Development and assessment of remotely derived variables in current southern pine beetle (*Dendroctonus frontalis* Zimm.) hazard mapping in North Carolina, USA

Jason E. Moan

Thesis submitted to the faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

In

Forestry

Randolph H. Wynne, Chair
Stephen P. Prisley
Scott M. Salom
John T. Nowak

July 24, 2008
Blacksburg, Virginia

Keywords: southern pine beetle, Landsat, CART®, Disturbance index

Development and assessment of remotely derived variables in current southern pine beetle (*Dendroctonus frontalis* Zimm.) hazard mapping in North Carolina, USA

Jason Edward Moan

Abstract

The southern pine beetle (SPB) (*Dendroctonus frontalis* Zimm.) is one of the most destructive forest insect pests in the southeastern United States and has historically had a large impact on the forests of North Carolina. Many characteristics of a forest can contribute to SPB susceptibility including stand density, growth rate, age, soil type, and position on the landscape. This work was undertaken in an effort to assist and improve on the current federal SPB hazard modeling being conducted for North Carolina by the USDA Forest Service – Forest Health Protection's Forest Health Technology Enterprise Team (FHTET). In our study, predictive SPB susceptibility models were developed for each physiographic region in North Carolina using two variables not currently included in the FHTET modeling, mean stand age and the in-stand percentage of sawtimber-sized pines. These variables were obtained from USDA Forest Service – Forest Inventory and Analysis (FIA) data and North Carolina Forest Service historical SPB records creating a dataset of both infested and non-infested stands and the models were developed using the CART® classification tree approach. Two model-derived age classes (older than and younger than 22 years) were identified on the landscape using current Landsat 5 Thematic Mapper (TM) imagery chronosequences of disturbance index (DI) – transformed scenes to identify stand-replacing disturbances, resulting in a kappa statistic of 0.6364 for the younger than 22 year age class and 0.7778 for the older than 22 years age class. A kappa value of 1 is ideal. The CART® modeling effort produced valid models in all three physiographic regions of North Carolina, though the complexity of the piedmont model makes it impractical for use in the field. The dependent variable in the classification tree was presence or absence of SPB outbreak and the test sample error percentages were similar across regions, with errors ranging between 23.76 - 34.95 percent. Overall prediction success, based on the software's internal cross-validation procedure, was likewise comparable across the regions with 72.28 - 89.56 percent correctly predicted. Based on our modeling, stand age and percent sawtimber should be included in future FHTET SPB hazard modeling efforts for the coastal plain and mountains, respectively. Age classes can be reasonably estimated using Landsat or other multispectral imagery.

Acknowledgements

There are many people who've made this project a success. I'd like to start by thanking those involved in my graduate committee. This project would not have succeeded were it not for the support, knowledge, and adaptability of my advisor, Dr. Randy Wynne. Dr. Wynne has always been willing to help and has always remained high spirited about the project even if things looked discouraging. He has also been a valuable resource through which to obtain the data needed for the project. I would like to thank my committee members Dr. Scott Salom, Dr. Steve Prisley, and Dr. John Nowak. Dr. Salom has always provided great feedback and is a wealth of information on the southern pine beetle. Dr. Prisley has provided relief to my GIS quandaries when there was no relief in sight. Dr. Nowak helped arrange and maintain our involvement with the USDA Forest Service - Forest Health Technology Enterprise Team (FHTET) and their southern pine beetle hazard modeling process. As well, I would like to thank Drs. Valerie Thomas and Christine Blinn for assistance in all aspects of LiDAR- and Landsat-related processing and model building.

I'd also like to thank our direct collaborators. The idea for this project was my brainchild as a Southern Pine Beetle Forester with the North Carolina Forest Service (NCFS). I would like to formally thank the NCFS for finding enough value in my proposed project to provide funding and technical oversight. Many thanks as well to the NCFS personnel scattered across the state that have provided SPB data or personal knowledge and feedback. FHTET in Fort Collins, Colorado, specifically Jim Ellenwood and Frank Krist, have readily shared their hazard models and datasets and have welcomed the idea of trying to fine-tune their SPB hazard models for North Carolina.

I'd like to thank a few organizations from which we gathered data, including the North Carolina Department of Transportation, North Carolina Floodplain Mapping Program, North Carolina Department of Emergency Management, and the USDA Forest Service - Forest Inventory and Analysis Program, specifically Joe McCollum; everyone was very courteous and went out of their way to be helpful.

Lastly, I want to send thanks to my friends and family. My friends, many of whom have no idea what it means to remotely sense, have nonetheless provided a multitude of support over the last two years. I'd like to thank my brother, Chris Moan, for giving me the "if my little brother can do it, so can I" mindset which led to my application to graduate school in the first place. I'd like to thank my parents, Linda and Ed Moan, for supporting me and my brother in our pursuits of great things and for always being interested in the research and asking questions. Daily thanks go out to my wife, Jessie, who pushed me towards the light at the end and who now knows more about remote sensing than she ever wanted to know.

Table of Contents

List of Figures

List of Tables

Glossary of acronyms

AOI	Area of interest
CART®	Classification and regression tree (CART 5.0)
DBH	Diameter at breast height
DEM	Digital elevation model
DI	Disturbance index
DOS	Dark object subtraction
ESUN	Earth-sun distance in astronomical units
ETM	Enhanced Thematic Mapper (Landsat)
FHP	Forest Health Protection
FHTET	Forest Health Technology Enterprise Team
FIA	Forest Inventory & Analysis
LIDAR	Light detection and ranging
MRLC	Multi-Resolution Land Characteristics Consortium
NAD	North American datum
NED	National Elevation Dataset
NC DOT	North Carolina Dept. of Transportation
NCFS	North Carolina Forest Service
NDMI	Normalized difference moisture index
NDVI	Normalized difference vegetation index
NIPF	Non-industrial private forests
NRCS	USDA-Natural Resource Conservation Service
QMD	Quadratic mean diameter
RMSE	Root mean square error
SPB	Southern pine beetle
SPBPRP	SPB Prevention and Restoration Program
TOA	Top of atmosphere reflectance or radiance
TC	Tasseled cap transformation
TM	Thematic Mapper (Landsat)
USGS	U.S. Geological Survey

Introduction

In previous years, while southern pine beetle (SPB), *Dendroctonus frontalis* Zimm. (Coleoptera: Curculionidae: Scolytinae), populations were in outbreak status across the southeastern United States, most SPB management recommendations were focused on suppression techniques. As the SPB outbreaks eventually subsided, the management focus has recently shifted towards SPB prevention (Nowak et al. *in press*). To facilitate the shift from a suppression-based approach to more proactive integrated approaches such as hazard mapping and preventive management, the Southern Pine Beetle Prevention and Restoration Program (SPBPRP) was created by the USDA Forest Service - Forest Health Protection section (FHP) and the Southern Group of State Foresters in 2003 (Nowak et al. *in press*). Hazard maps, specifically those being produced by the USDA Forest Service - Forest Health Protection's Forest Health Technology Enterprise Team (FHTET), have the potential to become a useful tool to assist natural resource managers in determining areas in which to focus SPB prevention activities. Current cooperative efforts between FHTET, the SPBPRP, and several southern state forestry agencies have led to the development of SPB hazard maps for the southeastern United States.

While several other southeastern states have developed SPB hazard rating systems specific to their states, North Carolina does not have an established method for rating the SPB hazard across the state and is thus relying heavily on the outcome of the federal hazard mapping process. Our research focuses on trying to improve and assess the SPB hazard modeling effort for North Carolina non-industrial private forests (NIPF), which account for 78% of the total forestland or 13.8 million acres (5.6 million hectares) in North Carolina (Brown 2004). Of these 13.8 million acres of NIPF, approximately 23% or 3.2 million acres (1.3 million hectares) are classified in the loblolly-shortleaf pine species group, the preferred host species for SPB, with another 2.2% or 304,500 acres (123,227 hectares) in other pine-dominated species groups (Brown 2004).

Hazard/Risk

In order to successfully predict and map southern pine beetle (SPB) hazard, one must first agree on a suitable definition of hazard. Hazard and risk are not synonymous, though they are often used interchangeably. Gottschalk (1995) noted that hazard is directly related to susceptibility and/or vulnerability and the probability of affecting the overall management objectives for a given area. Susceptibility, in our case, is the probability of a SPB attack, based on stand biological parameters, whereas vulnerability is the likelihood of stand damage during a successful attack. Risk, then, is the probability that the SPB will occur or reach outbreak populations in a given area (Gottschalk 1995). FHTET uses hazard and risk synonymously and classifies susceptibility and vulnerability as mortality potentials, not probabilities, to produce their hazard/risk maps (Krist Jr. et al. 2007). Their definitions of susceptibility and vulnerability follow those of Gottschalk (1995).

The southern pine beetle

The southern pine beetle *Dendroctonus frontalis* Zimm. (Coleoptera: Curculionidae: Scolytinae) has long been a major insect pest of pines in the eastern and southern United States, as well as in portions of Central America (Payne 1980). Records dating back to the early 1900s warn landowners that “There is probably no more serious enemy of pines in the Southern States then this beetle” (St. George and Beal 1929). To this day, the SPB continues to be the a major insect pest of southern pines (http://www.sfiwc.org/reports/SPB_losses98-05.pdf).

The SPB is found across the range of southern yellow pines, extending from southern New Jersey west to Oklahoma and south to northern Florida (Clarke and Nowak 2008). The SPB has also been found in Arizona and into Central America (Vite et al. 1974). The SPB typically infests and successfully colonizes all *Pinus* species within its range, but has been known to preferentially attack loblolly pine (*Pinus taeda* L.) and shortleaf pine (*P. echinata* Mill.). Eastern white pine (*P. strobus* L.) and slash pine (*P. elliottii* Englem.) are considered less susceptible than most pine species, but can be successfully colonized during large epidemics. Other susceptible species include Virginia pine (*P. virginiana* Mill.), pitch pine (*P. rigida* Mill.), pond pine (*P. serotina* Michx.), Table Mountain pine (*P. pungens* Lamb.), spruce pine (*P. glabra* Walt.), and even red spruce (*Picea rubens* Sarg.) and Norway spruce (*Picea abies* L.) (Payne 1980). A study conducted during an active outbreak period in North Carolina (1973-1975) showed host preference of shortleaf pine, Virginia pine, pitch pine, and loblolly pine, with shortleaf being the most preferred host (Anderson and Doggett 1993).

The question of what stand characteristics lead to SPB infestation has been researched nearly as long as the SPB has been recognized as a pest. There are a multitude of factors that lead to SPB outbreak and infestation, though our research focuses on only a fraction of them. Typically, SPB outbreaks occur in slow growing, dense stands, primarily composed of overmature sawtimber (Belanger et al. 1979; Flamm et al. 1988; Coulson et al. 1999). Age is a factor that has been found to affect a stand's susceptibility to the SPB in some studies and not in others. Lorio and Sommers (1981) concluded that loblolly pine stands had increased susceptibility to SPB attack after age 35 and a recent study conducted in Alabama concluded that the risk begins to decrease after age 40 (Ylioja et al. 2005).

Hazard mapping

The FHTET hazard modeling goal is to predict what areas will lose 25% or more of the total basal area in stems over 1 inch (2.54 cm) diameter due to forest pests within the next 15 years. FHTET has been developing SPB hazard designations using both remotely derived data and forest inventory and analysis data (FIA). The FHTET hazard modeling framework uses a multitude of data sources, including data from the 2001 National Land Cover Database, the USDA-Natural Resource Conservation Service's (NRCS) National Cooperative Soil Survey, the United States Geological Survey's (USGS) National Elevation Dataset, the PRISM 800 meter national climate dataset, and the USDA-Forest Service's FIA program (Ellenwood et al. *in press*). The output variables are weighted based on their importance, which has been determined from literature, expert opinion, and outbreak data where available. For the hazard map of North Carolina, no SPB outbreak-related stand variable data collected during previous SPB outbreaks were incorporated.

Using the FIA data, FHTET produced 47 layers to predict hazard with the layers of continuous variables being modeled using bilinear methods and the categorical variables being modeled using nearest neighbor methods. Individual models of basal area and trees per acre were created for each U.S. Geological Survey physiographic map zones (Figure 1) using Cubist 2.05 (RuleQuest Research 2007. Sydney, Australia). For infrequent tree species, FHTET developed a new procedure which they dubbed "extirpolation" (Ellenwood et al. *in press*). In the extirpolation technique, the basal area of the infrequent species is subtracted from the total forest basal area and the model is run. The result is then subtracted from the original total forest basal area to predict the basal area represented by the species of interest. The reported accuracies for this modeling process are $R^2 = 0.61 - 0.77$ for total basal area, $R^2 = 0.41 – 0.61$ for total trees per acre, $R^2 \leq 0.50$ for individual tree species (Ellenwood et al. *in press*).

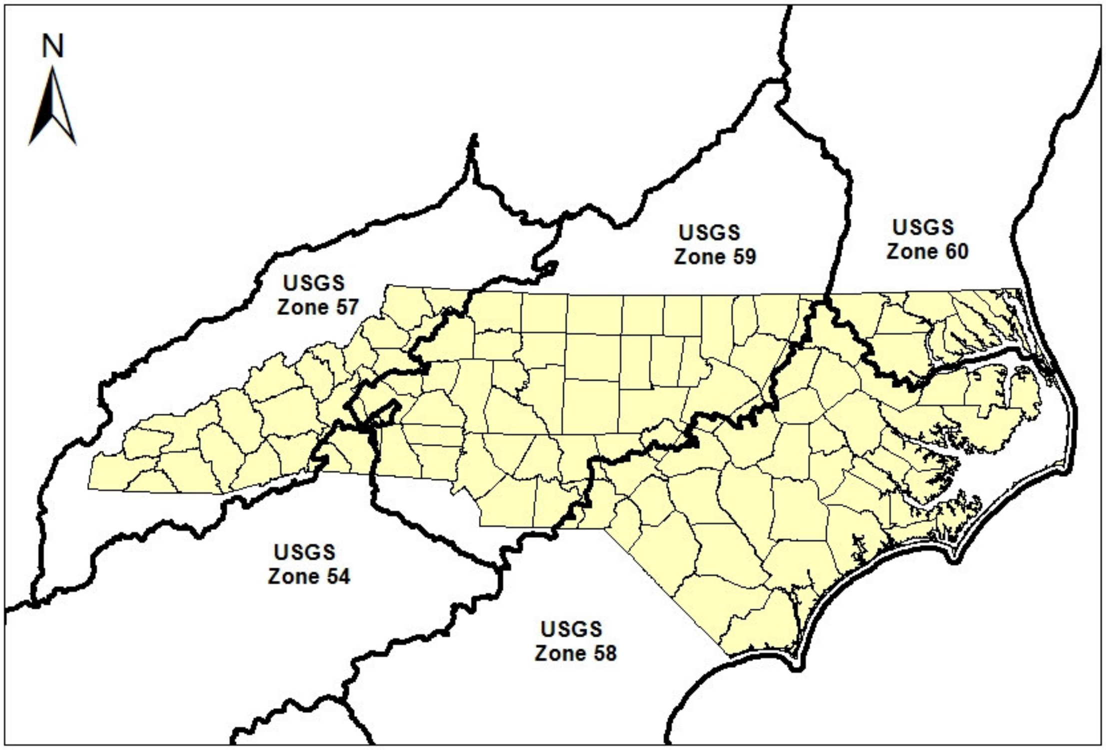


Figure 1. North Carolina showing the arrangement of U.S. Geological Survey map zones used by FHTET as boundaries for host and variable estimation.

Once host and variable layers have been developed the data are entered into one of five different hazard models that cover North Carolina, including three in the coastal plain. Each of these models is viewed as a spreadsheet which lists the factors affecting both susceptibility and vulnerability; the models for North Carolina are shown in Appendices 1-5. FHTET developed a risk/mortality scaling tool (Figure 2) which scales the output of each variable layer into a standardized 0-10 class, with 10 representing a 100% mortality rate. These standardized risk/hazard classes are then multiplied by the host basal area layer to produce an estimate of basal area loss.

Risk/Mortality Scaling Tool

To obtain eleven class values (for risk values, mortality thresholds), enter the risk begins and risk peaks values. Equal interval classes will be calculated.

	Input Value	Classes	Scaled Value
Risk Begins (0):	0	0	0
		9	1
		17	2
		26	3
		34	4
		43	5
		51	6
		60	7
		68	8
		77	9
Risk Peaks (10):	85	85	10

Figure 2. The FHTET risk/mortality scaling tool for standardizing risk or hazard classes (http://www.fs.fed.us/foresthealth/technology/nidrm_spb.shtml).

The susceptibility and vulnerability factors which occur in each of the FHTET regional models for North Carolina vary across the state and are decided upon by experts in the field and literature. Both coastal plain models include no susceptibility variables and are instead based solely on the vulnerability classifiers of stand density index and soil drainage index. In the piedmont, susceptibility is based on percent host, slope/aspect, and the percent clay in the soil, while vulnerability is based on host basal area. The mountain region is reasonably similar to the piedmont model but uses percent host, slope, and shadow effect to predict susceptibility and again uses host basal area to predict vulnerability. Constrained by data availability, stand age is not incorporated into the

FHTET modeling for any of the physiographic regions in North Carolina, despite its common inclusion in previous hazard and risk mapping research (Lorio Jr. 1978; Belanger et al. 1979; Belanger et al. 1993).

The FHTET SPB hazard ratings classify each 30 meter pixel of host into one of six possible hazard classes: little or none, low, moderate, moderate/high, high, and very high. The basis for this hazard rating system was developed from research conducted by Clemson University (Karpinski et al. 1997) and also that published by the USDA Forest Service (Hicks Jr. 1980).

Age class determination

Landsat imagery has been used in numerous studies to determine the time since harvest for forest stands (Wilson and Sader 2002; Wulder et al. 2004; Healey et al. 2005). Typically, Landsat imagery is transformed in some way before the date of harvests, and eventually age, can be extracted. Wilson and Sader (2002) converted Landsat scenes to the normalized difference vegetation index (NDVI) and the normalized difference moisture index (NDMI) before viewing the scenes as a color composite image to identify harvests. They compared harvest identification based on the number of years between scene acquisitions, with a maximum of 6 years and had the highest kappa value (0.76) for NDMI scenes acquired 2-3 years apart. In a study of forest carbon modeling, Wulder et al. (2004) used the tasseled cap transformation (TC) on Landsat Enhanced Thematic Mapper (ETM+) scenes to predict the age of lodgepole pine (*Pinus contorta* Douglas ex Louden) in British Columbia, Canada. Their highest R^2 value of 0.68 for age class determination was achieved using the TC wetness, ETM+ Band 7, ETM+ Band 3, ETM+ Band 5, and a constant, developed using a stepwise multivariate regression procedure.

Healey et al. (2005) developed a new method for determining time since harvest which they called the disturbance index (DI). The DI is derived from the tasseled cap indices and is based on the assumption that recently cleared stands have a higher relative tasseled cap brightness and lower tasseled cap greenness and tasseled cap wetness than that of an undisturbed stand (Healey et al. 2005). The longest time interval tested in this study was 9 years in a site in Russia, where regeneration time is slow. They achieved good results in the Russian study site with DI when using a 9-year interval, whereas another study site, this one in the fast-growing forests of western Washington, USA, did not perform as well using DI over an 8-year interval.

Objectives

The main objective of this research project was to develop a predictive southern pine beetle (SPB) hazard model specific to the forests of North Carolina. This model was based on aspatial historic SPB outbreak data and comparable non-outbreak data from the FIA program to determine the potential utility of stand age and percent sawtimber for SPB hazard modeling. Aspatial data was used out of necessity, as the majority of historical North Carolina SPB outbreak data did not include location information.

Though there are a multitude of other important variables which are being included in the FHTET hazard models, they are based on expert opinion and literature.

Our models were directly based on records from historical SPB outbreak data in North Carolina. After scouring all historical SPB outbreak data available from the North Carolina Forest Service (NCFS), there were only two consistently tracked variables shown to be related to SPB susceptibility while also being comparable with the FIA datasets: stand age and percent sawtimber (Belanger et al. 1979; Lorio and Sommers 1981; Ylioja et al. 2005).

Additional research objectives included the derivation of stand age classes through manipulation of Landsat Thematic Mapper (TM) imagery using the disturbance index (Healey et al. 2005) and comparison of the FHTET-produced biophysical parameters before and after incorporation of a LiDAR (Light detection and ranging) - derived digital elevation model (DEM) in the mountain region of North Carolina. For geographic areas in which we found that age was a good predictor of SPB susceptibility, the age classes of interest were obtained by manipulating the available Landsat TM imagery to identify approximate time of stand replacing disturbances. Previous research laid the groundwork for predicting age classes of pines in North Carolina using Landsat imagery. Our research attempted to classify pine age classes over long acquisition intervals (20+ years) using the disturbance index in an image differencing methodology, whereas Healey et al. (2005) and other researchers have used color composite analysis to identify the recent harvests in a scene. Our output age classes could be mapped and classified into hazard classes based on the FHTET framework.

The location and density of host species can be modeled across the landscape by comparing the measured forest composition of FIA plots with the topographic variables derived from the DEM (Ellenwood and Krist Jr. 2007). Our goal regarding the inclusion of the LiDAR DEM in the modeling process was to ensure the variables on which the hazard models were based, species basal area, total basal area, and total trees per acre, are as accurate as possible. Ellenwood and Krist Jr. (2007) report that they achieved the highest average R^2 values for total basal area (0.68), while trees per acre generated a R^2 value of 0.50 and they noted that individual species basal area R^2 was rarely over 0.50 as well.

We chose to study the DEM effect on the modeled biophysical parameters in the mountain region in part because the topographic differences are the most dramatic in this region. We were also told by FHTET that they were less confident about the accuracy of their current DEM in this region than in the other regions of the state (Jim Ellenwood and Frank Krist Jr., pers. comm., FHTET, May 15, 2007).

