq Z'}\{\psi_{\omega'}\}, \tag{147}$$

where $Z' \equiv \{\psi_{\omega'} : \psi_{\omega'}$ a partition of ω', $\{\psi_{\omega'}\} \subseteq \{\psi_{\omega_1}\}\bigcap(\bigcap_{i\in\omega_1} W^c_{r_1 i})\}$. Z' contains exactly one element; the partition of ω' wherein the subunits of ω are partitioned as per ψ_ω, and r_1 is in its own unit. This partition contains $|\psi_\omega| + 1$ units. Moreover, again because $R - \omega' = t$, by the induction hypothesis,

$$P(\{\psi_\omega\}\bigcap_{i\in\omega} W^c_{r_1 i}) = \sum_{\psi_{\omega'}\subseteq Z'} P\{\psi_{\omega'}\} = a^{|R|-(|\psi_\omega|+1)} \cdot \prod_{i=0}^{(|\psi_\omega|+1)-1}(1-ia). \tag{148}$$

Putting (144), (146), and (148) together, and noting again that partitions of ω' are mutually exclusive,

$$P\{\psi_\omega\} = |\psi_\omega| \cdot a^{|R|-|\psi_\omega|} \cdot \prod_{i=0}^{|\psi_\omega|-1}(1-ia) + a^{|R|-(|\psi_\omega|+1)} \cdot \prod_{i=0}^{(|\psi_\omega|+1)-1}(1-ia) =$$

$$= a^{(|R|-1)-|\psi_\omega|} \cdot \prod_{i=0}^{|\psi_\omega|-1}(1-ia) \cdot [|\psi_\omega| \cdot a + (1-|\psi_\omega| \cdot a)] = a^{(|R|-1)-|\psi_\omega|} \cdot \prod_{i=0}^{|\psi_\omega|-1}(1-ia). \quad (149)$$

As $|\omega| = |R| - 1$, the proof is complete. ■

Theorem 5. If $a \in (0, \frac{1}{|R|-1}) \cup \{\frac{1}{|R|-1}, \frac{1}{|R|-2}, ..., \frac{1}{2}, 1\}$, then

$$\sum_{\psi \in S} a^{|R|-|\psi|} \cdot \prod_{i=0}^{|\psi|-1}(1-ia) = 1.$$

Proof: Assume that $a \in (0, \frac{1}{|R|-1}) \cup \{\frac{1}{|R|-1}, \frac{1}{|R|-2}, ..., \frac{1}{2}, 1\}$. Fix $r_1, r_2 \in R$ with $r_1 \neq r_2$, and $\omega \in \Omega$ with $\omega = \{r_1, r_2\}$. Given any $\psi \in S$, by definition $\exists$ a partition ψ^*_ω of ω so that $\psi \in \{\psi^*_\omega\}$. Thus, letting $Z = \{\psi_\omega : \psi_\omega$ a partition of $\omega\}$,

$$S = \bigcup_{\psi_\omega \in Z} \{\psi_\omega\}. \quad (150)$$

Moreover, by definition, $S = \bigcup_{\psi \in S} \{\psi\}$, so

$$\bigcup_{\psi \in S} \{\psi\} = \bigcup_{\psi_\omega \in Z} \{\psi_\omega\}. \quad (151)$$

Because partitions of R are mutually exclusive (as are partitions of ω), by the additivity axiom,

$$\sum_{\psi \in S} P\{\psi\} = \sum_{\psi_\omega \in Z} P\{\psi_\omega\}. \quad (152)$$

$\omega = \{r_1, r_2\}$ implies that Z contains two elements: $\{r_1, r_2\}$ and $\{r_1\}\{r_2\}$. By Lemma 27,

$$P\{\{r_1, r_2\}\} + P\{\{r_1\}\{r_2\}\} = a + (1-a), \quad (153)$$

implying that $\sum_{\psi \in S} P\{\psi\} = 1$. ■

Lemma 28. For any distinct $r_1, r_2 \in R$ and $X, Y \subseteq E(K_{R-r_1})$, if

$P[(\bigcap_{\{i,j\}\in X} W_{ij})\bigcap(\bigcap_{\{i,j\}\in Y} W_{ij}^c)] > 0$, $a \in (0, \frac{1}{|R|-1}) \cup \{\frac{1}{|R|-1}, \frac{1}{|R|-2}, \dots, \frac{1}{2}, 1\}$, and for all $\psi \in S$,

$$P\{\psi\} = a^{|R|-|\psi|} \cdot \prod_{i=0}^{|\psi|-1} (1 - ia),$$

then

$$P[W_{r_1 r_2} \mid (\bigcap_{\{i,j\}\in X} W_{ij})\bigcap(\bigcap_{\{i,j\}\in Y} W_{ij}^c)] = a.$$

Proof: Given distinct $r_1, r_2 \in R$ and $X, Y \subseteq E(K_{R-r_1})$, assume that

$P[(\bigcap_{\{i,j\}\in X} W_{ij})\bigcap(\bigcap_{\{i,j\}\in Y} W_{ij}^c)] > 0$, $a \in (0, \frac{1}{|R|-1}) \cup \{\frac{1}{|R|-1}, \frac{1}{|R|-2}, \dots, \frac{1}{2}, 1\}$, and that for all $\psi \in S$, $P\{\psi\} = a^{|R|-|\psi|} \cdot \prod_{i=0}^{|\psi|-1} (1 - ia)$. Fix $\omega = R - r_1$. It follows from Lemma 23 that for any partition ψ_ω of ω, if $\exists$ $\psi \in \{\psi_\omega\}$ so that $\psi \in (\bigcap_{\{i,j\}\in X} W_{ij})\bigcap(\bigcap_{\{i,j\}\in Y} W_{ij}^c)$, then for every $\psi' \in \{\psi_\omega\}$, $\psi' \in (\bigcap_{\{i,j\}\in X} W_{ij})\bigcap(\bigcap_{\{i,j\}\in Y} W_{ij}^c)$. Hence, $\exists$ a set of partitions of ω ("*Z*") so that

$$(\bigcap_{\{i,j\}\in X} W_{ij})\bigcap(\bigcap_{\{i,j\}\in Y} W_{ij}^c) = \bigcup_{\psi_\omega \in Z} \{\psi_\omega\}. \tag{154}$$

Fix *Z*. It follows that

$$W_{r_1 r_2} \bigcap (\bigcap_{\{i,j\}\in X} W_{ij})\bigcap(\bigcap_{\{i,j\}\in Y} W_{ij}^c) = W_{r_1 r_2} \bigcap (\bigcup_{\psi_\omega \in Z} \{\psi_\omega\}) = \bigcup_{\psi_\omega \in Z} (W_{r_1 r_2} \bigcap \{\psi_\omega\}). \tag{155}$$

Clearly, given any value of $\psi_\omega \in Z$, $\exists$ exactly one partition ψ of R so that $\psi \in W_{r_1 r_2} \bigcap \{\psi_\omega\}$: the partition wherein the elements of ω are partitioned as per ψ_ω, and r_1 is in the same unit as r_2. This partition has $|\psi_\omega|$ units, so by the assumed distribution of $\{\psi\}$,

$$P(W_{r_1 r_2} \bigcap \{\psi_\omega\}) = a^{|R|-|\psi_\omega|} \cdot \prod_{i=0}^{|\psi_\omega|-1} (1-ia). \qquad (156)$$

Moreover, by Lemma 27,

$$P\{\psi_\omega\} = a^{|\omega|-|\psi_\omega|} \cdot \prod_{i=0}^{|\psi_\omega|-1} (1-ia). \qquad (157)$$

Because $|R| = |\omega| + 1$, (156) implies that

$$P(W_{r_1 r_2} \bigcap \{\psi_\omega\}) = a \cdot P\{\psi_\omega\}. \qquad (158)$$

Hence, by the additivity axiom,

$$P[\bigcup_{\psi_\omega \in Z} (W_{r_1 r_2} \bigcap \{\psi_\omega\})] = \sum_{\psi_\omega \in Z} P(W_{r_1 r_2} \bigcap \{\psi_\omega\}) = a \cdot \sum_{\psi_\omega \in Z} P\{\psi_\omega\}.$$

and by (155),

$$P[W_{r_1 r_2} \bigcap (\bigcap_{\{i,j\} \in X} W_{ij}) \bigcap (\bigcap_{\{i,j\} \in Y} W_{ij}^c)] = a \cdot \sum_{\psi_\omega \in Z} P\{\psi_\omega\} \qquad (159)$$

Also, by (154) and the additivity axiom,

$$P[(\bigcap_{\{i,j\} \in X} W_{ij}) \bigcap (\bigcap_{\{i,j\} \in Y} W_{ij}^c)] = P(\bigcup_{\psi_\omega \in Z} \{\psi_\omega\}) = \sum_{\psi_\omega \in Z} P\{\psi_\omega\}. \qquad (160)$$

Because $P[(\bigcap_{\{i,j\} \in X} W_{ij}) \bigcap (\bigcap_{\{i,j\} \in Y} W_{ij}^c)] > 0$, it follows from (160) that

$$\sum_{\psi_\omega \in Z} P\{\psi_\omega\} > 0. \qquad (161)$$

Therefore, by the definition of conditional probability, (159) and (160) imply that

$$P[W_{r_1 r_2} \mid (\bigcap_{\{i,j\} \in X} W_{ij}) \bigcap (\bigcap_{\{i,j\} \in Y} W_{ij}^c)] = \frac{P[W_{r_1 r_2} \bigcap (\bigcap_{\{i,j\} \in X} W_{ij}) \bigcap (\bigcap_{\{i,j\} \in Y} W_{ij}^c)]}{P[(\bigcap_{\{i,j\} \in X} W_{ij}) \bigcap (\bigcap_{\{i,j\} \in Y} W_{ij}^c)]} = a. \ \blacksquare$$

Theorem 6. If $a \in (0, \frac{1}{|R|-1}) \cup \{\frac{1}{|R|-1}, \frac{1}{|R|-2}, ..., \frac{1}{2}, 1\}$ and for all $\psi \in S$,

$$P\{\psi\} = a^{|R|-|\psi|} \cdot \prod_{i=0}^{|\psi|-1} (1-ia),$$

then H_0 is true.

Proof: Given distinct $r_1, r_2, r_3 \in R$ and $X, Y \subseteq E(K_{R-r_1})$, assume the aforementioned distribution. Assume also that

$$P[W_{r_2 r_3} \bigcap (\bigcap_{\{i,j\} \in X} W_{ij}) \bigcap (\bigcap_{\{i,j\} \in Y} W_{ij}^c)] > 0 \tag{162}$$

and that

$$P[W_{r_2 r_3}^c \bigcap (\bigcap_{\{i,j\} \in X} W_{ij}) \bigcap (\bigcap_{\{i,j\} \in Y} W_{ij}^c)] > 0. \tag{163}$$

Hence, by Lemma 28,

$$P[W_{r_1 r_2} \mid W_{r_2 r_3} \bigcap (\bigcap_{\{i,j\} \in X} W_{ij}) \bigcap (\bigcap_{\{i,j\} \in Y} W_{ij}^c)] = a \tag{164}$$

and

$$P[W_{r_1 r_2} \mid W_{r_2 r_3}^c \bigcap (\bigcap_{\{i,j\} \in X} W_{ij}) \bigcap (\bigcap_{\{i,j\} \in Y} W_{ij}^c)] = a, \tag{165}$$

so

$$P[W_{r_1 r_2} \mid W_{r_2 r_3} \bigcap (\bigcap_{\{i,j\} \in X} W_{ij}) \bigcap (\bigcap_{\{i,j\} \in Y} W_{ij}^c)] = P[W_{r_1 r_2} \mid W_{r_2 r_3}^c \bigcap (\bigcap_{\{i,j\} \in X} W_{ij}) \bigcap (\bigcap_{\{i,j\} \in Y} W_{ij}^c)]. \tag{166}$$

Therefore, for any $r_1, r_2, r_3 \in R$ and $X, Y \subseteq E(K_{R-r_1})$, if (162) and (163) are hold, then (166) will be true. This is precisely H_0. ■

DISCUSSION

Theorems 3 and 4 suggest two tests of the null hypothesis of (12). Suppose that a set of sites $\mathscr{S}$ is sampled. For site $s \in \mathscr{S}$, let R_s denote the set of species observed at this site, *h* denote $|R_s|$, $\mathbf{n}_s$ denote the value of $\mathbf{N} = (n_1, ..., n_h)$ obtained from R_s, and $|\mathbf{n}_s|_1$ denote the $L^1 - \text{norm}$ of $\mathbf{n}_s$. Moreover, define $\mathscr{S}_{h,k} \equiv \{s : s \in \mathscr{S}, |R_s| = h, |\mathbf{n}_s|_1 = k\}$, and let $\mathbf{M}_h$ be a multinomial random vector giving the observed distribution of **N** conditional on exactly *h* subunits occurring. For the first test, considering data only from the sites having *h* subunits, the maximum likelihood estimator for a is found from Theorems 1 and 3 by maximizing

$$L^h(a) = \prod_{k=1}^{h} \prod_{s \in \mathscr{S}_{h,k}} [a^{h-k} \cdot \prod_{i=1}^{k-1} (1 - ia)] \tag{167}$$

over $a \in (0, \frac{1}{|R|-1}) \cup \{\frac{1}{|R|-1}, \frac{1}{|R|-2}, ..., \frac{1}{2}, 1\}$. Hence the distribution of $\mathbf{M}_h$ can be found under H_0, permitting goodness-of-fit testing using either an exact likelihood ratio test or a chi-square test with $\left| \bigcup_{1 \le k \le h} \mathscr{S}_{h,k} \right| - 2$ degrees of freedom. Importantly, the results from these tests will pertain only to sites having *h* subunits; if $\exists s \in \mathscr{S}$ so that $|R_s| \ne h$, the tests will need to be repeated for each distinct value of $|R_s|$. To obtain an overall *p*-value, an overall multinomial likelihood ratio statistic can be computed, and tested exactly.

The second test fixes not just the number of subunits per site, but also the number of units. Specifically, let $\mathbf{M}_{h,k}$ denote a multinomial random vector giving the observed distribution of **N** conditional on both *h* subunits and k units occurring.

The estimated expected value of $\mathbf{M}_{h,k}$ can be found directly from Theorem 4, and goodness-of-fit testing follows immediately. Because the distribution of Theorem 4 is parameter-free, there will be $\left|\mathscr{S}_{h,k}\right| - 1$ degrees of freedom for the chi-square test. As in the first test, multiple p-values will again result here if sites differ in their numbers of subunits (or units). As above, an overall likelihood ratio statistic can be computed and tested to obtain an overall result.

Except for a few cases (Ladau, unpublished results), the power of these tests remains unknown. It depends on the number of sites sampled, numbers of subunits and units at the sites, and the specific deviation from the null hypothesis. Moreover, both tests will be inconsistent against alternatives involving deviations only in the relative frequencies of partitions giving the same values of **N**. For example, in both tests with $|R| = 4$ and $k = 2$, the frequencies of $\{\{r_1, r_2\}\{r_3, r_4\}\}$, $\{\{r_1, r_3\}\{r_2, r_4\}\}$, and $\{\{r_1, r_4\}\{r_2, r_3\}\}$ are grouped in the event $\{\mathbf{N} = (0,2,0,0)\}$, so the tests will have no power against alternatives that entail deviations only in the relative frequencies of these partitions. That said, Theorems 5 and 6 show that the tests are consistent against alternatives that entail deviations in the relative frequencies of **N**.

Using just presence-absence data, can any universally consistent tests be created? The answer appears to be no. For example, when $|R| = 3$, suppose that $\psi^* \equiv \{\{r_1\}\{r_2, r_3\}\} \in S$. For any $\psi \in S$, letting $P(\psi)$ and $P_0(\psi)$ denote the actual and null probabilities of ψ respectively, for universal consistency, H_0 must be rejected whenever $P(\psi^*) \neq P_0(\psi^*)$. However, this condition can be met only if the sample allows $P(\psi^*)$ to be specified, which is impossible with presence-absence data, because no information is contained therein about the relative frequencies of $\{\{r_1\}\{r_2, r_3\}\}$, $\{\{r_2\}\{r_1, r_3\}\}$, and $\{\{r_3\}\{r_1, r_2\}\}$ in the event $\{\mathbf{N} = (1,1,0)\}$. Because subunits are identified by their order of arrival, what is needed is information on the arrival sequence of the subunits.

However, the tests proposed here are reasonable because they are based on how **N** behaves under H_0. In other words, the tests are grouped data tests. Grouping of observations into classes is a well-known approach in goodness-of-fit testing; for instance, it is commonly employed when checking agreement with the Poisson distribution and when testing normality by grouping observations into intervals.

It could be suggested that because of the consistency issue, the tests are more appropriate for testing a null hypothesis both necessary and sufficient to derive the distributions of **N** than for testing the null hypothesis of (12). This redefined null hypothesis could consist of a subset of the equalities specified by (12). However, such a redefined null hypothesis would also follow from an absence of competitive effects, so the tests would give the same biological conclusions.

Importantly, the tests require that for all $r_1, r_2 \in R$, $P(W_{r_1 r_2}) > 0$. In some testing situations, this assumption may be unjustifiable. However, for such cases, it appears possible to construct assumption-free tests by extending the results derived here (Ladau, unpublished results). The first test also requires that sites having exactly the same number of subunits give independent values of **N**. Likewise, the second test requires independence between sites having common numbers of subunits and units. Thus, the tests do have an assumption of independent sampling. However, a "sample assumption" of this kind is unavoidable; it differs qualitatively from the "population assumptions" that must be overcome for robustness (Bradley 1968). Moreover, on both ecological and statistical grounds, it is usually justifiable to treat sites as being independent.

In ecological terms, what conclusions do the tests allow? Rejecting the null hypothesis in both tests implies the existence of either negative *or* positive conditional non-independence. Hence, rejection does not imply competitive structuring, because such structuring predicts only negative non-independence. Nonetheless, when power is sufficient, failure to reject the null hypothesis does indicate a lack of competitive structuring, because by definition, non-independence must occur whenever competitive structuring occurs. In some communities, analysis of large data sets does not lead to rejection of the null hypothesis (Chapter IV).

A one-sided test, in which rejecting the null hypothesis implies negative non-independence, would allow stronger conclusions about competitive structuring. Developing such a test will be the focus of further work.

ACKNOWLEDGMENTS

This work was funded by a National Science Foundation Graduate Research Fellowship. It benefited greatly from the advice of R. Knapp and H. K. Reeve.

LITERATURE CITED

Begon, M. L., L. Harper, and C. R. Townsend. 1996. Ecology: Third Edition, Blackwell Science, Oxford.

Bogart, K. P. 1983. Introductory Combinatorics. Pitman Publishing Inc., London.

Bradley, J. V. 1968. Distribution-Free Statistical Tests. Prentice-Hall, Englewood Cliffs, New Jersey.

Connell, J. H. 1961. The influence of interspecific competition and other factors on the distribution of the barnacle *Chthamalus stellatus*. Ecology **42:**710-723.

Connell, J. H. 1980. Diversity and the coevolution of competitors, or the ghost of competition past. Oikos **35:**131-138.

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

Connor, E. F. and D. Simberloff. 1983. Interspecific competition and species co-occurrence patterns on islands: null models and the evaluation of evidence. Oikos **41:**455-465.

Diamond, J. M. 1975. Assembly of species communities. Pages 342-344 *in* M. L. Cody and J. M. Diamond, eds. Ecology and Evolution of Communities. Harvard University Press, Cambridge.

Fox, B. J. 1987. Species assembly and the evolution of community structure. Evolutionary Ecology **1:**201-213.

