

SEARCHING FOR EASTERN OLD GROWTH:  
MODELING PRIMARY FOREST IN WESTERN NORTH CAROLINA USING TERRAIN  
ATTRIBUTES AND MULTISPECTRAL SATELLITE IMAGERY

By

Jennifer Hushaw  
Dr. Jennifer Swenson, Advisor  
December 2010

Masters project submitted in partial fulfillment of the  
requirements for the Master of Environmental Management and Master of Forestry degree in  
the Nicholas School of the Environment of  
Duke University

2010

## Table of Contents

ABSTRACT .....	3
INTRODUCTION .....	4
Ecological Importance of Old Growth Forests .....	4
State of Old Growth in the Eastern United States .....	6
Characterizing Primary Forests .....	7
Significance of this Research .....	9
Objectives .....	9
METHODS .....	10
Study Area .....	10
Environmental Predictor Variables .....	11
Predictive Modeling .....	13
RESULTS .....	16
Accuracy Assessment and Model Validation .....	16
Classification and Regression Tree (CART) .....	17
Maximum Entropy (MaxEnt) .....	17
DISCUSSION .....	20
Comparison of CART and MaxEnt .....	20
Environmental Variables .....	21
Implications of Study .....	23
Future Research .....	24
Conclusion .....	25
ACKNOWLEDGEMENTS .....	26
TABLES AND FIGURES .....	27
LITERATURE CITED .....	44
APPENDIX A: Data Acquisition and Software .....	47
APPENDIX B: Landsat Image Processing .....	54
APPENDIX C: CART Script .....	56
APPENDIX D: Complete MaxEnt Output .....	57

## **ABSTRACT**

After centuries of timber harvest and conversion of forest to farmland and development, only small pockets of old growth forest remain in the eastern United States. These remnant portions of older forest have intrinsic value as a rare forest type and they play an important ecological function on the landscape. However, old growth forests in the eastern U.S. are less well-studied and documented than their counterparts in the Pacific Northwest. This study was undertaken to predict the geographic location, ecological and spectral characteristics of existing old growth, specifically in the southern Appalachian forests of western North Carolina. Stands of old growth previously field validated by the Southern Appalachian Forest Coalition were used as the response variable. Predictor variables included a range of landscape, topographic, and satellite indices derived from Landsat TM 7 satellite imagery and terrain analysis. Predictions were made using Classification and Regression Tree (CART) and Maximum Entropy (MaxEnt) modeling techniques. Model results were successful based on validation with existing field data. However, the MaxEnt model produced the most realistic estimate of potential old growth area given the inherent rarity of this forest type and suitability of the MaxEnt modeling technique for predicting the distribution of rare species. Results highlight over 54,000 hectares of potential old growth to be investigated by researchers on the ground. This analysis will contribute to the relatively limited body of knowledge about old growth in the eastern U.S. and is unique in terms of its broad geographic extent. Continued research on these remnant eastern old growth stands must be done to increase our understanding of this rare forest stage and to better inform related management decisions on both public and private land in the eastern U.S.

## **INTRODUCTION**

Old growth forests are iconic in the history of the environmental movement. Concern over the fate of remaining old growth has sparked decades of debate over the conflicting values of stakeholders engaged in the discussion about how to manage (or not manage) these important natural resources. The issue and ecology of old growth has been most well-publicized and well-studied in the Pacific Northwest United States. However, some old growth forests remain in the eastern U.S. as well, with some estimates citing 482,000 acres in the southeast region alone (Davis 1996). The old growth issue is equally as important in the eastern region and the challenge of locating and identifying remaining old growth considerably more difficult. There is a need for large-scale analyses that will help land management agencies and conservation organizations identify remaining old growth stands, in order to facilitate the study and preservation of these unique forests. This analysis is meant to act in this capacity, by utilizing satellite imagery and terrain attribute data in a geographic information system (GIS) to develop predictive models of old growth forest that will serve to target potential old growth sites over a large geographic area

### **Ecological Importance of Old Growth Forests –**

Numerous studies focus on conditions of old growth forests, revealing that these forests are inherently different from anthropogenically altered second growth stands. These old growth forests can differ from early successional and secondary forest in a variety of ways, including structural complexity (Rapp 2003), species diversity, species composition (Jackson *et al.* 2009; Guyon *et al.* 2003), primary productivity (Gower *et al.* 1996), and carbon uptake (Zhou *et al.* 2006; Harmon *et al.* 1990). These differences are still an area of active research, but there is

general agreement that old growth forests provide unique habitat conditions that are important as part of the landscape mosaic.

Old growth forests are ecologically significant for many reasons. One important characteristic is their structural complexity, compared to younger and more managed forests. This complexity comes from the presence of snags and downed coarse woody debris, which accumulates as trees die and topple over. Structural complexity can also occur in the form of more intricate crown structures, especially in the long-lived tree species of the Pacific Northwest. Finally, old growth forests tend to have a more complex age structure than forests at other successional stages, with patches of different age classes present within one forest. This structural diversity is especially important because it leads to ecological diversity (Rapp 2003).

At the landscape scale, the presence of old growth forests, along with forests in earlier stages of succession, provides a range of habitat types that contribute to overall diversity (Rapp 2003). There are also instances in which it has been shown that old growth forests harbor greater species diversity or a higher incidence of endangered species than other forest conditions (Tikkanen *et al.* 2009). This may be especially true for species that need the structural characteristics found primarily in these types of forests, such as the well-known spotted owl (Mills *et al.* 1993). In fact, it has been stated that the loss of old-growth attributes in forests may result in a long-term loss of diversity (Halpern and Spies 1995).

Finally, old growth forests are important for scientific research because they represent reference conditions that are the necessary basis for both ecological studies and restoration efforts. It is difficult, if not impossible, to understand how humans are changing the landscape without having some untouched resource that acts as a point of comparison with current

conditions. Likewise, understanding the way ecosystems function when they remain in a relatively natural state is essential for developing old growth restoration plans.

### **State of Old Growth in the Eastern United States –**

In the United States, the well-known forests of the Pacific Northwest contain some the most documented and easily recognizable old growth in the world. In fact, it might be said that these forests have significantly influenced the common perception of what an old growth forest should look like. In actuality, old growth forests are difficult to describe in a generic way and they can look very different depending on forest type and disturbance history.

While old growth characteristics in the Pacific Northwest are a common research topic, there has been far less research done on the remaining old growth deciduous forests of the southeastern United States. This is partly because of a common conception that there is no old-growth left in the eastern U.S., beyond a select few preservation areas such as the Joyce Kilmer-Slickrock Wilderness Area.

Locating and identifying remaining old growth in the southeast can be problematic. The primary way to identify candidate old growth sites is to consider stand data, topography, adjacency to roads, and site nominations from local grass-roots ecology groups. Once potential sites have been pinpointed, field work is necessary to evaluate whether a particular stand qualifies as old growth (Messick 2004). Some of this work in the southern Appalachians was begun by Rob Messick in 2000 and has since been carried on by the Southern Appalachian Forest Coalition (SAFC).

The working definition of old growth used in this analysis is based on the classification system used by SAFC. Their system categorizes old growth sites into one of three classes: Class

A, Class B+, or Class B (the following descriptions are taken directly from Rob Messick's 2000 old growth forest report):

Class A: Old growth forest where no significant signs of human disturbance to the forest canopy or understory could be determined. Canopies are dominated by older trees generally over 150 years of age. (One hundred and fifty years is considered an appropriate coarse filter for old-growth candidacy as this corresponds to a period when logging was limited to areas near early settlement sites.)

Class B+: Old growth forests that have both class A and class B characteristics. Sites in this class tend to be large, with numerous forest communities, making it difficult to categorize the whole site. Uncut forests with canopy trees at or above 150 years may be present in these sites, yet the effects of disturbances such as blow downs, American chestnut blight, or fire may be present in other forest communities within the site.

Class B: old-growth forests exhibiting one of two different conditions:

- 1.) The canopy is dominated by old growth trees, yet signs of past human disturbance of the forest canopy or understory were found (generally a half century ago or longer). These stands have often been heavily impacted by American chestnut blight. Culling may also have occurred.
- 2.) No sign of past human disturbance could be confirmed, yet the forest canopy is dominated by younger forest. These stands can range from 100 to 150 years in age and were possibly affected by natural disturbances.

### **Characterizing Primary Forests –**

Heretofore, the term old growth has been used generally, but defining what actually constitutes old growth forest is an area of serious debate; a debate that warranted the Eastern

Old-Growth Definition Project, which manifested in a meeting with members of the US Forest Service and The Nature Conservancy in 1994. At this meeting, discussions led to list of criteria for defining old growth that included: density, basal area, number of four inch size classes, age and diameter of dominant and co-dominant trees, abundance of snags and coarse woody debris, and size and distribution of canopy gaps (White and Lloyd 1994).

As an example of the variability that exists in defining old growth, Shifley *et al.* (1995) describe three different schemes for describing/classifying old growth. They found that old growth has been classified in terms of the following characteristics: species diversity/species richness, age, size classes, number of canopy layers, presence of tree fall gaps, presence and amount of downed woody debris, and timber volume (in board feet). Another Forest Service publication by Rapp (2003) emphasizes that there are differing opinions among scientists and environmentalists about what constitutes old growth and that definitions must be flexible and dependent on the disturbance history in a given location.

One of the most commonly used indicators of old growth is the age of the trees. However, the exact age that constitutes old growth is not necessarily consistent or agreed upon. The Southern Appalachian Forest Coalition, which provided the old growth data used in this analysis, highlights several additional features that define old growth forest in the east and contribute to its unique character. These are: trees of great age, uneven-aged canopy structure, downed logs, standing snags, tree fall gaps, pit and mound topography, undisturbed soil, diversity of plants and animals, and little or no evidence of human disturbance (SAFC 2009). A goal in undertaking this analysis was to capture a number of these old growth characteristics using environmental variables derived from landscape, topographic, and satellite data.

## **Significance of this Research –**

This analysis will contribute to the search for old growth forest in western North Carolina by identifying areas with a high probability of being in old growth condition. The utilization of GIS and remote sensing allows for collection and analysis of data over a very large geographic area, with no cost for data acquisition and in far less time than it would take to conduct the same analysis via on-the-ground surveys. Lastly, the potential old growth areas delineated through this analysis will help direct future field surveys.

## **Objectives –**

My analysis focused on addressing two primary research questions:

(1) Where are forests in western North Carolina that remain in old growth condition, and in what quantity? (2) What are the attributes unique to these old growth forests, in terms of geographic distribution and spectral characteristics?

The overall objectives for this analysis include:

- 1.) Development of concrete metrics for the description of old growth in southern Appalachian forest types, related to the geographic pattern of their distribution, unique spectral characteristics, or ecological features.
- 2.) Development of a predictive old growth model that will have utility for land management agencies -- non-profits and the Forest Service alike. The predictions may be used to inform management on Forest Service lands by indicating areas having a likelihood of old growth remnants. These areas can then be ground-checked and excluded from logging projects.
- 3.) Development of effective environmental predictors and modeling techniques for prediction and delineation of old growth in the southeastern United States.

## **METHODS**

### **Study Area –**

The study area for this analysis is located in the mountain region of western North Carolina (Figure 1). It is 1,415,844 hectares in size and consists of 72% privately-owned lands, 25% federal lands, 2% state, and 1% lands held by conservation groups, land trusts, or other non-profits (Table 1). Other notable features within the study area include the developed area in and around the city of Asheville and portions of the Pisgah and Nantahala National Forests. The area is encompassed by one Landsat Thematic Mapper 7 (SLC-on) image (Path 18, Row 35) from May 23, 2003 with 0.06% cloud cover.

According to the US Geological Survey National Land Cover Dataset (USGS 2010), the dominant forest type in the area is deciduous forest. A more detailed map of ecological systems, the GAP Analysis project (SEGAP 2010), indicates that the most common forest types in this area are southern and central Appalachian oak forest and xeric oak forest (> 38 % of the study area; Table 2).

As stated previously, the SAFC old growth sites were identified through anecdotal evidence and on-the-ground surveys; they were subsequently mapped using GPS data collected in the field and by digitizing the boundaries of the old growth sites based on topographic maps. Examination of SAFC's old growth stands in a GIS, along with a USGS GAP Analysis map, revealed that the total amount of old growth forest that has been documented thus far in North Carolina consists primarily of southern and central Appalachian oak forest and xeric oak forest. The study area contains 22,340.83 hectares of old growth forest delineated by the SAFC, with 16,656.5 hectares of this classified as Class A or B+ and 5,684.3 hectares as Class B.

In order to capture the characteristics of old growth and non-old growth sites within the study area, I utilized a presence/absence sampling scheme. Using Hawth's Tools in ArcGIS I generated 1,000 random pseudo-absence points; these were located throughout the study area, with a minimum distance between points of 100 m. Similarly, I generated 1,000 random presence points; these were restricted to areas within SAFC Class A and B+ old growth sites, again with a minimum distance between points of 100 m. These sample points were loaded into the Sample Tool in ArcGIS and used to 'sample' a series of environmental variables (described below). The output of this sampling is a database file that was subsequently edited in Excel and saved as a comma delimited file for input into the following models.

### **Environmental Predictor Variables –**

The environmental variables used to predict old growth forest in this analysis can be divided into three broad categories: satellite (Figure 2), topographic (Figure 3), and landscape variables (Figure 4). These variables are aimed at capturing features of the vegetation itself or describing the topographic and/or ecological conditions where old growth remains in this region of western North Carolina. Variables derived from Landsat TM 7 imagery were used to describe the spectral characteristics of vegetation; these included Normalized Difference Vegetation index (NDVI), Enhanced Vegetation Index (EVI), Principle Component Analysis (Components 1-3), Tasseled Cap (Brightness, Greenness, and Wetness), Landsat TM 7, Band 8 (Panchromatic with a resolution of 15m) Texture Index, and Band 7 (Thermal).

The NDVI provides information about the level of photosynthetic activity across the landscape. EVI provides a similar type of information, but it is more sensitive in high biomass areas than NDVI (Huete *et al.* 1997). EVI was originally designed for use with MODIS data, but it was adapted for this analysis by utilizing the following equation, with coefficients taken from

Miettinen (2007):  $EVI = [(r_{nir} - r_{red}) / (r_{nir} + 6*r_{red} - 7.5*r_{blue} + 1)] * 2.5$ . PCA describes the majority of variation in the data and can be a useful tool for identifying patterns among environmental variables (Jolliffe 2002). The three Tasseled Cap components collectively account for 95% or more of the total variation in the data, where brightness is associated with soil reflectance characteristics, greenness characterizes the reflectance characteristics of green vegetation, and wetness corresponds with variation in soil moisture (Crist and Kauth 1986). Finally, texture within a 3 x 3 cell window on Band 8 (high resolution panchromatic) provides a measure of the variability of the data over a small spatial scale. It was included in this analysis because old growth forests are known to be 'patchier' than younger forest, due to the increased prevalence of tree fall gaps, and texture has the potential to act as a useful measure of this spatial variability. Further details of image processing are supplied in Appendix B.

The presence or absence of old growth forest will depend on human influence in terms of the history of logging or other harvest activity, as well as environmental factors, such as ice storms, fires, pests and diseases, all of which have the potential to cause catastrophic disturbance on the landscape, effectively wiping out entire stands.

As metrics for the likelihood of historic logging, I utilized slope, elevation standard deviation (at several fine and coarse spatial scales), and Euclidean distance to roads. Slope is important because very steep areas are less likely to have been logged historically, due to the difficulty of harvesting and transporting lumber on steep slopes before the advent of more modern, mechanized logging equipment. Similarly, elevation standard deviation acts as a measure of the roughness of the terrain, with rougher terrain having a larger standard deviation than more gentle terrain. Finally, distance to roads provides information about how remote a particular forest stand is from development and the level of access.

Normalized Difference Wetness Index (NDWI), Topographic Convergence Index (TCI), Euclidean distance to major and minor waterways, analytical hillshading, and radiation estimated moisture levels and sunlight across the landscape. While elevation, elevation focal mean (at fine and coarse spatial scales), slope position (ridge top, midslope, valley, etc.), terrain shape (convex or concave), and geology provide additional information about the terrain, potential exposure to wind and storms, moisture levels, and other ecologically significant characteristics of the landscape.

All layers were derived at a 30 m resolution, with the exception of the Band 8 texture (15 m resolution), and projected into WGS 1984, UTM Zone 17N. A correlation matrix was generated for all predictor variables based on the sampled values from the 2,000 presence/absence random sample points. This was used to determine whether any variables were highly correlated with each other. The matrix indicated that there were 19 pairs of variables with a correlation coefficient greater than 0.9; not surprisingly, 10 of these were correlations between mean elevation at different focal windows or between mean elevation and the unaltered elevation layer (Appendix A). These results suggest that the majority of variables used in predictive modeling will not confound the model predictions with correlation between variables. Also, as noted in the results, none of the highly correlated variables appeared as important predictors in either of the modeling approaches employed in this analysis.

### **Predictive Modeling –**

For the purposes of this analysis, I adapted two modeling approaches that are commonly used for species and habitat distribution modeling. The two models, Classification and Regression Tree (CART) and Maximum Entropy (MaxEnt), are non-parametric, meaning there are no assumptions made about the distribution of the data and the model structure is determined

from the data itself. Therefore, these models are well-suited for describing the complex attributes of vegetation distribution, especially when we have limited information about the actual distribution of a species or forest type, as in the case of our old growth data.

CART is a discriminative modeling algorithm that works to distinguish between response variable groups based on their relationships with the corresponding predictor variables and produces a dichotomous decision tree. In my analysis, the goal was to develop a classification tree that distinguished between the binary response groups of old growth and non-old growth. CART accomplishes this by recursively partitioning the dataset into ‘pure’ subgroups based on the dependent variables, providing a threshold value for each split. This recursive partitioning method makes CART particularly suitable for analyzing complex ecological interactions where the system would be described using many qualifiers (Urban 2006), such as old growth occurs on steep slopes at high elevations *or* low elevation locations that have a north facing aspect *and* valleys that are far from roads, etc. Literature supports the utility of CART for modeling complex ecological data and the distribution of rare species (De’Ath and Fabricius 2000; Vayssières *et al.* 2000; Bourg *et al.* 2005)

The CART analysis was carried out using the statistical software R and a corresponding script provided by Dr. Dean Urban (Appendix C). A comma-delimited file containing the 1,000 presence sample points and 1,000 absence sample points, along with the values corresponding to each of the environmental variables, provided the basis for the classification tree. The initial tree contained a total of six branches. To ensure the model was not over-fit to the data, a pruning function was used and a cross-validation assessment was performed.

The cross-validation assessment provides information about the performance of the model by dividing the dataset into ten subsets, with each subset containing 90% of the data.

These subsets are then used to generate 10 independent trees that are ‘tested’ on the remaining 10% of the data. The graphic output of this process is a distinct curve indicating the cross-validation error rate for trees of different sizes. An optimal tree can then be chosen based on the tree size that occurs where the error rate is minimized, i.e. where the ‘trough’ appears in the curve. The cross-validation error graph produced for my data indicated that trees with four to six branches would minimize the error rate and any tree seven branches or larger would be over-fitted (Figure 5). Based on the above assessment and pruning methods, a final tree was developed that consisted of five branches in total (Figure 6).

The final step of the CART analysis was to transform the dichotomous tree into a map output. The threshold values corresponding to each split (or branch) in the tree were put into a Single Output Map Algebra in ArcGIS and used to create a binary map output, where areas with a value of 1 represent old growth and areas with a value of zero represent non-old growth.

The second modeling technique used in this analysis was the maximum entropy method of habitat modeling, which exists as the software program MaxEnt (Phillips *et al.* 2006). MaxEnt is a generative modeling algorithm that seeks to characterize a set of habitat samples according to any number of environmental predictors measured at those sample sites. Consequently, MaxEnt uses only presence points to generate the model; in the case of my analysis, this consisted of the 1,000 points generated randomly within polygons defined as old growth. Additionally, I chose to set aside 25% of the input presence points to use as test data for validation of the MaxEnt model.

The primary assumption made by MaxEnt is that the presence samples represent an unknown probability distribution, which will be generated by the model with the constraint that the mean of the distribution of each variable will be the same as the mean of the observed

(presence-only) data. Application of the MaxEnt model to prediction of various other species distributions has demonstrated the effectiveness of this technique for accurately predicting habitat and species occurrence (Hernandez *et al.* 2006; Elith *et al.* 2006; Hernandez *et al.* 2008).

## **RESULTS**

### **Accuracy Assessment and Model Validation**

Both of the modeling algorithms used in this analysis performed well. For MaxEnt, this is evidenced by the area under the test data curve in the generated ROC, which has a value of 0.935 (Figure 7). This indicates that the model was successful at correctly predicting the location of old growth, based on the old growth presence validation points held back from the model for use as testing data. For CART, a similar measure of accuracy is not readily available. All validated old growth areas were correctly predicted by the CART model, indicating an absence of omission error. However, we do not have definitive data on all old growth absence areas; therefore it is difficult to determine whether any CART predicted old growth areas constitute commission error.