Throughout this project we put an emphasis on using readily available or soon-to-be readily available data sources where possible in an attempt to show the utility of low-cost and/or free data sources in SPB hazard modeling.

Materials and Methods

Study sites

The scope of this project and the predictive hazard modeling process encompassed the entire state of North Carolina. From the loblolly pine plantations and hardwood swamps of the coastal forests to the high elevation spruce (*Picea spp.*) and fir (*Abies spp.*) of the Appalachian mountains, SPB activity differs across the state (Doggett and Tweed 1994). To accommodate these differences in our modeling we have divided the state into the three physiographic regions identified by the FIA program: coastal plain, piedmont, and mountain (Figure 3). We then selected the area within one Landsat 5 TM scene in which to focus our age class analysis; this scene from path 18, row 35 (shown outlined in red in Figure 3) contains portions of 17 counties all within the mountain region. Approximately half of the scene covers Tennessee and a small portion of this scene fell within the piedmont region; these areas were consequently clipped to avoid modeling errors. For the entire state, each county's FIA region designation is shown in Appendix 6. We chose the path 18, row 35 Landsat 5 TM scene for several reasons, including the historic SPB outbreak occurrence, the mix of forest types, and the availability of high quality chronologically appropriate imagery. Future studies should include Landsat imagery from the piedmont and coastal plain regions.

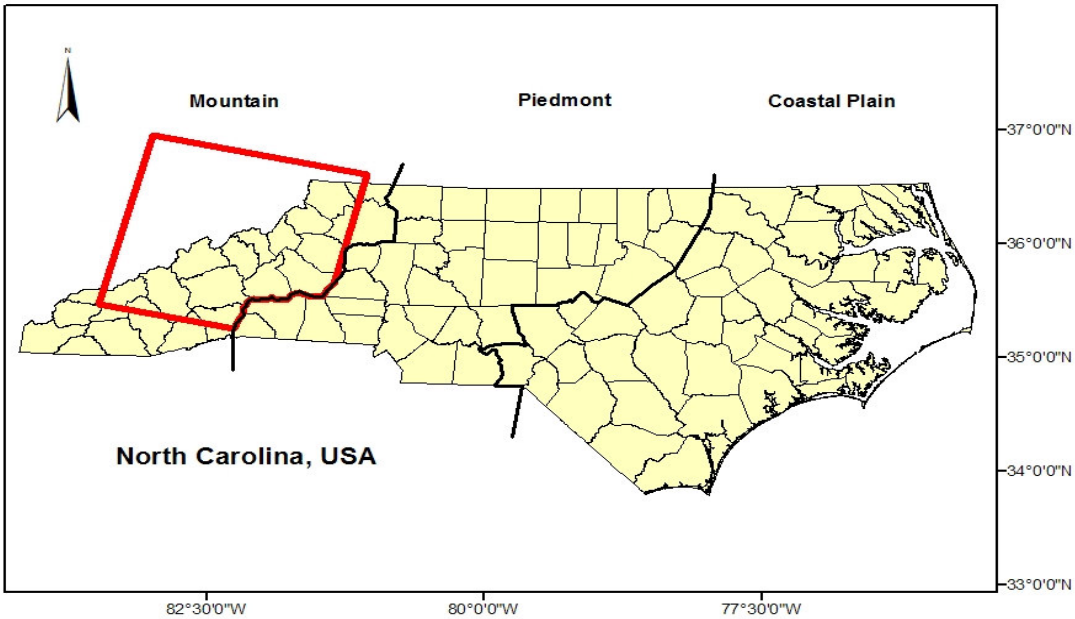


Figure 3. Three FIA physiographic regions of North Carolina (mountain, piedmont, and coastal plain) delineated by black lines and the area contained by Landsat 5 TM Path 18 Row 35 shown outlined in red. This Landsat scene was used in the age class determination.

SPB hazard modeling

A North Carolina SPB historical outbreak database was developed containing data from all areas monitored by the NCFS. Areas monitored by the NCFS generally consist of non-industrial private forests, however since most of these data were gathered via aerial survey they likely represent some industrial forests and state/federally owned forests. The database was created in Microsoft Access (Microsoft Corporation. 2007. Seattle, WA) and was compiled from a wide variety of historic NCFS SPB outbreak reporting forms, ranging from 1969 to the most recent date of 2005. All of these forms are non-spatially specific and were created to assist with SPB salvage operations, making tracked outbreak variables less than ideal for our modeling. The variables taken from these data and included in our modeling effort were percent sawtimber and stand age. As previously mentioned, these variables were not incorporated into the FHTET modeling and led us toward the creation of a separate model for this project. The SPB outbreak reporting forms from which our modeling variables were identified have varied slightly over time. Nevertheless, nearly all of the forms list the following additional variables for each SPB outbreak; these variables were not used in our modeling process: number of trees affected, acreage, average diameter at breast height (DBH) of sawtimber, number of sawlogs, DBH of pulpwood, height of pulpwood, and tree species present. Tree species is known to be related to SPB susceptibility (Payne 1980; Anderson and Doggett 1993; Veysey et al. 2003) but could not be included in our modeling due to incomplete reporting across the NCFS forms. Several tree species were often listed for each SPB outbreak, but the proportion of each was not consistently recorded.

As the SPB outbreak database was being developed, we obtained aspatial FIA data from the 1974 periodic survey of North Carolina only. We chose the 1974 survey because NCFS records compiled by former NCFS Pest Control Branch Head Coleman Doggett suggested that the majority of the counties in the state were in SPB outbreak over much of the 1970s. The choice of the 1974 data was further confirmed by a study which identified 1973-1975 as having been a period of substantial SPB outbreaks in North Carolina (Anderson and Doggett 1993). We felt that comparing our NCFS outbreak data to the FIA stands that were not attacked during an outbreak period would provide us with a more robust model. To ensure our FIA sample plots were not infested by SPB, we filtered out any plots that listed any type of damage during this survey whether it was insect related or otherwise.

We wanted to identify variables recorded by the FIA program that would be comparable to NCFS SPB outbreak variables. Stand age was gathered directly in both the NCFS data and the FIA data. Percent sawtimber, however, while listed for the NCFS data, was calculated manually in the FIA data using a value of eleven inches or greater as the definition for sawtimber. We concluded that eleven inches was an acceptable cutoff value for sawtimber because the official NCFS definition is twelve inches with the ten to twelve inch diameter classes being classified as sawtimber depending on tree form and the demands of the timber markets (Brian Council, pers. comm., North Carolina Forest Service, Jan. 11, 2008).

The FIA and NCFS data described above were the inputs for our modeling analysis, using our two most comparable variables: mean plot age and percent of sawtimber sized pines in a given plot. These datasets are binary, where each entry was either attacked by SPB or not attacked. Sample sizes were equal for both attacked and non-attacked stands for each region, but sample sizes varied across regions based on data availability. Numbers of observations were 202, 632, and 206 for the coastal plain, piedmont, and mountain region, respectively. These sample sizes were based on data availability, and the higher numbers in the piedmont do not necessarily reflect a higher incidence of SPB attacks.

To determine the importance of our input variables for each region, we ran a classification tree analysis using CART 5.0 (Salford Systems 2003-2004, San Diego, CA). CART has been shown in previous research to be an acceptable nonparametric method for dealing with binary data (Lindbladh et al. 2002; Lindbladh et al. 2003; Shifley et al. 2006). CART 5.0 separates the binary dataset into a decision tree by identifying patterns within the data and it can be used as a predictive tool (Salford Systems 2002). The decision tree produces parent nodes that divide the data into smaller groups and terminal nodes which show how well CART could separate the data. An example to illustrate this is shown in Figure 4. Both age and percent sawtimber were predictor (independent) variables in our model with the target (dependent) variable set as presence/absence of SPB infestation, while all other inputs to the modeling were the defaults. CART's default settings include running a 10-fold internal cross validation, assuming non-weighted variables, and utilizing the Gini splitting rule (Salford Systems 2002).

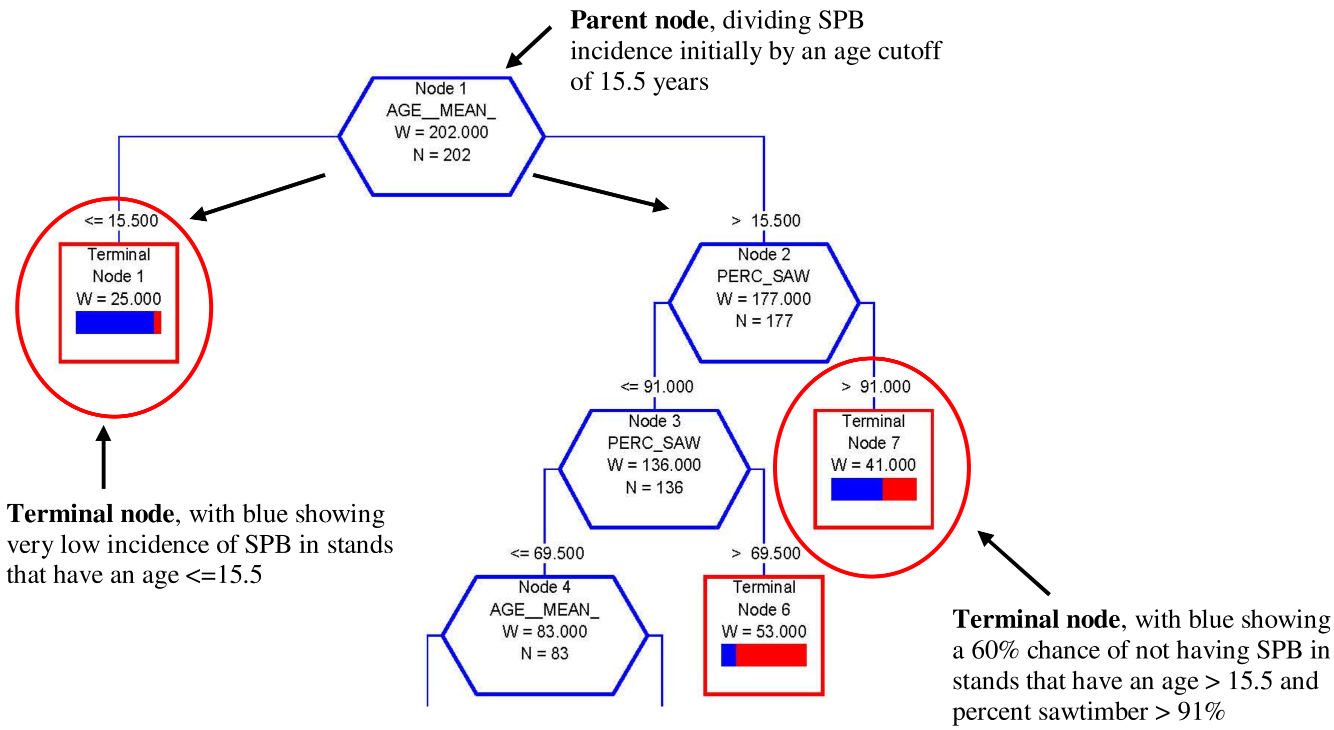


Figure 4. An example of a CART® output tree. Red in the terminal nodes represents SPB incidence, while blue represent a lack of SPB incidence.

Age class determination

Terrain-corrected Landsat 5 TM images for use in stand age determination were obtained from multiple sources, including the Multi-Resolution Land Characteristics Consortium (MRLC) database and the U.S. Geological Survey (USGS), and were all processed using Erdas Imagine 9.0 (Leica Geosystems Geospatial Imaging LLC. 2006. Norcross, GA). Based on our interest in pine, we tried to exclusively use leaf-off (preferably winter) images in our study. Interval length between scenes was determined by availability of acceptable imagery and the results of the database modeling effort. Two Landsat 5 TM scenes from path 18, row 35 were used in the mountain region, one from January 20, 2007 and one from February 08, 1985, creating an interval of approximately

22 years. This 22-year interval was used due to data availability despite our modeled age class cutoff being slightly higher at 26.5 years. Although the imagery was terrain-corrected, the Autosync co-registration tool in Erdas Imagine 9.0 was used to improve the registration, resulting in a final co-registration root mean square error of 0.0555 pixels.

To conform to the projection information being used by FHTET, all scenes were reprojected into NAD83 Albers Conical Equal Area (USGS version) prior to all other preprocessing; Erdas Imagine does not contain information on this projection and so a new projection had to be created (Appendix 7). All scenes were converted to top of atmosphere (TOA) reflectance using a model we developed in Erdas Imagine 9.0 (Leica Geosystems Geospatial Imaging LLC. 2006. Norcross, GA). The model has two steps, the first of which is the conversion of the digital numbers to TOA radiance (Appendix 8) and then secondly the conversion from TOA radiance to TOA reflectance (Appendix 9). These conversions are known as radiometric corrections and are used as a way to normalize the data so that they are comparable across scenes from a variety of years and sensors (Chander et al. 2007). The equations for conversion from digital numbers to TOA radiance (1) and then to TOA reflectance (2) are shown below (Chander and Markham 2003). Table 1 shows the inputs for both earth-sun distance and solar zenith angle for our two scenes in the TOA reflectance model.

(1) $L_i = G_{rescale} * Q_{cal} + B_{rescale}$

Where L_i: Spectral radiance at sensor for band i (watts/(m^2 sr μm))

Q_{cal}: Calibrated pixel digital number

$G_{rescale}$: Gain factor rescaled for each band [(watts/(m^2 sr μm))/ DN]

$B_{rescale}$: Bias factor rescaled for each band [(watts/(m^2 sr μm))/ DN]

(2) $R_i = (\pi L_i d^2)/(ESUN_i \cos\theta)$

Where R_i: Top of atmosphere reflectance (watts/(m^2 sr μm))

L_i: Spectral radiance at sensor for band i (watts/(m^2 sr μm))

d: Earth-sun distance in astronomical units

$ESUN_i$: Mean solar exoatmospheric irradiance for band i (watts/(m^2 μm))

θ: Solar zenith angle (degrees)

Note: Our TOA reflectance model requires θ to be in radians and not degrees, so all θ values were multiplied by 0.0174532925 to convert them to radians prior to running the model.

Table 1. Earth-sun distance (Au) and solar zenith angle (º) inputs used in the model to calculate top of atmosphere reflectance of the Landsat scenes.

Scene Date	**Earth-Sun distance (Au)**	**Solar zenith angle (º)**
20 JAN 2007	0.98535	60.8
08 FEB 1985	0.9856	59.0

Once all the scenes were converted to TOA reflectance, we rescaled them from a float single image to a signed 16-bit image to minimize file sizes. To rescale the images, we multiplied each band by 10,000 and added 100 to each band to ensure positive values. This rescaling was completed using a rescale model (Appendix 10) created in Erdas Imagine 9.0 (Leica Geosystems Geospatial Imaging LLC. 2006. Norcross, GA). We chose not to proceed with further radiometric corrections such as a dark object subtraction (DOS). Dark object subtraction and other corrections are typically used to normalize the reflectance values of the images, however Cohen et al. (1998) noted that the difference in brightness values between forest and clearcut forest is noticeable enough without these corrections. Our scenes were rescaled for normalization in a latter portion of the age class determination as outlined by Healey et al. (2005).

To derive stand age classes, we compiled two Landsat 5 TM scenes for path 18, row 35 as previously mentioned. These two scenes were clipped to the state boundary of North Carolina, the mountain FIA region, and also to each other to ensure complete overlap of the two scenes. After clipping the scenes, an acquisition date-specific mask was applied to each scene to remove all non-forest areas and pure deciduous forests. This masking was accomplished by creating a training dataset of non-forest and deciduous forests through manual interpretation of the Landsat scenes with 1998 color-infrared leaf-off aerial photographs used as a reference. The AOIs for each scene were classified into three land use classes, pine (class 1), other forest (class 2), and non-forest/noise (class 3). The spectral signatures for each AOI were assessed for separability using a contingency table and were edited to improve the classification results. Once the training data were developed, a supervised classification was performed on each scene using a maximum likelihood classifier until a visually suitable classification was created. Each scene was recoded to represent only the three land use classes of interest and the image was then prepared for accuracy assessment before any masking was completed.

The supervised classification accuracy was assessed using non-randomly assigned points of known land use with 40 points per class scattered throughout the scene. All accuracy assessment was completed in Erdas Imagine 9.0. Classification success was determined based on both the user's accuracy (error of commission) and the kappa statistic for Class 1. Errors of commission equate to designating, for example, an actual area of forest to the non-forest class in the classified image (Campbell 2006). The kappa statistic is a commonly used measure of classification accuracy, ranging between 0.0 and 1.0, which compares the given classification to that of randomly assigned classes (Cohen et al. 1995; Hayes and Sader 2001; Healey et al. 2005). A kappa value of 1 is ideal. Table 2 shows the user's accuracy and the kappa statistics for Class 1 in each scene, while Table 3 shows the error matrices for each of the two scenes.

Table 2. Classification accuracies for the maximum likelihood supervised classification to determine the extent of pine in the Landsat scenes.

Scene Date	**User's Accuracy (pine)**	**Kappa statistic (pine)**	**Kappa statistic (overall)**
20 JAN 2007	93.10%	0.8966	0.7500
08 FEB 1985	100.00%	1.0000	0.8375

Table 3. Classification error matrices for the maximum likelihood supervised classification to determine the extent of pine in the Landsat scenes. Class 1 is pine, Class 2 is other forest, and Class 3 is non-forest/noise. The values in each cell represent the number of accuracy assessment points (n=40) that were classified in each of the three classes.

20 JAN 2007			
Classified Data	**Class 1**	**Class 2**	**Class 3**
Class 1	27	2	0
Class 2	9	36	3
Class 3	4	2	37
08 FEB 1985			
Classified Data	**Class 1**	**Class 2**	**Class 3**
Class 1	28	0	0
Class 2	12	40	1
Class 3	0	0	39

Once pine could be identified and all other land types filtered out, the scenes were processed using the tasseled cap transformation. The tasseled cap transformation was performed and the resulting indices were subset to represent only the three bands on interest. Each scene was then rescaled and transformed into the disturbance index.

Before the disturbance index can be calculated, each tasseled cap image must be rescaled to its standard deviation above or below its respective mean forest value, which acts as a normalization procedure. We used the rescaling equations (1) and the subsequent disturbance index (DI) equation (2) as outlined in Healey et al. (2005).

(1)

$$B_r = (B - B_\mu)/B_\sigma$$

$$G_r = (G - G_\mu)/G_\sigma$$

$$W_r = (W - W_\mu)/W_\sigma$$

(2)

$$\mathrm{DI} = B_r - (G_r + W_r)$$

Where B, G, W represent brightness, greenness, and wetness, respectively, while r is the rescaled value for each; μ is the mean forest value for each; and σ is the standard deviation of the forest value for each. Our disturbance index model developed in Erdas Imagine 9.0 is shown in Appendix 11.

The two tasseled cap Landsat TM scenes were converted to the disturbance index, which resulted in values ranging between -14 to 10. The disturbance index values are integers because our preprocessing steps and disturbance index calculation steps designated them as such to provide for easier interpretation of the results and to save on data storage space. The raster calculator tool in ArcGIS 9.2 was used to subtract the 1985 disturbance index scene from the 2007 disturbance index scene to create a resulting difference image with disturbance index values from -13 to 15. We visually inspected the resulting difference image to determine a suitable cutoff value for highly disturbed forest and found that areas that had been cleared within our 22-year time period had the values ranging from -13 to -3; values of -2 picked up numerous partial harvests and values of -1, 0, or 1 could be considered static or relatively so. Disturbance index values of 2-15 were visually identified as being undisturbed pixels. Due to the relatively small range of disturbance index values, a small change in our threshold values for harvested and no-harvest pixels was magnified in its effect. Figure 5 shows the raw 1985 and 2007 Landsat scenes in an area that was harvested within the 22-year interval, as well as the resulting disturbance index difference image.

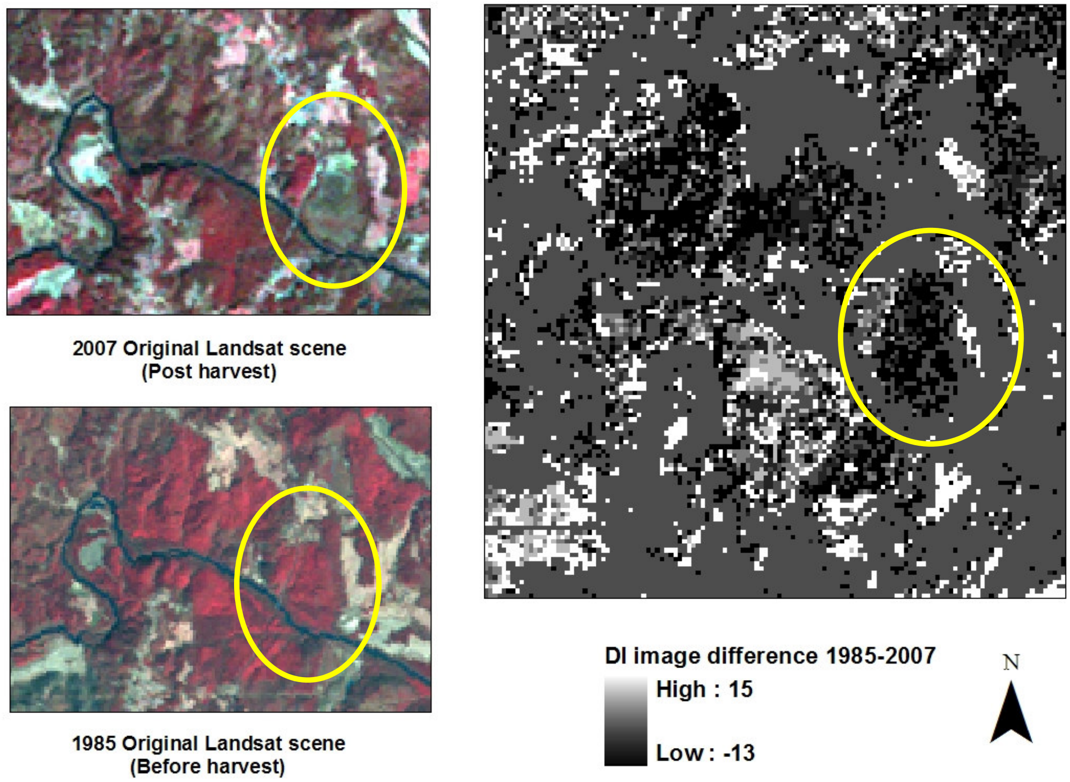


Figure 5. Raw 1985 and 2007 Landsat scenes in an area that was harvested within the 22 years interval, as well as the resulting DI difference image. The yellow circle on the bottom left is an uncut stand from the 1985 scene, the same stand is circled after being recently harvested in the 2007 scene. This same stand is identified at right within the disturbance index difference image. Notice the dark colors in the harvested areas which correspond to disturbance index values in the highly negative range.

