Bio 112 | Community Analysis Final Lab Report
Dahl Winters
12/5/05

Introduction

Ecologists often seek to identify patterns in species distribution, and to explain why those patterns exist. To identify distributional patterns, it is helpful to classify the landscape into various plant communities based on variations in their species composition. To explain why those patterns exist, we must look for environmental gradients that might have led to these differences in species composition.

Ordination, followed by cluster analysis and indicator species analysis, is a powerful method for doing just this. Samples plotted in ordination space allow a qualitative analysis of which plots are more similar to each other, and with an overlay of environmental variables we can begin to guess what environmental variables might be important to the arrangement of plots. We can begin to characterize communities from the ordination results using **cluster analysis**, which selects statistically significant groupings that represent possible community types. Our selection of groupings can be further refined using **indicator species analysis (ISA)**, which tells us both the best number of groupings needed to characterize our plots, as well as which species are diagnostic for each group to a high degree of statistical significance. These indicator species can then be used to define the different community types in the study area.

To demonstrate the usefulness of this method, we sought to characterize the communities in Duke Forest, a research and recreational forest near the eastern edge of the Piedmont Plateau in North Carolina. Various patches within this forest are in different stages of secondary succession due to past disturbances, which include an extensive history of agricultural land use and relatively recent windthrows due to Hurricane Fran.

Methods

Data

We used two data sets. The first was species composition data derived from stem counts of 56 woody species in 106 plots in Duke Forest (Treelong.wk1). The second was a list of values for 17 environmental variables measured for all 106 plots (Envlong.wk1). We performed our ordination and cluster analyses using PCOrd, and our indicator species analysis using PCOrd and Excel.

Table 1: Species and environmental data used in the analysis.

Code	Species Name					Env. vars.	Values (no units given)
ACNE	*Acer negundo*	CRAT	*Crataegus sp.*	PLOC	*Platanus occidentalis*	pH	3.5-6.5
ACRU	*Acer rubrum*	DIVI	*Diospyros virginianus*	PRAM	*Prunus americana*	Ca-A	0-12
ACSA	*Acer saccharum*	FAGR	*Fagus grandifolia*	PRSE	*Prunus serotina*	Mg-A	0-7
AMAR	*Amelanchier arboreum*	FRAX	*Fraxinus sp.*	QUAL	*Quercus alba*	K-A	0.05-0.8
BENI	*Betula nigra*	ILAM	*Ilex ambigua*	QUCO	*Quercus coccinea*	Al	0-800
CACR	*Carpinus caroliniana*	ILDE	*Ilex decidua*	QUFA	*Quercus falcata*	Mn	0-1400
CACA	*Carya carolinae-septentrionalis*	ILOP	*Ilex opaca*	QUMA	*Quercus marilandica*	PO4	0-13
CACO	*Carya cordiformis*	JUNI	*Juglans nigra*	QUMI	*Quercus michauxii*	Organic	3-20
CAGL	*Carya glabra*	JUVI	*Juniperus virginiana*	QUNI	*Quercus nigra*	Sand	20-80
		LIST	*Liquidambar styraciflua*	QUPH	*Quercus phellos*		

CAOL	*Carya ovalis*	LITU	*Liriodendron tulipifera*	QUPR	*Quercus prinus*	Silt	20-60
CAOV	*Carya ovata*	MATR	*Magnolia tripetala*	QURU	*Quercus rubra*	Clay	3-30
CAPA	*Carya pallida*	MORU	*Morus rubra*	QUSH	*Quercus shumardii*	Av-H2O	2-19
CATO	*Carya tomentosa*	NYSY	*Nyssa sylvatica*	QUST	*Quercus stellata*	Slope	0-50
CECA	*Cercis canadensis*	OSVI	*Ostrya virginiana*	QUVE	*Quercus velutina*	Aspect-t	0-200
CEOC	*Celtis occidentalis*	OXAR	*Oxydendrum arboreum*	SAAL	*Sassafras albidum*	Solar	20-60
COFL	*Cornus florida*	PITA	*Pinus taeda*	ULAL	*Ulmus alata*	Dist-H2O	0-6
CRMA	*Crataegus marshallii*	PIEC	*Pinus echinata*	ULAM	*Ulmus americana*	Elev	200-900
CRUN	*Crataegus uniflora*	PIVI	*Pinus virginiana*	ULRU	*Ulmus rubra*		

