



## Letter to the Editor

**Bias in null model analyses of species co-occurrence: A response to Gotelli and Ulrich (2011)**
**Keywords:**

Null model  
Species co-occurrence  
Competition  
Type 1 error  
Type 2 error  
Randomisation algorithm  
Sequential swap  
C-score

In a recent article we investigated the statistical error rates of one of the most widespread techniques for assessing the structure of ecological binary presence/absence matrices (Fayle and Manica, 2010). The method involves the calculation of a metric for the observed data set, the C-score, which gives a measure of the degree to which pairs of species tend not to co-occur. In order to assess the probability of obtaining this value of the metric (or one more extreme) under the null hypothesis of species being distributed independently of one-another, a set of null matrices is constructed. Often these null matrices are constrained to have the same number of species per site and the same number of species occurrences per species as the observed matrix. These stringent constraints often lead to difficulties in constructing sufficient numbers of null matrices, which can be overcome to an extent by using a “sequential swap” (Gotelli and Entsminger, 2003). We show that the sequential (non-independent) nature of the null matrices generated means that a surprisingly large number are needed to obtain unbiased estimates of significance values, in particular for large matrices and for those with more even relative abundance distributions. Gotelli and Ulrich (2011), in their response to our article, provide some interesting comments on other approaches that might deal with the bias of the sequential swap algorithm and seem to agree with our overall results, but not with our interpretation of their importance.

Gotelli and Ulrich (2011) present a range of alternative approaches to deal with the bias other than increasing the number of swaps. We appreciate that there are many different options for reducing this bias, but the majority of people do not use them, and advocating an increase in the number of randomisations used seems to be the simplest solution and therefore the one most likely to be adopted. Perhaps as a result of the difficulty in having an intuitive grasp for how many randomisations are required when those randomisations are not independent, studies to date have rarely used more than 5000 swaps, and there seems to be no correlation between size of matrix and number of swaps used (Fayle and Manica, 2010: Appendix D). It should also be noted that other approaches, such as thinning and the use of multiple randomisation chains present similar issues of how many replicates can be taken from each chain, although obviously fewer matrices will be required overall. If the thinning approach is to be used then analyses

can be conducted using the R code provided in Appendix G of our article (Fayle and Manica, 2010), by editing line 575 to read for any required thinning parameter (this is currently set to 1 to emulate the implementation in EcoSim). This code uses the commssimulator function in the R package vegan (Oksanen et al., 2008).

“Burn-in” is certainly a useful method to ensure that the null matrices are sufficiently independent of the starting matrix, but a large burn-in by itself is unlikely to solve the type 1 problem as Gotelli and Ulrich suggest, since sequential matrices are still non-independent. This is demonstrated by the fact that even for small simulated matrices, when using a (presumably sufficiently large) burn-in of 30,000 swaps, increasing the number of null matrices generated still decreases the type 1 error rates (Figure 3; Fayle and Manica, 2010).

However, the main objection by Gotelli and Ulrich (2011) seems to be that the degree of bias we report is not large enough to be a concern in these kinds of analysis and that consequently our claim that previous studies may have over-reported structuring in biotic communities is unfounded. Here we contend that the degree of bias is indeed large enough to be an issue.

The statistic that Gotelli and Ulrich (2011) quote of 2 out of 100 cases in our analysis of real datasets exhibiting false positives is misleading, in part due to an error of ours in the original paper. Seven out of ten datasets were clearly non-random, and thus would not give any risk of over reporting. In the three datasets which seem to be stochastically structured (where over reporting was an issue), two were incorrectly classed as significant in 3 out of 30 runs (not 2 as we incorrectly state in the original article: the points overlap in our Figure 4). If these three non-significant matrices are representative then this results in a type 1 error rate of 10%, which is in the region predicted by our extensive analyses of simulated matrices. Given that the critical significance level used in these tests is 5%, this represents quite a worrying degree of bias. Indeed, this clearly exceeds the upper acceptable rate of 7% as stated by Gotelli and Ulrich (2011). This will result in 1 in 10 real (ecologically unstructured) matrices as being falsely reported as having significant structure, as opposed to 1 in 20 without this bias. However, the proportion of false reporting in published analyses will depend greatly on the proportion of analysed matrices that really do have structure, and the reporting bias between significant and non-significant analyses. This is, of course, true for all statistical analyses, but increases in the rates of type 1 errors are likely to exacerbate the problem. However, our purpose in conducting the analyses of real datasets was not to provide an estimate of the degree of bias, which would be an unrealistic aim with only 10 datasets, and is clearly demonstrated in our analyses of simulated datasets. Instead we wished to demonstrate the mechanism by which the bias occurs at low numbers of swaps, i.e. that increasing variability between runs can lead to some runs becoming significant at low swap numbers where all runs are non-significant at high swap numbers.