A new lookup table raster had to be created for the layer in ArcGIS 9.2 using the Spatial Analyst – Reclass tool set before the classified image could be assessed for accuracy in Erdas Imagine. Once in Erdas Imagine 9.0, the image had to be converted to a thematic layer before accuracy assessment. This was accomplished by creating 15 areas of interest (AOI) and signatures for each class (harvest and no-harvest) and performing a maximum likelihood supervised classification. The AOIs were created by manually identifying stands in the original images that had complete harvests within the 22-year period. Research in clearcut identification with Landsat imagery has shown that increasing the number of training polygons over 15 does not significantly increase accuracy (Healey et al. 2005). In a separate study to determine degree of forest clearing, the disturbance index was found to be more strongly correlated to forest clearing than was the normalized difference vegetation index (Healey et al. 2006). A visual interpretation of the supervised classification image suggested that the classification represented the two change classes fairly well. The image was consequently recoded to represent only the two change classes for accuracy assessment.

The harvest, no-harvest recode image was validated using the approach outlined by Cohen et al. (1998). In this method, a sample of randomly generated points from the classified image is overlaid on the raw Landsat scenes and each point is visually interpreted as change or no-change. This method produced accuracies comparable to those found using independent reference data (Cohen et al. 1998). In our case, 60 equalized stratified random points, 30 for each class, were randomly generated using the accuracy assessment tool in Erdas Imagine 9.0. These points were manually assessed as harvested or not harvested using binary classes of 1 (harvest) and 2 (no-harvest) and were then used to assess the accuracy of the supervised classification. Table 4 shows the accuracy assessment outputs.

Table 4. Disturbance index harvest, no-harvest maximum likelihood supervised classification accuracy assessment results.

DI Difference Image	User's Accuracy	Kappa statistic
Harvest (Class 1)	81.82%	0.6364
No-Harvest (Class 2)	88.89%	0.7778

Percent sawtimber determination

Within our current methodology, we cannot directly predict percent sawtimber within a stand. Our goal in this regard was simply to model whether or not percent sawtimber was an acceptable SPB hazard predictor. FHTET has the capability to predict quadratic mean diameter (QMD) and therefore could determine percent sawtimber using their current modeling techniques.

Biophysical parameter estimates

FHTET mentioned to us in early methodology talks that they were interested in testing an improved digital elevation model (DEM) in the mountain region of North Carolina. The DEM they were using came from the National Elevation Dataset (NED) which was developed from many data sources, and they felt that their host layers could potentially be improved using a more accurate DEM. The accuracy for the NED varies based on the source data used for its development (http://ned.usgs.gov/Ned/accuracy.asp) and FHTET had expressed some concern in the mountain region about the accuracy of the original elevation mapping from which the NED was assembled (Jim Ellenwood and Frank Krist Jr., pers. comm., FHTET, May 15, 2007).

We determined that the best way to test the quality of a new DEM would be to compare the modeled estimates of biophysical parameters with each DEM separately. We used the LiDAR-derived DEM available from the North Carolina Department of Transportation (NC DOT) (www.ncdot.org/it/GIS) and the North Carolina Flood Mapping Program (www.ncfloodmaps.com). The LiDAR DEM was derived using a proprietary bare earth algorithm from the LiDAR data acquired by EarthData for the NC Floodplain Mapping Program. This LiDAR dataset, as reported in Hodgson et al. (2003), featured an average point density of 0.0183 points per square meter. Additional reports state that the LiDAR data were acquired during leaf-off periods (North Carolina Floodplain Mapping Program 2003) to improve the likelihood of achieving the vertical accuracy requirements of a root mean square error (RMSE) of 8 inches (20 cm) and 10 inches (25 cm) for coastal counties and inland counties, respectively (Thompson and Maune 2001).

The NC DOT offers the DEM, released in May of 2007, on a statewide level at a resolution of 80 feet (24.4 meters) and a county level resolution of 20 feet (6.1 meters), with all elevation measurements rounded to the nearest foot. The NC DOT website lists the vertical accuracy to be approximately 10 inches (25 centimeters) and the data are projected in NAD 83 North Carolina State Plane Feet.

We downloaded each county DEM at the 20-foot (6.1 meter) resolution and then mosaicked the counties back into a statewide DEM at a 20-foot (6.1 meter) resolution using ArcGIS Desktop 9.2 (ESRI. 1999-2006. Redlands, CA). The 20-foot (6.1 meter) statewide DEM was then visually assessed for edge matching. This DEM was shipped to the FHTET headquarters in Fort Collins, Colorado for biophysical parameter comparative analysis in North Carolina.

The LiDAR-derived DEM obtained from the NC DOT was sent to FHTET to compare the modeled biophysical parameters with those derived from the NED DEM. FHTET personnel input the NED and the LiDAR DEM terrain variables into their models to determine the DEM effects on total basal area for USGS Zones 57 and 59, shown in Figure 6. These two USGS map zones loosely correspond to the mountain and piedmont regions of North Carolina. The variables derived from the DEMs include slope, slope*sine(aspect), slope*cosine(aspect), slope position index, and the digital elevation model itself.

In order for the DEMs to have more emphasis in the modeling, the climate layer was not included in the modeling process. Additionally, FHTET determined the effect of the NED and LiDAR DEM terrain variables on two individual tree species, white pine and yellow birch. These individual species were assessed for changes in predicted basal area in the mountain region (USGS zone 57) where terrain difference is most dramatic, to see how individual species basal area was affected by the LiDAR DEM.

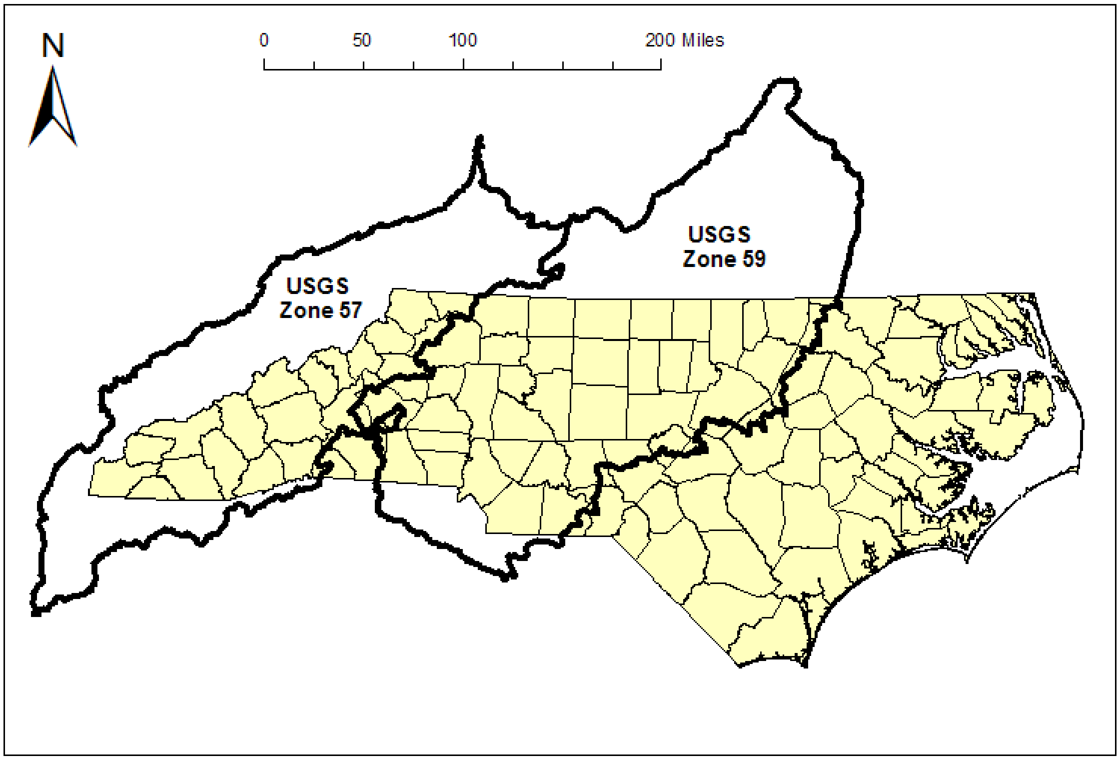


Figure 6. USGS Zones 57 and 59. These are the two zones for which FHTET ran biophysical parameter analyses using the LiDAR-derived DEM comparatively against the NED DEM.

Results and Discussion

SPB hazard modeling

The CART 5.0 software (Salford Systems 2003-2004, San Diego, CA) was able to successfully divide the datasets into incidence of SPB based solely on stand age and percent sawtimber in all three regions, though there are clearly other factors that influence SPB hazard. These outputs resulted in varying probabilities of SPB incidence based on specific age classes and/or percent sawtimber values. The CART output trees for these regions are shown as Appendices 12a, 13a, and 14a. Despite multiple attempts, we found that the CART output in the piedmont region was far too complex to be practical for determination of SPB hazard-related age classes and percent sawtimber values in the field (See Appendix 13). In order to retain an acceptable error percentage in the modeling, the CART output for the piedmont required far more divisions in the output tree than could be effectively interpreted to coincide with SPB incidence. We attribute the complexity of the CART results in the piedmont to other biophysical processes occurring in the piedmont which may also impact on SPB susceptibility, these may include coexistence of littleleaf disease (Belanger et al. 1986; Belanger et al. 1993); soil type and landuse history (Lorio Jr. and Sommers 1985; Doggett and Tweed 1994; Van Lear et al. 2004); or factors directly related to the current stand composition, including competition of understory vegetation, host tree radial growth, or elevated stand densities (Lorio Jr. and Hodges 1974b; Kushmaul et al. 1979; Hicks Jr. 1980). None of the afore mentioned variables were available for the CART analysis. There is clearly a complex of factors at work in the piedmont region of North Carolina as evidenced by its historically heavy SPB infestations (Doggett and Tweed 1994).

With respect to predictor variable importance, in the coastal plain and piedmont regions, stand age was the most important variable whereas the situation was opposite in the mountain region where percent sawtimber was more than twice as important as stand age. Variable importance is simply a measure of how influential each of the input variables is in determining where the splitting of the classification tree occurs (Salford Systems 2003-2004, San Diego, CA). CART was able to separate the data into SPB classes of Yes (outbreak) and No (no outbreak) with reasonable accuracy in all three regions. Tables 5, 6, and 7 list the misclassification percent error in the test samples as well as the overall prediction success as percent correctly classified from the CART analysis for all three regions, regardless of the impractical nature of the piedmont results. Tabular age classes, percent sawtimber values, and percent of the total regional dataset represented by each CART node, as identified by the CART analysis, are shown in Appendices 12b, 13b, and 14b, with visual representations presented in Figures 7, 8, and 9.

Table 5. Coastal plain CART® results showing total number of terminal nodes and relative accuracies.

Coastal Plain		
Total # Nodes		*7*
Test Sample	No	*23.76*
% Error	Yes	*27.72*
Predicted Success	No	*84.16*
% Correct	Yes	*72.28*

Table 6. Piedmont CART® results showing total number of terminal nodes and relative accuracies.

Piedmont		
Total # Nodes		*45*
Test Sample	No	*32.59*
% Error	Yes	*25*
Predicted Success	No	*78.17*
% Correct	Yes	*89.56*

Table 7. Mountain CART® results showing total number of terminal nodes and relative accuracies.

Mountain		
Total # Nodes		*13*
Test Sample	No	*34.95*
% Error	Yes	*29.13*
Predicted Success	No	*74.76*
% Correct	Yes	*88.35*

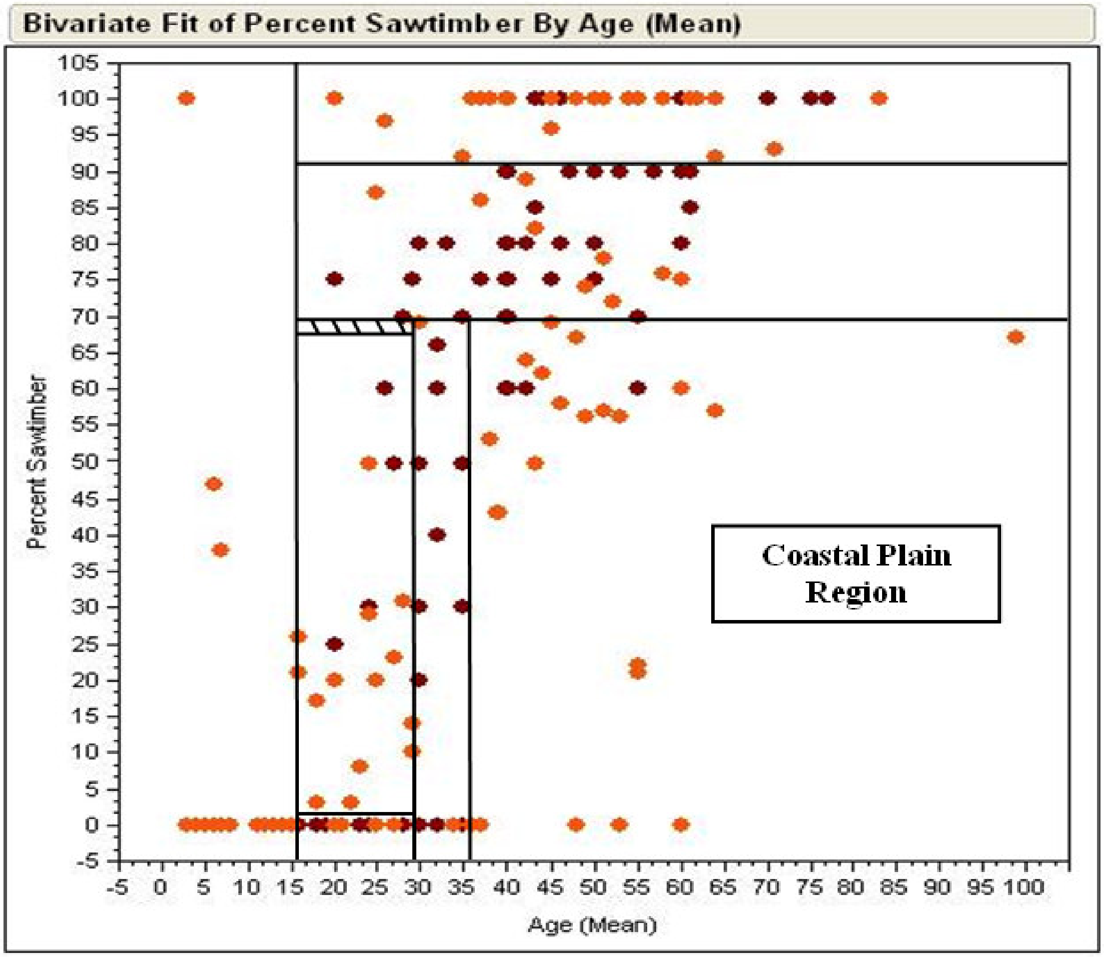


Figure 7. Coastal plain region scatterplot with SPB prediction classes. Maroon points represent N.C. Forest Service plots that were in SPB outbreak at the time of measurement and orange points represent plots not in outbreak during the known outbreak period of 1974. The black lines represent the age class and percent sawtimber divisions as identified by the CART® analysis.

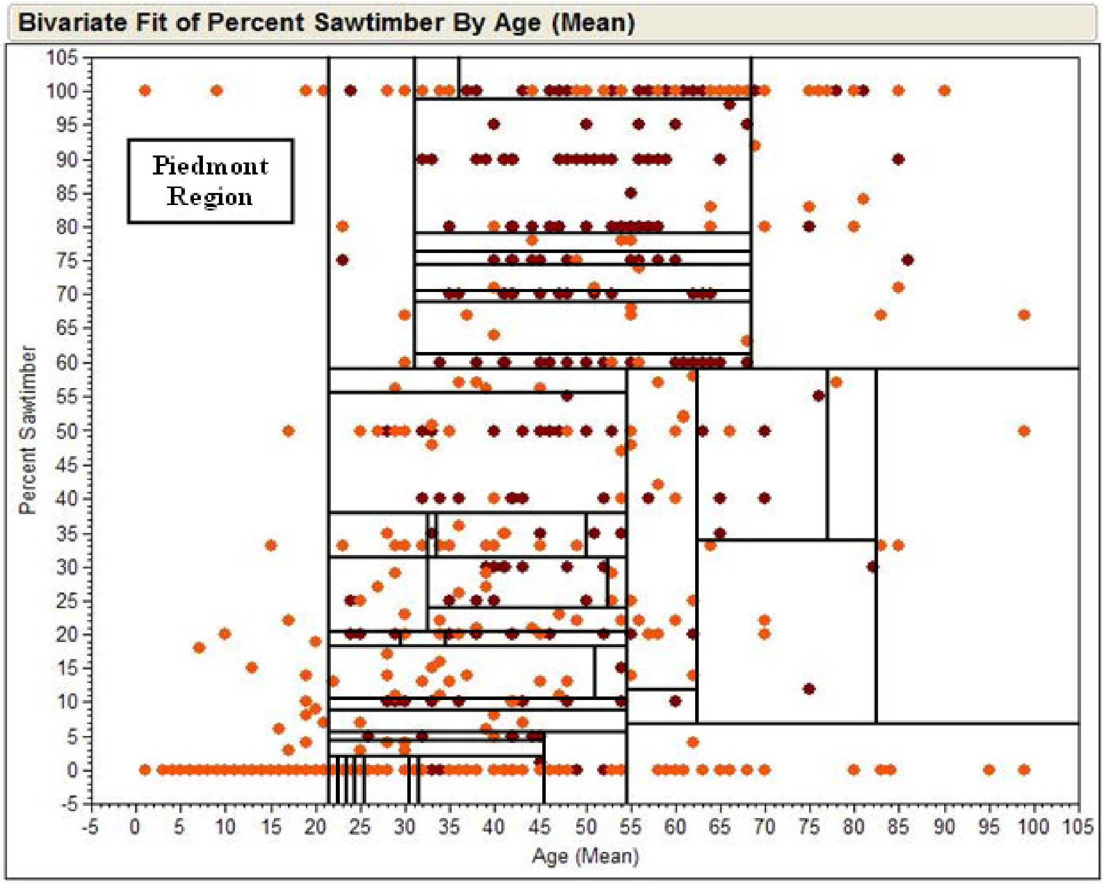


Figure 8. Piedmont region scatterplot with SPB prediction classes. Maroon points represent N.C. Forest Service plots that were in SPB outbreak at the time of measurement and orange points represent plots not in outbreak during the known outbreak period of 1974. The black lines represent the age class and percent sawtimber divisions as identified by the CART® analysis.

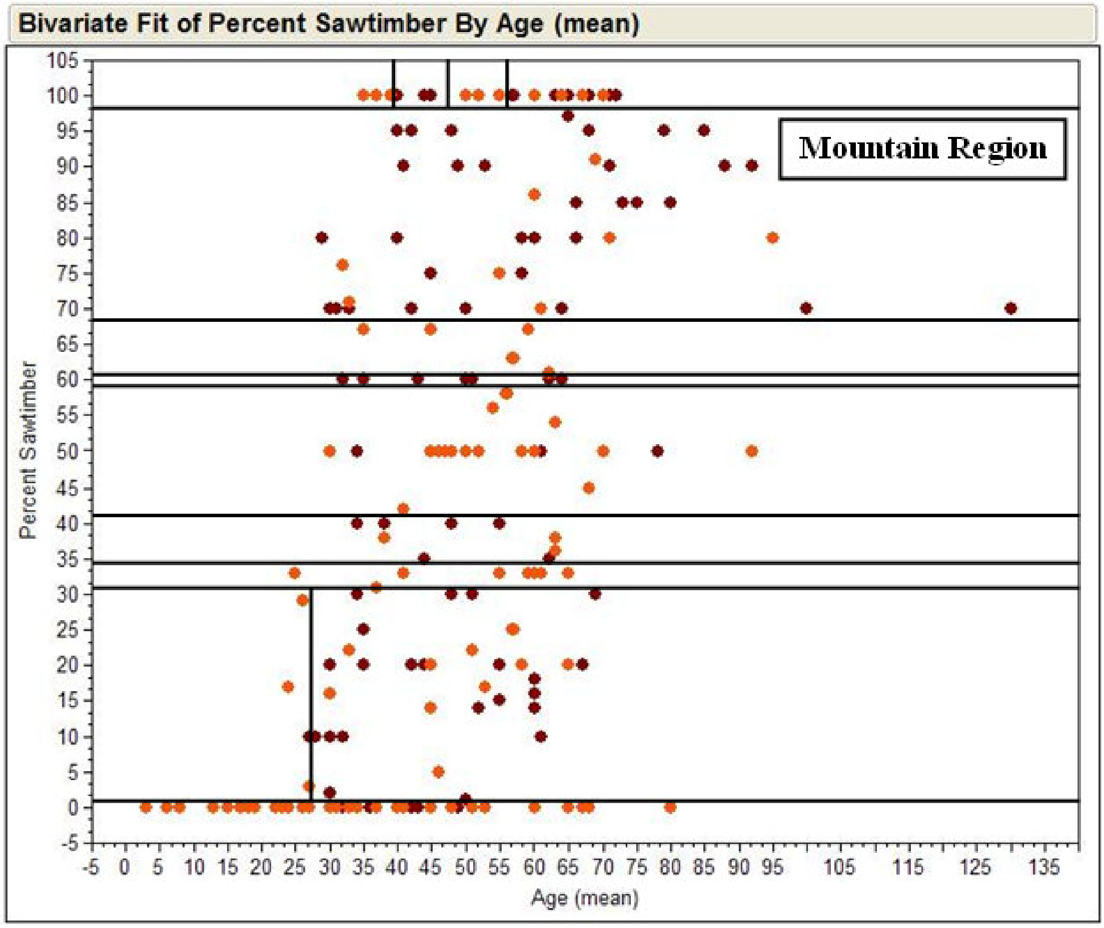


Figure 9. Mountain region scatterplot with SPB prediction classes. Maroon points represent N.C. Forest Service plots that were in SPB outbreak at the time of measurement and orange points represent plots not in outbreak during the known outbreak period of 1974. The black lines represent the age class and percent sawtimber divisions as identified by the CART® analysis.