Analysis

Ordination

We did an NMS (nonmetric multidimensional scaling) ordination, using the Sorensen (Bray-Curtis) distance measure, to group the 106 plots by the similarity of their species composition. First we did a step-down ordination to find the optimal number of axes needed in our ordination space to preserve the greatest amount of information with the least number of axes. We were able to judge this from the scree plot, which showed we could best reduce the number of dimensions from 68 (the number of species) to 3 without too much information loss.

Next, we ran a focal ordination using the same method. Instead of doing a step-down ordination, we explicitly selected 3 axes for our ordination space because the scree plot suggested 3 axes would give us the best data reduction. This caused plots that were more similar to each other to be arranged closer together in the 3D ordination space than more dissimilar plots. We also overlaid the various quantities of 17 environmental variables measured for the 106 plots to try to interpret which environmental variables might have led to the differences in species composition (and thus the placement in ordination space) of the plots.

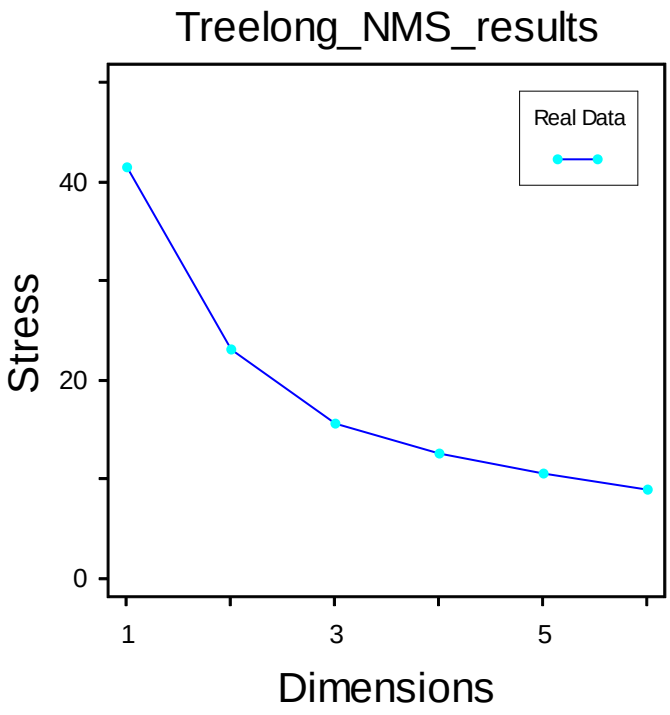

Fig. 1: Scree plot suggesting 3 axes.

Cluster Analysis

After performing our ordination, we needed to identify groups within the data based on species composition that would serve as the basis for our subsequent indicator species analysis.

We performed a cluster analysis, using the Sorensen (Bray-Curtis) distance measure and the flexible beta group linkage method, to divide our plots into 6 groups (Fig. 4). Six groups were selected because it seemed like a good choice that would mimimize the number of groups while maximizing the information remaining. These 6 groups were added to both the species and environmental data files. This allowed us to display the groups within our ordination results for a qualitative interpretation of which axes correlated best with each of the six plot groupings. Using

a biplot with our species data showing 6 groups as the main matrix (Treelong_6groups) and our environmental data showing 6 groups as the second matrix (Envlong_6groups), we were also able to interpret which environmental variables correlated best with each of the six plot groupings.

Indicator Species Analysis (ISA)

After performing a cluster analysis and making qualitative interpretations of the groupings, we used ISA to perform more quantitative interpretations of these groupings. Specifically, ISA allowed us to identify which species defined the groups we found in our data with a high level of statistical significance. In ISA, species abundance and frequency values are combined to determine the degree that each species is diagnostic for a group of plots. This degree is given by the species' indicator value (IV). As a species approaches being a perfect indicator for a particular group, it will approach having 100% of its abundance found on plots belonging to that group, and it will have a frequency of 100% (meaning it occurs on all the plots within that group).

