

Reducing over-reporting of deterministic co-occurrence patterns in biotic communities

Tom M. Fayle^{a,*},¹, Andrea Manica^b,^{1,2}

^a *Insect Ecology Group, University Museum of Zoology Cambridge, University of Cambridge, Downing Street, Cambridge, CB2 3EJ, UK*

^b *Evolutionary Ecology Group, Department of Zoology, University of Cambridge, Cambridge, CB2 3EJ, UK*

ARTICLE INFO

Article history:

Received 21 February 2010
Received in revised form 7 June 2010
Accepted 10 June 2010
Available online 23 July 2010

Keywords:

Type 1 error
Type 2 error
Power
Null model
Competition
Swap algorithm
C-score

ABSTRACT

Null models of species co-occurrence are widely used to infer the existence of various ecological processes. Here we investigate the susceptibility of the most commonly used of these models (the C-score in conjunction with the sequential swap algorithm) to type 1 and type 2 errors. To do this we use simulated datasets with a range of numbers of sites, species and coefficients of variation (CV) in species abundance. We find that this model is particularly susceptible to type 1 errors when applied to large matrices and those with low CV in species abundance. As expected, type 2 error rates decrease with increasing numbers of sites and species, although they increase with increasing CV in species abundance. Despite this, power remains acceptable over a wide range of parameter combinations. The susceptibility of this analytical method to type 1 errors indicates that many previous studies may have incorrectly reported the existence of deterministic patterns of species co-occurrence. We demonstrate that in order to overcome the problem of high type 1 error rates, the number of swaps used to generate null distributions for smaller matrices needs to be increased to over 50,000 swaps (well beyond the 5000 commonly used in published analyses and the 30,000 suggested by Lehsten and Harmand, 2006). We also show that this approach reduces type 1 error rates in real datasets. However, even using this solution, larger datasets still suffer from high type 1 error rates. Such datasets therefore require the use of very large numbers of swaps, which calls for improvements in the most commonly used software. In general, users of this powerful analytical method must be aware that they need surprisingly large numbers of swaps to obtain unbiased estimates of structuring in biotic communities.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Most ecological processes, including competition (e.g. Gotelli et al., 2010), predation (e.g. Bell et al., 2010), parasitism (e.g. Philpott, 2005), mutualism (e.g. Frederickson et al., 2005), and differences or similarities in resource use (e.g. Ellwood et al., 2009; Holdaway and Sparrow, 2006) lead to species not occurring at random with respect to one another. It is common for ecological studies to infer these processes from patterns of species presence and absence across a number of locations. There are several metrics that can be used to describe patterns of co-occurrence, and statistical testing is usually carried out by comparing an observed metric to a distribution of the metric under the null hypothesis (Connor and Simberloff, 1978). The use of such null models of species co-occurrence is on the increase (Appendix A), making

it particularly important to understand their statistical properties.

Despite the existence of several analytical methods for assessing patterns of species co-occurrence, the majority of analyses employ the checkerboard score, or C-score, in conjunction with the sequential swap algorithm (Stone and Roberts, 1990). The C-score counts the number of pairs of species occurring exclusively on separate sites for all pairs of species and sites, while the sequential swap is used to create null matrices while maintaining the number of species per site and the number of occurrences per species. Out of a total of 78 data-driven studies of ecological co-occurrence found during a search of the literature (Appendix A), 54 used the C-score metric, and out of the 51 of these that reported the randomisation algorithm, 40 used the sequential swap. This is due in part to the availability of free software to carry out null model analyses (Gotelli and Entsminger, 2007) for which the C-score and sequential swap (fixed-fixed algorithm) are the most powerful analytical method (Gotelli, 2000).

The studies employing the C-score and sequential swap cover a wide range of ecological concepts and taxonomic groups. For example, the method has been used to study the effects of habi-

* Corresponding author. Tel.: +44 01223 331768.