We can visually see that in the coastal plain and mountain regions the CART output age classes and percent sawtimber classes can be identified with relative ease, while the piedmont lacks any observable pattern. We had to carefully interpret the nodes in each region to determine both if they made sense from a SPB hazard perspective, but more importantly if they made sense from a stand growth perspective. That is, the age class/percent sawtimber classes for all the regions show some unlikely SPB hazard designations, some relating to age classes, others to percent sawtimber, and others with both.

All regional outputs identified older stands with very little sawtimber as not being attacked. Our interpretation is this is more likely an issue of management practices and that in all likelihood private landowners generally don't let their trees grow much past financial maturity. If more landowners left their stands growing past financial maturity, when growth rates slow, these stands would likely show high incidence of SPB attack during an outbreak period (Kushmaul et al. 1979; Coulson et al. 1999).

There were some nodes in the piedmont which identified stands of very young trees with very high sawtimber values as not being attacked. We attribute this to possible errors in the data collection process. There were entries in both the FIA and the NCFS data where a young stand had sawtimber-sized trees listed in the plot tally and when there were few trees listed for the plot; as such the sawtimber-sized trees became very influential. It is not hard to imagine a relatively young stand, perhaps a seed tree prescription, in which there are scattered sawtimber sized trees in an otherwise young stand. Presumably these stands would in fact be less susceptible to attack.

Some nodes show the difference between attacked and not attacked to be less than 5 year age difference and less than 5 percent sawtimber difference. In some cases, particularly the piedmont, this created highly unrealistic age class and percent sawtimber cutoff values. We initially attributed this to the fact that percent sawtimber was rounded to the nearest 5 percent in the NCFS data, but we ran the CART process with the NCFS data rounded to the nearest 5 percent and found no significant differences in the outputs.

Despite the CART output issues mentioned above, the outcome of this analysis showed definite future predictive capabilities. The CART output trees give the strength of the prediction for each node by giving the percentage of the total data that each node represents. From this it would not be difficult to convert the age classes and percent sawtimber classes to the hazard designations outlined by the FHTET hazard modeling protocol (Krist Jr. et al. 2007). This conversion could, of course, not be completed until age classes and percent sawtimber can be accurately mapped. An improved dataset, one with spatial specificity, would help reinforce our findings that age and percent sawtimber are important factors for SPB hazard prediction in the North Carolina coastal plain and mountain regions, and may help simplify that of the piedmont region.

Age class determination

Converting Landsat 5 TM scenes to the tasseled cap and eventually to the disturbance index gave us two comparable scenes on which to perform our analysis. The results of the image differencing and harvest, no-harvest classification were very successful and we were able to produce a user's accuracy of 81.82% for the harvest class and 88.89% for the no-harvest class. The kappa statistics were 0.6364 for the harvest class and 0.7778 for the no-harvest class. The results from this analysis, and the resulting kappa statistics, could have been much more accurate had we not encountered the setbacks discussed below in our initial pine classification. Figure 10 shows the final designation of our two age classes, less than 22 years (shown in green) and greater than 22 years (shown in red).

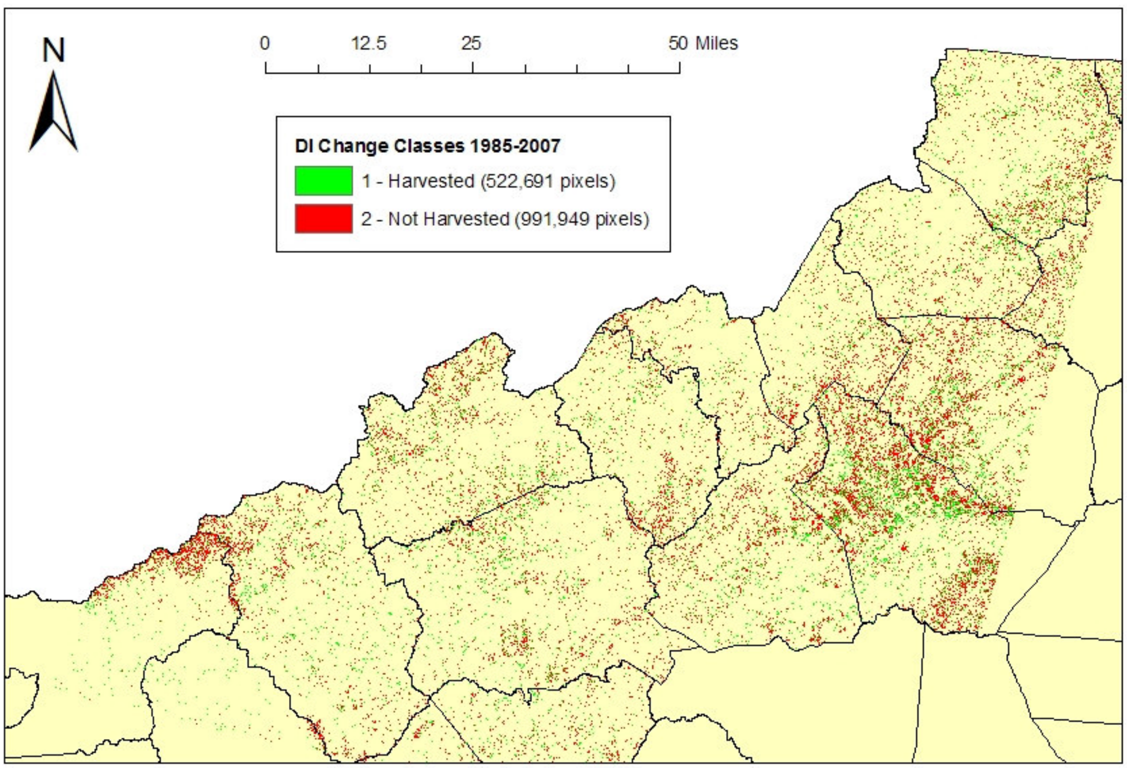


Figure 10. Disturbance Index difference image, reclassified to show stands harvested within the 22-year period (green) and stands not harvested within the 22-year period (red). In the course of this project, no field validation of this output was performed.

There were several issues that became apparent as we worked through the age class determination analysis. The biggest issue was that of the spectral separability in the initial pine identification supervised classification. For all intents and purposes, all the evergreen species looked the same spectrally within our Landsat scenes. The mountain region was a good example in this regard because there are several non-SPB host species, or only very rarely attacked host species, present. We used 1998 color-infrared digital orthophoto quarter quads to visually assess species composition when there was confusion in the classification process.

The mountain region is well known for its Fraser fir (*Abies fraseri* (Pursh) Poir.) Christmas tree industry and we were unable to consistently mask out the Christmas tree farms. Likewise, Fraser fir and red spruce (*Picea rubens* Sarg.) dominate the high elevation Appalachian ecosystem in North Carolina and we could not consistently separate these areas from the pine spectrally either. Lastly, the mountain region often has a midstory composed of various evergreen shrubs and these areas, generally riparian, are commonly associated with eastern hemlock (*Tsuga canadensis* (L.) Carr.) and eastern white pine (Hedman and Vanlear 1995). Even armed with the visual certainty of pixels that were entirely evergreen shrubs, we could not separate these pixels from the pine, nor could we distinguish between eastern hemlock and eastern white pine.

The unfortunate result is that we could not, despite using leaf-off scenes and multiple attempts, create a pine mask that did not include a seemingly large amount of non-host pixels. The accuracy assessment on this classification was based on manually entered points of known classes and did not include any pixels from classes that could not be definitively identified. So, while the accuracy was acceptable and the user's accuracies were maximized, the producer's accuracies were much less accurate as shown in Table 8. As previously mentioned, user's accuracy equates to designating, for example, an actual area of forest to the non-forest class in the classified image (Campbell 2006). Conversely, an example of producer's accuracy is the percentage of the actual forest on the ground that was correctly classified (Campbell 2006).

Table 8. Producer and user classification accuracies for the maximum likelihood supervised classification to determine the extent of pine in the Landsat scenes.

Scene Date	Producer's Accuracy (pine)	User's Accuracy (pine)
20 JAN 2007	67.50%	93.10%
08 FEB 1985	70.00%	100.00%

As mentioned in the CART analysis results, an improved dataset could alleviate the issues encountered with the pine mask classification. Most importantly, a spatially specific database could provide us with accurate land types across the region and provide a much improved layer for accuracy assessment.

Another notable issue we faced was that of older age classes. The first Landsat satellite was launched in 1972, so at the very best we could only predict specific age classes of 35 years or younger and would only be able to say that stands were 35 year old or older. In addition, actual pine stands harvested early in our 22-year interval may bias the age class results if they are completely reforested by the end of the interval and thus identified during accuracy assessment as unchanged. An option might be to divide larger time intervals into smaller additive intervals. If we had divided our 22-year interval into two smaller intervals we would be able to view the three resulting Disturbance index scenes in a RGB color composite stack as outlined by Healey et al. (2005) and have higher certainty we were only classifying scenes as being 22 years old that actually are 22 years old.

Our disturbance index cutoff values were visually determined by comparing the disturbance index values to the original scenes. We determined that -3 or less was an acceptable cutoff value and that it captured a number of noticeable complete harvests. However, we also visually identified some heavy partial harvests that were being classified in the harvest class. Attempts to make the cutoff value -4 resulted in the loss of several harvested pixels that should have been included, so the cutoff of -3 was used. In most cases of clearcuts that were reasonably recent, the initial pine masks did not classify these areas as pine at all; the spectral signatures likened these areas to the non-forest class. This resulted in errors in the disturbance index image for the 2007 Landsat scene due to recent clearcuts not being included in the pine mask and thus not being identified as disturbed pixels.

Our methodology should be applied in the future to determine young pine age classes in the coastal plain and piedmont regions to assess the success in these regions.

These regions may provide more accurate age class determination, as there are few other evergreen species with which to confuse the pine.

Biophysical parameter estimates

As mentioned in the methods section, the climate layer was not included in this modeling process. Regarding total basal area in the piedmont region (USGS zone 59), the NED DEM produced a R^2 of 0.67, and the LiDAR DEM produced a R^2 of 0.68. Total basal area in the mountain region (USGS zone 57) has similar results with R^2 values of 0.55 for the original DEM and 0.56 for the new DEM. Individual species basal area estimates also showed little difference between DEMs with white pine basal area R^2 values of 0.59 and 0.60 and yellow birch basal area R^2 values of 0.53 and 0.55 for the original and new DEMs respectively. The Cubist 2.05 model variable definitions and model outputs can be found in Appendices 15-23 and the individual species are identified by their FIA species codes, 129 for white pine and 371 for yellow birch.

Conclusion

SPB hazard modeling

Our study identified weaknesses in the SPB reporting system in place within the NCFS. This could potentially be an issue within other state forestry agencies as well. The NCFS forms for recording SPB outbreak data have hardly changed since the 1960s. At our suggestion, the SPB reporting system in North Carolina will be updated in the near future to reflect additional SPB outbreak tracking variables including basal area, average diameter at breast height, tree height, slope, aspect, growth rate, and soil type/drainage. As well, there is a need for the data to be spatially specific both for the benefit of the NCFS SPB record keeping and also for future accuracy assessment in the modeling process. These data, if updated regularly, could provide many more opportunities for hazard modeling research whether it be within the NCFS or at the federal level.

We recognize that age class and percent sawtimber represent only a fraction of the factors that affect a stand's susceptibility to SPB attack and may be highly correlated with variables already being included in hazard modeling efforts. Our goal with this project was to identify additional variables from actual SPB outbreak data that could influence a stand's susceptibility to SPB attack. Our modeling effort using CART produced valid predictive SPB models capable of use in the coastal plain and mountain regions using variables not currently included in the FHTET hazard modeling. Based on our modeling results, age class and percent sawtimber should be considered as additional FHTET modeling variables in the coastal plain and mountain regions respectively.

Age class determination

As mentioned in several areas, there is a great need for better data. Accurate host presence/absence will be instrumental in any future successes for a project of this type. Inaccurate delineation and identification of susceptible pine types can introduce large amounts of error by suggesting the hazard may be high in areas that are actually non-host.

Though it was not tested statistically, visual comparison of our pine host maps and of the FHTET SPB hazard map V1.0, both showed the target host present in an area that was clearly identifiable in aerial photographs as being high elevation spruce-fir forest. This may not pose as much of an issue in the piedmont and coastal plain regions where there are less non-pine evergreen species. With accurate host type mapping, our method for aging stands could be applied to a variety of forestry objectives, both related to forest health and also general forest management.

The DI calculations produced acceptable results, despite the long interval length 22 years. Using such lengthy acquisition intervals could present problems in areas that regenerate quickly, such as the North Carolina coastal plain. If an area was stable at the time of the first acquisition, cut shortly thereafter, and had regenerated completely within the 22-year period, it may appear to have been unchanged across the interval and thus classified in the incorrect age class. With the upcoming release of all Landsat data to the public, we suggest using shorter intervals to ensure proper age class determination. We have shown that stand age classes can be reasonably identified using Landsat or other multispectral imagery.

Biophysical parameter estimates

FHTET remains confident that the new LiDAR-derived DEM will improve their modeled host maps, despite the minimal differences reported earlier. They are currently in the process of updating their climate layer using the new DEM and are also updating plot data to help improve the results. In the past, they have noted increased R^2 values in their results with the updated plot data. While these updates are not incorporated in the current effort, they will be included in the future (Frank Krist Jr., pers. comm., FHTET, May 29, 2008).

Literature Cited

Anderson, R.F., and C.A. Doggett. 1993. Host preference of southern pine beetle in North Carolina. N.C. Division of Forest Resources - Forestry Note No. 66.

Belanger, R.P., R.L. Hedden, and P.L. Lorio, Jr. 1993. Management strategies to reduce losses from the southern pine beetle. Southern Journal of Applied Forestry 17(3):150-154.

Belanger, R.P., R.L. Hedden, and F.H. Tainter. 1986. Managing Piedmont forests to reduce losses from the littleleaf disease-southern pine beetle complex. P. 19 pp. USDA Forest Service Agriculture Handbook 649.

Belanger, R.P., E.A. Osgood, and G.E. Hatchell. 1979. Stand, soil, and site characteristics associated with southern pine beetle infestations in the southern Appalachians. USDA Forest Service Research Paper SE-198.

Brown, M.J. 2004. Forest statistics for North Carolina 2002. P. 78. USDA Forest Service Resource Bulletin SRS-088.

Campbell, J.B. 2006. Introduction to Remote Sensing. Guilford Press. 626 p.

Chander, G., and B. Markham. 2003. Revised Landsat-5 TM radiometric calibration procedures and postcalibration dynamic ranges. IEEE Transactions on Geoscience and Remote Sensing 41(11):2674-2677.

Chander, G., B.L. Markham, and J.A. Barsi. 2007. Revised Landsat-5 Thematic Mapper radiometric calibration. IEEE Geoscience and Remote Sensing Letters 4(3):490-494.

Clarke, S.R., and J.T. Nowak. 2008. Southern pine beetle. P. 7. USDA Forest Service Forest Insect and Disease Leaflet.

Cohen, W.B., M. Fiorella, J. Gray, E. Helmer, and K. Anderson. 1998. An efficient and accurate method for mapping forest clearcuts in the Pacific Northwest using Landsat imagery. Photogrammetric Engineering and Remote Sensing 64(4):293-300.

Cohen, W.B., T.A. Spies, and M. Fiorella. 1995. Estimating the age and structure of forests in a multi-ownership landscape of western Oregon, USA. International Journal of Remote Sensing 16(4):721-746.

Coulson, R.N., B.A. McFadden, P.E. Pulley, C.N. Lovelady, J.W. Fitzgerald, and S.B. Jack. 1999. Heterogeneity of forest landscapes and the distribution and abundance of the southern pine beetle. Forest Ecology and Management 114(2-3):471-485.

Council, B. Jan. 11, 2008. [Personal communication]. N.C. Forest Service.

Doggett, C.A., and D.R. Tweed. 1994. Geographical intensity of southern pine beetle infestations. Southern Journal of Applied Forestry 18(4):145-146.

Ellenwood, J., and F. Krist Jr. May 15, 2007. [Personal communication]. USDA Forest Service - Forest Health Technology Enterprise Team.

Ellenwood, J., and F.J. Krist Jr. 2007. Building a nationwide 30-meter forest parameter dataset for forest health risk assessments. in ForestSat. Cemagref - UMR TETIS, Montpellier, France.

Ellenwood, J., F.J. Krist Jr., F.J. Sapio, and B.M. Tkacz. *in press*. Assessing forest health risk: Building regional datasets from national datasets. USDA Forest Service FHTET.

ESRI. 1999-2006. ArcGIS Desktop 9.2. [Computer software], Redlands, CA.

Flamm, R.O., R.N. Coulson, and T.L. Payne. 1988. The southern pine beetle. In Dynamics of Forest Insect Populations, Berryman, A.A. (ed.). Plenum Press, New York.

Gottschalk, K. 1995. Using silviculture to improve health in northeastern conifer and eastern hardwood forests. P. 219-226 in 1995 National Silviculture Workshop, Eskew, L.G. (ed.). USDA Forest Service RM-GTR-267, Mescalero, New Mexico.

Hayes, D.J., and S.A. Sader. 2001. Comparison of change-detection techniques for monitoring tropical forest clearing and vegetation regrowth in a time series. Photogrammetric Engineering and Remote Sensing 67(9):1067-1075.

Healey, S.P., W.B. Cohen, Y. Zhiqiang, and O.N. Krankina. 2005. Comparison of tasseled cap-based landsat data structures for use in forest disturbance detection. Remote Sensing of Environment 97(3):301 – 310.

Healey, S.P., Y. Zhiqiang, W.B. Cohen, and D.J. Pierce. 2006. Application of two regression-based methods to estimate the effects of partial harvest on forest structure using Landsat data. Remote Sensing of Environment 101(1):115–126.

Hedman, C.W., and D.H. Vanlear. 1995. Vegetative structure and composition of southern Appalachian riparian forests. Bulletin of the Torrey Botanical Club 122(2):134-144.

Hicks Jr., R.R. 1980. Climate, site, and stand factors. P. 55-68 In The Southern Pine Beetle, Thatcher, R.C., J.L. Searcy, J.E. Coster, and G.D. Hertel (eds.). USDA Forest Service Technical Bulletin 1631, Pineville, LA.

Hodgson, M.E., J.R. Jensen, L. Schmidt, S. Schill, and B. Davis. 2003. An evaluation of LIDAR- and IFSAR-derived digital elevation models in leaf-on conditions with USGS Level 1 and Level 2 DEMs. Remote Sensing of Environment 84(2):295-308.

Karpinski, C., D.L. Ham, and R.L. Hedden. 1997. Predicting potential loss from southern pine beetle in the piedmont of South Carolina. Clemson University Extension Forestry Leaflet No. 11.

Krist Jr., F. May 29, 2008. [Personal communication]. USDA Forest Service - Forest Health Technology Enterprise Team

Krist Jr., F.J., F.J. Sapio, and B.M. Tkacz. 2007. Mapping risk from insects and diseases. USDA Forest Service - FHTET.

Kushmaul, R.J., M.D. Cain, C.E. Rowell, and R.L. Porterfield. 1979. Stand and site conditions related to southern pine beetle susceptibility. Forest Science 25(4):656-664.

Leica Geosystems Geospatial Imaging LLC. 2006. Erdas Imagine 9.0. [Computer software], Norcross, GA.

Lindbladh, M., G.L. Jacobson, and M. Schauffler. 2003. The postglacial history of three Picea species in New England, USA. Quaternary Research 59(1):61-69.

Lindbladh, M.S., R. O'Connor, and G.L. Jacobson. 2002. Morphometric analysis of pollen grains for paleoecological studies: Classification of Picea from eastern North America. American Journal of Botany 89(9):1459-1467.

Lorio Jr., P. 1978. Developing stand risk classes for the southern pine beetle. USDA Forest Service Research Paper SO-144.

Lorio Jr., P.L., and J.D. Hodges. 1974b. Host and site factors in southern pine beetle infestations. P. 32-34 *In* Southern pine beetle symposium, Payne, T.L., R.N. Coulson, and R.C. Thatcher (eds.). College Station, Texas, Texas A&M University Memorial Student Center.

Lorio Jr., P.L., and R.A. Sommers. 1985. Potential use of soil maps to estimate southern pine beetle risk. P. 239-245. USDA Forest Service GTR-SO-56.