Unlike cluster analysis, where we chose the number of groups to use based on what we thought might give a good minimization of group number and maximization of information remaining, ISA allows us to quantitatively determine the optimal number of groups based on the information contained in the species data. To find this optimal number, we performed a Monte Carlo test with 1000 runs for 3, 4, 5, and 6 groups, and obtained p-values for each number of groups. The optimal number of groups would have both the lowest average p-value and the highest number of significant indicator species (meaning the p-value of the indicators were below 0.05). As shown in Fig. 2, only 6 groups met the criteria. A higher number of groups could possibly have been better, but we only tested 3-6 groups to keep the number low for simplicity.

Next, we opened the Monte Carlo test results file for 6 groups, then transferred to an Excel spreadsheet the relative abundance, relative frequency, indicator values, and p-values. Each species had an average and maximum indicator value (IV), the group where the maximum value occurred, and a p-value. We divided the data into the 6 groups, then sorted each group by descending average value. The species with the highest average IVs were at the top. Indicator species were selected by choosing IV values > 30 (Table 1).

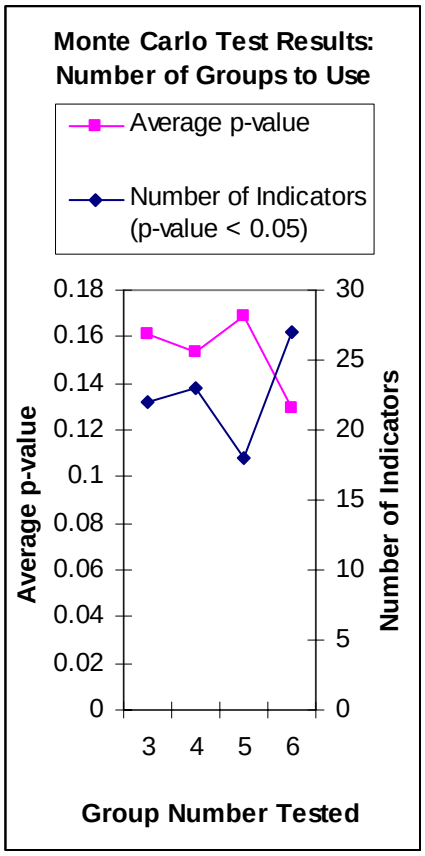

Fig. 2: Results from Monte Carlo test.

Results

Ordination

For each of the three axes of our ordination space, the following species responded most strongly to them. This was determined by a strong correlation (long vector) on the biplot in the direction of that axis, as well as a high r-value above 0.3.

Axis 1: *Q. prinus, A. rubrum, O. arboreum*
Axis 2: *Q. alba, L. styraciflua, C. caroliniana, U. alata, O. arboreum*
Axis 3: *Q. stellata, J. virginiana, F. grandifolia, L. tulipifera*

Axis 1 had (-) correlations for pH, Mn, and Ca-A, and a (+) correlation for Al (see Fig. 3). **This axis likely involves soil pH, which can affect nutrient availability.** Acid pH can increase Mn and Al availability to toxic levels, which would explain the association of Mn and Al with this axis, and it also lowers Ca and Mg availability, which could explain the skewing of the Ca and Mg vectors in the direction of the pH vector.

Axis 2 had (-) correlations for Dist-H2O (distance to water) and Elev, and (+) correlations for Ca-A and Mg-A. **This axis might involve soil factors (water and nutrients) that control plant growth.** The Ca-A and Mg-A lines fall along Axis 2 just as Dist-H2O does, and both these nutrients are as essential as water for plant growth. The Elev line might be explained by its link to soil moisture if, as elevation increases, we have the same effect on plant growth as increasing Dist-H2O. This is expected since water usually drains to lower elevations.

