Choosing the Best Similarity Index when Performing Fuzzy Set Ordination on Abundance Data

Richard L. Boyce
Department of Biological Sciences, Northern Kentucky University, Highland Heights, KY 41099

ABSTRACT

Ordination consists of a set of multivariate techniques which reduce multiple variables in a community to a few dimensions, reflecting the most important patterns in a data set. Fuzzy Set Ordination (FSO) is a relatively new technique that has not been evaluated with abundance data. It requires similarity indices that return values between 0, when two sites have no species in common, and 1, when they are alike. After initial screening we selected three indices with this property. Simulated plant community data sets along a single gradient were created by COMPAS. Factors that where varied included beta-diversities, ecological response shapes, trends in sample, total qualitative and quantitative noise. Sites were sampled along a regular sampling grid. Preliminary results indicated that most indices performed at low levels of beta-diversity (R<1). The Yule and Baroni-Urbani & Buser (BUB) indices also performed well at intermediate and high beta-diversity levels, followed by the Horn coefficient. The Yule and BUB indices were by far the best for skewed species responses and noisy data. The BUB index was robust for quantitative noise. A step-across algorithm minimized the curlover effect, a distortion in FSO at high beta-diversity, and improved the fit with the Yule and BUB indices. Thus the Yule and BUB indices are recommended for FSO performed on abundance data.

INTRODUCTION & OBJECTIVES

Fuzzy set ordination (FSO) is a direct gradient analysis technique introduced by Roberts (1986). FSO needs further development before it becomes as useful as methods like canonical correspondence analysis. FSO requires the use of a similarity index (SI), which is the complement of the distance measures used in some eigenanalysis techniques. Results using simulated binary (presence/absence) data sets show that some SIs perform much better than others (Boyce 2000, Boyce and Ellison 2001). Furthermore, FSO suffers from the "curlover effect" (Boyce 1998), a distortion similar to the horseshoe effect encountered in principal components analysis (PCA). A flexible shortest path adjustment (FSPA) (e.g., Williamson 1983, De'ath 1999) has been shown to greatly improve this distortion for binary data.

The objective of this study is to find the best similarity indices for use with abundance data, using simulated vegetation data sets along a single gradient generated by COMPAS, a plant community simulator (Minchin, 1987). FSO was performed using routines written in SAS that are freely available at http://www.nku.edu/~boycer/fso/. Ordinations were evaluated by comparing ecological positions of sampled points against their actual gradient position, using Spearman r to quantify the performance of the ordination. Because the apparent position of sample points near the gradient ends can be pulled in or "curled-over," the mean # of curlover points was also assessed for each similarity index & level of beta diversity.

Table 1. Similarity indices evaluated, with formula, between two sites j and k, based upon species, i=1 to S, where S is the total number of species, and x_{ij} is the abundance of species i at site j. $N_j = \sum_{i=1}^{S} x_{ij}$ and $N_k = \sum_{i=1}^{S} x_{ik}$.