Our statement that it is an unusual situation for it to be “worse to have larger sample size” was meant more broadly than merely within the context of error rates. In general, ecologists would be pleased to have a larger sample size, since this decreases type 2 error rates! We agree with Gotelli and Ulrich (2011) (who cite Anderson et al. (2000)) that the problem that this presents is common to all frequentist analyses, but we contend that this is true only in situations where there really is a difference (no matter how small, and lacking in biological meaning) between the null hypothesis and the real data. Indeed, Anderson et al. (2000) themselves state that “One can always reject a null hypothesis with a large enough sample size, even if the true difference is trivially small”. There does need to be a difference, no matter how small, for Gotelli and Ulrich’s statement to be correct. Where there is no difference, i.e. the data are truly drawn from the distribution assumed by the null hypothesis, as they are in our simulations, then this objection is not valid: increasing the sample size will not increase the proportion of type 1 errors.

Gotelli and Ulrich state that the finding that increasing the number of swaps decreases type 1 error rates is expected from elementary statistics. While an increased number of randomisations decreases error rates in general, the key point of our original paper was that, because of the biased nature of the sequential swap algorithm, an unusually high level of replication is required to get good estimates. Most simple randomisation procedures (which perform a uniform, uncorrelated sampling of matrix space) usually require 1 to 10,000 randomisations to get good estimates, making it unsurprising that many authors routinely choose 5000 randomisation for the sequential algorithm. As shown in our original paper and in the analysis presented above, as well as in previous work by other authors, the biased nature of the sequential swap increases the type 1 error above the accepted threshold with these levels of replication. While we do not imply that all previous reports of structuring in communities have to be discounted, as the bias is of moderate size, we still state emphatically that the number of swaps should be raised to avoid the possibility of erroneous conclusions. Gotelli and Ulrich, while claiming that there is “no major issue” about over-reporting, also state that “increasing the number

of sequential swaps is a useful prescription”. So, while we might disagree on how “major” the issue is, we all seem to agree that 5000 swaps are not enough, and that 50,000 (or even more) is a much more suitable level.

## 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.

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., Entsminger, G.L., 2003. Swap algorithms in null model analysis. *Ecology* 84, 532–535.

Gotelli, N.J., Ulrich, W., 2011. Over-reporting bias in null model analysis: a response to Fayle and Manica (2010). *Ecological Modelling*.

Oksanen, J., Kindt, R., Legendre, P., O’Hara, B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Wagner, H., 2008. vegan: Community Ecology Package. R Package Version 1.15-1. <http://cran.r-project.org/>, <http://vegan.r-forge.r-project.org/>.

Tom M. Fayle <sup>a,b,\*</sup>

<sup>a</sup> Entomology Department, Natural History Museum, Cromwell Road, London SW6 5BD, UK

<sup>b</sup> Insect Ecology Group, University Museum of Zoology Cambridge, Downing Street, Cambridge CB2 3EJ, UK

Andrea Manica <sup>1</sup>

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

\* Corresponding author at: Insect Ecology Group, University Museum of Zoology Cambridge, Downing Street, Cambridge CB2 3EJ, UK. Tel.: +44 07816 092365.

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

<sup>1</sup> Tel.: +44 01223 336627.

2 December 2010

Available online 12 February 2011