BIOL 540 — Analysis of Ecological Communities

Monday 9:00 Lewis 407, Tuesday-Thursday 9:00-11:00, Lewis 407

Dave Roberts
Ecology Department
117 AJM Johnson Hall
droberts@montana.edu

Course Description

This course is designed to familiarize community ecologists with the analysis of ecological community data. Such data are generated by sampling multiple species at multiple sites, e.g. plant species abundances at sample plots across a range of environments, aquatic invertebrate species (or taxa) abundances from multiple benthic samples, or bird species abundances from pre-dawn surveys across a range of vegetation structures. In each case, there are observations of multiple species and observations of multiple attributes of the environment at each sample unit. Depending on class interest, we may also cover the analysis of individual species distributions. However, this is not a class in experimental design or time series, and is not appropriate for students with structured data requiring ANOVA-like analyses.

This is a practical, hands-on course emphasizing the interpretation of ecological analyses, and covers the majority of multivariate analyses in common use by community ecologists. The course is largely a lab course, and all grades are based on the quality of lab reports submitted; there are no exams. Lab reports are expected to be presented in a formal form with extensive annotated tabular and graphical output from the analyses. Submission of lab reports in HTML or PDF is strongly encouraged. Students are encourage to use RMarkdown or other software like knitr or Sweave to write reports, but other software is certainly acceptable.

We will be using the R computer environment to conduct all analyses. Students will be expected to load their personal computers with a copy of the R package and all necessary material. No prior experience with R is necessary, but general familiarity with computer spreadsheets and data handling is helpful. Students are encouraged to bring their own datasets, but data will be made available to those who need it.

There will be a general lecture/literature review on Mondays, a lab introduction to the software and analyses on Tuesdays, and hands-on analysis Thursdays. There is no text, but readings from the primary literature will be made available.

Modules

- Introduction to the R language
    - obtaining and installing R
    - S language and syntax
    - loading datasets and saving results
    - functions and packages
- Graphical/tabular examination/presentation of multivariate data
    - scatterplots and boxplots
    - multi-layer plots, colors, and glyphs
- Species and community distribution models
    - generalized linear models (GLM)
    - generalized additive models (GAM)
    - classification and regression trees (CART/RANDOM FORESTS)
- Ordination
    - Dissimilarity and distance indices
    - Principal coordinates analysis (PCO)
    - Non-metric multidimensional scaling (NMDS)
    - Correspondence analysis and detrended correspondence analysis (CA and DCA)
    - Canonical correspondence analysis (CCA)
    - Fuzzy set ordination (FSO)
- Cluster analysis and discriminant analysis
    - hierarchical cluster analysis
    - PAM, optpart, and partana

Materials

There are no computer workstations in the lab. Rather, students will be expected to work on their own laptop. You will need to install R and several packages. In recent years many student have chosen to install Rstudio. I'm not a fan, but you're welcome to work in that environment if you choose.

Work Flow

This a demanding course, and requires continual effort. It is critical that you stay up-to-date as new material is presented every week. You must make sure that your data are loaded fully and correctly as soon as possible. You must also be sure to save all your results when you first obtain them to avoid having to repeat analyses. R makes this very easy, but it requires a little discipline on your part.

Lab Reports

You will be graded on the quality of your lab reports. There are no exams. You will write one or two lab reports for each module, depending on the length of that module. I will want to see extensive results, both tabular and graphical, so you will want to make sure you know how to capture and save your results as you go.

Lab reports should be written in a nearly publishable format:

1) abstract (question, general approach, major findings)
2) introduction
    general ecological question of interest
    review of previous results if relevant
3) data
    description of data to be analyzed
4) methods
    description in statistical terms of approach
    description of software packages or functions employed
5) results
    extensive presentation of annotated tabular and graphical output
6) interpretation
    ecological perspective — did you learn anything about the species or communities
    methodological perspective — how well did the analyses work to answer your question, compared to previous methods where relevant.

General References

There is no text for this course. The material on the S language and the R program is available on the web, and will be presented to you on CD-ROM. Statistics texts that are likely to be helpful include:

<u>Statistical Ecology</u>

Legendre, P. and L. Legendre. 2012. Numerical Ecology. Elsevier

Kent, M. 2012. Vegetation Description and Analysis: a Practical Approach. Wiley–Blackwell

McCune, B. and J.B. Grace. 2002. Analysis of Ecological Communities. MJM Press

Manly, B.F.J. 2004. Multivariate statistical methods: a primer. Chapman & Hall

Venables, W.N. and B.D. Ripley. 2002. Modern applied statistics with S. Springer-Verlag

Chambers, J.M. and T.J. Hastie. 1992. Statistical models in S. CRC Press

Digby, P.G.N. and R.A. Kempton. 1987. Multivariate analysis of ecological communities. Chapman & Hall

Pielou, E.C. 1984. The interpretation of ecological data: a primer on classification and ordination. Wiley-Interscience

<u>R</u>

Adler, J. 2012. R in a Nutshell. O'Reilly

Matloff, N. 2011. The Art of R Programming. No Starch Press.

Borcard, D., F. Gillet, and P. Legendre. 2011. Numerical Ecology with R [electronic resource]. Springer

Spector, P. 2008. Data manipulation with R [electronic resource]. Springer.

Everitt, B. and T. Hothorn. An Introduction to Applied Multivariate Analysis with R.

Specific Readings

R Syntax and Functions

Spector. Chap 1. Data in R

Spector. Chap 2. Reading and Writing Data

Spector. Chap 5. Factors

Spector. Chap 6. Subscripting

Everitt and Hothorn. Chap 2. Looking at Multivariate Data Sets

Gradient Response Models

Guisan, A. and Zimmermann, N.E. 2000. Predictive habitat distribution models in ecology. Ecol. Model. 135:147-186.