Fox, B. J. 1989. Small-mammal community pattern in Australian heathland: a taxonomically-based rule for species assembly. Pages 91-103 *in* D. W. Morris, Z. Abramsky, B. J. Fox, and M. Willig, eds. Patterns in the structure of mammalian communities. Special Publications of the Museum, Texas Tech. University, no. 28, Lubbock.

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 DC.

Hairston, N. G. 1980. The exponential test of an analysis of field distributions: competition in terrestrial salamanders. Ecology **61:**817-826.

Hrbacek, K. and T. Jech. 1999. Introduction to Set Theory: 3rd Edition (Monographs and textbooks in pure and applied mathematics 220). Marcel Dekker, Inc., New York.

Root, R. B. 2001. Guilds. Pages 295-302 *in* P. Kareiva and S. Levin, eds. Are there expendable species? Princeton University Press, Princeton.

Simberloff, D. S. 1970. Taxonomic diversity of island biotas. Evolution **24:**23-47.

Sokal, R. R. and E. J. Rohlf. 1995. Biometry: Third Edition. W. H. Freeman and Company, New York.

Tansley, A. G. 1917. On competition between *Galium sylvestre* Poll. (*G. aspermum* Schreb.) on different types of soil. Journal of Ecology **5:**173-179.

Wilson, J. B. 1989. A null model of guild proportionality, applied to stratification of a New Zealand temperate rain forest. Oecologia **80:**263-267.

— CHAPTER IV —

NULL MODEL TESTS OF SPECIES CO-OCCURRENCE: APPLICATION OF ROBUST METHODS

Abstract An important question in ecology is to what extent competitive interactions affect the composition of communities. Null model testing has been used extensively to address this question, but its conclusions have been unreliable. Here, using a new robust null model test, I analyze seven large data sets. I show that this test gives consistent results in the absence of major biological fluctuations. For all of the data sets analyzed, the test does not indicate effects of competitive interactions. Although simplistic, the null model accounts remarkably well for much of the variability in community composition (median $R^2 > 0.99$). Hence, in addition to being useful for testing for the effects of competitive interactions, the model appears to be a step towards providing a simple, general theory of community assembly.

INTRODUCTION

What influence do interspecific interactions have on the composition of ecological communities? Gleason (1926) proposed that the influence of interactions is minimal, with stochastic processes and the abiotic environment being the prime determinants of community composition. By contrast, Diamond (1975), McNab (1971), and others proposed that interspecific interactions – particularly competition – play a much larger role. The matter remains unresolved, but bears broadly on basic and applied ecology.

The most widely applicable approach for assessing the effects of competition has been null model testing. The applicability of null model testing derives from the data that are employed: lists of species occurring at sets of sites. These data are broadly available and easily collectible. Null model testing begins with the premise that competition does not affect the composition of communities. This premise is translated into a quantitative statement, or “null hypothesis,” which, together with quantitative assumptions, comprise the “null model.” The model is then used to generate a distribution of a statistic summarizing the lists of observed species. If the observed statistic (or one more extreme) is sufficiently unlikely to have come from the null model’s distribution, then the null hypothesis is rejected, and an effect of competition is inferred. Null model testing has been applied to numerous communities, for instance those of desert rodents (Fox and Brown 1993, Stone et al 1996), island birds (Connor and Simberloff 1979, Gotelli and Abele 1982, Feeley 2003), and fish parasites (Gotelli and Rohde 2002). It has also been applied to understand the effects of exotic species introductions (Sanders et al 2003). The findings of these applications have been mixed.

Unfortunately, the null model tests applied in these investigations suffer from two critical flaws. Like all statistical tests, null model tests should be powerful and robust. Power refers to the test’s sensitivity, with the most powerful tests rejecting false null hypotheses most frequently. Robustness refers to the distortion of significance levels when assumptions are violated, with robust tests only negligibly affected by violations (Bradley 1968, Sokal and Rohlf 1995). Thus, null model tests should be powerful and, if assumptions cannot be independently verified, robust. Existing null model tests lack adequate power or robustness (Chapter II).

It is the issue of robustness that is particularly troubling. It means that when assumptions cannot be verified – as is usually the case – tests will incorrectly indicate competitive effects unacceptably often. Thus, existing tests, and any inferences from them are unreliable (Chapter II). However, to resolve the problem, a test free of parametric assumptions has been recently proposed (Chapter III).

The proposed test is based on two premises: First, regardless of how competition acts – evolutionarily or ecologically, and extrinsically or intrinsically – it will reduce the co-occurrence of ecologically similar organisms. Second, ecological similarity can be specified using a hierarchical classification system. "Subunits" of organisms can be classified into "units," with subunits in the same unit being more ecologically similar to each other than those in different units. Hence, defining W_{ij} as the event that the *i*th and *j*th subunit to arrive at a community belong to the same unit, it follows that competitive effects predict that $P(W_{13} \mid W_{12}) < P(W_{13} \mid W_{12}^c)$, $P(W_{23} \mid W_{12}) < P(W_{23} \mid W_{12}^c)$, or $P(W_{34} \mid W_{23} \bigcap W_{13}^c) < P(W_{34} \mid W_{23}^c \bigcap W_{13}^c)$ for instance. Therefore, an appropriate null hypothesis for testing for the absence of competitive effects specifies equality between all relevant pairs of conditional probabilities. This null hypothesis is sufficient to specify a distribution of the partition of subunits into units (Chapter III), implying that to test robustly for the absence of competition, it is sufficient to check whether the observed and predicted distributions of partitions match.

Here, I apply a test a test similar to the parametric assumption-free test, which for simplicity assumes that for all *i* and *j*, $P(W_{ij}) > 0$ (Chapter III; hereafter referred to as "Ladau and Schwager's test"). Examining two plant and five animal communities, I first assess the overall frequency with which the null hypothesis is rejected. I then investigate whether rejections of the null hypothesis were due to (i) random fluctuations, unreflective of biological processes, or (ii) effects of interspecific competition, and I compare results to those of a parametric alternative. Last, I use the null model to assess the relative importance of interspecific interactions in composing communities.

MATERIALS AND METHODS

Data Sets

The data that Ladau and Schwager's test employ consist of lists of species present at sets of sites. I obtained these lists from seven sources.

North American Breeding Bird Survey (BBS)

The North American Breeding Bird Survey began in 1966 and is ongoing. Data are collected once per year in May or June along approximately 3000 "routes" – randomly situated 24.5 mile segments of roadway in the United States and Canada. Every 0.5 miles along each route, observers stop for 3 minutes and record abundance data on all bird species that they observe within 0.25 miles. Each route is sampled entirely in 1 day. Here, I considered data from 1997 and 2004 surveys – a total of 1,836,897 species-occurrences. I used data from 60,000 stops, randomly chosen from the routes that were rated to be of high quality (Sauer et al 2004). For analyses, I took individual stops as sites.

Barro Colorado Island (BCI) Vegetation

The 50 ha Barro Colorado Island Plot in Panama (UTM coordinates 625754, 1011569) has been censused 4 times since 1980. In each census, all trees at least 1 cm in diameter at breast height were identified and mapped to the nearest 0.1 m, providing records for over 3.25×10^6 individual trees (Condit 1998, Hubbell et al 1999, Condit et al 2005). Here, I used data from the 1982 and 1995 censuses, and I considered data only from live trees. For sites, I divided the 50 ha plot into 2 m, 5 m, or 10 m quadrats. I analyzed each type of site separately.

Sherman Plot Vegetation

The Sherman Plot is located in Panama, near the Panama Canal (UTM coordinates 612610, 1026067). The plot occupies 5.96 ha, with a 1 ha quadrat having been recently cleared. All trees of at least 1 cm diameter at breast height were identified and mapped to the nearest 0.1 m in censuses in 1996, 1997-1998, and 1999, for a total of 24,454 distinct individuals (see Condit et al 2004). Here, I analyzed data from the 1996 and 1999 censuses, considering only data from undisturbed portion of the plot. As with the BCI plot, I divided the Sherman plot into 2 m, 5 m, and 10 m quadrats, and analyzed each type of site separately.

Yosemite Zooplankton

This data set consists of zooplankton records from 279 randomly selected lakes in Yosemite National Park. Selection of lakes was stratified by physical characteristics of lakes and biological disturbance levels. All samples were collected in 2000 and 2001 using a conical plankton net (diameter 29.5 cm and mesh size 64-μm). To collect samples, in lakes less than 1 m deep, tows were taken horizontally; otherwise, vertical tows through the entire water column were collected from the deepest part of lakes. All zooplankton in two 1 mL subsamples of each sample were then identified to species. Over 101,000 individuals were identified (Knapp et al 2005). Here, I considered each lake a site.

John Muir Wilderness and Sequoia and Kings Canyon (JMW-SEKI) Zooplankton

This data set consists of zooplankton records from 357 randomly selected lakes in the John Muir Wilderness and Sequoia and Kings Canyon National Parks. Data were collected between 1995 and 1997 following the same protocols as for the Yosemite zooplankton data set. Over 251,500 individuals were identified (Knapp et al 2001). I considered each lake a site.

Yosemite Benthic Insects

This data set was created by sampling benthic insects in 295 randomly selected lakes in Yosemite National Park. As in the zooplankton sampling, lake selection was stratified, and samples were collected during the summers of 2000-2001. From each lake, 15 consecutive sweeps were collected over 1 m of littoral substrate using an aquatic D-net with mesh size 0.5 mm. At each lake, habitats were sampled in rough proportion to their relative abundances throughout the entire littoral area of the lake, with sweeps being combined as the sample was collected. All specimens were sorted from debris in the samples in the field, to be identified later, although when the amount of debris was excessive, only a portion of the sample was sorted. Where possible, I identified specimens to genus. A total of over 80,000 individuals were identified (Knapp et al 2005). Sites were taken as lakes.

JMW-SEKI Benthic Insects

This data set was created following the same protocols as the Yosemite Benthic Insect data set, but samples were collected from 314 randomly selected lakes in JMW-SEKI during the summers of 1995-1997. A total of over 15,000 individuals were identified (Knapp et al 2001). Sites were again taken as lakes.

Null Hypothesis Test

For each data set and each site, I first assembled a list of species or other taxa present. I then defined "subunits" and "units" so that species within the same subunit were more ecologically similar to each other than to those in different units. Hence, for example, in some cases, I defined subunits as genera and units as orders. I used only taxonomic groupings for subunit and unit definitions (Table 5), although other groupings were possible (see Discussion). Given a particular definition of subunits and units, I next considered the subset of sites having exactly a specified number of subunits ("$|R|$"). For each site, I defined the random vector **N** so that $\mathbf{N} = (n_1, n_2, ..., n_{|R|})$ if and only if n_i units had exactly *i* subunits, $i = 1, 2, ..., |R|$. Thus, if a site were observed to have a total of three subunits, with two sharing a unit and the third in a different unit, then the observed value of **N** would be $(1,1,0)$. Likewise, if a site had four subunits all sharing the same unit, then $\mathbf{N} = (0,0,0,1)$.

I next assumed that for all *i* and *j*, $P(W_{ij}) > 0$. Under this assumption, the aforementioned null hypothesis alone implied that within the set of sites having exactly $|R|$ subunits,

$$P[\mathbf{N} = (n_1, n_2, ..., n_{|R|})] = \frac{|R|!}{\prod_{i=1}^{|R|} (i!)^{n_i} \cdot n_i!} \cdot a^{|R| - |(n_1, n_2, ..., n_{|R|})|_1} \cdot \prod_{i=0}^{|(n_1, n_2, ..., n_{|R|})|_1 - 1} (1 - ia), \qquad (1)$$

where $|(n_1, n_2, ..., n_{|R|})|_1 = \sum_{i=1}^{|R|} n_i$ and $a \in (0, \frac{1}{|R|-1}) \cup \{\frac{1}{|R|-1}, \frac{1}{|R|-2}, ..., \frac{1}{2}, 1\}$ (Chapter III). Hence for example, among the sites having exactly three subunits, the fractions of sites having $\mathbf{N} = (0,0,1)$, $\mathbf{N} = (1,1,0)$, and $\mathbf{N} = (3,0,0)$ were predicted to be a^2, $3a(1-a)$, and $(1-a)(1-2a)$, respectively. To test the null hypothesis, I compared

Table 5. *P*-values[1] for data set and subunit/unit/quadrat-size combinations.

		***P*-Value**		
Data Set	**Subunits/Units**	**Full Data Set**	**Reduced Data Set**	**Reduced Data Set - Overall**
Breeding Bird Survey 2004	Families/Orders	0*	0*	<0.001*
Breeding Bird Survey 2004	Genera/Families	0*	0*	
Breeding Bird Survey 2004	Genera/Orders	0*	0*	
Breeding Bird Survey 2004	Species/Families	0*	0*	
Breeding Bird Survey 2004	Species/Genera	0*	0.019	
Breeding Bird Survey 2004	Species/Orders	0*	0*	
Barro Colorado Island 1995 2m	Genera/Families	0*	0*	<0.001*
Barro Colorado Island 1995 2m	Species/Families	0*	0*	
Barro Colorado Island 1995 2m	Species/Genera	0.216	0.421	
Barro Colorado Island 1995 5m	Genera/Families	0*	0*	
Barro Colorado Island 1995 5m	Species/Families	0*	0*	
Barro Colorado Island 1995 5m	Species/Genera	0*	0.431	
Barro Colorado Island 1995 10m	Genera/Families	0*	0.775	
Barro Colorado Island 1995 10m	Species/Families	0*	0.495	
Barro Colorado Island 1995 10m	Species/Genera	0*	1	
Sherman 1999 2m	Genera/Families	0.67	0.547	0.567
Sherman 1999 2m	Species/Families	0.659	0.441	
Sherman 1999 2m	Species/Genera	0.895	0.694	
Sherman 1999 5m	Genera/Families	0.85	0.981	
Sherman 1999 5m	Species/Families	0*	0.085	
Sherman 1999 5m	Species/Genera	0.855	1	
Sherman 1999 10m	Genera/Families	0*	1	
Sherman 1999 10m	Species/Families	0*	0.364	
Sherman 1999 10m	Species/Genera	0.552	1	
JMW-SEKI Zooplankton	Families/Orders	0*	1	0.015
JMW-SEKI Zooplankton	Genera/Families	0.073	1	
JMW-SEKI Zooplankton	Genera/Orders	0.013	0.002	
JMW-SEKI Zooplankton	Species/Families	0.572	1	
JMW-SEKI Zooplankton	Species/Genera	0.952	0.919	
JMW-SEKI Zooplankton	Species/Orders	0*	0.002	
Yosemite Zooplankton	Families/Orders	0*	0*	<0.001*
Yosemite Zooplankton	Genera/Families	0*	1	
Yosemite Zooplankton	Genera/Orders	0*	0*	
Yosemite Zooplankton	Species/Families	0.098	0.757	
Yosemite Zooplankton	Species/Genera	0*	1	
Yosemite Zooplankton	Species/Orders	0*	0*	
JMW-SEKI Benthic Insects	Families/Orders	0.654	0.612	0.979
JMW-SEKI Benthic Insects	Genera/Families	0*	0.343	
JMW-SEKI Benthic Insects	Genera/Orders	0*	0.575	
Yosemite Benthic Insects	Families/Orders	0.991	0.999	0.818
Yosemite Benthic Insects	Genera/Families	0*	0.041	
Yosemite Benthic Insects	Genera/Orders	0.047	0.557	

[1] Results significant after correction for multiple comparisons are starred.

the observed and predicted distributions of **N**. If the distributions differed significantly, then I rejected the null hypothesis.

Two issues complicated the comparisons. First, because they contain the parameter a, the aforementioned fractions could not be used directly to find a predicted distribution of **N**. I resolved this problem by using maximum likelihood estimation for a. Second, the aforementioned distribution of **N** pertained only to the set of sites having exactly three subunits. For sets of sites having different numbers of subunits, (1) gives different distributions of **N**. For each of these sets of sites, I estimated a, and computed the appropriate likelihood ratio statistic. The product of the likelihood ratio statistics over all sets of sites gave an overall likelihood ratio statistic, whose significance I checked using Monte Carlo simulations. 1000 statistics were simulated per observed statistic.

For each data set, I repeated the test using different combinations of subunits and unit definitions. For the BCI and Sherman data sets, I additionally repeated the test using each definition of "sites;" i.e., 2 m, 5 m, and 10 m quadrats. Hence, a total of 91 tests were performed, and I set the individual significance level at 0.0005 to give an overall significance level of 0.05. To obtain an overall result for each data set, I used a likelihood ratio test (Sokal and Rohlf 1995, p.794). Procedures were implemented using custom-written Visual Basic 6.0 software.

Causes of Null Hypothesis Rejection

Assumption Violations

Here I used a test that assumes that $P(W_{ij}) > 0$ for all i and j. Hence, assumption violations may have caused the true Type I error rate to exceed the nominal rate of 0.05. To minimize this problem, I postulated that if two subunits i and *j* were observed to occur in the same unit somewhere, then $P(W_{ij})$ might reasonably be supposed nonzero everywhere. Hence, for each data set and subunit/unit/quadrat-size combination, I found the largest number of subunits to occur in the same unit ("*M*"), and then constructed a "reduced" data set, which included only data from sites having *M* or fewer subunits. With these data sets, $P(W_{ij})$ was likely to be nonzero for most choices of *i* and *j*, and tests were likely to have Type I error rates close to 0.05. Except where otherwise noted, in analyses I used the reduced data sets.

Random Fluctuations

Using the reduced data sets, the null hypothesis may have been rejected due to either random fluctuations or underlying biological processes. If random fluctuations were the cause, then inconsistent results should have been obtained in the absence of major perturbations. Between 1997 and 2004, 1982 and 1995, and 1996 and 1999, no drastic perturbations occurred in area surveyed in Breeding Bird Survey, the BCI plot, or the Sherman plot, respectively. Hence, to check for effects of random fluctuations, I checked whether the null hypothesis was rejected for the same subunit/unit/quadrat-size combinations in different years for these data sets. I used a test of Cohen's Kappa statistic to assess consistency (Siegel 1988).

Non-Competitive Interactions

The test applied here is a two-tailed test; that is, for example, the null hypothesis might be false because $P(W_{13} \mid W_{12}) < P(W_{13} \mid W_{12}^{c})$, $P(W_{23} \mid W_{12}) < P(W_{23} \mid W_{12}^{c})$, and $P(W_{34} \mid W_{23} \bigcap W_{12}^{c}) < P(W_{34} \mid W_{23}^{c} \bigcap W_{12}^{c})$, or because $P(W_{13} \mid W_{12}) > P(W_{13} \mid W_{12}^{c})$, $P(W_{23} \mid W_{12}) > P(W_{23} \mid W_{12}^{c})$, or $P(W_{34} \mid W_{23} \bigcap W_{13}^{c}) > P(W_{34} \mid W_{23}^{c} \bigcap W_{13}^{c})$. However, only the former is consistent with competitive effects. To check whether competitive effects were responsible for rejections of the null hypothesis, I considered that if they were the cause, then the frequency of rejections should be inversely proportional to the inclusiveness of the units. Hence, for example, because species in the same genera should on average be less ecologically similar than those in the same families, competitive effects predict that the null hypothesis should be rejected less often when units are defined as families than when they are defined as genera. I checked this possibility using a χ^2 test for trend (Zar 1999, p. 565).

Comparison to a Standard Parametric Test

Using the JMW-SEKI and Yosemite data, I compared the results of Ladau and Schwager's test to those of a parametric test. Because the sites in these data sets were small, I employed a parametric test that uses a *C-score* statistic and a *fixed rows – equiprobable columns* randomization algorithm (see Sanders et al 2003). For each data set, this parametric test gave a single result, so for comparison I used the overall result of Ladau and Schwager's test (see "Null Hypothesis Test," above). To perform parametric testing, I employed EcoSim Version 7.70, Build 120404 (Gotelli and Entsminger 2001).