Certain predictor variables also appeared as important contributors to both models, these included elevation standard deviation at 9 ha and 900 ha focal windows and Tasseled Cap brightness. These variables appeared in the CART decision tree and contributed greater than 1% to the MaxEnt model. The usefulness of these variables in both modeling techniques suggests that they are especially important for describing the location of old growth in this region.

## **Classification and Regression Tree (CART)**

In total, the CART model predicted 387,028.53 hectares of old growth in the study area (Figure 8). The most useful variables for discriminating between old growth and non-old growth, i.e. the variables that were utilized in developing the decision tree, were large-scale elevation standard deviation (at 9 ha and 900 ha focal windows), Euclidean distance to USGS 24k hydrology, and Tasseled Cap Brightness (Figure 6). According to a GAP Analysis map, 64% of the area predicted to be old growth by CART is southern and central Appalachian oak forest and xeric oak forest. These are also the two most dominant forest types within the study area, in terms of percent area. Interestingly, 2% of the area predicted to be old growth by CART is actually classified as developed open space by the GAP Analysis land cover map, which may be an indication that CART is over predicting the likelihood of old growth within the study area (Table 3). Additionally, 55% of the CART old growth area is privately owned, while 39% of it is owned and managed by the federal government, primarily the US Forest Service, and only 5% is held by state of North Carolina, conservation groups, land trusts, or other non-profit organizations (Table 4).

## **Maximum Entropy (MaxEnt)**

The MaxEnt output includes an ASCII file specifying a probability distribution map (Figure 9) and several graphic and tabular measures of variable contribution and model predictive strength (Appendix D). However, in the interest of comparison between CART and MaxEnt, I also converted this probability map into a binary (old growth presence/absence) map (Figure 10) based on a probability threshold (0.35) where 90% of the random old growth presence points were correctly predicted (Figure 11).

At a probability threshold of 0.35, MaxEnt predicted a total of 76,443.30 hectares of old growth. However, the total area of old growth predicted will vary depending on the probability threshold, with the hectares of old growth dropping drastically as the probability required for classification as old growth increases (Figure 12). The most useful variable for describing old growth forest, i.e. the variable with the highest percent contribution to the model, was large scale elevation standard deviation (at a 900 ha focal window), with a percent contribution of 40.7%. This was followed by geology, with a percent contribution of 26.6%. Since geology proved to be an important variable for describing old growth forest, it is interesting to see the dominant geologic formations in confirmed old growth areas. In fact, over 25% of all SAFC old growth sites in the study area have felsic metavolcanic rock as their majority geologic formation (Figure 13); this geologic formation is associated with an abundance of quartz and feldspar (USGS 2010). The third most important variable was elevation standard deviation at a 9 ha focal window with a percent contribution of 16.7%. Other variables contributing 1% or greater to the model included, mean elevation at a 900 ha focal window, Euclidean distance to roads, Tasseled Cap brightness, NDVI, and Euclidean distance to USGS 24k hydrology (Table 5).

The importance of these variables in terms of their unique contribution is also supported by the jackknife analysis carried out by the MaxEnt program. In a jackknife analysis, two scenarios are created for each environmental variable, one where the environmental variable is the only variable used in the model and one where a model is created that uses every variable except that environmental variable. This provides information about the importance of each predictor variable; environmental variables that present a high training gain when used individually are useful predictors that will perform fairly well even in the absence of other predictor variables. Likewise, if the training gain decreases when a particular variable is left out

of the model, this is an indication that that variable is providing unique information to the model, and the reduction in training gain is a measure of how much the predictive power of the model would be reduced if that variable were not included. Based on this interpretation, the graphic output of the jackknife analysis supports the assertion that geology and large scale elevation standard deviation (900 ha focal window) are important variables and corroborates the utility of the other variables that presented with a high percent contribution (Figure 14).

The old growth area predicted by MaxEnt at the 0.35 probability threshold consists of 62% southern and central Appalachian oak forest and xeric oak forest, according to a GAP Analysis map. The remaining MaxEnt old growth area consists of the same forest types that appeared in the CART old growth predicted areas, and in relatively similar proportions to those found in the CART results (Table 3). In contrast to CART, where the majority of area predicted to be old growth was privately owned, 72% of the MaxEnt old growth area is owned and managed by the federal government, primarily the US Forest Service. Another 16% is privately owned and 10% is held by conservation groups, land trusts, or other non-profit organizations (Table 4). As an example, focusing in on the area of predicted old growth near Mt. Mitchell State Park displays the prevalence of old growth within areas that are held by conservation organizations, which will be an important factor in our ability to investigate and protect any old growth forest remaining in those locations (Figure 15).

## DISCUSSION

### Comparison of CART and MaxEnt

A comparison of the old growth area predicted by the CART and MaxEnt modeling algorithms highlights the variability in results that can be achieved when using the same environmental predictors. Even at a very low probability threshold of 0.35, MaxEnt still predicts a far smaller area (310,585 fewer hectares) of old growth than the CART model. Determining which result is more accurate requires some expert knowledge of the region and familiarity with the study landscape, as well as an understanding of how these two modeling methods operate. Studies have demonstrated that MaxEnt performs well with a small number of presence samples (Hernandez *et al.* 2006) and that *Random Forests*, an iterative version of the CART modeling method, may perform better when the species (or in this analysis, forest type) being predicted exists in separate and discontinuous populations or groups. Therefore, each method has some advantage in terms of predicting old growth because it is a rare forest type and many additional locations are unknown. However, a significant advantage of the MaxEnt modeling method is that continuous probability predictions allow the user considerable flexibility in choosing a probability threshold that is reasonable based on expert knowledge of the subject, a benefit also recognized by Hernandez *et al.* 2008. All areas predicted to be old growth by the MaxEnt model at a 0.35 threshold were also predicted to be old growth by CART, which lends credence to the validity of the final MaxEnt predictions (Figure 16). My assessment of the results is that the MaxEnt 0.35 threshold prediction is the most reasonable estimate of old growth area in this analysis, based on the inherent rarity of this forest type and the more conservative area predicted by the MaxEnt model. If we use this estimate of old growth area and remove those areas that we

have already identified in the field, we are left with a total of 54,100 hectares of potential old growth that remains to be discovered within the study area (Table 6).

## **Environmental Variables**

Certain environmental predictors emerged as particularly important or useful for describing old growth or discriminating between old growth and non-old growth. These included large-scale elevation standard deviation (at 9 ha and 900 ha focal windows), mean elevation at a 900 ha focal window, geology, distance to roads and waterways, Tasseled Cap brightness, and NDVI.

The large-scale measures of elevation standard deviation are most likely important because these variables are a measure of terrain roughness, which will be related to the likelihood that an area was logged in the past. The same might be true for the variable of large-scale mean elevation. Distance to roads is meant to be a measure of remoteness and ease of access, which is also related to the likelihood of past logging. Due to the tendency for human development and agriculture to occur near available water, it may be that distance from waterways is useful for the same reason that distance to roads would be useful; areas further from water are less likely to be developed and therefore less likely to have experienced any human alteration of the forest landscape. Tasseled Cap brightness and NDVI probably appear in the models because they are useful for distinguishing between forested and unforested areas. Tassel Cap brightness is a measure of reflectance and areas of development or bare earth will appear much brighter than forested areas. Likewise, forest canopies will have a high value of NDVI, while water and other non-photosynthesizing surfaces will have a much lower value. For the most part, the appearance of these variables as useful predictors is not overly surprising given

that old growth forests exist in old growth condition because favorable environmental conditions and the absence of human influence have allowed them to persist for more than a century.

Less transparent, however, is the appearance of geology as the variable with the second highest percent contribution to the MaxEnt model. I hypothesize that this may have something to do with the effect of geology on underlying soil characteristics. Geology is one of the five soil forming factors; it provides the parent material for soil formation and has a significant influence on soil texture and the chemical and mineralogical composition of soils (Brady and Weil 2010). Soil characteristics will have an effect on site quality, which may influence the longevity and vigor of forest stands growing in an area, as well as the likelihood that an area was desirable for agricultural production (conversion of forest to farmland). As stated previously, 25% of all the SAFC old growth sites have felsic metavolcanic rock as the majority geology type. This type of rock contains large amounts of granite and feldspars, and these feldspars weather into the water-holding clays typical of many North Carolina soils (Buol 2010). The utility of geology as a predictor variable may also be related to the fact that known old growth sites tend to be located in more mountainous areas, which are also associated with particular geological formations. In fact, many of the rocks in the piedmont and mountain regions of the state are granitic and gneissic (Buol 2010), which is reflected in the predominance of these rock types in the assessment of common geology types associated with SAFC old growth sites (Figure 13).

It is interesting that the satellite-derived variables related to vegetation characteristics did not play a bigger role in either model. I would attribute this to the level of detail available in the spectral signatures from 7-band Landsat Thematic Mapper imagery, which may not have been sufficient to detect differences between forests of different age classes. In contrast, there have been studies that demonstrate the utility of hyperspectral satellite imagery for distinguishing

between old growth and younger forests. For example, a study by Franklin *et al.* (2001) successfully distinguished between different age classes of Douglas-fir using texture analysis of IKONOS panchromatic satellite imagery. Another study by Ustin and Trabucco (2000) utilized 224-band airborne visible infrared imaging spectrometer (AVIRIS) and found that the spectral signature for coniferous old growth forests was distinct from the signatures for herbaceous vegetation and young forest canopies; young forests had a higher reflectance across the spectrum than old growth. Both of these studies were conducted with a focus on coniferous forests. However, it is possible that the efficacy of the satellite-derived variables used in this analysis might be improved through the use of hyper-, rather than multi-spectral satellite imagery.

### **Implications of Study –**

In the future, the old growth predictions developed in this analysis may be used to target sites for field validation. This analysis is not meant to replace on-the-ground surveys for old growth in the Southern Appalachians. Instead, it is meant to provide an estimate of the geographic distribution of old growth in the region, in order to help guide field investigations. Potential sites must be visited on the ground and examined for the characteristics usually employed in defining old growth sites, such as tree age, evidence of human disturbance, presence of an uneven-aged canopy, tree-fall gaps, downed woody debris, and plant and animal diversity.

The assessment of land ownership in old growth areas has important implications for the future of any old growth stands remaining in these areas. Of the MaxEnt predicted old growth area, 10% is held by conservation groups, land trusts, and other non-profit organization. This is encouraging because it suggests that these areas are likely to be permanently protected from development pressure and are probably being managed with an eye toward conservation. Taking note of the different types of land ownership associated with predicted old growth areas may also

provide us with an opportunity to prioritize on-the-ground efforts, by focusing first on the potential old growth areas that are at greatest risk from logging or development.

### **Future Research –**

There is great potential for the use of Light Detection and Ranging (LiDAR) technology in describing the structural characteristics of old growth in the Southern Appalachian region. LiDAR, also known as laser altimetry, is an active remote sensing technology, in which a laser light pulse is emitted from a sensor, it reflects off a target and then returns to the sensor; allowing for a calculation of the distance and location of the target based on the travel time of the laser beam. The end result is a 3-D point cloud that can provide an accurate image of three-dimensional forest canopy structure. LiDAR remote sensing systems are often mounted on airborne platforms in planes, but there are some space-borne LiDAR systems currently in place on satellites (Lefsky *et al.* 2002).

Although LiDAR is often flown commercially for the purpose of creating accurate renderings of the topography of the earth's surface, researchers have recently discovered its usefulness in a variety of forest research and inventory applications. LiDAR has utility for many ecological applications, including assessment of forest canopy structure (Maier *et al.* 2009), forest biomass (Zhao *et al.* 2009), tree species identification (Kim *et al.* 2009), and other forestry applications (Dubayah and Drake 2009).

North Carolina is fortunate to have a state-wide LiDAR dataset, which might be utilized for assessing forest canopy characteristics in the state. However, the purpose of this LiDAR dataset was to create more accurate floodplain maps for the state, so the data were collected during leaf-off conditions in order to facilitate ground-returns (NC Floodplain Mapping 2003). Given that the majority of my study area consists of deciduous forest, this particular dataset may

not provide a very accurate picture of deciduous forest canopies, young or old. However, LiDAR-derived variables of forest stand structure, including vegetation density and canopy height, may be helpful predictors to include in any future iterations of this analysis.

## **Conclusion**

This analysis provides an encouraging assessment of the potential area of old growth forest remaining in western North Carolina. The maps of potential distribution will be useful for guiding the location of future field surveys in the region. The environmental variables that emerged as particularly useful for modeling old growth forests were not overly surprising; they support our existing knowledge and assumptions about the role of topography and proximity to human influence in determining the location of remaining old growth.

These forests have great cultural, ecological, and scientific value. Given their importance as part of the landscape mosaic, the task of identifying and protecting these surviving pockets of untouched forest is critical. This analysis demonstrates that these environmental variables and the MaxEnt model have great potential to further the search for eastern old growth through the identification of potential old growth ‘hotspots.’

## **ACKNOWLEDGEMENTS**

I would like to extend a special thank you to my advisors, Dr. Jennifer Swenson and Dr. Dean Urban, for their assistance with this project. Thanks to Jennifer Swenson for providing me with continual guidance and support from conception to completion of this project and Dean Urban for lending me his advice and expertise, especially with regard to modeling techniques. I would also like to thank Mr. Fred Stanback and the Stanback Internship Program for funding my summer internship with the Southern Appalachian Forest Coalition in Asheville, North Carolina. Also, thank you to Hugh Irwin and Josh Kelly of the Forest Coalition for inspiring this project and providing valuable input throughout the analysis. Finally, a huge thank you to my family and friends, who have provided me with support and encouragement every step of the way.

TABLES AND FIGURES

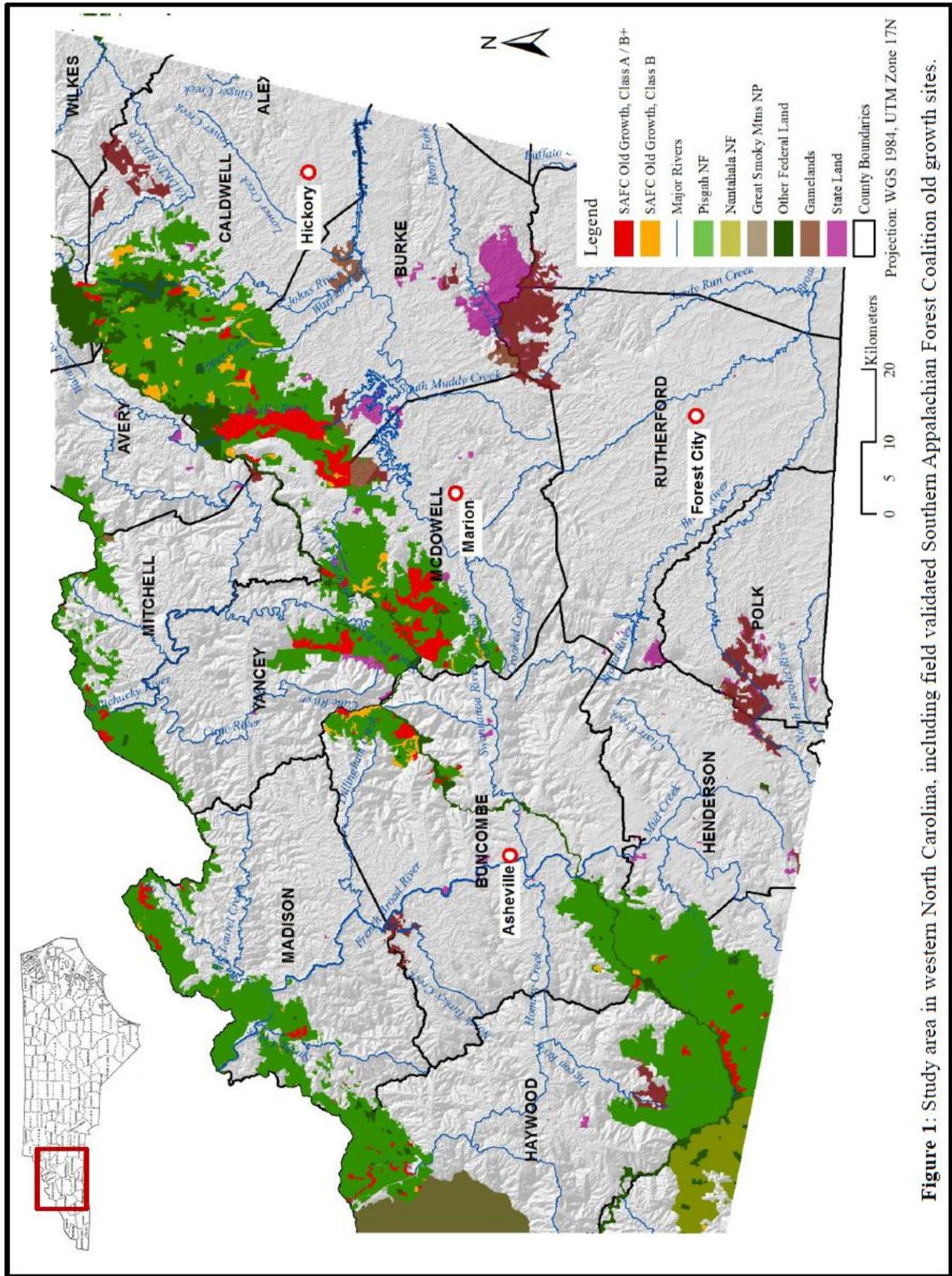


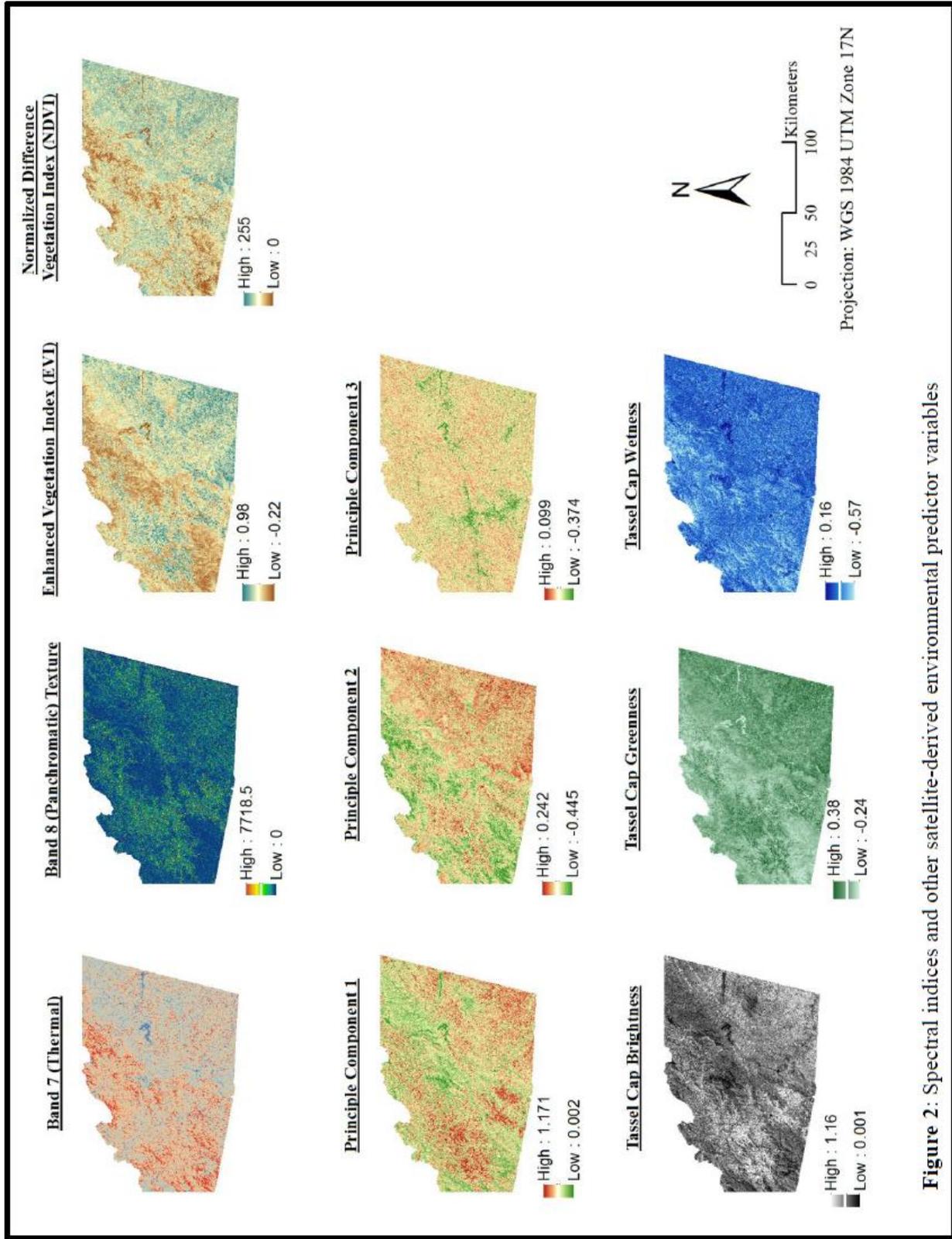
Figure 1: Study area in western North Carolina, including field validated Southern Appalachian Forest Coalition old growth sites.