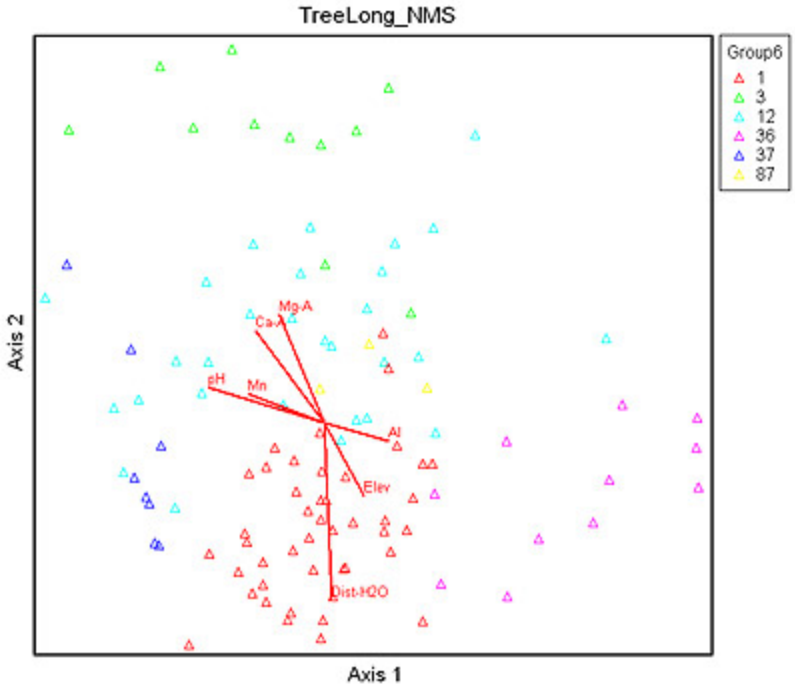

Fig. 3: Environmental variable vectors overlaid on a 2D view (Axis 2 vs. Axis 1) of the 3D ordination space. (Also shows the 6 groups found in cluster analysis.)

Dist-H2O lines up very well with Axis 2. It was the most significant environmental variable since it has the longest vector length. Thus, as distance from water increases, this has more of an effect than any other variable on a plot's species composition (Fig. 4). *Q. alba, L. styraciflua, C. caroliniana, and U. alata* all strongly corresponded with Axis 2, suggesting these 4 species were sensitive to distance from water.

Axis 3 lacked any strong correlations with the environmental variables. K-A had the only correlation above 0.3 for this axis. Like Ca and Mg for Axis 2, K is another important plant nutrient.

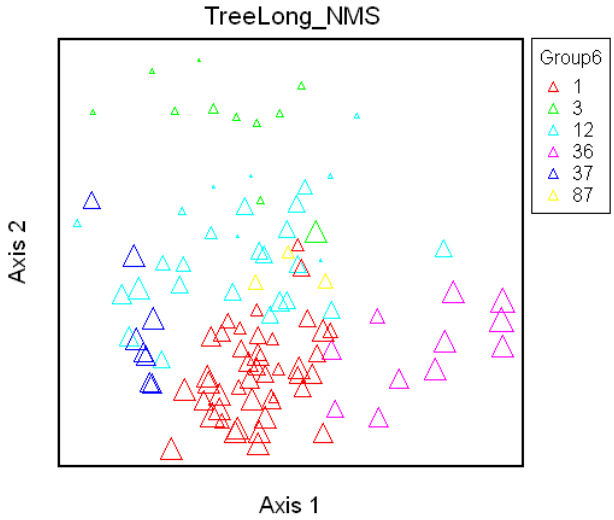

Fig. 4: Dist-H2O is high for all groups except green, indicating that most plots in the green group are closer to lakes or streams.

Cluster Analysis

On the following page are the 6 groups found from cluster analysis (Fig. 5). A description of their characteristics and how they correspond to both the environmental variables and species composition will be given in the next section, when these groups are discussed along with their indicator species.