Predictive Power of the Null Model

It was of interest to assess not just whether the null hypothesis was rejected, but the extent of deviations from it – to assess the relative importance of interspecific interactions in determining community composition. To quantify the extent of deviations, I calculated the proportion of variation in frequencies of partitions that could be accounted for by the null hypothesis (R^2). Specifically, for a particular data set and subunit/unit/quadrat-size combination *i*, set of sites having exactly *j* subunits, and value **n** of **N**: let $y_{ij\mathbf{n}}$ and $\hat{y}_{ij\mathbf{n}}$ denote the observed and predicted frequencies of **n**, and let $\overline{y}_{ij\boldsymbol{\cdot}}$ denote mean frequency for the set of sites having *j* subunits. For data set and subunit/unit/quadrat-size combination *i*

$$R^2 = 1 - \frac{\sum_{j,\mathbf{n}} (y_{ij\mathbf{n}} - \hat{y}_{ij\mathbf{n}})^2}{\sum_{j,\mathbf{n}} (y_{ij\mathbf{n}} - \overline{y}_{ij\mathbf{n}})^2} \qquad (2)$$

(Kvalseth 1985). In cases where the model performed extremely poorly, (2) gave negative values of R^2 – in effect, the mean predicted values of $y_{ij\mathbf{n}}$ better than the model. In these cases, I set $R^2 = 0$. Thus, R^2 could consistently be interpreted in terms of the amount of variation for which the model accounted. I considered one potential predictor of R^2: the inclusiveness of the units. I used a Pitman correlation test to check this possibility (Bradley 1968).

RESULTS

Null Hypothesis Tests

Using the full data sets, 27 of the 42 tests performed were significant, while using the reduced data sets, only 12 of 42 tests were significant (Table 5). The parametric test gave non-significant results in all four cases examined ($p = 1$ in all cases), consistent with an absence of competitive effects. Using the reduced data sets, Ladau and Schwager's test gave a significant result in one case (significant: Yosemite zooplankton; non-significant: Yosemite benthic insects, JMW-SEKI benthic insects, and JMW-SEKI zooplankton). However, this significant result was likely due to non-competitive effects, because the null hypothesis was rejected more often when orders, rather than genera or families, were used as units (Table 5).

Using Ladau and Schwager's test with the reduced data sets, results were consistent between years $(K = 0.909, \quad z = 6.547, p < 0.001)$, suggesting that random fluctuations were not influential in rejecting the null hypothesis. As unit inclusiveness increased, the null hypothesis was rejected more frequently $(X^2 = 5.437, \text{ df } = 1, p = 0.021)$, an outcome inconsistent with competitive effects.

Predictive Power of the Null Model

Using the full data sets, and considering the values of R^2 for each data set and subunit/unit/quadrat-size combination individually, the median R^2 was 0.942. Using the reduced data sets, the median was 0.993 (Figure 6 and Figure 7). Considering the overall R^2 values for each data set, the medians remained high (0.937 and 0.962 for the full and reduced data sets, respectively), but R^2 was less than 0.5 for the BBS data, and 0.637 and 0.598 for the Yosemite zooplankton data with the full and reduced data sets, respectively. (R^2 was greater than 0.8 for all other cases.) In the BBS data, the lack of fit stemmed from cases where units were defined as orders ($R^2 = 0$ using both the full and reduced BBS data sets). Overall, there was a negative relationship between R^2 and the inclusiveness of the definition of units (for the full and reduced data sets respectively, $r = -0.467$, $p = 0.001$; $r = -0.454$, $p < 0.001$). However,

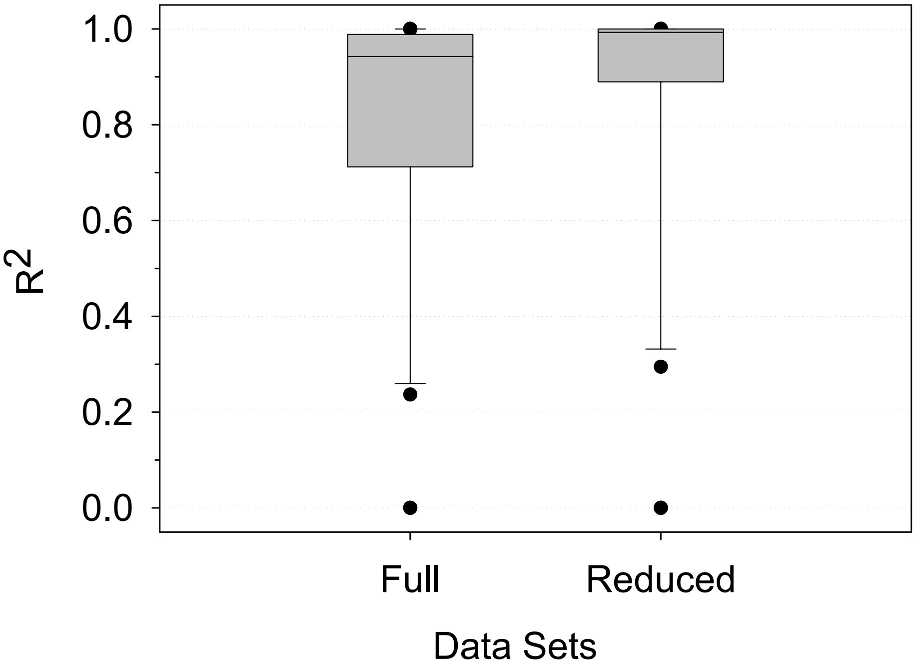


Figure 6. Box plots of the observed distributions of R^2. 42 values of R^2 were used for both the reduced and full data sets – each corresponding to a data set and subunit/unit/quadrat-size combination. Because results were similar between years, only the BBS, BCI, and Sherman data from 2004, 1995, and 1999 are plotted, respectively.

Figure 7. Plots showing the observed and predicted distributions of **N** for the 2004 Breeding Bird Survey data. Subunits are defined as species, and units as families. Each plot gives the distributions for the subset of sites having exactly the listed number of species. Some sites had more than 13 species, but for clarity and brevity such sites are omitted. Although the observed and predicted distributions differ significantly $(p < 0.001)$, the null model accounts for a large portion of the variation in the frequencies of **N** $(R^2 = 0.963)$.

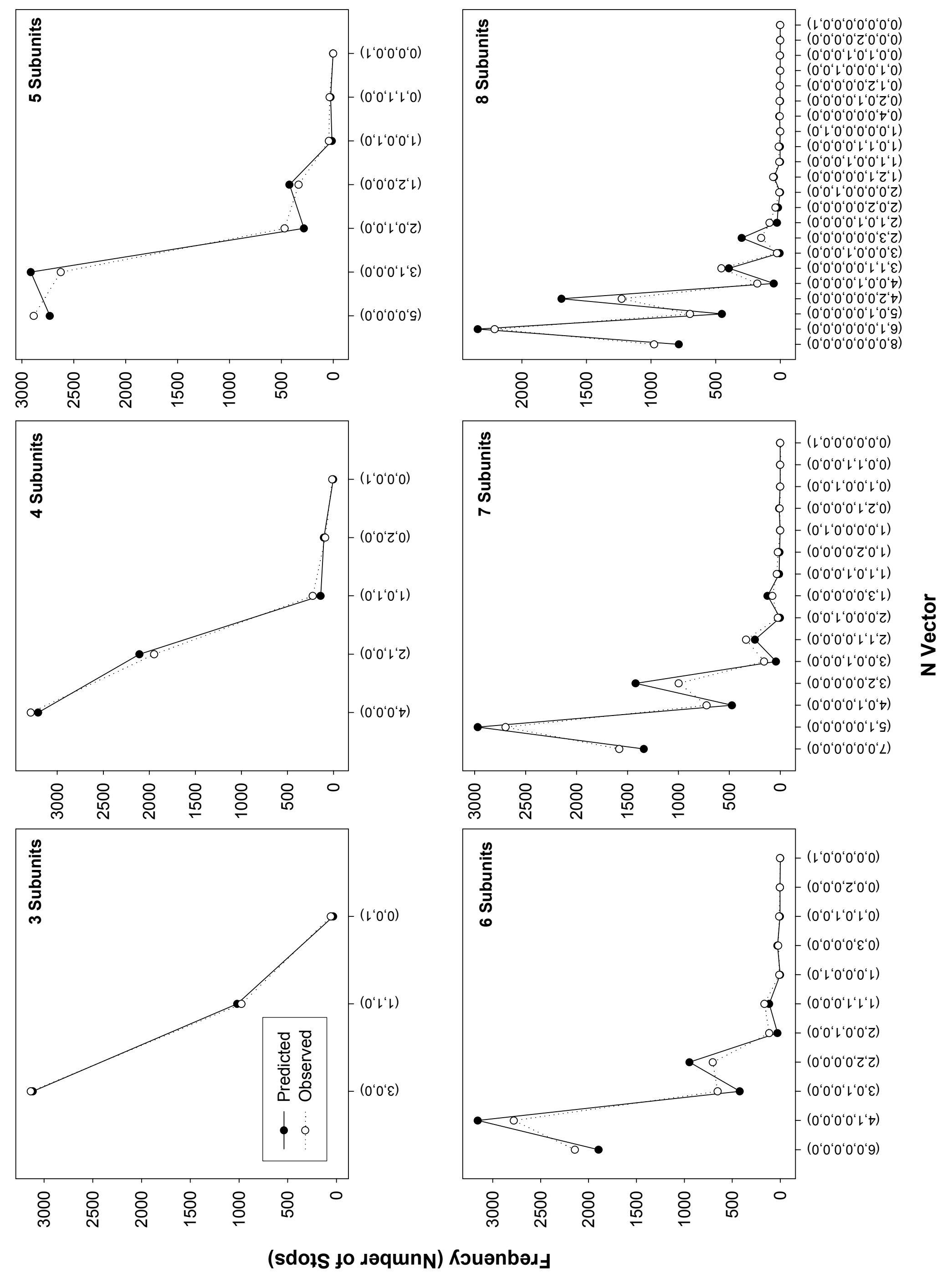

Frequency (Number of Stops)
N Vector
3 Subunits
Predicted
Observed
0
500
1000
1500
2000
2500
3000
(3,0,0)
(1,1,0)
(0,0,1)
4 Subunits
(4,0,0,0)
(2,1,0,0)
(1,0,1,0)
(0,2,0,0)
(0,0,0,1)
5 Subunits
(5,0,0,0,0)
(3,1,0,0,0)
(2,0,1,0,0)
(1,2,0,0,0)
(1,0,0,1,0)
(0,1,1,0,0)
(0,0,0,0,1)
6 Subunits
(6,0,0,0,0,0)
(4,1,0,0,0,0)
(3,0,1,0,0,0)
(2,2,0,0,0,0)
(2,0,0,1,0,0)
(1,1,1,0,0,0)
(1,0,0,0,1,0)
(0,3,0,0,0,0)
(0,1,0,1,0,0)
(0,0,2,0,0,0)
(0,0,0,0,0,1)
7 Subunits
(7,0,0,0,0,0,0)
(5,1,0,0,0,0,0)
(4,0,1,0,0,0,0)
(3,2,0,0,0,0,0)
(3,0,0,1,0,0,0)
(2,1,1,0,0,0,0)
(2,0,0,0,1,0,0)
(1,3,0,0,0,0,0)
(1,1,0,1,0,0,0)
(1,0,2,0,0,0,0)
(1,0,0,0,0,1,0)
(0,2,1,0,0,0,0)
(0,1,0,0,1,0,0)
(0,0,1,1,0,0,0)
(0,0,0,0,0,0,1)
8 Subunits
(8,0,0,0,0,0,0,0)
(6,1,0,0,0,0,0,0)
(5,0,1,0,0,0,0,0)
(4,2,0,0,0,0,0,0)
(4,0,0,1,0,0,0,0)
(3,1,1,0,0,0,0,0)
(3,0,0,0,1,0,0,0)
(2,3,0,0,0,0,0,0)
(2,1,0,1,0,0,0,0)
(2,0,2,0,0,0,0,0)
(2,0,0,0,0,1,0,0)
(1,2,1,0,0,0,0,0)
(1,1,0,0,1,0,0,0)
(1,0,1,1,0,0,0,0)
(1,0,0,0,0,0,1,0)
(0,4,0,0,0,0,0,0)
(0,2,0,1,0,0,0,0)
(0,1,2,0,0,0,0,0)
(0,1,0,0,0,1,0,0)
(0,0,1,0,1,0,0,0)
(0,0,0,2,0,0,0,0)
(0,0,0,0,0,0,0,1)

Figure 7 (Continued)

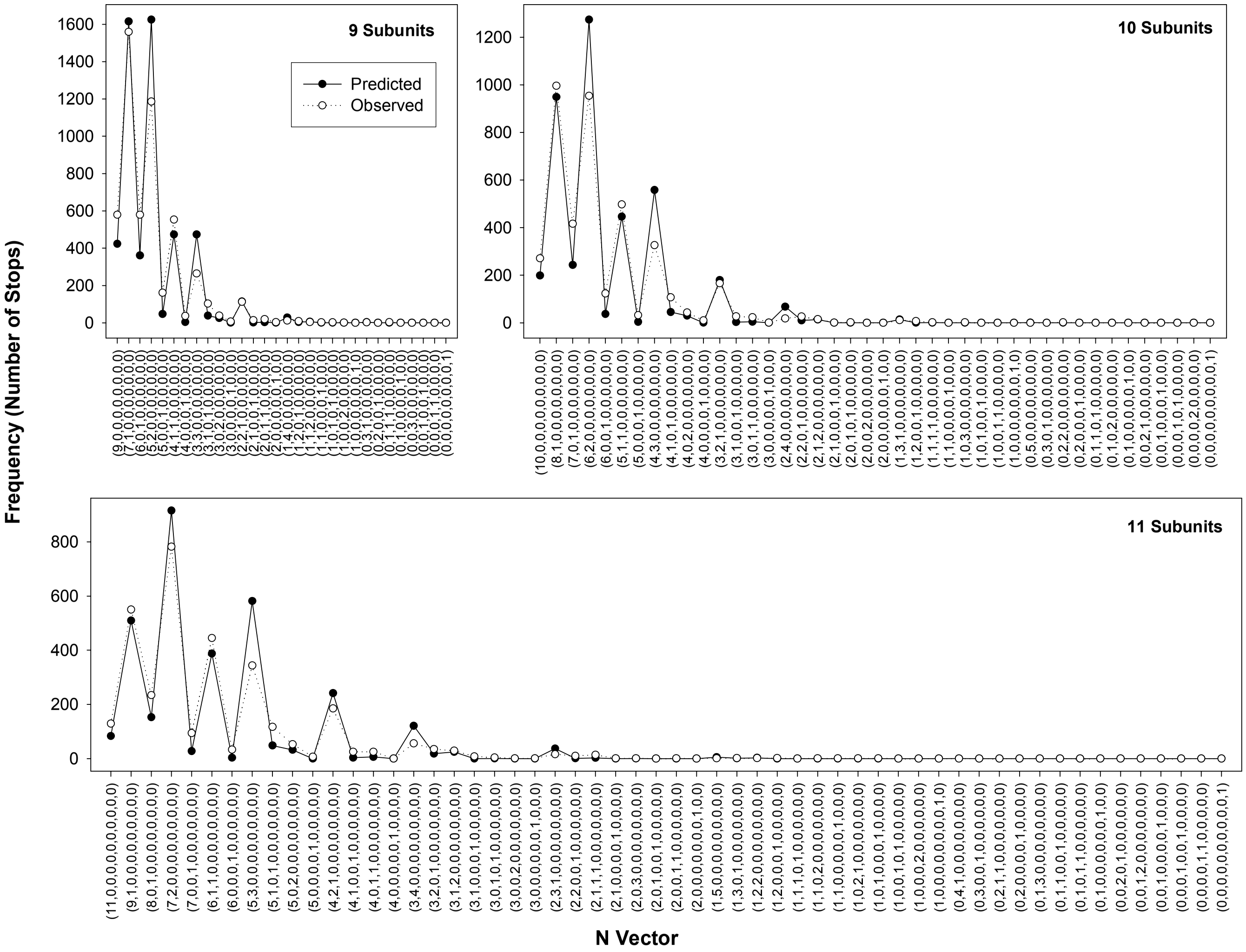
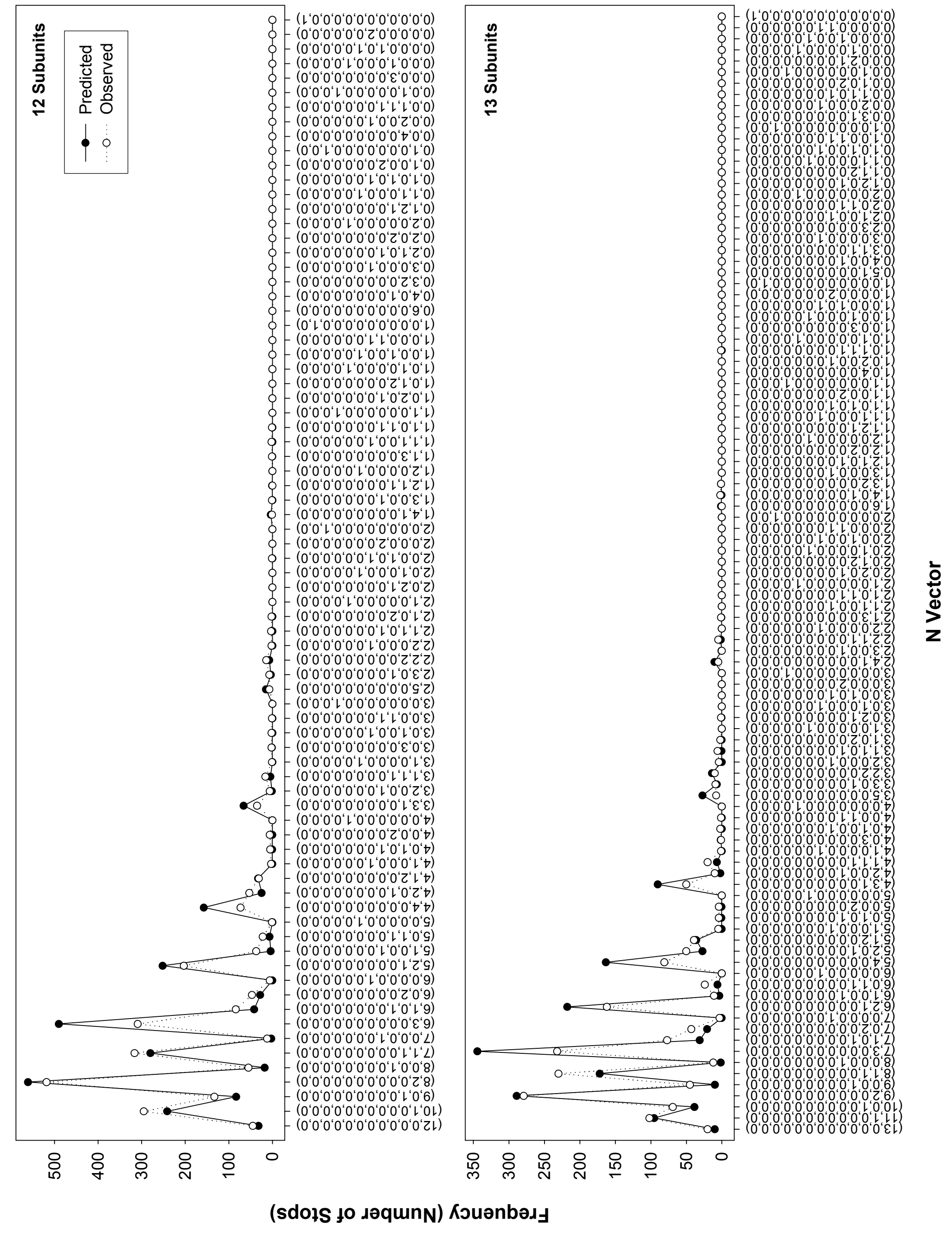


Figure 7 (Continued)



there were notable exceptions to this trend – for example, with the JMW-SEKI benthic insect data set, R^2 values were greater than 0.925 when units were taken as orders.

