Community Ecology in Coweeta Long Term Ecological Research

https://github.com/mtgaffney/BihariGaffneyOakley_ENV872_EDA_ FinalProject.git

Michael Gaffney, Eni Bihari, & Cal Oakley

Contents

List of Figures

1 Rationale and Research Questions

Coweeta Long Term Ecological Research (LTER) is one of the oldest long term ecological research sites in the United States. It's located in the eastern deciduous forest of the southern Appalachian Mountains, specifically Macon County, North Carolina. This ecosystem is ecologically valuable because of its combination of old age, high productivity, temperate biodiversity, & high environmental variation (NSF 2018). These last two factors are the ones on which our investigation will focus.

The southern Appalachians are home to numerous woody plant species. These species each have unique characteristics that make them better suited for certain sites more than others (think of shade tolerance, drought tolerance, or nutrient requirements). Also, the southern Appalachian mountains are so topographically diverse that species with vastly different site requirements can often be found within close proximity to one another, while groups of the same species are found far apart. This presents a challenge to researchers interested in community ecology because unless you collect extensive environmental measurements at every research site they cannot understand why a species assemblages occurs where it does. That is why we want to perform a comprehensive analysis of environmental and species occurrence data to determine if there are more general environmental trends that can explain the distribution of woody species at Coweeta.

2   Dataset Information

Three datasets in total were used for this analysis, all provided by Dean Urban and acquired between 1997 and 1999. Data collection was organized through 108 sample quadrats (20x20m) arranged in clusters of 3 to 4 along transects that covered an elevation gradient. Species data consists of the basal area of 18 species across 108 plots in Coweeta; realtively uncommon species were removed prior to analysis. Environmental data consists of 20 environmental variables across the same 108 sites. Finally, we were provided geospatial data for the sites.

Table 1. Tree species present in the data set.

Species Code	Scientific Name	Common Name
ACRU	Acer rubrum	Red Maple
AMAR	Amelanchier arborea	Common Serviceberry
BELE	Betula lenta	Sweet Birch
CAGL	Carya glabra	Pignut Hickory
CATO	Carya tomentosa	Mockernut Hickory
COFL	Cornus florida	Flowering Dogwood
KALA	Kalmia latifolia	Mountain Laurel
LITU	Liriodendron tulipifera	Tulip Tree
NYSY	Nyssa sylvatica	Black Tupelo
OXAR	Oxydendron arboreum	Sourwood
PIRI	Pinus rigida	Pitch Pine
QUAL	Quercus alba	White Oak
QUCO	Quercus coccinea	Scarlet Oak
QUPR	Quercus prinus	Chestnut Oak
QURU	Quercus rubra	Red Oak
QUVE	Quercus velutina	Black Oak
RHMA	Rhododendron maximum	Great Laurel
ROPS	Robinia pseudoacacia	Black Locust

Table 2. Environmental variables present in the data set.

| Variable	Description
Elevation	Elevation (M)
SlopeDEM	Slope
TAspectD	Transformed Aspect
TCI	Topograhpic Convergence Index
xDepth	Mean soil depth (cm)
sDepth	Standard deviation of depth
pH	pH
Acidity	Acidity
Ca	Calcium (cmol(+)/kg)
K	Potassium (cmol(+)/kg)
Mg	Magnesium (cmol(+)/kg)
P	Phosphorous (g/g)
C	Soil carbon (%)
N	Nitrogen (%)
C:N	Carbon:Nitrogen ratio
ECEC	Effective Cation Exchange Capacity
BS	Base Saturation
Clay	Clay (%)
Silt	Silt (%)
Sand	Sand (%)
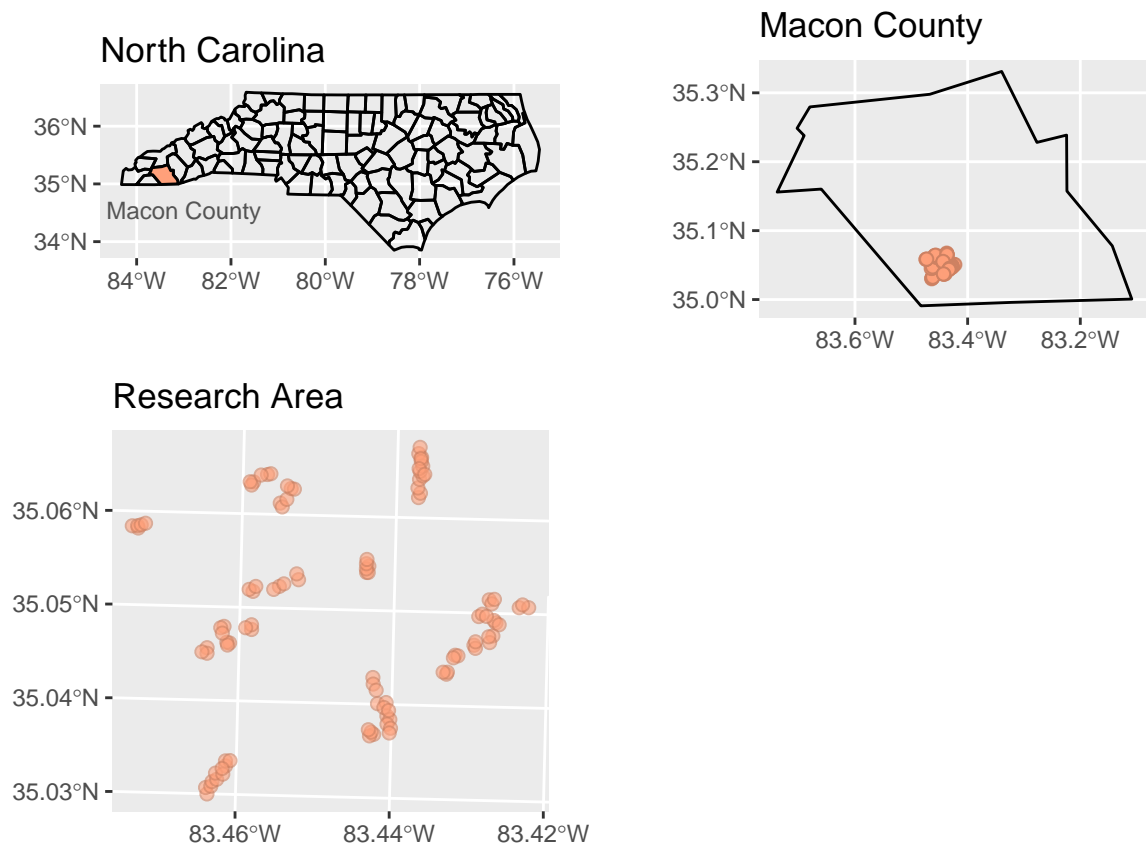
Figure 1: Geographic distribution of the 108 Coweeta plots.

