HW4- Multivariate Models

Sarah Wiegreffe

February 10, 2016

For this assignment will be analyzing data on the Vegetation and Environment in Dutch Dune Meadows.

To import the data and read the metadata run the following:

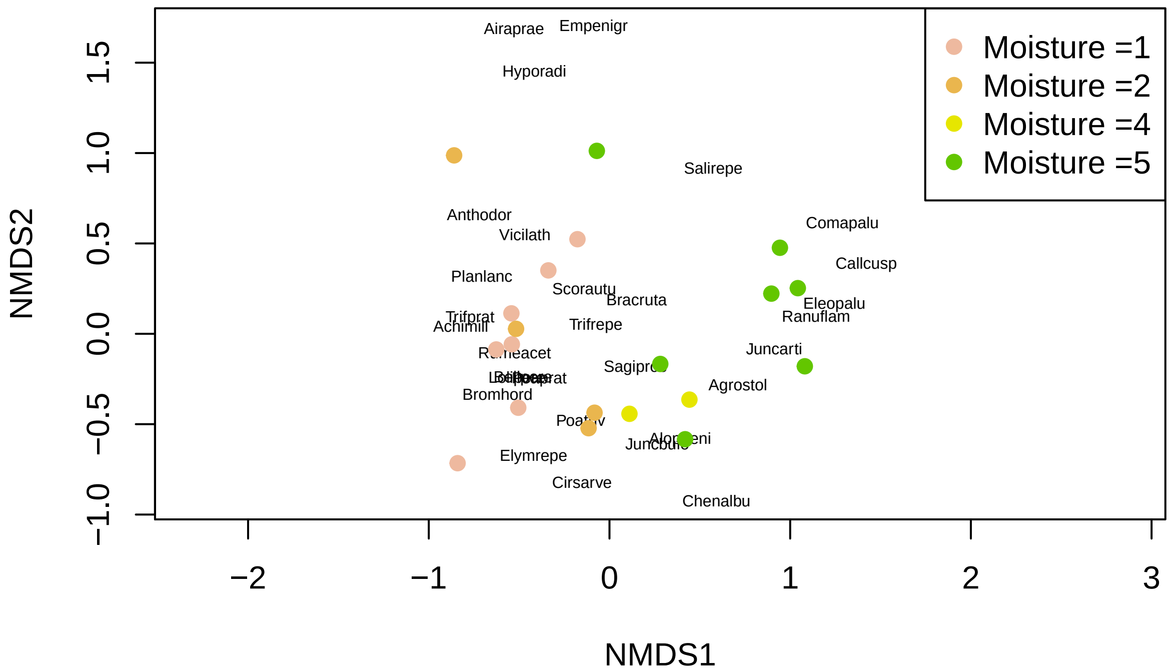
1. Conduct an indirect ordination on the dune plant community. Specifically, visually examine a NMDS plot using the bray-curtis distance metric. Below is some code to help you develop a potential plot that emphasizes the role of the environmental variable "Moisture". Describe how you interpret the graphic. What is the goal of creating such a plot? Does this analysis suggest any interesting findings with respect to the dune vegetation?

```r
#Create community matrix
#Already done

#NMDS analysis
dune_mds = metaMDS(dune)
```

```
## Run 0 stress 0.1192678
## Run 1 stress 0.1183186
## ... New best solution
## ... procrustes: rmse 0.02027125  max resid 0.06496075
## Run 2 stress 0.1192678
## Run 3 stress 0.1990349
## Run 4 stress 0.1192683
## Run 5 stress 0.1183186
## ... procrustes: rmse 3.592382e-05  max resid 0.0001203306
## *** Solution reached
```

```r
plot(dune_mds, type='n')
text(dune_mds, 'sp', cex=.5)
# generate vector of colors
color_vect = rev(terrain.colors(6))[-1]
points(dune_mds, 'sites', pch=19,
  col=color_vect[dune.env$Moisture])
legend('topright', paste("Moisture =", c(1,2,4,5), sep=''),
  col=color_vect, pch=19)
```




```
#repeat plot for other environmental variables by changing the argument to the color vector...
```

