



## Short communication

## Over-reporting bias in null model analysis: A response to Fayle and Manica (2010)

Nicholas J. Gotelli<sup>a,\*</sup>, Werner Ulrich<sup>b,1</sup><sup>a</sup> University of Vermont, Department of Biology, Burlington, VT 05405 USA<sup>b</sup> Nicolaus Copernicus University in Toruń, Department of Animal Ecology, Gagarina 9, 87-100 Toruń, Poland

## ARTICLE INFO

## Article history:

Received 1 September 2010

Received in revised form 4 November 2010

Accepted 5 November 2010

Available online 9 December 2010

## Key words:

Null model analysis

Randomization test

Sequential swap

Binary presence–absence matrix

Co-occurrence

Nestedness

Fayle and Manica (2010) explored the behavior of the “sequential swap algorithm” in null model analysis. This procedure uses a binary presence–absence matrix (rows = species, columns = sites or species) to test for non-random patterns of species co-occurrence. One standard approach is to compare an observed matrix to a set of random matrices that have the same row and column totals (the “fixed–fixed” model of Gotelli (2000)). Creating a set of such random matrices is challenging. A simple and popular algorithm is to begin with the observed matrix and randomly select two rows and two columns. If the resulting  $2 \times 2$  submatrix is of the form [0101] or [1010], the elements in the submatrix can be swapped, which creates a new matrix that retains the observed row and column totals. If the matrix is repeatedly swapped in this way, a large set of distinct matrices can be obtained that all have the same row and column totals (Connor and Simberloff, 1979).

Fayle and Manica (2010) showed that the results of the analysis are potentially sensitive to the number of swaps used and recommend using at least 50,000 swaps to achieve stable results. In this kind of analysis, it is typical to use a “burn-in” series of initial swaps to remove transient effects. For example, the EcoSim software (Gotelli and Entsminger, 2010) uses 30,000 initial swaps before retaining matrices in its implementation of the sequential swap algorithm. Thus, reported results from EcoSim that use 5000 swaps actually represent the final 5000 swaps in a sequence of

35,000. For matrices that are larger than 30,000 cells, EcoSim uses a burn in series of  $m \times n$  in initial swaps where  $m$  is the number of rows and  $n$  is the number of columns.

Although Fayle and Manica (2010) did not report it in their paper, their analysis also used a burn-in of 30,000 initial swaps (A. Manica, pers. comm.), so their results are directly comparable, for small matrices, to those using the EcoSim algorithm. It is the length of this transient period, rather than the number of swaps per se, that seems to be responsible for the performance of the swap.

Fayle and Manica's (2010) results suggest that the behavior of the model with empirical matrices is not very sensitive to the number of swaps used. For matrices with less than 100 species and 100 sites, Fayle and Manica (2010:2240) report that a co-occurrence analysis using 1000 versus 50,000 swaps gave qualitatively different results ( $p \leq 0.05$  versus  $p > 0.05$ ) in only 2 of 100 trials (Figure 4 in Fayle and Manica (2010)). In those two trials, their small-sample analysis generated  $p$  values that were only marginally significant; such cases should always be treated with caution. For the 3 non-significant matrices in their Figure 4, the Type I error rate would be 2 in 30 (0.067). For this sample of empirical data matrices from the Atmar and Patterson (1993) collection, the estimated frequency of “over-reporting”—the false discovery rate—is  $((2/30) - 1.0) = 0.029$ . Error rates between 0.02 and 0.07 are acceptable and are comparable to those generated by more formal benchmark testing with artificial matrices (Gotelli, 2000; Ulrich and Gotelli, 2007a,b).

Moreover, this is a worst-case analysis because it uses only 1000 swaps. Since 2001, the EcoSim software has used a default setting of 5000 swaps. We re-analyzed the set of empirical matrices used by

\* Corresponding author. Tel.: +1 802 656 0450.

E-mail addresses: [ngotelli@uvm.edu](mailto:ngotelli@uvm.edu) (N.J. Gotelli), [ulrichw@umk.pl](mailto:ulrichw@umk.pl) (W. Ulrich).<sup>1</sup> Tel.: +48 56 611 4469.

Fayle and Manica (2010; their Figure 4) with these default settings in EcoSim. There were no misclassifications, and the results with only 5000 swaps were qualitatively identical to those reported by Fayle and Manica (2010) with 50,000 swaps. So, there is little evidence that any “over-reporting” would have occurred for EcoSim users who retained the default settings during the past 10 years. Gotelli and McCabe’s (2002) meta-analysis of 96 presence–absence matrices from the Atmar and Patterson (1993) compilation used only 1000 swaps, but the results were presented in terms of standardized effect sizes, not critical  $p$ -values.