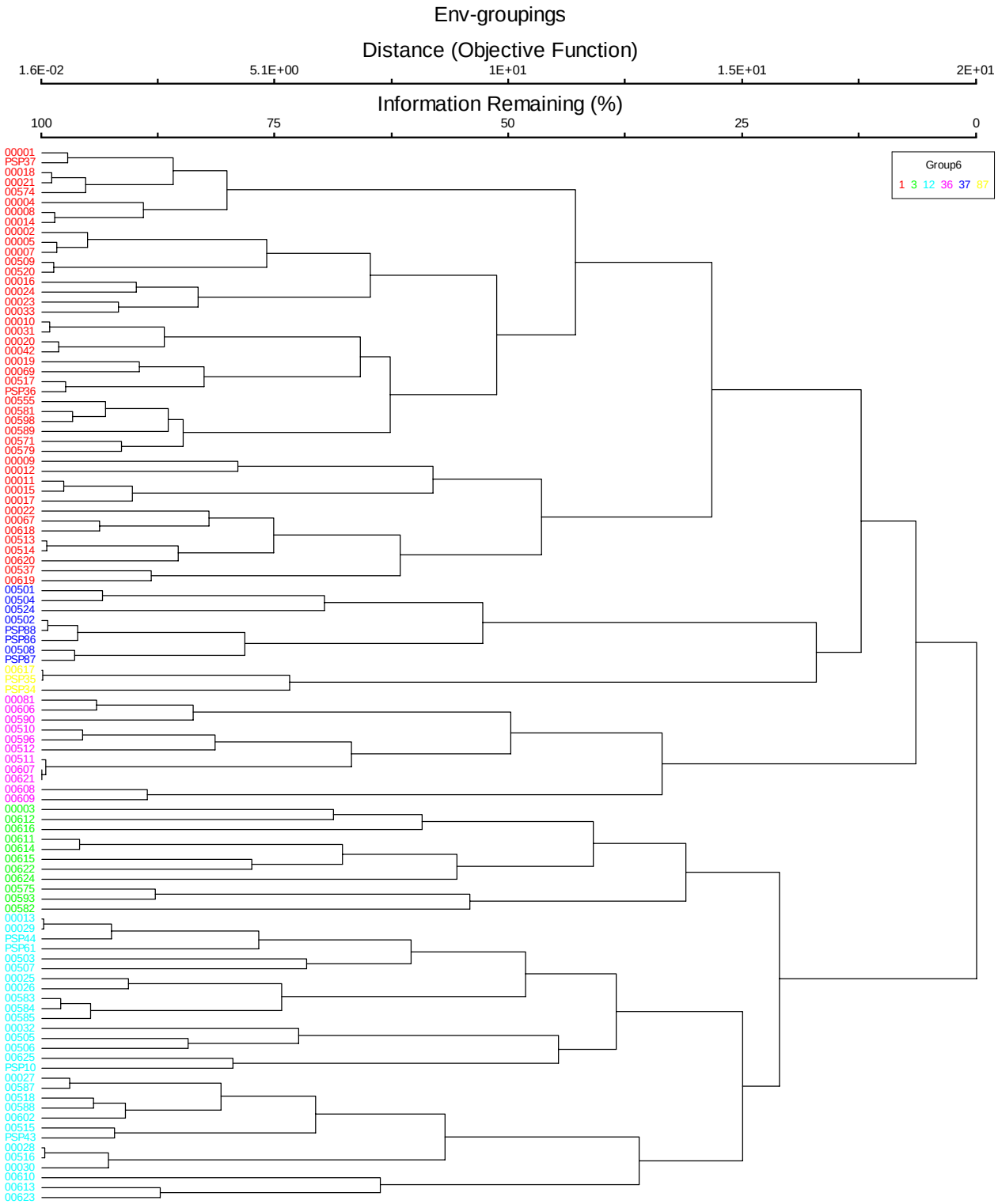

Fig. 5: Dendrogram of ordination results divided into 6 groups by cluster analysis. The first split separates the Dark Blue (37) and Yellow (87) groups from the Red group (1), indicating the Dark Blue and Yellow groups are more closely related to each other. The second split separates the Purple group (36) from the remaining three. The third split separates the Green (3) and Light Blue (12) groups from the other four, indicating these are most closely related to each other. It takes almost all the available information to define the Green and Light Blue groups, and nearly 75% of the information to make the first split between Red (1) and the others.

Indicator Species Analysis

Table 2 below lists the results of our analysis. We chose to list all species with an indicator value (IV) greater than or equal to 30 as indicator species for each group, though some species with lower IVs appear for multiple groups (like Q. alba for Groups 1 and 37). Species with an IV > 66 were very significant indicators for that particular group (*U. alata* for Group 3, *F. grandifolia* for Group 12, *Q. prinus* for Group 36, *F. americana* for Group 37, and *P. taeda* for Group 87).

