2010

Predicting species composition in an eastern hardwood forest with the use of digitally derived terrain variables

Richard D. Flanigan
West Virginia University

PREDICTING SPECIES COMPOSITION IN AN EASTERN HARDWOOD FOREST WITH THE USE OF DIGITALLY DERIVED TERRAIN VARIABLES

By

Richard D Flanigan

A Thesis submitted to the
Davis College of Agriculture, Natural Resources and Design
at West Virginia University

in partial fulfillment of the requirements
for the degree of

Master of Science
In
Forestry

John Brooks Ph.D. co-Chair
Michael P. Strager, Ph.D. co-Chair
Tim Warner, Ph.D.
Jeff Simcoe, M.S.

Division of Forestry and Natural Resources

Morgantown, WV

2010

Keywords: Forest species composition, landform modeling, terrain analysis,
GIS /spatial analysis

Abstract

PREDICTING SPECIES COMPOSITION IN AN EASTERN HARDWOOD FOREST WITH THE USE OF DIGITALLY DERIVED TERRAIN VARIABLES

Richard D. Flanigan

This thesis addresses the need for improved classification of remotely sensed imagery in the complex hardwood forests of West Virginia. A geographic information system (GIS) was used in conjunction with forest plot data to develop a model to predict species composition in the eastern hardwood forest of West Virginia. The study area was located on the West Virginia University Research Forest (WVURF) in northern West Virginia. Terrain variables including aspect, curvature and slope change drastically at a local scale within the forest to greatly influence species composition. Light Detection and Ranging (LiDAR) data was collected for the entire WVURF, which produced an extremely detailed digital elevation model (DEM), with 1 m spatial resolution. Individual tree crown polygons were created from the LiDAR data so that individual trees could be co-registered to the DEM eliminating the bias of misplaced inventory points. Forest-plot data was collected and each individual tree crown polygon that was created from the LiDAR was assigned actual ground data. Terrain variable values were then sampled for each plot. The data was analyzed using a classification and regression tree (CART) to produce a binomial decision tree that was used within GIS to create a prediction grid of species distribution based on terrain variables. With the decreasing price of data acquisition and with new technology this method is likely to become more widespread and useful to various management agencies.

Acknowledgments

I would like to thank Dr. John Brooks, Dr. Mike Strager, Dr. Tim Warner and Jeff Simcoe for their patience and guidance throughout this entire process. I would like to thank Kara Fisher for putting up with me and providing me with the motivation to continue and finish this project. I would also like to thank the ImageTree Corporation for working in conjunction with this project and supplying all the data that made this work possible. Last but not least I would like to thank and praise my family and friends for their love, support and understanding while working on this project.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

C h a p t e r 1

INTRODUCTION

Tree species of the Appalachian region are ecologically notable for their comparatively high diversity and high economic value. Yellow-poplar (*Liriodendron tulipifera*), black cherry (*Prunus serotina*), northern red oak (*Quercus rubra*), white oak (*Quercus alba*), chestnut oak (*Quercus prinus*), American beech (*Fagus grandifolia*), black birch (*Betula nigra*), black oak (*Quercus velutina*), cucumber-tree (*Magnolia acuminata*), hickory spp (*Caryas spp.*), red maple (*Acer rubrum*), scarlet oak (*Quercus coccinea*), sugar maple (*Acer saccharum*) and American basswood (*Tillia americana*) are major components of forests throughout the Appalachian region. The complexity and interaction of these communities is problematic and costly for land managers in meeting their goals and objectives. It is important for land managers to have working knowledge about the spatial distribution of these species across the landscape in order to minimize the cost of acquiring field data and to make sound management decisions.

Traditional methods of inventory often include various types of intensive field work that can be costly and time consuming for large scale landowners on an annual basis. In the highly complex forest of the Appalachian region, the ability to achieve a suitable statistical inventory over large areas can be a very intensive projectundertaking, typically limiting owners to place their holding on an approximate ten year inventory cycle. Approximately 10% of their properties are evaluated in any given year, and of the 10% inventoried, assumptions are generated between inventory years to determine compositions within these areas.

In conjunction with technological advances being made in Geographic Information Systems (GIS) and remote sensing, it is imperative that processes be developed to best utilize these tools. Within the disciplines of forestry and resource management the improvement of species classification opens the door to a more complete and accurate inventory process through the use of more sophisticated remotely sensed data such as LiDAR and high resolution aerial imagery. As these technologies become more affordable and cost effective, the use of them in everyday forest management should become more cost effective giving land managers more robust and accurate data.

The goal of this study is to investigate the use of terrain variables in a GIS with Classification and Regression Tree (CART) analysis, to predict the occurrences of the species in the Appalachian region, and to provide a time efficient snap shot of the species composition of the entire forest.

CHAPTER 2: LITERATURE REVIEW

The distribution of plants and animals in space and time has been a focus of many biogeographical and ecological studies (Franklin 1995; Guisan and Zimmermann 2000; Guisan and Thuiller 2005). Species distribution modeling is founded in the quantification of species-environment relationships, where species and community distributions are explained by topographic and climatic variables (Franklin 1995; Guisan and Zimmermann 2000; Guisan and Thuiller 2005). Advances in geographic information science have produced alternatives for mapping vegetation beyond traditional methods, such as field surveying and photo interpretation. As a result, predictive modeling of species distribution has become a widespread tool in the areas of conservation biology, climate change research, land-use/land-cover change assessment, and biodiversity estimates (Guisan and Zimmermann 2000; Guisan and Thuiller 2005). Predictive vegetation modeling is defined as predicting the vegetation distribution across a landscape based on the spatial correlation of vegetation with environmental variables (Franklin 1995; Guisan and Zimmermann 2000). An increasing number of machine learning and statistical methods have been integrated with mapped environmental data to model distributions of species and other biodiversity attributes important for conservation planning across multiple scales (Franklin 1995; Guisan and Zimmermann 2000).