Several authors have independently demonstrated that the sequential swap does not provide a truly random sample of matrices with the same row and column totals (Zaman and Simberloff, 2002; Miklós and Podani, 2004; Artzy-Randrup and Stone, 2005). Instead, the swap is slightly biased towards finding matrices that are segregated because there are more pathways from swapping that reach these kind of matrices. Lehsten and Harmand (2006) conducted the most thorough analysis of 271 published presence–absence matrices, 96 of which were used in an earlier meta-analysis by Gotelli and McCabe (2002). Lehsten and Harmand (2006) used an unbiased version of the sequential swap with 30,000 swaps and found that in only 5 of 271 cases were the results changed, and in all cases the  $p$  values were close for the corrected and uncorrected swap. The frequency of altered results in their analysis is comparable to Fayle and Manica’s (2010) results in Figure 4 with 50,000 swaps using the biased version of the sequential swap.

One interesting property of the sequential swap is that, no matter how many swaps are used, each matrix in the resulting set differs from the one before it by only 4 matrix elements. As a consequence, there is an inherent serial correlation in the set of matrices created by the sequential swap. It is therefore no surprise that by increasing the number of replicates used, the variability between the runs is reduced (Figure 4 in Fayle and Manica (2010)).

An alternative approach to avoiding serial dependence is to use an “independent swap” algorithm (Gotelli and Entsminger, 2010; Gotelli et al., 2010). In this algorithm, each of the 5000 matrices is created *independently* by taking 30,000 swaps of the initial matrix. In the sequential swap, after 30,000 initial swaps, each consecutive matrix from 5000 swaps is retained. Of course the independent swap is much slower than the sequential swap, but the resulting matrices exhibit no serial correlation, and each random matrix differs by considerably more than 4 elements from the next one generated. The independent swap eliminates much of the between-run variability seen with low replication of the sequential swap and is a standard option (for the past 10 years) in the EcoSim software application (Gotelli and Entsminger, 2010).

Yet another algorithm to reduce serial correlation in swapped matrices is to use a “thinned” series, in which every  $n$ th matrix from a sequential swap series is retained and the intervening  $n - 1$  swapped matrices are discarded. Employing a sampling interval of every  $10 \times R \times C$  matrices ( $R$  is number of matrix rows,  $C$  is number of matrix columns), this is the swapping method used in Ulrich and Gotelli (2007a,b) in their benchmark tests and analyses of empirical matrices.

In our experience, all of these variations make little difference when applied to real matrices of the size typically collected by field ecologists (<100 species and <100 sites). It is very rare to find an empirical matrix for which the resulting  $p$  value differs substantially when using an independent swap, a sequential swap (biased or corrected), a thinned sequential swap, or whether using 1000, 5000, or 50,000 replicates. Thus, there is little evidence for Fayle and Manica’s (2010) claim of “over-reporting” of significant results in previous studies with small ecological matrices.

Using extensive simulations, Fayle and Manica (2010) showed that large random matrices incorrectly reject the null hypothesis

more than 30% of the time with the sequential swap. Ulrich and Gotelli (2007b, Figure 3) previously noted a very similar relationship between matrix size and probability of Type I error for random matrices tested with the fixed–fixed algorithm and reported a type I error rate of 20% for matrix sizes  $k$  above 2500 ( $k = R \times C$ ). Fayle and Manica (2010) consider it an “unusual situation” that larger sample sizes may lead to higher error rates. However, this result is not specific to the sequential swap or even to null model analysis. Instead, it is a general phenomenon that characterizes all frequentist analyses that use a probability value for a formal test of a null hypothesis: with a large enough sample size, the null hypothesis will always be rejected (Anderson et al., 2000)! The sequential swap is not a process-based simulation of community assembly (Ulrich and Gotelli, 2010). Therefore, any empirical distribution—or any simulated distribution other than one generated by the sequential swap algorithm itself—will inevitably diverge from the null model predictions if the sample size is large enough.

The analysis of large data sets is becoming more common in macroecology (Gotelli, 2008) and in emerging subdisciplines such as microbial ecology (Ramette, 2007). For large data sets, some different strategies than traditional null model analysis may be needed. First, the null model analysis assumes that the sites represent random, independent samples of species. However, in many large macroecology data sets, the species occurrence data are derived from contiguous cells in gridded terrestrial maps, which may not represent independent replicates of local assemblages (Gotelli and Graves, 1996). In such cases, random subsampling or aggregation of occurrences from contiguous grid cells will reduce the apparent sample size and may ensure more true independence of the samples. However, data aggregation also changes the spatial grain of the analysis (Rahbek and Graves, 2001), which is important when considering mechanisms of species interaction (Gotelli et al., 2010). A second approach is to tease apart the pattern of non-randomness in a large data set into smaller elements. For example, Gotelli and Ulrich (2010) used an empirical Bayes approach to detect which of the thousands of pairs of species in a co-occurrence analysis are contributing to the overall pattern of non-randomness. Finally, the philosophical stance of parsimony—which is the basis for null model analysis—can be abandoned. For large data sets, some ecologists advocate pattern-oriented modeling (Grimm et al., 2005) to explicitly incorporate “bottom-up” mechanisms of interest in stochastic simulations (Gotelli et al., 2009).