E-mail addresses: tmf26@cam.ac.uk (T.M. Fayle), am315@cam.ac.uk (A. Manica).

¹ Both these authors contributed equally to this work.

² Tel.: +44 01223 336627.

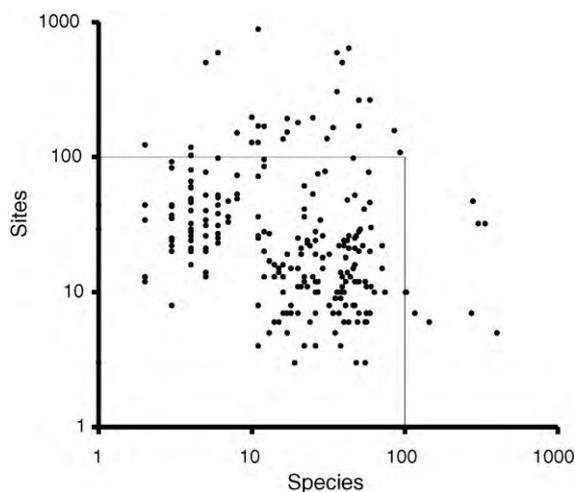


Fig. 1. The distribution of 271 matrix sizes from 47 publications (see supplementary Appendix F for sources). The broken line denotes the range of matrix sizes investigated in this study (10–100 sites, 10–100 species). Previous studies of error rates have used small matrices: Gotelli (2000), 17 species 19 sites and 11 species 25 sites; Lehsten and Harmand (2006), 201/291 matrices with fewer than 50 species and sites. Note the logarithmic scales on the *x* and *y* axes.

tat fragmentation on geckos (Luiselli et al., 2007), the impacts of invasive ant species on native communities (Gotelli and Arnett, 2000), the effects of fire regimes on ungulates (Klop and Goethem, 2008) and to identify areas of species endemism (Giokas and Sfenthourakis, 2008). Gotelli and McCabe (2002) carried out a meta-analysis of published presence/absence matrices using the *C*-score and sequential swap and found that there was less co-occurrence in biological communities than one would expect by chance.

Despite this extensive use, only limited work has been carried out to investigate how susceptible this analytical method is to type 1 and type 2 errors. Gotelli (2000) assessed error rates for a range of metrics and swap algorithms, including the *C*-score and sequential swap, but did so using only a single small data matrix (<30 sites and <30 species) for each type of error. However, the current literature on null model analysis of species co-occurrence includes many analyses of much larger matrices (Fig. 1). Although the sequential swap has been shown to be slightly conservative compared to completely randomized matrices (Lehsten and Harmand, 2006), this was not looked in relation to any properties of the matrices themselves and the matrices used in this analysis were rather small (201/291 with fewer than 50 sites and species). It has been noted that for large matrix sizes (large numbers of species and sites) the *C*-score becomes particularly susceptible to type 1 errors (Ulrich and Gotelli, 2007) but this has not been formally investigated. Nor has there been any work carried out on the effects of variance in species abundance on error rates for this method or indeed any other metric/algorithm combination for analysing patterns of species co-occurrence. It is therefore of vital importance to know how often this widely used method gives false positive results and how powerful it is at detecting trends when applied to a variety of datasets.

Here we investigate type 1 and type 2 error rates for the *C*-score when used with the sequential swap algorithm. We do this by applying it to large numbers of simulated matrices in which species are either randomly assigned to samples (to assess type 1 errors) or where there is some structure to the matrices (to assess type 2 errors). We vary the size of the data matrix (number of species and sites) and the coefficient of variation (CV) in species abundance in order to assess the effects of these parameters on error rates. To counter problems with high type 1 error rates we investigate the effect of increasing the number of swaps used to

generate the null distribution against which the observed value of the metric is compared. This is done for both real and simulated datasets. We also generate and test some extremely large matrices, to see whether this further exacerbates the problems with type 1 errors.