Table 2: Species with IV > 30 for each of the 6 groups.

Indicator Values

Max: actual max indicator value

Maxgrp: group that the max value occurs in

p-value: anything under.05 significant

Group 1 - contains 44 of 106 plots (42%)

Red Group

Species	Code	Avg IV	MaxGrp	p-value
Quercus alba	QUAL	44	1	0.001
Oxydendrum arboreum	OXAR	37	1	0.027
Quercus velutina	QUVE	35	1	0.047
Carya tomentosa	CATO	34	1	0.118
Cornus florida	COFL	33	12	0.022

Group 3 - contains 11 of 106 plots (10%)

Green Group

Species	Code	Avg IV	MaxGrp	p-value
Ulmus alata	ULAL	66	3	0.005
Ilex decidua	ILDE	62	3	0.004
Liquidambar styraciflua	LIST	60	3	0.002
Carpinus caroliniana	CACR	49	3	0.044
Ulmus rubra	ULRU	49	3	0.009
Morus rubra	MORU	43	3	0.029
Carya ovata	CAOV	39	3	0.053
Quercus michauxii	QUMI	35	3	0.005

Group 12 - contains 29 of 106 plots (27%)

Light Blue Group

White ash, tulip poplar, dogwood, red oak - disturbed area?

Species	Code	Avg IV	MaxGrp	p-value
Fagus grandifolia	FAGR	73	12	0.002
Liriodendron tulipifera	LITU	58	12	0.003
Cornus florida	COFL	34	12	0.022
Quercus rubra	QURU	30	37	0.055

Group 36 - contains 11 of 106 plots (10%)

Purple Group: *Q. prinus* strongly diagnostic

Species	Code	Avg IV	MaxGrp	p-value
Quercus prinus	QUPR	99	36	0.001
Oxydendrum arboreum	OXAR	37	1	0.027
Quercus coccinea	QUCO	37	36	0.034
Acer rubrum	ACRU	33	36	0.076

Group 37 - contains 8 of 106 plots (8%)

Dark Blue Group

Species	Code	Avg IV	MaxGrp	p-value
Fraxinus americana	FRAX	74	37	0.001
Cercis canadensis	CECA	59	37	0.003
Ostrya virginiana	OSVI	45	37	0.018
Quercus rubra	QURU	38	37	0.055
Quercus alba	QUAL	35	1	0.001
Carya glabra	CAGL	34	37	0.071
Prunus serotina	PRSE	33	37	0.186

Group 87 - contains 3 of 106 plots (3%)

Yellow Group

Mostly pines and *J. virginiana* (conifers), and *Q. stellata*.

Species	Code	Avg IV	MaxGrp	p-value
Pinus taeda	PITA	84	87	0.023
Pinus virginiana	PIVI	69	87	0.002
Juniperus virginiana	JUVI	63	87	0.002
Quercus stellata	QUST	59	87	0.001
Pinus echinata	PIEC	39	87	0.001

We used these indicator species results together with an overlay of environmental variables over the 6 groups to characterize the 6 types of communities represented in the 106 Duke Forest plots, and to determine which environmental gradients might be responsible for differences in species composition between the groups. Axes 1 and 2 of our ordination graph did a good job of capturing the important environmental gradients (Fig. 6).

Communities Represented by the Duke Forest Plots

Group 1 (Red), White Oak/Sourwood community – the highest indicator value is 44 for *Q. alba*, meaning that this group isn't as well-defined as the others which have higher indicator values. See Fig. 7 for a visual description of why *Q. alba* was not a good indicator species. This group has high Dist-H2O and Elev (found on high, dry sites), and grows in soils with low Ca-A and Mg-A values, indicating they are somewhat nutrient-poor.

Group 3 (Green), Elm/Deciduous Holly/Tulip Tree community – this group is characterized by low Dist-H2O, Elev, and Slope, and to a lesser extent (by only some plots in this group) by high Ca-A and Mg-A. These sites are likely low-lying areas near lakes or streams.