Predictive vegetation modeling requires digital maps of the environmental variables of interest, in addition to spatially attributed vegetation data (usually sample locations) (Franklin 1995; Guisan and Zimmermann 2000). Some statistical methods are based on the idea of

Gaussian species responses along environmental gradients as used in ordination-based regression models (Thuiller et al. 2003). Recent work illustrates that asymmetric and other complex (non-Gaussian) response curves are more frequently observed in vegetation associations with environmental variables (Thuiller et al. 2003). Statistical methods that can effectively model non-Gaussian and non-linear species responses to indirect (e.g. slope, aspect) and direct (e.g. moisture, temperature) environmental variables are highly beneficial in predictive vegetation modeling. The number of methods used is becoming increasingly flexible in order to describe complex response curves (Austin and Smith 1989).

There are different approaches to incorporating ancillary data into a classification process to improve classification accuracy, such as geographical stratification, post-classification sorting and classifier operations (Jensen, 2005). In this study, the classification and regression tree (CART) model was selected based on the following considerations; traditional expert classification systems are often hindered by the lack of expert knowledge or difficulties in defining classification rules; the CART model can be used to help develop a rule-based classification system, where expert knowledge is inadequate, using training data and machine learning processes; and CART tools are available in common statistical software.

CART is a nonparametric procedure that creates binary split rules in a stepwise method. A training data set is needed to serve as the input data for the CART model. Classification results are highly dependent on the selection of the training set. The classification accuracy can potentially be increased significantly if sample points are selected using expert knowledge of the most representative areas of the classes (Domac and Süzen 2006). In a study performed by Fekedulegn et al. (2004) it was determined that aspect has a profound effect on species composition in an Appalachian forest. Yellow-poplar and black cherry showed a strong preference for north and east oriented aspects while chestnut oak and white oak showed a strong preference for south and west oriented aspects. Northern red oak had mild aspect preference indicating its ability to grow and compete in a variety of environments (Fekedulegn et al., 2004). Terrain shape has also been found to be strongly related to vegetation distributions in a forested setting (Elliott et al., 1999). Bolstad et al. (1998) found a significant relationship between basal area and terrain shape for both yellow-poplar and chestnut oak. In addition, they observed that species composition changes gradually with terrain shape resulting in notable overlap between forest types (Bolstad et al., 1998, McNabb, 1993; and McNabb 1989). This indicates that species such as yellow-poplar and chestnut oak will not be confined to coves or ridges but will also be found in some abundance outside these areas. Soil organic matter content, elevation, and terrain shape were shown to explain 33% of the variation in species distributions in a study performed on the Coweeta Hydrologic Laboratory in North Carolina (Elliot et al., 1999).

A separate study conducted by Iverson et al. (1997) showed a direct relationship between soil moisture and species distribution in a southeast Ohio study. Elliot et al. (1999) describes soil moisture as a direct function of precipitation, terrain shape, and soil characteristics. Slopes with an aspect of north-northeast have the highest moisture levels while slopes with an aspect of south-southwest have the lowest moisture levels due to high amounts of direct solar radiation (Iverson et al., 1997). Chestnut oak and the oak genus are generally more tolerant of moisture stress, allowing them to persist on drier sites (Fekedulegn et al., 2003; and Iverson et al., 1997). Yellow-poplar was found to be exploitive and concentrated in more moist areas with concave shapes (Elliot et al., 1999; Fekedulegn et al., 2003; Iverson et al., 1997). Tree height of yellow-poplar has also been shown to be directly related to terrain shape (McNabb, 1989).

CARTs have been shown to be adequate for predicting the presence and absence of four species of oak in California (Vayssieres et al., 2000). One advantage in using CART analysis is that it is a non-parametric data-driven approach that eliminates the potential for user-introduced bias, and thus reduces the risk of using simplifying assumptions (Vayssieres et al., 2000). The output includes a tree diagram, with the branches determined by splitting rules, and a series of terminal nodes that contain the prediction (Prasad et al., 2006). CART is designed to work with data that might have multiple variables for a single outcome, as is the case in a forested environment, rather than trying to force the outcome to be derived from a single overriding variable, commonly employed by many parametric techniques (Prasad et al., 2006; Vayssieres et al., 2000). Prasad et al. (2006) found that CART analyses are favorable among biological applications, such as species prediction. Some studies have found that classification trees can sometimes be too complex to effectively interpret species distributions (Guisan and Zimmermann 2000; Muñoz and Felicísimo 2004).

CHAPTER 3: METHODOLOGY

3.1 Study Area

West Virginia University (WVU) Research Forest is a 3,123.06 hectare forest located in Monongalia and Preston counties of West Virginia (Figure 1).