**Table 1:** Percentage of study area in different land ownership classes.

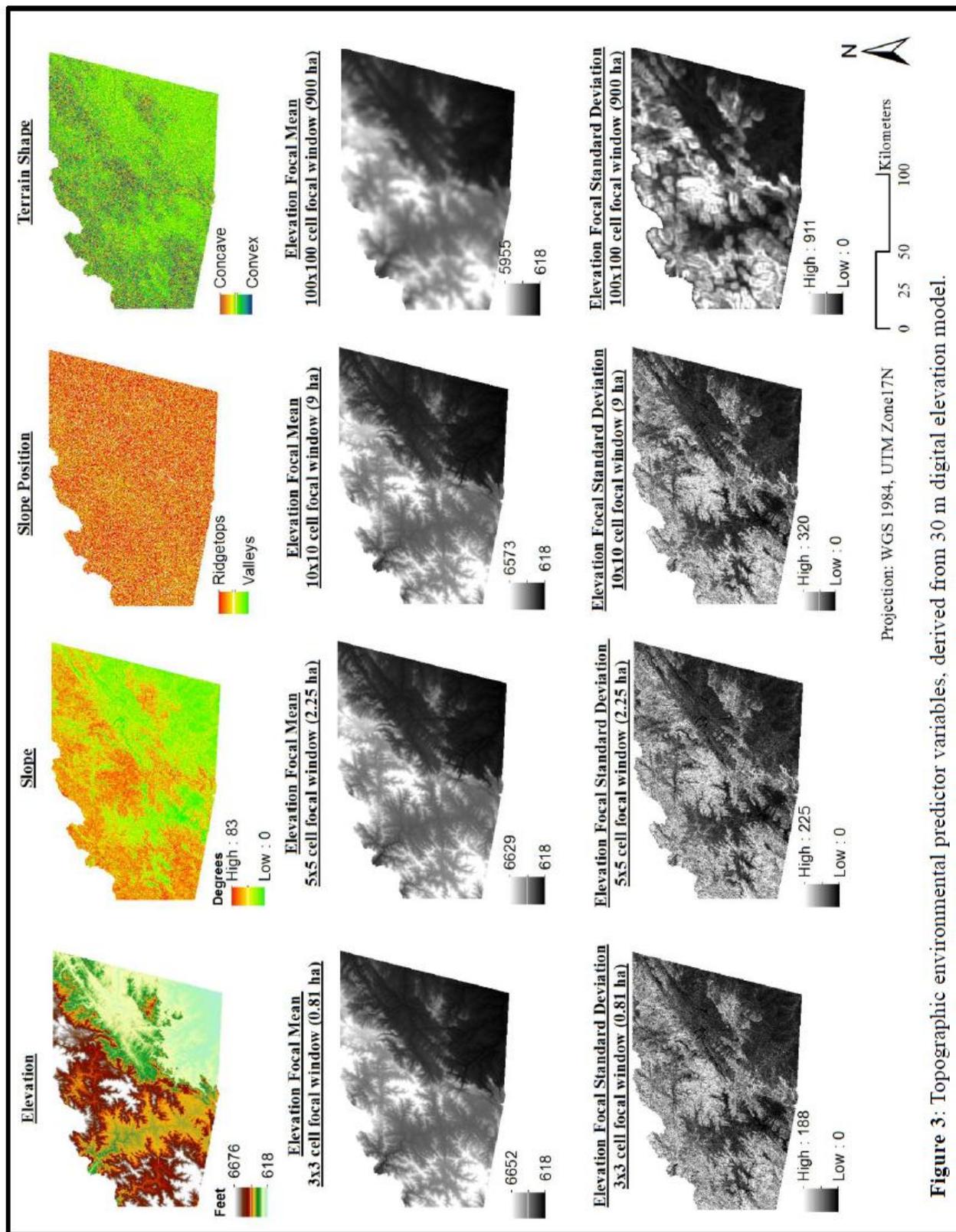
Land Ownership	Area (hectares)	Percentage of Study Area
Private	1,022,787.32	72%
Federal	353,939.96	25%
State	24,949.92	2%
Conservation Group/Land Trust /Other Non-Profit	14,167.11	1%

**Table 2:** Percentage of study area in different GAP Analysis land cover/forest type categories.

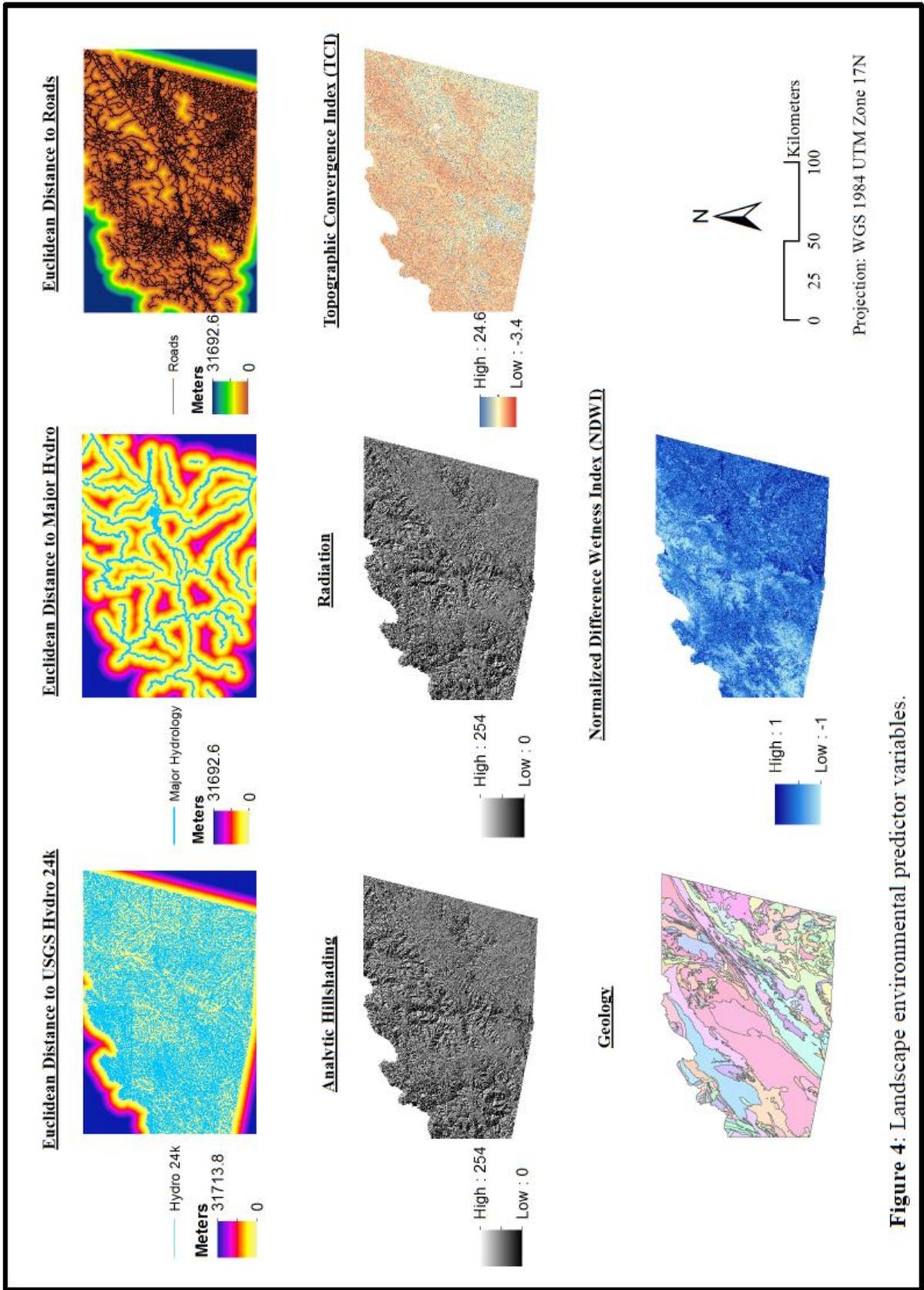
GAP Analysis Land Cover Type	Area (hectares)	Percentage of Study Area
Southern and Central Appalachian Oak Forest	314725.50	22.2%
Southern and Central Appalachian Oak Forest - Xeric	229240.62	16.2%
Pasture/Hay	172457.73	12.2%
Southern and Central Appalachian Cove Forest	110614.95	7.8%
Southern Piedmont Dry Oak-(Pine) - Hardwood Modifier	106087.59	7.5%
Developed Open Space	101846.34	7.2%
Evergreen Plantations or Managed Pine	49478.49	3.5%
Central and Southern Appalachian Montane Oak Forest	45879.03	3.2%
Other - Herbaceous	33184.53	2.3%
Central and Southern Appalachian N Hardwood Forest	32639.40	2.3%
S Appalachian Low Mountain Pine Forest	32270.31	2.3%
Appalachian Hemlock-Hardwood Forest	28339.65	2.0%
Low Intensity Developed	22087.17	1.6%
Successional Shrub/Scrub (Utility Swath)	16919.82	1.2%
Successional Shrub/Scrub (Other)	16523.37	1.2%
South-Central Interior Small Stream & Riparian	15498.63	1.1%
Southern Piedmont Dry Oak-(Pine) - Loblolly Modifier	13824.09	1.0%



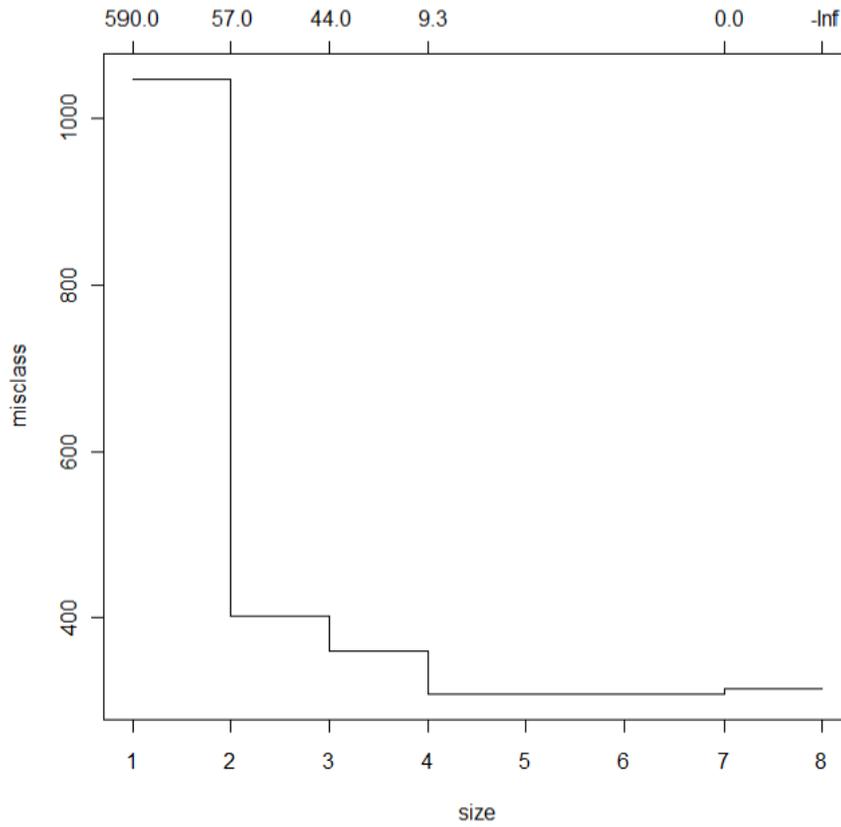
**Figure 2:** Spectral indices and other satellite-derived environmental predictor variables



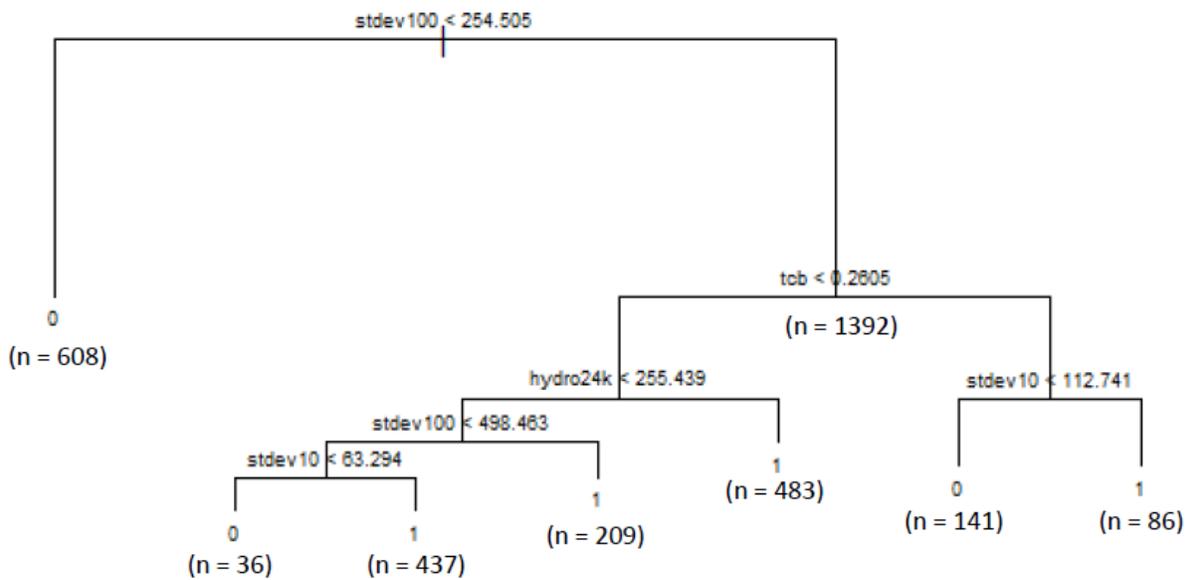
**Figure 3:** Topographic environmental predictor variables, derived from 30 m digital elevation model.



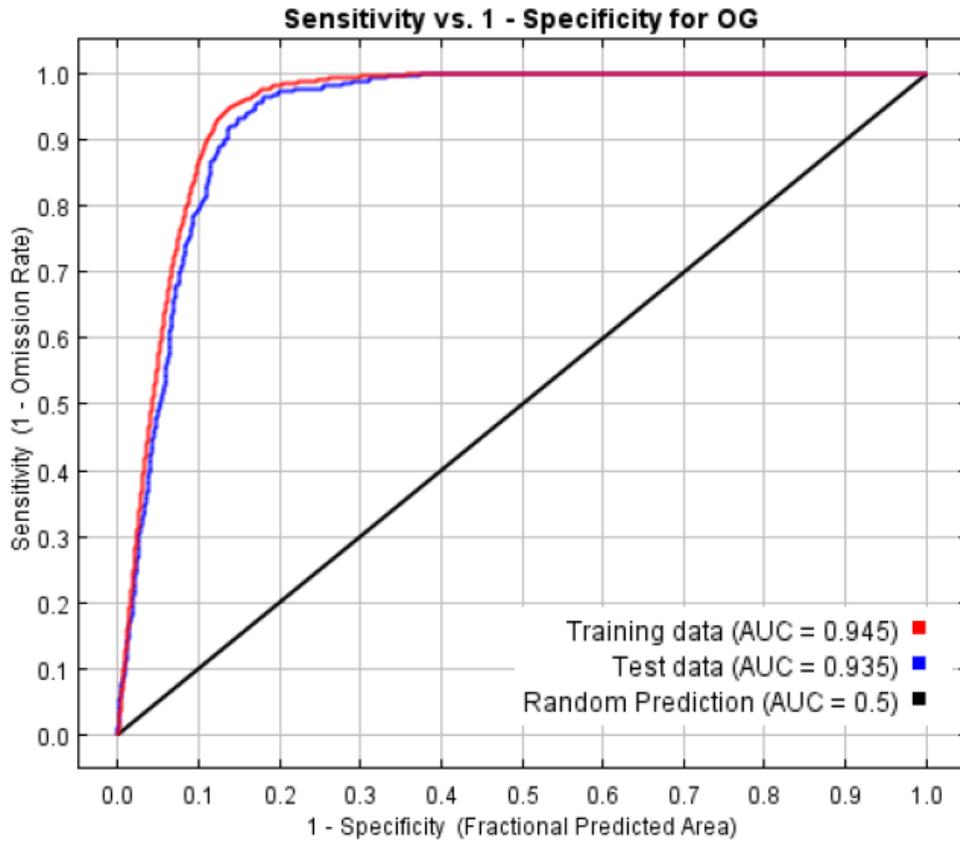
**Figure 4:** Landscape environmental predictor variables.



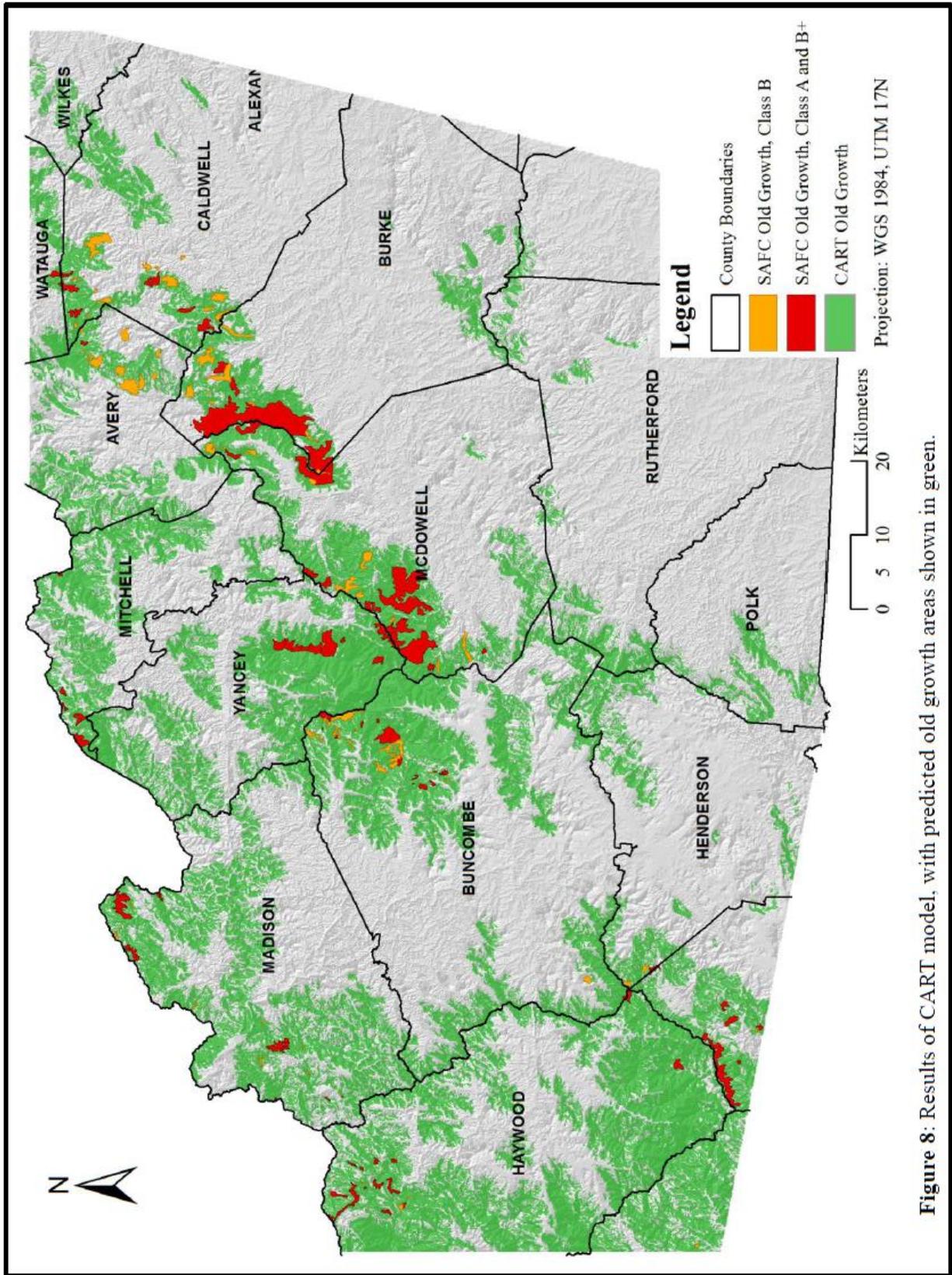
**Figure 5:** Cross-validation error graph, showing minimization of misclass error with trees 4 to 6 branches in size.