3 Exploratory Analysis

We first explored the species composition of the Coweeta plots by looking at the mean basal area (BA) per species. We did this separately for the canopy and understory species, since these species play very different ecological roles in these ecosystems. The canopy and understory are made up of very different species, and the trees have very different sizes and abundances. Canopy trees tend to be sparser and larger, while understory trees tend to be denser and smaller. Chestnut oak (QUPR), red maple (ACRU), northern red oak (QURU), and scarlet oak (QUCO) are the trees that make up the most BA among overstory trees, which is to be expected in these oak-hickory dominated mountainous ecosystems. The highest BA understory species is a native rhododendron (RHMA), which is also to be expected in these ecosystems.
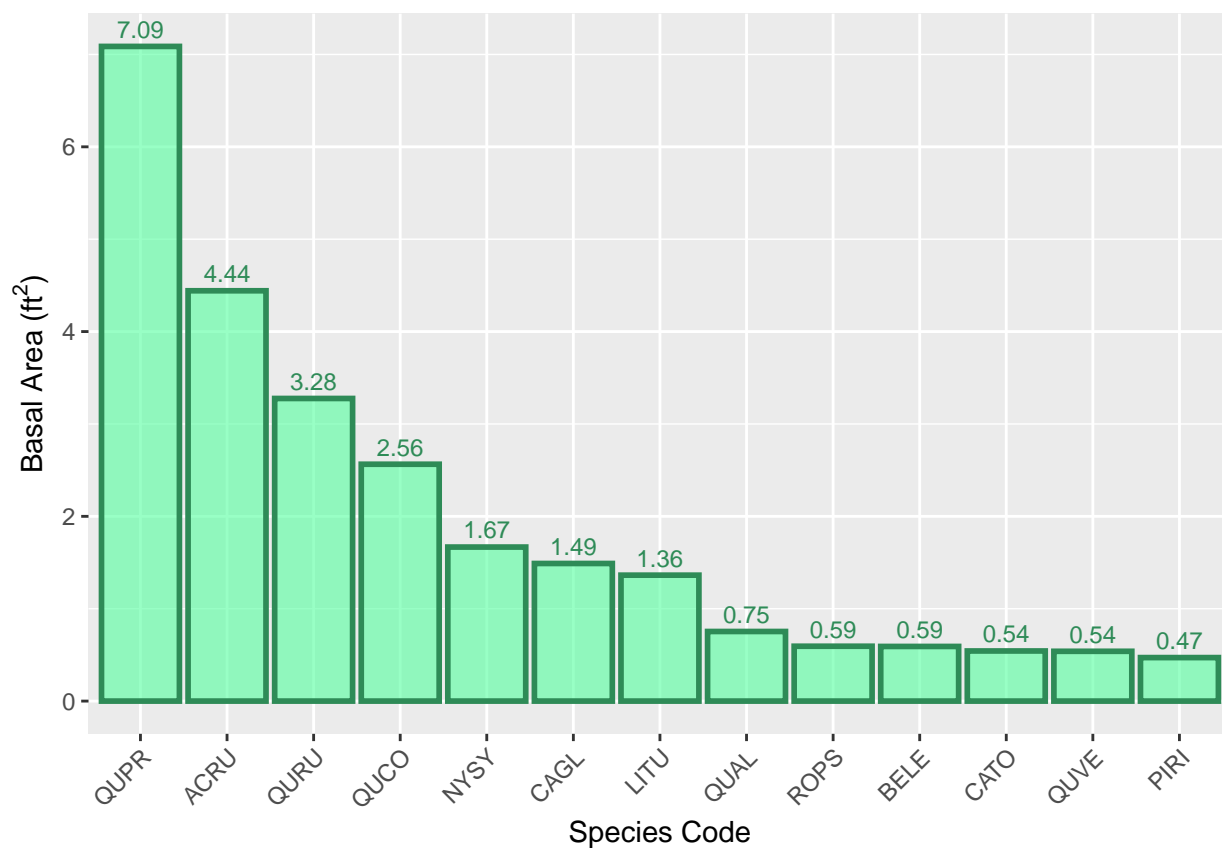