2. Materials and methods

Error rates were tested using computer-generated species/sites matrices of a range of dimensions (10–100 species, 10–100 sites, both in steps of 10). Species abundances in these matrices were log-normally distributed with a range of coefficients of variation (CV: 0, 1, and 10). Species abundances from a wide range of communities follow a log-normal distribution (Magurran, 2004). The mean number of occurrences per row for any matrix was half the number of sites for that matrix, i.e. the fill was 50% for all matrices. Matrix fill is known to have an effect on the outcome of null model analyses of species co-occurrence (Ulrich and Gotelli, 2007). Consequently, we concentrate here on the lesser-known effects of matrix size and CV in species abundance.

Type 2 errors were assessed by applying the null model test to ordered matrices that were then progressively disordered and retested. Ordered matrices (noise = 0) were constructed as “perfect squares” with two sets of species co-occurring exclusively within their own sets (Appendix B). For CVs greater than 0, this was not possible and the equivalent maximally ordered matrices were used (Appendix B). To introduce noise we swapped randomly selected pairs of cells within each row (i.e. we swapped around the sites at which species were observed) following Gotelli (2000). Each pair was only swapped once and the noise of any given matrix is defined as the proportion of cells swapped in each row. For example, for a 20 × 20 matrix, five pairs of cells would be swapped to give a noise level of 0.5. Noise ranged from 0.0 to 0.8 in steps of 0.2 (a noise level of 1.0 being equivalent to the completely random matrices used in testing the type 1 error (see below)). One thousand matrices were generated for each combination of parameters, making a total of 2 million matrices. Each matrix was tested for a significant *C*-score using 5000 (the default for EcoSim) and 30,000 (recommended by Lehsten and Harmand, 2006) randomizations per matrix, fixed column and row sums (as recommended by Ulrich and Gotelli, 2007), and the sequential swap randomization algorithm. We estimated power by recording the number of times $p < 0.05$ for either tail.

Type 1 error rates were assessed by applying the *C*-score test to randomly generated matrices. All matrices were randomly filled with the constraint of following the relevant species abundance distribution. Five thousand matrices were generated for each combination of species/sites/CV giving a total of 2 million matrices. Each matrix was tested using the same settings as for type 2 errors. To give an estimation of the type 1 error, we recorded the number of times $p < 0.05$ for each tail for each combination of parameters. Proportions of matrices falling into the upper and lower tails did not differ (results not presented) and so the combined p -values are presented here. For such a test the type 1 error should therefore be in the region of 0.1.

We tested how the number of randomizations affects type 1 error rates by varying the number of swaps used to generate null distributions of the *C*-score. Specifically, we used 5000 swaps (the default in the widely used software EcoSim (Gotelli and Entsminger, 2007) and the most commonly used number of swaps in published studies (Appendix D)), 30,000 (which has been suggested by Lehsten and Harmand, 2006 as being sufficient for the majority of matrices), and 50,000 (the largest number possible in EcoSim). We also tested ten real data matrices ranging in size from 31 to 93 species and 30 to 147 locations. These were