DISCUSSION

Using the full data sets, the null hypothesis was rejected at a rate higher than the nominal significance level, 0.05. Because assumptions may have been invalid for the full data sets, the high rate may have been partially due to a high Type I error rate. However, the Type I error rate was closer to 0.05 for the reduced data sets, where assumptions were better justified. As the null hypothesis was still rejected at a rate exceeding 0.05 for these data sets, the null hypothesis was likely indeed violated in some cases. Importantly, this conclusion is based on the premise that if two subunits have nonzero probability of occurring in the same unit somewhere, then they have nonzero probability of occurring in the same unit everywhere – a premise that may not always be true. Nonetheless, preliminary analyses with the assumption-free test suggest that the null hypothesis was indeed often rejected correctly, particularly for the BBS data (Ladau, unpublished results).

I investigated two factors that could have caused violations of the null hypothesis: random fluctuations unrelated to interspecific interactions, and effects of interspecific competition. Random fluctuations appear to have had little influence on null hypothesis rejection, because in stable communities, the null hypothesis was consistently rejected for the same subunit/unit/quadrat-size combinations. Likewise, competitive effects appear to have been un-influential, as the frequency of null hypothesis rejection was negatively related to the inclusiveness of units.

Nonetheless, despite the lack of evidence, competitive effects may still have influenced community composition. For subunits and units, I used the groupings of species, genera, families and orders. However, for the test applied here to detect competitive effects, subunits within the same unit must be more ecologically similar than those in different units. Ecological similarity may be better reflected by other taxonomic groupings, or by groupings based explicitly on ecological similarity – for instance, functional groups or guilds (Fox 1987). Moreover, competitive effects may be manifest at different spatial scales than those examined here. It is possible, for instance, that the effects are evident only in sites smaller than the ones used here.

In the four cases examined, the parametric test indicated a lack of effects of competitive interactions, like Ladau and Schwager's test. However, such a result cannot always be expected. Although both tests check for effects of competitive interactions, those effects are non-equivalent; it is possible for one test's null hypothesis to be true but not the other. Moreover, the true Type I error rate of the parametric test will often exceed the nominal significance level, in which case the null hypothesis will be frequently rejected when it is true (Chapter II). So long as the assumption that for all i and j, $P(W_{ij}) > 0$ is justified, Ladau and Schwager's test will be immune to this problem (Bradley 1968).

Overall, the null hypothesis could be rejected for the BBS, BCI, and Yosemite zooplankton communities, but not for the Sherman, JMW-SEKI zooplankton, JMW-SEKI benthic insect, or Yosemite benthic insect communities. This outcome may have been due to two factors. First, a simple evolutionary birth process can invalidate the null hypothesis (Ladau, unpublished results), although why such a process would operate in just the former communities is unclear. Second, facilitative interspecific interactions may be concentrated between ecologically dissimilar species (but see Stachowicz 2001, p.237), potentially accounting for the violations of the null hypothesis. Consistent with this explanation, facilitative interactions have been documented in bird (Kilham 1971, Daily et al 1993, Blendinger 1999, Pejchar and Jeffrey 2004, Schlatter and Vergara 2005) and zooplankton communities (Nandini and Sarma 2001, Nandini and Sarma 2002). They are also widespread in plant communities (Stachowicz 2001), with the failure to reject the null hypothesis for Sherman plot perhaps because of small sample sizes relative to the BCI plot. By contrast, aquatic insect communities may be structured primarily by interspecific predation (Merrit and Cummins 1996, p. 70), whose effects on the test's outcome will be variable.

Overall, the null model predicted the composition of the communities almost exactly. By one measure – the vector **N** with the reduced data sets – the model predicted a median of over 99% of the variation in community composition. Particularly remarkable was the null model's ability to account for the "bumps" and irregularities in the observed frequency distributions (Figure 7). Similar predictive power of a null model was found by Connor and Simberloff (1979), but later shown to be trivial, as the fitting procedures employed therein made it mathematically impossible to obtain low values of R^2 (Diamond and Gilpin 1982). However, it can be shown analytically that poor fits are possible here (Ladau, unpublished results). Moreover, using both the full and reduced data sets, in three cases the model predicted ~0% of the variation in community composition, and in six and three cases respectively, the model predicted less than 50% of the variation. Thus, the nearly perfect overall agreement between observation and prediction was not a trivial result of the fitting procedure.

Although there were exceptions, R^2 was highest (>0.95) when genera and families were used as units, and lowest when orders were used. Hence, at least when units are defined as genera and families, the following model of community assembly appears to apply: By definition, the first type of subunit ("I") to arrive and persist at a community must be in its own unit. The second subunit ("II") might then be in I's unit, or belong to its own unit. When the third subunit ("III") arrives, it can belong to I's unit, II's unit, or be in its own unit. According to the model, the probability of III belonging to II's unit is the same regardless of whether II is in the same unit as I. Likewise, III is equally likely to belong to I's unit regardless of whether I shares its unit with II. Corresponding equalities then hold for subunits that arrive later. This model – an informal description of the null hypothesis – is null to competitive effects inasmuch that, for instance, when competitive effects occur, III should be less likely to belong to II's unit when it shares that unit with I than when it does not.

In cases where R^2 was high and the null hypothesis could not be rejected, the model potentially could account entirely for the observed compositions of the communities as measured by **N**. Where R^2 was high and the null hypothesis was rejected, the model could account *almost* entirely. Thus, the tests using genera or families as units suggested that interspecific interactions influence community composition at most minimally. For the North American breeding bird communities, the tests using orders as units suggested that this influence may be more substantial.

Recently, considerable interest has focused on developing simple theories to account for broad and complex macroecological patterns (e.g., Harte et al 1999, Hubbell 2001, Harte et al 2005). As put succinctly by Harte (2003), the aim is to develop theories whose "seemingly preposterous assumption… yields amazingly accurate predictions of a range of phenomena." The strength of such theories lies both in their predictive power and the inferences that can be drawn the failures of their predictions. The results presented here suggest that the null model of Ladau and Schwager constitutes a simple, but biologically relevant and highly predictive theory of community composition.

ACKNOWLEDGEMENTS

This work benefited from the helpful suggestions of Dr. R. Knapp and Dr. H. K. Reeve. Thousands of volunteers collected the BBS data, and data entry and quality control were performed by the USGS Patuxent Wildlife Research Center. The BCI and Sherman data are kindly provided publicly by the Center for Tropical Forest Studies. Many field workers spent countless hours collecting the JMW-SEKI and Yosemite data. The Sierra Nevada Aquatic Research Laboratory and Dr. Roland Knapp graciously provided material support for this work. Funding was provided by a National Science Foundation Graduate Research Fellowship.

LITERATURE CITED

Bledinger, P.G. 1999. Facilitation of sap-feeding birds by the White-Fronted Woodpecker in the Monte Desert, Argentina. The Condor **101:**402-407.

Bradley, J. V. 1968. Distribution-Free Statistical Tests. Prentice-Hall, Englewood Cliffs, New Jersey.

Condit, R. 1998. Tropical Forest Census Plots. Springer-Verlag and R. G. Landes Company, Berlin, Germany, and Georgetown, Texas.

Condit, R., S. Aguilar, A. Hernández, R. Pérez, S. Lao, G. Angehr, S.P. Hubbell, and R.B. Foster. 2004. Tropical forest dynamics across a rainfall gradient and the impact of an El Niño dry season. Journal of Tropical Ecology **20:**51-72.

Condit, R., S.P. Hubbell, and R.B. Foster. 2005. Barro Colorado Forest Census Plot Data. URL http://ctfs.si/edu/datasets/bci.

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

Daily, G. C., P. R. Ehrlich, and N. M. Haddad. 1993. Double keystone bird in a keystone species complex. Proceedings of the National Academy of Science USA **90:**592-594.

Diamond, J. M. 1975. Assembly of species communities. In Ecology and Evolution of Communities, ed. M. L. Cody and J. M. Diamond, 342-344. Harvard University Press, Cambridge.

Diamond, J.M. and M.E. Gilpin. 1982. Examination of the “Null” Model of Connor and Simberloff for Species Co-Occurrence on Islands. Oecologia **52:**64-74.

Feeley, K. 2003. Analysis of avian communities in Lake Guri, Venezuela, using multiple assembly rule models. Oecologia **137:**104-113.

Fox, B. J. 1987. Species assembly and the evolution of community structure.

Evolutionary Ecology **1:**201-213.

Fox, B. J. and J. H. Brown. 1993. Assembly rules for functional groups in North American desert rodent communities. Oikos **67:**358-370.

Gleason, H. A. 1926. The individualistic concept of plant association. Bulletin of the Torey Botanical Club **53:**7-26.

Gotelli, N. J., and K. Rohde. 2002. Co-occurrence of ectoparasites of marine fishes: a null model analysis. Ecology Letters **5:**86-94.

Gotelli, N. J., and L. G. Abele. 1982. Statistical distributions of West Indian land bird families. Journal of Biogeography **9:**421-435.

Gotelli, N.J. and G.L. Entsminger. 2001. EcoSim: Null models software for ecology. Version 7.0. Acquired Intelligence Inc. & Kesey-Bear. http://homepages.together.net/~gentsmin/ecosim.htm..

Harte, J. 2003. Tail of death and resurrection. Nature **424:**1006-1007.

Harte, J., A. Kinzig, J. Green. 1999. Self-Similarity in the Distribution and Abundance of Species. Science **284:**334-336.

Harte, J., E. Conlisk, A. Ostling, J.L. Green, and A.B. Smith. 2005. A theory of spatial structure in ecological communities at multiple spatial scales. Ecological Monographs **75:**179-197.

Hubbel, S.P. 2001. The Unified Neutral Theory of Biodiversity and Biogeography. Monographs in Population Biology 32. Princeton University Press, Princeton, New Jersey.

Hubbell, S.P., R.B. Foster, S.T. O'Brien, K.E. Harms, R. Condit, B. Wechsler, S.J. Wright, and S. Loo de Lao. 1999. Light gap disturbances, recruitment limitation, and tree diversity in a neotropical forest. Science **283:**554-557.

Kilham, L. 1971. Reproductive behavior of Yellow-Bellied Sapsuckers: I. Preference for nesting in *Fomes*-infected aspens and nest hole interrelations with flying squirrels, raccoons, and other animals. The Wilson Bulletin **83:**159-171.

Knapp, R. A., K. R. Matthews, and O. Sarnelle. 2001. Resistance and resilience of alpine lake fauna to fish introductions. Ecological Monographs **71:**401–421.

Knapp, R.A., C.P. Hawkins, J. Ladau, J.G. McClory. 2005. Fauna of Yosemite National Park has low resistance but high resilience to fish introductions. Ecological Applications **15:**835-847.

Kvalseth, T.O. 1985. Cautionary Note about R^2. The American Statistician **39:**279-285.

McNab, B.K. 1971. The structure of tropical bat faunas. Ecology **52:**352-358.

Merrit, R.W. and K.W. Cummins. 1996. An Introduction to the Aquatic Insects of North America: Third Edition. Kendall/Hunt, Dubuque, Iowa.

Nandini, S. and S.S.S. Sarma. 2001. Population growth of *Lepadella patella* (O. F. Müller, 1786) at different algal (*Chlorella vulgaris*) densities and in association with *Philodina roseola* Ehrenberg, 1832. Hydrobiologia **446/447:**63-69.

Nandini, S. and S.S.S. Sarma. 2002. Competition between *Moina macrocopa* and *Ceriodaphnia dubia*: a Life Table Demography Study. International Review of Hydrobiology **87:**85-95.

Pejchar, L. and J. Jeffrey. 2004. Sap-feeding behavior and tree selection in the endangered Akiapolauu (*Hemignathus munroi*) in Hawaii. The Auk **121:**548-556.

Sanders, N. J., N. J. Gotelli, N. E. Heller, and D. M. Gordon. 2003. Community disassembly by an invasive species. Proceedings of the National Academy of Science USA **100:**2474-2477.

Sauer, J. R., J. E. Hines, and J. Fallon. 2004. The North American Breeding Bird Survey, Results and Analysis 1966 - 2004. Version 2004.0. USGS Patuxent Wildlife Research Center, Laurel, MD. ftp://ftpext.usgs.gov/pub/er/md/laurel/BBS/DataFiles/.

Schlatter R. P. and P. Vergara. 2005. Magellanic Woodpecker (*Campephilus magellanicus*) sap feeding and its role in the Tierra del Fuego forest bird assemblage. Journal of Ornithology **146:**188-190.

Siegel, S. 1988. Nonparametric Statistics for the Behavioral Sciences: Second Edition. McGraw-Hill, New York.

Sokal, R. R., and F. J. Rohlf. 1995. Biometry: Third Edition. W. H. Freeman and Company, New York.

Stachowicz, J. J. 2001. Mutualism, Facilitation, and the Structure of Ecological Communities. BioScience **51:**235-246.

Stone, L., T. Dayan, and D. Simberloff. 1996. Community-wide assembly patterns unmasked: the importance of species' differing geographical ranges. The American Naturalist **148:**997-1015.

Zar, J. H. 1999. Biostatistical Analysis: Fourth Edition. Prentice Hall, Upper Saddle River, New Jersey.

— APPENDIX A —

PROPERTIES USED TO CHECK FOR UNCONDITIONAL INDEPENDENCE

DEFINITIONS

Let $\mathscr{S}$ be the set of sites sampled, and let the sample space $S \equiv \{\mathbf{M} : \mathbf{M}$ a binary matrix with $|\mathscr{S}|$ columns$\}$. Let $\mathbf{R}$ be a binary row vector with $|\mathscr{S}|$ columns. For any **R**, define the event $E_i(\mathbf{R}) = \{\mathbf{M} : \mathbf{M} \in S,$ row i of **M** given by $\mathbf{R}\}$. Also, define the event C so that $\mathbf{M} \in C$ if and only if **M** meets arbitrary predefined criteria. The following is an example of these definitions:

Suppose that two sites are sampled. Then $|\mathscr{S}| = 2$, and S is the set of all binary matrices with 2 columns. For instance, $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 1 \end{pmatrix} \in S$. If $\mathbf{M} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, then $\mathbf{M} \in E_1[(1,0)]$ and $\mathbf{M} \in E_2[(0,1)]$, while if $\mathbf{M} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$, then $\mathbf{M} \in E_1[(0,0)]$ and $\mathbf{M} \in E_2[(0,0)]$. If the criterion for event C is that all row and column totals are nonzero, then $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \in C$, while $\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \notin C$.

By definition, $E_i(\mathbf{R})$ is the event that species i has the distribution given by **R**. For example, if four sites are sampled and $E_1[(1,0,0,1)]$ occurs, then species 1 is found only at the first and fourth sites. Therefore, Null Hypothesis 1 (below) is a formalization of the null hypothesis of unconditional independence.

RESULTS

Null Hypothesis 1. H_0: for any $\rho \subseteq \{1, 2, ...\}$ and set of binary row vectors $\{\mathbf{R}_i : i \in \rho, \mathbf{R}_i \text{ with } |\mathscr{S}| \text{ columns}\}$,

$$P[\bigcap_{i \in \rho} E_i(\mathbf{R}_i)] = \prod_{i \in \rho} P[E_i(\mathbf{R}_i)] \tag{1}$$

Theorem 1. Under H_0, for any arbitrary predefined criteria for **M**, disjoint $\rho_1, \rho_2, \rho_3 \subseteq \{1, 2, ...\}$ with $\rho_1 \neq \varnothing$ and $\rho_2 \neq \varnothing$ (ρ_3 possibly empty), and unique sets of binary row vectors $\{\mathbf{R}_i : \mathbf{R}_i \text{ with } |\mathscr{S}| \text{ columns}, i \in \{1, 2, ...\}\}$ and $\{\mathbf{R}'_i : \mathbf{R}'_i \text{ with } |\mathscr{S}| \text{ columns}, i \in \{1, 2, ...\}\}$, if:

$$P[C \mid [\bigcap_{i \in \rho_1} E_i(\mathbf{R}_i)] \bigcap [\bigcap_{i \in \rho_2} E_i(\mathbf{R}_i)] \bigcap [\bigcap_{i \in \rho_3} E_i(\mathbf{R}_i)]] = 1, \tag{2}$$

$$P[C \mid [\bigcap_{i \in \rho_1} E_i(\mathbf{R}_i)] \bigcap [\bigcap_{i \in \rho_2} E_i(\mathbf{R}'_i)] \bigcap [\bigcap_{i \in \rho_3} E_i(\mathbf{R}_i)]] = 1, \tag{3}$$

$$P[C \mid [\bigcap_{i \in \rho_1} E_i(\mathbf{R}'_i)] \bigcap [\bigcap_{i \in \rho_2} E_i(\mathbf{R}_i)] \bigcap [\bigcap_{i \in \rho_3} E_i(\mathbf{R}_i)]] = 1, \tag{4}$$

$$P[C \mid [\bigcap_{i \in \rho_1} E_i(\mathbf{R}'_i)] \bigcap [\bigcap_{i \in \rho_2} E_i(\mathbf{R}'_i)] \bigcap [\bigcap_{i \in \rho_3} E_i(\mathbf{R}_i)]] = 1, \tag{5}$$

then

$$\frac{P[[\bigcap_{i \in \rho_1} E_i(\mathbf{R}_i)] \bigcap [\bigcap_{i \in \rho_2} E_i(\mathbf{R}_i)] \bigcap [\bigcap_{i \in \rho_3} E_i(\mathbf{R}_i)] \mid C]}{P[[\bigcap_{i \in \rho_1} E_i(\mathbf{R}'_i)] \bigcap [\bigcap_{i \in \rho_2} E_i(\mathbf{R}_i)] \bigcap [\bigcap_{i \in \rho_3} E_i(\mathbf{R}_i)] \mid C]}$$

$$= \frac{P[[\bigcap_{i\in\rho_1} E_i(\mathbf{R}_i)]\bigcap[\bigcap_{i\in\rho_2} E_i(\mathbf{R'}_i)]\bigcap[\bigcap_{i\in\rho_3} E_i(\mathbf{R}_i)] \mid C]}{P[[\bigcap_{i\in\rho_1} E_i(\mathbf{R'}_i)]\bigcap[\bigcap_{i\in\rho_2} E_i(\mathbf{R'}_i)]\bigcap[\bigcap_{i\in\rho_3} E_i(\mathbf{R}_i)] \mid C]}.$$