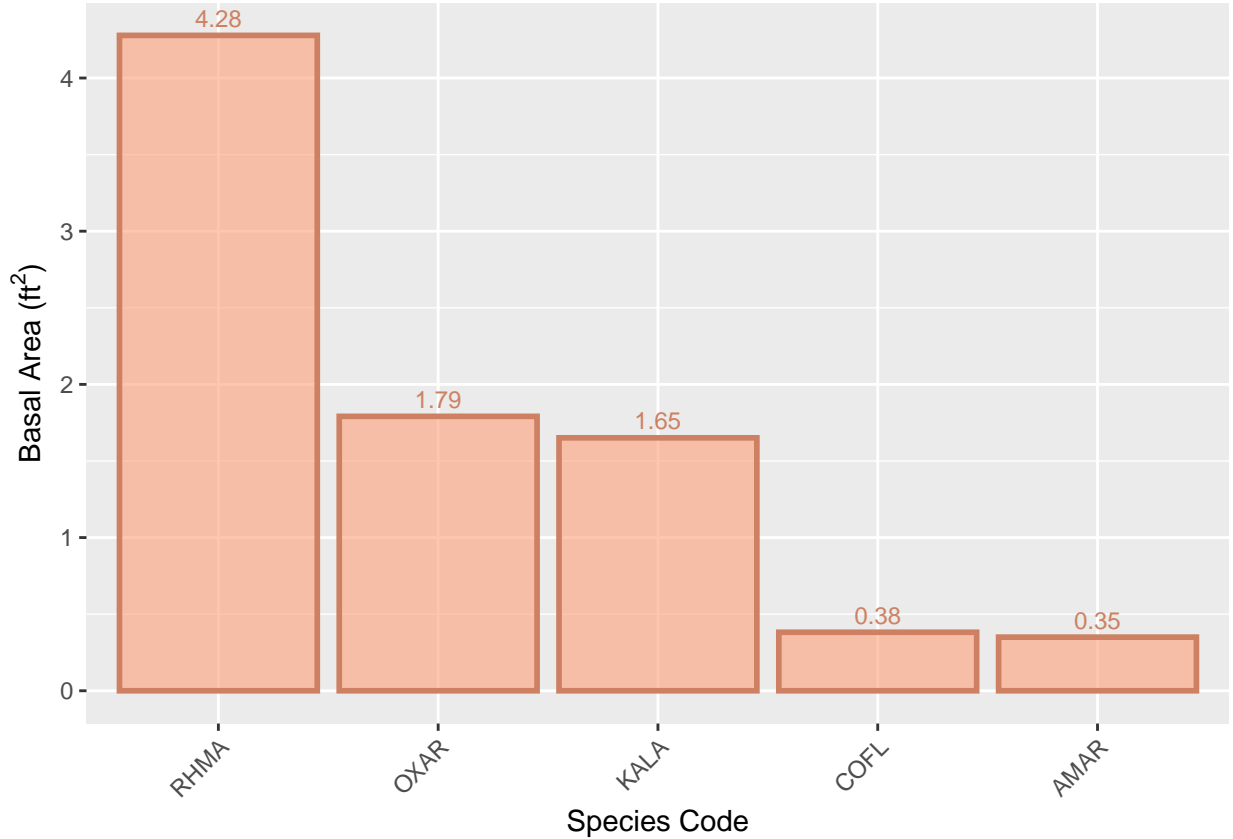Figure 2: Mean basal area per plot for canopy species at Coweeta LTERS.

Figure 3: Mean basal area per plot for understory species at Coweeta LTERS.

We also explored the basal area distributions individually for some of the most common canopy species in Coweeta. These distributions are sharply skewed right for all 4 species, which can be expected for the dominant trees in such ecosystems. This simply means that there are very many small trees and very few large trees, which makes sense based on their reproductive strategy. These trees produce a lot of seeds and stump sprouts, but only a fraction of these survive into old age. We created a cowplot of all of these species alongside one another for easy comparison.