From the graphic, one can conclude that moisture is a somewhat adequate explanation of variance among dune meadow vegetation among species in the dataset, since there is for most datapoints a distance between those of lower moisture (levels 1 and 2) and higher (levels 4 and 5). The goal of creating this plot is to determine what environmental variables explain the most variance in the dataset, which can be done by observing the position of the points in the dataset with respect to the two axes and their distance apart from one another, to see if there is a variable that explains the most variance. Examining the relationship between the points by grouping them by the other environmental variables (A1, Management, Use and Manure) did not produce the same type of visual relationship as with Moisture.

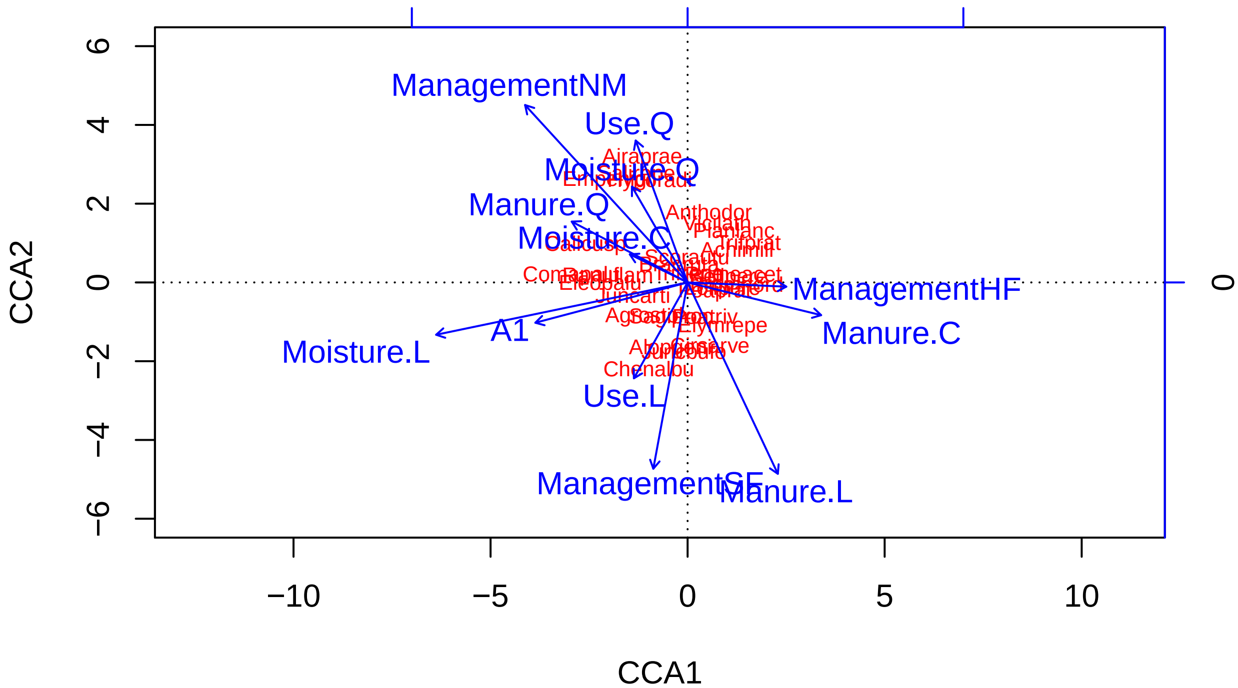
2. Carry out a direct ordination using CCA in order to test any potential hypotheses that you developed after examining the MDS plot. Specifically, carry out a test of the entire model (i.e., including all constrained axes) and also carry out tests at the scale of individual explanatory variables you included in your model if you included more than one variable. Plot your results.

```
dune_cca = cca(dune ~ ., data = dune.env)
dune_cca
```

```
## Call: cca(formula = dune ~ A1 + Moisture + Management + Use +
## Manure, data = dune.env)
##
##               Inertia Proportion Rank
## Total          2.1153     1.0000
## Constrained    1.5032     0.7106   12
## Unconstrained  0.6121     0.2894    7
## Inertia is mean squared contingency coefficient
## Some constraints were aliased because they were collinear (redundant)
##
## Eigenvalues for constrained axes:
##    CCA1    CCA2    CCA3    CCA4    CCA5    CCA6    CCA7    CCA8    CCA9   CCA10
```

```
## 0.4671 0.3410 0.1761 0.1532 0.0953 0.0703 0.0589 0.0499 0.0318 0.0260
##  CCA11  CCA12
## 0.0228 0.0108
##
## Eigenvalues for unconstrained axes:
##     CA1     CA2     CA3     CA4     CA5     CA6     CA7
## 0.27237 0.10876 0.08975 0.06305 0.03489 0.02529 0.01798
```

```r
#plot on all environmental variables
plot(dune_cca, ylim=c(-6, 6), display=c('sp','bp'), scaling=1)
```



