Journal of Vegetation Science 14: 927-930, 2003

Computer program review

VEGAN, a package of R functions for community ecology

Dixon, Philip

Department of Statistics, Iowa State University, Ames, IA 50011-1210, USA;
Fax +15152944040; E-mail pdixon@iastate.edu

Abstract. VEGAN adds vegetation analysis functions to the general-purpose statistical program R. Both R and VEGAN can be downloaded for free. VEGAN implements several ordination methods, including Canonical Correspondence Analysis and Non-metric Multidimensional Scaling, vector fitting of environmental variables, randomization tests, and various other analyses of vegetation data. It can be used for large data. Graphical output can be customized using the R language's extensive graphics capabilities. VEGAN is appropriate for routine and research use, if you are willing to learn some R.

Keywords: ANOSIM; Mantel test; Multidimensional scaling; Ordination; Procrustes rotation.

Abbreviation: MDS = Multidimensional Scaling.

Description

R is a very powerful general-purpose statistical language that provides excellent graphics and interactive data analysis tools. It is public-domain software with a language and capabilities that are very similar to Splus. Both programs are popular with statisticians because they can be easily expanded by writing functions and packages of functions. Many state-of-the-art statistical techniques are available as add-on packages.

VEGAN, developed by Jari Oksanen, is an add-on package for R that provides useful functions for community ecologists. All the commonly used procedures for vegetation analysis are either included as functions or can be quickly implemented by combining these functions with base R commands. VEGAN provides various ordination methods, including Correspondence Analysis (Reciprocal Averaging), Detrended Correspondence Analysis, Canonical Correspondence Analysis, Principal Components Analysis, Redundancy Analysis, and Non-metric Multidimensional Scaling. Addition results include Procrustes rotations, vector and surface fitting to relate environmental information to ordination plots, and various randomization tests. VEGAN also includes functions for working with distance matrices, including a choice of vegetation distance measures, Mantel tests, and ANOSIM using Clarke's R. Finally, VEGAN provides various utility functions to read data files in Cornell Ecology Program (CANOCO) format, standardize data, write compact species × site tables, and compute diversity measures.

R is an object-oriented language, so most functions in the VEGAN package do not produce output directly; instead, they produce a data structure that stores the results. This output can be manipulated using additional VEGAN or R functions. The output can also be plotted, printed, or saved in an external file.

R provides fantastic graphics facilities with many options. Customizing a graph is easy, once you learn what options or commands you need to use. It is also possible to use the mouse to interact with a graph, for example, to identify unusual or interesting points.

R is free, public-domain software that is available for Linux, Mac and Windows PC operating systems. Add-on packages are also free and in the public domain. Installing R on a Windows PC involves downloading a self-installing binary file from one of the Comprehensive R Archive Network (CRAN) sites (http://cran.r-project.org). Once R is installed, add-on packages are downloaded using the package drop-down menu. One choice on that menu is 'Install package from CRAN'. Click and a long list of available packages (196 in May 2003) appears. Click VEGAN, then click OK and the package is downloaded, unzipped, and installed. The next time you need to use the package, you only need to type library(vegan) or use the 'load package' option from the package menu. It is also easy to download updates, both for the base R distribution and add-on packages. I evaluated VEGAN version 1.4-4, using R version 1.6-2.

R is documented by various .pdf manuals and introductory guides. Alternatively, any of the introductory documentation for Splus (various guides available on the web, or textbooks) will be suitable introductions to R. The VEGAN package is documented by a 42 page .pdf file that describes the purpose, syntax, and details of each function, and gives examples of its use.

On-line documentation includes the help files with syntax, explanation, and examples. Most of the examples can be executed within R by typing example (name of command).

The functions in VEGAN take advantage of the extensive library of methods available in R and other add-on packages. For example, the MASS library in the VR package includes a good non-metric multidimensional scaling (nMDS) algorithm. That is not duplicated in the VEGAN package. Instead, VEGAN provides additional functions, initMDS and postMDS, to use nMDS intelligently on vegetation data, including randomly selecting starting values and post-processing the results. To use all the functions in VEGAN, you need to load four libraries, MASS, akima, mva, and mgcv from three downloadable packages: VR, akima, and mgcv.

R is especially useful for randomization tests. Some functions in VEGAN provides randomization tests upon request. Something more unusual, e.g. the randomization test for trend in species composition proposed by Philippi et al. (1996), can be easily programmed.