Finally, we explored some of the environmental variables. Although our method of analysis for this data set will not require us to eliminate correlated variables for statistical purposes, it is worth exploring what some of the correlations are.
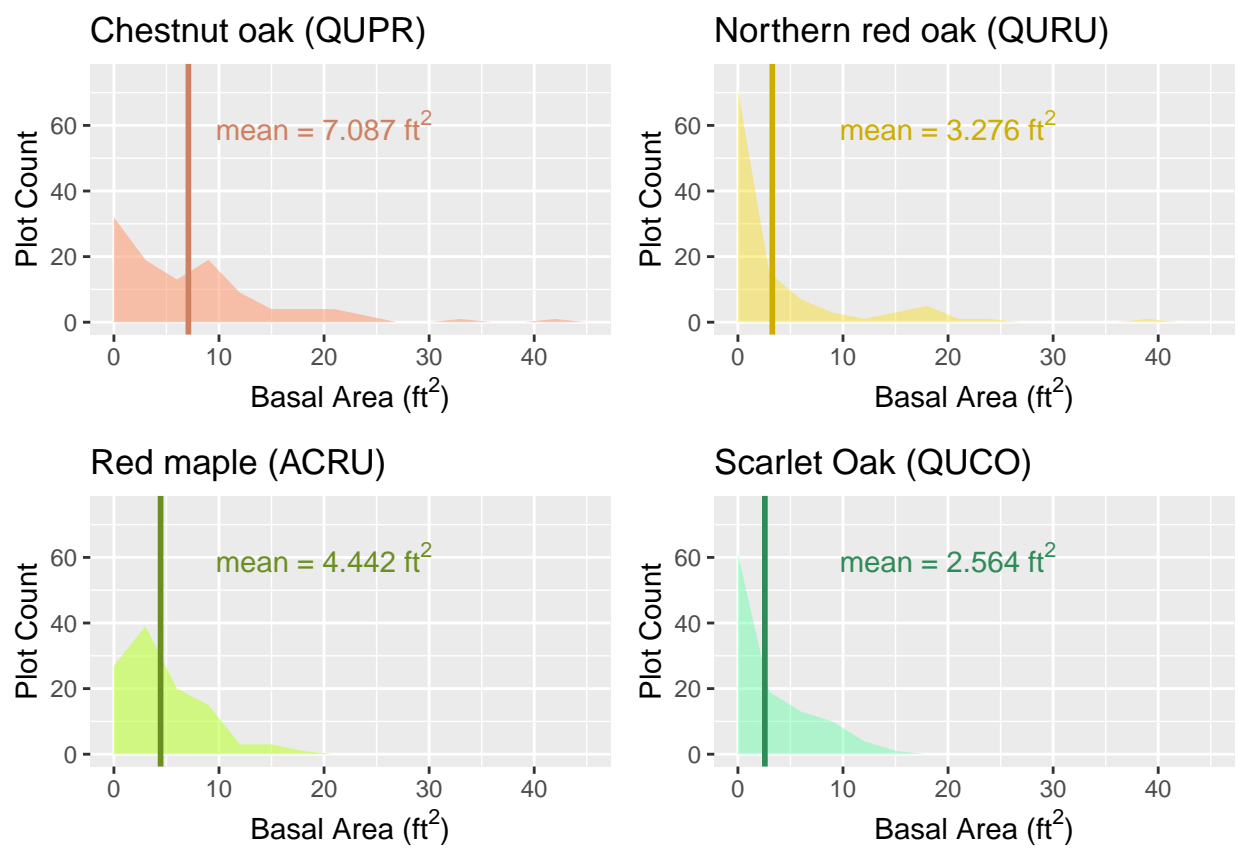
Figure 4: Frequency distribution of areas for dominant canopy trees in plots at Coweeta LTRS.

4 Analysis

Question:

Our question for this analysis is: what environmental variables, or complex of variables, primarily sorts tree communities in Coweeta? Our initial hypothesis, following a similiar study conducted by Urban et al. (2000), is that elevation, serving as a proxy for both temperature and moisture, will function as the primary sorting mechanism for tree species. In order to answer this question, we elected to use Nonmetric Multidimensional Scaling (NMS). NMS is an ordination technique that can visualize trends in complex, non-parametric, multivariate datasets. As a technique, NMS simplifies complex data by plotting sites in ordination space. The aim is to draw sites close together on a graph that are close together ecologically. It converts ecological distance, in other words, to graph distance.

NMS works through an iterative process. It creates a series of numerical approximations that capture trends within the data. The process first randomly scatters the plots, regresses the ordination distances and ecological distances between the points, and then "jitters" the data once again to create an improved configuration. A successful ordination will create an ordination where ecological distance should correspond with a 1:1 ratio with ordination distance; far ecological distance should be far ordination distance.

NMS is particularly appropriate for this data set because, unlike other statistical tests, it makes no parametric assumptions about input data sources; as tree basal are not normally distributed, this is particularly advantageous. Additionally, NMS works particularly well when there are a great deal of zeroes in the data, which is often the case for species data on tree plots.

The NMS process began by screening the environmental variables for correlations and removing the correlated variables. Although the actual NMS analysis was run on the species data, removing correlations is important for visual purposes later on. We then relativized the original species data to account for the vast difference in basal area between some of the species. With this relativized data, we then created a distance matrix from the species data. This is done using the ecodist package, using the Bray Curtis method (Urban et al. 2007. This produces a matrix that details how far apart each plot is ecologically from one another. In other words, plots that are ecologically very similar—having similar tree communities, species richness, and basal area—will have different scores than those that are very far apart.