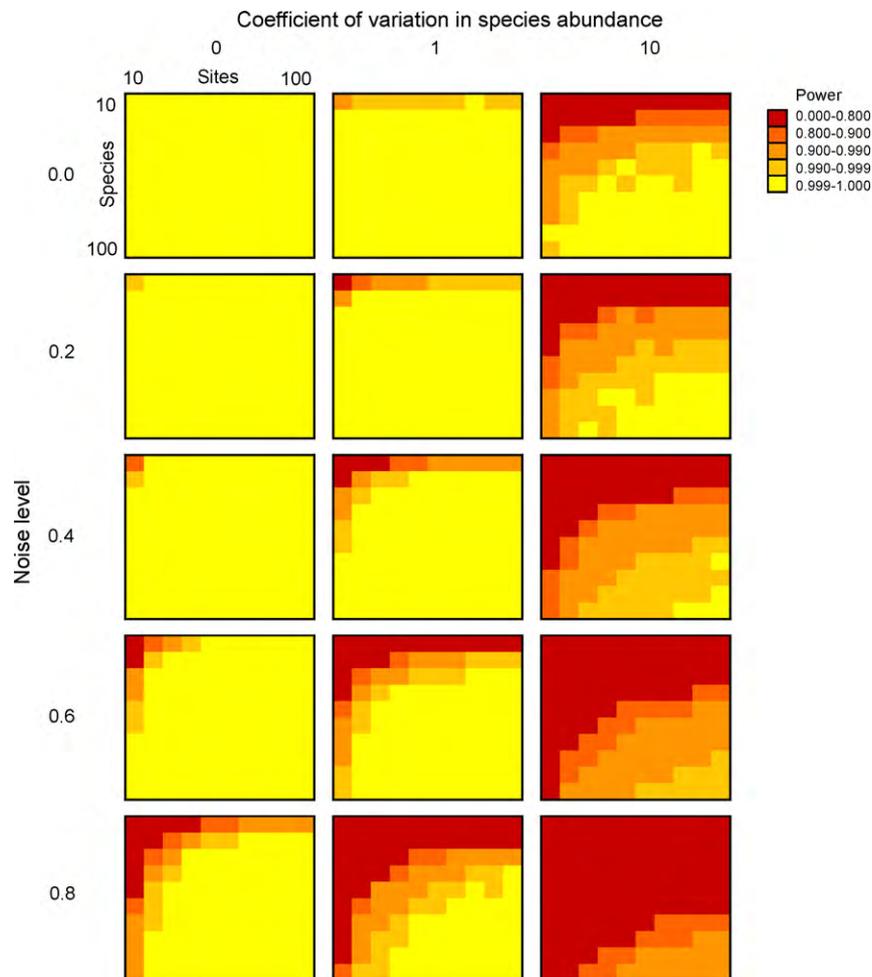


Fig. 2. Type 2 errors incurred by the C-score method of assessing patterns of species co-occurrence for a range of numbers of sites, numbers of species and coefficients of variation in species abundance. Type 2 errors were estimated by progressively randomizing ordered matrices. All errors are given for a critical p -value of 0.1 (two-tailed). 5000 swaps were used to generate null matrices for testing the significance of the metric. Increasing numbers of swaps beyond 5000 had no effect on the power of the tests (Appendix C).

collated from the datasets available with the freeware Nestedness Temperature Calculator Program (Atmar and Patterson, 1993). We used the subset of datasets with more than 30 species distributed across more than 30 sites. In order to test the effect of the number of swaps on type 1 error rates, we used 1000 swaps, following the methods used in a meta-analysis conducted by Gotelli and McCabe (2002), and 50,000 swaps, the largest number possible in EcoSim.

While the majority of analyses in the literature are conducted on matrices within the range of 10–100 sites and 10–100 species (Fig. 1), the analysis of much larger datasets is becoming more common. We tested the type 1 error rates for 1000 randomly generated matrices with 900 sites and 400 species. The largest number of sites in a matrix analyzed using null model techniques found during our literature survey (Fig. 1) was 890 (Helmus et al., 2007) while the largest number of species was 399 (Behnke et al., 2005). A matrix of 900 sites and 400 species therefore represents the maximum size of matrices currently analyzed using null models. We employed an intermediate CV in species abundance of 1, and carried out the analysis once using 30,000 swaps to generate null matrices and once using 50,000 swaps.

Generation and randomization of matrices was carried out with a simple C program run on CamGrid, a distributed computing resource maintained by the University of Cambridge, UK. Analysis of real datasets was carried out in R, using the *oecosimu* function in the *vegan* package (Oksanen et al., 2009).