Index Name	Equation
1. Bray-Curtis (Percentage Similarity)	$\frac{2\sum_{i=1}^{S}\min(x_{ij},x_{ik})}{\sum_{i=1}^{S}(x_{ij}+x_{ik})}$
2. Baroni-Urbani & Buser	$\frac{\sum_{i=1}^{S}\min(x_{ij},x_{ik})+\sqrt{\left(\sum_{i=1}^{S}\min(x_{ij},x_{ik})\right)\left(\sum_{i=1}^{S}(\max_j(x_{ij})-\max(x_{ij},x_{ik}))\right)}}{\sum_{i=1}^{S}\max(x_{ij},x_{ik})+\sqrt{\left(\sum_{i=1}^{S}\min(x_{ij},x_{ik})\right)\left(\sum_{i=1}^{S}(\max_j(x_{ij})-\max(x_{ij},x_{ik}))\right)}}$
3. Canberra Metric	$1-\frac{1}{S}\sum_{i=1}^{S}\left(\frac{\lvert x_{ij}-x_{ik}\rvert}{x_{ij}+x_{ik}}\right)$
4. Horn	$\frac{\sum_{i=1}^{S}[(x_{ij}+x_{ik})\log(x_{ij}+x_{ik})]-\sum_{i=1}^{S}(x_{ij}\log x_{ij})-\sum_{i=1}^{S}(x_{ik}\log x_{ik})}{[(N_j+N_k)\log(N_j+N_k)]-N_j\log N_j-N_k\log N_k}$
5. Kulczynski	$0.5\left[\frac{\sum_{i=1}^{S}\min(x_{ij},x_{ik})}{\sum_{i=1}^{S}x_{ij}}+\frac{\sum_{i=1}^{S}\min(x_{ij},x_{ik})}{\sum_{i=1}^{S}x_{ik}}\right]$
6. Modified Chord	$1-\sqrt{1-\frac{\sum_{i=1}^{S}x_{ij}x_{ik}}{\sqrt{\sum_{i=1}^{S}(x_{ij})^2\sum_{i=1}^{S}(x_{ik})^2}}}$
7. Moritsa, Simplified	$\frac{2\sum_{i=1}^{S}x_{ij}x_{ik}}{\left[\left(\sum_{i=1}^{S}x_{ij}^2\right)/N_j^2+\left(\sum_{i=1}^{S}x_{ik}^2\right)/N_k^2\right]N_jN_k}$
8. Ochiai (Stander, Cosine)	$\frac{\sum_{i=1}^{S}x_{ij}x_{ik}}{\sqrt{\sum_{i=1}^{S}x_{ij}^2\sum_{i=1}^{S}x_{ik}^2}}$
9. Roberts	$\frac{\sum_{i=1}^{S}\left[(x_{ij}+x_{ik})\frac{\min(x_{ij},x_{ik})}{\max(x_{ij},x_{ik})}\right]}{\sum_{i=1}^{S}(x_{ij}+x_{ik})}$
10. Ruzicka	$\frac{\sum_{i=1}^{S}\min(x_{ij},x_{ik})}{\sum_{i=1}^{S}\max(x_{ij},x_{ik})}$
11. Similarity Ratio	$\frac{\sum_{i=1}^{S}x_{ij}x_{ik}}{\sum_{i=1}^{S}(x_{ij})^2+\sum_{i=1}^{S}(x_{ik})^2-\sum_{i=1}^{S}x_{ij}x_{ik}}$
12. Yule, Modified	$\frac{\sqrt{\left(\sum\min(x_{ij},x_{ik})\right)\left(\sum(\max_j(x_{ij})-\max(x_{ij},x_{ik}))\right)}+\sqrt{\left(\sum\min(x_{ij},x_{ik})\right)\left(\sum(\max(x_{ij},x_{ik})-x_{ik})\right)}}{\sqrt{\left(\sum\min(x_{ij},x_{ik})\right)\left(\sum(\max_j(x_{ij})-\max(x_{ij},x_{ik}))\right)}+\sqrt{\left(\sum\min(x_{ij},x_{ik})\right)\left(\sum(\max(x_{ij},x_{ik})-x_{ij})\right)}}$

References

Boyce, RL. 1998. Fuzzy set ordination along an ecological gradient on a mountain in Vermont, USA. J. Veg. Sci. 9:191-200.

Boyce, R.L. 2000. Choosing the best similarity index when performing fuzzy set ordination on binary data. ESA 2000 Annual Meeting Abstracts 252.

Boyce, R.L., & P.C. Ellison. 2001. Choosing the best similarity index when performing fuzzy set ordination on binary data. J. Veg. Sci. 12:711-720.

De'ath, G. 1999. Extended dissimilarity: a method of robust estimation of ecological distances from high beta diversity data. Plant Ecol. 144:191-199.

