**Nonmetric Multidimensional Scaling (NMDS) Ordination**

**Vegan package in R**

**Katherine Hovanes**

**BIOL 7901 – R seminar – Spring 2013**

Nonmetric Multidimensional Scaling ordination can be used to plot samples in “ecological space” using a community dissimilarity matrix based on species composition. Communities that have very similar species composition will appear as points near each other whereas communities that have very different species compositions will be further away from each other on the plot. Points can be color coded by categorical factors so you can see the grouping of samples based on categorical environmental factors. In order to determine whether samples are “significantly” different by one of the categorical environmental factors you can perform a permutational ANOVA (PERMANOVA) on your presence/absence matrix. I will cover NMDS ordination using the function “metaMDS” (packages vegan and MASS required). Plots of the ordination will be done with the plot function and “ordisymbol” (from package BiodiversityR). The PERMANOVA test is done with the function “adonis” (from vegan). The dataset “KNF\_mod.txt” is the groundcover vegetation species composition of 89 1-m2 plots in the Kisatchie National Forest; this has been modified from the original dataset, so it isn’t “real” data, just something to play with for this tutorial.

There is much more to the vegan package than ordination. Check out this source to see all of the functions in vegan: <http://cran.r-project.org/web/packages/vegan/vegan.pdf>

*First load the necessary libraries; make sure you have downloaded and installed vegan, BiodiversityR, and MASS.*

> #libraries

> library(vegan)

> library(BiodiversityR)

> library(MASS)

*Now read in the data and check the first few rows and columns. You’ll see that the first 5 columns are categorical data about each plot (plot #, soil type, slope position, fire season, and presence vs. absence of the dominant bunchgrass Schizachyrium scoparium). The presence/absence matrix of species composition begins with column 6. There are 182 species in the data set and most of them are rare (they only occur in a handful of the plots).*

> #read in data

> comm=read.table("C:/.../KNF\_mod.txt", header=T)

> comm[1:5,1:10]

plot soil slope fire Ssco ACALGRA ACERRUB ALLICAN AMBRPSI ANDRGER

1 19 p u e p 1 0 0 0 0

2 21 p u e p 0 0 0 0 0

3 22 p m e p 0 0 0 0 0

4 23 p m e p 0 0 0 0 0

5 31 p m e p 0 0 0 0 0

*Now perform the ordination! I am naming it “ord” so I can use it later in with the plot function. For this you enter the name of the data set and exclude the first 5 columns (ordination should only be done on the presence absence matrix which starts with column 6). Then specify the “ecological distance” measure you want to use (distance = …). I am using jaccard distance, but many other options are available; check out specifics in the vegan guide to see what some other options do. Next set the number of dimensions (k=…); I just used 2. Trymax refers to the number of runs R will attempt until it converges on a solution. Set “autotransform” to TRUE so R can transform uncooperative data.*

> ord<-metaMDS(comm[,6:187],distance = "jaccard",k=2,trymax=1000,autotransform=TRUE,expand=FALSE, plot=FALSE)

*This function chooses a random configuration of samples, calculates the pairwise distances between all the points (samples) and then compares it to the actual distances given the community data. The differences between the random configuration distances and the actual distances are the residuals. MetaMDS calculates the Sum of Squares of the residuals for the random configuration and then in the next run changes the configuration to try to reduce the Sum of Squares of the residuals. With every run, R is trying a new configuration of points in order to minimize the Sum of Squares. When it says “new best solution reached” that means it has found a configuration that fits the data better than any of the previous configurations. R will continue to try new random configurations until it converges on a solution. This can take 10 iterations, 100, or 1000; it all depends on the starting random configuration.*

Using step-across dissimilarities:

Too long or NA distances: 800 out of 3916 (20.4%)

Stepping across 3916 dissimilarities...

Run 0 stress 27.20167

Run 1 stress 27.52355

Run 2 stress 26.92146

... New best solution

... procrustes: rmse 0.05475202 max resid 0.2572422

Run 3 stress 27.10583

Run 4 stress 26.87087

... New best solution

(...)