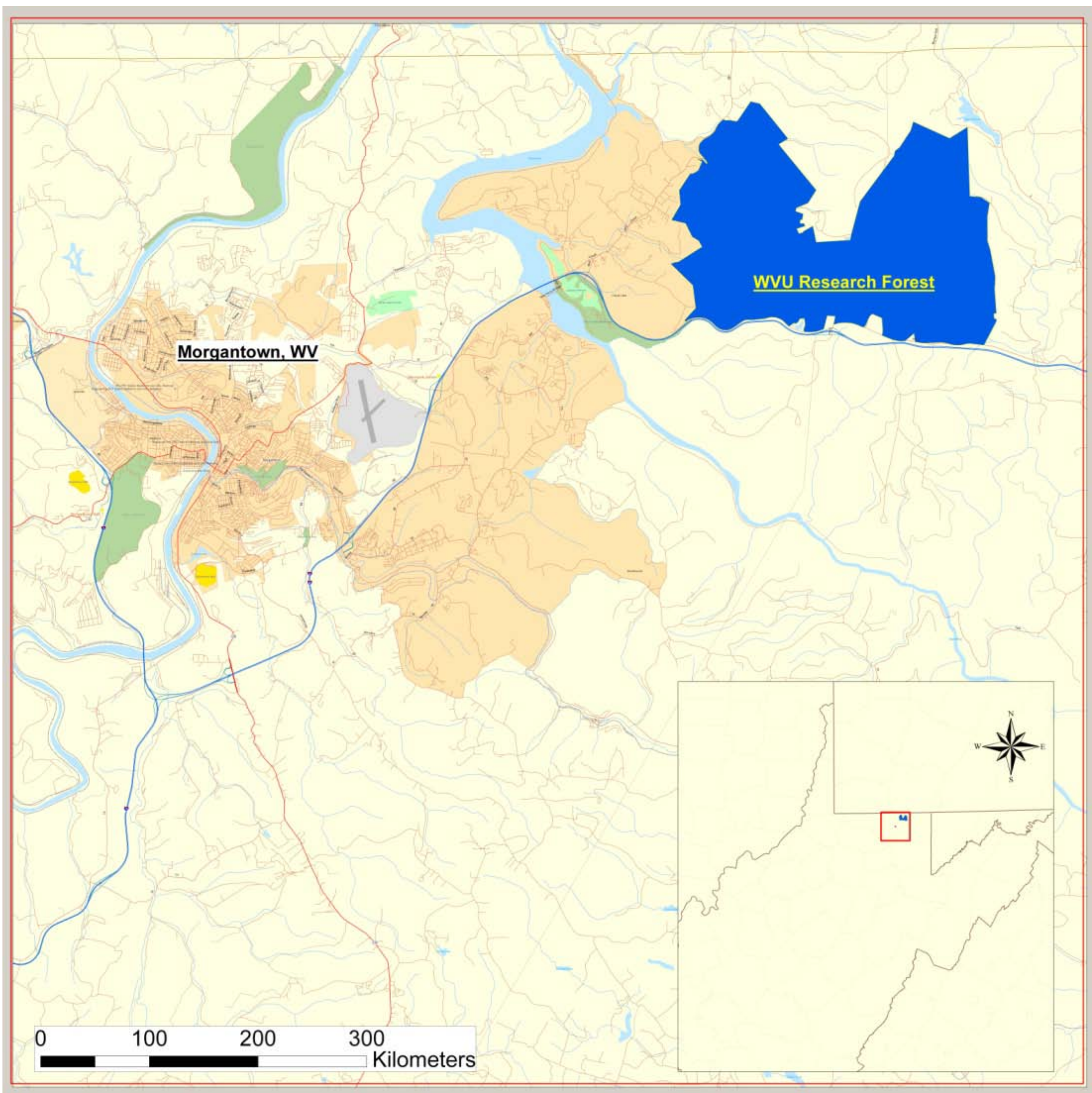

Figure 1. Location of WVU Research Forest in WV

By the year 1936, the entire region encompassing the WVU Research Forest had been completely clearcut with repeated heavy burns that still affect the forest we see today (Carvell, 1973). The loss of American chestnut (*Castanea dentata*), due to chestnut blight, was critical since it comprised the most commercially valuable component of these stands (Carvell, 1973). Today's forest consists of 70-year-old stands that have had various silvicultural treatments applied to them. The transitions between the different forest types on the WVU Research Forest are more gradual than is typical in other areas of the state, and this gradual phenomenon is primarily due to relatively high mean annual precipitation of 129 cm that is evenly distributed throughout the year (Fekedulegn et al. 2004). This distinctive characteristic of the study area and the surrounding region results in large overlaps of species that typically would not be found in such abundance on common sites.

3.2 Data

3.2.1 Creating Predictors from Terrain

A one meter Digital Elevation Model (DEM) was derived from LiDAR data collected on the study area in 2006. The DEM is a high resolution continuous elevation data set for the entire study site. This elevation grid layer was then resampled from its original one meter cell size to a ten meter cell size by cubic convolution; to closely resemble the size of the fixed area inventory points. Cubic convolution produces an output cell value that is computed by fitting a smooth surface to the nearest 16 (4x4) input cells (Wu et al, 2008). The cubic convolution methodology preserves much of the high level detail of the original one meter data set. All subsequent grids and analysis were based on the resampled 10m DEM.

Five terrain variable grids were created using tools from the Spatial Analyst Toolbox found in ArcGIS 9.2 (ESRI 2007). These resulting grids included continuous aspect (Figure 2), slope (Figure 3), overall curvature (Figure 4), plan curvature (Figure 5), and profile curvature (Figure 6). A brief description of these variables and their ranges for the study area can be found in **Table 1.**