Austin, M.P., Nicholls, A.O. and Margules, C.R. 1990. Determining the realized qualitative niche: environmental niches of five Eucalyptus species. Ecol. Monogr. 60:135-159.

Yee, T.W. and Mitchell, N.D. 1991. Generalized additive models in plant ecology. J. Veg. Sci. 2:587-602.

Huisman, J., Olff, H. and Fresco, L.F.M. 1993. A hierarchical set of models for species response analysis. J. Veg. Sci. 4:37-46.

Brzeziecki, B., Kienast, F. and Wildi, O. 1993. A simulated map of the potential natural vegetation of Switzerland. J. Veg. Sci. 4:499-508.

Leathwick, J. 1995. Climatic relationships of some New Zealand forest tree species. J. Veg. Sci. 6:237-248.

Austin, M.P., Nicholls, A.O., Doherty, M.D. and Myers, J.A. 1994. Determining species response functions to an environmental gradient by means of a beta-function. J. Veg. Sci. 5:215-228.

Franklin, J. 1995. Predictive vegetation mapping: Geographic modeling of biospatial patterns in relation to environmental gradients. Progress in Phys. Geog. 19:474-499.

Lenihan, J.M. 1993. Ecological response surfaces for North American boreal tree species and their use in forest classification. J. Veg. Sci. 4:667-680.

Tree Classifiers

Franklin, J. 1998. Predicting the distribution of shrub species in southern California from climate and terrain-derived variables. J. Veg. Sci. 9:733-748.

Vayssieres, MP, Plant, RE, Allen-Diaz, BH. 2000. Classification trees: An alternative non-parametric approach for predicting species distributions. J. Veg. Sci. 11:679-694.

Miller, J, Franklin, J. 2002. Modeling the distribution of four vegetation alliances using generalized linear models and classification trees with spatial dependence. Ecol. Model 157:227-247.

Similarity, Dissimilarity, and Distance

Kent, M. 1992. The nature and properties of vegetation data. Chap 4. Pages 101-122 ONLY. Wiley–Blackwell.

Legendre, P. and Legendre, L. 2012. Chapter 7 – Ecological resemblance. In: Legendre, P. and Legendre, L. 1998, Numerical ecology. Elsevier.

McCune, B. and Grace, J.B. 2002. Chapter 6 – Distance Measures. In: McCune. B. and Grace, J.B. Analysis of Ecological Communities. pp. 45-57. MJM Publishers.

Goodall, D.W. 1978a. Sample similarity and species correlation. In: Whittaker, R.H. (ed.) *Ordination of plant communities*, pp. 99-149. Dr. W. Junk, the Hague.

van der Maarel, E. 1979a. Multivariate methods in phytosociology, with reference to the Netherlands. In: Werger, M.J.A. (ed.) *The study of vegetation,* pp. 182-194 ONLY. Dr. W. Junk, the Hague.

van der Maarel, E. 1979b. Transformation of cover-abundance values in phytosociology and its effects on community similarity. *Vegetatio* 39:97-114.

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

Ordination

Kent, M. 2012. Chap. 6 – Ordination methods. Pages 171–271.

Everitt, B. and T. Hothorn. Chaps 3–4. PCA and NMDS.

Borcard, Gillet and Legendre. Unconstrained Ordination. Chap 5

Borcard, Gillet, and Legendre. Canonical Ordination. Chap 6

Hill, M.O. 1980. Detrended correspondence analysis: An improved ordination technique. Vegetatio 42:47-58.

Prentice, I.C. 1980. Vegetation analysis and order invariant gradient models. Vegetatio 42:27-34.

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

Roberts, D.W. 2008. Statistical analysis of multidimensional fuzzy set ordination. Ecology 89:1246-1260.

Ter Braak, C.J.F. 1987. The analysis of vegetation-environment relations by canonical correspondence analysis. Vegetatio 69:69-77.

Wartenberg, D., Ferson, S. and Rohlf, F.J. 1987. Putting things in order: A critique of detrended correspondence analysis. Amer. Nat. 129:434-448.

Peet, R.K., Knox, R.G., Case, J.S. and Allen, R.B. 1988. Putting things in order: The advantage of detrended correspondence analysis. Amer. Nat. 131:924-934.

Minchin, P.R. 1987. An evaluation of the relative robustness of techniques for ecological ordination. Vegetatio 69:89-107.

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

Cluster Analysis

Legendre, P. and Legendre, L. 2012. Chapter 8 – Cluster analysis. In: Legendre, P. and Legendre, L. 1998. Numerical ecology. pp. 337–423. Elsevier.

Kent, M. Numerical classification, cluster analysis, and phytosociology. Chap.

Roberts, D.W. 2015. Vegetation classification by two new iterative reallocation optimization algorithms. Plant Ecology 216(5):741–758.

Aho, K., D.W. Roberts, and T.W.Weaver. 2008. Using geometric and non-geometric internal evaluators to compare eight vegetation classification methods. J. Veg. Sci. 19:549-562.

McCune, B. and Grace, J.B. 2002. Chapter 11 – Hierarchical clustering. In: McCune. B. and Grace, J.B. Analysis of Ecological Communities. pp. 86-96. MJM Publishers.

Goodall, D.W. 1973. Numerical classification. In: Whittaker, R.H. (ed.) *Ordination and classification of communities*, pp. 575-615.

Anderberg, M.R. 1973. Cluster analysis for applications. Academic Press.

Ludwig, J.A. and Reynolds, J.F. 1988. Cluster analysis. Chapter 16 In: Statistical Ecology. John Wiley and Sons.