Run 32 stress 26.89532

Run 33 stress 26.77503

... New best solution

... procrustes: rmse 0.000936491 max resid 0.005675025

\*\*\* Solution reached

*To look at the basic stats call “ord”; R will tell you what distance measure it used, the number of dimensions, how many iterations it went through to reach a solution, and the stress. Stress is a measure of the fit of the data and will be reported on a range of 0-1 or 1-100. In this example it is 1-100. The stress is 26.77503, which would be 0.2677503 if it were in a range of 0-1. Ordinations with stress higher than 0.3 or 30 can’t be reliably interpreted; lower stress means the solution fits the data better.*

> ord

Call:

metaMDS(comm = comm[, 6:187], distance = "jaccard", k = 2, trymax = 1000, autotransform = TRUE, expand = FALSE, plot = FALSE)

Nonmetric Multidimensional Scaling using isoMDS (MASS package)

Data: comm[, 6:187]

Distance: jaccard shortest

Dimensions: 2

Stress: 26.77503

Two convergent solutions found after 33 tries

Scaling: centring, PC rotation, halfchange scaling

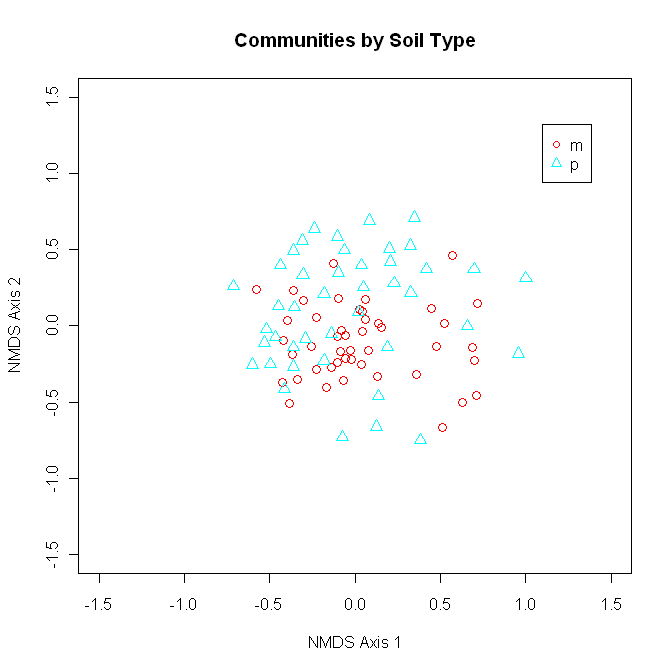
Species: non-expanded scores based on ‘comm[, 6:187]’

*Now that we have a solution we can graph it to look at how the plots are grouped with respect to each other. I expect species composition to differ by soil type, so I’ll try that first. I start with a blank graph, label the axes, set the x and y limits, and then using function “ordisymbol” I color code the points by soil type. This function doesn’t let you choose the colors, but will assign a different color to each category (in this case Miocene soil plots and Pleistocene soil plots). Each plot appears as a point. In the ordisymbol function you list the name of the ordination, the name of the data set, and the name of the factor in quotation marks. “Cex” determines the size of the points and “rainbow=T” assigns the colors. You should set “legend” to T or TRUE so you will know which colors correspond to each category. After running this function you have to click on the body of the graph to make the legend box appear; R will absolutely refuse to do anything else until you click on the graph!!!*

> plot(ord$points[,2],ord$points[,1],type="n", main="Communities by Soil Type",

+ xlab="NMDS Axis 1",ylab="NMDS Axis 2",xlim=c(-1.5,1.5),ylim=c(-1.5,1.5))

> ordisymbol(ord,comm,factor="soil",cex=1.25, rainbow=T,legend=T)