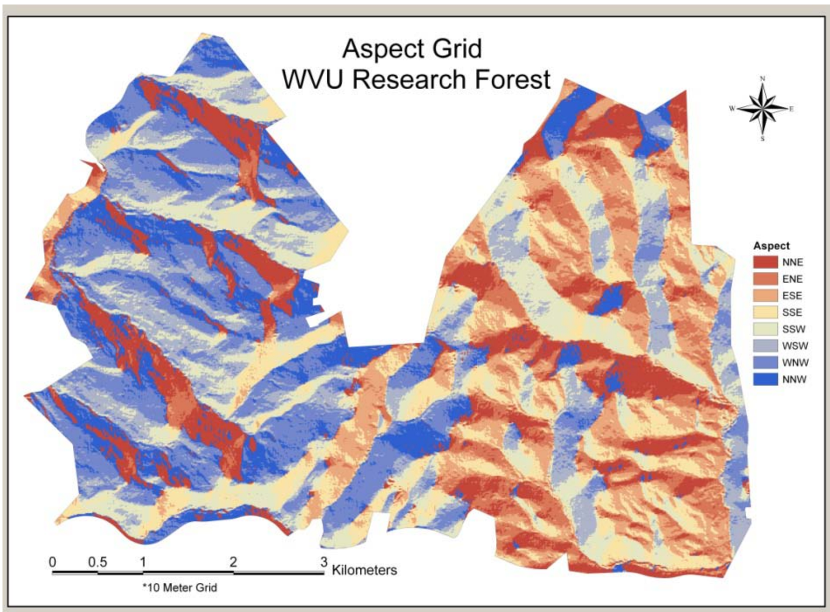

Figure 2. Aspect Grid

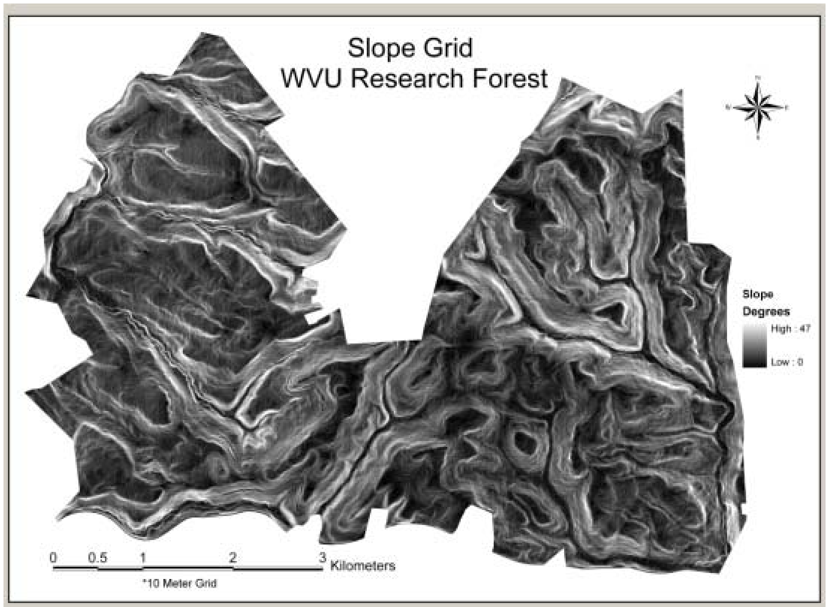

Figure 3. Slope Grid

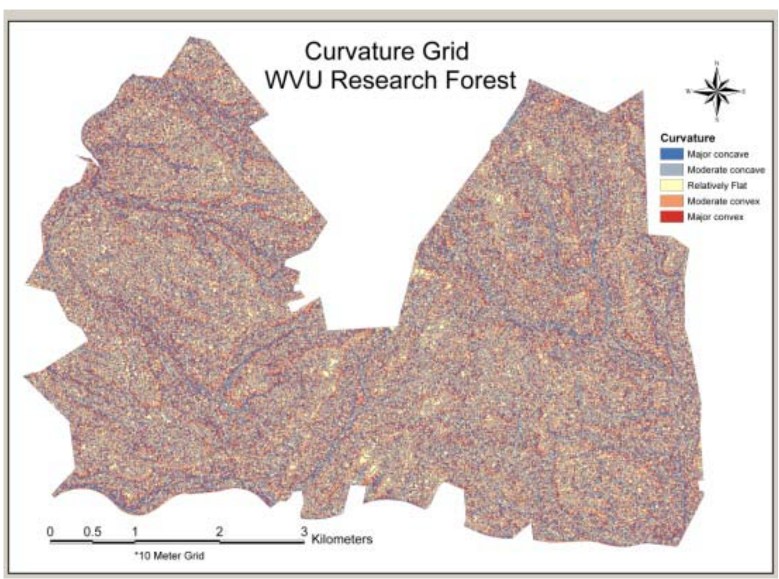

Figure 4. Curvature Grid

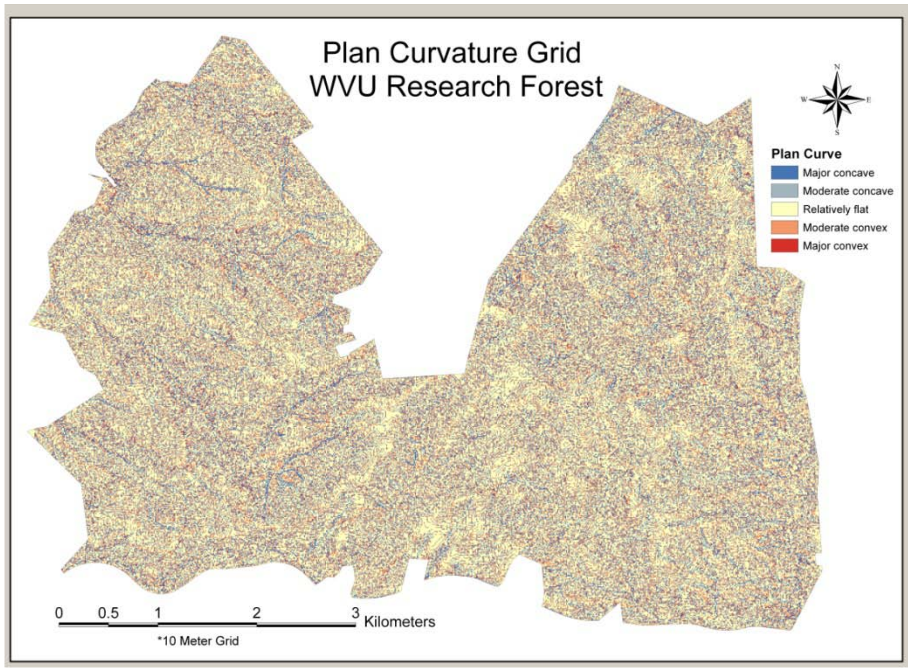

Figure 5. Plan Grid

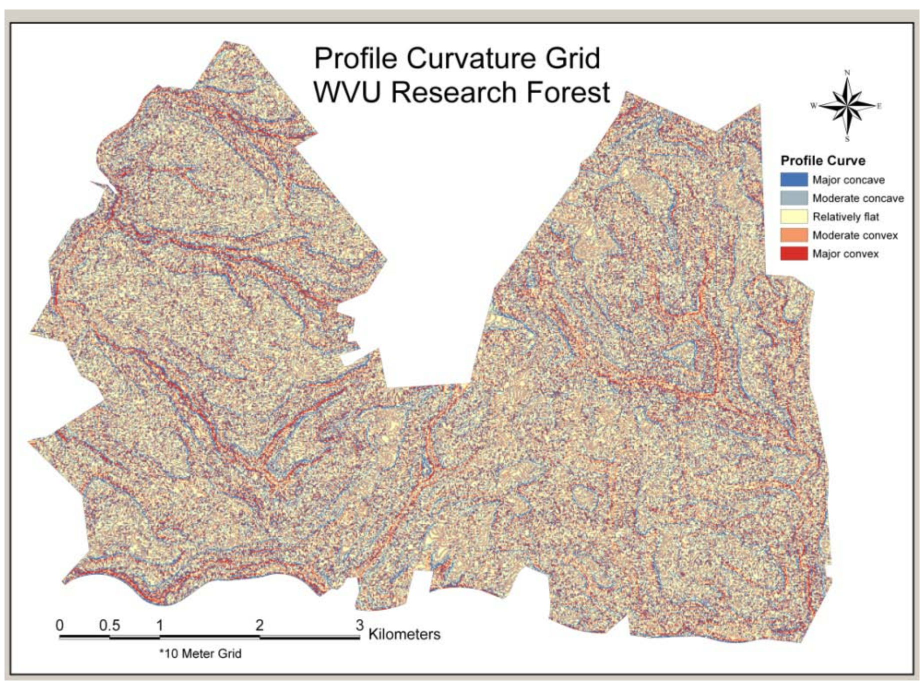

Figure 6. Profile Grid

Variable Name	Variable	Range of Values
Aspect	The aspect grid represents the orientation of the surface in regard to its cardinal direction (Beers et. Al. 1966)	0-360^{o}
Slope	The departure from a completely horizontal plane given an increase or decrease in Y for a unit change in X	0 – 47^{o}
Curvature	Representation of a grid cell elevation in relation to the eight adjacent grid cells around it. Curvature is calculated by subtracting the each individual adjacent cell from the center cell and summing the values. A grid cell that is found to be at a lower position than the grid cells around would result in a negative value and be considered concave	(-29) - 31
Plan	The curvature of the surface perpendicular to the slope direction	(-12) – 16
Profile	The curvature of the surface in the direction of the slope	(-21) – 20