Proof: Assume H_0. Given arbitrary predefined criteria for **M**, disjoint $\rho_1, \rho_2, \rho_3 \subseteq \{1,2,...\}$ with $\rho_1, \rho_2 \neq \varnothing$ (ρ_3 possibly empty), and the sets of row vectors listed above, assume the aforementioned conditions. By (4), (5), and the definition of conditional probability,

$$P[[\bigcap_{i\in\rho_1} E_i(\mathbf{R'}_i)]\bigcap[\bigcap_{i\in\rho_2} E_i(\mathbf{R}_i)]\bigcap[\bigcap_{i\in\rho_3} E_i(\mathbf{R}_i)]] > 0, \tag{6}$$

$$P[[\bigcap_{i\in\rho_1} E_i(\mathbf{R'}_i)]\bigcap[\bigcap_{i\in\rho_2} E_i(\mathbf{R'}_i)]\bigcap[\bigcap_{i\in\rho_3} E_i(\mathbf{R}_i)]] > 0, \tag{7}$$

and

$$P(C) > 0. \tag{8}$$

By H_0,

$$\begin{aligned} &P[[\bigcap_{i\in\rho_1} E_i(\mathbf{R'}_i)]\bigcap[\bigcap_{i\in\rho_2} E_i(\mathbf{R}_i)]\bigcap[\bigcap_{i\in\rho_3} E_i(\mathbf{R}_i)]] \\ &\qquad = P[\bigcap_{i\in\rho_1} E_i(\mathbf{R'}_i)]\cdot P[\bigcap_{i\in\rho_2} E_i(\mathbf{R}_i)]\cdot P[\bigcap_{i\in\rho_3} E_i(\mathbf{R}_i)] \end{aligned} \tag{9}$$

and

$$\begin{aligned} &P[[\bigcap_{i\in\rho_1} E_i(\mathbf{R'}_i)]\bigcap[\bigcap_{i\in\rho_2} E_i(\mathbf{R'}_i)]\bigcap[\bigcap_{i\in\rho_3} E_i(\mathbf{R}_i)]] \\ &\qquad = P[\bigcap_{i\in\rho_1} E_i(\mathbf{R'}_i)]\cdot P[\bigcap_{i\in\rho_2} E_i(\mathbf{R'}_i)]\cdot P[\bigcap_{i\in\rho_3} E_i(\mathbf{R}_i)] \end{aligned} \tag{10}$$

Hence,

$$\frac{P[\bigcap_{i\in\rho_1} E_i(\mathbf{R}_i)]\cdot P[\bigcap_{i\in\rho_2} E_i(\mathbf{R}_i)]\cdot P[\bigcap_{i\in\rho_3} E_i(\mathbf{R}_i)] \Big/ P(C)}{P[\bigcap_{i\in\rho_1} E_i(\mathbf{R'}_i)]\cdot P[\bigcap_{i\in\rho_2} E_i(\mathbf{R}_i)]\cdot P[\bigcap_{i\in\rho_3} E_i(\mathbf{R}_i)] \Big/ P(C)}$$

$$= \frac{P[\bigcap_{i\in\rho_1} E_i(\mathbf{R}_i)]\cdot P[\bigcap_{i\in\rho_2} E_i(\mathbf{R}'_i)]\cdot P[\bigcap_{i\in\rho_3} E_i(\mathbf{R}_i)] \Big/ P(C)}{P[\bigcap_{i\in\rho_1} E_i(\mathbf{R}'_i)]\cdot P[\bigcap_{i\in\rho_2} E_i(\mathbf{R}'_i)]\cdot P[\bigcap_{i\in\rho_3} E_i(\mathbf{R}_i)] \Big/ P(C)}. \tag{11}$$

Also by H_0,

$$\begin{aligned} &P[\bigcap_{i\in\rho_1} E_i(\mathbf{R}_i)]\cdot P[\bigcap_{i\in\rho_2} E_i(\mathbf{R}_i)]\cdot P[\bigcap_{i\in\rho_3} E_i(\mathbf{R}_i)] \\ &\qquad = P[[\bigcap_{i\in\rho_1} E_i(\mathbf{R}_i)]\bigcap[\bigcap_{i\in\rho_2} E_i(\mathbf{R}_i)]\bigcap[\bigcap_{i\in\rho_3} E_i(\mathbf{R}_i)]] \end{aligned} \tag{12}$$

implying by (2) and the definition of conditional probability that

$$\begin{aligned} &P[\bigcap_{i\in\rho_1} E_i(\mathbf{R}_i)]\cdot P[\bigcap_{i\in\rho_2} E_i(\mathbf{R}_i)]\cdot P[\bigcap_{i\in\rho_3} E_i(\mathbf{R}_i)] \\ &\qquad = P[C\bigcap[\bigcap_{i\in\rho_1} E_i(\mathbf{R}_i)]\bigcap[\bigcap_{i\in\rho_2} E_i(\mathbf{R}_i)]\bigcap[\bigcap_{i\in\rho_3} E_i(\mathbf{R}_i)]]. \end{aligned} \tag{13}$$

Thus, by the definition of conditional probability,

$$\begin{aligned} &P[\bigcap_{i\in\rho_1} E_i(\mathbf{R}_i)]\cdot P[\bigcap_{i\in\rho_2} E_i(\mathbf{R}_i)]\cdot P[\bigcap_{i\in\rho_3} E_i(\mathbf{R}_i)]/P(C) \\ &\qquad = P[[\bigcap_{i\in\rho_1} E_i(\mathbf{R}_i)]\bigcap[\bigcap_{i\in\rho_2} E_i(\mathbf{R}_i)]\bigcap[\bigcap_{i\in\rho_3} E_i(\mathbf{R}_i)]\,|\,C]. \end{aligned} \tag{14}$$

By similar arguments with (3) – (5), it follows that

$$\begin{aligned} &P[\bigcap_{i\in\rho_1} E_i(\mathbf{R}_i)]\cdot P[\bigcap_{i\in\rho_2} E_i(\mathbf{R}'_i)]\cdot P[\bigcap_{i\in\rho_3} E_i(\mathbf{R}_i)]/P(C) \\ &\qquad = P[[\bigcap_{i\in\rho_1} E_i(\mathbf{R}_i)]\bigcap[\bigcap_{i\in\rho_2} E_i(\mathbf{R}'_i)]\bigcap[\bigcap_{i\in\rho_3} E_i(\mathbf{R}_i)]\,|\,C], \end{aligned} \tag{15}$$

$$\begin{aligned} &P[\bigcap_{i\in\rho_1} E_i(\mathbf{R}'_i)]\cdot P[\bigcap_{i\in\rho_2} E_i(\mathbf{R}_i)]\cdot P[\bigcap_{i\in\rho_3} E_i(\mathbf{R}_i)]/P(C) \\ &\qquad = P[[\bigcap_{i\in\rho_1} E_i(\mathbf{R}'_i)]\bigcap[\bigcap_{i\in\rho_2} E_i(\mathbf{R}_i)]\bigcap[\bigcap_{i\in\rho_3} E_i(\mathbf{R}_i)]\,|\,C], \end{aligned} \tag{16}$$

and

$$\begin{aligned} &P[\bigcap_{i\in\rho_1} E_i(\mathbf{R}'_i)]\cdot P[\bigcap_{i\in\rho_2} E_i(\mathbf{R}'_i)]\cdot P[\bigcap_{i\in\rho_3} E_i(\mathbf{R}_i)]/P(C) \\ &\qquad = P[[\bigcap_{i\in\rho_1} E_i(\mathbf{R}'_i)]\bigcap[\bigcap_{i\in\rho_2} E_i(\mathbf{R}'_i)]\bigcap[\bigcap_{i\in\rho_3} E_i(\mathbf{R}_i)]\,|\,C]. \end{aligned} \tag{17}$$

The result follows directly from applying (14)-(17) to (11). ■

— APPENDIX B —

QUARTETS OF MATRICES EXAMINED

Following the definitions from Appendix A, for certain predefined criteria for **M**, it may be possible to choose $\{\mathbf{R}_i : \mathbf{R}_i$ a binary row vector with $|\mathscr{S}|$ columns, $i \in \{1,2,...\}\}$ and $\{\mathbf{R}'_i : \mathbf{R}'_i$ a binary row vector with $|\mathscr{S}|$ columns, $i \in \{1,2,...\}\}$ so that

$$\left| C \bigcap [\bigcap_{i \in \rho_1} E_i(\mathbf{R}_i)] \bigcap [\bigcap_{i \in \rho_2} E_i(\mathbf{R}_i)] \bigcap [\bigcap_{i \in \rho_3} E_i(\mathbf{R}_i)] \right| = 1 \,, \tag{1}$$

$$\left| C \bigcap [\bigcap_{i \in \rho_1} E_i(\mathbf{R}_i)] \bigcap [\bigcap_{i \in \rho_2} E_i(\mathbf{R}'_i)] \bigcap [\bigcap_{i \in \rho_3} E_i(\mathbf{R}_i)] \right| = 1 \,, \tag{2}$$

$$\left| C \bigcap [\bigcap_{i \in \rho_1} E_i(\mathbf{R}'_i)] \bigcap [\bigcap_{i \in \rho_2} E_i(\mathbf{R}_i)] \bigcap [\bigcap_{i \in \rho_3} E_i(\mathbf{R}_i)] \right| = 1 \,, \tag{3}$$

and

$$\left| C \bigcap [\bigcap_{i \in \rho_1} E_i(\mathbf{R}'_i)] \bigcap [\bigcap_{i \in \rho_2} E_i(\mathbf{R}'_i)] \bigcap [\bigcap_{i \in \rho_3} E_i(\mathbf{R}_i)] \right| = 1 \,. \tag{4}$$

Under these circumstances, fixing

$$\mathbf{M}_1 \in C \bigcap [\bigcap_{i \in \rho_1} E_i(\mathbf{R}_i)] \bigcap [\bigcap_{i \in \rho_2} E_i(\mathbf{R}_i)] \bigcap [\bigcap_{i \in \rho_3} E_i(\mathbf{R}_i)] \,, \tag{5}$$

$$\mathbf{M}_2 \in C \bigcap [\bigcap_{i \in \rho_1} E_i(\mathbf{R}'_i)] \bigcap [\bigcap_{i \in \rho_2} E_i(\mathbf{R}'_i)] \bigcap [\bigcap_{i \in \rho_3} E_i(\mathbf{R}_i) \,, \tag{6}$$

$$\mathbf{M}_3 \in C \bigcap [\bigcap_{i \in \rho_1} E_i(\mathbf{R}'_i)] \bigcap [\bigcap_{i \in \rho_2} E_i(\mathbf{R}_i)] \bigcap [\bigcap_{i \in \rho_3} E_i(\mathbf{R}_i) \,, \tag{7}$$

and

$$\mathbf{M}_4 \in C \bigcap [\bigcap_{i \in \rho_1} E_i(\mathbf{R}_i)] \bigcap [\bigcap_{i \in \rho_2} E_i(\mathbf{R}'_i)] \bigcap [\bigcap_{i \in \rho_3} E_i(\mathbf{R}_i) \,, \tag{8}$$

by Theorem 1, under H_0,

$$\frac{P(\mathbf{M}_1 \mid C)}{P(\mathbf{M}_3 \mid C)} = \frac{P(\mathbf{M}_4 \mid C)}{P(\mathbf{M}_2 \mid C)}. \tag{9}$$

In Chapter I, I investigate whether (9) holds for various $\{\mathbf{M}_i : i = 1, 2, 3, 4\}$. Table 6 lists the $\{\mathbf{M}_i : i = 1, 2, 3, 4\}$ that I use.

Table 6. Quartets of matrices used to evaluate Gotelli's and Graves's models.

	M_1[1]	**M_2**	**M_3**	**M_4**
Matrix 1	Equiprobable Rows, Proportional Columns			
	001,010,110	010,001,101	010,001,110	001,010,101
	001,110,010	110,001,100	110,001,010	001,110,100
	010,101,001	101,010,100	101,010,001	010,101,100
	100,001,011	001,100,110	001,100,011	100,001,110
	100,010,101	010,100,011	010,100,101	100,010,011
	Proportional Rows, Equiprobable Columns			
	001,010,110	010,001,101	010,001,110	001,010,101
	001,110,001	010,101,100	010,101,001	001,110,100
	010,100,101	100,010,011	100,010,101	010,100,011
	011,100,010	110,001,100	110,001,010	011,100,100
	Proportional Rows, Proportional Columns			
	010,001,110	001,010,101	001,001,110	010,010,101
	100,001,011	101,100,010	100,100,011	101,001,010
	001,100,011	010,011,100	010,100,011	001,011,100
	011,100,100	010,001,101	011,001,100	010,100,101
	001,101,010	010,110,001	001,110,010	010,101,001
Matrix 2	Equiprobable Rows, Proportional Columns			
	110,111,100	111,011,001	111,011,100	110,111,001
	011,110,101	100,011,111	100,110,111	011,011,101
	111,100,011	110,001,111	111,001,011	110,100,111
	011,001,111	111,100,110	111,001,110	011,100,111
	110,110,011	001,011,111	110,011,011	001,110,111
	Proportional Rows, Equiprobable Columns			
	001,111,011	101,101,110	101,101,011	001,111,110
	011,011,101	111,010,110	111,010,101	011,011,110
	011,101,011	110,011,110	110,011,011	011,101,110
	011,110,011	111,001,101	111,001,011	011,110,101
	011,111,001	101,111,010	101,111,001	011,111,010
	Proportional Rows, Proportional Columns			
	001,011,111	011,110,110	001,110,111	011,011,110
	111,010,110	001,111,011	111,010,011	001,111,110
	010,110,111	011,011,101	010,011,111	011,110,101
	101,011,011	101,110,110	101,011,110	101,110,011
	100,011,111	111,110,010	111,011,010	100,110,111
Matrix3	Equiprobable Rows, Proportional Columns			
	110,110,111	101,111,011	101,110,111	110,111,011
	011,011,111	111,110,110	111,011,110	011,110,111
	Proportional Rows, Equiprobable Columns			
	011,111,011	111,011,101	111,011,011	011,111,101
	101,111,110	110,111,011	110,111,110	101,111,011
	111,110,011	111,101,101	111,101,011	111,110,101
	Proportional Rows, Proportional Columns			
	011,111,011	111,110,110	111,110,011	011,111,110
	101,111,110	111,101,101	101,111,101	111,101,110
	011,011,111	111,110,101	111,011,101	011,110,111
	111,011,110	110,111,011	111,011,011	110,111,110

[1] Commas delimit rows of matrices.

Table 6 (Continued)

	M$_1$	**M**$_2$	**M**$_3$	**M**$_4$
Matrix 4	Equiprobable Rows, Proportional Columns			
	100,111,001,001	001,110,011,100	100,110,011,100	001,111,001,001
	001,100,110,011	100,101,001,011	100,100,110,011	001,101,001,011
	010,101,100,110	001,011,101,001	010,101,101,001	001,011,100,110
	110,100,001,110	101,001,001,011	101,001,001,110	110,100,001,011
	110,011,010,010	011,101,001,001	110,011,001,001	011,101,010,010
	Proportional Rows, Equiprobable Columns			
	100,010,101,011	101,011,100,010	100,011,101,010	101,010,100,011
	010,100,011,011	110,101,001,001	010,101,011,001	110,100,001,011
	110,100,101,001	001,101,010,011	001,100,101,011	110,101,010,001
	101,100,110,010	001,110,010,011	001,100,110,011	101,110,010,010
	101,010,011,100	010,011,010,101	101,011,010,100	010,010,011,101
	Proportional Rows, Proportional Columns			
	001,011,011,100	011,010,010,101	001,010,011,101	011,011,010,100
	001,110,010,011	101,100,011,010	001,100,011,011	101,110,010,010
	011,001,110,010	001,010,111,001	011,001,110,001	001,010,111,010
	001,001,111,001	100,011,110,010	001,001,111,010	100,011,110,001
	101,001,011,001	011,100,110,010	101,001,011,010	011,100,110,001
	Proportional Rows, Fixed Columns			
	010,001,011,101	001,011,001,110	001,001,011,110	010,011,001,101
	001,010,001,111	011,001,010,101	001,001,010,111	011,010,001,101
	001,010,101,011	011,001,110,001	001,001,110,011	011,010,101,001
	100,001,011,011	110,011,001,001	110,001,001,011	100,011,011,001
	011,110,001,001	001,100,011,011	001,110,011,001	011,100,001,011
Matrix 5	Equiprobable Rows, Proportional Columns			
	100,011,100,101	011,100,110,100	100,011,110,100	011,100,100,101
	110,001,010,110	001,101,011,100	110,001,011,100	001,101,010,110
	001,101,011,010	110,101,001,001	110,101,001,010	001,101,011,001
	100,101,010,110	011,100,011,100	100,100,011,110	011,101,010,100
	001,011,001,101	010,001,101,011	010,011,001,101	001,001,101,011
	Proportional Rows, Equiprobable Columns			
	100,100,011,101	101,101,010,100	101,100,011,100	100,101,010,101
	010,010,101,101	110,011,001,100	110,010,101,100	010,011,001,101
	100,100,110,101	101,110,010,100	101,100,110,100	100,110,010,101
	011,101,010,001	001,100,101,011	011,100,101,001	001,101,010,011
	001,011,110,100	110,010,100,101	001,010,110,101	110,011,100,100
	Proportional Rows, Proportional Columns			
	100,010,011,110	110,011,010,100	100,011,010,110	110,010,011,100
	100,010,011,101	011,011,001,100	100,011,011,100	011,010,001,101
	001,100,110,011	110,110,010,001	110,100,110,001	001,110,010,011
	100,100,110,101	011,101,010,100	100,101,010,101	011,100,110,100
	100,001,011,110	110,011,100,100	110,001,011,100	100,011,100,110
	Proportional Rows, Fixed Columns			
	001,001,110,110	010,010,101,101	010,001,101,110	001,010,110,101
	101,010,010,101	110,001,001,110	110,010,001,101	101,001,010,110
	001,110,110,001	010,101,101,010	010,110,101,001	001,101,110,010
	001,110,010,101	010,101,001,110	010,110,001,101	001,101,010,110
	101,001,110,010	110,010,101,001	101,010,101,010	110,001,110,001

— APPENDIX C —

GOTELLI AND GRAVES'S MODELS RECAST IN TERMS OF INDIVIDUALS

In the following, let $\boldsymbol{\varphi}$ denote the matrix of probabilities used to place the first species in Gotelli and Graves's models.

Proposition 1. The following algorithm is equivalent to SIM6:

1. The first individual is placed according to $\boldsymbol{\varphi}$.
2. Additional individuals are placed independently and according to $\boldsymbol{\varphi}$, until the simulated and observed numbers of total species-occurrences match.

Justification: Rather than provide a rigorous proof, I will outline arguments with an example, from which generalizations follow. Let

$$\boldsymbol{\varphi} \equiv \begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{pmatrix} \tag{10}$$

where each p_{ij} is a probability less than 1, and $\sum_{i,j} p_{ij} = 1$, for $i = 1,2,3$ and $j = 1,2,3$.

SIM6 implies that for the presence-absence matrix

$$\mathbf{M} \equiv \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{11}$$

$$P\{\mathbf{M}\} = \sum_{(x,y,z)\in S} p_x \cdot \frac{p_y}{1-p_x} \cdot \frac{p_z}{1-p_x-p_y} \tag{12}$$

where $S \equiv \{(x,y,z) : x,y,z \in \{12, 21, 33\}; x \neq y \neq z\}$. Moreover, the proposed algorithm indicates that the probability of observing an arrival sequence of (i) species 1 at site 2, then (ii) species 2 at site 1, and finally (iii) species 3 at site 3 is given by

$$p_{12} \cdot (\sum_{n=0}^{\infty} p_{12}^{n}) \cdot p_{21} \cdot [\sum_{n=0}^{\infty} (p_{12} + p_{21})^{n}] \cdot p_{33} \tag{13}$$

Importantly, (13) will hold only if individuals arrive independently of each other, and all according to $\boldsymbol{\varphi}$. Under the same conditions, the proposed algorithm implies that

$$P\{\mathbf{M}\} = \sum_{(x,y,z)\in S} [p_x \cdot (\sum_{n=0}^{\infty} p_x^{n}) \cdot p_y \cdot (\sum_{n=0}^{\infty} (p_x + p_y)^{n}) \cdot p_z] \tag{14}$$

Moreover, because for any geometric sequence

$$\sum_{n=0}^{\infty} a^{n} = \frac{1}{1-a}, \tag{15}$$

$$\sum_{n=0}^{\infty} p_x^{n} = \frac{1}{1-p_x} \tag{16}$$

and

$$\sum_{n=0}^{\infty} (p_x + p_y)^{n} = \frac{1}{1-p_x-p_y}, \tag{17}$$

implying that the right hand sides of (12) and (14) must be equal. Generalizing, it can be shown that equality holds regardless of the initial probabilities and presence-absence matrix. Hence, the proposition follows. ■

Proposition 2. The following algorithm is equivalent to SIM5:

1. The first individual in each column is placed according to $\boldsymbol{\varphi}$.
2. Additional individuals are placed independently and according to $\boldsymbol{\varphi}$, until the simulated numbers of species in each column match the marginal totals.