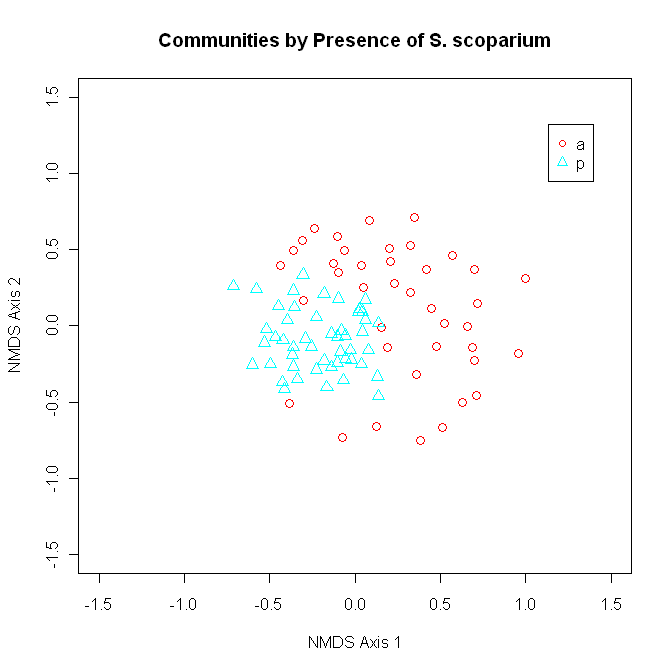


*Now I want to look at a plot that has the points color coded by the presence or absence of the dominant bunchgrass species Schizachyrium scoparium. Everything is the same as in the last command except for the main graph label and the factor (“Ssco” instead of “soil”).*

> plot(ord$points[,2],ord$points[,1],type="n", main="Communities by Presence of S. scoparium",

+ xlab="NMDS Axis 1",ylab="NMDS Axis 2",xlim=c(-1.5,1.5),ylim=c(-1.5,1.5))

> ordisymbol(ord,comm,factor="Ssco",cex=1.25, rainbow=T,legend=T)



*Based on my plots it looks like I might have some significant differences in species composition, but everybody loves p-values. To check for significant differences in species composition use the function adonis. Enter the presence/absence matrix (response matrix)~the categorical environmental variable of interest (independent variable), and the distance measure. R will give you an ANOVA table. I’ll test for difference by soil type first.*

> #now test for significant differences by soil type#

> adonis(comm[,6:187]~comm$soil,method="jaccard")

Call:

adonis(formula = comm[, 6:187] ~ comm$soil, method = "jaccard")

Df SumsOfSqs MeanSqs F.Model R2 Pr(>F)

comm$soil 1 1.454 1.4542 3.6758 0.04054 0.001 \*\*\*

Residuals 87 34.417 0.3956 0.95946

Total 88 35.871 1.00000

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

*My primary interest is in the grasses, and it looks like there is stronger grouping by presence or absence of S. scoparium, so now I’ll test for differences there.*

> #by presence/absence of S. sco#

> adonis(comm[,6:187]~comm$Ssco,method="jaccard")

Call:

adonis(formula = comm[, 6:187] ~ comm$Ssco, method = "jaccard")

Df SumsOfSqs MeanSqs F.Model R2 Pr(>F)

comm$Ssco 1 2.520 2.51956 6.5724 0.07024 0.001 \*\*\*

Residuals 87 33.352 0.38335 0.92976

Total 88 35.871 1.00000

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

*When tested independently both factors are significant! Luckily, adonis comes with the command “strata” which lets you nest factors. I am going to nest Ssco in soil type. For this test R will only permute rows and columns within soil types.*

> #nested in soil type#

> adonis(comm[,6:187]~comm$Ssco, permutations=999,method="jaccard",

+ strata=comm$soil)

Call:

adonis(formula = comm[, 6:187] ~ comm$Ssco, permutations = 999, method = "jaccard", strata = comm$soil)

Df SumsOfSqs MeanSqs F.Model R2 Pr(>F)

comm$Ssco 1 2.520 2.51956 6.5724 0.07024 0.001 \*\*\*

Residuals 87 33.352 0.38335 0.92976

Total 88 35.871 1.00000

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

*I want to have a little more freedom to make figures with this ordination (i.e. I want to be able to choose my own colors) so I am exporting the x- and y-coordinates of each point using the “write.table” function. The points are in the same order as the data, so the first set of coordinates corresponds to the first plot in the original dataset (plot # 19), the second set of coordinates corresponds to the second plot listed in the dataset (plot # 21), etc. Now you can plot your ordination using different software or different graphing functions in R.*

> write.table(ord$points, file=”C:/.../KNF.ord.txt")