The next step in running the NMS was to start a stepdown procedure. This involved producing a full 6 dimensional NMS object, and then evaluating how well it performed. An NMS will produce two important values: and R^2 and stress values (Figure 5). With a full ordination, we plotted these values to explore what the optimal number of dimensions for the final NMS would be. The optimal number captures a significant portion of the variance but also has low stress. In this case, 2 axes appears to be the best configuration (with a stress value below .3 and an R^2 around .6).

We then ran a second final NMS with two dimensions. In this case, we have R generate not one but twenty versions of this same NMS, so that we can choose the one that performs best.
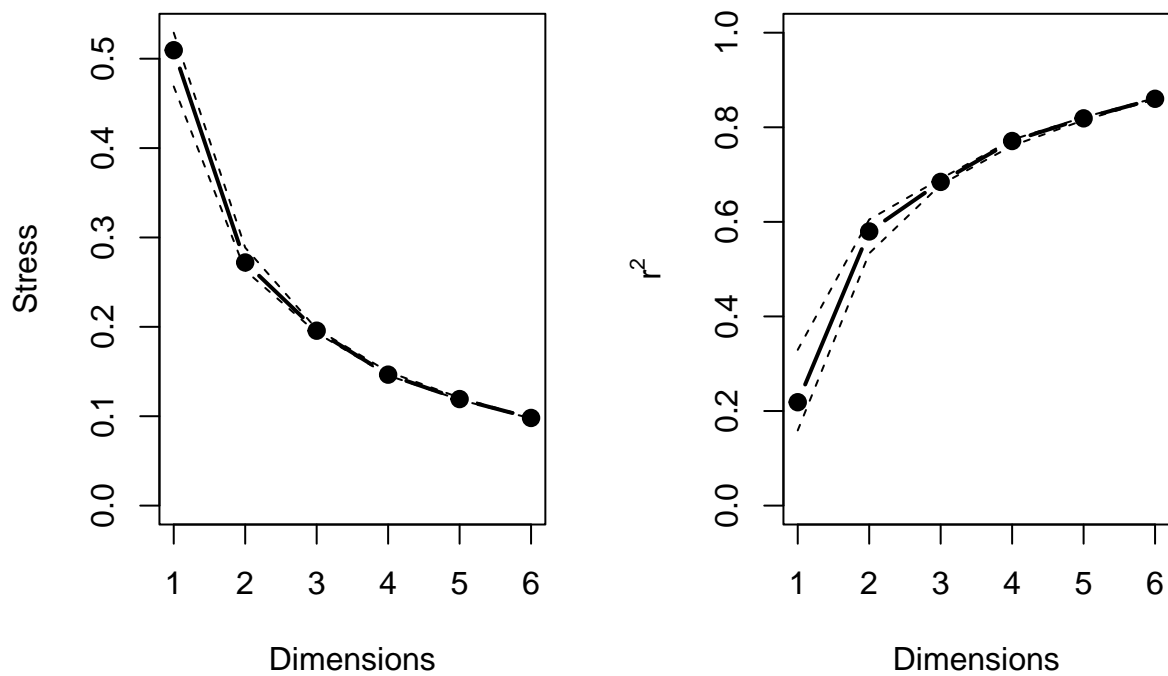
Figure 5: Stress test and R^2 for 6-dimensional NMS.

The final NMS object, however, is not directly intelligible. This is because NMS generates all of the axes at the same time; in doing so, it does not pay attention to importance. In practice, this means that in order to properly understand what the axes are, it is necessary to run a Principle Component Analysis (PCA) on the axes. The PCA effectively rotates the initial NMS ordination in such a way so as to force the points to align with the axes. It also makes axis 1 contain the majority of the variance.