Justification: The arguments from Proposition 1 can immediately be extended, by treating each column as a separate presence-absence matrix, into which individuals are placed until the number of species matches the corresponding marginal total. ■

— APPENDIX D —

ASSUMPTION VIOLATIONS FOR GOTELLI’S (2000) MODELS

This appendix lists the assumption violations that I considered for each of Gotelli's (2000) models. The assumptions are divided into two sets: the “limited” set, consisting of violations suggested by the models (Table 7), and the “full” set, including additional violations constructed here (Table 8).

Table 7. Assumption violations belonging to the "limited set" for Gotelli's (2000) models.

Model	Assumption Violations[1]		
SIM1	T_j/NR	S_i/NC	S_iT_j/N^2
SIM2	T_j/N		
SIM3	S_i/N		
SIM4	$1/C$		
SIM5	$1/R$		
SIM6	$1/RC$	S_i/NC	S_iT_j/N^2
SIM7	$1/RC$	T_j/NR	S_iT_j/N^2
SIM8	$1/RC$	T_j/NR	S_i/NC
SIM9	T_j/N	S_i/N	$\sim S_iT_j/N$

[1] Variables defined in Figure 5.

Table 8. Assumption violations belonging only to the “full set” for Gotelli’s (2000) models.

Matrix	Model(s)	Violation									
I	SIM1, SIM6, SIM7, SIM8	0.028	0.028	0.028	0.028	0.028	0.001	0.001	0.001	0.001	0.001
		0.028	0.028	0.028	0.028	0.028	0.001	0.001	0.001	0.001	0.001
		0.028	0.028	0.028	0.028	0.028	0.001	0.001	0.001	0.001	0.001
		0.028	0.028	0.028	0.028	0.028	0.001	0.001	0.001	0.001	0.001
		0.001	0.001	0.001	0.001	0.001	0.028	0.028	0.028	0.028	0.028
		0.001	0.001	0.001	0.001	0.001	0.028	0.028	0.028	0.028	0.028
		0.001	0.001	0.001	0.001	0.001	0.028	0.028	0.028	0.028	0.028
I	SIM1, SIM6, SIM7, SIM8	0.005	0.005	0.005	0.026	0.005	0.026	0.005	0.026	0.026	0.005
		0.005	0.005	0.005	0.026	0.026	0.026	0.005	0.026	0.026	0.005
		0.026	0.026	0.026	0.005	0.026	0.005	0.005	0.005	0.005	0.005
		0.005	0.005	0.005	0.026	0.005	0.026	0.026	0.026	0.026	0.026
		0.005	0.005	0.005	0.005	0.026	0.005	0.026	0.005	0.005	0.026
		0.005	0.005	0.005	0.026	0.005	0.026	0.005	0.026	0.026	0.005
		0.005	0.005	0.005	0.026	0.026	0.026	0.005	0.026	0.026	0.005
I	SIM3, SIM5	0.243	0.243	0.243	0.243	0.243	0.013	0.013	0.013	0.013	0.013
		0.243	0.243	0.243	0.243	0.243	0.013	0.013	0.013	0.013	0.013
		0.243	0.243	0.243	0.243	0.243	0.013	0.013	0.013	0.013	0.013
		0.243	0.243	0.243	0.243	0.243	0.013	0.013	0.013	0.013	0.013
		0.01	0.01	0.01	0.01	0.01	0.317	0.317	0.317	0.317	0.317
		0.01	0.01	0.01	0.01	0.01	0.317	0.317	0.317	0.317	0.317
		0.01	0.01	0.01	0.01	0.01	0.317	0.317	0.317	0.317	0.317
I	SIM3, SIM5	0.091	0.091	0.091	0.185	0.044	0.185	0.067	0.185	0.185	0.067
		0.091	0.091	0.091	0.185	0.217	0.185	0.067	0.185	0.185	0.067
		0.453	0.453	0.453	0.037	0.217	0.037	0.067	0.037	0.037	0.067
		0.091	0.091	0.091	0.185	0.044	0.185	0.333	0.185	0.185	0.333
		0.091	0.091	0.091	0.037	0.217	0.037	0.333	0.037	0.037	0.333
		0.091	0.091	0.091	0.185	0.044	0.185	0.067	0.185	0.185	0.067
		0.091	0.091	0.091	0.185	0.217	0.185	0.067	0.185	0.185	0.067

Table 8 (Continued)

Matrix	Model(s)	Violation											
I	SIM2, SIM4	0.192	0.192	0.192	0.192	0.192	0.008	0.008	0.008	0.008	0.008		
		0.192	0.192	0.192	0.192	0.192	0.008	0.008	0.008	0.008	0.008		
		0.192	0.192	0.192	0.192	0.192	0.008	0.008	0.008	0.008	0.008		
		0.192	0.192	0.192	0.192	0.192	0.008	0.008	0.008	0.008	0.008		
		0.008	0.008	0.008	0.008	0.008	0.192	0.192	0.192	0.192	0.192		
		0.008	0.008	0.008	0.008	0.008	0.192	0.192	0.192	0.192	0.192		
		0.008	0.008	0.008	0.008	0.008	0.192	0.192	0.192	0.192	0.192		
I	SIM2, SIM4	0.039	0.039	0.039	0.192	0.039	0.192	0.039	0.192	0.192	0.039		
		0.034	0.034	0.034	0.167	0.167	0.167	0.034	0.167	0.167	0.034		
		0.192	0.192	0.192	0.039	0.192	0.039	0.039	0.039	0.039	0.039		
		0.03	0.03	0.03	0.147	0.03	0.147	0.147	0.147	0.147	0.147		
		0.046	0.046	0.046	0.046	0.227	0.046	0.227	0.046	0.046	0.227		
		0.039	0.039	0.039	0.192	0.039	0.192	0.039	0.192	0.192	0.039		
		0.034	0.034	0.034	0.167	0.167	0.167	0.034	0.167	0.167	0.034		
II	SIM1, SIM6, SIM7, SIM8	0.027	0.027	0.027	0.027	0.027	0.027	0.001	0.001	0.001	0.001	0.001	0.001
		0.027	0.027	0.027	0.027	0.027	0.027	0.001	0.001	0.001	0.001	0.001	0.001
		0.027	0.027	0.027	0.027	0.027	0.027	0.001	0.001	0.001	0.001	0.001	0.001
		0.001	0.001	0.001	0.001	0.001	0.001	0.027	0.027	0.027	0.027	0.027	0.027
		0.001	0.001	0.001	0.001	0.001	0.001	0.027	0.027	0.027	0.027	0.027	0.027
		0.001	0.001	0.001	0.001	0.001	0.001	0.027	0.027	0.027	0.027	0.027	0.027
II	SIM3, SIM5	0.32	0.32	0.32	0.32	0.32	0.32	0.013	0.013	0.013	0.013	0.013	0.013
		0.32	0.32	0.32	0.32	0.32	0.32	0.013	0.013	0.013	0.013	0.013	0.013
		0.32	0.32	0.32	0.32	0.32	0.32	0.013	0.013	0.013	0.013	0.013	0.013
		0.013	0.013	0.013	0.013	0.013	0.013	0.32	0.32	0.32	0.32	0.32	0.32
		0.013	0.013	0.013	0.013	0.013	0.013	0.32	0.32	0.32	0.32	0.32	0.32
		0.013	0.013	0.013	0.013	0.013	0.013	0.32	0.32	0.32	0.32	0.32	0.32

Table 8 (Continued)

Matrix	Model(s)	Violation																		
II	SIM2, SIM4	0.16	0.16	0.16	0.16	0.16	0.16	0.007	0.007	0.007	0.007	0.007	0.007							
		0.16	0.16	0.16	0.16	0.16	0.16	0.007	0.007	0.007	0.007	0.007	0.007							
		0.16	0.16	0.16	0.16	0.16	0.16	0.007	0.007	0.007	0.007	0.007	0.007							
		0.007	0.007	0.007	0.007	0.007	0.007	0.16	0.16	0.16	0.16	0.16	0.16							
		0.007	0.007	0.007	0.007	0.007	0.007	0.16	0.16	0.16	0.16	0.16	0.16							
		0.007	0.007	0.007	0.007	0.007	0.007	0.16	0.16	0.16	0.16	0.16	0.16							
III	SIM1, SIM6, SIM7, SIM8	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04
		0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04
		0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04
		0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04
		0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04
		0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04
		0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04
		0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04
		0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04
		2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006
		2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006
		2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006
		2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006
		2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006
		2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006
		2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006
		2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006

Table 8 (Continued)

Matrix	Model(s)	Violation																		
III	SIM1, SIM6, SIM7, SIM8	0.006	0.006	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04
		0.006	0.006	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04
		0.006	0.006	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04
		0.006	0.006	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04
		0.006	0.006	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04
		0.006	0.006	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04
		0.006	0.006	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04
		0.006	0.006	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04
		0.006	0.006	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04
		2E-04	2E-04	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006
		2E-04	2E-04	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006
		2E-04	2E-04	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006
		2E-04	2E-04	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006
		2E-04	2E-04	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006
		2E-04	2E-04	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006
		2E-04	2E-04	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006
		2E-04	2E-04	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006

Table 8 (Continued)

Matrix	Model(s)	Violation																		
III	SIM1, SIM6, SIM7, SIM8	0.011	0.011	0.011	0.011	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04
		0.011	0.011	0.011	0.011	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04
		0.011	0.011	0.011	0.011	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04
		0.011	0.011	0.011	0.011	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04
		4E-04	4E-04	4E-04	4E-04	0.011	0.011	0.011	0.011	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04
		4E-04	4E-04	4E-04	4E-04	0.011	0.011	0.011	0.011	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04
		4E-04	4E-04	4E-04	4E-04	0.011	0.011	0.011	0.011	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04
		4E-04	4E-04	4E-04	4E-04	0.011	0.011	0.011	0.011	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04
		4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	0.011	0.011	0.011	0.011	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04
		4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	0.011	0.011	0.011	0.011	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04
		4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	0.011	0.011	0.011	0.011	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04
		4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	0.011	0.011	0.011	0.011	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04
		4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	0.011	0.011	0.011	0.011	0.011	0.011	0.011
		4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	0.011	0.011	0.011	0.011	0.011	0.011	0.011
		4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	0.011	0.011	0.011	0.011	0.011	0.011	0.011
		4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	0.011	0.011	0.011	0.011	0.011	0.011	0.011
		4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	0.011	0.011	0.011	0.011	0.011	0.011	0.011

Table 8 (Continued)

Matrix	Model(s)	Violation																		
III	SIM1, SIM6, SIM7, SIM8	4E-04	0.011	4E-04	4E-04	4E-04	4E-04	4E-04	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011
		4E-04	0.011	4E-04	4E-04	4E-04	4E-04	4E-04	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011
		4E-04	0.011	4E-04	4E-04	4E-04	4E-04	4E-04	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011
		4E-04	0.011	4E-04	4E-04	4E-04	4E-04	4E-04	0.011	4E-04	4E-04	0.011	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04
		4E-04	0.011	4E-04	4E-04	4E-04	4E-04	4E-04	0.011	4E-04	4E-04	0.011	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04
		4E-04	0.011	4E-04	4E-04	4E-04	4E-04	4E-04	0.011	4E-04	4E-04	0.011	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04
		4E-04	0.011	4E-04	4E-04	4E-04	4E-04	4E-04	0.011	4E-04	4E-04	0.011	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04
		4E-04	4E-04	0.011	0.011	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04
		4E-04	4E-04	0.011	0.011	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04
		4E-04	4E-04	0.011	0.011	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04
		4E-04	4E-04	0.011	0.011	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04
		4E-04	4E-04	0.011	0.011	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04
		4E-04	4E-04	0.011	0.011	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04
		4E-04	4E-04	0.011	0.011	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04
		0.011	4E-04	4E-04	4E-04	0.011	0.011	0.011	4E-04	0.011	0.011	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04
		0.011	4E-04	4E-04	4E-04	0.011	0.011	0.011	4E-04	0.011	0.011	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04
		0.011	4E-04	4E-04	4E-04	0.011	0.011	0.011	4E-04	0.011	0.011	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04	4E-04

Table 8 (Continued)

Matrix	Model(s)	Violation																		
III	SIM1, SIM6, SIM7, SIM8	0.016	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04
		0.016	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04
		0.016	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04
		0.016	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04
		0.016	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04
		0.016	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04
		0.016	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04
		6E-04	6E-04	6E-04	0.016	0.016	0.016	0.016	0.016	0.016	0.016	0.016	0.016	0.016	0.016	0.016	0.016	0.016	0.016	0.016
		0.016	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04
		0.016	0.016	0.016	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04
		6E-04	6E-04	6E-04	0.016	0.016	0.016	0.016	0.016	0.016	0.016	0.016	0.016	0.016	0.016	0.016	0.016	0.016	0.016	0.016
		0.016	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04
		0.016	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04
		0.016	0.016	0.016	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04
		0.016	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04
		0.016	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04
		0.016	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04

Table 8 (Continued)

Matrix	Model(s)	Violation																		
III	SIM1, SIM6, SIM7, SIM8	0.013	0.013	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04
		0.013	0.013	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04
		5E-04	5E-04	0.013	0.013	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04
		5E-04	5E-04	0.013	0.013	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04
		5E-04	5E-04	5E-04	5E-04	0.013	0.013	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04
		5E-04	5E-04	5E-04	5E-04	0.013	0.013	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04
		5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	0.013	0.013	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04
		0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013
		5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	0.013	0.013	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04
		5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	0.013	0.013	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04
		0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013
		5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	0.013	0.013	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04
		5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	0.013	0.013	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04
		5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	0.013	0.013	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04
		5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	0.013	0.013	5E-04	5E-04	5E-04	5E-04	5E-04
		5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	0.013	0.013	5E-04	5E-04	5E-04
		5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	5E-04	0.013	0.013	5E-04

Table 8 (Continued)

Matrix	Model(s)	Violation																		
III	SIM3, SIM5	0.108	0.108	0.108	0.108	0.108	0.108	0.108	0.108	0.108	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
		0.108	0.108	0.108	0.108	0.108	0.108	0.108	0.108	0.108	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
		0.108	0.108	0.108	0.108	0.108	0.108	0.108	0.108	0.108	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
		0.108	0.108	0.108	0.108	0.108	0.108	0.108	0.108	0.108	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
		0.108	0.108	0.108	0.108	0.108	0.108	0.108	0.108	0.108	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
		0.108	0.108	0.108	0.108	0.108	0.108	0.108	0.108	0.108	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
		0.108	0.108	0.108	0.108	0.108	0.108	0.108	0.108	0.108	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
		0.108	0.108	0.108	0.108	0.108	0.108	0.108	0.108	0.108	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
		0.108	0.108	0.108	0.108	0.108	0.108	0.108	0.108	0.108	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
		0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121
		0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121
		0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121
		0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121
		0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121
		0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121
		0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121
		0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121

Table 8 (Continued)

Matrix	Model(s)	Violation																		
III	SIM3, SIM5	0.108	0.108	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
		0.108	0.108	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
		0.108	0.108	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
		0.108	0.108	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
		0.108	0.108	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
		0.108	0.108	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
		0.108	0.108	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
		0.108	0.108	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
		0.108	0.108	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
		0.004	0.004	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121
		0.004	0.004	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121
		0.004	0.004	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121
		0.004	0.004	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121
		0.004	0.004	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121
		0.004	0.004	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121
		0.004	0.004	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121
		0.004	0.004	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121

Table 8 (Continued)

Matrix	Model(s)	Violation																		
III	SIM3, SIM5	0.223	0.223	0.223	0.223	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.007	0.007	0.007	0.007	0.007	0.007	0.007
		0.223	0.223	0.223	0.223	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.007	0.007	0.007	0.007	0.007	0.007	0.007
		0.223	0.223	0.223	0.223	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.007	0.007	0.007	0.007	0.007	0.007	0.007
		0.223	0.223	0.223	0.223	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.007	0.007	0.007	0.007	0.007	0.007	0.007
		0.008	0.008	0.008	0.008	0.223	0.223	0.223	0.223	0.008	0.008	0.008	0.008	0.007	0.007	0.007	0.007	0.007	0.007	0.007
		0.008	0.008	0.008	0.008	0.223	0.223	0.223	0.223	0.008	0.008	0.008	0.008	0.007	0.007	0.007	0.007	0.007	0.007	0.007
		0.008	0.008	0.008	0.008	0.223	0.223	0.223	0.223	0.008	0.008	0.008	0.008	0.007	0.007	0.007	0.007	0.007	0.007	0.007
		0.008	0.008	0.008	0.008	0.223	0.223	0.223	0.223	0.008	0.008	0.008	0.008	0.007	0.007	0.007	0.007	0.007	0.007	0.007
		0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.223	0.223	0.223	0.223	0.007	0.007	0.007	0.007	0.007	0.007	0.007
		0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.223	0.223	0.223	0.223	0.007	0.007	0.007	0.007	0.007	0.007	0.007
		0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.223	0.223	0.223	0.223	0.007	0.007	0.007	0.007	0.007	0.007	0.007
		0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.223	0.223	0.223	0.223	0.007	0.007	0.007	0.007	0.007	0.007	0.007
		0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.184	0.184	0.184	0.184	0.184	0.184	0.184
		0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.184	0.184	0.184	0.184	0.184	0.184	0.184
		0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.184	0.184	0.184	0.184	0.184	0.184	0.184
		0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.184	0.184	0.184	0.184	0.184	0.184	0.184
		0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.184	0.184	0.184	0.184	0.184	0.184	0.184

Table 8 (Continued)

Matrix	Model(s)	Violation																		
III	SIM3, SIM5	0.011	0.136	0.005	0.005	0.011	0.011	0.011	0.136	0.156	0.156	0.136	0.284	0.284	0.284	0.284	0.284	0.284	0.284	0.284
		0.011	0.136	0.005	0.005	0.011	0.011	0.011	0.136	0.156	0.156	0.136	0.284	0.284	0.284	0.284	0.284	0.284	0.284	0.284
		0.011	0.136	0.005	0.005	0.011	0.011	0.011	0.136	0.156	0.156	0.136	0.284	0.284	0.284	0.284	0.284	0.284	0.284	0.284
		0.011	0.136	0.005	0.005	0.011	0.011	0.011	0.136	0.006	0.006	0.136	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011
		0.011	0.136	0.005	0.005	0.011	0.011	0.011	0.136	0.006	0.006	0.136	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011
		0.011	0.136	0.005	0.005	0.011	0.011	0.011	0.136	0.006	0.006	0.136	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011
		0.011	0.136	0.005	0.005	0.011	0.011	0.011	0.136	0.006	0.006	0.136	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011
		0.011	0.005	0.136	0.136	0.011	0.011	0.011	0.005	0.006	0.006	0.005	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011
		0.011	0.005	0.136	0.136	0.011	0.011	0.011	0.005	0.006	0.006	0.005	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011
		0.011	0.005	0.136	0.136	0.011	0.011	0.011	0.005	0.006	0.006	0.005	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011
		0.011	0.005	0.136	0.136	0.011	0.011	0.011	0.005	0.006	0.006	0.005	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011
		0.011	0.005	0.136	0.136	0.011	0.011	0.011	0.005	0.006	0.006	0.005	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011
		0.011	0.005	0.136	0.136	0.011	0.011	0.011	0.005	0.006	0.006	0.005	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011
		0.011	0.005	0.136	0.136	0.011	0.011	0.011	0.005	0.006	0.006	0.005	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011
		0.284	0.005	0.005	0.005	0.284	0.284	0.284	0.005	0.156	0.156	0.005	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011
		0.284	0.005	0.005	0.005	0.284	0.284	0.284	0.005	0.156	0.156	0.005	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011
		0.284	0.005	0.005	0.005	0.284	0.284	0.284	0.005	0.156	0.156	0.005	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011