Table 1. Variable Ranges

3.2.2 Field Data Collection

Field data was collected and provided by ImageTree Corporation located in Morgantown, WV. A stratified random sample approach was used to select the sampling points within 181 fixed area sample plots (Figure 7). Sample plots had of a radius of 12.44m and covered an area of 0.0486 hectares. A survey grade Trimble GPS unit was used to navigate to the predefined points generated prior to the foresters entering the field.

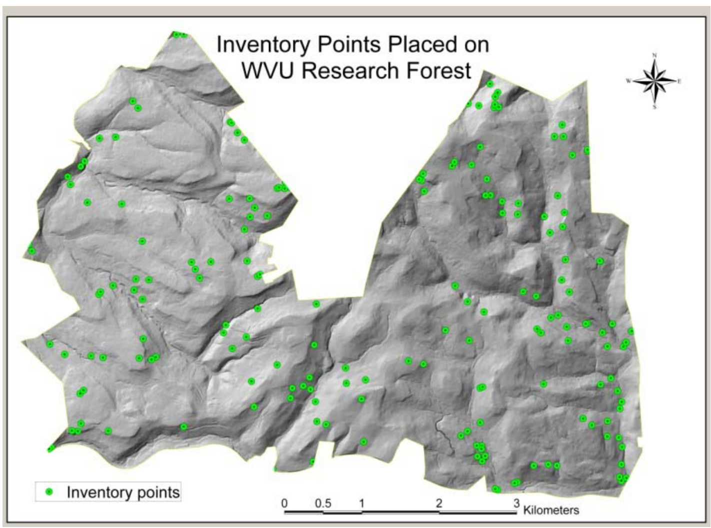

Figure 7. Sample Plots

Once plot center was located from the GPS, data collection commenced starting with the most northern tree on the sample plot. A Hagloff laser was used to determine the slope and horizontal distance from plot center to each tree to ensure it was within the boundary of the plot. Diameter at breast height (DBH) was measured for all tally trees on the plot and it was noted if the individual trees were dead or alive. Tree crowns were matched with tree crown polygons that were derived from the first return LiDAR data by Imagetree Corporation technicians. All tree crowns, or groups of crowns that were visible, were matched with the polygons to help mitigate any co-registration issues between the tree crown polygons and the

LiDAR data. The registration process is outlined in figure 8. The red blocky polygons represent automated generated tree crown polygons from the LIDAR, the blue circles are proportionally sized to represent the DBH of measured primary trees. The yellow boxes and corresponding lines represent secondary or understory trees that were measured with the yellow line indicating what primary tree the secondary trees were located under.

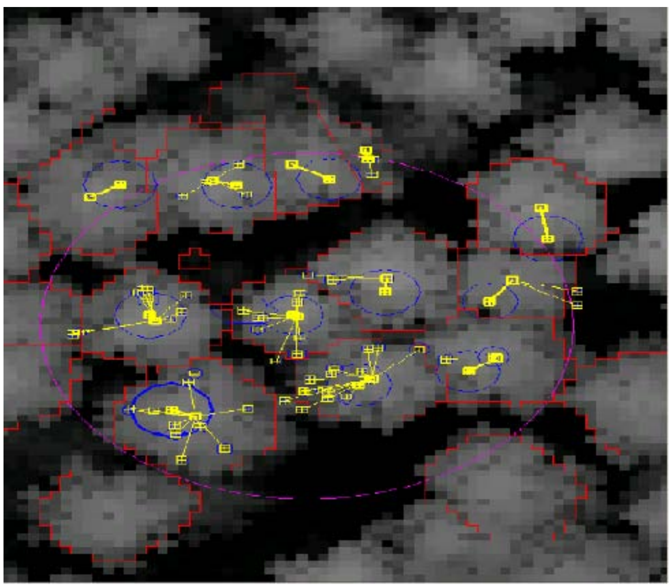

Figure 8. Tree Match

The total basal area for each individual inventory point was calculated by summing the basal area for all species tallied within the inventory point, this allowed species-specific basal area percentages to be calculated for each inventory point. For example; 11.14 sq meters total basal area and 5.57 sq meters basal area of Yellow-poplar would equate to Yellow-poplar making up 50% of the total basal area of the inventory point. Inventory points were then overlaid on the five terrain input grids which were used as predictor variables for species composition. The value of each grid was captured for each inventory point and recorded to a table for further analysis.

3.2.3 Classification and Regression Tree

Classification and regression trees (CART) are nonparametric, data-driven algorithms that generate a tree through binary recursive partitioning, where a node, representing a single variable, is split in order to divide the data into increasingly homogeneous subsets (Muñoz and Felicísimo 2004). The variables are automatically tested until the split is found for which the resulting branches are the most homogeneous or a minimum number of observations remain in the subset (Miller 2005). The end of the branch, known as the terminal node, is defined by the hierarchical rules that precede it. The CART classification was set up so that the terminal nodes indicate the percentage of basal area that is predicted to be present based on the preceding variables.

The parameters for the regression tree were set so that the final tree size was determined with the use of Mallow's Cp statistic, which determines the best statistical fit of all possible model combinations. The tree models resulted in terminal nodes ranging from 11-19 for the prominent species found on the research forest (yellow-poplar, black cherry, chestnut oak, northern red oak and white oak). An example of a final tree is shown in Figure 9.