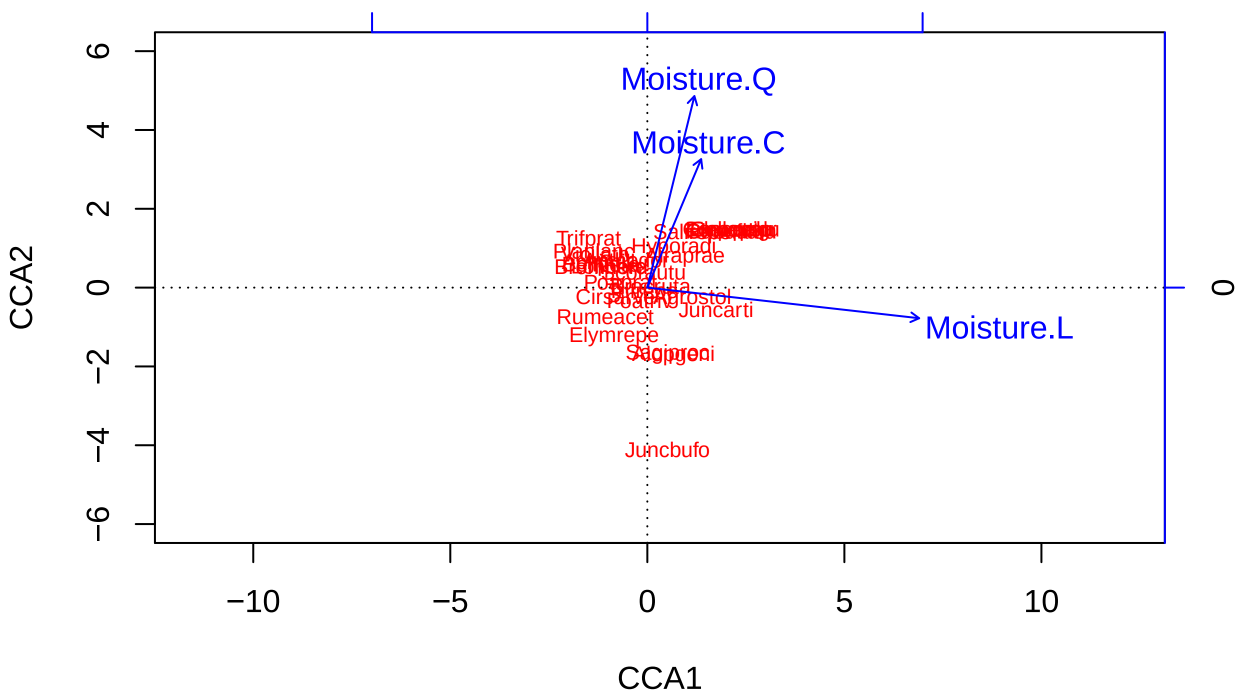

```r
#plot only on moisture and other env variables
plot(cca(dune ~ dune.env$Moisture), ylim=c(-6, 6), display=c('sp','bp'), scaling=1)
cca(dune ~ dune.env$Moisture)
plot(cca(dune ~ dune.env$A1), ylim=c(-6, 6), display=c('sp','bp'), scaling=1)
cca(dune ~ dune.env$A1)
plot(cca(dune ~ dune.env$Management), ylim=c(-6, 6), display=c('sp','bp'), scaling=1)
cca(dune ~ dune.env$Management)
plot(cca(dune ~ dune.env$Use), ylim=c(-6, 6), display=c('sp','bp'), scaling=1)
cca(dune ~ dune.env$Use)
plot(cca(dune ~ dune.env$Manure), ylim=c(-6, 6), display=c('sp','bp'), scaling=1)
cca(dune ~ dune.env$Manure)
```

```r
#Step to produce model
mod0 = cca(dune ~ 1, data = dune.env)
mod = step(mod0, scope = formula(dune_cca), test="perm", perm.max=100)
```

```
## Start:  AIC=87.66
## dune ~ 1
##
```

```
##               Df    AIC      F Pr(>F)
## + Moisture     3 86.608 2.2536  0.005 **
## + Management   3 86.935 2.1307  0.010 **
## + A1           1 87.411 2.1400  0.030 *
## <none>           87.657
## + Manure       4 88.832 1.5251  0.020 *
## + Use          2 89.134 1.1431  0.225
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=86.61
## dune ~ Moisture
##
##               Df    AIC      F Pr(>F)
## <none>           86.608
## + Management   3 86.813 1.4565  0.015 *
## + A1           1 86.992 1.2624  0.225
## + Use          2 87.259 1.2760  0.090 .
## + Manure       4 87.342 1.3143  0.085 .
## - Moisture     3 87.657 2.2536  0.005 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
plot(mod, ylim=c(-6, 6), display=c('sp','bp'), scaling=1)
```




```r
#Get details on model
mod
```

```
## Call: cca(formula = dune ~ Moisture, data = dune.env)
##
##                 Inertia Proportion Rank
```

```