Table 8 (Continued)

Matrix	Model(s)	Violation																		
III	SIM3, SIM5	0.066	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015
		0.066	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015
		0.066	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015
		0.066	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015
		0.066	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015
		0.066	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015
		0.066	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015
		0.003	0.015	0.015	0.391	0.391	0.391	0.391	0.391	0.391	0.391	0.391	0.391	0.391	0.391	0.391	0.391	0.391	0.391	0.391
		0.066	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015
		0.066	0.391	0.391	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015
		0.003	0.015	0.015	0.391	0.391	0.391	0.391	0.391	0.391	0.391	0.391	0.391	0.391	0.391	0.391	0.391	0.391	0.391	0.391
		0.066	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015
		0.066	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015
		0.066	0.391	0.391	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015
		0.066	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015
		0.066	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015
		0.066	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015

Table 8 (Continued)

Matrix	Model(s)	Violation																		
III	SIM3, SIM5	0.222	0.222	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.011	0.011	0.011	0.011	0.011	0.011	0.015
		0.222	0.222	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.011	0.011	0.011	0.011	0.011	0.011	0.015
		0.009	0.009	0.222	0.222	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.011	0.011	0.011	0.011	0.011	0.011	0.015
		0.009	0.009	0.222	0.222	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.011	0.011	0.011	0.011	0.011	0.011	0.015
		0.009	0.009	0.009	0.009	0.222	0.222	0.009	0.009	0.009	0.009	0.009	0.009	0.011	0.011	0.011	0.011	0.011	0.011	0.015
		0.009	0.009	0.009	0.009	0.222	0.222	0.009	0.009	0.009	0.009	0.009	0.009	0.011	0.011	0.011	0.011	0.011	0.011	0.015
		0.009	0.009	0.009	0.009	0.009	0.009	0.222	0.222	0.009	0.009	0.009	0.009	0.011	0.011	0.011	0.011	0.011	0.011	0.015
		0.222	0.222	0.222	0.222	0.222	0.222	0.222	0.222	0.222	0.222	0.222	0.222	0.282	0.282	0.282	0.282	0.282	0.282	0.387
		0.009	0.009	0.009	0.009	0.009	0.009	0.222	0.222	0.009	0.009	0.009	0.009	0.011	0.011	0.011	0.011	0.011	0.011	0.015
		0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.222	0.222	0.009	0.009	0.011	0.011	0.011	0.011	0.011	0.011	0.015
		0.222	0.222	0.222	0.222	0.222	0.222	0.222	0.222	0.222	0.222	0.222	0.222	0.282	0.282	0.282	0.282	0.282	0.282	0.387
		0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.222	0.222	0.009	0.009	0.011	0.011	0.011	0.011	0.011	0.011	0.015
		0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.222	0.222	0.011	0.011	0.011	0.011	0.011	0.011	0.015
		0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.222	0.222	0.011	0.011	0.011	0.011	0.011	0.011	0.015
		0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.282	0.282	0.011	0.011	0.011	0.011	0.015
		0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.011	0.011	0.282	0.282	0.011	0.011	0.015
		0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.011	0.011	0.011	0.011	0.282	0.282	0.015

Table 8 (Continued)

Matrix	Model(s)	Violation																		
		0.107	0.107	0.107	0.107	0.107	0.107	0.107	0.107	0.107	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
		0.107	0.107	0.107	0.107	0.107	0.107	0.107	0.107	0.107	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
		0.107	0.107	0.107	0.107	0.107	0.107	0.107	0.107	0.107	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
		0.107	0.107	0.107	0.107	0.107	0.107	0.107	0.107	0.107	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
		0.107	0.107	0.107	0.107	0.107	0.107	0.107	0.107	0.107	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
		0.107	0.107	0.107	0.107	0.107	0.107	0.107	0.107	0.107	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
		0.107	0.107	0.107	0.107	0.107	0.107	0.107	0.107	0.107	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
		0.107	0.107	0.107	0.107	0.107	0.107	0.107	0.107	0.107	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
III	SIM2, SIM4	0.107	0.107	0.107	0.107	0.107	0.107	0.107	0.107	0.107	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
		0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.097	0.097	0.097	0.097	0.097	0.097	0.097	0.097	0.097	0.097
		0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.097	0.097	0.097	0.097	0.097	0.097	0.097	0.097	0.097	0.097
		0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.097	0.097	0.097	0.097	0.097	0.097	0.097	0.097	0.097	0.097
		0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.097	0.097	0.097	0.097	0.097	0.097	0.097	0.097	0.097	0.097
		0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.097	0.097	0.097	0.097	0.097	0.097	0.097	0.097	0.097	0.097
		0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.097	0.097	0.097	0.097	0.097	0.097	0.097	0.097	0.097	0.097
		0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.097	0.097	0.097	0.097	0.097	0.097	0.097	0.097	0.097	0.097
		0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.097	0.097	0.097	0.097	0.097	0.097	0.097	0.097	0.097	0.097

Table 8 (Continued)

Matrix	Model(s)	Violation																		
III	SIM2, SIM4	0.392	0.392	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013
		0.392	0.392	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013
		0.392	0.392	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013
		0.392	0.392	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013
		0.392	0.392	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013
		0.392	0.392	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013
		0.392	0.392	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013
		0.392	0.392	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013
		0.392	0.392	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013
		0.002	0.002	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059
		0.002	0.002	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059
		0.002	0.002	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059
		0.002	0.002	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059
		0.002	0.002	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059
		0.002	0.002	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059
		0.002	0.002	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059
		0.002	0.002	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059

Table 8 (Continued)

Matrix	Model(s)	Violation																		
III	SIM2, SIM4	0.22	0.22	0.22	0.22	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008
		0.22	0.22	0.22	0.22	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008
		0.22	0.22	0.22	0.22	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008
		0.22	0.22	0.22	0.22	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008
		0.008	0.008	0.008	0.008	0.22	0.22	0.22	0.22	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008
		0.008	0.008	0.008	0.008	0.22	0.22	0.22	0.22	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008
		0.008	0.008	0.008	0.008	0.22	0.22	0.22	0.22	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008
		0.008	0.008	0.008	0.008	0.22	0.22	0.22	0.22	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008
		0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.22	0.22	0.22	0.22	0.008	0.008	0.008	0.008	0.008	0.008	0.008
		0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.22	0.22	0.22	0.22	0.008	0.008	0.008	0.008	0.008	0.008	0.008
		0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.22	0.22	0.22	0.22	0.008	0.008	0.008	0.008	0.008	0.008	0.008
		0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.22	0.22	0.22	0.22	0.008	0.008	0.008	0.008	0.008	0.008	0.008
		0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.134	0.134	0.134	0.134	0.134	0.134	0.134
		0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.134	0.134	0.134	0.134	0.134	0.134	0.134
		0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.134	0.134	0.134	0.134	0.134	0.134	0.134
		0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.134	0.134	0.134	0.134	0.134	0.134	0.134
		0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.134	0.134	0.134	0.134	0.134	0.134	0.134

Table 8 (Continued)

Matrix	Model(s)	Violation																		
III	SIM2, SIM4	0.003	0.076	0.003	0.003	0.003	0.003	0.003	0.076	0.076	0.076	0.076	0.076	0.076	0.076	0.076	0.076	0.076	0.076	0.076
		0.003	0.076	0.003	0.003	0.003	0.003	0.003	0.076	0.076	0.076	0.076	0.076	0.076	0.076	0.076	0.076	0.076	0.076	0.076
		0.003	0.076	0.003	0.003	0.003	0.003	0.003	0.076	0.076	0.076	0.076	0.076	0.076	0.076	0.076	0.076	0.076	0.076	0.076
		0.01	0.278	0.01	0.01	0.01	0.01	0.01	0.278	0.01	0.01	0.278	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
		0.01	0.278	0.01	0.01	0.01	0.01	0.01	0.278	0.01	0.01	0.278	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
		0.01	0.278	0.01	0.01	0.01	0.01	0.01	0.278	0.01	0.01	0.278	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
		0.01	0.278	0.01	0.01	0.01	0.01	0.01	0.278	0.01	0.01	0.278	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
		0.014	0.014	0.38	0.38	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014
		0.014	0.014	0.38	0.38	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014
		0.014	0.014	0.38	0.38	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014
		0.014	0.014	0.38	0.38	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014
		0.014	0.014	0.38	0.38	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014
		0.014	0.014	0.38	0.38	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014
		0.014	0.014	0.38	0.38	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014
		0.154	0.006	0.006	0.006	0.154	0.154	0.154	0.006	0.154	0.154	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006
		0.154	0.006	0.006	0.006	0.154	0.154	0.154	0.006	0.154	0.154	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006
		0.154	0.006	0.006	0.006	0.154	0.154	0.154	0.006	0.154	0.154	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006

Table 8 (Continued)

Matrix	Model(s)	Violation																		
III	SIM2, SIM4	0.6	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022
		0.6	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022
		0.6	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022
		0.6	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022
		0.6	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022
		0.6	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022
		0.6	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022
		0.002	0.002	0.002	0.062	0.062	0.062	0.062	0.062	0.062	0.062	0.062	0.062	0.062	0.062	0.062	0.062	0.062	0.062	0.062
		0.6	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022
		0.278	0.278	0.278	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
		0.002	0.002	0.002	0.062	0.062	0.062	0.062	0.062	0.062	0.062	0.062	0.062	0.062	0.062	0.062	0.062	0.062	0.062	0.062
		0.6	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022
		0.6	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022
		0.278	0.278	0.278	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
		0.6	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022
		0.6	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022
		0.6	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022

Table 8 (Continued)

Matrix	Model(s)	Violation																		
III	SIM2, SIM4	0.375	0.375	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015
		0.375	0.375	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015
		0.015	0.015	0.375	0.375	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015
		0.015	0.015	0.375	0.375	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015
		0.015	0.015	0.015	0.015	0.375	0.375	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015
		0.015	0.015	0.015	0.015	0.375	0.375	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015
		0.015	0.015	0.015	0.015	0.015	0.015	0.375	0.375	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015
		0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.053
		0.015	0.015	0.015	0.015	0.015	0.015	0.375	0.375	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015
		0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.375	0.375	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015
		0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.053
		0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.375	0.375	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015
		0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.375	0.375	0.015	0.015	0.015	0.015	0.015	0.015	0.015
		0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.375	0.375	0.015	0.015	0.015	0.015	0.015	0.015	0.015
		0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.375	0.375	0.015	0.015	0.015	0.015	0.015
		0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.375	0.375	0.015	0.015	0.015
		0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.375	0.375	0.015
IV	SIM1, SIM6, SIM7, SIM8	0.048	0.048	0.048	0.048	0.002	0.002	0.002	0.002											
		0.048	0.048	0.048	0.048	0.002	0.002	0.002	0.002											
		0.048	0.048	0.048	0.048	0.002	0.002	0.002	0.002											
		0.002	0.002	0.002	0.002	0.048	0.048	0.048	0.048											
		0.002	0.002	0.002	0.002	0.048	0.048	0.048	0.048											
IV	SIM1, SIM6, SIM7, SIM8	0.071	0.071	0.071	0.003	0.003	0.003	0.003	0.003											
		0.071	0.071	0.071	0.003	0.003	0.003	0.003	0.003											
		0.003	0.003	0.003	0.071	0.071	0.003	0.003	0.003											
		0.003	0.003	0.003	0.071	0.071	0.003	0.003	0.003											
		0.003	0.003	0.003	0.003	0.003	0.071	0.071	0.071											

Table 8 (Continued)

Matrix	Model(s)	Violation							
IV	SIM1, SIM6, SIM7, SIM8	0.076	0.076	0.076	0.076	0.003	0.003	0.003	0.003
		0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003
		0.003	0.003	0.003	0.003	0.076	0.076	0.076	0.076
		0.076	0.076	0.076	0.076	0.003	0.003	0.003	0.003
		0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003
IV	SIM1, SIM6, SIM7, SIM8	0.003	0.003	0.003	0.003	0.063	0.003	0.063	0.063
		0.003	0.003	0.003	0.003	0.063	0.003	0.063	0.063
		0.003	0.063	0.063	0.063	0.003	0.003	0.003	0.003
		0.003	0.003	0.003	0.003	0.063	0.003	0.063	0.063
		0.003	0.003	0.003	0.003	0.063	0.003	0.063	0.063
IV	SIM3, SIM5	0.325	0.325	0.325	0.325	0.019	0.019	0.019	0.019
		0.325	0.325	0.325	0.325	0.019	0.019	0.019	0.019
		0.325	0.325	0.325	0.325	0.019	0.019	0.019	0.019
		0.013	0.013	0.013	0.013	0.472	0.472	0.472	0.472
		0.013	0.013	0.013	0.013	0.472	0.472	0.472	0.472
IV	SIM3, SIM5	0.472	0.472	0.472	0.019	0.019	0.034	0.034	0.034
		0.472	0.472	0.472	0.019	0.019	0.034	0.034	0.034
		0.019	0.019	0.019	0.472	0.472	0.034	0.034	0.034
		0.019	0.019	0.019	0.472	0.472	0.034	0.034	0.034
		0.019	0.019	0.019	0.019	0.019	0.864	0.864	0.864
IV	SIM3, SIM5	0.472	0.472	0.472	0.472	0.034	0.034	0.034	0.034
		0.019	0.019	0.019	0.019	0.034	0.034	0.034	0.034
		0.019	0.019	0.019	0.019	0.864	0.864	0.864	0.864
		0.472	0.472	0.472	0.472	0.034	0.034	0.034	0.034
		0.019	0.019	0.019	0.019	0.034	0.034	0.034	0.034
IV	SIM3, SIM5	0.2	0.035	0.035	0.035	0.248	0.2	0.248	0.248
		0.2	0.035	0.035	0.035	0.248	0.2	0.248	0.248
		0.2	0.862	0.862	0.862	0.01	0.2	0.01	0.01
		0.2	0.035	0.035	0.035	0.248	0.2	0.248	0.248
		0.2	0.035	0.035	0.035	0.248	0.2	0.248	0.248

Table 8 (Continued)

Matrix	Model(s)	Violation							
IV	SIM2, SIM4	0.241	0.241	0.241	0.241	0.01	0.01	0.01	0.01
		0.241	0.241	0.241	0.241	0.01	0.01	0.01	0.01
		0.241	0.241	0.241	0.241	0.01	0.01	0.01	0.01
		0.01	0.01	0.01	0.01	0.241	0.241	0.241	0.241
		0.01	0.01	0.01	0.01	0.241	0.241	0.241	0.241
IV	SIM2, SIM4	0.313	0.313	0.313	0.012	0.012	0.012	0.012	0.012
		0.313	0.313	0.313	0.012	0.012	0.012	0.012	0.012
		0.018	0.018	0.018	0.447	0.447	0.018	0.018	0.018
		0.018	0.018	0.018	0.447	0.447	0.018	0.018	0.018
		0.012	0.012	0.012	0.012	0.012	0.313	0.313	0.313
IV	SIM2, SIM4	0.241	0.241	0.241	0.241	0.01	0.01	0.01	0.01
		0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
		0.01	0.01	0.01	0.01	0.241	0.241	0.241	0.241
		0.241	0.241	0.241	0.241	0.01	0.01	0.01	0.01
		0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
IV	SIM2, SIM4	0.013	0.013	0.013	0.013	0.313	0.013	0.313	0.313
		0.013	0.013	0.013	0.013	0.313	0.013	0.313	0.313
		0.013	0.313	0.313	0.313	0.013	0.013	0.013	0.013
		0.013	0.013	0.013	0.013	0.313	0.013	0.313	0.313
		0.013	0.013	0.013	0.013	0.313	0.013	0.313	0.313
V	SIM1, SIM6, SIM7, SIM8	0.053	0.053	0.053	0.002	0.002	0.002		
		0.053	0.053	0.053	0.002	0.002	0.002		
		0.053	0.053	0.053	0.002	0.002	0.002		
		0.002	0.002	0.002	0.053	0.053	0.053		
		0.002	0.002	0.002	0.053	0.053	0.053		
		0.002	0.002	0.002	0.053	0.053	0.053		

Table 8 (Continued)

Matrix	Model(s)	Violation					
V	SIM1, SIM6, SIM7, SIM8	0.041	0.041	0.041	0.041	0.041	0.041
		0.041	0.041	0.041	0.041	0.041	0.041
		0.041	0.041	0.041	0.041	0.041	0.041
		0.041	0.041	0.041	0.002	0.002	0.002
		0.002	0.002	0.002	0.041	0.041	0.041
		0.002	0.002	0.002	0.002	0.002	0.002
V	SIM1, SIM6, SIM7, SIM8	0.041	0.041	0.041	0.041	0.041	0.041
		0.041	0.041	0.041	0.002	0.002	0.002
		0.041	0.041	0.041	0.041	0.041	0.041
		0.002	0.002	0.002	0.041	0.041	0.041
		0.041	0.041	0.041	0.002	0.002	0.002
		0.002	0.002	0.002	0.041	0.041	0.041
V	SIM1, SIM6, SIM7, SIM8	0.053	0.053	0.053	0.002	0.002	0.002
		0.053	0.053	0.053	0.002	0.002	0.002
		0.002	0.002	0.002	0.053	0.053	0.053
		0.053	0.053	0.053	0.002	0.002	0.002
		0.053	0.053	0.053	0.002	0.002	0.002
		0.053	0.053	0.053	0.002	0.002	0.002
V	SIM3, SIM5	0.321	0.321	0.321	0.013	0.013	0.013
		0.321	0.321	0.321	0.013	0.013	0.013
		0.321	0.321	0.321	0.013	0.013	0.013
		0.013	0.013	0.013	0.321	0.321	0.321
		0.013	0.013	0.013	0.321	0.321	0.321
		0.013	0.013	0.013	0.321	0.321	0.321
V	SIM3, SIM5	0.245	0.245	0.245	0.245	0.245	0.245
		0.245	0.245	0.245	0.245	0.245	0.245
		0.245	0.245	0.245	0.245	0.245	0.245
		0.245	0.245	0.245	0.01	0.01	0.01
		0.01	0.01	0.01	0.245	0.245	0.245
		0.01	0.01	0.01	0.01	0.01	0.01

Table 8 (Continued)

Matrix	Model(s)	Violation					
V	SIM3, SIM5	0.245	0.245	0.245	0.245	0.245	0.245
		0.245	0.245	0.245	0.01	0.01	0.01
		0.245	0.245	0.245	0.245	0.245	0.245
		0.01	0.01	0.01	0.245	0.245	0.245
		0.245	0.245	0.245	0.01	0.01	0.01
		0.01	0.01	0.01	0.245	0.245	0.245
V	SIM3, SIM5	0.321	0.321	0.321	0.013	0.013	0.013
		0.321	0.321	0.321	0.013	0.013	0.013
		0.013	0.013	0.013	0.321	0.321	0.321
		0.321	0.321	0.321	0.013	0.013	0.013
		0.321	0.321	0.321	0.013	0.013	0.013
		0.321	0.321	0.321	0.013	0.013	0.013
V	SIM2, SIM4	0.321	0.321	0.321	0.013	0.013	0.013
		0.321	0.321	0.321	0.013	0.013	0.013
		0.321	0.321	0.321	0.013	0.013	0.013
		0.013	0.013	0.013	0.321	0.321	0.321
		0.013	0.013	0.013	0.321	0.321	0.321
		0.013	0.013	0.013	0.321	0.321	0.321
V	SIM2, SIM4	0.167	0.167	0.167	0.167	0.167	0.167
		0.167	0.167	0.167	0.167	0.167	0.167
		0.167	0.167	0.167	0.167	0.167	0.167
		0.321	0.321	0.321	0.013	0.013	0.013
		0.013	0.013	0.013	0.321	0.321	0.321
		0.167	0.167	0.167	0.167	0.167	0.167
V	SIM2, SIM4	0.167	0.167	0.167	0.167	0.167	0.167
		0.321	0.321	0.321	0.013	0.013	0.013
		0.167	0.167	0.167	0.167	0.167	0.167
		0.013	0.013	0.013	0.321	0.321	0.321
		0.321	0.321	0.321	0.013	0.013	0.013
		0.013	0.013	0.013	0.321	0.321	0.321

Table 8 (Continued)

Matrix	Model(s)	Violation					
V	SIM2, SIM4	0.321	0.321	0.321	0.013	0.013	0.013
		0.321	0.321	0.321	0.013	0.013	0.013
		0.013	0.013	0.013	0.321	0.321	0.321
		0.321	0.321	0.321	0.013	0.013	0.013
		0.321	0.321	0.321	0.013	0.013	0.013
		0.321	0.321	0.321	0.013	0.013	0.013

— APPENDIX E —

ASSUMPTION VIOLATIONS FOR THE MODELS NOT CONSIDERED BY GOTELLI (2000)

This appendix lists the assumption violations that I considered for each of the models not considered in Gotelli (2000). As in Appendix D, the assumptions are divided into “limited” (Table 9) and “full” sets (Table 10).