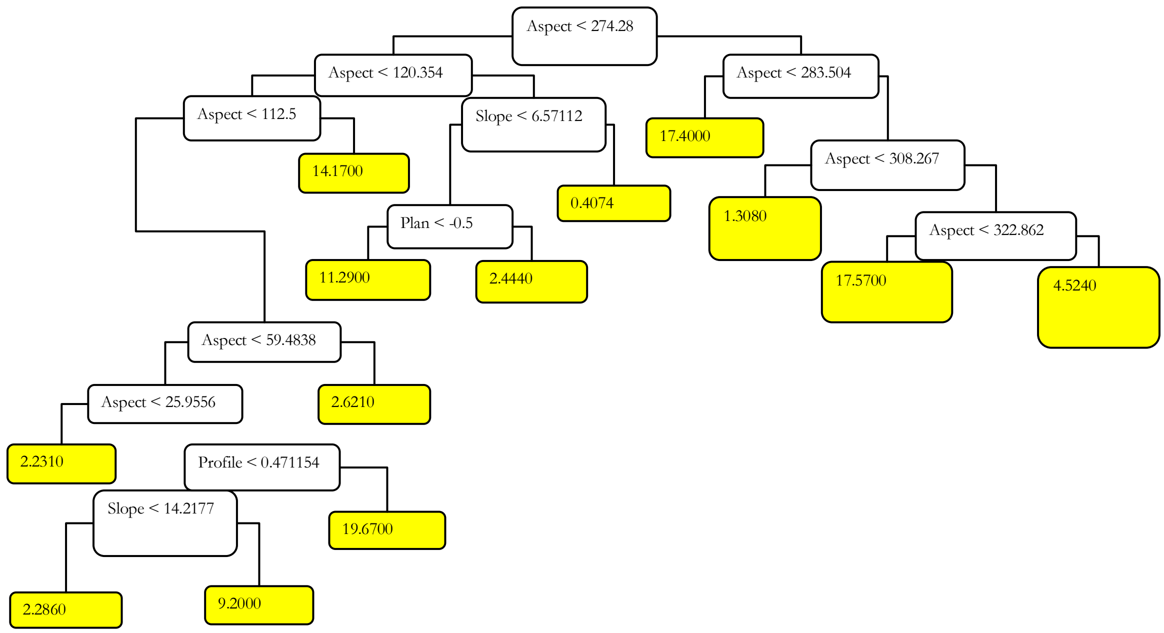

Figure 9: Black Cherry Pruned Tree

<u>3.2.4 Prediction grid creation</u>

Once a final tree was determined within S-Plus (Insightful 2008) the tree was then imported into ArcView 3.3 (ESRI 2002) via the StatMod extension (Garrard 2002) producing a prediction grid for each individual tree species. Grids were produced at the same 10m resolution as the previously created predictor variable grids. The resulting prediction grids show the estimated percentage of basal area as determined by the CART analysis. Zonal statistics were then calculated to summarize the average predicted percentage of species and the standard deviation within each of the validation tracts.

3.2.5 Validation data

Validation data was comprised of three individual stands on the WVURF that were 100% tallied prior to harvest activity. These stands are known as the Lick Run tract 14.67 hectares, the Blaney tract 57.47 hectares and the Fire Tower tract 14.16 hectares (Figure 10). Total basal area was calculated for each individual tract. The percentage that each individual species contributed to each tract was then calculated.

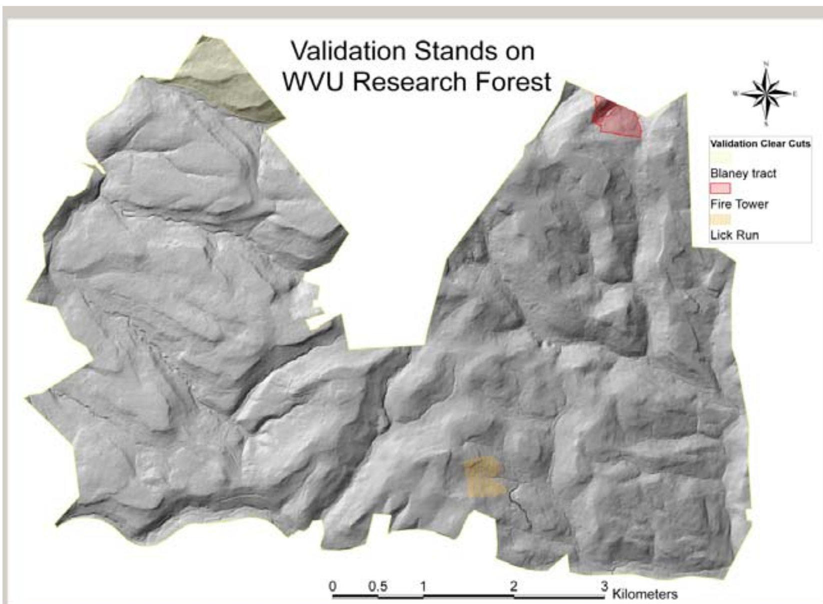

Figure 10. Validation Stands

CHAPTER 4: RESULTS

Table 2 shows the measured and predicted estimates for each of the validation stands. Yellow-poplar was the most prominent species within two of the three test tracts and was found to have an r-square value of 0.851 and a root mean square error (RMSE) of 33.47. Figure 11 shows the prediction grid for yellow-poplar. Northern red oak had an r-square value of 0.940 with a RMSE of 16.80. Table 3 shows the linear regression outputs for all species and validation stands.