After rotating the NMS, we checked for how well the analysis performed. We plotted a Shepard Diagram, which illustrates the relationship between ecological distance and ordination distance (Figure 6). Ideally, this is a one to one relationship; here, we see that the origination performed relatively well in aligning ecological distance with ordination space. Here, we also checked the overall R^2 for the NMS, which was roughly 60% depending on the different runs. Of the variance explained, roughly 36% was captured by NMS 1, and the other 24% was captured by NMS 2.
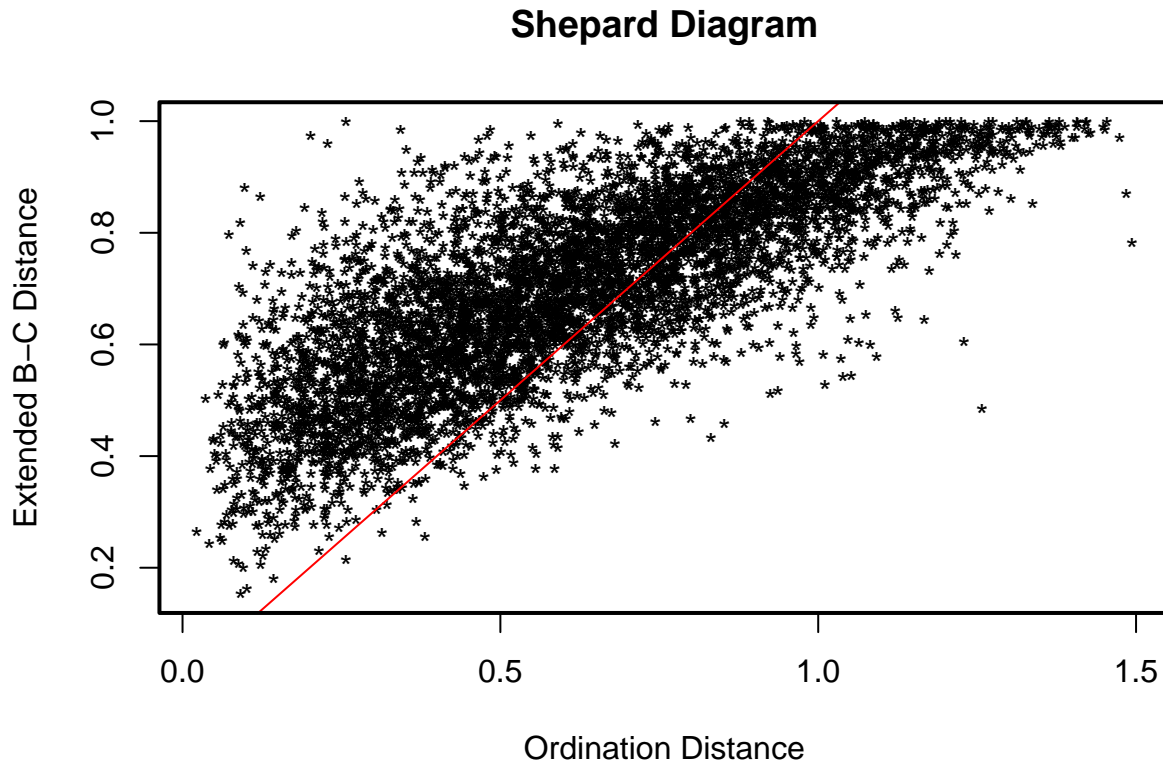

Figure 6: Shepard Diagram showing the relationship between ordination distance and ecological distance.

In order to fully understand the ecological implications of the ordination we then created three different plots. We began by preparing the data to be plotted with the ggplot package.

Then we created the plots. Figure 7 below shows a graph of the 108 Coweeta plots in ordination space. The first NMS axes (with most of the variance) is on the y-axis, the second (with the rest of the variance) is on the X axis. Placed on top of this ordination are correlation vectors for all of the environmental variables. Effectively, they show how well correlated the environmental variables are with the ordination axes that were generated from the species data: the direction of the vector shows the direction of the correlation (positive/negative), and the length shows how strong the correlation is. Here, the C:N ratio is strongly correlated with the first NMS axis, where K (potassium) is negatively correlated. This suggests that NMS 1 likely represents a gradient of soil properties, or, given the Coweeta system, it likely represents a gradient of soil moisture. While NMS 2 contained far less variance, it appears to be positively correlated with elevation, meaning that it likely represents a rough elevation gradient.
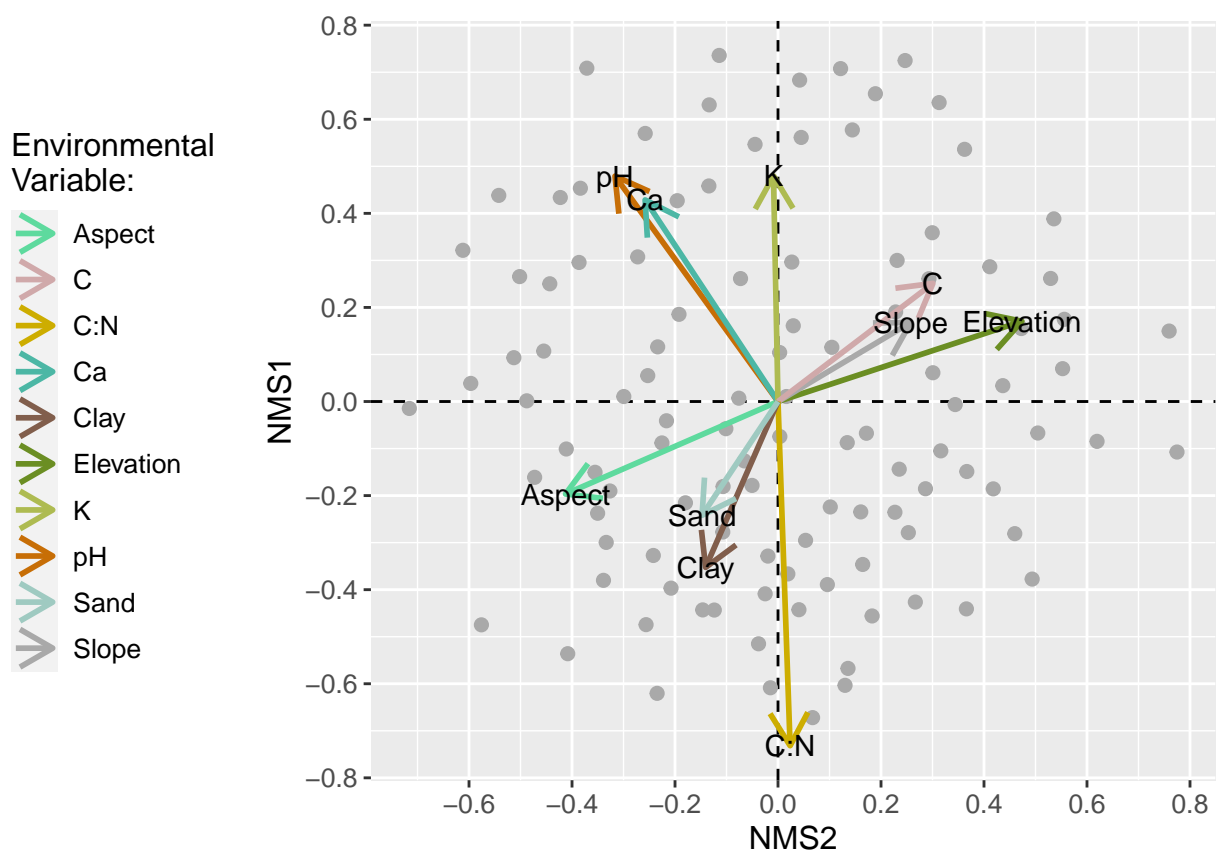