Lorio, P.L., Jr., and R.A. Sommers. 1981. Use of available resource data to rate stands for southern pine beetle risk. General Technical Report, USDA Forest Service, Washington, DC(No. WO-27):75-78.

Microsoft Corporation. 2007. Microsoft Access 12. [Computer software], Seattle, WA.

North Carolina Floodplain Mapping Program. 2003. LiDAR and digital elevation data fact sheet. P. 6.

Nowak, J., C. Asaro, K. Klepzig, and R. Billings. *in press*. The southern pine beetle prevention initiative: Working for healthier forests. Journal of Forestry.

Payne, T.L. 1980. Life History and Habits. In The Southern Pine Beetle Thatcher, R.C., J.L. Searcy, J.E. Coster, and G.D. Hertel (eds.). USDA Forest Service Technical Bulletin 1631, Pineville, LA.

RuleQuest Research. 2007. Cubist 2.05. [Computer software], Sydney, Australia.

Salford Systems. 2002. CART for Windows - User's Guide. San Diego, CA.

Salford Systems. 2003-2004. CART 5.0. [Computer software], San Diego, CA.

Shifley, S.R., Z.F. Fan, J.M. Kabrick, and R.G. Jensen. 2006. Oak mortality risk factors and mortality estimation. Forest Ecology and Management 229(1-3):16-26.

St. George, R.A., and J.A. Beal. 1929. The southern pine beetle: a serious enemy of pines in the south. P. 18. USDA Farmers' Bulletin 1188.

Thompson, G., and D. Maune. 2001. Quality control of LiDAR elevation data in North Carolina: North Carolina Cooperating Technical State Mapping Program Issue 5.

Van Lear, D.H., R.A. Harper, P.R. Kapeluck, and W.D. Carroll. 2004. History of piedmont forests: Implications for current pine management. P. 127-131. USDA Forest Service GTR-SRS-71.

Veysey, J.S., M.P. Ayres, M.J. Lombardero, R.W. Hofstetter, and K. Klepzig. 2003. Relative suitability of Virginia pine and loblolly pine as host species for *Dendroctonus frontalis* (Coleoptera: Scolytidae). Environmental Entomology(32):668-679.

Vite, J.P., S.F. Islas, J.A.A. Renwick, P.R. Hughes, and R.A. Kliefoth. 1974. Biochemical and biological variation of southern pine beetle populations in North and Central America. Zeitschrift fur Angewandte Entomologie 75(4):422-435.

Wilson, E.H., and S.A. Sader. 2002. Detection of forest harvest type using multiple dates of Landsat TM imagery. Remote Sensing of Environment 80(3):385-396.

Wulder, M.A., R.S. Skakun, W.A. Kurz, and J.C. White. 2004. Estimating time since forest harvest using segmented Landsat ETM+ imagery. Remote Sensing of Environment 93(1-2):179-187.

Ylioja, T., D.H. Slone, and M.P. Ayres. 2005. Mismatch between herbivore behavior and demographics contributes to scale-dependence of host susceptibility in two pine species. Forest Science 51(6):522-531.

Appendix 1. FHTET SPB hazard model worksheet for the northern coastal plain, containing weighted criterion chosen as SPB outbreak factors in this location. (http://www.fs.fed.us/foresthealth/technology/nidrm_spb.shtml)

Risk Model Worksheet - South

Risk Agent(s):	Southern Pine Beetle	**Host(s):**	Pitch, Loblolly, Virgina, Pond
Model Extent:	Parts of: AR, DE, MD, NC, NJ, OK, VA	**Max Percent Mortality:**	85%

Susceptibility

Rank/Weight		Criterion	Risk Begins (a)	Risk Peaks (b)	Risk Decreases	Risk Ends (d)	Curve	Rank	Weight
	0%								
Criteria 1									
Criteria 2									
Criteria 3									
Criteria 4									
Criteria 5									
Criteria 6									
Criteria 7									
Criteria 8									
Criteria 9									
Criteria 10									

Vulnerability

Rank/Weight		Criterion	Risk Begins (a)	Risk Peaks (b)	Risk Decreases	Risk Ends (d)	Curve	Rank	Weight
1	100%								
Criteria 1		Stand Density Index (SDI)	80	270	270	270	Linear	1	75%
Criteria 2		Soil Drainage Index (Soil Moisture)	54	85	85	85	Linear	1/3	25%
Criteria 3									
Criteria 4									
Criteria 5									
Criteria 6									
Criteria 7									
Criteria 8									
Criteria 9									
Criteria 10									

Constraints	Host must be present.	**Comments**	
Citations		**Model Certainity**	4 - Expert Opinion

Appendix 2. FHTET SPB hazard model worksheet for the central coastal plain, containing weighted criterion chosen as SPB outbreak factors in this location. (http://www.fs.fed.us/foresthealth/technology/nidrm_spb.shtml)

Risk Model Worksheet - South

Risk Agent(s):	Southern Pine Beetle	Host(s):	Shortleaf, Loblolly, Virgina, Spruce, Pond
Model Extent:	Parts of: Eastern NC	Max Percent Mortality:	85%

Susceptibility

Rank/Weight		Criterion	Risk Begins (a)	Risk Peaks (b)	Risk Decreases	Risk Ends (d)	Curve	Rank	Weight
	0%								
Criteria 1									
Criteria 2									
Criteria 3									
Criteria 4									
Criteria 5									
Criteria 6									
Criteria 7									
Criteria 8									
Criteria 9									
Criteria 10									

Vulnerability

Rank/Weight		Criterion	Risk Begins (a)	Risk Peaks (b)	Risk Decreases	Risk Ends (d)	Curve	Rank	Weight
1	100%								
Criteria 1		Stand Density Index	80	270	270	270	Linear	1	100%
Criteria 2									
Criteria 3									
Criteria 4									
Criteria 5									
Criteria 6									
Criteria 7									
Criteria 8									
Criteria 9									
Criteria 10									

Constraints	Host must be present.	Comments	
Citations		Model Certainity	4 - Expert Opinion

Appendix 3. FHTET SPB hazard model worksheet for the southern coastal plain, containing weighted criterion chosen as SPB outbreak factors in this location. (http://www.fs.fed.us/foresthealth/technology/nidrm_spb.shtml)

Risk Model Worksheet - South

Risk Agent(s): Southern Pine Beetle

Host(s): Shortleaf, Loblolly, Virgina, Spruce, Pond

Model Extent: Parts of: NC Only

Max Percent Mortality: 85%

Susceptibility

Rank/Weight		Criterion	Risk Begins (a)	Risk Peaks (b)	Risk Decreases	Risk Ends (d)	Curve	Rank	Weight
	0%								
Criteria 1									
Criteria 2									
Criteria 3									
Criteria 4									
Criteria 5									
Criteria 6									
Criteria 7									
Criteria 8									
Criteria 9									
Criteria 10									

Vulnerability

Rank/Weight		Criterion	Risk Begins (a)	Risk Peaks (b)	Risk Decreases	Risk Ends (d)	Curve	Rank	Weight
1	100%								
Criteria 1		Stand Density Index (SDI)	50	120	120	120	Linear	1	75%
Criteria 2		Soil Drainage Index (Soil Moisture Wet)	54	85	85	85	Linear	1/3	25%
Criteria 3									
Criteria 4									
Criteria 5									
Criteria 6									
Criteria 7									
Criteria 8									
Criteria 9									
Criteria 10									

Constraints: Host must be present

Comments:

Citations:

Model Certainity: 4 - Expert Opinion

Appendix 4. FHTET SPB hazard model worksheet for the piedmont region, containing weighted criterion chosen as SPB outbreak factors in this location. (http://www.fs.fed.us/foresthealth/technology/nidrm_spb.shtml)

Risk Model Worksheet - South

Risk Agent(s): Southern Pine Beetle

Host(s): Shortleaf, Loblolly, Virgina, Pitch Pines

Model Extent: Parts of: AL, GA, KY, NC, SC, TN, VA

Max Percent Mortality: 85%

Susceptibility

Rank/Weight		Criterion	Risk Begins (a)	Risk Peaks (b)	Risk Decreases	Risk Ends (d)	Curve	Rank	Weight
1	50%								
Criteria 1		Pct. Host	5%	50%	50%	50%	Linear	1	25%
Criteria 2		Slope	3%	10%	10%	10%	Linear	1	25%
Criteria 3		Aspect (Degrees)	135	180	270	310	Linear	1	25%
Criteria 4		Pct. Clay	10	32	32	32	Linear	1	25%
Criteria 5									
Criteria 6									
Criteria 7									
Criteria 8									
Criteria 9									
Criteria 10									

Vulnerability

Rank/Weight		Criterion	Risk Begins (a)	Risk Peaks (b)	Risk Decreases	Risk Ends (d)	Curve	Rank	Weight
1	50%								
Criteria 1		Host BA	60	120	120	120	Linear	1	100%
Criteria 2									
Criteria 3									
Criteria 4									
Criteria 5									
Criteria 6									
Criteria 7									
Criteria 8									
Criteria 9									
Criteria 10									

Constraints: Host must be present.

Comments:

Citations: 1,2,3,4

Model Certainity: 4 - Expert Opinion

Appendix 5. FHTET SPB hazard model hazard for the mountain region, containing weighted criterion chosen as SPB outbreak factors in this location. (http://www.fs.fed.us/foresthealth/technology/nidrm_spb.shtml)

Risk Model Worksheet - South

Risk Agent(s):	Southern Pine Beetle
Host(s):	Shortleaf, Loblolly, Virgina, Pitch Pines
Model Extent:	Parts of:Al, GA, KY, NC, SC, TN, VA, WV
Max Percent Mortality:	85%

Susceptibility

Rank/Weight		Criterion	Risk Begins (a)	Risk Peaks (b)	Risk Decreases	Risk Ends (d)	Curve	Rank	Weight
1	50%								
Criteria 1		Pct. Host	5%	60%	60%	60%	Linear	1	33%
Criteria 2		Slope	3%	10%	10%	10%	Linear	1	33%
Criteria 3		Shadow Effect	Some	Heavy	Heavy	Heavy	Linear	1	33%
Criteria 4									
Criteria 5									
Criteria 6									
Criteria 7									
Criteria 8									
Criteria 9									
Criteria 10									

Vulnerability

Rank/Weight		Criterion	Risk Begins (a)	Risk Peaks (b)	Risk Decreases	Risk Ends (d)	Curve	Rank	Weight
1	50%								
Criteria 1		Host BA	5	60	60	60	Linear	1	100%
Criteria 2									
Criteria 3									
Criteria 4									
Criteria 5									
Criteria 6									
Criteria 7									
Criteria 8									
Criteria 9									
Criteria 10									

Constraints	Host must be present.
Comments	Shadow Effect was calculated within ArcGIS using a 30-meter DEM, average sun angle based on latitude. Host BA lowered, host maps do not show heavy enough BA.
Citations	1,3,4
Model Certainity	4 - Expert Opinion

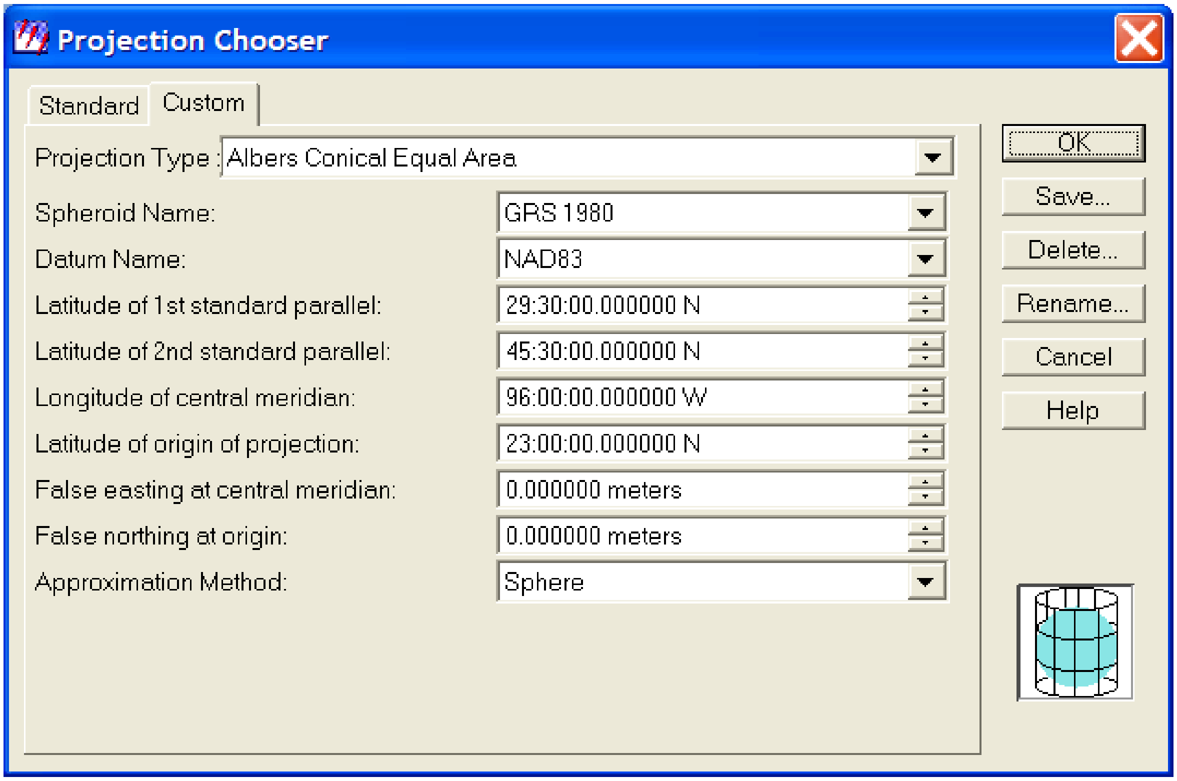
Appendix 6. North Carolina counties shown by FIA region (compiled from the FIA Handbook). These regional designations were the basis for our SPB hazard modeling.

County	FIA Region
BEAUFORT	Coastal Plain
BERTIE	Coastal Plain
BLADEN	Coastal Plain
BRUNSWICK	Coastal Plain
CAMDEN	Coastal Plain
CARTERET	Coastal Plain
CHOWAN	Coastal Plain
COLUMBUS	Coastal Plain
CRAVEN	Coastal Plain
CUMBERLAND	Coastal Plain
CURRITUCK	Coastal Plain
DARE	Coastal Plain
DUPLIN	Coastal Plain
EDGECOMBE	Coastal Plain
GATES	Coastal Plain
GREENE	Coastal Plain
HALIFAX	Coastal Plain
HARNETT	Coastal Plain
HERTFORD	Coastal Plain
HOKE	Coastal Plain
HYDE	Coastal Plain
JOHNSTON	Coastal Plain
JONES	Coastal Plain
LEE	Coastal Plain
LENOIR	Coastal Plain
MARTIN	Coastal Plain
MOORE	Coastal Plain
NASH	Coastal Plain
NEW HANOVER	Coastal Plain
NORTHAMPTON	Coastal Plain
ONSLOW	Coastal Plain
PAMLICO	Coastal Plain
PASQUOTANK	Coastal Plain
PENDER	Coastal Plain
PERQUIMANS	Coastal Plain
PITT	Coastal Plain
RICHMOND	Coastal Plain
ROBESON	Coastal Plain
SAMPSON	Coastal Plain

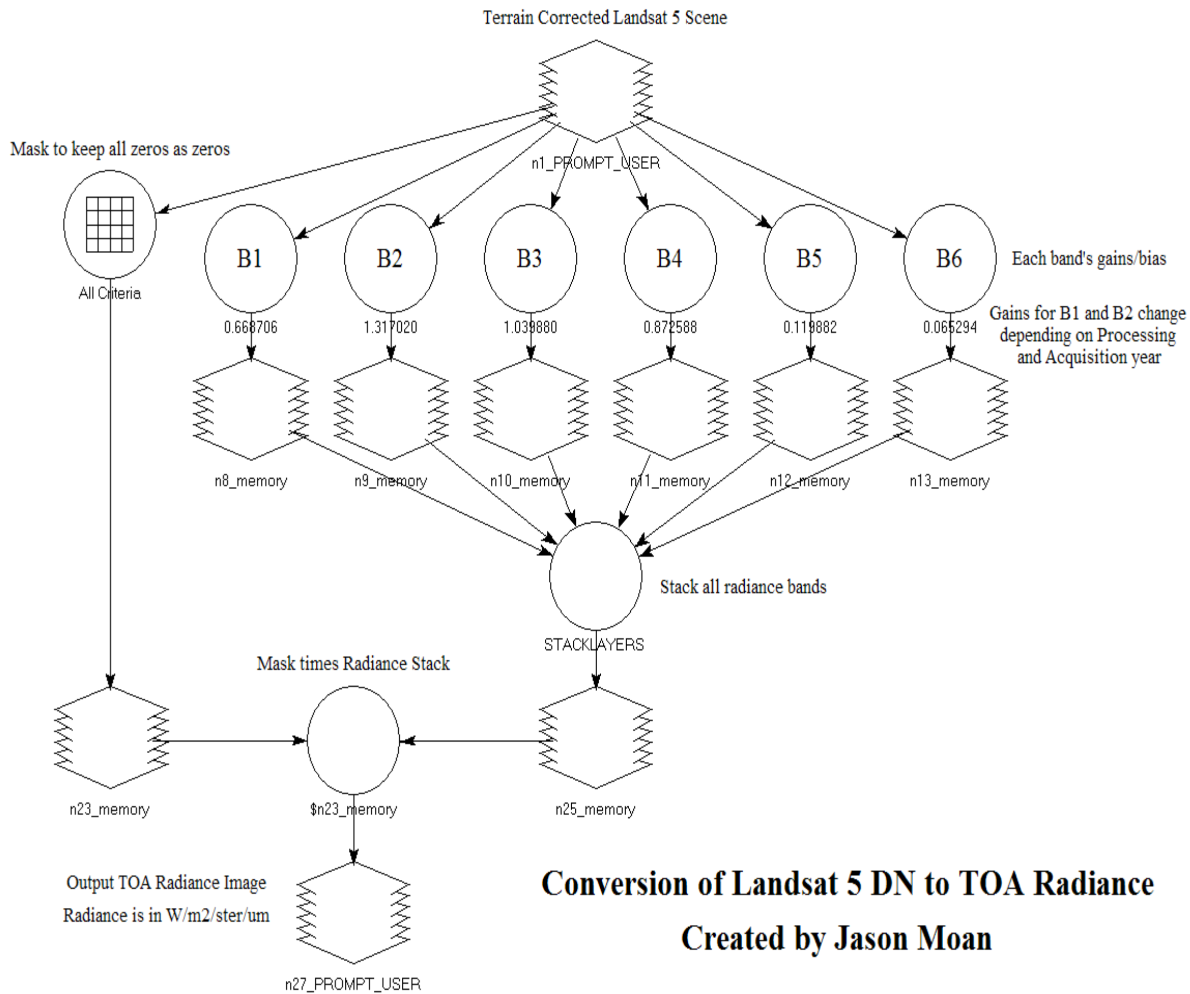
County	FIA Region
SCOTLAND	Coastal Plain
TYRRELL	Coastal Plain
WASHINGTON	Coastal Plain
WAYNE	Coastal Plain
WILSON	Coastal Plain
ALAMANCE	Piedmont
ALEXANDER	Piedmont
ANSON	Piedmont
CABARRUS	Piedmont
CASWELL	Piedmont
CATAWBA	Piedmont
CHATHAM	Piedmont
CLEVELAND	Piedmont
DAVIDSON	Piedmont
DAVIE	Piedmont
DURHAM	Piedmont
FORSYTH	Piedmont
FRANKLIN	Piedmont
GASTON	Piedmont
GRANVILLE	Piedmont
GUILFORD	Piedmont
IREDELL	Piedmont
LINCOLN	Piedmont
MECKLENBURG	Piedmont
MONTGOMERY	Piedmont
ORANGE	Piedmont
PERSON	Piedmont
POLK	Piedmont
RANDOLPH	Piedmont
ROCKINGHAM	Piedmont
ROWAN	Piedmont
RUTHERFORD	Piedmont
STANLY	Piedmont
STOKES	Piedmont
SURRY	Piedmont
UNION	Piedmont
VANCE	Piedmont
WAKE	Piedmont
WARREN	Piedmont
YADKIN	Piedmont
ALLEGHANY	Mountain
ASHE	Mountain

County	FIA Region
AVERY	Mountain
BUNCOMBE	Mountain
BURKE	Mountain
CALDWELL	Mountain
CHEROKEE	Mountain
CLAY	Mountain
GRAHAM	Mountain
HAYWOOD	Mountain
HENDERSON	Mountain
JACKSON	Mountain
MACON	Mountain
MADISON	Mountain
MCDOWELL	Mountain
MITCHELL	Mountain
SWAIN	Mountain
TRANSYLVANIA	Mountain
WATAUGA	Mountain
WILKES	Mountain
YANCEY	Mountain

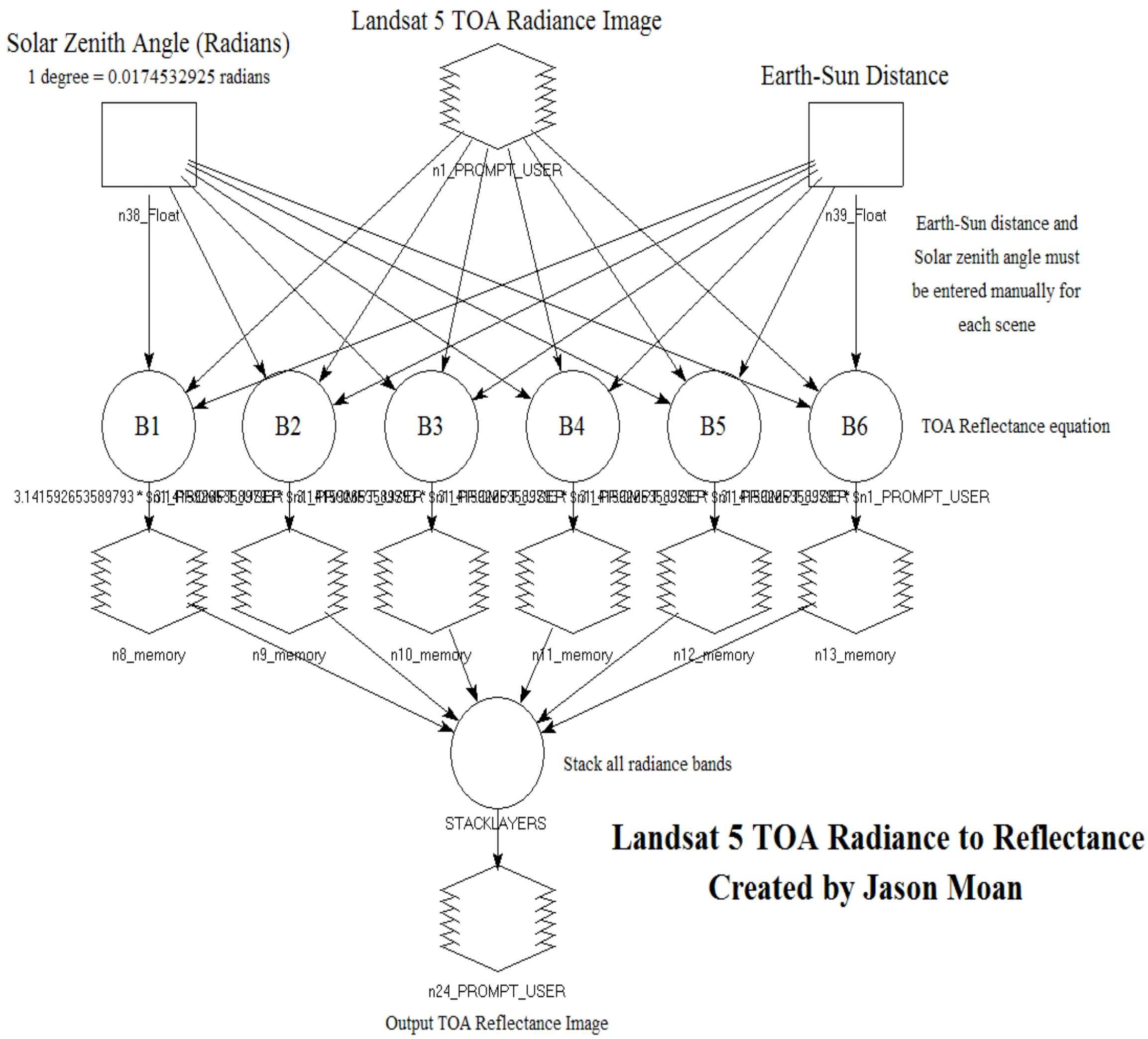
Appendix 7. Creation of Albers Conical Equal Area USGS-Version Projection in Erdas Imagine V9.0. This projection is not a standard projection included with Erdas Imagine V9.0.



