



Short communication

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

Nicholas J. Gotelli^{a,*}, Werner Ulrich^{b,1}^a University of Vermont, Department of Biology, Burlington, VT 05405 USA^b 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 2×2 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 (

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 used,

- 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. Merns -353(Goteies)-260(associationsersion)- asLsiontp://garyentaasLsi Ecow-14.7(ic5..)JTJas18r-60(Ecow-02.6ety)-2s, ayons, 212(Patter