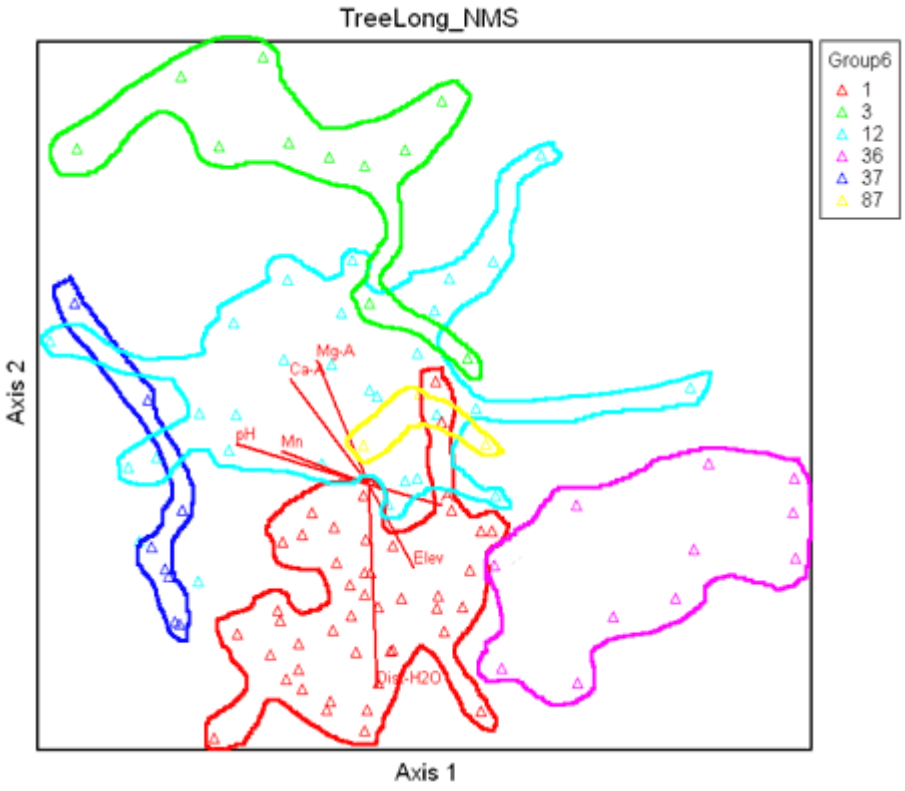

Fig. 6: A 2D view (Axis 2 vs. Axis 1) of the 3D ordination space with each community outlined to show how they correlate with the axes. Axis 1 (pH/nutrients) increases from right (acid soils) to left (alkaline soils). Axis 2 (nutrients, water) increases from bottom (dry, nutrient-poor) to top (wet, nutrient-rich).

Group 12 (Light Blue), Beech/Sweetgum community – this group is largely between the red and green groups, indicating that the plots are intermediate in elevation and distance to water. This community must occur under a wide range of environmental conditions since there are no strong correlations with any environmental variable, only a somewhat significant one with sites having low Al.

Group 36 (Purple), *Q. prinus* community – this group can be defined by the presence of *Q. prinus*, which has the highest indicator value of any species, 99. See Fig. 8 for a visual description of why *Q. prinus* was an excellent indicator species. This group has an overall lower pH, Ca-A, Mg-A, Mn-A, and K-A than other groups, indicating high acid, nutrient-poor soils in these plots. Dist-H2O is also high here, so plots in this group must be dry.

Group 37 (Dark Blue) – White ash/redbud community – This group grows on soils having high pH, Mn-A, and Ca-A, and low Al values, which indicate alkaline, mostly nutrient rich soils.

Group 87 (Yellow) – Pine community – best characterized by the presence of *P. taeda*, and is a conifer-dominated community. *Q. stellata* is the only hardwood with an average IV >= 30. This group is found in areas low in pH, organic matter, slope, and with very nutrient-poor soils (low in Ca-A, Mg-A, Mn-A, and K-A).