Figure 7: Environmental vectors showing correlations in NMS space.

We further explored the meaning of the gradients by plotting individual tree species by relative basal area into NMS space. Figure 8 below shows 6 species in ordination space sorted by the soil moisture gradient. What is particularly evident here is the way that sweet birch (BELE) and tulip trees (LITU) appear in the areas of relatively high soil moisture, where pitch pine (PIRI), scarlet oak (QUCO) and mountain laurel tend to appear in places with drier soils. Red maple (ACRU), a generalist, appears all over and does not seem to favor any particular soil type.
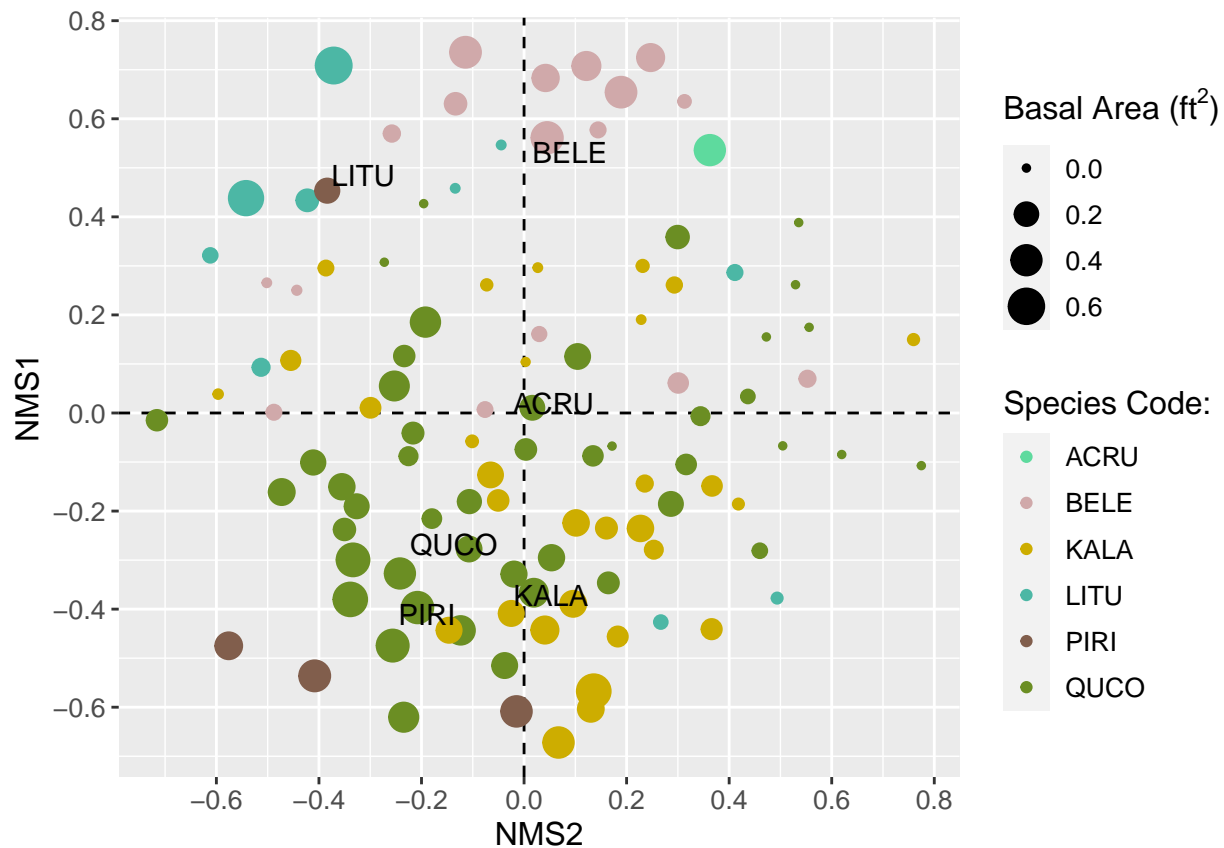
Figure 9 emphasizes species that sorted particular well on the elevation gradient. White oak (QUAL) appears with flowering dogwood (COFL) more often in lower elevation sites, where great laurel (RHMA) and black locust (ROPS) appear at much higher elevations.
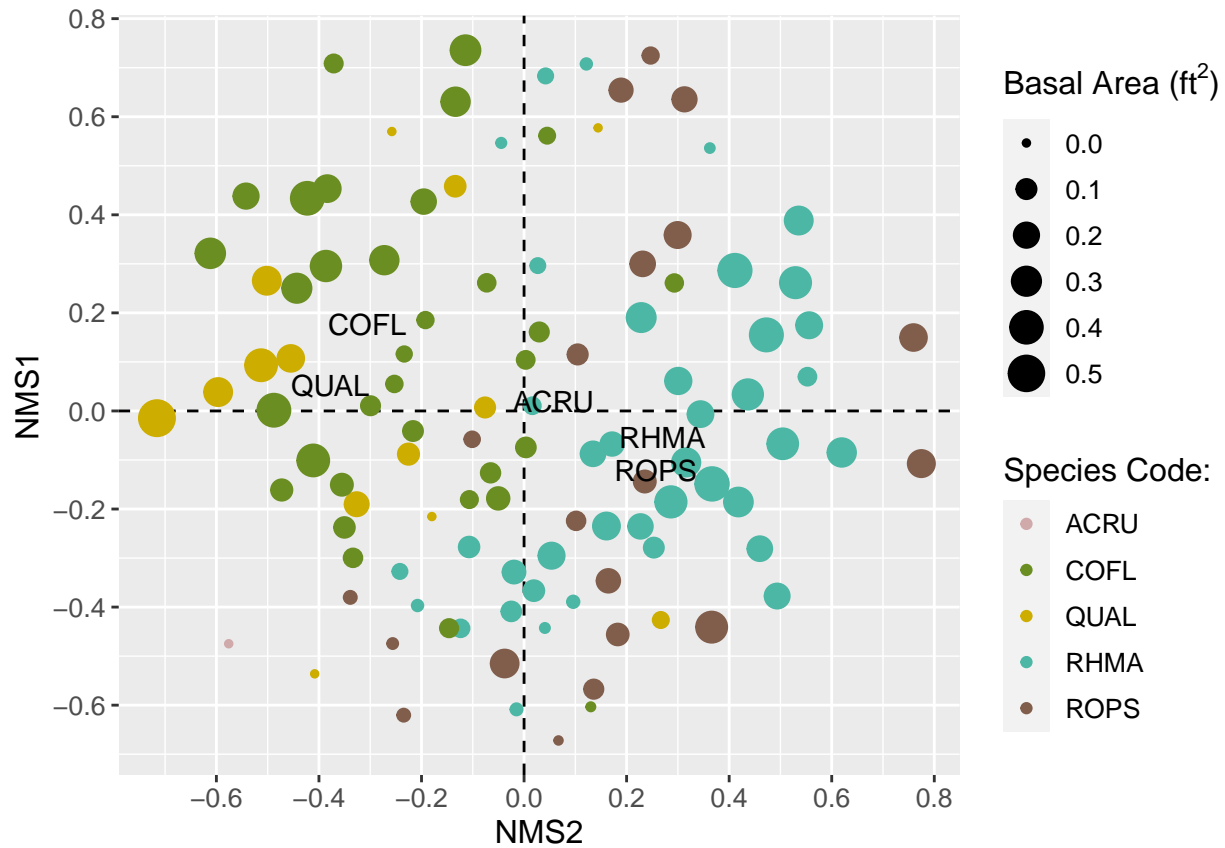