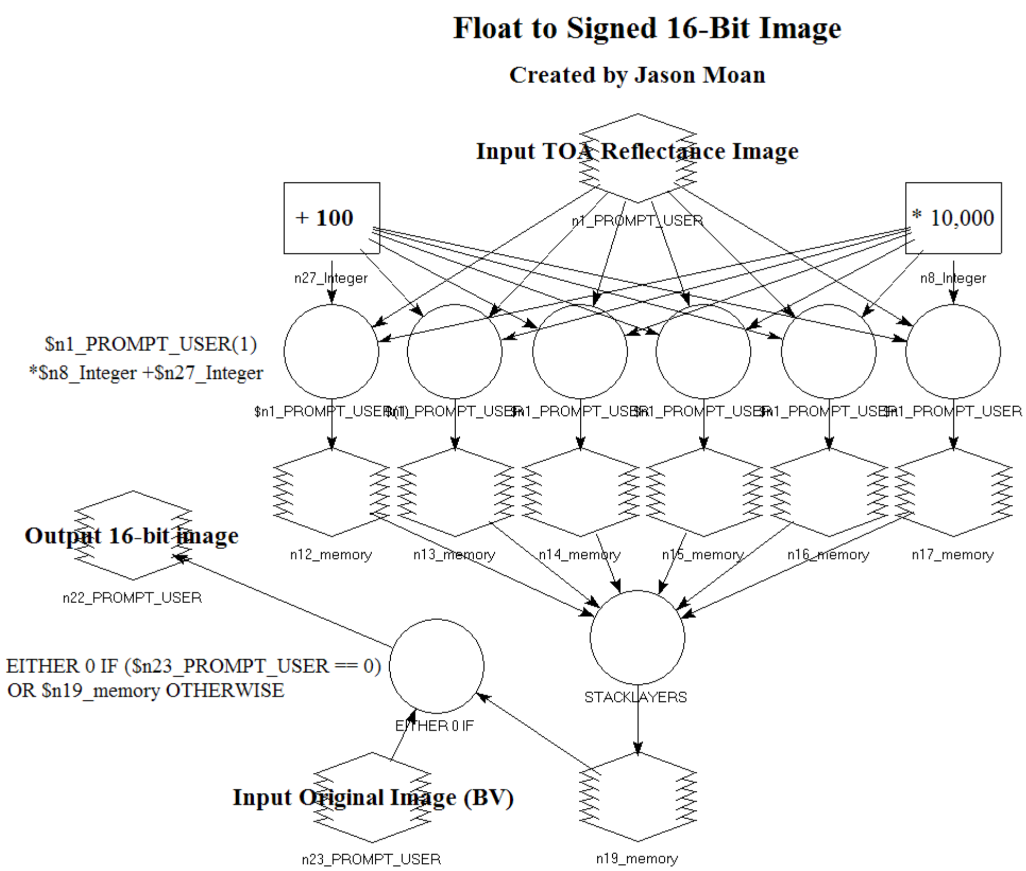
Appendix 8. Landsat 5 digital number to top of atmosphere reflectance model (Part 1) using coefficients outlined in Chander et al. (2007) and equations from Chander and Markham (2003) for an image acquired in January 2007 and processed after April 2, 2007.



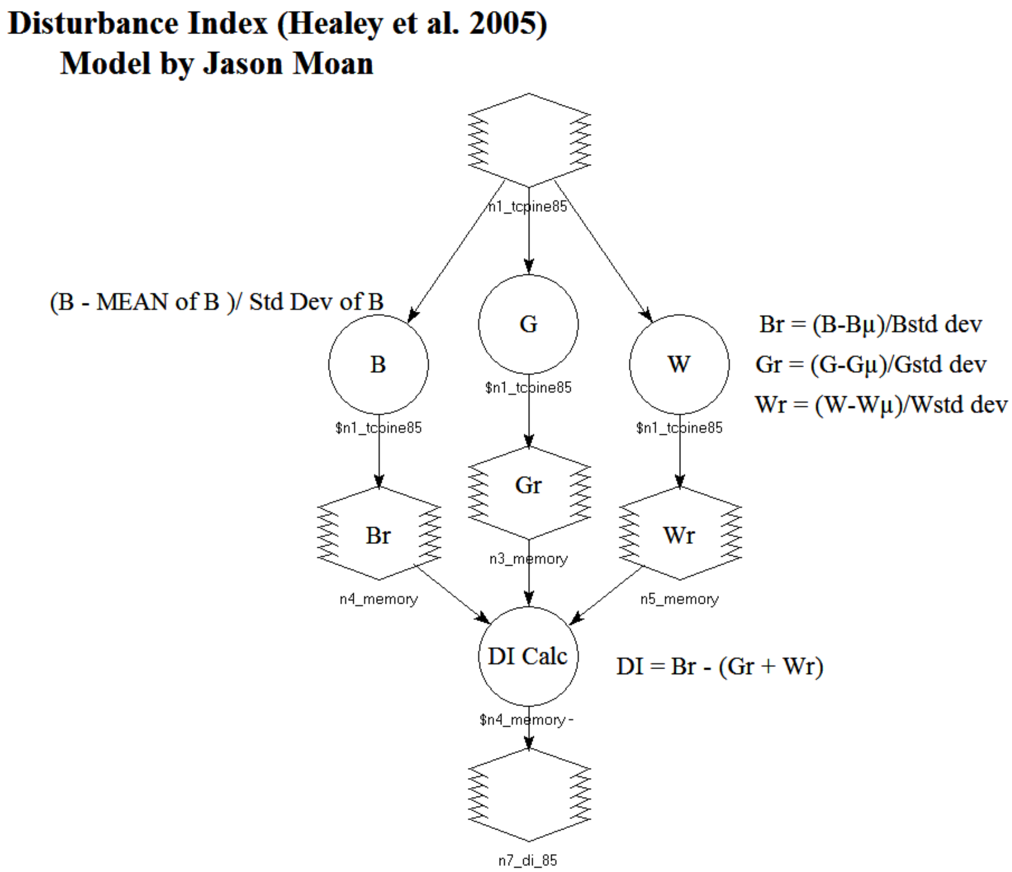
Appendix 9. Landsat 5 digital number to top of atmosphere reflectance model (Part 2) using coefficients outlined in Chander et al. (2007) and equations from Chander and Markham (2003) for an image acquired in January 2007 and processed after April 2, 2007.



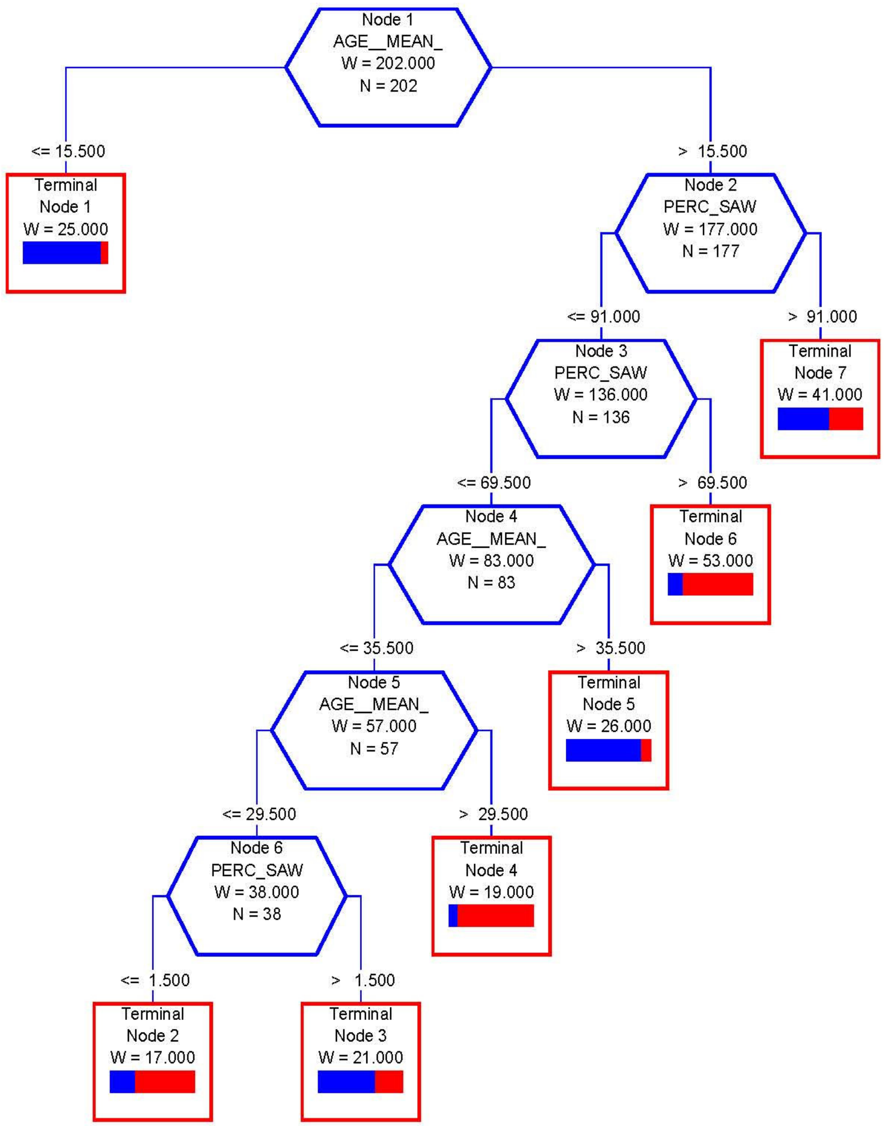
Appendix 10. Landsat 5 rescale model to convert images from float single to signed 16-bit top of atmosphere reflectance images.



Appendix 11. Disturbance index model used in our image differencing methodology to determine pine age classes susceptible to SPB in Landsat TM imagery.



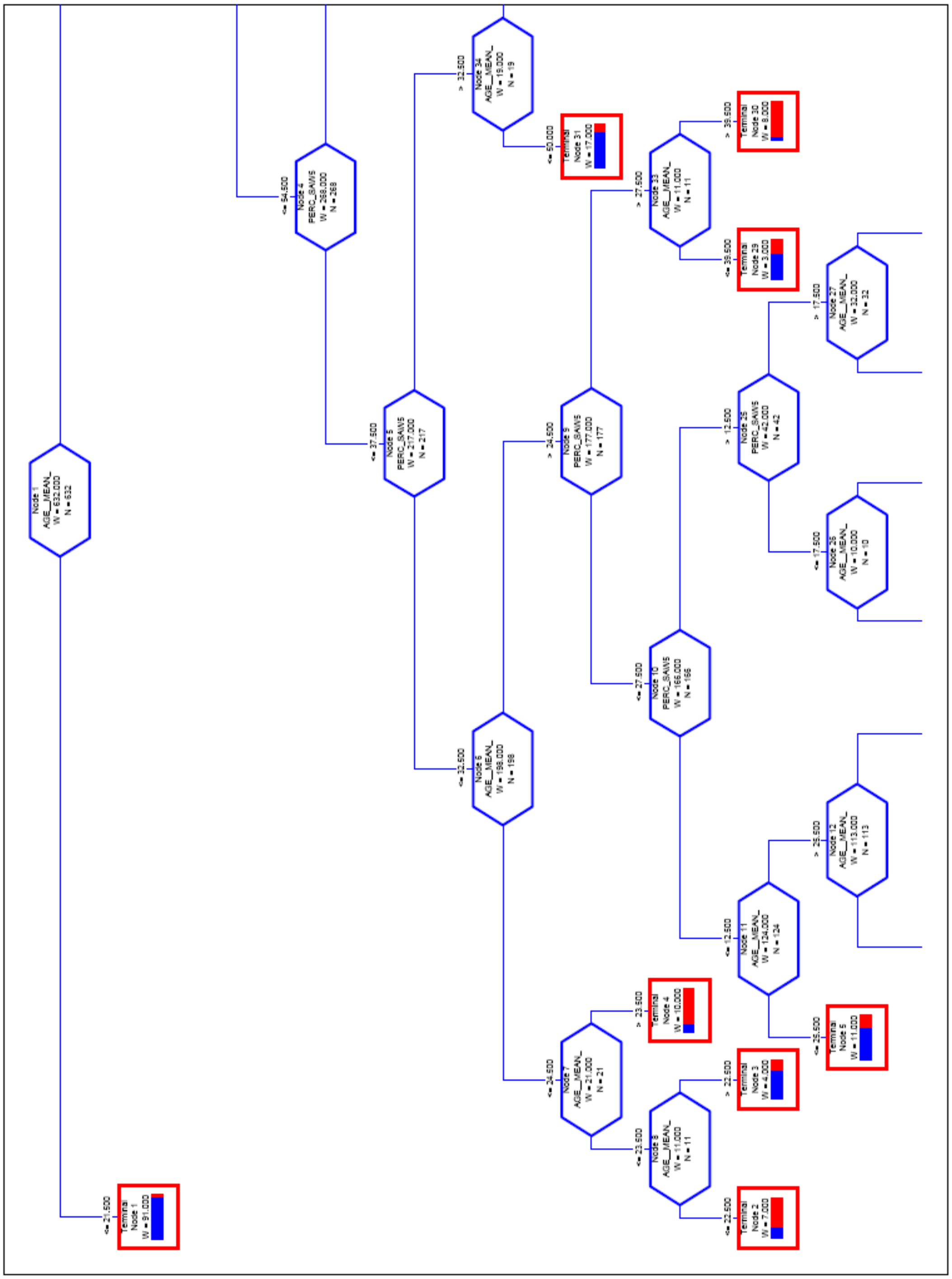
Appendix 12a. Coastal plain region CART® output classification tree. Red is SPB incidence, blue is no SPB.



Appendix 12b. Coastal Plain region CART® output age classes and percent sawtimber. Predicted SPB represents the whether or not each terminal node was classified as SPB outbreak or not based on the input data.

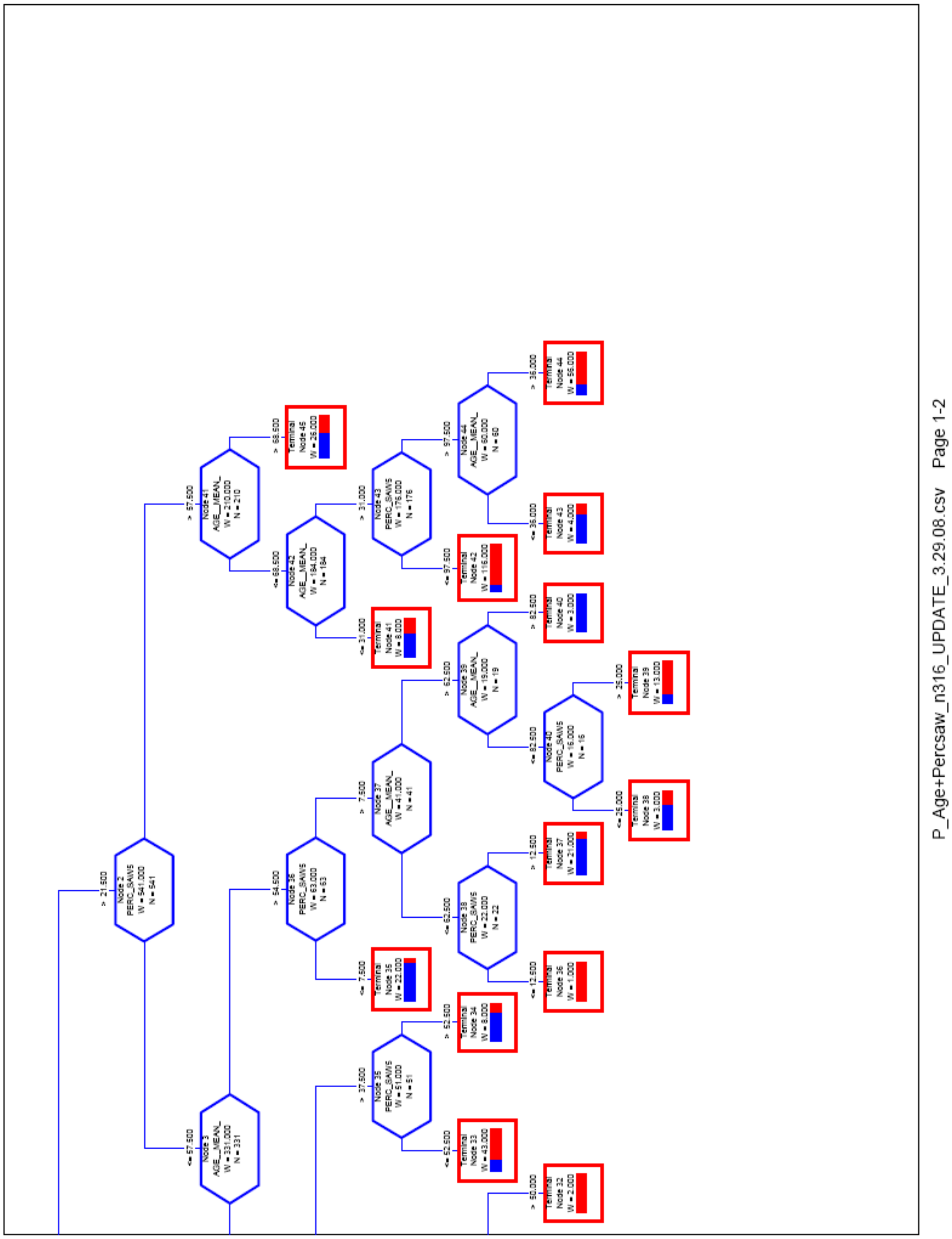
Coastal Plain						
Node	**Age Class**	**% Sawtimber**	**Predicted SPB (Yes/No)**	**# of Cases**	**% of Total Data**	**% in Predicted Class**
1	<= 15.5	NA	No	25	12.40	92.00
2	15.5 <= 29.5	<= 1.5	Yes	17	8.40	70.60
3	15.5 <= 29.5	1.5 <= 67.5	No	21	10.40	66.70
4	29.5 <= 35.5	<= 69.5	Yes	19	9.40	89.50
5	> 35.5	<= 69.5	No	26	12.90	88.50
6	> 15.5	69.5 <= 91	Yes	53	26.20	83.00
7	> 15.5	> 91	No	41	20.30	61.00

Appendix 13a. Piedmont region CART® output classification tree (Page 1-1). Red is SPB incidence, blue is no SPB.