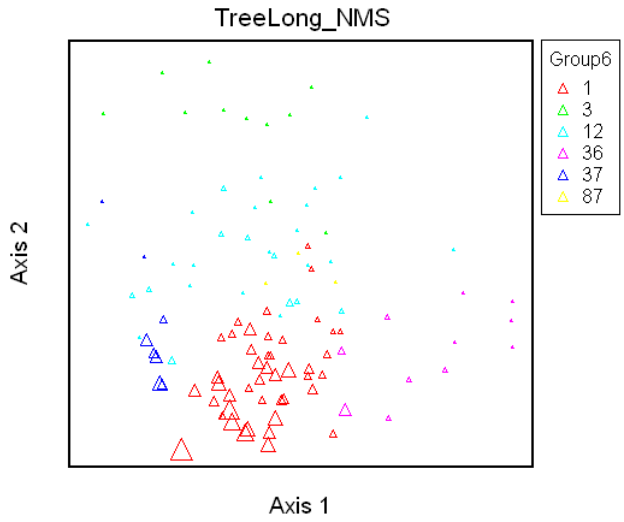

Fig. 7: *Q. alba* is a somewhat good indicator species for the red group, since some *Q. alba* abundance (given by large triangles) is also found in the blue and purple groups.

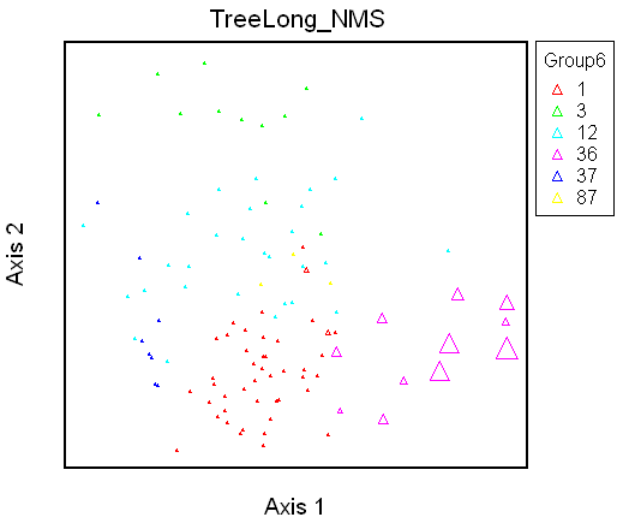

Fig. 8: Greater abundance only in the purple group (36) shows *Q. prinus* is a better indicator species for that group than *Q. alba* is for the red group (1).

Discussion

There were several strong environmental gradients that were likely responsible for the differences in community composition observed in the 106 Duke Forest plots. The most important gradients involved soil nutrients and moisture in some way, which are factors we know to be important for plant growth. These multiple gradients seem to have been effectively captured by Axis 1 (soil pH, which affects nutrient availability) and Axis 2 (water and nutrients) during the data reduction phase of ordination.

Plots with a high abundance of white ash and redbud (Group 37) likely also contained alkaline, nutrient-rich soils. Plots with high conifer abundance (Group 87) were likely to contain acid, nutrient-poor soils, and if the plots were also dry, *Q. prinus* (Group 36) was likely to be found. Low-lying areas near lakes or streams were most likely to contain a high abundance of elms, deciduous hollies, and tulip trees, indicator species for Group 3. Group 1 (white oak/sourwood) was found on high, dry, moderately nutrient-poor sites, and Group 12 (beech/sweetgum) was found over a wide range of environmental gradients.

Our analysis led us to find relatively good indicator species for each of the groups, ranging from *Q. prinus* (IV = 99) to *Q. alba* (IV = 44). However, most species were not good indicators. These were likely either rare species that didn't have high abundance in the plots, or they were present in many of the plots, so that their distribution didn't correspond as highly to the environmental gradients as those of the indicator species.

Several improvements could be made to future analyses of the Duke Forest plots. First, the possibility of having more than 6 groups should be tested during ISA. The species composition data might have been better represented by a larger number of groups, considering that almost ¾ of the available information was needed to make the first split between Group 1 and the others. Second, other environmental variables should be included that might be significant in explaining species compositional variations in the plots, such as microclimate and re-growth in recently disturbed areas. Lastly, other types of ordination should be tried. The community types found in this study would be better supported if analyses were done using other ordination methods and the same types were found.