Figure 8: Tree species plotted in NMS space to illustrate sorting along the elevation gradient (roughly NMS2).

5   Summary and Conclusions

Overall, these results suggest that in the Coweeta system, soil moisture and composition plays a substantially stronger role in sorting species than elevation does. The relatively light elevation gradient in the southern Appalachians (compared to the profound gradient in the Sierran system as Urban et al. studied, for example), clearly affects the pattern of different trees, but not as substantially as soil. With data of environmental variables more attuned specifically to soil moisture, it might be possible to verify more specifically that this factor is what sorts tree species in this space.

6 References

Goslee, S.C. and Urban, D.L. 2007. The ecodist package for dissimilarity-based analysis of ecological data. Journal of Statistical Software 22(7):1-19.

Jari Oksanen, F. Guillaume Blanchet, Michael Friendly, Roeland Kindt, Pierre Legendre, Dan McGlinn, Peter R. Minchin, R. B. O'Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens, Eduard Szoecs and Helene Wagner (2020). vegan: Community Ecology Package. R package version 2.5-7. https://CRAN.R-project.org/package=vegan

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K.(2021). Cluster: Cluster Analysis Basics and Extensions. R package version 2.1.2.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Urban, D. L., Miller, C., Halpin, P. N., & Stephenson, N. L. (2000). Forest gradient response in Sierran landscapes: the physical template. Landscape Ecology, 15(7), 603-620