In summary, Fayle and Manica (2010) showed that limited replication in the sequential swap test leads to variable results and perhaps to a small increase in Type I error rates. For large data sets, they showed that the null hypothesis will often be rejected for a randomly assembled matrix. Both findings are expected from elementary statistical principles, and neither is unique to the sequential swap or to null model analysis. For small data sets, the results of previous empirical analyses are robust to variations in the implementation details and replication of the swap algorithm. We agree with Fayle and Manica (2010) that increasing the number of sequential swaps is a useful prescription, but analyses based on 5000 replicates perform just as well as with 50,000. We do not see this as a major issue in null model analysis or as cause for concern about “over-reporting” of significant results.

## References

- Anderson, D.R., Burnham, K.P., Thompson, W.L., 2000. Null hypothesis testing: problems, prevalence, and an alternative. *Journal of Wildlife Management* 64, 912–923.
- Artzy-Randrup, Y., Stone, L., 2005. Generating uniformly distributed random networks. *Physical Review E* 72 (5), 72, 7 pages.
- Atmar, W., Patterson, B.D., 1993. The measure of order and disorder in the distribution of species in fragmented habitat. *Oecologia* 96, 373–382.
- Connor, E.F., Simberloff, D., 1979. The assembly of species communities: chance or competition? *Ecology* 60, 1132–1140.

- Fayle, T.M., Manica, A., 2010. Reducing over-reporting of deterministic co-occurrence patterns in biotic communities. *Ecological Modelling* 221, 2237–2242.
- Gotelli, N.J., Ulrich, W., 2010. The empirical Bayes distribution as a tool to identify non-random species associations. *Oecologia* 162, 463–477.
- Gotelli, N.J., 2000. Null model analysis of species co-occurrence patterns. *Ecology* 81, 2606–2621.
- Gotelli, N.J., 2008. Perspectives in biogeography: hypothesis testing, curve fitting, and data mining in macroecology. *International Biogeography Society Newsletter* 6, 1–7.
- Gotelli, N.J., Anderson, M.J., Arita, H.T., Chao, A., Colwell, R.K., Connolly, S.R., Currie, D.J., Dunn, R.R., Graves, G.R., Green, J.L., Grytnes, J.A., Jiang, Y.-H., Jetz, W., Lyons, S.K., McCain, C.M., Magurran, A.E., Rahbek, C., Rangel, T.F.L.V.B., Soberon, J., Webb, C.O., Willig, M.R., 2009. Patterns and causes of species richness: a general simulation model for macroecology. *Ecology Letters* 12, 873–886.
- Gotelli, N.J., Entsminger, G.L., 2010. EcoSim: Null Models Software for Ecology. Version 7. Acquired Intelligence Inc. & Kesey-Bear, Jericho, VT 05465, <http://garyentsminger.com/ecosim/index.htm>.
- Gotelli, N.J., Graves, G.R., 1996. *Null Models in Ecology*. Smithsonian Institution Press, Washington, DC.
- 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 of America* 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.
- Grimm, V., Revilla, E., Berger, U., Jeltsch, F., Mooij, W.M., Railsback, S.F., Thulke, H.-H., Weiner, J., Wiegand, T., DeAngelis, D.L., 2005. Pattern-oriented modeling of agent-based complex systems: lessons from ecology. *Science* 310, 987–991.
- 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.
- Miklós, I., Podani, J., 2004. Randomization of presence–absence matrices: comments and new algorithms. *Ecology* 85, 86–92.
- Rahbek, C., Graves, G.R., 2001. Multiscale assessment of patterns of avian species richness. *Proceedings of the National Academy of Sciences of the United States of America* 98, 4534–4539.
- Ramette, A., 2007. Multivariate analyses in microbial ecology. *Microbial Ecology* 62, 142–160.
- Ulrich, W., Gotelli, N.J., 2007a. Null model analysis of species nestedness patterns. *Ecology* 88, 1824–1831.
- Ulrich, W., Gotelli, N.J., 2007b. Disentangling community patterns of nestedness and species co-occurrence. *Oikos* 116, 2053–2061.
- Ulrich, W., Gotelli, N.J., 2010. Null model analysis of species associations using abundance data. *Ecology* 91, 3384–3397.
- Zaman, A., Simberloff, D., 2002. Random binary matrices in biogeographical ecology—instituting a good neighbor policy. *Environmental and Ecological Statistics* 9, 405–421.