**Figure 6:** Final CART decision tree, with threshold values and number of observations at each node.



**Figure 7:** Receiver Operating Characteristic (ROC) curve indicating strong performance of the MaxEnt model on test data based on an AUC of 0.935.



**Figure 8:** Results of CART model, with predicted old growth areas shown in green.

**Table 3:** Percentage of different GAP Analysis land cover types in all of North Carolina, the study area, CART predicted old growth, and MaxEnt predicted old growth.

GAP Analysis Land Cover Type	Percent Area of			
	North Carolina	Study Area	CART Old Growth	MaxEnt Old Growth (0.35 Threshold)
Southern and Central Appalachian Oak Forest	5%	22%	37%	36%
Southern and Central Appalachian Oak Forest - Xeric	3%	16%	27%	26%
Pasture/Hay	10%	12%		
Southern and Central Appalachian Cove Forest	2%	8%	11%	8%
Southern Piedmont Dry Oak-(Pine) - Hardwood Modifier	10%	7%		
Developed Open Space	6%	7%	2%	
Evergreen Plantations or Managed Pine	7%	3%		
Central and Southern Appalachian Montane Oak Forest	1%	3%	7%	6%
Other - Herbaceous	3%	2%		
Central and Southern Appalachian N Hardwood Forest	<1%	2%	7%	9%
S Appalachian Low Mountain Pine Forest	<1%	2%	2%	3%
Appalachian Hemlock-Hardwood Forest	<1%	2%	3%	5%
Low Intensity Developed	2%	2%		
Successional Shrub/Scrub (Utility Swath)	<1%	1%		
Successional Shrub/Scrub (Other)	2%	1%		
South-Central Interior Small Stream & Riparian	<1%	1%		
Southern Piedmont Dry Oak-(Pine) - Loblolly Modifier	1%	1%		
Central and Southern Appalachian Spruce-Fir Forest	<1%	<1%	1%	2%

**Table 4:** Percentage of predicted CART and MaxEnt old growth areas under different types of land ownership.

Land Ownership	CART		MaxEnt	
	Area (hectares)	Percentage of Predicted Old Growth	Area (hectares)	Percentage of Predicted Old Growth
Private	211,484.16	55%	12,608.46	16%
Federal	152,127.00	39%	55,206.00	72%
State	7,539.48	2%	152.46	0%
Conservation Group / Land Trust / Other Non-Profit	11,746.35	3%	7,641.63	10%

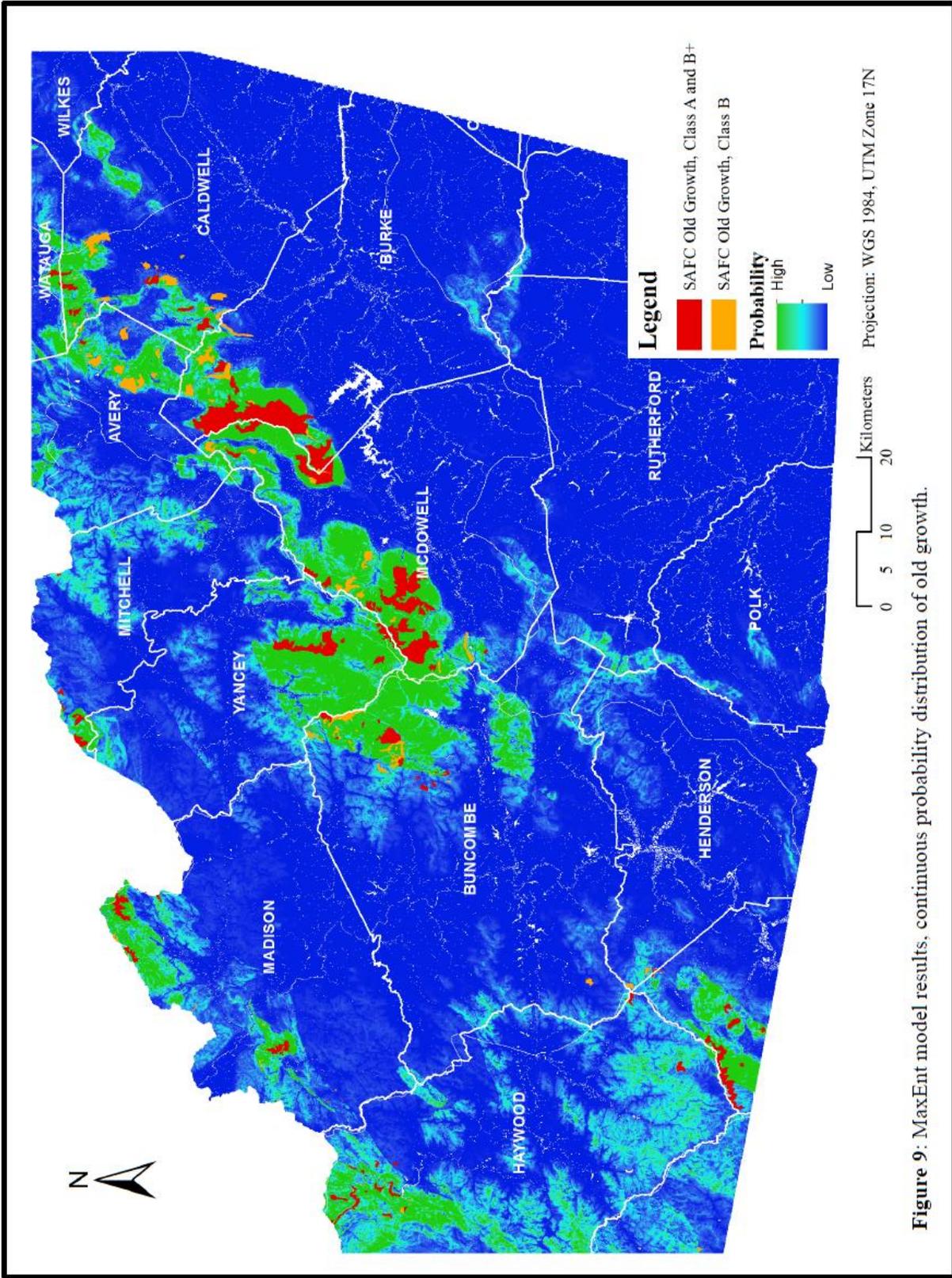
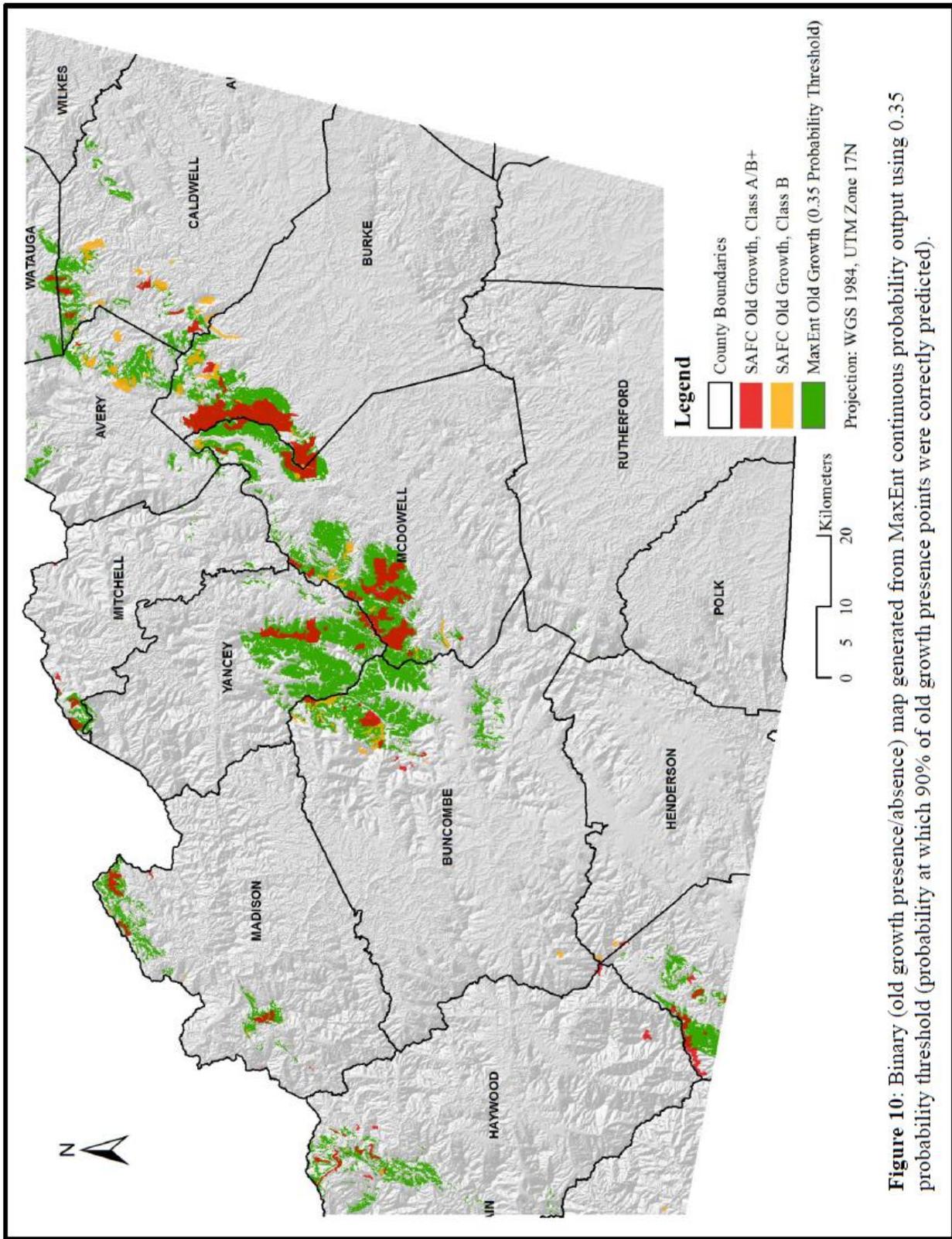
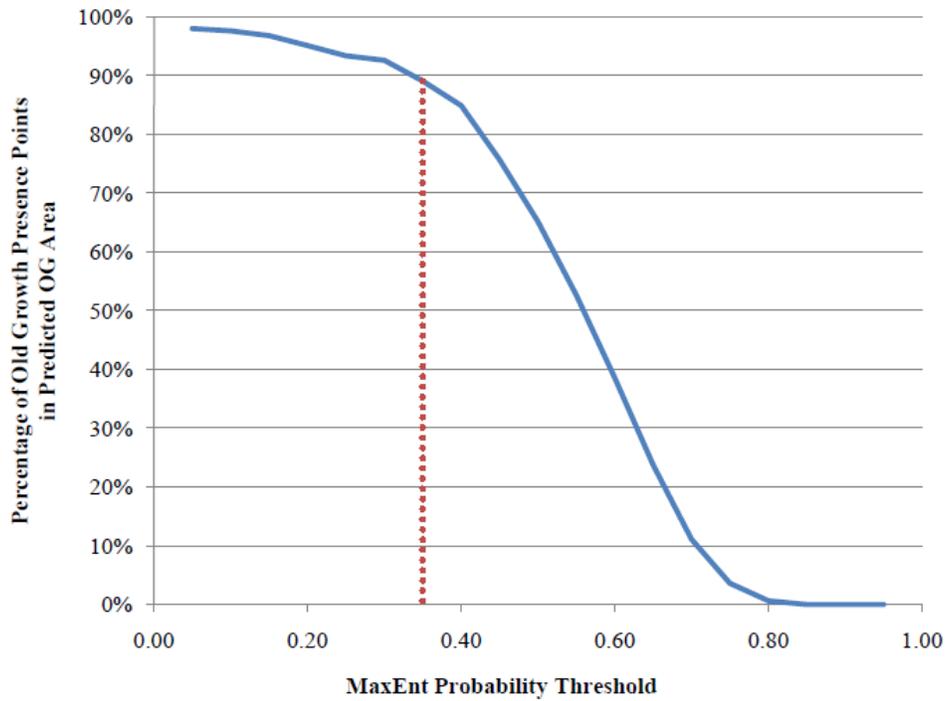


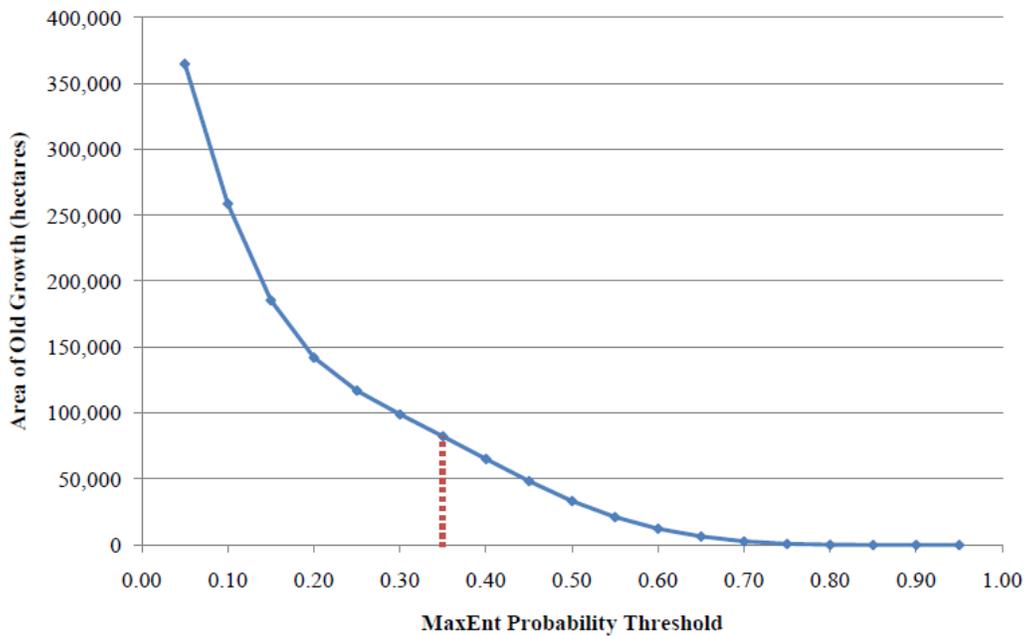
Figure 9: MaxEnt model results, continuous probability distribution of old growth.



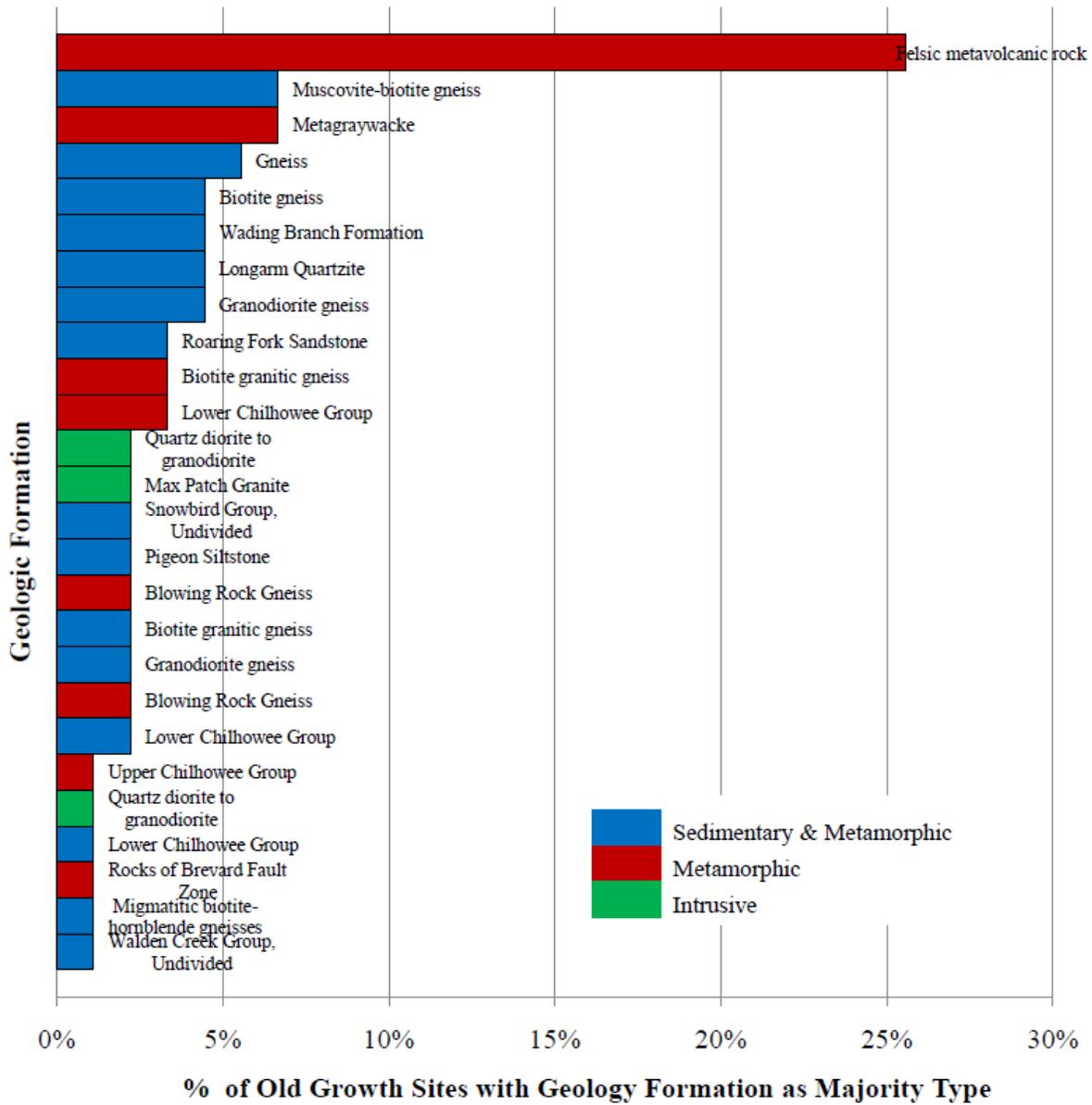
**Figure 10:** Binary (old growth presence/absence) map generated from MaxEnt continuous probability output using 0.35 probability threshold (probability at which 90% of old growth presence points were correctly predicted).



**Figure 11:** Graph of the change in percent of old growth validation (presence) points correctly predicted by the model at varying probability thresholds. Chosen threshold value of 0.35 (90% correct) indicated by dotted line.



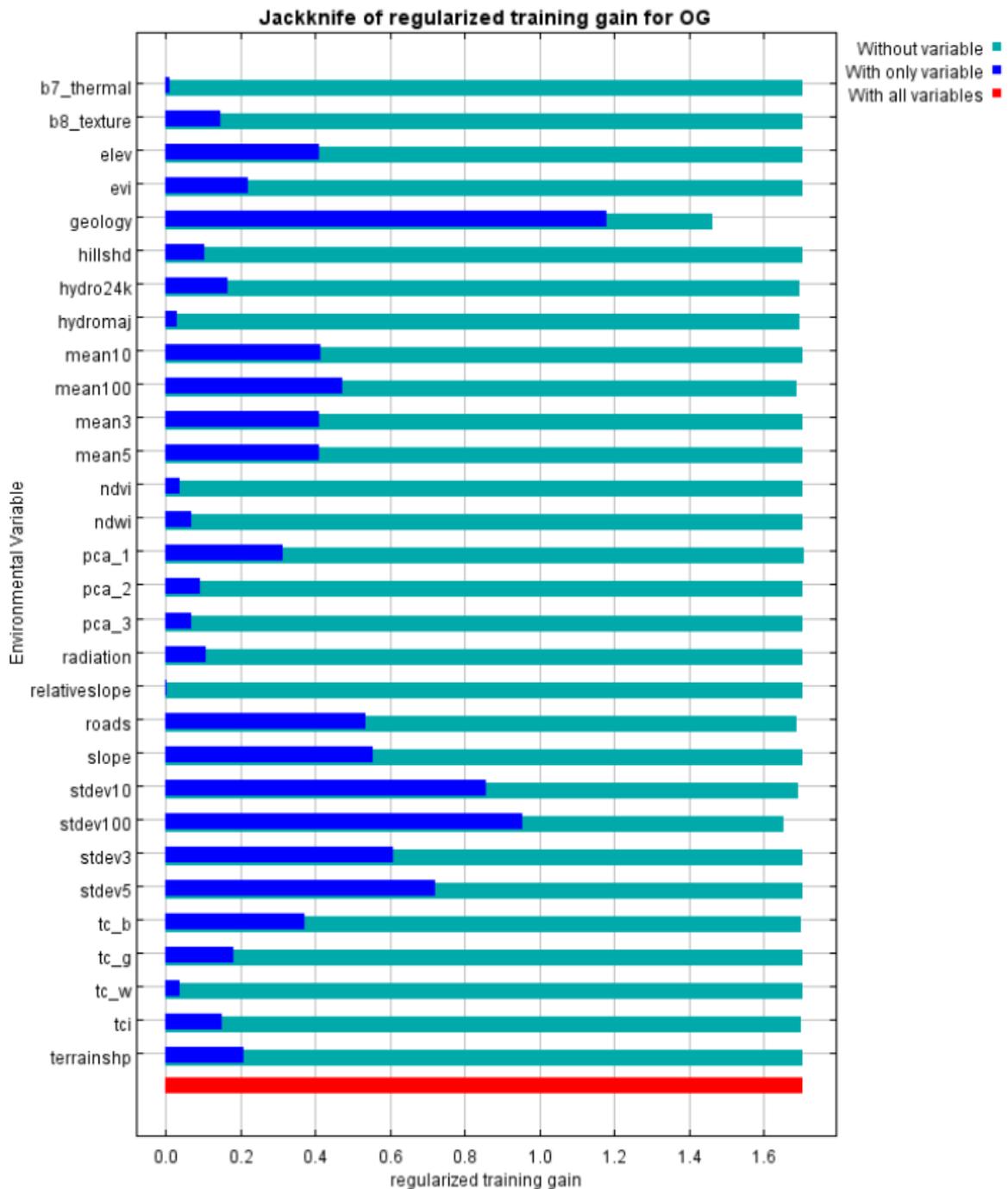
**Figure 12:** Change in hectares of old growth predicted with increasing MaxEnt probability threshold. Chosen probability threshold of 0.35 indicated by dotted line.