Table 9. Assumption violations belonging to the "limited set" for the models other than those of Gotelli (2000).

Nominal Probabilities	Assumption Violations[1]		
Equiprobable Rows, Equiprobable Columns	T_j/N	S_i/N	$\sim S_i T_j/N$
Equiprobable Rows, Proportional Columns	N/RC	S_i/N	$\sim S_i T_j/N$
Proportional Rows, Equiprobable Columns	N/RC	T_j/N	$\sim S_i T_j/N$
Proportional Rows, Proportional Columns	N/RC	T_j/N	S_i/N

[1] Variables defined in Figure 5.

Table 10. Assumption violations belonging only to the “full set” for the models other than those of Gotelli (2000).

Matrix	Model(s)[1]	Violation											
I	All but Fr, Fc	0.88	0.88	0.88	0.88	0.88	0.04	0.04	0.04	0.04	0.04		
		0.88	0.88	0.88	0.88	0.88	0.04	0.04	0.04	0.04	0.04		
		0.88	0.88	0.88	0.88	0.88	0.04	0.04	0.04	0.04	0.04		
		0.88	0.88	0.88	0.88	0.88	0.04	0.04	0.04	0.04	0.04		
		0.04	0.04	0.04	0.04	0.04	0.88	0.88	0.88	0.88	0.88		
		0.04	0.04	0.04	0.04	0.04	0.88	0.88	0.88	0.88	0.88		
		0.04	0.04	0.04	0.04	0.04	0.88	0.88	0.88	0.88	0.88		
I	All but Fr, Fc	0.16	0.16	0.16	0.82	0.16	0.82	0.16	0.82	0.82	0.16		
		0.16	0.16	0.16	0.82	0.82	0.82	0.16	0.82	0.82	0.16		
		0.82	0.82	0.82	0.16	0.82	0.16	0.16	0.16	0.16	0.16		
		0.16	0.16	0.16	0.82	0.16	0.82	0.82	0.82	0.82	0.82		
		0.16	0.16	0.16	0.16	0.82	0.16	0.82	0.16	0.16	0.82		
		0.16	0.16	0.16	0.82	0.16	0.82	0.16	0.82	0.82	0.16		
		0.16	0.16	0.16	0.82	0.82	0.82	0.16	0.82	0.82	0.16		
II	All but Fr, Fc	0.83	0.83	0.83	0.83	0.83	0.83	0.03	0.03	0.03	0.03	0.03	0.03
		0.83	0.83	0.83	0.83	0.83	0.83	0.03	0.03	0.03	0.03	0.03	0.03
		0.83	0.83	0.83	0.83	0.83	0.83	0.03	0.03	0.03	0.03	0.03	0.03
		0.03	0.03	0.03	0.03	0.03	0.03	0.83	0.83	0.83	0.83	0.83	0.83
		0.03	0.03	0.03	0.03	0.03	0.03	0.83	0.83	0.83	0.83	0.83	0.83
		0.03	0.03	0.03	0.03	0.03	0.03	0.83	0.83	0.83	0.83	0.83	0.83

[1] Abbreviations for tests defined in Table 4. * and ** denote the following tests:

* Er, Pe; Fr, Uc; C
Er, Pc; Fr, Uc; C
Er, Pe; Ur, Fc; C
Er, Pc; Ur, Fc; C
Pr, Pc; Ur, Fc; C
Er, Pe; Fr, Uc; Ch
Er, Pc; Fr, Uc; Ch
Er, Pe; Ur, Fc; Ch
Er, Pc; Ur, Fc; Ch
Pr, Pc; Ur, Fc; Ch
Er, Pe; Fr, Uc; Co
Er, Pc; Fr, Uc; Co
Er, Pe; Ur, Fc; Co
Er, Pc; Ur, Fc; Co
Pr, Pc; Ur, Fc; Co
Er, Pe; Fr, Uc; V
Er, Pc; Fr, Uc; V
Er, Pe; Ur, Fc; V
Er, Pc; Ur, Fc; V
Pr, Pc; Ur, Fc; V

** Er, Pe; Fr, Uc; C
Er, Pc; Fr, Uc; C
Er, Pe; Fr, Uc; Ch
Er, Pc; Fr, Uc; Ch
Er, Pe; Fr, Uc; Co
Er, Pc; Fr, Uc; Co
Er, Pe; Fr, Uc; V
Er, Pc; Fr, Uc; V

Table 10 (Continued)

Matrix	Model(s)	Violation																		
III	All but Fr, Fc; Fr, Uc	0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013
		0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013
		0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013
		0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013
		0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013
		0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013
		0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013
		0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013
		0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013
		0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.328
		0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.328
		0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.328
		0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.328
		0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.328
		0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.328
		0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.328
		0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.328	0.328

Table 10 (Continued)

Matrix	Model(s)	Violation																		
III	All but Fr, Fc; Fr, Uc	0.342	0.342	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014
		0.342	0.342	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014
		0.342	0.342	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014
		0.342	0.342	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014
		0.342	0.342	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014
		0.342	0.342	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014
		0.342	0.342	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014
		0.342	0.342	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014
		0.342	0.342	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014
		0.014	0.014	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342
		0.014	0.014	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342
		0.014	0.014	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342
		0.014	0.014	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342
		0.014	0.014	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342
		0.014	0.014	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342
		0.014	0.014	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342
		0.014	0.014	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342

Table 10 (Continued)

Matrix	Model(s)	Violation																		
III	All but Fr, Fc; Fr, Uc	0.594	0.594	0.594	0.594	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024
		0.594	0.594	0.594	0.594	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024
		0.594	0.594	0.594	0.594	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024
		0.594	0.594	0.594	0.594	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024
		0.024	0.024	0.024	0.024	0.594	0.594	0.594	0.594	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024
		0.024	0.024	0.024	0.024	0.594	0.594	0.594	0.594	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024
		0.024	0.024	0.024	0.024	0.594	0.594	0.594	0.594	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024
		0.024	0.024	0.024	0.024	0.594	0.594	0.594	0.594	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024
		0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.594	0.594	0.594	0.594	0.024	0.024	0.024	0.024	0.024	0.024	0.024
		0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.594	0.594	0.594	0.594	0.024	0.024	0.024	0.024	0.024	0.024	0.024
		0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.594	0.594	0.594	0.594	0.024	0.024	0.024	0.024	0.024	0.024	0.024
		0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.594	0.594	0.594	0.594	0.024	0.024	0.024	0.024	0.024	0.024	0.024
		0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.594	0.594	0.594	0.594	0.594	0.594	0.594
		0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.594	0.594	0.594	0.594	0.594	0.594	0.594
		0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.594	0.594	0.594	0.594	0.594	0.594	0.594
		0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.594	0.594	0.594	0.594	0.594	0.594	0.594
		0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.594	0.594	0.594	0.594	0.594	0.594	0.594

Table 10 (Continued)

Matrix	Model(s)	Violation																		
III	All but Fr, Fc; Fr, Uc	0.024	0.594	0.024	0.024	0.024	0.024	0.024	0.594	0.594	0.594	0.594	0.594	0.594	0.594	0.594	0.594	0.594	0.594	0.594
		0.024	0.594	0.024	0.024	0.024	0.024	0.024	0.594	0.594	0.594	0.594	0.594	0.594	0.594	0.594	0.594	0.594	0.594	0.594
		0.024	0.594	0.024	0.024	0.024	0.024	0.024	0.594	0.594	0.594	0.594	0.594	0.594	0.594	0.594	0.594	0.594	0.594	0.594
		0.024	0.594	0.024	0.024	0.024	0.024	0.024	0.594	0.024	0.024	0.594	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024
		0.024	0.594	0.024	0.024	0.024	0.024	0.024	0.594	0.024	0.024	0.594	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024
		0.024	0.594	0.024	0.024	0.024	0.024	0.024	0.594	0.024	0.024	0.594	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024
		0.024	0.594	0.024	0.024	0.024	0.024	0.024	0.594	0.024	0.024	0.594	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024
		0.024	0.024	0.594	0.594	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024
		0.024	0.024	0.594	0.594	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024
		0.024	0.024	0.594	0.594	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024
		0.024	0.024	0.594	0.594	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024
		0.024	0.024	0.594	0.594	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024
		0.024	0.024	0.594	0.594	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024
		0.024	0.024	0.594	0.594	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024
		0.594	0.024	0.024	0.024	0.594	0.594	0.594	0.024	0.594	0.594	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024
		0.594	0.024	0.024	0.024	0.594	0.594	0.594	0.024	0.594	0.594	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024
		0.594	0.024	0.024	0.024	0.594	0.594	0.594	0.024	0.594	0.594	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024

Table 10 (Continued)

Matrix	Model(s)	Violation																		
III	All but Fr, Fc; *	0.889	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036
		0.889	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036
		0.889	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036
		0.889	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036
		0.889	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036
		0.889	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036
		0.889	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036
		0.036	0.036	0.036	0.889	0.889	0.889	0.889	0.889	0.889	0.889	0.889	0.889	0.889	0.889	0.889	0.889	0.889	0.889	0.889
		0.889	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036
		0.889	0.889	0.889	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036
		0.036	0.036	0.036	0.889	0.889	0.889	0.889	0.889	0.889	0.889	0.889	0.889	0.889	0.889	0.889	0.889	0.889	0.889	0.889
		0.889	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036
		0.889	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036
		0.889	0.889	0.889	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036
		0.889	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036
		0.889	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036
		0.889	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036

Table 10 (Continued)

Matrix	Model(s)	Violation																		
III	All but Fr, Fc; **	0.703	0.703	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028
		0.703	0.703	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028
		0.028	0.028	0.703	0.703	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028
		0.028	0.028	0.703	0.703	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028
		0.028	0.028	0.028	0.028	0.703	0.703	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028
		0.028	0.028	0.028	0.028	0.703	0.703	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028
		0.028	0.028	0.028	0.028	0.028	0.028	0.703	0.703	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028
		0.703	0.703	0.703	0.703	0.703	0.703	0.703	0.703	0.703	0.703	0.703	0.703	0.703	0.703	0.703	0.703	0.703	0.703	0.703
		0.028	0.028	0.028	0.028	0.028	0.028	0.703	0.703	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028
		0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.703	0.703	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028
		0.703	0.703	0.703	0.703	0.703	0.703	0.703	0.703	0.703	0.703	0.703	0.703	0.703	0.703	0.703	0.703	0.703	0.703	0.703
		0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.703	0.703	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028
		0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.703	0.703	0.028	0.028	0.028	0.028	0.028	0.028	0.028
		0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.703	0.703	0.028	0.028	0.028	0.028	0.028	0.028	0.028
		0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.703	0.703	0.028	0.028	0.028	0.028	0.028
		0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.703	0.703	0.028	0.028	0.028
		0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.703	0.703	0.028
IV	All but Fr, Fc	0.529	0.529	0.529	0.529	0.021	0.021	0.021	0.021											
		0.529	0.529	0.529	0.529	0.021	0.021	0.021	0.021											
		0.529	0.529	0.529	0.529	0.021	0.021	0.021	0.021											
		0.021	0.021	0.021	0.021	0.529	0.529	0.529	0.529											
		0.021	0.021	0.021	0.021	0.529	0.529	0.529	0.529											
IV	All but Fr, Fc	0.781	0.781	0.781	0.031	0.031	0.031	0.031	0.031											
		0.781	0.781	0.781	0.031	0.031	0.031	0.031	0.031											
		0.031	0.031	0.031	0.781	0.781	0.031	0.031	0.031											
		0.031	0.031	0.031	0.781	0.781	0.031	0.031	0.031											
		0.031	0.031	0.031	0.031	0.031	0.781	0.781	0.781											

Table 10 (Continued)

Matrix	Model(s)	Violation							
IV	All	0.838	0.838	0.838	0.838	0.034	0.034	0.034	0.034
		0.034	0.034	0.034	0.034	0.034	0.034	0.034	0.034
		0.034	0.034	0.034	0.034	0.838	0.838	0.838	0.838
		0.838	0.838	0.838	0.838	0.034	0.034	0.034	0.034
		0.034	0.034	0.034	0.034	0.034	0.034	0.034	0.034
IV	All but Ur, Uc, Ft w ith V	0.028	0.028	0.028	0.028	0.688	0.028	0.688	0.688
		0.028	0.028	0.028	0.028	0.688	0.028	0.688	0.688
		0.028	0.688	0.688	0.688	0.028	0.028	0.028	0.028
		0.028	0.028	0.028	0.028	0.688	0.028	0.688	0.688
		0.028	0.028	0.028	0.028	0.688	0.028	0.688	0.688
V	All but Fr, Fc	0.926	0.926	0.926	0.463	0.463	0.463		
		0.926	0.926	0.926	0.463	0.463	0.463		
		0.926	0.926	0.926	0.463	0.463	0.463		
		0.463	0.463	0.463	0.926	0.926	0.926		
		0.463	0.463	0.463	0.926	0.926	0.926		
		0.463	0.463	0.463	0.926	0.926	0.926		
V	All but Fr, Fc	0.833	0.833	0.833	0.833	0.833	0.833		
		0.833	0.833	0.833	0.833	0.833	0.833		
		0.833	0.833	0.833	0.833	0.833	0.833		
		0.833	0.833	0.833	0.417	0.417	0.417		
		0.417	0.417	0.417	0.833	0.833	0.833		
		0.417	0.417	0.417	0.417	0.417	0.417		
V	All but Fr, Fc	0.947	0.947	0.947	0.947	0.947	0.947		
		0.947	0.947	0.947	0.189	0.189	0.189		
		0.947	0.947	0.947	0.947	0.947	0.947		
		0.189	0.189	0.189	0.947	0.947	0.947		
		0.947	0.947	0.947	0.189	0.189	0.189		
		0.189	0.189	0.189	0.947	0.947	0.947		

Table 10 (Continued)

Matrix	Model(s)	Violation					
V	All but Fr, Fc	0.926	0.926	0.926	0.463	0.463	0.463
		0.926	0.926	0.926	0.463	0.463	0.463
		0.463	0.463	0.463	0.926	0.926	0.926
		0.926	0.926	0.926	0.463	0.463	0.463
		0.926	0.926	0.926	0.463	0.463	0.463
		0.926	0.926	0.926	0.463	0.463	0.463

— SUMMARY[1] —

For seventy years ecologists have debated to what extent competition affects the composition of ecological communities. At one extreme, the "Gleasonian" viewpoint posits that species assemble randomly, assorting by chance and abiotic factors (Gleason 1926). At the other extreme, putative "assembly rules" suggest that competition almost entirely determines the composition of communities (e.g., Diamond 1975). Differentiating between the two viewpoints has been challenging, as experimentation is often impractical and unethical at the spatial scale of communities (Connor and Simberloff 1986).

To compensate for the lack of experimental evidence, the most promising approach has been null model testing. The testing asks what pattern would have been observed in the absence of competitive effects. If the observed pattern differs from the prediction, then the effects are inferred. Central to null model testing is the species "presence-absence matrix," in which each row represents a species while each column represents a site. If species i was observed at site j, then the i,jth entry is a one; otherwise it is a zero. To perform a test, one begins by assuming a null hypothesis reflective of an absence of competitive effects. The null hypothesis is then used to simulate a distribution of presence-absence matrices. If the observed presence-absence matrix (or one more extreme) is sufficiently unlikely to have come from the simulated distribution of matrices, then the null hypothesis is rejected and an effect of competition is inferred. Null model testing is a form of statistical hypothesis testing (Gotelli and Graves 1996).

In Chapters I and II, I show that existing null model tests suffer from two critical flaws. Like all statistical tests, null model tests should be powerful and robust.

---

[1] This summary is an expansion of the abstract. It also contains elements of chapter introductions.

Power refers to the sensitivity of the test, with the most powerful tests rejecting false null hypotheses most frequently. Robustness refers to the distortion of significance levels when assumptions are violated, with robust tests only negligibly affected by violations. Thus, null model tests should be powerful and, if assumptions cannot be independently verified, robust. Chapters I and II show that all existing null model analyses lack adequate power or robustness.

It is the issue of robustness that is particularly troubling. It means that when assumptions cannot be verified – as is usually the case – existing tests will incorrectly indicate competitive effects unacceptably often. Thus, the tests are unreliable. To resolve the problem, in Chapter III, I develop robust, assumption-free tests.

The tests that I develop are based on two premises: First, regardless of how competition acts – evolutionarily or ecologically, and extrinsically or intrinsically – it will reduce the co-occurrence of ecologically similar organisms. Second, ecological similarity can be specified using a hierarchical classification system. “Subunits” of organisms can be classified into “units,” with subunits in the same unit being more ecologically similar to each other than those in different units. Hence, defining the W_{ij} as the event that the *i*th and *j*th subunit to arrive at a community belong to the same unit, it follows that competitive effects predict that $P(W_{13} \mid W_{12}) < P(W_{13} \mid W_{12}^c)$, $P(W_{23} \mid W_{12}) < P(W_{23} \mid W_{12}^c)$, or $P(W_{34} \mid W_{23} \bigcap W_{12}^c) < P(W_{34} \mid W_{23}^c \bigcap W_{12}^c)$, for instance. Thus, an appropriate null hypothesis for testing for the absence of competitive effects specifies equality between all relevant pairs of conditional probabilities. Using random graph theoretic techniques, in Chapter III, I show that this null hypothesis is sufficient to specify a distribution of the partition of subunits into units. Hence, to test robustly for the absence of competitive effects, it is sufficient to check whether the observed and predicted distributions of partitions match.

In Chapter IV, I apply a test from Chapter III, which for simplicity relies on one assumption. For the seven communities that I examine, competitive effects are not indicated, although effects of other interspecific interactions (e.g., facilitation) sometimes are. Overall, the predictions of the null model account for over 95% of the variation in observed frequencies of partitions. Hence, these applications suggest that although interspecific interactions may discernibly affect the composition of communities, those effects are minor.

LITERATURE CITED

Connor, E. F., and D. Simberloff. 1986. Competition, Scientific Method, and Null Models in Ecology. American Scientist **74:**155-162.

Diamond, J. M. 1975. Assembly of species communities. Pages 342-344 *in* M. L. Cody and J. M. Diamond, eds. Ecology and Evolution of Communities. Harvard University Press, Cambridge.

Gleason, H. A. 1926. The individualistic concept of plant association. Bulletin of the Torey Botanical Club **53:**7-26.

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