R keeps all objects in memory, so the maximum data set size is determined by the available memory. Using a PC with 128Mb of memory, I had no trouble computing a principal components analysis and a non-metric multidimensional scaling on a test data with 500 species and 500 sites. Principal Components Analysis of a larger data set with 1000 species and 1000 sites would not run on the 128Mb machine, but it did run on a PC with 640Mb of memory.

Illustration

To illustrate the use of VEGAN, the code in Table 1 recreates the PCA analysis of species composition for a subset of the sites in the Park Grass experiment at Rothamsted (Digby & Kempton 1987, Fig. 3.6). Fig. 1 is the triplot (site scores, species scores, and regressions of site scores on environmental variables) produced by this code. The code in Table 2 analyses the same data using non-metric Multidimensional Scaling. In addition, contours of one environmental variable (plot yield) are superimposed on the triplot (Fig. 2).

Table 1. VEGAN and R code to load data sets with species composition and environmental data, compute a principal components analysis, and plot a triplot of species scores, sites scores, and environmental variables. Lines starting with # are comments.

```
library(vegan)

# read in the species composition and environmental data
parkg <- read.table('g:/philip/vegdata/park.txt', header=T)
parke <- read.table('g:/philip/vegdata/parkenv.txt', header=T)

# do a PCA on log(abundance+1) transformed data, then plot 1st and 2nd
#axes
#  sites and species scores are plotted by default, using row and column
#names
parkpca <- rda(log(parkg+1))
plot(parkpca)

#regress environmental vars on scores, do randomization test with 999reps.
parkenv <- envfit(parkpca,parke,999)

# print out coefficients and tests of the environmental fits
parkenv

# plot the environmental vectors on the ordination
plot(parkenv,add=T)
# could use arrows() function to connect sites with same fertilizer
#treatments.
```

Table 2. VEGAN and R code to load data sets with species composition and environmental data, compute a non-metric multidimensional scaling in two dimensions, plot species scores, sites scores, and environmental variables, then superimpose yield contours. Lines starting with # are comments.

```
# repeat analysis using Bray-Curtis distance and nMDS
#   assumes that the species and environmental data have been loaded
#   into parkg and parke

# save the species and site names
sitenames <- dimnames(parkg)[[1]]
spnames <- dimnames(parkg)[[2]]

#  isoMDS is the nMDS function from the MASS library, requires the mva
#library
library(MASS)
library(mva)

# compute distances between sites using Bray-Curtis distance
parkdist <- vegdist(parkg,'bray')

#  fit nMDS in 2 dimensions
parkmds <- isoMDS(parkdist, k=2)

# clean up the MDS results
parkmds <- postMDS(parkmds,parkdist)

# plot points, labeling each by the site name
plot(parkmds$points,type='n',asp=1,xlab='Dim 1', ylab='Dim 2')
text(parkmds$points,sitenames)

# regress environmental variables on configuration of site points
parkenv2 <- envfit(parkmds$points,parke,999)

# then add the environmental regressions to the species plot
plot(parkenv2,add=T)

# use weighted averages to compute species scores, then plot them
parkmdssp <- wascores(parkmds$points,parkg)
text(parkmdssp,spnames,col=2)

# fit the surface of productivity, then add it to the plot.
ordisurf(parkmds$points,parkprod,add=T)
```