## Total          2.1153     1.0000
## Constrained    0.6283     0.2970    3
## Unconstrained  1.4870     0.7030   16
## Inertia is mean squared contingency coefficient
##
## Eigenvalues for constrained axes:
##   CCA1   CCA2   CCA3
## 0.4187 0.1330 0.0766
##
## Eigenvalues for unconstrained axes:
##    CA1    CA2    CA3    CA4    CA5    CA6    CA7    CA8    CA9   CA10
## 0.4098 0.2259 0.1761 0.1234 0.1082 0.0908 0.0859 0.0609 0.0566 0.0467
##   CA11   CA12   CA13   CA14   CA15   CA16
## 0.0419 0.0201 0.0143 0.0099 0.0085 0.0080
```

From the first plot with all environmental variables, it is difficult to see which are the most significant in explaining variation- 2 dummy variables for management as well as a factor variable for both moisture and manure have the longest vectors. However this is not the best visualization because of potential correlation between variables. By using the step function and plotting the result of it, it appears that the most significant variable is Moisture alone, as we found in part 1. However, the statistics on this model shows that not much variance is explained by the constrained variables compared with our model including all environmental variables, although it is better than or at least as good as the performance of the other environmental variables.

3. Do your two analyses agree with one another or complement one another or do these two analyses seem to be suggesting different take home messages? Which analysis do you find to be more useful?

Both analyses seem to support the same results, although the outputs from the CCA provide more useful information about what numerical differences there are between modeling the different environmental variables. It is also helpful to get visualizations of the vectors of the environmental variables from the CCA to see their strength and direction.