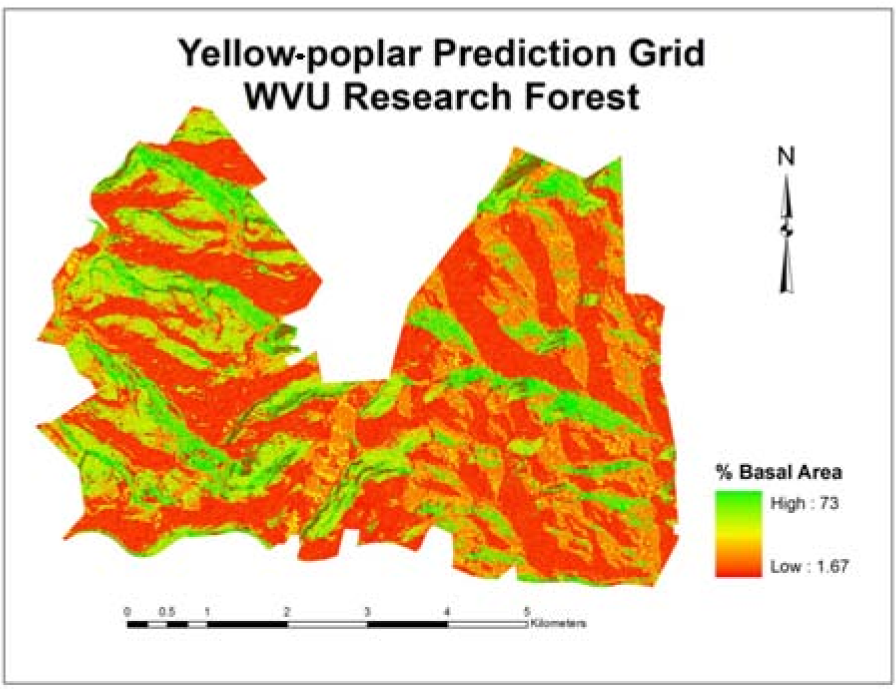

Figure 11. Yellow poplar prediction grid

	Lick Run Tract		Blaney Hollow Tract		Fire Tower Tract	
SPECIES	Measured % Basal Area	Predicted % Basal Area	Measured % Basal Area	Predicted % Basal Area	Measured % Basal Area	Predicted % Basal Area
American Beech	0.04	0.14	0.01	0.17	0.00	0.19
Black Birch	1.51	5.01	0.30	2.39	2.68	5.89
Black Cherry	7.18	4.59	0.40	3.65	5.26	6.50
Black Oak	1.82	2.12	2.96	2.81	0.47	0.30
Chestnut Oak	5.38	5.97	16.23	15.08	3.98	6.11
Cucumber tree	2.27	0.17	0.01	0.22	1.01	0.27
Hickory	0.01	0.54	0.02	0.35	1.49	0.17
Red Maple	6.53	12.75	4.17	11.56	24.94	12.76
Northern Red Oak	24.00	25.30	2.04	29.89	30.01	21.69
Scarlet Oak	8.15	1.36	5.70	9.35	0.07	1.89
Sugar Maple	0.06	0.33	0.00	0.54	3.92	0.25
White Oak	0.49	6.34	0.18	5.02	0.45	4.98
Yellow-poplar	41.22	22.20	64.87	18.58	17.04	46.30
Basswood	0.00	0.19	0.01	0.16	0.00	0.27

Table 2. Measured and Predicted Estimates

Species	r^2	Line Formula	RMSE
American Beech	0.988	y = -1.1945x + 0.1872	0.15
Black birch	0.930	y = 1.4731x + 2.2248	3.00
Black cherry	0.296	y = 0.226x + 3.948	2.50
Black oak	0.958	y = 1.0187x - 0.0357	0.22
Chestnut Oak	0.986	y = 0.7728x + 2.4642	1.44
Cucumber	0.320	y = -0.0232x + 0.2457	1.29
Hickory spp.	0.738	y = -0.1883x + 0.4472	0.85
Red Maple	0.351	y = 0.0359x + 11.931	8.98
Northern Red Oak	0.940	y = -0.2705x + 30.683	16.80
Scarlet Oak	0.027	y = 0.1757x + 3.382	4.57
Sugar Maple	0.514	y = -0.0482x + 0.4383	2.15
White Oak	0.348	y = 2.6426x + 4.4583	5.10
Yellow-poplar	0.851	y = -0.5811x + 52.877	33.47
Basswood	0.544	y = -6.1542x + 0.233	0.21

Table 3. Linear Regression Outputs

CHAPTER 5: DISCUSSION AND CONCLUSION

The results indicate that the terrain variables used in this project are reasonably effective in predicting major tree species distribution and abundance for the majority of the study area. American beech and chestnut oak proved to be the most accurately modeled species in this study in terms of the best linear fit. This is most likely due to the site specific characteristics preferred by these species (ridge tops and generally drier sites). In an earlier study, similar findings were recorded by Fekedulegn et al. (2004). American beech was consistently over predicted in the model but it occupied such a small portion of the study area the difference it represented did not have a major effect on its r-square value of 0.988. Chestnut oak conversely comprised much more of the test stands and was predicted relatively well. This is primarily observed because chestnut oak is predominately confined to specific areas, which helps to somewhat eliminate confusion when producing the final model for this species based on terrain variables.

The abundance or site specificity of a given species could potentially create a problem with the technique presented here as evident with the confusion observed between yellow-poplar and northern red oak. Northern red oak and yellow-poplar proved to be the most difficult of the species to accurately model and predict in terms of their abundance and distribution. Northern red oak and yellow-poplar comprised 68% of the total combined basal area for the validation stands and make up a major component of the study area as a whole. Although both species grow best and are typically found in coves and moist sites, the 129.4 cm of annual precipitation received on the study area allows them to compete in the drier more shallow soils of the forest; this effectively allows them to compete in virtually all areas within the study. Undoubtedly this caused confusion between the two species as observed in the results. Both exhibit a relatively strong r-square value, 0.940 for northern red oak and 0.851 for yellow poplar, although both species also comprise the largest RMSE within the study. With the exception of northern red oak on the Lick Run tract, neither of these species was predicted accurately at the validation stand level (table 2).

Species that occur in a low abundance or occur at site specific locations could be potentially missed during field sampling, or detected in the initial sampling and result in a false prediction throughout the forest. Although CART is a non-parametric model, this research highlights a major flaw in the models tendency to falsely predict species occurrences where the species undoubtedly do not occur. American basswood is an excellent example of this occurring. This species was recorded on only one of the 181 inventory plots, but was still predicted to occur over the entire study area. Species that can exploit and compete over a wide range of terrain and environmental variables, such as red maple, also proved difficult to accurately predict on the validation stands. This can be attributed to the ability of red maple to grow and compete in all areas of the forest, resulting in the model being unable to predict this species based on terrain variables.

The complexities of the eastern hardwood forest make it difficult to accurately classify individual tree distributions throughout the forest. It has been shown that the terrain variables used can predict, with some certainty, the composition of some of the major forest species at the stand level. With the ability to predict where a given species will or will not occur, a specific forest composition for a given area can be estimated to aid in management planning.