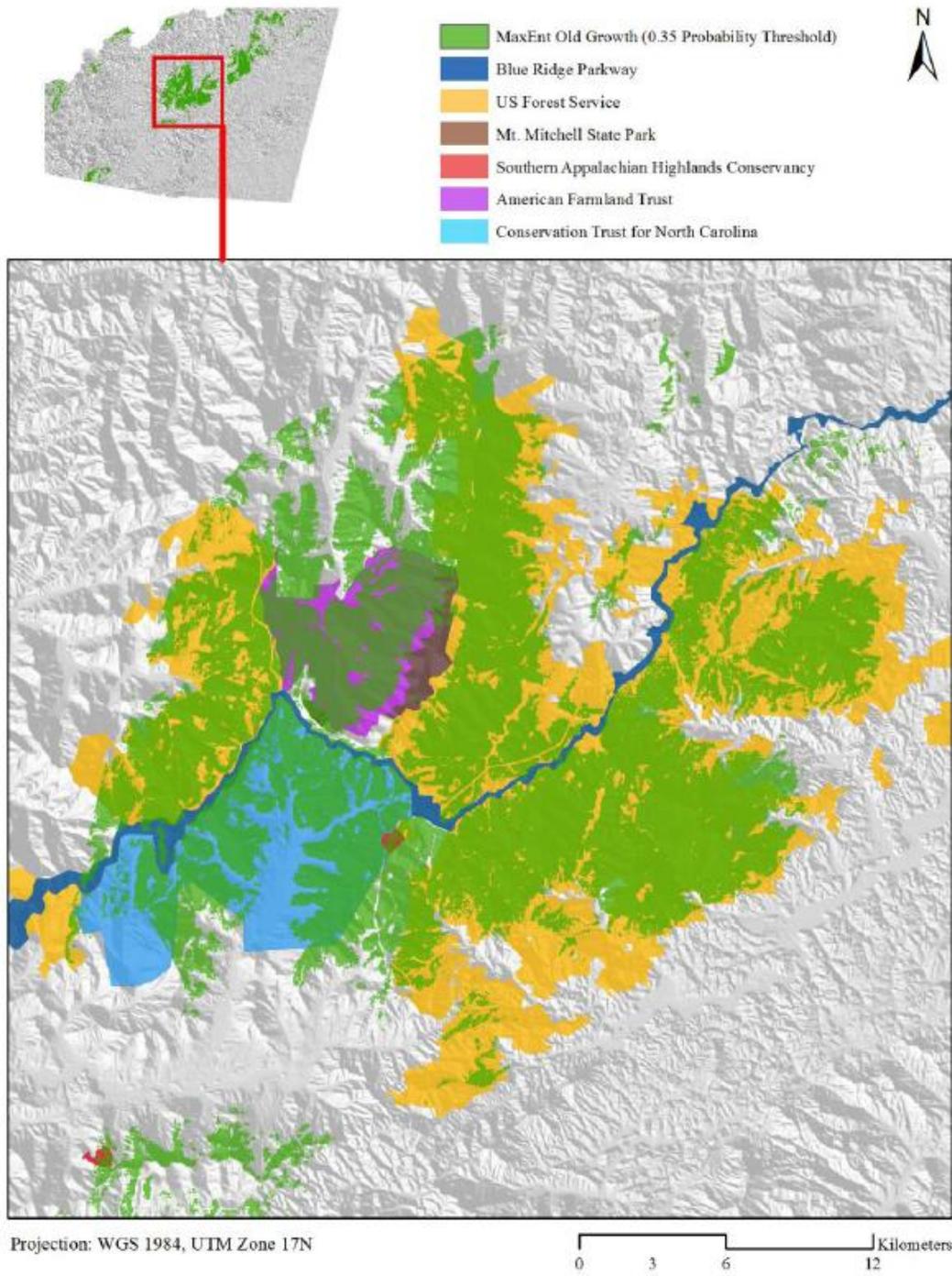
**Figure 13:** Percentage of Southern Appalachian Forest Coalition old growth sites with different geologic formations as the majority geology type within the site.

**Table 5:** Percent contribution of all environmental variables contributing 1% or more to the MaxEnt model.

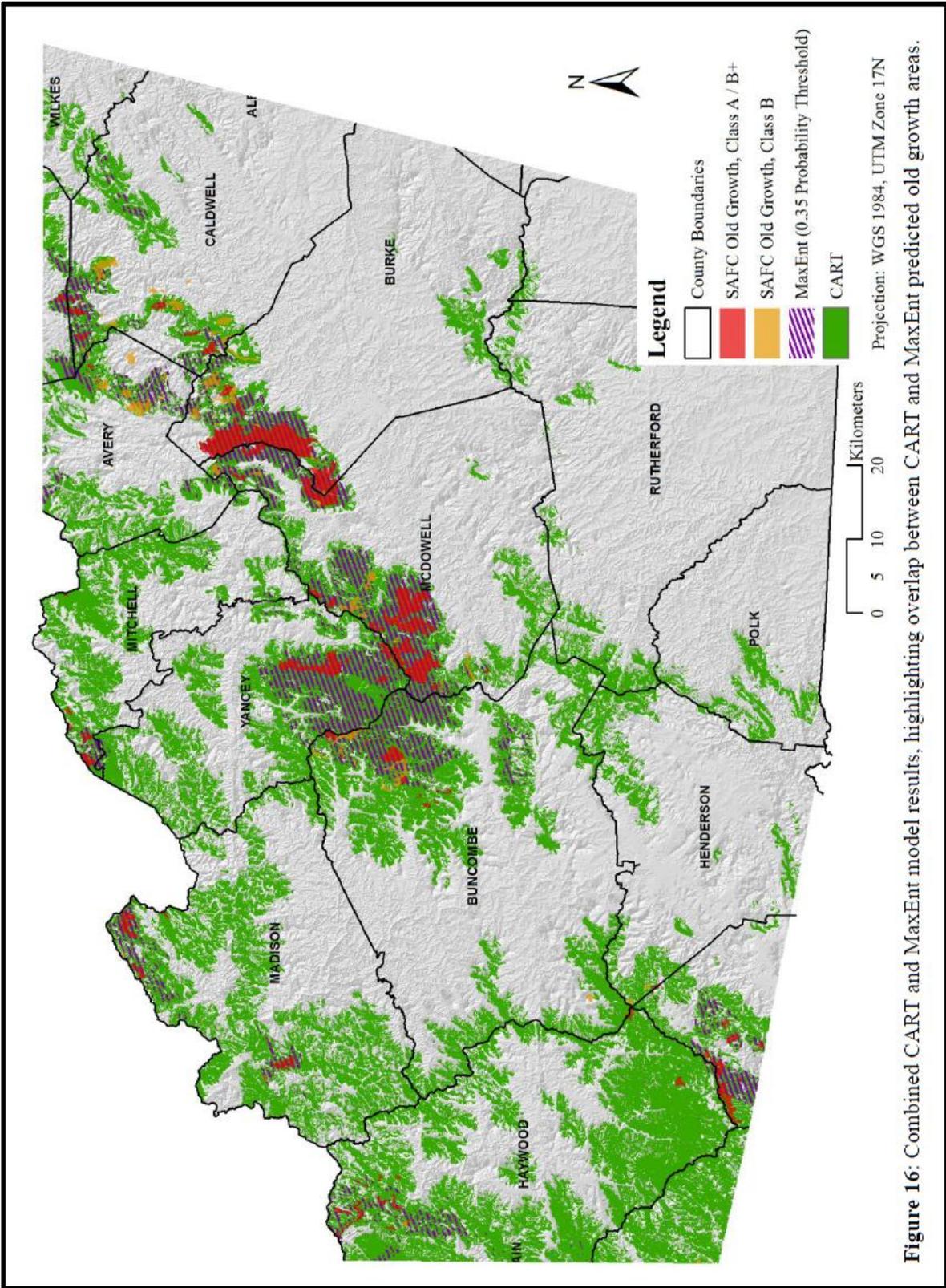
<b>Environmental Variable</b>	<b>% Contribution</b>
Elevation Standard Deviation (900 ha window)	40.7
Geology	26.6
Elevation Standard Deviation (9 ha window)	16.7
Mean Elevation (900 ha window)	4.9
Euclidean Distance to Roads	3
Tasseled Cap Brightness	3
NDVI	1
Euclidean Distance to Major Hydrology	1



**Figure 14:** MaxEnt jackknife test of variable importance. The increase in training gain observed when variables are used individually indicates the amount of information they contain in isolation. The decrease in training gain when a variable is removed indicates the amount of unique information contained by that variable, which is not provided by other variables in the model. Geology has the largest gain when used individually and the largest decrease when removed, indicating a high degree of unique information.



**Figure 15:** Example of different land ownership classes in MaxEnt predicted old growth areas.



**Figure 16:** Combined C.A.R.T. and MaxEnt model results, highlighting overlap between C.A.R.T. and MaxEnt predicted old growth areas.

**Table 6:** Summary of old growth predicted and validated areas.

Area of SAFC Old Growth in Study Area (hectares)		
Class A/B+	Class B	Total Area
16,656.53	5,684.30	22,340.83

Area of Old Growth Predicted (hectares)		Total Area of "Undiscovered" Old Growth Predicted (hectares)
CART	MaxEnt	
387,028.53	76,443.30	54,103.30

## LITERATURE CITED

- Bourg, N.A., McShea, W.J., Gill, D.E. 2005. Putting a CART before the search: Successful habitat prediction for a rare forest herb. *Ecology*. 86(10): 2793-2804.
- Brady, N.C. and Weil, R.R. 2010. "Factors Influencing Soil Formation" in *Elements of the Nature and Properties of Soils*, 3<sup>rd</sup> Ed.. Prentice Hall. Pg. 32-33.
- Buol, S.W. 2010. Formation of Soils in North Carolina. Accessed online at: <<http://www.soil.ncsu.edu/about/century/formation.html>>. [last accessed 2010 Dec 10]
- Crist, E.P. and Kauth, R.J. 1986. The Tasseled Cap De-Mystified. *Photogrammetric Engineering and Remote Sensing*. 52(1): 81-86.
- Davis, M.B., ed. 1996. Eastern Old-Growth Forests: Prospects for Rediscovery and Recovery. Island Press. Covelo, California. Pgs. 26-27.
- Dubayah, R.O. and Drake, J.B. Lidar Remote Sensing for Forestry Applications. Accessed online at: <http://www.geog.umd.edu/vcl/pubs/jof.pdf>. [last accessed 2009 Oct 19]
- Elith J., Graham, C.H., Anderson, R.P., Dudi'k, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R.E., Schapire, R.E., Sobero'n, J., Williams, S., Wisz, M.S., Zimmermann, N.E. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29:129–151.
- Franklin, S.E., Wulder, M.A., Gerylo, G.R. 2001. Texture analysis of IKONOS panchromatic data for Douglas-fir forest age class separability in British Columbia. *Int. J. Remote Sensing*. 22(13): 2627-2632.
- Guyon, L.J., Rolfe, G.L., Edgington, J.M., Mendoza, G.A. 2003. A comparative analysis of the diversity of woody vegetation in old-growth and secondary southern Appalachian cove forests. In: Van Sambeek, J. W.; Dawson, Jeffery O.; Ponder Jr., Felix; Loewenstein,

- Edward F.; Fralish, James S., eds. Proceedings of the 13th Central Hardwood Forest Conference; Gen. Tech. Rep. NC-234. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station: 75-87.
- Gower, S.T., McMertrie, R.E., Murty, D. 1996. Aboveground net primary production decline with stand age: potential causes. *TREE* 11(9): 378-382.
- Halpern, C.B. and Spies, T.A. 1995. Plants Species Diversity in Natural and Managed Forests of the Pacific Northwest. *Ecological Applications*. 5(4): 913-934.
- Harmon, M.E., Ferrell, W.K., Franklin, J.F. 1990. Effects on Carbon Storage of Conversion of Old-Growth Forests to Young Forests. *Science*, New Series. 247(4943): 699-702.
- Hernandez, P.A., Graham, C.H., Master, L.L., Albert, D.L. 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography*. 29:773-285.
- Hernandez, P.A., Frank, I., Herzog, S.K., Pacheco, V., Paniagua, L., Quintana, H.L., Soto, A., Swenson, J.J., Tovar, C., Valqui, T.H., Vargas, J., Young, B.E. 2008. Predicting species distributions in poorly-studied landscapes. *Biodiversity Conservation*. 17:1353-1366.
- Huete, A.R., HuiQing, L., van Leeuwen, W.J.D. 1997. The use of vegetation indices in forested regions: issues of linearity and saturation. IGARSS '97 – 1997 International Geoscience and Remote Sensing Symposium, Proceedings Vols I-IV: 1996-1968.
- Huang, C., Wylie, B., Yang, L., Homer, C., Zylstra, G. 2002. Derivation of a tasseled cap transformation based on Landsat 7 at-satellite reflectance. *International Journal of Remote Sensing* 23: 1741-1748.
- Irons, J.R. and Petersen, G.W. 1981. Texture transforms of remote sensing data. *Remote Sensing of Environment*. 11: 359-370.
- Jackson, C., Pitillo, D., Allen, L., Wentworth, T.R., Bullock, B.P., Loftis, D.L. 2009. Species diversity and composition in old growth and second growth rich coves of the southern Appalachian Mountains. *Castanea*. 74(1): 27–38.
- Jolliffe, I.T. 2002. Principle Component Analysis, 2<sup>nd</sup> Ed. Springer-Verlag, New York, Inc. Accessed online on: <[http://kolho3.tiera.ru/M\\_Mathematics/MV\\_Probability/MVsa\\_Statistics%20and%20applications/Jolliffe%20I.%20Principal%20Component%20Analysis%20\(2ed.,%20Springer,%202002\)\(518s\)\\_MVsa\\_.pdf](http://kolho3.tiera.ru/M_Mathematics/MV_Probability/MVsa_Statistics%20and%20applications/Jolliffe%20I.%20Principal%20Component%20Analysis%20(2ed.,%20Springer,%202002)(518s)_MVsa_.pdf)>. [last accessed 2010 Dec 10]
- Kim, S., McGaughey, R.J., Andersen, H. Schreuder, G. 2009. Tree species differentiation using intensity data derived from leaf-on and leaf-off airborne laser scanner data. *Remote Sensing of Environment*. 113: 1575-1586.
- Lefsky, M.A., Cohen, W.B., Parker, G.G. Harding, D.J. 2002. Lidar Remote Sensing for Ecosystem Studies. *BioScience*. 52(1): 19 – 30.

- Maier, B., Tiede, D., Dorren, L. Chapter 7.2: Characterising mountain forest structure using landscape metrics on LiDAR-based canopy surface models. pp. 625-644. Accessed online at: [http://ecorisq.org/docs/Maier\\_Tiede\\_Dorren\\_2008.pdf](http://ecorisq.org/docs/Maier_Tiede_Dorren_2008.pdf). [last accessed 2009 Oct 19]
- Messick, R. 2000. Old-Growth Forest Communities in the Nantahala-Pisgah National Forest: Final Report. Provided by Southern Appalachian Forest Coalition.
- Miettinen, J. 2007. Variability of fire-induced changes in MODIS surface reflectance by land-cover type in Borneo. *International Journal of Remote Sensing*. 28(22):4967- 4984.
- Mills, L.S., Fredrickson, R.J., Moorhead, B.B. 1993. Characteristics of Old-Growth Forests Associated With Northern Spotted Owls in Olympic National Park. *J. Wildl. Manage.* 57(2):315-321.
- Moore, I.D., Grayson, R.B., Ladson, A.R. 1991. Digital Terrain Modelling: A Review of Hydrological, Geomorphological, and Biological Applications. *Hydrological Processes*. 5:3-30.
- North Carolina Floodplain Mapping Program. 2003. LIDAR and Digital Elevation Data. Accessed online at: [http://www.ncfloodmaps.com/pubdocs/lidar\\_final\\_jan03.pdf](http://www.ncfloodmaps.com/pubdocs/lidar_final_jan03.pdf). [last accessed 2010 Dec 8]
- Phillips, S.J., Anderson, R.P., Schapire, R.E. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling*. 190:231-259.
- Pierce, K.B., Lookingbill, T., Urban, D. 2005. A simple method for estimating potential relative radiation (PRR) for landscape-scale vegetation analysis. *Landscape Ecology*. 20:137-147.
- Rapp, V. 2003. New findings about old-growth forests. Science Update 4. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. 12 p
- SEGAP (Southeast GAP Analysis Project). 2010. "Land Cover Mapping: GAP Land Cover Dataset." Accessed online at: <http://www.basic.ncsu.edu/segap/index.html>. [last accessed 2010 Dec 05]
- Shifley, S.R.; Roovers, L.M.; Brookshire, B.L. 1995. Structural and compositional differences between old-growth and mature second-growth forests in the Missouri Ozarks. In: Gottschalk, Kurt W.; Fosbroke, Sandra L. C., ed. Proceedings, 10th Central Hardwood Forest Conference; 1995 March 5-8; Morgantown, WV.: Gen. Tech. Rep. NE-197. Radnor, PA: U.S. Department of Agriculture, Forest Service, Northeastern Forest Experiment Station. 23-36

- Southern Appalachian Forest Coalition. 2009. "Characteristics of Old Growth." <Accessed online at: [http://www.safc.org/campaigns/old\\_growth/index.php](http://www.safc.org/campaigns/old_growth/index.php)> [last accessed 2010 Oct 19]
- Tikkanen, O. Punttila, P., Heikkila, R. 2009. Species-area relationships of red-listed species in old boreal forests: a large-scale data analysis. *Diversity and Distributions*. 15: 852-862.
- U.S. Geological Survey. U.S. Department of Interior. "Mineral resources on-line spatial data: Geologic units containing felsic metavolcanic rock." Page last modified: 18 Nov 2010. Accessed online at: <<http://tin.er.usgs.gov/geology/state/sgmclith.php?text=felsic+metavolcanic+rock>>. [last accessed 2010 Dec 07]
- U.S. Geological Survey. U.S. Department of Interior. "Seamless data warehouse: what is land cover?" Page last modified: 11 Mar 2010. Accessed online at: <[http://seamless.usgs.gov/about\\_landcover.php](http://seamless.usgs.gov/about_landcover.php)>. [last accessed 2010 Dec 08]
- Ustin, S.L. and Trabucco, A. 2000. Using hyperspectral data to assess forest structure. *Journal of Forestry*. 98(6): 47-49.
- Vayssières, M.P., Plant, R.E., Allen-Diaz, B.H. 2000. Classification trees: An alternative non-parametric approach to predicting species distributions. *Journal of Vegetation Science*. 11: 679-694.
- White, D.L. and Lloyd, F.T. 1994. Defining Old Growth: Implications for Management. Paper presented at the Eighth Biennial Southern Silvicultural Research Conference, Auburn, AL, Nov. 1-3, 1994.
- Zhao, K., Popescu, S., Nelson, R. 2009. Lidar remote sensing of forest biomass: A scale-invariant estimation approach using airborne lasers. *Remote Sensing of Environment*. 113: 182-196.
- Zhou, G., Liu, S., Li, Z., Zhang, D., Tang, X., Zhou, C., Yan, J., Mo, J. 2006. Old-growth Forests can Accumulate Carbon in Soils. *Science*. 314(5804): 1417.