3. Results

The C-score statistic proved to be very powerful at detecting patterns, even at high noise levels (Fig. 2). As would be expected, power decreased with increasing noise and decreasing numbers of sites and species, but also decreased with increasing CV in species abundance. Power was greater than 0.8 for most combinations of sites and species up to a noise level of 0.6. Even with 80% of the cells in each row randomized, power was still high for larger matrices. The number of swaps had little effect on the power of the test (Fig. 2, Appendix C).

Using 5000 swaps to generate null matrices, the proportion of type 1 errors increased both with number of sites and number of species but decreased with higher CV in species abundance (Fig. 3). As the critical p -value for the two-tailed test is 0.1, the acceptable type 1 error for this test is also 0.1. The region with error of less than 0.1 was restricted to matrices with very small numbers of species and sites. For large matrices, particularly at low CVs in species abundance, type 1 errors were unacceptably large (Fig. 3).

Running the analysis using 30,000 swaps instead of 5000 reduced the type 1 error rate substantially (Fig. 3) without affecting the power of the test (Appendix C). Type 1 error remained acceptable over a much wider range of species/site combinations, while still being lowest at high CV in species abundance. For example, for a matrix with 100 species and 100 locations with a CV in species abundance of 1, type 1 errors decrease from 0.164 with

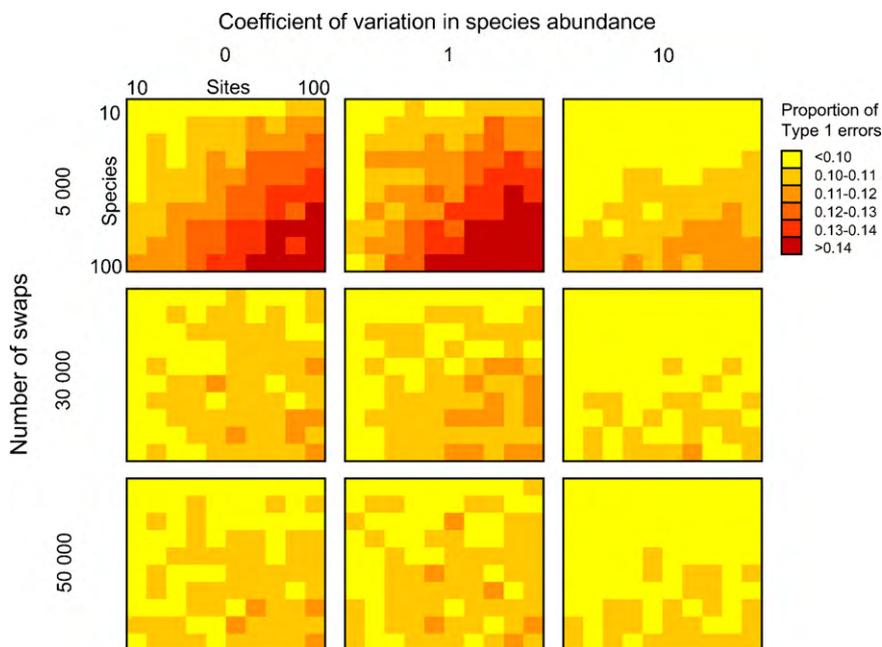


Fig. 3. Type 1 errors incurred by the *C*-score method of assessing species co-occurrence for the same range of matrix sizes and coefficients of variation in species abundance as presented in Fig. 2, but for a range of numbers of swaps used to generate null matrices.

5000 swaps to 0.116 with 30,000 swaps. However, type 1 errors still increased slowly with matrix size, with error rates reaching 0.12 for the largest matrices at CVs in species abundance of 0 and 1. Carrying out the analysis using 50,000 swaps further decreased type 1 error rates, but large matrices still exhibited errors above the 0.1 threshold (Fig. 3).

Analysis of ten real data matrices also showed type 1 errors to be a smaller problem when larger numbers of swaps are used (Fig. 4). For two of the three datasets for which the 50,000 swaps consensus was for non-significance, one of the ten 1000 swap tests on the same datasets returned a significant result. For the other seven datasets for which the 50,000 swaps consensus was for significance all of the 1000 swap tests also returned significant results.

