A Comparison of Randomization-based Multivariate Tests for Ecological Difference

Cameron Hurst
Janet Chaseling
Carla Catterall
Environmental Sciences, Griffith University
Nathan, Qld 4111, Australia
C.Hurst@mailbox.gu.edu.au

1. Introduction

Community ecologists seek to determine whether the assemblage of species that occur in one environmental situation differs from that in another, by comparing two data matrices, that contain abundance values for a set of sampled species at each of some number of replicate sites within each environment. There is a need for multivariate tests for difference that can deal with properties that often characterize ecological assemblage data sets, such as data matrix sparseness, non-normality (Clarke, 1993), heterogeneous group dispersion and associations (Biondini *et al.*, 1988). Multivariate, randomization tests for difference present a possible solution to these problems (Legendre and Anderson, 1999). However, little research has been done on the comparative performance of such techniques. This paper compares the performance of four randomization-based, multivariate techniques in detecting differences among ecological assemblages: (1) Multi-response permutation procedure (MRPP; Mielke *et al.*, 1976); (2) Nonparametric Multivariate Analysis of Variance (NP-MANOVA; Anderson, 2001); (3) Analysis of Similarities (ANOSIM; Clarke, 1993) and (4) distance-based Redundancy Analysis (db-RDA; Legendre and Anderson, 1999). The aspect of performance considered is the power of tests in detecting differences in species turnover (the proportion of species shared among two assemblages).

2. Methods

To compare the four techniques, tests were run on simulated data sets which represented varying degrees of ecological difference, in terms of species turnover. Empirical power (*sensu* Tracey & Khan, 1989) was then calculated. Simulated data sets were generated by resampling from a modified version of a field-based data set comprising the abundances of each of 41 bird species at each of 20 replicate 0.3 ha sites, within large remnants of forest near Brisbane, Australia (see Sewell & Catterall, 1998). To obtain two *base assemblages* (each with 41 species across 20 replicate sites, but differing in the identity of a specified proportion of species) an enlarged species pool was created by generating additional hypothetical species, whose relative abundance and sampling distribution were realistic, resembling those in the source data. The number of additional species generated was the number that would be unique to each of the two base assemblages. The two assemblages were then generated by randomly drawing species (without replacement) and allocating them to one of three groups: those common to both assemblages, those only in the first and those only in the second.

The most extreme level of species turnover difference (represented by Set 1) was set as that observed between the bird assemblages of large and small forest remnant areas (71% of species shared) in Sewell and Catterall's (1998) study. When the tests used in this paper were applied to these data, all gave a highly significant difference ($p < 0.001$). Sets 2 and 3 simulated decreasing levels of species turnover: 78% and 85% of species shared among the two assemblages, respectively. From each base assemblage, 500 simulated datasets were then generated by randomly drawing (with replacement) the abundance of all species at a site; repeated 20 times to give 20 observations (sites). The abundance for each species, at each site, was then replaced by a poisson deviate whose lambda corresponded to this original abundance. The two sets of assemblage observations (each a matrix of 20 sites by 41 species) were then statistically compared using each of the four multivariate tests. The significance of each test statistic was evaluated using 1000 randomizations. Power was calculated as the proportion of simulations (n = 500), in which the null hypothesis of no difference was rejected, with an alpha of 0.05. In all cases dissimilarity among sites was calculated using the Bray-Curtis distance measure (Bray & Curtis, 1957). All simulation and data analysis was conducted using software written in Java™ by the first author.

3. Results and Discussion

For the largest difference between assemblages (71% of species shared) there was little difference in empirical power between the four tests (Table 1). As the assemblages became more compositionally similar, db-RDA and ANOSIM showed a higher power than MRPP and NP-MANOVA. All tests showed reasonably high power (> 0.9) at detecting assemblage differences even when 78% of the species were shared (with a sample size of 20 per assemblage).

Table 1. Level of species turnover associated with each set of simulations and empirical power of the randomization tests for the three levels of species turnover.

Technique (number of simulations)	Set 1(500)	Set 2(500)	Set 3(500)
Species common to both assemblages	29/41 (71%)	32/41 (78%)	35/41 (85%)
Species unique to each assemblage	12/41 (29%)	9/41 (22%)	6/41 (15%)
Power of MRPP	0.98	0.92	0.67
Power of NP-MANOVA	0.98	0.92	0.66
Power of ANOSIM	0.98	0.94	0.73
Power of db-RDA	1.00	0.99	0.88

* Power is the proportion of simulations (n=500) where H_0 of no difference is rejected at $\alpha = 0.05$

MRPP, NP-MANOVA and ANOSIM are all examples of Mantel-type methods. i.e. those which test the hypothesis of multivariate equality of populations through a univariate test for difference using inter-observation (between site) distances (Manly, 1991; Legendre & Anderson, 1999). This may explain why these techniques exhibited approximately the same level of power. The higher power of ANOSIM, when compared to MRPP and NP-MANOVA, may stem from this test's use of rank-transformed distances rather than unmodified distances.

The consistently higher power associated with db-RDA may be due to it being a canonical ordination technique. These methods involve the imposition of a model on an ordination derived from some eigenanalysis-based unconstrained ordination technique. In this case, the model being fit to the ordination scores is a one-way MANOVA. Further work is needed to identify the factors that underlie the higher power of db-RDA. The power of these tests when applied to ecological assemblage data may be strongly affected by other factors, including the type of ecological difference between assemblages, sample size and distance measure used.

REFERENCES

Anderson, M.J. (2001) *Austral Ecology* **26**, 32-46

Biondini, M.E., Mielke, P.W. and Redente, E.F. (1988) *Coenoses* **3**, 155-174.

Bray, J.R. and Curtis, J.T. (1957) *Ecological Monographs* **27**, 325-349.

Clarke, K.R. (1993) *Australian Journal of Ecology* **18**, 117-143.

Legendre, P. and Anderson, M.J. (1999) *Ecological Monographs* **69**, 1-24.

Manly, B.F.J. (1991) *Randomization and Monte Carlo Methods in Biology*, Chapman and Hall.

Mielke, P.W. Berry, K.J., Johnson, E.S. (1976) *Communications in Statistics: Theory and Methods* **5**,1409-24

Sewell, S. R. and Catterall, C.P. (1998) *Wildlife Research* **25**, 41-63

Tracey, D.S. and Khan, K.A. (1989) *Communications in Statistics: Simulation and Computation* **18**, 729-756.

RESUME

Quatre epreuves pour difference utilisant des techniques de randomization multivariable: MRPP, NP-MANOVA, ANOSIM et db-RDA pour la detection de differences entre le changement des especes sout considerees. L'epreuve db-RDA etait la plus forte. Eutre les dufae epreuves ANOSIM etait la plus forte.