## **APPENDIX A: Data Acquisition and Software**

### **Software used in this analysis:**

- Environmental Systems Research Institute (ESRI), ArcGIS 9.3; Redlands, CA
- Earth Resource Data Analysis System (ERDAS) Imagine 9.3; Atlanta, GA
- "R" version 2.7.2; R Foundation for Statistical Computing
- Maximum Entropy Modeling of Species Geographic Distributions; Version 3.3.3a; Phillips *et al.* (2006)

Environmental Predictor Variable	Source/Description of Derivation Method
Normalized Difference Vegetation Index (NDVI)	Generated in ERDAS Imagine 9.3 Spatial Modeler, from Landsat TM 7 (SLC-on) image from 2003, with the following equation: $\frac{[\text{Near Infrared} - \text{Red}]}{[\text{Near Infrared} + \text{Red}]}$
Enhanced Vegetation Index (EVI)	Generated in ERDAS Imagine 9.3 Spatial Modeler, from Landsat TM 7 (SLC-on) image from 2003, with the following equation (from Miettinen (2007)): $\text{EVI} = \frac{[(r_{\text{nir}} - r_{\text{red}}) / (r_{\text{nir}} + 6 * r_{\text{red}} - 7.5 * r_{\text{blue}} + 1)] * 2.5}$
Principle Components Analysis: Components 1-3	Generated in ERDAS Imagine 9.3 from Landsat TM 7 (SLC-on) image from 2003.
Tasseled Cap: Brightness, Greenness, Wetness	Generated in ERDAS Imagine 9.3 from Landsat TM 7 (SLC-on) image from 2003.
Landsat TM Band 8 Texture	Generated using Texture tool in ERDAS Imagine 9.3 from Landsat TM 7 (SLC-on) image from 2003. “Four algorithms are currently utilized for texture enhancement in ERDAS IMAGINE: mean Euclidean distance (1st-order), variance (2nd-order), skewness (3rd-order), kurtosis (4th-order)” (from C:\ERDAS\Geospatial Imaging 9.3\help\html\image_interpreter\texture.html)  This texture layer was created using the <i>variance</i> algorithm with the following equation (from Irons and Petersen (1981)):  Where: $\text{Variance} = \frac{\sum (x_{ij} - M)^2}{n - 1}$ $\text{Mean} = \frac{\sum x_{ij}}{n}$ $x_{ij}$ = DN value of pixel (i,j) n = number of pixels in a window M = Mean of the moving window, where:
Landsat TM Band 7 (Thermal)	Raw Band 7 of Landsat TM 7 (SLC-on) image from 2003.
Normalized Difference Water Index	Generated in ERDAS Imagine 9.3 Spatial Modeler, from Landsat TM 7 (SLC-on) image from 2003, with the following equation: $\frac{[\text{Near-Infrared} - \text{Short Wave Infrared}]}{[\text{Near-Infrared} + \text{Short Wave Infrared}]}$
Elevation	Digital Elevation Model resampled to 30 m resolution from LiDAR-derived state-wide DEM; Generated by Duke University, Landscape Ecology Lab.
Slope (degrees)	Calculated in degrees; Created using <i>Slope</i> tool in ArcGIS 9.3.
Analytic Hillshading	Created using <i>Hillshade</i> tool in ArcGIS 9.3 with the following parameters: azimuth = 225, declination angle = 30

Environmental Predictor Variable	Source/Description of Derivation Method
Slope Position	Slope Position calculated using TARDEM DEM tools developed by David Tarboton, Utah State University (Available online at: <a href="http://hydrology.usu.edu/tardem/tardem4.0/index.html">http://hydrology.usu.edu/tardem/tardem4.0/index.html</a> )
Terrain Shape Index	Terrain Shape Index calculated using TARDEM DEM tools developed by David Tarboton, Utah State University (Available online at: <a href="http://hydrology.usu.edu/tardem/tardem4.0/index.html">http://hydrology.usu.edu/tardem/tardem4.0/index.html</a> )
Topographic Convergence Index (TCI)	Generated based on the following equation from Moore <i>et al.</i> (1991): Created by Duke University, Landscape Ecology Lab. $w = \ln\left(\frac{A_s}{\tan \beta}\right)$
Radiation	Created according to analytic hillshading techniques outlined by Pierce <i>et al.</i> (2005); Generated by Duke University, Landscape Ecology Lab
Euclidean Distance to Roads	Euclidean distance from statewide primary and secondary road arcs from North Carolina Department of Transportation roads layer, which includes: “interstate, US, NC, secondary routes, and ramps and some non-state maintained and projected roads required for reporting purposes.” (Available online at: <a href="http://www.ncdot.org/it/gis/DataDistribution/DOTData/default.html">http://www.ncdot.org/it/gis/DataDistribution/DOTData/default.html</a> ).
Euclidean Distance to Major waterways	Euclidean distance from major hydrography calculated in ArcGIS 9.3; major hydrography layer from the North Carolina Corporate Geographic Database, representing the locations and identities of surface waters consisting of streams and rivers, lakes, ponds, and shorelines, as depicted on USGS 1:100,000-scale digital line graphs and supplemented to reflect hydrographic features classified by the state but not represented on the original DLG linework. (Available online at: <a href="http://www.nconemap.com">www.nconemap.com</a> )
Euclidean Distance to USGS 24k dataset	Euclidean distance from statewide 24k USGS DLG hydro dataset for North Carolina. (Available online at: <a href="http://www.nconemap.com">www.nconemap.com</a> )
Elevation Focal Mean at: 3x3 cell window (0.81 ha) 5x5 cell window (2.25 ha) 10x10 cell window (9 ha) 100x100 cell window (900 ha)	Focal mean elevation derived using <i>Focal Statistics</i> tool in ArcGIS 9.3 and 30 m DEM.

Environmental Predictor Variable	Source/Description of Derivation Method
Elevation Focal Standard Deviation at: 3x3 cell window (0.81 ha) 5x5 cell window (2.25 ha) 10x10 cell window (9 ha) 100x100 cell window (900 ha)	Focal elevation standard deviation derived using <i>Focal Statistics</i> tool in ArcGIS 9.3 and 30 m DEM.
Geology	Geology layer from <a href="http://www.nconemap.com">www.nconemap.com</a> ; GIS data set version of the geology of North Carolina created by North Carolina Department of Environment and Natural Resources, Division of Land Resources, NC Geological Survey, in cooperation with the North Carolina Center for Geographic Information and Analysis. The data represents the digital equivalent of the official State Geology map (1:500,000-scale), but was digitized from (1:250,000-scale) base maps. These data include three feature classes: geologic formations; diabase dikes; and geologic structures (faults, folds, scarps, contacts, and structural symbols).

Correlation matrix of all environmental variables  
(with correlation coefficients greater than 0.9 highlighted in yellow):

	<i>ndvi</i>	<i>evi</i>	<i>pca_1</i>	<i>pca_2</i>	<i>pca_3</i>	<i>tc_b</i>	<i>tc_g</i>	<i>tc_w</i>	<i>b8_texture</i>	<i>b7_thermal</i>
<i>ndvi</i>	1.000									
<i>evi</i>	0.7995	1.0000								
<i>pca_1</i>	-0.2110	0.3827	1.0000							
<i>pca_2</i>	0.9233	0.6548	-0.4302	1.0000						
<i>pca_3</i>	-0.0268	0.2125	0.4686	-0.1389	1.0000					
<i>tc_b</i>	-0.1579	0.4396	0.9960	-0.3671	0.4392	1.0000				
<i>tc_g</i>	0.8757	0.9661	0.1966	0.7958	0.2170	0.2563	1.0000			
<i>tc_w</i>	0.7144	0.2441	-0.7910	0.8896	-0.3415	-0.7431	0.4322	1.0000		
<i>b8_texture</i>	0.0913	0.0743	-0.0666	0.1265	-0.1339	-0.0509	0.0785	0.1252	1.0000	
<i>b7_thermal</i>	-0.7221	-0.3200	0.6781	-0.8957	0.0517	0.6312	-0.5219	-0.9505	-0.0991	1.0000
<i>ndwi</i>	0.9444	0.7005	-0.3346	0.9557	-0.2418	-0.2695	0.8063	0.8096	0.1273	-0.7744
<i>elev</i>	-0.2434	-0.2806	-0.0060	-0.2312	0.1092	-0.0424	-0.2212	-0.1734	-0.0818	0.1748
<i>slope</i>	-0.0371	-0.0786	-0.0453	-0.0280	-0.0054	-0.0518	-0.0547	0.0003	-0.0959	0.0070
<i>hillshd</i>	-0.1827	-0.2310	-0.1131	-0.1370	-0.1038	-0.1207	-0.2329	-0.0291	0.0118	0.0552
<i>relativeslope</i>	-0.1764	-0.1621	0.0397	-0.1882	0.0361	0.0261	-0.1759	-0.1463	-0.0509	0.1417
<i>terrainshp</i>	-0.2148	-0.1787	0.0443	-0.2042	0.0158	0.0309	-0.1925	-0.1585	-0.0431	0.1597
<i>radiation</i>	-0.1841	-0.2224	-0.0947	-0.1446	-0.0911	-0.1028	-0.2283	-0.0439	-0.0051	0.0674
<i>roads</i>	0.0648	0.0203	-0.0487	0.0670	0.0126	-0.0461	0.0423	0.0693	-0.0239	-0.0776
<i>hydromaj</i>	-0.0397	-0.0248	0.0940	-0.0903	0.1969	0.0805	-0.0188	-0.1146	-0.1153	0.0651
<i>hydro24k</i>	-0.2884	-0.2886	-0.0177	-0.2636	0.0436	-0.0436	-0.2869	-0.1735	-0.0451	0.1771
<i>mean3</i>	-0.2432	-0.2802	-0.0055	-0.2312	0.1097	-0.0419	-0.2208	-0.1736	-0.0821	0.1749
<i>mean5</i>	-0.2424	-0.2789	-0.0048	-0.2304	0.1105	-0.0412	-0.2195	-0.1735	-0.0818	0.1746
<i>mean10</i>	-0.2370	-0.2807	-0.0164	-0.2226	0.1052	-0.0526	-0.2186	-0.1625	-0.0799	0.1649
<i>mean100</i>	-0.1195	-0.1759	-0.0298	-0.1104	0.0915	-0.0574	-0.1071	-0.0802	-0.0631	0.0821
<i>stdev3</i>	-0.0361	-0.0613	-0.0483	-0.0057	-0.0432	-0.0484	-0.0411	0.0220	0.0137	-0.0116
<i>stdev5</i>	-0.0543	-0.0810	-0.0591	-0.0153	-0.0431	-0.0615	-0.0563	0.0194	0.0378	-0.0064
<i>stdev10</i>	-0.1257	-0.1411	-0.0565	-0.0763	-0.0288	-0.0659	-0.1162	-0.0258	0.0578	0.0411
<i>stdev100</i>	-0.0844	-0.0997	-0.0098	-0.0693	0.0357	-0.0234	-0.0674	-0.0511	-0.0359	0.0546
<i>tci</i>	0.0932	0.1533	0.0708	0.0851	-0.0084	0.0829	0.1323	0.0248	0.0757	-0.0278
<i>geology</i>	0.0021	-0.0292	-0.0373	0.0016	-0.0492	-0.0380	-0.0234	0.0190	0.0020	-0.0031

	<i>ndwi</i>	<i>elev</i>	<i>slope</i>	<i>hillshd</i>	<i>relativeslope</i>	<i>terrainshp</i>	<i>radiation</i>	<i>roads</i>	<i>hydromaj</i>	<i>hydro24k</i>
ndvi										
evi										
pca_1										
pca_2										
pca_3										
tc_b										
tc_g										
tc_w										
b8_texture										
b7_thermal										
ndwi	1.0000									
elev	-0.2090	1.0000								
slope	-0.0243	0.1750	1.0000							
hillshd	-0.1615	-0.0173	-0.0890	1.0000						
relativeslope	-0.1809	0.1627	0.0873	0.0353	1.0000					
terrainshp	-0.2113	0.2005	-0.0095	0.0847	0.6577	1.0000				
radiation	-0.1680	-0.0303	-0.0948	0.9856	0.0303	0.0796	1.0000			
roads	0.0634	0.0093	0.0075	0.0244	0.0327	0.0498	0.0388	1.0000		
hydromaj	-0.0805	0.2602	0.0571	-0.0007	0.1130	0.0972	0.0068	0.4058	1.0000	
hydro24k	-0.2901	0.2147	0.0103	-0.0623	0.2395	0.2677	-0.0589	-0.0089	0.0243	1.0000
mean3	-0.2090	0.9999	0.1756	-0.0172	0.1561	0.1942	-0.0302	0.0095	0.2603	0.2140
mean5	-0.2083	0.9997	0.1764	-0.0174	0.1476	0.1841	-0.0303	0.0092	0.2600	0.2124
mean10	-0.2014	0.9985	0.1782	-0.0158	0.1270	0.1525	-0.0291	0.0080	0.2579	0.2054
mean100	-0.0791	0.9373	0.1758	-0.0251	0.0017	0.0112	-0.0370	-0.0220	0.1691	0.0230
stdev3	-0.0101	0.1364	0.8898	-0.0995	0.0695	0.0095	-0.1043	-0.0090	0.0292	0.0059
stdev5	-0.0283	0.1707	0.8420	-0.0932	0.0745	0.0273	-0.0964	-0.0186	0.0241	0.0348
stdev10	-0.1038	0.2420	0.6756	-0.0750	0.0703	0.0641	-0.0786	-0.0201	0.0043	0.1481
stdev100	-0.0639	0.3924	0.1621	-0.0508	-0.0660	-0.0092	-0.0503	0.0146	-0.0561	0.0622
tci	0.0854	-0.1776	-0.3236	0.0027	-0.5959	-0.4875	0.0275	-0.0162	-0.0915	-0.1362
geology	0.0241	0.0967	-0.0186	0.0349	-0.1113	-0.0986	0.0279	-0.1280	-0.1967	0.0838

	<i>mean3</i>	<i>mean5</i>	<i>mean10</i>	<i>mean100</i>	<i>stdev3</i>	<i>stdev5</i>	<i>stdev10</i>	<i>stdev100</i>	<i>tci</i>	<i>geology</i>
ndvi										
evi										
pca_1										
pca_2										
pca_3										
tc_b										
tc_g										
tc_w										
b8_texture										
b7_thermal										
ndwi										
elev										
slope										
hillshd										
relativeslope										
terrainshp										
radiation										
roads										
hydromaj										
hydro24k										
mean3	1.0000									
mean5	0.9999	1.0000								
mean10	0.9989	0.9993	1.0000							
mean100	0.9381	0.9394	0.9436	1.0000						
stdev3	0.1372	0.1380	0.1388	0.1350	1.0000					
stdev5	0.1714	0.1720	0.1725	0.1645	0.9537	1.0000				
stdev10	0.2427	0.2432	0.2425	0.2166	0.7605	0.8763	1.0000			
stdev100	0.3935	0.3945	0.3946	0.4604	0.1371	0.1722	0.2743	1.0000		
tci	-0.1719	-0.1649	-0.1497	-0.0722	-0.3278	-0.2970	-0.2169	0.0332	1.0000	
geology	0.0974	0.0983	0.1019	0.1731	-0.0317	-0.0142	-0.0095	0.0441	0.0716	1.0000

## **APPENDIX B: Landsat Image Processing**

### **Data Acquisition and Software**

The basis of my analysis was a Landsat TM 7 (SLC-on) satellite image, Path 18 Row 35, from March 23, 2003, with 0.06% cloud cover. In addition, several GIS layers were utilized in this analysis. The software packages used to carry out this analysis included ArcGIS 9.3 and ERDAS IMAGINE 9.3

### **Preprocessing**

I began image processing by subsetting the original Landsat image to the southeast corner, which included a portion of western North Carolina. In this analysis, I used a GIS shapefile of old-growth polygons; therefore, it was necessary to georectify the Landsat image before continuing with any additional image processing. I performed an image-to-map rectification using the Landsat image subset and two corresponding GIS layers of roads and hydrography. I then used a 2<sup>nd</sup> order polynomial transformation, due to the amount of topography present in the region of interest. The initial 20 control points were checked with 5 checkpoints, which resulted in an RMS checkpoint error of 0.0045m. These 5 checkpoints were then converted to controls, for a final RMS Error of 0.7386 pixels. Finally, the image was resampled using bilinear interpolation.

The final preprocessing step was radiometric correction of my study image. I began by converting all input bands to at-satellite radiance using the information found in the image header file and the following equation:  $L_{\lambda} = (L_{MAX\lambda} - L_{MIN\lambda} / Q_{CALMAX} - Q_{CALMIN}) * (Q_{CAL} - Q_{CALMIN}) + L_{MIN\lambda}$ . Then I converted bands 1 through 5 and 7 to at-surface reflectance values, using the following equation:  $\rho_{surface} = \pi * d^2 * (L_{\lambda} - L_{haze}) / ESUN * \cos(\theta)$ .

## Spectral and Spatial Enhancement

After correcting the study image I created several spectral enhancement indices and component analyses, which included a Normalized Difference Vegetation Index (NDVI), a Principle Components Analysis (PCA), a Tasseled Cap, an Enhanced Vegetation Index (EVI), and a Normalized Difference Water Index (NDWI). The Tasseled Cap was carried out using the “Landsat 7 Tasseled Cap Transformation for Reflectance Data” model, taken from Huang *et al.* (2002). NDVI was carried out using the standard equation:  $[NIR - RED]/[NIR + RED]$ . EVI was originally designed for use with MODIS data, but it was adapted for this analysis. The following equation was used to calculate EVI, with coefficients taken from Miettinen (2007):  $EVI = [(r_{nir} - r_{red}) / (r_{nir} + 6 * r_{red} - 7.5 * r_{blue} + 1)] * 2.5$ . NDWI was created according to the following equation:  $[NIR - SWIR]/[NIR + SWIR]$ . Finally, I performed one spatial enhancement on the image by inputting Band 8 (high resolution panchromatic) in the Texture tool in IMAGINE, using a 3x3 cell focal window.