For very large simulated matrices (900 locations, 400 species), type 1 errors were very high. Using 30,000 swaps to generate null distributions of the *C*-score metric, 40.0% of random matrices gave false positive results, while using 50,000 swaps, 32.3% of random matrices gave positive results.

4. Discussion

As expected, the test was more powerful for larger matrices for a given noise level. But the probability of incorrectly detecting a signal in truly random data (type 1 error) also increased with matrix size, confirming the suspicions of Ulrich and Gotelli (2007). This is problematic, as we have the unusual situation in which it is worse to have a larger sample size (i.e. more locations and species). It may be the case that the sequential swap algorithm does not sample the total possible set of matrices with fixed row and sum totals in an unbiased way. Such a bias in the algorithm was noted before by Manly and Sanderson (2002), who claimed that the sequential nature of the randomized matrices produced means that they are not independent and therefore do not sample the entire universe of possible matrices. But Gotelli and Entsminger (2003) were unable to replicate this result, and suggested that this was due in part to an incorrect implementation of the method by Manly and Sanderson (2002) who used means and standard errors of null distributions to test significance values, rather than the distributions themselves. Gotelli and Entsminger (2003) therefore reasserted that the sequential swap is a suitable algorithm to use for the generation of null matrices. However, both Manly and Sanderson (2002) and Gotelli and Entsminger (2003) used only very small matrices (15 species at 15 sites), for which we found type 1 errors to be acceptable in our study. If the randomization algorithm over-sampled the middle ground of possible *C*-score value matrices for larger matrices, this would result in a greater probability of accepting the observed *C*-score as being significant. The bias would be greater the larger the matrix, as the number of iterations carried out by the sequential swap algorithm remains the same, but the number of possible null matrices increases.

For matrices with high CV in species abundance, type 1 errors are less pronounced than in those with low CV, which is reassur-

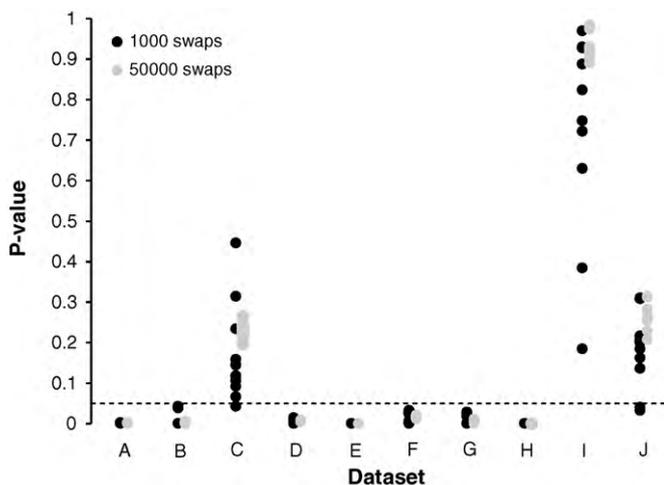


Fig. 4. Null model analyses of ten large real matrices using 1000 swaps and 50,000 swaps. Each analysis was carried out ten times for both numbers of swaps. The broken line indicates the critical *p*-value of 0.05. Datasets from Atmar and Patterson (1993)—A: Amazfish, B: Bajahall, C: Gbfish, D: Newzbird, E: Perthp, F: Polyante, G: Polyantt, H: Qcbirds, I: Sofinbrd, and J: Wausnail.

ing as most real datasets fall into this category. But it is also more difficult to detect real trends in high-CV matrices. This overall tendency towards acceptance of the null hypothesis when dealing with high CVs in species abundances may be due to the smaller number of potential checkerboard pairs in these matrices. In the extreme case, if half of the species were present in all sites and the other half were absent from all sites, the observed number of checkerboard pairs would be zero and all possible randomizations would yield the same result. If most species occur at all or very few sites, then there will be a low probability of getting a significant result, regardless of whether or not a signal is really present in the data. This pattern also indirectly indicates the effects that different species abundance distributions (SADs) will have on statistical error rates. Null model analysis of SADs with high evenness (e.g. the broken stick) will be more susceptible to type 1 errors than those with low evenness (e.g. the log series).