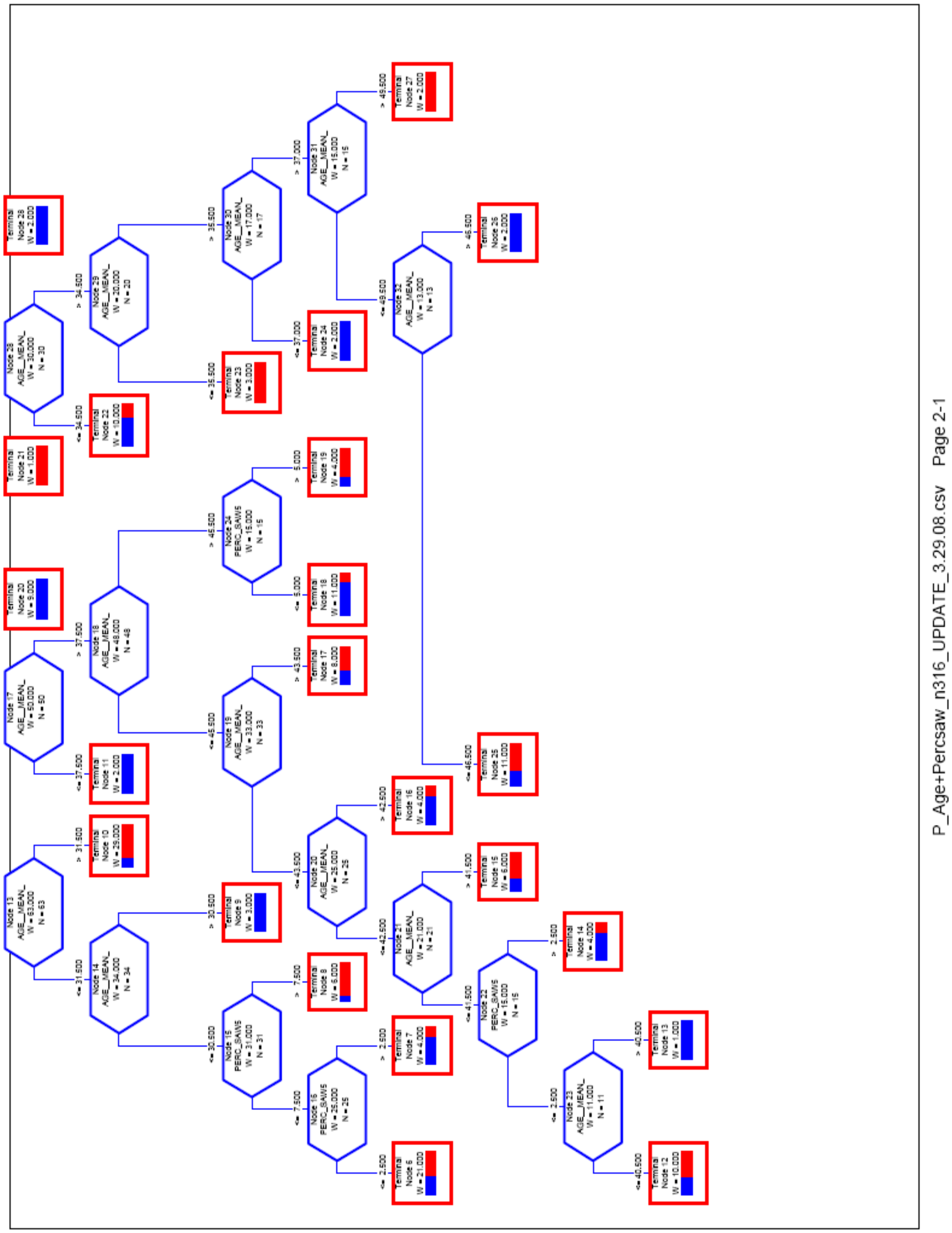


P_Age+Percsaw_n316_UPDATE_3.29.08.csv Page 1-1

Appendix 13a. Piedmont region CART® output classification tree (Page 1-2). Red is SPB incidence, blue is no SPB.



Appendix 13a. Piedmont region CART® output classification tree (Page 2-1). Red is SPB incidence, blue is no SPB.



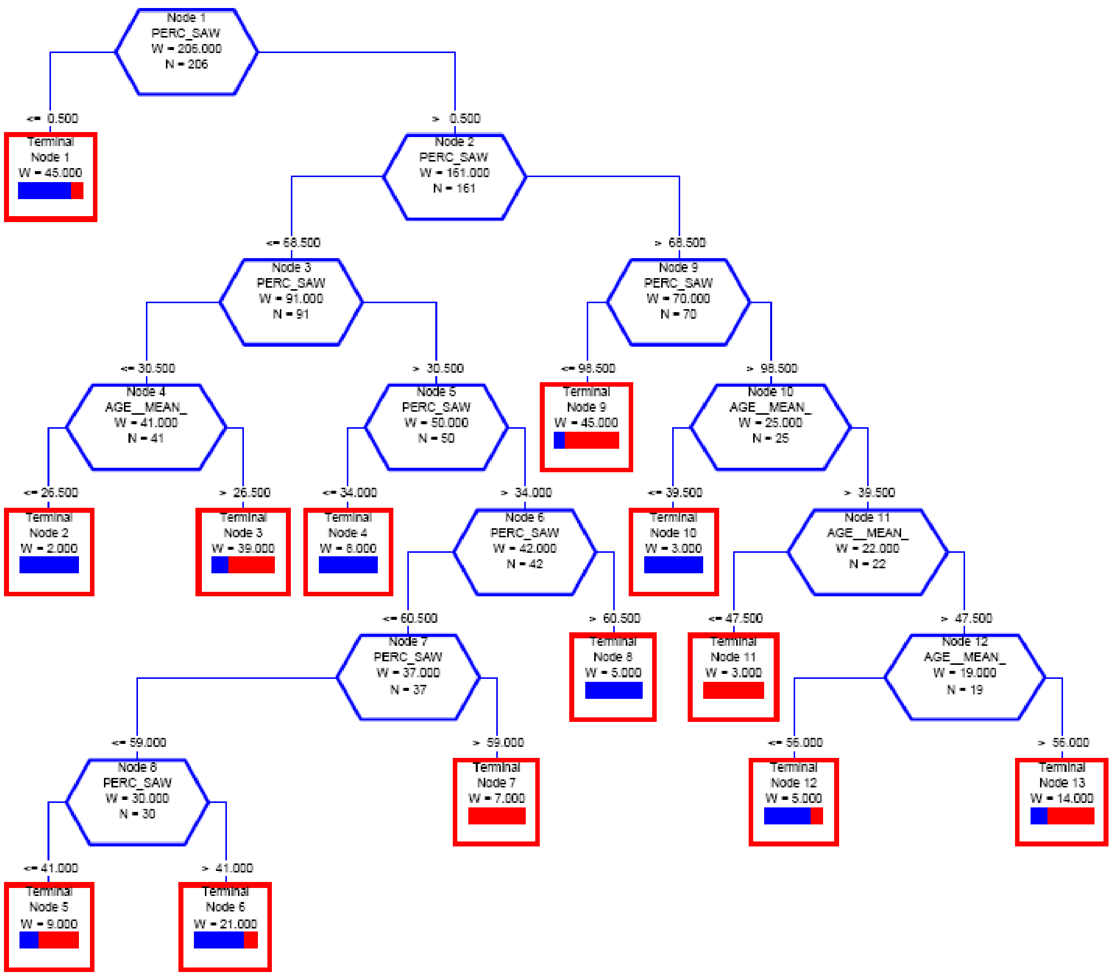
Appendix 13b. Piedmont region CART® output age classes and percent sawtimber. Predicted SPB represents the whether or not each terminal node was classified as SPB outbreak or not based on the input data.

Piedmont

Node	Age Class	% Sawtimber	Predicted SPB (Yes/No)	# of Cases	% of Total Data	% in Predicted Class
1	<= 21.5	NA	No	91	14.40	96.70
2	21.5 <= 22.5	<= 2	Yes	6	0.90	83.30
3	22.5 <= 23.5	<= 2	No	4	0.60	75.00
4	23.5 <= 24.5	<= 2	Yes	8	1.30	75.00
5	24.5 <= 25.5	<= 2	No	9	1.40	66.70
6	25.5 <= 30.5	<= 2	Yes	21	3.30	57.10
7	30.5 <= 31.5	<= 2	No	3	0.50	100.00
8	31.5 <= 45.5	<= 2	Yes	48	7.60	62.50
9	21.5 <= 45.5	2 <= 4.5	No	4	0.60	100.00
10	21.5 <= 45.5	4.5 <= 5.5	Yes	7	1.10	85.70
11	45.5 <= 54.5	<= 5.5	No	11	1.70	81.80
12	21.5 <= 54.5	5.5 <= 9	No	5	0.80	100.00
13	21.5 <= 54.5	9 <= 10.5	Yes	13	2.10	92.30
14	21.5 <= 51	10.5 <= 18.5	No	13	2.10	100.00
15	51 <= 54.5	10.5 <= 18.5	Yes	1	2.00	100.00
16	21.5 <= 29.5	18.5 <= 20.5	Yes	4	0.60	100.00
17	29.5 <= 34.5	18.5 <= 20.5	No	2	0.30	100.00
18	34.5 <= 54.5	18.5 <= 20.5	Yes	9	1.40	77.80
19	21.5 <= 32.5	20.5 <= 38	No	12	1.90	91.70
20	32.5 <= 54.5	20.5 <= 24	No	6	0.90	100.00
21	32.5 <= 52.5	24 <= 31.5	Yes	16	2.50	81.30
22	52.5 <= 54.5	24 <= 31.5	No	2	0.30	100.00
23	32.5 <= 33.5	31.5 <= 38	Yes	1	0.20	100.00
24	33.5 <= 50	31.5 <= 38	No	10	1.60	80.00
25	50 <= 54.5	31.5 <= 38	Yes	2	0.30	100.00
26	21.5 <= 54.5	38 <= 55.5	Yes	45	7.10	73.30
27	21.5 <= 54.5	55.5 <= 59	No	6	0.90	100.00
28	> 54.5	<= 7	No	22	3.50	90.90
29	54.5 <= 62.5	7 <= 12	Yes	1	0.20	100.00
30	54.5 <= 62.5	12 <= 59	No	22	3.50	86.40
31	62.5 <= 82.5	7 <= 34	No	5	0.80	60.00
32	62.5 <= 77	34 <= 59	Yes	10	1.60	90.00
33	77 <= 82.5	34 <= 59	No	1	0.20	100.00

Node	Age Class	% Sawtimber	Predicted SPB (Yes/No)	# of Cases	% of Total Data	% in Predicted Class
34	> 82.5	7 <= 59	No	3	0.50	100.00
35	21.5 <= 31	> 59	No	8	1.30	62.50
36	31 <= 68.5	59 <= 61.5	Yes	23	3.60	91.30
37	31 <= 68.5	61.5 < 69	No	5	0.80	100.00
38	31 <= 68.5	69 <=70.5	Yes	14	2.20	100.00
39	31 <= 68.5	70.5 <= 74.5	No	3	0.50	100.00
40	31 <= 68.5	74.5 <= 76.5	Yes	14	2.20	92.90
41	31 <= 68.5	76.5 <= 79	No	3	0.50	100.00
42	31 <= 68.5	79 <= 99	Yes	54	8.50	94.40
43	31 <= 36	> 99	No	4	0.60	75.00
44	36 <= 68.5	> 99	Yes	55	8.70	76.40
45	> 68.5	> 59	No	26	4.10	61.50

Appendix 14a. Mountain region CART® output classification tree. Red is SPB incidence, blue is no SPB.



Appendix 14b. Mountain region CART® output age classes and percent sawtimber. Predicted SPB represents the whether or not each terminal node was classified as SPB outbreak or not based on the input data.

Mountain						
Node	**Age Class**	**% Sawtimber**	**Predicted SPB (Yes/No)**	**# of Cases**	**% of Total Data**	**% in Predicted Class**
1	NA	<= 0.5%	No	45	21.80	84.40
2	<= 26.5	0.5 <= 30.5	No	2	1.00	100.00
3	> 26.5	0.5 <= 30.5	Yes	39	18.90	71.80
4	NA	30.5 <= 34	No	8	3.90	100.00
5	NA	34 <= 41	Yes	9	4.40	66.70
6	NA	41 <= 59	No	21	10.20	81.00
7	NA	59 <= 60.5	Yes	7	3.40	100.00
8	NA	60.5 <= 68.5	No	5	2.40	100.00
9	NA	68.5 <= 98.5	Yes	45	21.80	82.20
10	<= 39.5	> 98.5	No	3	1.50	100.00
11	39.5 <= 47.5	> 98.5	Yes	3	1.50	100.00
12	47.5 <= 56	> 98.5	No	5	2.40	80.00
13	> 56	> 98.5	Yes	14	6.80	71.40

Appendix 15. Definition of variables used in Cubist 2.05 models for digital elevation model comparison, courtesy of FHTET.

Cubist Variable	Definition
cc	canopy cover from NLCD 2002
curv2	curvature Arc/Info from dem
dem	digital elevation model - NED 30 m
dem2	digital elevation model - NC Lidar Resample
di	drainage index
ewslp	slope * cos(aspect) - from NED 30 m
ewslp2	slope * cos(aspect) - from NC Lidar
imp	impervious surface from NLCD 2002
lc	land cover from NLCD 2002
lofc1	leafoff tasselcap band 1
lofc2	leafoff tasselcap band 2
lofc3	leafoff tasselcap band 3
lofdy	leafoff imagery julian day
lofrc1	leafoff reflectance band 1
lofrc2	leafoff reflectance band 2
lofrc3	leafoff reflectance band 3
lofrc4	leafoff reflectance band 4
lofrc5	leafoff reflectance band 5
lofrc6	leafoff reflectance band 6
lonc1	leafon tasselcap band 1
lonc2	leafon tasselcap band 2
lonc3	leafon tasselcap band 3
londy	leafon imagery julian day
lonrc1	leafon reflectance band 1
lonrc2	leafon reflectance band 2
lonrc3	leafon reflectance band 3
lonrc4	leafon reflectance band 4
lonrc5	leafon reflectance band 5
lonrc6	leafon reflectance band 6
nsslp	slope * sin(aspect) - from NED 30 m
nsslp2	slope * sin(aspect) - from NC Lidar
posidx	slope position index - NLCD NED 30m
posidx2	slope position index - NLCD Lidar 30m
slp	slope
slp2	slope
sprc1	spring tasselcap band 1
sprc2	spring tasselcap band 2
sprc3	spring tasselcap band 3

Cubist Variable	**Definition**
sprdy	spring imagery julian day
sprrc1	spring reflectance band 1
sprrc2	spring reflectance band 2
sprrc3	spring reflectance band 3
sprrc4	spring reflectance band 4
sprrc5	spring reflectance band 5
sprrc6	spring reflectance band 6