## APPENDIX C: CART Script

```
wd <- "F:/Working_MP_folder/CART_files"  
setwd(wd)
```

```
OG.data <- read.csv(file="old_growth.csv")  
names(OG.data)
```

```
dim(OG.data)
```

```
OG.data2 <- OG.data[,c(1,3:30)]
```

```
library(tree)  
OG.tree <- tree(as.factor(OG)~.,data=OG.data2)  
plot(OG.tree)  
text(OG.tree, cex=0.6)
```

```
OG.tree.prune <- prune.tree(OG.tree, method="misclass")  
OG.tree.prune <- prune.misclass(OG.tree)  
OG.tree.prune  
Plot(prune.tree(OG.tree, method="misclass"))
```

```
OG.tree.cv <- cv.tree(OG.tree, FUN=prune.misclass)  
Plot(OG.tree.cv)
```

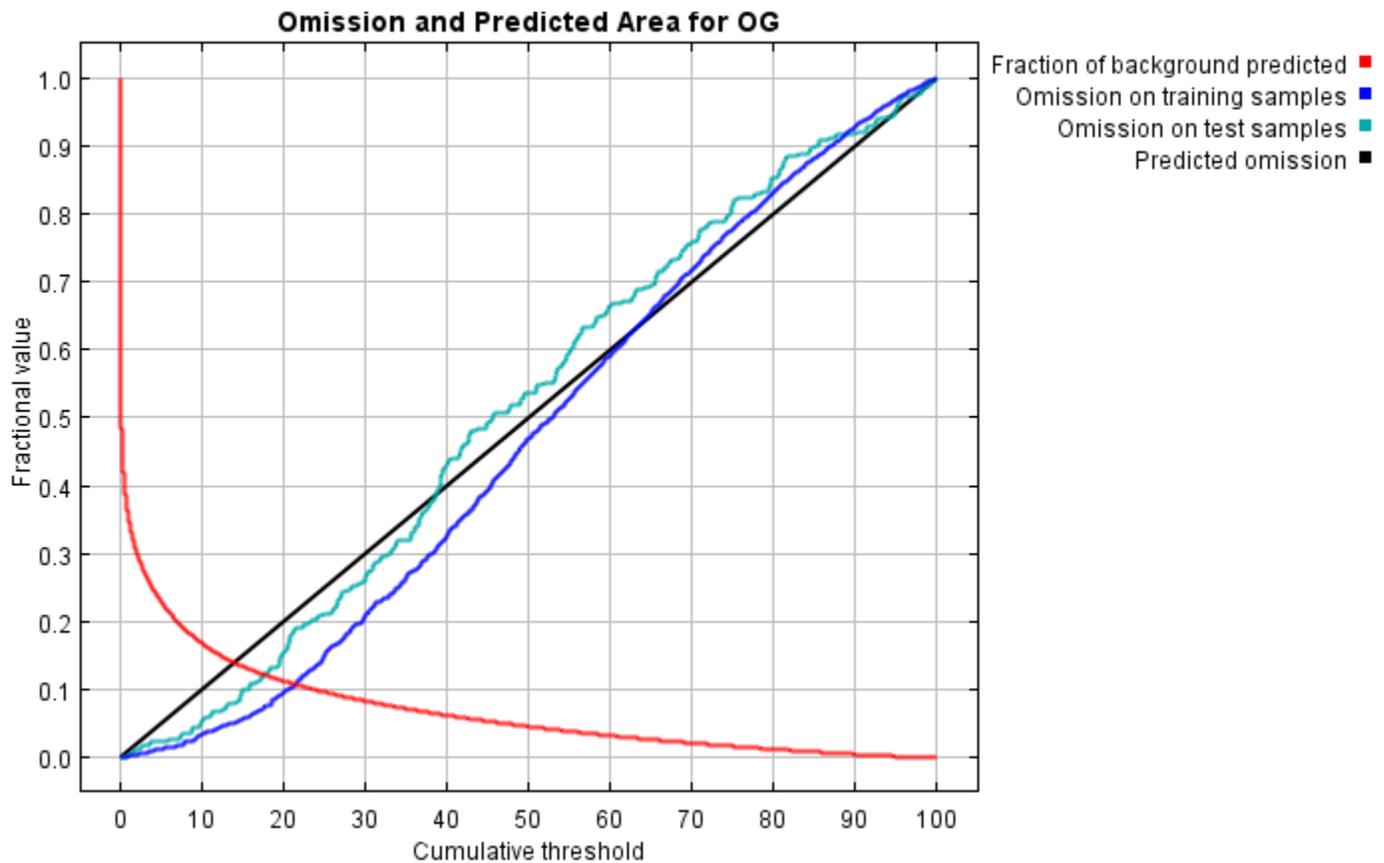
## APPENDIX D: MaxEnt Output

The following is a reproduction of portions of the detailed html output file generated by the MaxEnt program:

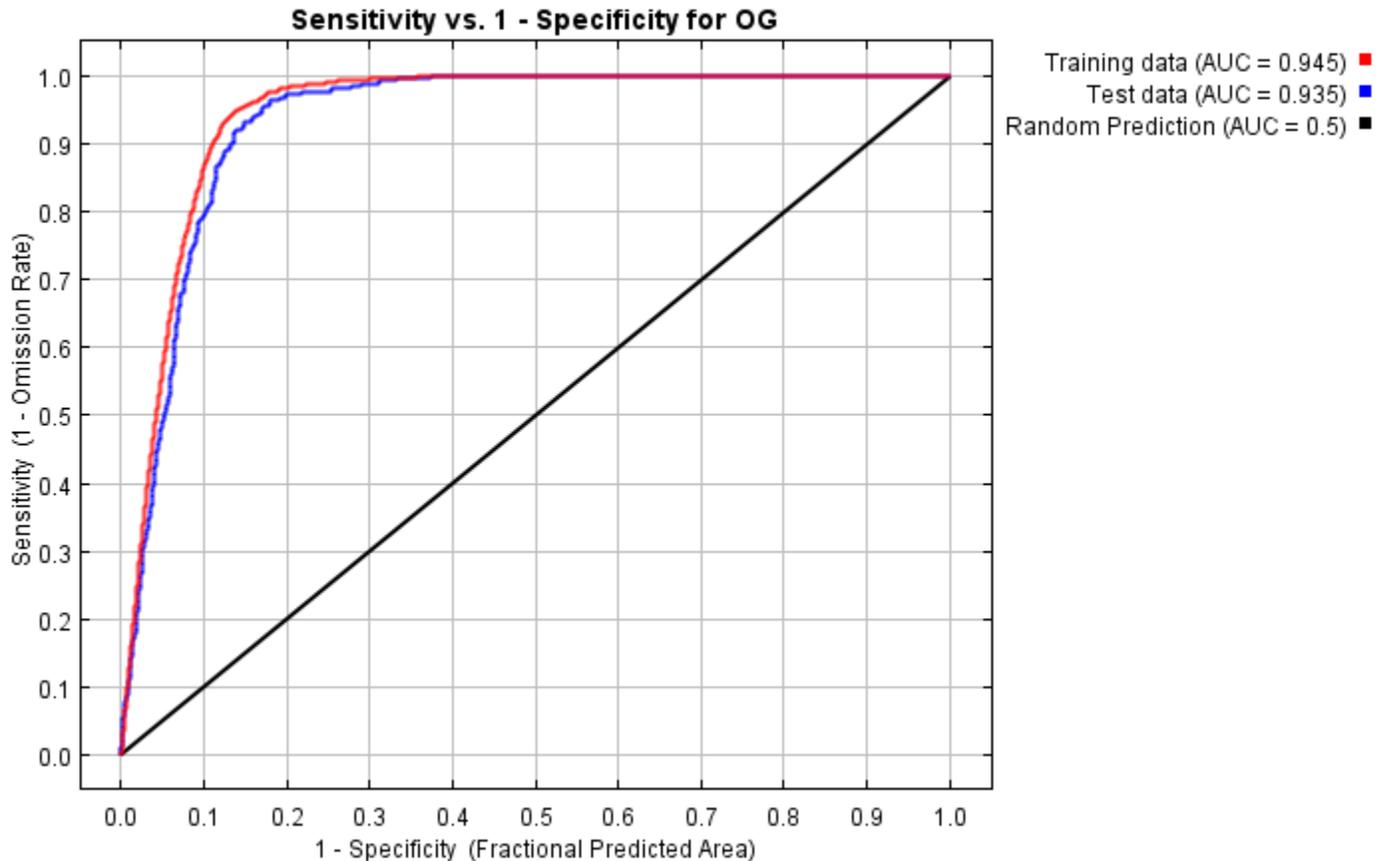
This page contains some analysis of the Maxent model for OG, created Thu Nov 04 12:52:56 EDT 2010 using Maxent version 3.3.3a. If you would like to do further analyses, the raw data used here is linked to at the end of this page.

### Analysis of omission/commission

The following picture shows the omission rate and predicted area as a function of the cumulative threshold. The omission rate is calculated both on the training presence records, and (if test data are used) on the test records. The omission rate should be close to the predicted omission, because of the definition of the cumulative threshold.



The next picture is the receiver operating characteristic (ROC) curve for the same data. Note that the specificity is defined using predicted area, rather than true commission (see the paper by Phillips, Anderson and Schapire cited on the help page for discussion of what this means). This implies that the maximum achievable AUC is less than 1. If test data is drawn from the Maxent distribution itself, then the maximum possible test AUC would be 0.929 rather than 1; in practice the test AUC may exceed this bound.



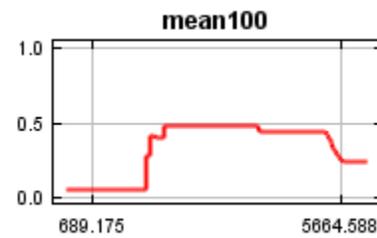
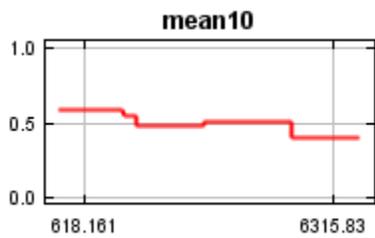
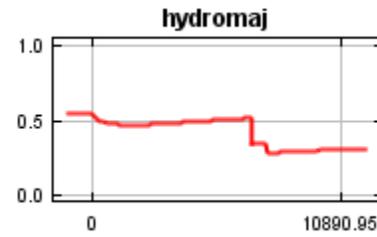
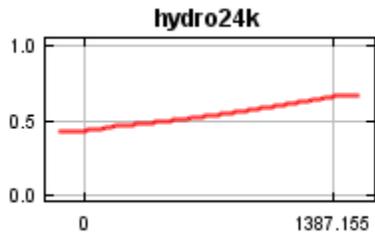
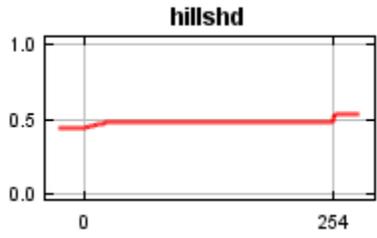
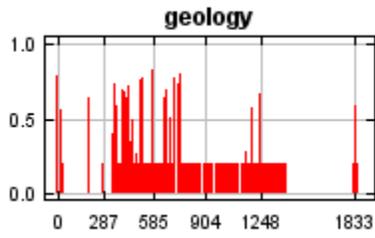
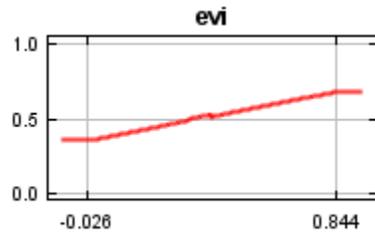
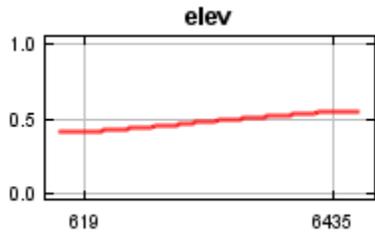
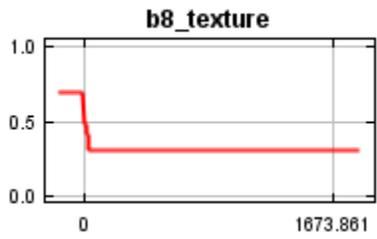
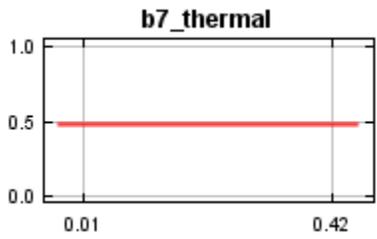
Some common thresholds and corresponding omission rates are as follows. If test data are available, binomial probabilities are calculated exactly if the number of test samples is at most 25, otherwise using a normal approximation to the binomial. These are 1-sided p-values for the null hypothesis that test points are predicted no better than by a random prediction with the same fractional predicted area. The "Balance" threshold minimizes  $6 * \text{training omission rate} + .04 * \text{cumulative threshold} + 1.6 * \text{fractional predicted area}$ .

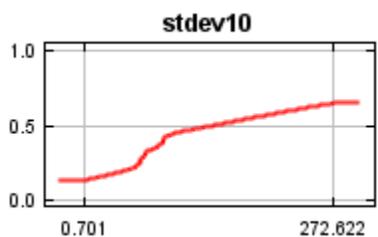
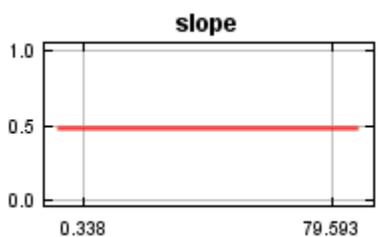
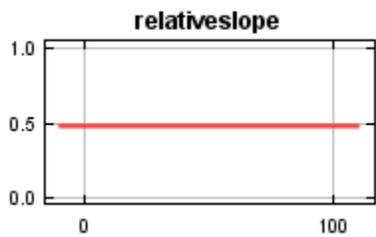
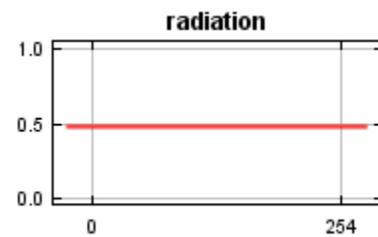
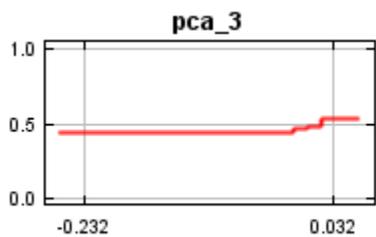
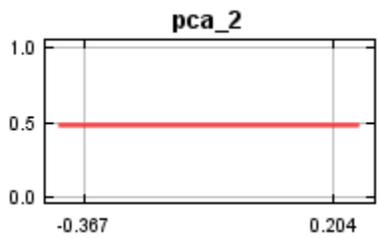
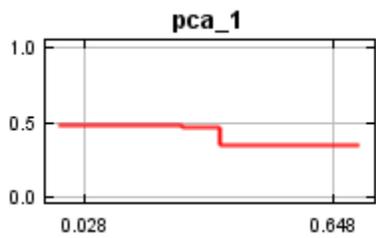
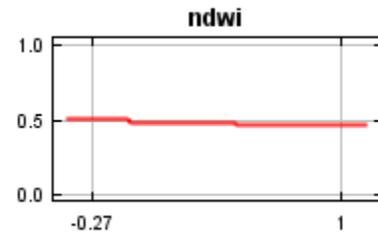
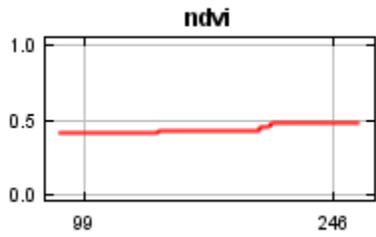
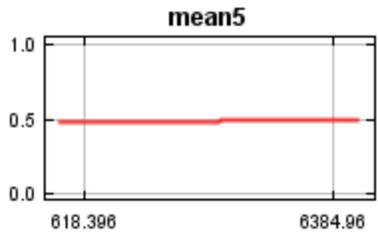
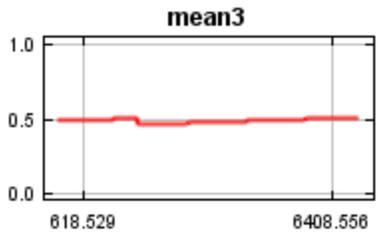
Cumulative threshold	Logistic threshold	Description	Fractional predicted area	Training omission rate	Test omission rate	P-value
1.000	0.027	Fixed cumulative value 1	0.353	0.003	0.004	0E0
5.000	0.101	Fixed cumulative value 5	0.230	0.012	0.024	0E0
10.000	0.177	Fixed cumulative value 10	0.168	0.036	0.056	0E0
0.775	0.022	Minimum training presence	0.370	0.000	0.004	0E0
20.673	0.350	10 percentile training presence	0.111	0.100	0.168	0E0
21.478	0.359	Equal training sensitivity and specificity	0.108	0.108	0.188	0E0
15.828	0.286	Maximum training sensitivity plus specificity	0.131	0.060	0.100	0E0
17.557	0.312	Equal test sensitivity and specificity	0.123	0.072	0.124	0E0
9.661	0.171	Maximum test sensitivity plus specificity	0.171	0.032	0.044	0E0
1.969	0.047	Balance training omission, predicted area and threshold value	0.304	0.004	0.012	0E0
7.812	0.142	Equate entropy of thresholded and original distributions	0.191	0.021	0.036	0E0

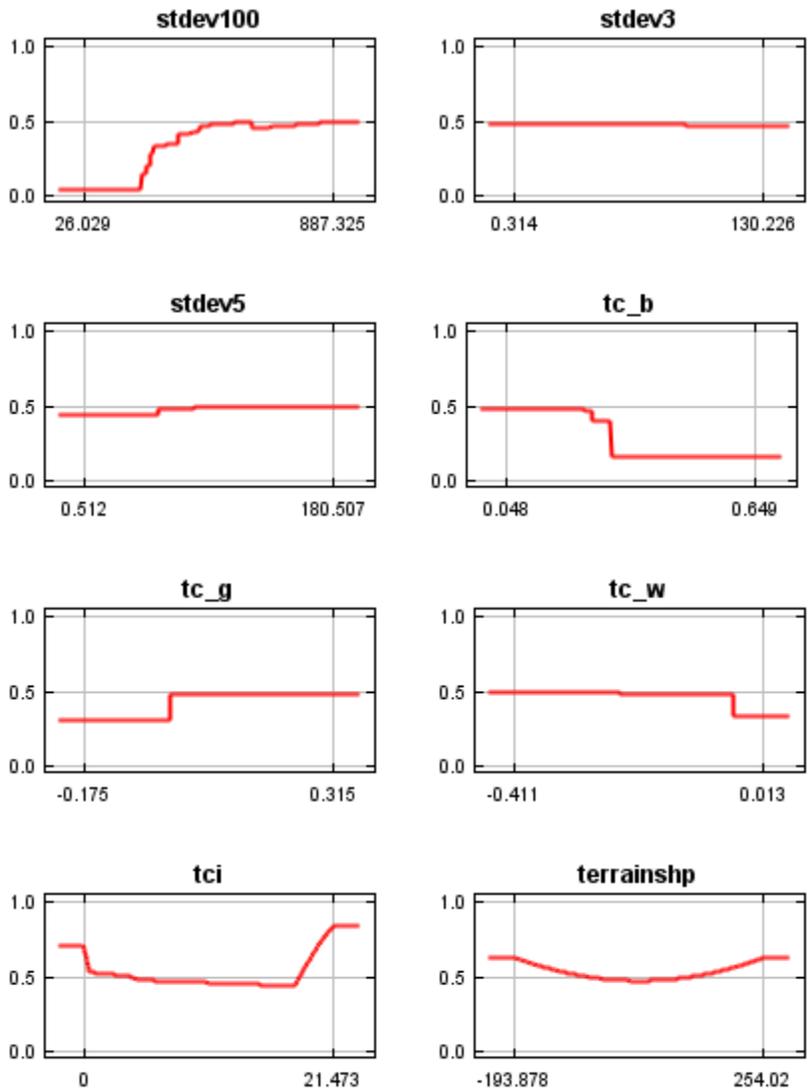
---

## Response curves

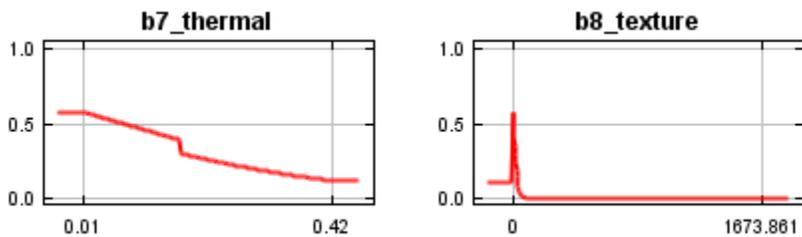
These curves show how each environmental variable affects the Maxent prediction. The curves show how the logistic prediction changes as each environmental variable is varied, keeping all other environmental variables at their average sample value. Click on a response curve to see a larger version. Note that the curves can be hard to interpret if you have strongly correlated variables, as the model may depend on the correlations in ways that are not evident in the curves. In other words, the curves show the marginal effect of changing exactly one variable, whereas the model may take advantage of sets of variables changing together.

