

Testing model assumptions, models and predictions of biodiversity

Piers Dunstan, Scott Foster

December 12, 2008

The project's goal is to test models and assumptions used when analysing biological data for biodiversity. Also, subsequent prediction performance of these models can be assessed. The information gained will be an invaluable resource for future modelling / prediction of biodiversity. It is envisaged that the results will be used in multiple present and future projects as they will indicate plausible and / or effective assumptions for modelling biodiversity. One of the central aspects of this proposed work is to empirically assess the effect of sampling, the effect of varying sampling effort, and the effect of replication. This is particularly important and urgent as it has direct ramifications for future survey designs.

The stages of the project are:

1. Develop simulation models of single and multispecies distributions along environmental gradients. Potential methods may include those relating to Minchin 1987. Distributions of species must be able to be generated independently of each other and with species dependences. The distribution of a species will be simulated as function of 'environmental gradients', which will exhibit continuous and threshold patterns. To mimic reality as much as possible the environmental data should be taken from expansive real data.
2. Sampling techniques will be simulated to generate 'samples' of assemblages as would be found in real physical sampling protocols. In every case the 'true' sample will be known (i.e the actual number of individuals that would have been sampled) and hence the effect of sampling can be empirically evaluated.

Multivariate and univariate modelling techniques will be compared using the sampled data and the complete data (the 'truth').

3. The effectiveness of indices of biodiversity can be compared, both as measures of the sampled assemblages and as useful predictors when conditioned with covariates.

4. The assumptions underpinning multivariate descriptions can be assessed and guidelines can be asserted. Previous work has suggested that distance measures (i.e. non-euclidean) are the only metrics that accurately describe patterns in communities (e.g. Faith et al. 1987). This simulation set-up enables the testing of this assertion with appropriately formulated multivariate models.

The project will have links to CERF biodiversity and biodiversity analysis programs conducted in CSIRO (CMAR, CSE, CLW, + potentially others).

Minchin, P.R. 1987. Simulation of multidimensional community patterns: towards a comprehensive model. *Vegetatio* 71:145-156.

Faith D.P, Minchin, P.R and Belbin, L. 1987. Compositional dissimilarity as a robust measure of ecological distance. *Vegetatio* 69: 57-68.