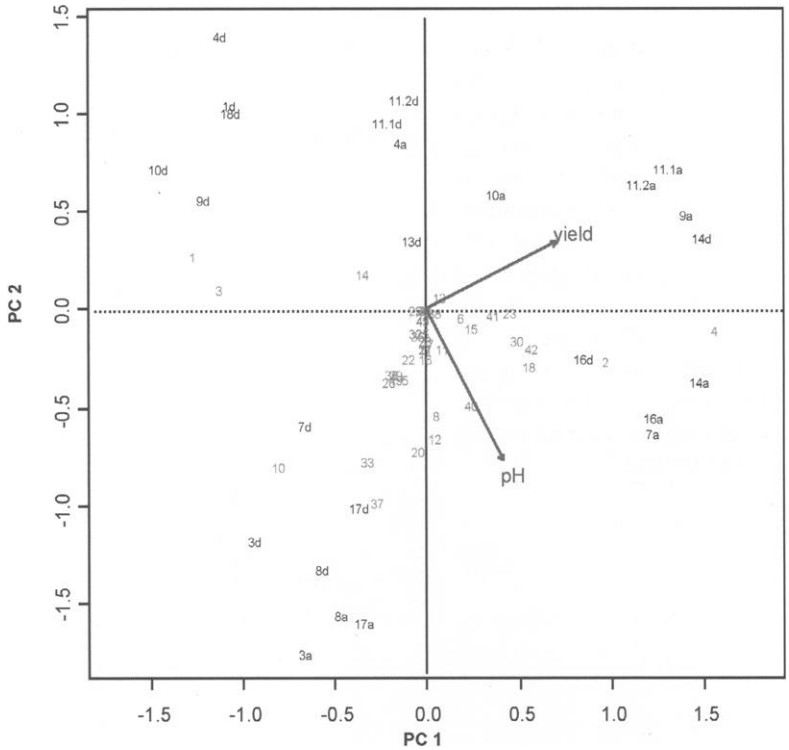

Fig. 1. Triplot from a Principal Components Analysis of a subset of the Park Grass data set. This plot was produced by the code in Table 1. Species are numbered 1 through 44. Sites are labeled as in Digby & Kempton (1987).

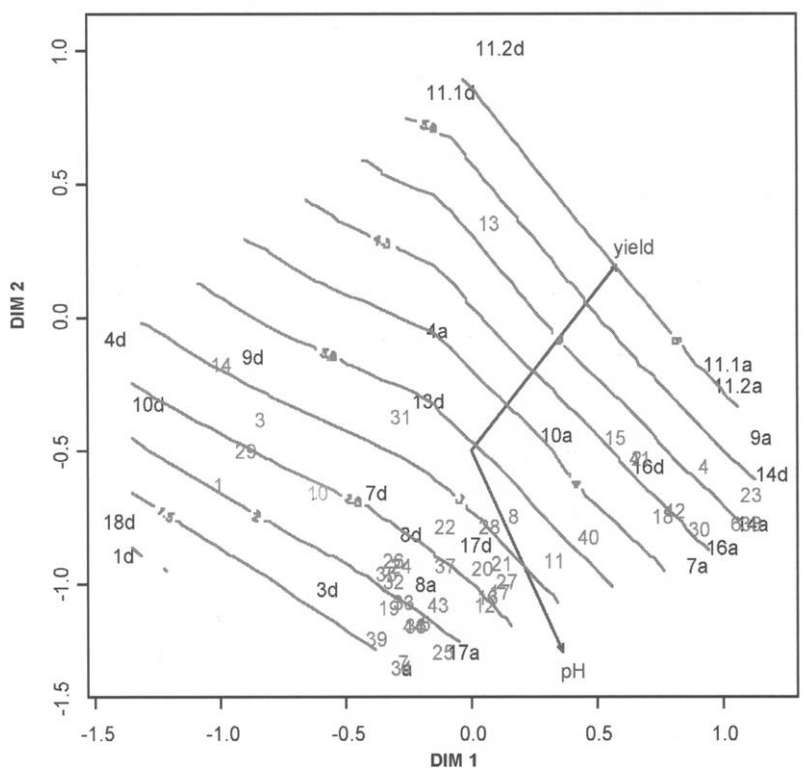

Fig. 2. Triplot and fitted surface of yield from a Non-metric Multidimensional Scaling of a subset of the Park Grass data set. This plot was produced by the code in Table 2. Species are numbered 1 through 44. Sites are labeled as in Digby & Kempton (1987).

Comments

The connection to R is the strongest and weakest part of the VEGAN package. R is incredibly powerful but not especially easy to learn. Neither R nor VEGAN are point-and-click programs. If something goes wrong (e.g. R decides a variable is a factor instead of a numeric vector or you failed to load one of the necessary libraries), the error messages are often cryptic and the solution not obvious. If you are looking for an easy to use 'canned' package or a menu-driven program, R and VEGAN are not for you. However, if you are looking for something that does standard analyses quickly and easily, while providing the flexibility and power to implement non-standard analyses or develop new methods, I highly recommend VEGAN and R.

References

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

Philippi, T.E., Dixon, P.M. & Taylor, B.E. 1998. Detecting trends in species composition. *Ecol. Appl.* 8: 300-308.

Received 27 May 2003;
Revision received 8 September 2003;
Accepted 9 October 2003.
Co-ordinating Editor: M.W. Palmer.