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

Roberts, D.W. 1986. Ordination on the basis of fuzzy set theory. Vegetatio 66:123-131.

Tamás, J., J. Podani, & P. Csontos, P. 2001. An extension of presence/absence coefficients to abundance data: a new look at absence. J. Veg. Sci. 12: 401-410.

Williamson, M.H. 1983. The ordination of incidence data. J. Ecol. 66:911-920.

Unimodal & Symmetric Distribution, No Noise

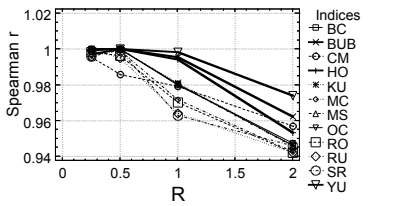

Fig. 1. Mean Spearman r of each similarity index vs. R. R is a measure of beta diversity across the gradient, such that R=1 means the mean species range is the same as the gradient length. Best-performing indices are marked with thick lines. N=3 for each point.

Effect of Step-Across at R=2

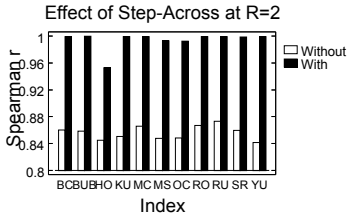

Fig. 2. Effect of step-across algorithm on Spearman r for each similarity index. Note that this improves the Spearman r of most indices to or near 1.0. Best-performing indices are marked with thick lines. N=3 for each point.

Curlover

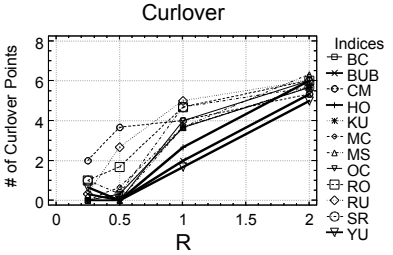

Fig. 3. Mean # of curlover points vs. R. A low # of curlover points is a better fit. Best-performing indices are marked with thick lines. N=3 for each point.

Acknowledgments: This work was funded by a Kentucky NSF EPSCoR Research Enhancement Grant. Gretchen Weisenburger assisted with many of the simulations.

Three-Fold Increase in Species Abundance

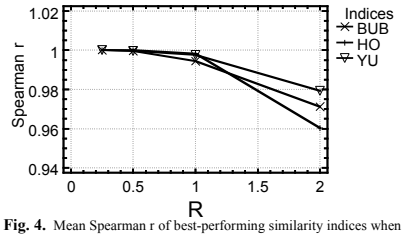

Fig. 4. Mean Spearman r of best-performing similarity indices when species abundance has increased three-fold across the ecological gradient

Skewed (Asymmetric) Species Distributions

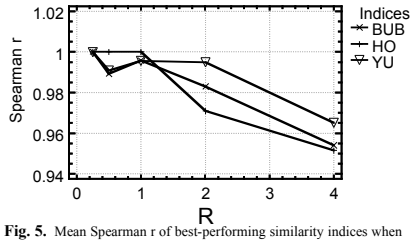

Fig. 5. Mean Spearman r of best-performing similarity indices when species distributions across gradient are skewed or asymmetric.

C0NCLUSIONS

- The modified Yule and the Baroni-Urbani & Buser indices, followed by the Horn index, performed the best at both low & high levels of beta diversity. These indices also perform well when species abundances change across the gradient and when species distributions are skewed. The Yule index performed slightly but consistently better when species turnover is high (R>1).
- These three indices also minimize curlover better than other indices, particularly at R>0.5.
- However, the Horn index does not respond to the step-across algorithm as well as the Yule and Baroni-Urbani & Buser indices. Thus, the Yule and Baroni-Urbani & Buser indices are recommended for use in fuzzy set ordination.
- Data standardization had little effect on the performance of any index (not shown).
- Future directions:
  1) Effect of noise
  2) Performance of synthetic data sets with two gradients