The level of accuracy achieved in this research would not have been obtainable without the high level of detail provided by the LiDAR data set. This data set virtually eliminated co-registration errors between the individual sample plots and the input variable grids that could have occurred if another DEM source were used. This type of external data could potentially influence the modest results that were obtained, but also presents the question that with the vast amount of public data available currently could the results on this study be replicated with the use of another form of DEM data.

Future strategies to improve model accuracy might include a different ground sampling strategy, the incorporation of auxiliary data from other sources such as aerial imagery, and the incorporation and the use of all LiDAR return data rather than the use of the last return only data. Incorporation of these variables has the potential to reduce the confusion in predicting between species with similar growing characteristics.

To enhance the ground sampling strategy, when stratifying the study area, marginal sites should have additional points allocated to them to account for the high variability that was not potentially captured in the initial inventory on the study area. A different and perhaps a more efficient enhancement to improve the sampling strategy would be the use of a uniform grid that could improve complete coverage of all areas of the forest and possibly help mitigate the potential of over or under sampling areas. The incorporation of the LiDAR first and mid returns could show vertical stratification differences between species that may help to refine predictor models between species.

The focus of this research was to produce predictor variables derived solely from the LiDAR point data additional research should also incorporate the use of aerial imagery as a predictor variable. The addition of aerial imagery may help differentiate between species based on color signature or texture.

The 10m DEM chosen in this project appeared to offer a sufficient resolution to capture the terrain characteristics that affect the complex species composition in an eastern hardwood forest setting. Some interesting further exploration would be to examine the effect of grid cell size on the prediction of species. Would a finer resolution of input grids increase the accuracy and precision of the models or add more noise and error? If a smaller grid was used to more closely represent the footprint of individual trees would it be possible to predict and classify where individual tree species occur rather than generalizing at the stand/landscape scale as represented in this study? The LiDAR data available for this study along with the ability to map individual tree crowns allows these additional possibilities in the optimal cell size or resolution for a wide array of studies. Furthermore, would the addition of more variables such as vertical stratification derived from the LiDAR help to classify individual forest species more accurately?

REFERENCES

Austin, M. and T. Smith. 1989. A new model for the continuum concept. Plant Ecology 83, 35. 47.

Beers, T. W., P. E. Dress, and L. C. Wensel. 1966. Aspect transformation in site productivity research. American Scientist 54: 691-692.

Bolstad, P.V., W. Swank, and J. Vose. 1998. Predicting Appalachian overstory vegetation with digital terrain data. Landscape Ecology. 13: 271-283.

Carvell, K.L. 1973. Stand Composition: Effects of the Past on the Present. Morgantown, WV: West Virginia University Agricultural and Forest Experiment Station. Forestry Notes. 1: 3-5.

Domac, A., and M.L. Süzen. 2006. Integration of environmental variables with satellite images in regional scale vegetation classification. International Journal of Remote Sensing. 27: 1329-1350.

Elliott, K.J., J.M. Vose, W.T. Swank, and P.V. Bolstad. 1999. Long-term patterns in vegetation site relationships in a southern Appalachian forest. Journal of the Torrey Botanical Society. 126: 320-334.

ESRI. 2002. ArcView GIS 3.3. Redlands, CA: Environmental Systems Research Institute, Inc.

ESRI. 2007. ArcView GIS 9.2. Redlands, CA: Environmental Systems Research Institute, Inc.

Fekedulegn, D., J. Colbert, J. Rentch, and K. Gottschalk. 2004. Aspect induced differences in vegetation, soil, and microclimatic characteristics of an Appalachian watershed. Castanea. 69: 92-108.

Fekedulegn, D., R. Hicks, and J. Colbert. 2003. Influence of topographic aspect, precipitation and drought on radial growth of four major tree species in an Appalachian watershed. Forest Ecology and Management. 177: 409-425.

Franklin, J. 1995. Predictive vegetation mapping: Geographic modeling of biospatial patterns in relation to environmental gradients. Progress in Physical Geography 19: 474.99.

Garrard, C.M. 2002. StatMod zone user's guide. Available at: http://www.gis.usu.edu/~chrisg/avext/downloads/statmodz.pdf. Accessed 2 May 2008.

Guisan, A., and N. Zimmermann. 2000. Predictive habitat distribution models in ecology. Ecological Modelling 135:147.86.

Guisan, A. and W. Thuiller. 2005. Predicting species distribution: offering more than simple habitat models. Ecology Letters 8: 993.1009.

Insightful Corp. 2008. S-Plus. Available at: http://www.insightful.com/ Accessed 2 May 2008.

Iverson, L.R., M.E. Dale, C.T. Scott, and A. Prasad. 1997. A GIS-derived integrated moisture index to predict forest composition and productivity of Ohio forests. Landscape Ecology. 12: 331-348. 176

Jensen, J.R. 2005. Introductory Digital Image Processing: A Remote Sensing Perspective, Third Edition. Upper Saddle River, NJ: Prentice Hall.

McNabb, W.H. 1989. Terrain shape index: quantifying effect of minor landforms on tree height. Forest Science. 35: 91-104.

McNab, W.H. 1993. A topographic index to quantify the effect of mesoscale landform on site productivity. Canadian Journal of Forest Research. 23: 1100-1107.

Miller, Jennifer, 2005. Incorporating Spatial Dependence in Predictive Vegetation Models: Residual Interpolation Methods. The Professional Geographer, 57: 169-184.

Muñoz, J. and Felicísimo, A.M 2004. Comparison of statistical methods commonly used in predictive modeling. Journal of Vegetation Science, 15: 285-292.

Prasad, A.M., L.R. Iverson, and A. Liaw. 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems. 9: 181-199.

Thuiller, W., Araujo, M., Lavorel, 2003. S. Journal of Vegetation Science 14: 669-680,.

Vayssières, M. P., R. E. Plant, Allen-Diaz, B.H. 2000. Classification trees: an alternative non-parametric approach for predicting species distributions. Journal of Vegetation Science 11: 679-694. 177

Wu, S., J. Li, and G.H. Huan. 2008. A study on DEM-derived primary topographic attributes for hydrologic applications: Sensitivity to elevation data resolution. Applied Geography 28: 210-223.