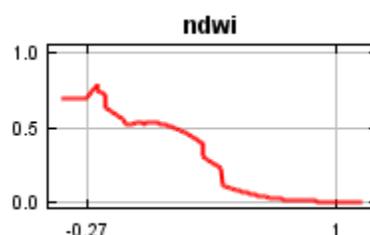
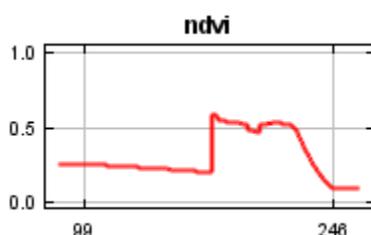
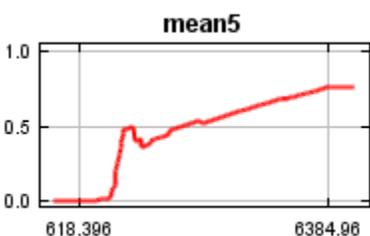
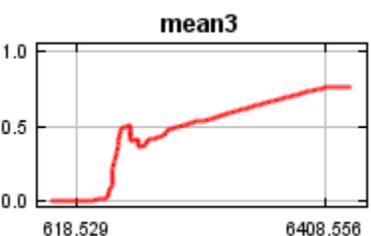
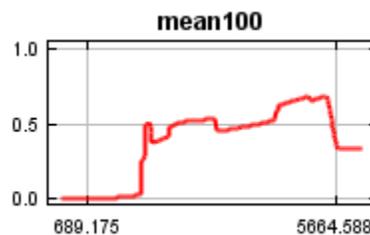
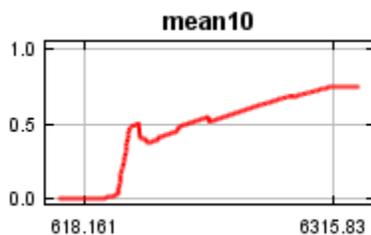
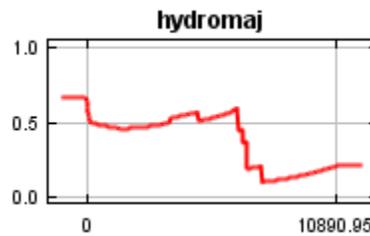
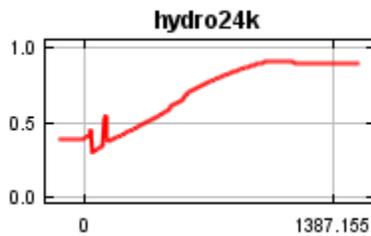
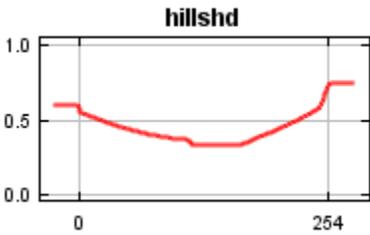
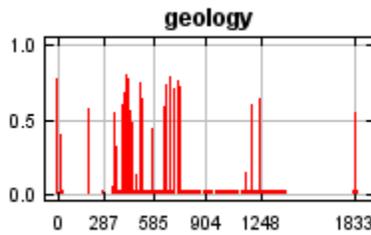
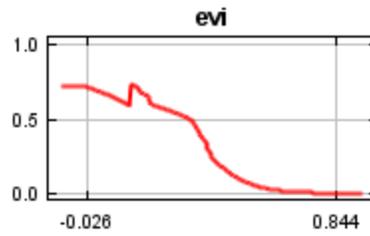


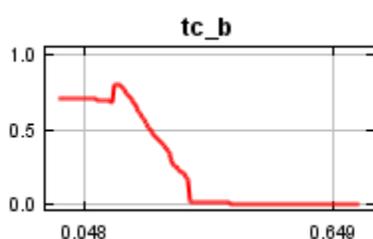
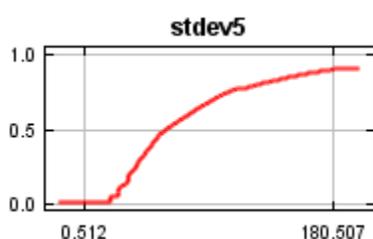
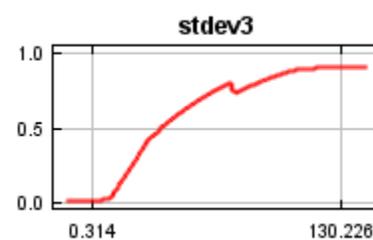
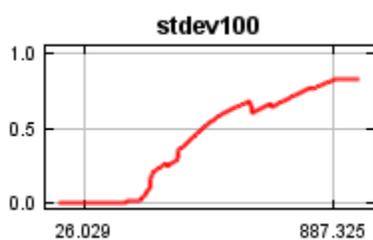
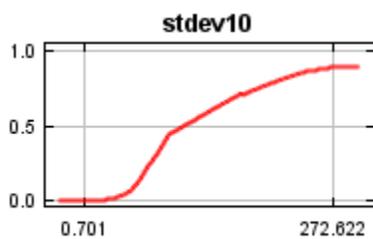
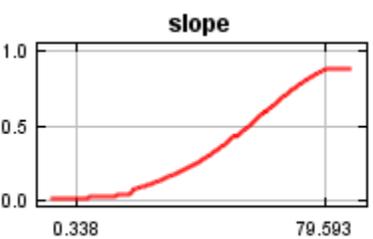
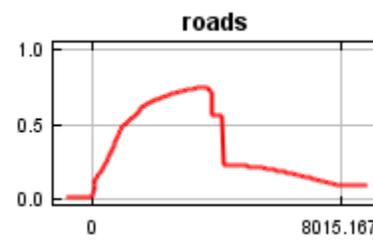
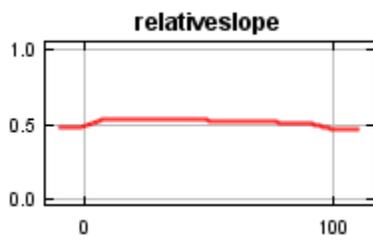
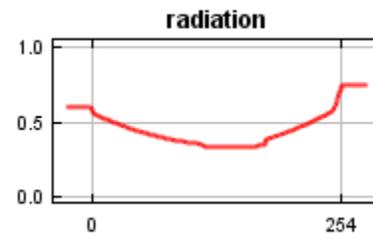
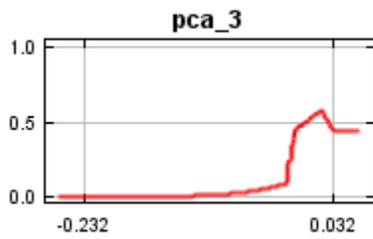
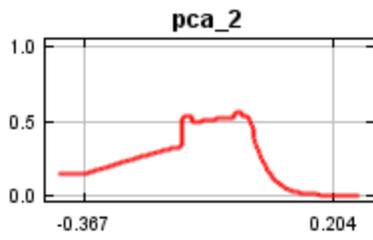
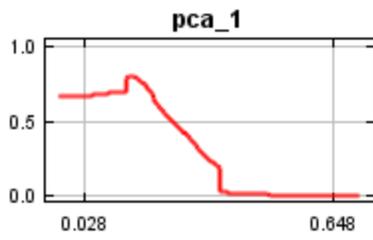


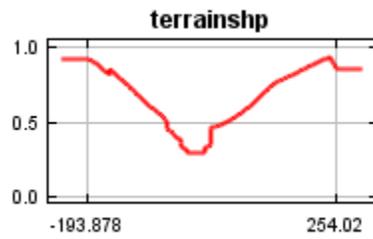
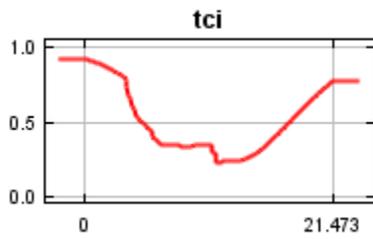
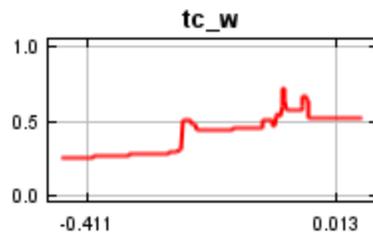
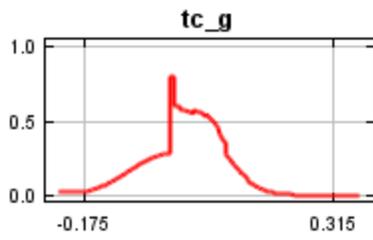


In contrast to the above marginal response curves, each of the following curves represents a different model, namely, a Maxent model created using only the corresponding variable. These plots reflect the dependence of predicted suitability both on the selected variable and on dependencies induced by correlations between the selected variable and other variables. They may be easier to interpret if there are strong correlations between variables.







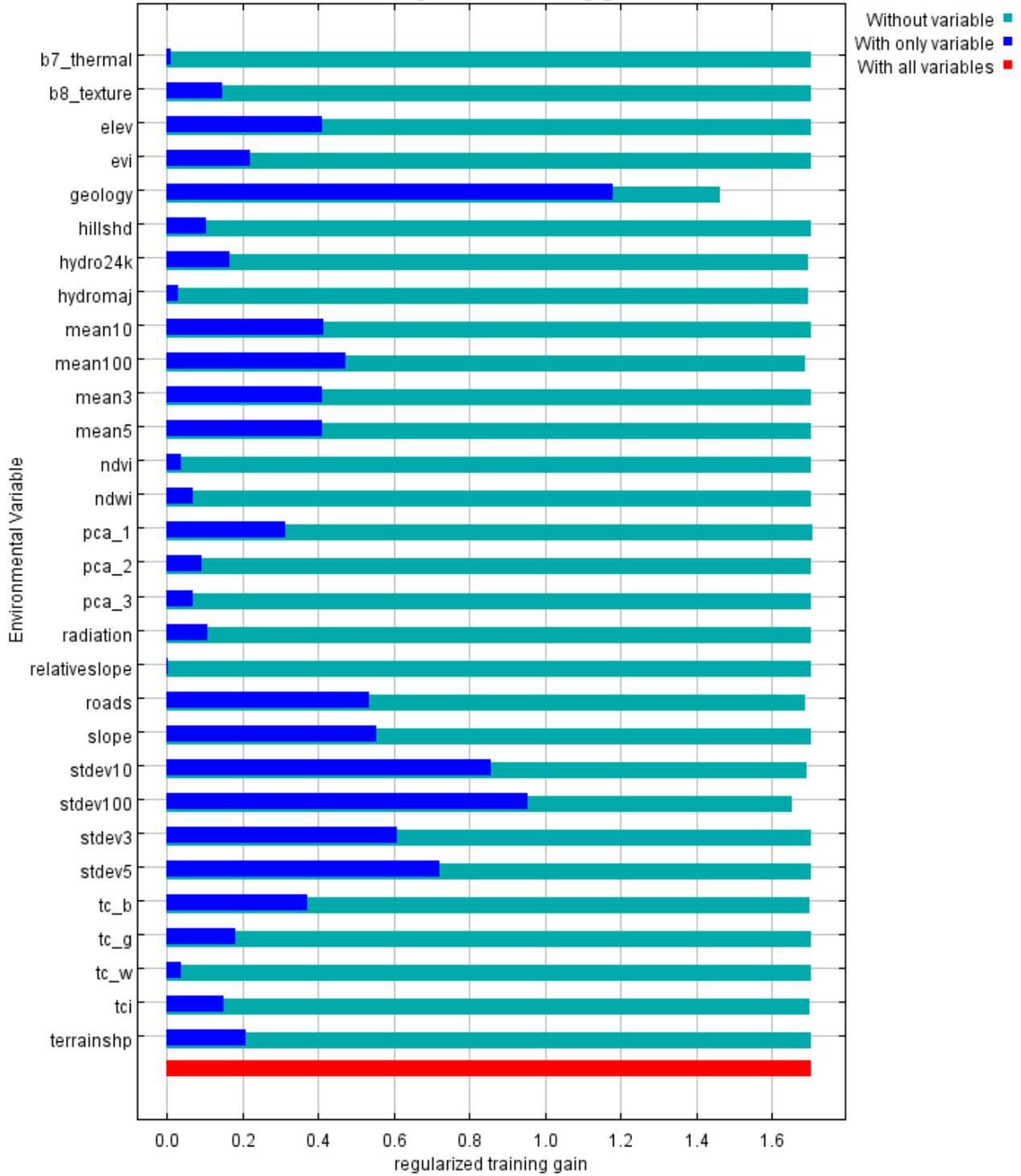


## Analysis of variable contributions

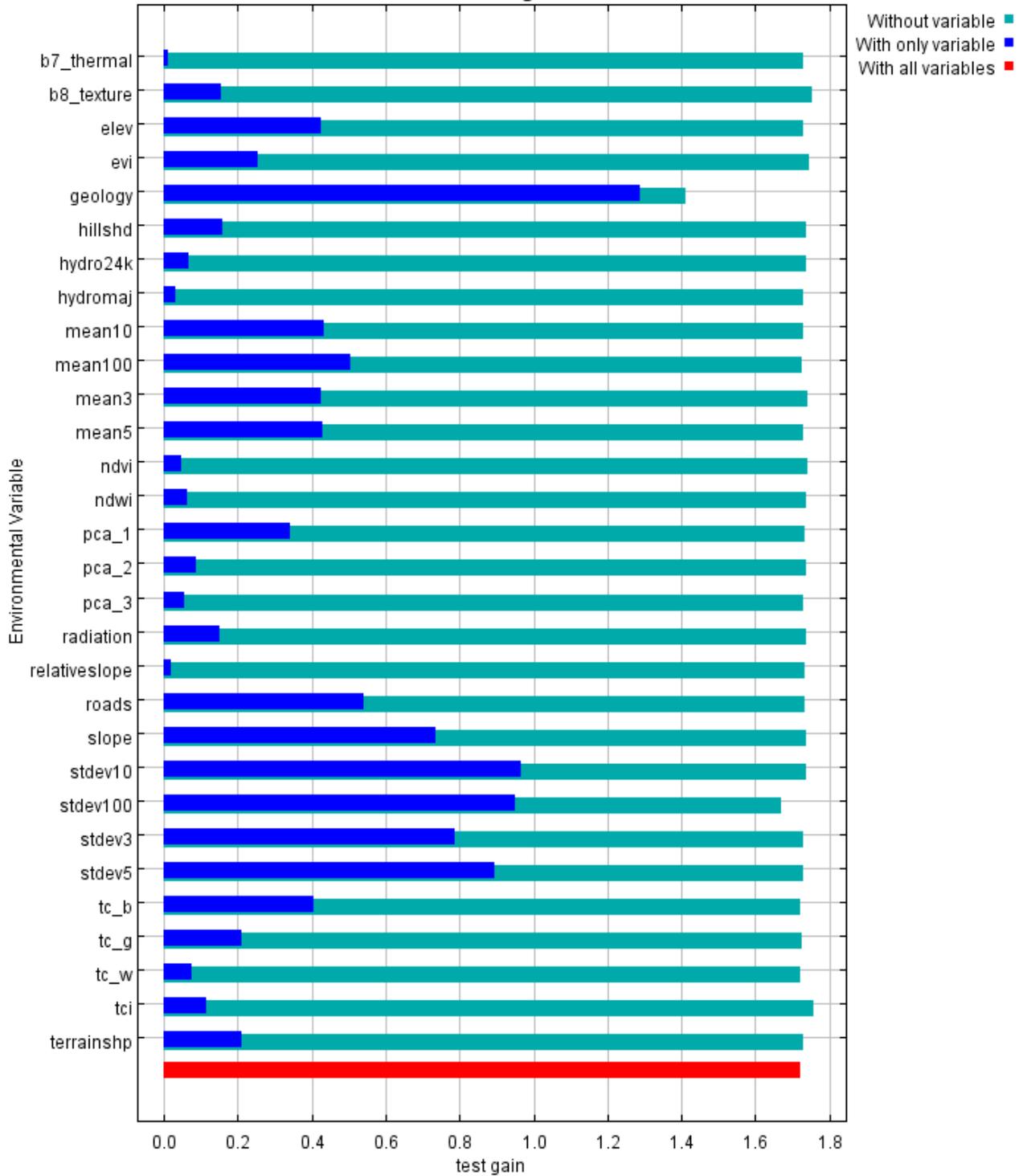
The following table gives estimates of relative contributions of the environmental variables to the Maxent model. To determine the first estimate, in each iteration of the training algorithm, the increase in regularized gain is added to the contribution of the corresponding variable, or subtracted from it if the change to the absolute value of lambda is negative. For the second estimate, for each environmental variable in turn, the values of that variable on training presence and background data are randomly permuted. The model is reevaluated on the permuted data, and the resulting drop in training AUC is shown in the table, normalized to percentages. As with the variable jackknife, variable contributions should be interpreted with caution when the predictor variables are correlated.

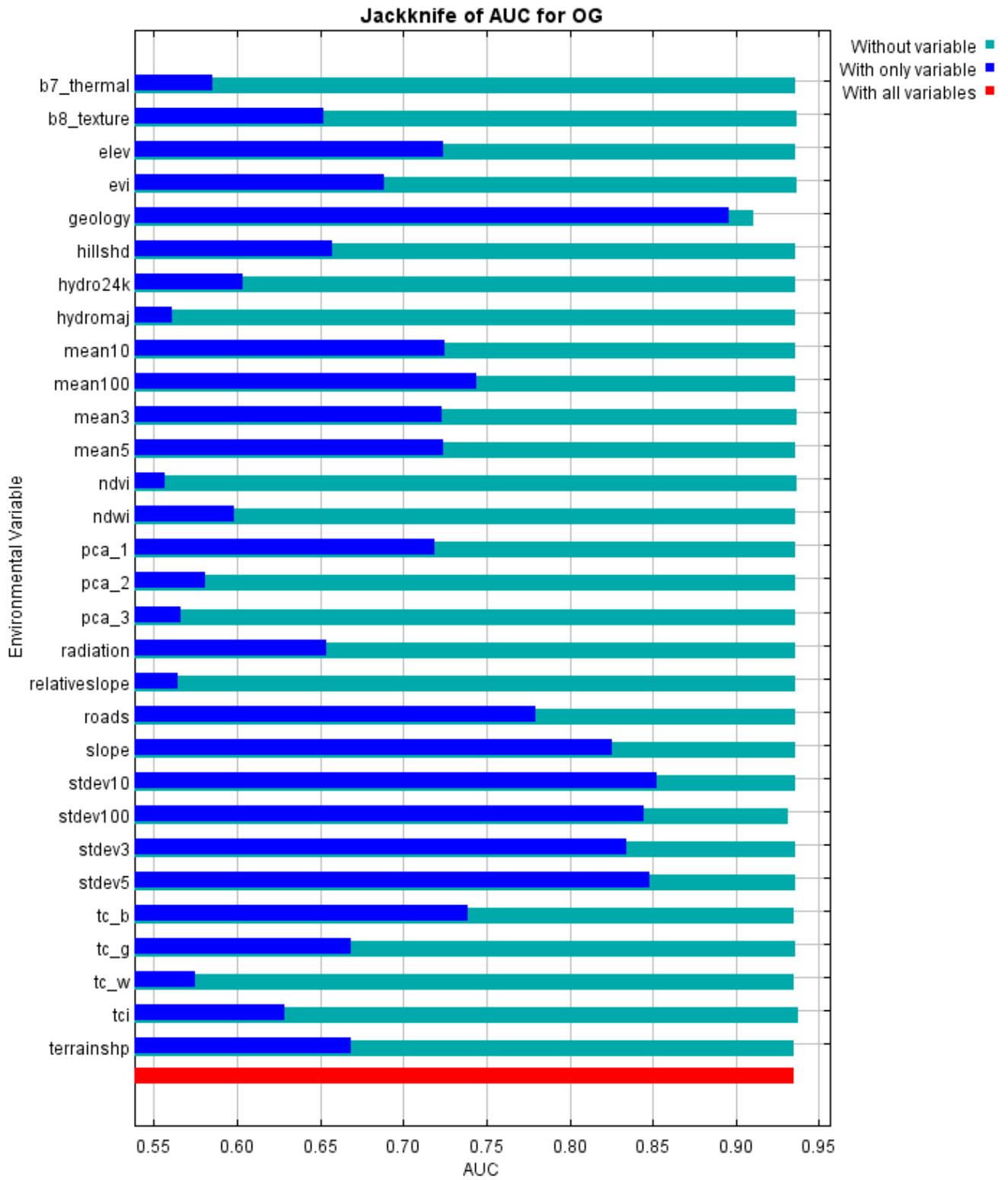
Variable	Percent contribution	Permutation importance
stdev100	40.7	28.5
geology	26.6	33.8
stdev10	16.7	7.7
mean100	4.9	18.3
roads	3	2.3
tc_b	3	3.9
ndvi	1	0.2
hydromaj	1	0.7
hydro24k	0.7	0.3
tci	0.5	0.3
mean10	0.3	0.8
elev	0.2	0.4
terrainshp	0.2	0.4
ndwi	0.2	0.4
b8_texture	0.2	0.5
mean3	0.2	0.2
tc_w	0.1	0.2
evi	0.1	0.4
hillshd	0.1	0.1
pca_3	0.1	0.1
stdev5	0.1	0.1
tc_g	0.1	0.1
pca_1	0	0.4
mean5	0	0
stdev3	0	0
slope	0	0
relativeslope	0	0
radiation	0	0
pca_2	0	0
b7_thermal	0	0

### Jackknife of regularized training gain for OG



Jackknife of test gain for OG





## Raw data outputs and control parameters

Regularized training gain is 1.704, training AUC is 0.945, unregularized training gain is 1.864. Unregularized test gain is 1.724.

Test AUC is 0.935, standard deviation is 0.004 (calculated as in DeLong, DeLong & Clarke-Pearson 1988, equation 2).

Algorithm terminated after 500 iterations (52 seconds).

The follow settings were used during the run:

750 presence records used for training, 250 for testing.

10750 points used to determine the Maxent distribution (background points and presence points).

Environmental layers used: b7\_thermal b8\_texture elev evi geology(categorical) hillshd hydro24k hydromaj mean10 mean100 mean3 mean5 ndvi ndwi pca\_1 pca\_2 pca\_3 radiation relativeslope roads slope stdev10 stdev100 stdev3 stdev5 tc\_b tc\_g tc\_w tci terrainshp

Regularization values: linear/quadratic/product: 0.050, categorical: 0.250, threshold: 1.000, hinge: 0.500

Feature types used: product linear quadratic hinge threshold

responsecurves: true

jackknife: true

outputdirectory: F:\Working\_MP\_folder\MaxEnt\MaxEnt\_Output\_2

samplesfile: F:\Working\_MP\_folder\MaxEnt\OG\_pts.csv

environmentallayers: F:\Working\_MP\_folder\MaxEnt\asc\_files

randomtestpoints: 25

Command line used:

Command line to repeat this species model: java density.MaxEnt nowarnings noprefixes -E "" -E

OG responsecurves jackknife

outputdirectory=F:\Working\_MP\_folder\MaxEnt\MaxEnt\_Output\_2

samplesfile=F:\Working\_MP\_folder\MaxEnt\OG\_pts.csv

environmentallayers=F:\Working\_MP\_folder\MaxEnt\asc\_files randomtestpoints=25 -t geology