Taxa that are highly species-rich or those that are often sampled across a large number of locations will be particularly susceptible to type 1 errors. In the literature survey that we carried out (Fig. 1), there were nine publications for which the number of species plus the number of sites was greater than 300. These comprised publications involving surveys of plants, ants, fishes, diatoms, parasitic helminths and terrestrial isopods. In addition, studies comparing the species co-occurrence patterns between areas with differing relative abundance distributions will find that areas with high CV in species abundance are less likely to be structured than those with low CV in species abundance. For example, disturbed habitats often support communities with less even relative abundance distributions than those supported by pristine habitats (Hill and Hamer, 1998). Consequently, analyses of the changes in patterns of species co-occurrence in response to habitat conversion should be carried out with an awareness of this problem.

The tendency to reject falsely the null hypothesis can be overcome by using a larger number of swaps to produce the null distribution of the C-score. This has only a minimal effect on the power of the test. The number of swaps needed to give acceptable type 1 error rates for a particular dataset, given its size and CV in species abundance, can be judged by inspection of Fig. 3 or by using the code provided in Appendix G. It is of vital importance that this analytical improvement be implemented, as most analyses currently use only 5000 swaps and there is no overall correlation between the size of matrices and the number of swaps used in analyses (Appendix D).

However, for large datasets, even the large number of swaps investigated in our study (50,000) is not enough: type 1 errors remain unacceptably high. Therefore, we would recommend that when analyzing large datasets, a correspondingly large number of swaps are used. In these cases, the analysis should be repeated with progressively larger numbers of swaps until the *p*-value stabilizes (Appendix E). The current EcoSim software (Gotelli and Entsminger, 2007), does not allow for numbers of swaps larger than 50,000. For such analyses, the R function *oecosimu* (Oksanen et al., 2009) in the *vegan* package can be used, although it is relatively slow due to its implementation in pure R code (as opposed to a pre-compiled Delphi routine). Alternatively, a custom version of the EcoSim software incorporating larger numbers of swaps is currently available on request from its authors and a version will be publicly released in the summer of 2010 (Gary Entsminger, pers. comm.).

5. Conclusions

We have demonstrated that the most widely used method of detecting non-random species co-occurrence is highly vulnerable

to type 1 errors, particularly when dealing with large numbers of sites and species, and with low CVs in species abundance. This indicates that many previous analyses may have reported structure in matrices in which species occur at random with respect to each other. We recommend increasing the number of swaps to over 50,000 in order to overcome this problem and allow the continued use of this powerful community analysis tool.

Acknowledgements

TMF was funded by a Natural Environment Research Council PhD studentship and a Natural History Museum CASE studentship during this work. AM acknowledges support from a British Ecological Society Early Career Grant. We would like to thank Neal Maskell and Nanna Evers for allowing access to extra computing resources, and to Jenny Barna for help in running the code on CamGRID. We are also grateful to William Foster, Paul Eggleton, Gary Entsminger and three anonymous reviewers for helpful comments on the manuscript.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ecolmodel.2010.06.013.

References

- Atmar, W., Patterson, B.D., 1993. The measure of order and disorder in the distribution of species in fragmented habitat. *Oecologia* 96, 373–382.
- Behnke, J.M., Gilbert, F.S., Abu-Madi, M.A., Lewis, J.W., 2005. Do the helminth parasites of wood mice interact? *Journal of Animal Ecology* 74, 982–993.
- Bell, J.R., King, R.A., Bohan, D.A., Symondson, W.O.C., 2010. Spatial co-occurrence networks predict the feeding histories of polyphagous arthropod predators at field scales. *Ecography* 33, 64–72.
- Connor, E.F., Simberloff, D., 1978. Species number and compositional similarity of the Galapagos flora and avifauna. *Ecological Monographs* 48, 219–248.
- Ellwood, M.D.F., Manica, A., Foster, W.A., 2009. Stochastic and deterministic processes jointly structure tropical arthropod communities. *Ecology Letters* 12, 277–284.
- Frederickson, M.E., Greene, M.J., Gordon, D.M., 2005. 'Devil's gardens' bedevilled by ants. *Nature* 437, 495–496.
- Giokas, S., Sfenthourakis, S., 2008. An improved method for the identification of areas of endemism using species co-occurrences. *Journal of Biogeography* 35, 893–902.
- Gotelli, N.J., 2000. Null model analysis of species co-occurrence patterns. *Ecology* 81, 2606–2621.
- Gotelli, N.J., Arnett, A.E., 2000. Biogeographic effects of fire ant invasion. *Ecology Letters* 3, 257–261.
- Gotelli, N.J., Entsminger, G.L., 2003. Swap algorithms in null model analysis. *Ecology* 84, 532–535.
- Gotelli, N.J., Entsminger, G.L., 2007. EcoSim: null models software for ecology. Version 7. Acquired Intelligence Inc. & Kesey-Bear, Jericho, VT 05465.
- Gotelli, N.J., Graves, G.R., Rahbek, C., 2010. Macroecological signals of species interactions in the Danish avifauna. *Proceedings of the National Academy of Sciences* 107, 5030–5035.
- Gotelli, N.J., McCabe, D.J., 2002. Species co-occurrence: a meta-analysis of J.M. Diamond's assembly rules model. *Ecology* 83, 2091–2096.
- Helmus, M.R., Savage, K., Diebel, M.W., Maxted, J.T., Ives, A.R., 2007. Separating the determinants of phylogenetic community structure. *Ecology Letters* 10, 917–925.
- Hill, J.K., Hamer, K.C., 1998. Using species abundance curves as indicators of habitat disturbance in tropical forests. *Journal of Applied Ecology* 35, 458–460.
- Holdaway, R.J., Sparrow, A.D., 2006. Assembly rules operating along a primary riverbed-grassland successional sequence. *Journal of Ecology* 94, 1092–1102.
- Klop, E., Goethem, J.v., 2008. Savanna fires govern community structure of ungulates in Benoue National Park, Cameroon. *Journal of Tropical Ecology* 24, 39–47.
- Lehsten, V., Harmand, P., 2006. Null models for species co-occurrence patterns: assessing bias and minimum iteration number for the sequential swap. *Ecography* 29, 786–792.
- Luiselli, L., Eniang, E.A., Akani, G.C., 2007. Non-random structure of a guild of geckos in a fragmented, human-altered, African rainforest. *Ecological Research* 22, 593–603.
- Magurran, A.E., 2004. *Measuring Biological Diversity*. Blackwell Science Limited, Oxford.
- Manly, B., Sanderson, J.G., 2002. A note on null models: justifying the methodology. *Ecology* 83, 580–582.

- Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Wagner, H., 2009. Vegan: Community Ecology Package. <http://cran.r-project.org/web/packages/vegan/index.html>.
- Philpott, S.M., 2005. Trait-mediated effects of parasitic phorid flies (Diptera: Phoridae) on ant (Hymenoptera: Formicidae) competition and resource access in coffee agro-ecosystems. *Environmental Entomology* 34, 1089–1094.
- Stone, L., Roberts, A., 1990. The checkerboard score and species distributions. *Oecologia* 85, 74–79.
- Ulrich, W., Gotelli, N.J., 2007. Disentangling community patterns of nestedness and species co-occurrence. *Oikos* 116, 2053–2061.
- Web of Science, 2009. ISI Web of Science for UK Education. Thomson Scientific Products, <http://wos.mimas.ac.uk/>.