Appendix 16. Cubist 2.05 total basal area model for USGS zone 59 (piedmont region) with National Elevation Dataset DEM

```
Cubist [Release 2.05]  Wed May 28 10:14:09 2008

    Target attribute `Tba'

Read 7750 cases (59 attributes) from z59tba.data

Model:

  Rule 1: [2236 cases, mean 6.227, range 0 to 291.4, est err 6.233]

    if
        cc <= 3
    then
        Tba = 0

  Rule 2: [58 cases, mean 8.134, range 0 to 119.93, est err 14.626]

    if
        cc > 3
        imp > 6
        lonc1 <= 92
        lofrc4 <= 111
        lonrc6 <= 26
    then
        Tba = -256.028 + 1.98 lofc3 + 1.26 lofrc5 + 0.62 cc - 1.05
lofc2
              + 1.03 sprc2 + 2.2 sprrc1 - 2.1 lofrc1 - 0.56 sprrc4 -
0.51 lonrc6
              + 0.43 lofrc4 + 0.26 lonc2 - 0.27 sprc3 - 0.62 lonrc2 +
0.6 lonrc1
              - 0.21 sprc1 - 0.2 lofc1 - 0.13 lonrc5 + 0.46 lofrc2 +
0.28 lofrc3
              + 0.035 dem - 0.19 lonc1 - 0.22 lonrc3 - 0.1 lonrc4 -
0.24 imp
              + 0.1 lofrc6 + 0.07 sprrc5 - 0.08 lonc3 - 0.15 sprrc3
              + 0.003 londy

  Rule 3: [38 cases, mean 9.026, range 0 to 146.38, est err 14.242]

    if
        cc > 3
        imp > 14
        sprc1 <= 62
        lonrc6 > 26
    then
        Tba = 495.01 - 4.21 lofc3 - 3.22 lofrc5 + 3.26 sprc3 + 2.43
sprrc5
              - 2.84 lofrc6 + 2.73 lofc1 - 2.09 sprc1 - 1.34 lonc2 +
1.86 sprrc6
              + 1.11 lonrc4 - 1.62 lofc2 + 0.96 sprrc4 - 1.47 sprrc3 +
0.27 cc
              - 0.66 lonc1 - 0.59 sprc2 - 1.3 lofrc1 - 0.61 imp

  Rule 4: [25 cases, mean 15.092, range 0 to 134.94, est err 19.753]

    if
        cc > 3
        lofrc4 <= 51
        lonrc6 <= 26
        sprrc3 > 20
    then
        Tba = -65.54 + 0.85 cc + 0.32 lonc2 - 0.45 lonc1 + 0.5 lonrc1
              + 0.22 lofrc3 + 0.08 lofc3

  Rule 5: [1244 cases, mean 30.434, range 0 to 331.25, est err 31.407]

    if
        cc > 3
        sprc1 > 62
        lonrc6 > 26
    then
        Tba = 226.74 - 2.21 sprc3 - 1.19 sprrc5 + 1.27 lofc1 - 1.16
sprc1
              + 0.7 sprrc4 + 0.68 lofc3 - 0.64 lofrc4 - 0.67 sprrc6
```

```
              - 1.13 lofrc2 + 0.21 cc + 0.84 sprrc2 + 0.58 sprrc3 +
0.16 lofrc5
              - 0.29 lonrc2 + 0.13 lonrc3 + 0.08 lonrc6 - 0.06 lonrc5
              + 0.2 lonrc1 - 0.11 imp + 0.2 sprrc1 - 0.2 lofrc1 + 0.09
lofrc3
              + 0.05 lofrc6 - 0.04 lonc3 - 0.05 lonc1

  Rule 6: [1013 cases, mean 35.254, range 0 to 331.25, est err 37.754]

    if
        cc > 3
        imp <= 5
        sprc1 > 62
        lonrc6 > 26
    then
        Tba = -23.58 + 4.02 lofc1 + 2.46 lonc3 - 1.81 sprc3 - 1.81
lofrc4
              + 1.05 lonrc5 - 1.04 lofrc5 - 3.02 lonrc2 - 0.96 sprrc5
              + 1.12 lonrc6 - 3.01 lofrc2 + 1.42 lonc1 - 0.95 sprc1
              - 0.48 lonrc4 + 0.57 sprrc4 - 0.57 sprrc6 + 0.18 cc +
0.29 lofc3
              + 0.52 sprrc3 + 0.72 sprrc2 - 0.11 lofrc6 - 0.07 lonc2 -
0.14 imp
              - 0.08 lofc2 + 0.06 lonrc3

  Rule 7: [164 cases, mean 35.538, range 0 to 228.55, est err 36.553]

    if
        cc > 3
        dem > 98
        dem <= 217
        sprc1 <= 62
        lofrc3 > 18
        lofrc4 > 74
        lonrc1 <= 41
        lonrc6 > 26
    then
        Tba = 1104.56 - 9.02 lofc3 - 7.42 lofc1 - 4.75 lofrc5 + 10.11
lofrc3
              + 3.06 lonc2 - 4.23 lofrc6 + 3.16 lofrc4 - 2.19 lonrc4
              + 1.76 lonrc6

  Rule 8: [56 cases, mean 41.709, range 0 to 286.68, est err 43.354]

    if
        cc > 3
        imp <= 14
        sprc1 <= 62
        lofrc6 > 35
        lonrc6 > 26
    then
        Tba = 953.56 - 5.1 lofc3 - 3.33 lofrc5 + 3.04 lofc1 - 2.33
lofc2
              - 1.66 lofrc6 - 1.06 lonc2 + 1.38 lonc1 + 0.54 cc - 1.81
lonrc2
              - 1.26 sprrc3

  Rule 9: [93 cases, mean 54.665, range 0 to 281.65, est err 49.346]

    if
        di > 52
        cc > 3
        imp <= 6
        dem > 92
        lonc1 <= 92
        lofrc4 > 51
        lonrc3 > 19
        lonrc6 <= 26
    then
        Tba = -729.376 + 11.23 lonrc3 + 1.58 cc + 1.32 lonc2 - 1.66
sprrc4
              + 1.44 sprc1 - 1.25 lonc1 + 1.5 sprc2 + 0.63 sprrc5 +
0.81 sprc3
              + 2 lonrc1 + 0.36 lofc3 - 0.27 lonrc4 - 0.31 lonrc6 +
0.57 lofrc3
              - 0.18 lonc3 + 0.14 lofrc5 + 0.6 sprrc1 - 0.14 lofc1 -
0.13 lofc2
              + 0.013 dem - 0.2 lofrc1 + 0.05 lofrc4

  Rule 10: [95 cases, mean 54.750, range 0 to 236.21, est err 50.208]
```

```
    if
        cc > 3
        imp <= 14
        dem > 217
        sprc1 <= 62
        sprc2 > 96
        lonrc6 > 26
    then
        Tba = 1204.88 - 9.35 lofrc5 - 8.14 lofc3 + 5.17 sprc3 + 3.95
sprrc5
              - 4.55 sprc1 + 4.43 lofc1 + 3.41 sprrc6 - 2.09 lonc2 -
10.3 lofrc1
              + 5.82 lofrc2 - 2.51 lofc2 + 1.81 sprrc4 + 1.45 lonrc4
              - 1.14 lofrc6 + 0.52 cc - 1.42 sprrc3 - 0.64 lonc1 -
0.57 sprc2
              - 0.59 imp - 0.2 lonrc1

  Rule 11: [26 cases, mean 55.908, range 0 to 214.36, est err 60.363]

    if
        di > 83
        cc > 3
        lonc1 <= 92
        lofrc4 > 51
        lonrc3 > 19
        lonrc6 <= 26
    then
        Tba = 46.89

  Rule 12: [387 cases, mean 58.690, range 0 to 326.15, est err 44.701]

    if
        di <= 83
        cc > 3
        imp <= 6
        dem > 55
        lonc3 <= 123
        lofrc4 <= 111
        lonrc3 > 19
        lonrc6 <= 26
        sprrc3 > 20
        sprrc4 > 71
    then
        Tba = 389.594 - 6.83 lonrc6 - 8.65 lonrc3 - 3.38 lonc2 + 1.08
lonrc4
              - 1.28 sprrc4 + 1.59 lonc1 + 1.06 lofc3 - 0.92 lonc3 +
1.04 sprc1
              + 0.49 cc + 0.71 lofrc5 + 0.65 sprc3 + 0.92 sprc2 + 0.4
sprrc5
              - 0.65 lofc2 + 0.31 lofrc4 - 1 lofrc1 + 0.8 lonrc1 + 0.9
sprrc1
              - 0.26 sprrc3 + 0.023 dem + 0.004 londy + 0.05 sprrc6

  Rule 13: [21 cases, mean 59.011, range 0 to 195.89, est err 52.687]

    if
        dem > 55
        ewslp <= 130
        lonc3 > 123
        lofrc4 <= 111
        sprrc3 > 20
        sprrc4 > 71
    then
        Tba = 1410.018 - 11.86 lonc3 - 0.87 lonrc6 + 0.89 lofc3 - 1.15
lonrc3
              + 0.55 lofrc5 - 0.53 sprrc4 - 0.69 lofc2 + 0.77 sprc2 +
0.24 cc
              - 0.28 lonc2 + 0.26 sprrc5 + 1.2 sprrc1 + 0.28 lofrc4 -
1 lofrc1
              + 0.32 lonc1 + 0.21 sprc3 + 0.15 sprrc6 + 0.013 dem

  Rule 14: [71 cases, mean 59.749, range 0 to 255.93, est err 34.655]

    if
        cc > 3
        imp <= 14
        dem <= 217
        sprc1 <= 62
        lofrc3 > 18
        lofrc6 <= 35
        lonrc1 > 41
```

```
        lonrc6 > 26
    then
        Tba = -172.49 + 8.2 lonrc1 + 1.36 lonrc6 - 1.1 lofc3 - 0.72
lofrc5
              - 0.99 lofrc6 + 0.64 lofc1 - 0.52 lofc2 + 0.13 cc - 0.15
lonc2
              + 0.2 lonc1 - 0.26 lonrc2 - 0.18 sprrc3

  Rule 15: [34 cases, mean 74.678, range 0 to 261.47, est err 49.307]

    if
        dem <= 98
        sprc1 <= 62
        lofrc4 > 74
        lonrc1 <= 41
        lonrc6 > 26
    then
        Tba = 437.753 + 5.411 dem + 12.95 lofrc6 - 11.1 sprc1 - 7.45
lofrc5
              - 7.5 lonrc6 + 4.76 lonrc5 - 1.63 ewslp - 0.47 lofc3 +
0.21 sprrc5
              + 0.26 sprc3 + 0.21 lofc1 + 0.19 sprrc6 - 0.09 lonc2 +
0.11 sprrc4
              - 0.11 lofc2 + 0.03 cc + 0.04 lonrc4

  Rule 16: [680 cases, mean 78.735, range 0 to 332.1, est err 45.451]

    if
        cc > 3
        lonc1 > 92
        lofrc4 <= 111
        lonrc6 <= 26
        sprrc3 > 20
    then
        Tba = 567.502 - 5.39 sprrc5 - 6.22 sprc3 + 3.85 lonc2 - 2.09
lonrc4
              + 6.7 lonrc1 + 2.31 sprc1 + 1.1 cc - 4.26 lonrc2 + 1.3
lofc3
              - 1.49 sprrc6 - 1.35 lonc1 + 0.82 lofrc5 - 1.06 lofrc6
              - 1.12 lofc2 - 0.75 sprc2 - 0.39 lonc3 + 0.45 lofc1 +
0.67 lofrc3
              + 1.3 sprrc1 - 1 lofrc1 - 0.22 lonrc6 + 0.29 lonrc3 -
0.34 sprrc3
              - 0.14 lonrc5 + 0.49 lofrc2 + 0.038 dem + 0.14 lofrc4 -
0.25 imp
              + 0.006 londy

  Rule 17: [33 cases, mean 80.629, range 0 to 326.38, est err 49.720]

    if
        cc > 3
        dem <= 55
        lonrc6 <= 26
    then
        Tba = -528.085 + 8.395 dem + 1.35 cc + 0.58 lonc2 - 0.39
lonrc4
              + 0.22 lofc3 + 0.14 lofrc5 + 0.5 lonrc1 + 0.17 lofrc3 -
0.06 lofc1
              - 0.04 lonrc6 - 0.05 lonc1

  Rule 18: [80 cases, mean 81.308, range 0 to 332.1, est err 66.619]

    if
        di <= 52
        imp <= 6
        posidx <= 26
        sprc2 <= 97
        lofrc4 <= 111
        lofrc6 > 21
        lonrc3 > 19
        lonrc6 <= 26
        sprrc4 <= 71
    then
        Tba = -1243.503 + 23.12 sprc3 + 12.55 sprrc5 + 14.79 sprrc4
              - 21.34 lonrc3 - 19.81 sprc2 - 12.16 sprrc3 + 3.73 slp
              + 0.52 lonrc4 - 0.81 lonc1 - 0.51 sprrc6 + 0.017 dem

  Rule 19: [29 cases, mean 86.400, range 0 to 270.45, est err 55.118]

    if
```

```
        cc > 3
        dem <= 217
        lofrc3 > 18
        lofrc4 <= 74
        lofrc6 <= 35
        lonrc1 <= 41
        lonrc6 > 26
        lonrc6 <= 29
    then
        Tba = 1893.83 - 61.06 lonrc6 + 37.1 lonrc2 - 36.4 lonrc1 -
13.18 lofrc6
              + 19.88 lofrc3 - 5.85 lonrc5 + 5.68 sprc1

  Rule 20: [204 cases, mean 87.143, range 0 to 278.98, est err 55.178]

    if
        di <= 52
        imp <= 6
        dem > 92
        posidx > 26
        lonc1 <= 92
        sprc2 <= 97
        sprc3 <= 103
        lofrc4 <= 111
        lonrc3 > 19
        lonrc6 <= 26
        sprrc4 <= 71
    then
        Tba = -1057.944 + 4.99 lofc3 + 3.02 lofrc5 - 3.83 lonc1 + 1.93
lonrc4
              + 1.61 sprrc5 + 1.7 sprc3 + 1.8 lofrc6 - 1.42 lonrc6 +
0.75 cc
              + 0.84 lonrc5 - 1.04 sprrc4 - 3.6 lofrc1 + 0.53 lonc2 +
1.12 sprc2
              + 0.52 lonc3 + 1.7 sprrc1 + 0.45 sprc1 - 0.48 lofc2 -
0.31 sprrc6
              - 0.39 lonrc3 + 0.7 lonrc1 + 0.03 dem - 0.22 sprrc3 +
0.004 londy
              - 0.04 lofrc4

  Rule 21: [141 cases, mean 92.221, range 0 to 441.12, est err 66.572]

    if
        cc > 3
        imp <= 14
        dem > 217
        sprc1 <= 62
        sprc2 <= 96
        lonrc6 > 26
    then
        Tba = 2280.93 - 7.43 lonrc5 - 8.76 lonc3 - 2.23 lofrc5 - 2.5
lofc3
              - 1.37 sprc3 - 0.79 lonc2 - 0.72 sprrc5 - 0.9 sprc1 -
0.81 lofrc6
              - 1.9 lonrc1 + 0.64 lofc1 + 0.58 sprrc6 + 0.32 sprrc4 -
0.34 lofc2
              + 0.09 cc + 0.13 lonrc4

  Rule 22: [258 cases, mean 93.339, range 0 to 257.43, est err 47.596]

    if
        di <= 52
        imp <= 6
        dem > 92
        sprc3 > 103
        lofrc4 > 51
        lofrc4 <= 111
        lonrc3 > 19
        lonrc6 <= 26
        sprrc4 <= 71
    then
        Tba = 768.34 - 10.06 lonc3 - 5.59 lonrc5 - 5.16 lonrc6 + 1.7
cc
              + 3.09 lofc3 + 2.52 sprc3 + 1.8 lofrc5 + 1.64 lofrc6 +
0.55 sprrc5
              - 0.49 lofrc4 + 0.57 lofc2 - 1.3 lofrc1 - 0.36 sprrc6

  Rule 23: [174 cases, mean 94.815, range 0 to 297.19, est err 49.257]

    if
```

```
        cc > 3
        lofrc4 > 111
        lonrc6 <= 26
        sprrc3 > 20
    then
        Tba = -925.95 + 4.92 lonc2 - 5.45 lonc1 + 7.4 lonrc1 + 3.56
lofrc3
              + 2.45 lofc2 - 1.58 lofrc4 + 1.44 lofc3 + 1 lofrc5 + 0.5
cc
              - 1.29 sprrc3 - 0.43 lonrc6 + 1.5 sprrc1 - 0.3 lonrc4 -
0.3 lonc3
              + 0.53 sprc2 - 0.32 sprrc4 - 0.28 lofc1 - 0.7 lofrc1 +
0.04 dem
              - 0.14 sprc3 + 0.005 londy

  Rule 24: [44 cases, mean 101.009, range 0 to 249.29, est err 60.589]

    if
        cc > 3
        sprc1 <= 62
        lofrc4 <= 74
        lofrc6 <= 35
        lonrc1 <= 41
        lonrc6 > 29
    then
        Tba = -1278.448 + 11.6 lonc2 - 8.5 lonrc4 + 9.57 lonrc6 +
23.69 sprrc2
              - 7.99 sprc1 - 8.21 imp + 3.12 lofrc4 + 0.677 dem - 1.11
lofc3
              - 0.77 lofrc5 - 0.81 lofrc6 + 0.49 lofc1 + 0.35 sprc3
              + 0.26 sprrc5 - 0.34 lofc2 + 0.22 sprrc6 + 0.08 cc +
0.12 sprrc4
              + 0.14 lofrc3 - 0.09 sprrc3

  Rule 25: [24 cases, mean 101.244, range 0 to 191.15, est err 35.855]

    if
        dem <= 217
        lofrc3 <= 18
        lonrc6 > 26
    then
        Tba = -667.406 - 13.47 lofrc6 - 8.43 lofrc3 + 15.9 sprrc1 +
3.99 sprc2
              - 4.55 imp + 0.296 dem + 0.84 posidx - 0.09 lofc3 - 0.06
lofrc5
              + 0.05 lofc1

  Rule 26: [66 cases, mean 102.085, range 0 to 319.41, est err 61.043]

    if
        di <= 52
        cc > 3
        dem > 92
        lonc1 <= 92
        sprc2 > 97
        sprc3 <= 103
        lofrc4 <= 111
        lonrc3 > 19
        lonrc6 <= 26
        sprrc4 <= 71
    then
        Tba = -6229.29 + 28.74 lonc3 + 13.14 lonrc5 + 15 lonrc6 +
18.96 sprc2
              - 2.31 sprc3 + 1.95 lofc3 + 1.13 lofrc5 + 1.04 lofrc6
              + 0.53 sprrc5 - 0.31 lofrc4 + 0.36 lofc2 - 1 lofrc1 +
0.13 cc
              - 0.29 lonc1 + 0.16 lonrc4 - 0.23 sprrc6

  Rule 27: [139 cases, mean 103.005, range 0 to 446.42, est err
54.598]

    if
        di <= 52
        imp <= 6
        dem > 92
        posidx <= 26
        lonc1 <= 92
        sprc3 <= 103
        lofrc6 <= 21
        lonrc6 <= 26
        sprrc4 <= 71
```

```
    then
        Tba = 331.48 - 12.85 lofrc6

  Rule 28: [1303 cases, mean 103.170, range 0 to 482.18, est err
48.391]

    if
        cc > 3
        dem > 55
        lonc1 <= 92
        lonrc3 <= 19
        lonrc6 <= 26
        sprrc3 > 20
    then
        Tba = -669.628 + 3.26 lofc3 + 1.51 cc + 2.08 lofrc5 - 2.23
lofc2
              - 1.58 lonrc6 + 2.68 sprc2 - 1.46 sprrc4 + 5.3 sprrc1 +
1.4 lofrc4
              + 0.83 lonc2 - 1.14 lofc1 - 3.3 lofrc1 - 0.94 sprc1 +
0.69 sprrc5
              + 1.56 lofrc3 + 0.98 sprrc6 - 0.92 lonrc3 - 0.75 lonc1
              + 0.51 sprc3 + 1.1 lonrc1 - 0.27 lonrc4 - 0.28 lonc3 +
0.058 dem
              - 0.28 sprrc3 + 0.005 londy

  Rule 29: [182 cases, mean 108.837, range 0 to 281.65, est err
45.087]

    if
        lonrc4 > 113
        lonrc6 <= 26
        sprrc3 <= 20
    then
        Tba = -1792.2 + 8.65 lofrc5 + 9.75 lofc3 + 8.59 lofrc3 - 3.27
lonrc5
              - 3.95 lofc1 + 10.6 sprrc1 - 3.05 sprc1 + 2.11 lofc2 +
0.42 ewslp
              + 1.02 slp + 0.11 lonc2 - 0.06 lonrc6 - 0.08 lonc1 +
0.03 cc
              - 0.05 lonc3 - 0.04 lonrc4 + 0.1 lonrc1

  Rule 30: [71 cases, mean 109.758, range 0 to 285.68, est err 66.659]

    if
        di <= 83
        imp <= 6
        dem > 55
        dem <= 92
        lonc1 <= 92
        lofrc4 > 51
        lonrc3 > 19
        lonrc6 <= 26
        sprrc4 <= 71
    then
        Tba = 1181.205 - 20.26 lonrc3 - 8.5 lonc2 + 2.209 dem + 9.57
lonc1
              - 4.55 lonrc6 - 0.458 lofdy + 1.54 cc - 0.66 sprrc4 +
0.66 sprc1
              + 0.38 sprc3 + 0.26 sprrc5 + 0.54 sprc2 + 0.4 lonrc1

  Rule 31: [29 cases, mean 139.975, range 0 to 266.98, est err 48.647]

    if
        cc > 3
        ewslp > 130
        lonc1 <= 92
        lonc3 > 123
        lonrc3 > 19
        sprrc4 > 71
    then
        Tba = 1698.59 - 10.21 lonc3 - 5.81 lonrc3 - 3.14 lonrc6 - 1.89
lonc2
              + 2.21 lonc1 + 0.2 lofc3 + 0.14 lofrc5 + 0.05 cc - 0.07
lonrc4
              - 0.11 lofc2 + 0.2 lonrc1 + 0.06 lofrc4 - 0.2 lofrc1 -
0.07 sprrc3


  Rule 32: [201 cases, mean 145.785, range 0 to 325.61, est err
52.218]
```

```
    if
        cc > 3
        lonrc4 <= 113
        lonrc6 <= 26
        sprrc3 <= 20
    then
        Tba = 145.575 - 4.29 lonrc5 - 8.5 sprrc3 + 2.65 lonrc4 + 0.039
londy
              + 0.35 cc


Evaluation on training data (7750 cases):

    Average  |error|              31.442
    Relative |error|                0.53
    Correlation coefficient         0.67


        Attribute usage:
          Conds  Model

            89%    71%    cc
            75%    67%    lonrc6
            34%    35%    dem
            32%    62%    sprrc3
            31%    61%    sprc1
            29%    35%    imp
            29%    64%    lonc1
            27%    56%    lonrc3
            26%    62%    lofrc4
            14%           di
            14%    54%    sprrc4
             7%    65%    sprc3
             6%    36%    sprc2
             5%    63%    lonc3
             5%           posidx
             5%    47%    lofrc6
             4%    54%    lonrc4
             4%    52%    lonrc1
             3%    44%    lofrc3
                    2%    ewslp
                   70%    lofc3
                   70%    lofrc5
                   63%    sprrc5
                   61%    sprrc6
                   59%    lofc1
                   54%    lofc2
                   53%    lonc2
                   50%    lofrc1
                   47%    sprrc1
                   45%    lonrc5
                   34%    lonrc2
                   34%    lofrc2
                   33%    londy
                   25%    sprrc2
                    3%    slp


Time: 0.9 secs
```

Appendix 17. Cubist 2.05 total basal area model for USGS zone 59 (piedmont region) with LiDAR DEM

```
Cubist [Release 2.05]  Wed May 28 10:16:12 2008

    Target attribute `Tba'

Read 7750 cases (59 attributes) from z59tba.data

Model:

  Rule 1: [2236 cases, mean 6.227, range 0 to 291.4, est err 6.233]

    if
        cc <= 3
    then
        Tba = 0

  Rule 2: [57 cases, mean 8.276, range 0 to 119.93, est err 13.954]

    if
        cc > 3
        imp > 6
        lonc1 <= 92
        lofrc4 <= 111
        lonrc6 <= 26
        sprrc3 > 20
    then
        Tba = -422.412 + 3.11 lofc3 + 2 lofrc5 + 0.66 cc - 0.7 lonrc6
              + 1.12 sprc2 + 0.75 lofrc6 - 0.62 sprrc4 - 2.1 lofrc1 +
2 sprrc1
              - 0.61 lofc2 - 0.32 sprc3 - 0.67 lonrc2 - 0.28 sprc1 +
0.31 lonrc3
              - 0.13 lonrc5 + 0.25 lofrc3 - 0.26 imp + 0.029 dem2 -
0.14 lofc1
              + 0.3 lonrc1 - 0.09 lonc3 + 0.26 lofrc2 + 0.07 lonc2 -
0.11 lonc1
              + 0.05 sprrc6 + 0.03 sprrc5 - 0.06 sprrc3

  Rule 3: [38 cases, mean 9.026, range 0 to 146.38, est err 13.518]

    if
        cc > 3
        imp > 14
        sprc1 <= 62
        lonrc6 > 26
    then
        Tba = 331.17 - 3.48 lofc3 - 2.61 lofrc5 + 2.67 sprc3 + 1.96
sprrc5
              - 2.32 lofrc6 + 2.2 lofc1 - 1.67 sprc1 + 1.42 sprrc6 -
1.28 lofc2
              + 0.83 sprrc4 - 1.3 sprrc3 - 0.42 lonc2 + 0.23 cc - 0.67
sprc2
              - 1.1 lofrc1 + 0.34 lonc1 - 0.47 imp + 0.08 ewslp2

  Rule 4: [31 cases, mean 16.407, range 0 to 134.94, est err 22.330]

    if
        cc > 3
        lofrc4 <= 51
        sprrc3 > 20
    then
        Tba = -100.39 + 0.91 cc + 0.67 lofc3 + 0.39 lofrc5 + 0.35
lonc2
              - 0.44 lonc1 - 0.25 sprrc5 - 0.32 sprc3 - 0.64 lonrc2 +
0.8 lonrc1
              + 0.42 lofrc3 - 0.24 sprc1 + 0.28 lonrc3 - 0.13 lonrc5
              + 0.6 sprrc1 + 0.14 lonrc6 - 0.27 imp - 0.16 lofc1 -
0.12 lonc3
              - 0.4 lofrc1 + 0.11 lofrc6 + 0.26 lofrc2 + 0.08 sprrc4 +
0.1 sprc2
              - 0.07 sprrc3 - 0.05 lofc2 - 0.03 lonrc4 + 0.04 sprrc6


  Rule 5: [98 cases, mean 24.897, range 0 to 326.15, est err 29.302]
```

```
    if
        cc > 3
        imp > 3
        lonc1 <= 92
        lonrc6 <= 26
    then
        Tba = -172.676 + 1.39 lofc3 + 0.9 lofrc5 + 0.41 cc - 0.48
sprc3
              - 0.31 sprrc5 + 0.29 lonc2 - 0.92 lonrc2 + 1.4 sprrc1
              - 0.25 lonrc5 - 0.34 sprc1 - 0.29 lonc3 + 0.45 sprc2 +
0.5 lofrc3
              - 0.35 lonc1 + 0.41 lonrc3 - 0.31 lofc1 + 0.8 lonrc1 -
0.9 lofrc1
              + 0.23 lofrc6 - 0.39 imp - 0.2 lofc2 + 0.37 lofrc2 -
0.09 sprrc4
              + 0.09 sprrc6 + 0.016 dem2 - 0.06 lonrc6 - 0.1 sprrc3
              - 0.04 lonrc4 + 0.05 lofrc4 + 0.03 ewslp2

  Rule 6: [1244 cases, mean 30.434, range 0 to 331.25, est err 31.401]

    if
        cc > 3
        sprc1 > 62
        lonrc6 > 26
    then
        Tba = 88.58 - 1.79 sprc3 - 1.58 sprc1 - 0.84 sprrc5 + 0.8
lonc1
              + 0.69 lofc1 + 1 sprrc3 + 0.79 sprc2 + 1.15 sprrc2 -
0.34 lonrc4
              - 0.41 lofrc4 + 0.34 lonc3 - 0.41 sprrc6 - 0.26 lofrc5
              - 0.75 lonrc2 + 1.1 sprrc1 - 0.77 lofrc2 + 0.24 lonrc6
              + 0.25 sprrc4 + 0.42 lofrc3 - 0.26 lonrc3 - 0.32 imp +
0.08 cc
              + 0.15 lofc2 + 0.05 ewslp2 - 0.07 nsslp2 + 0.03 posidx2

  Rule 7: [1285 cases, mean 30.961, range 0 to 441.12, est err 32.207]

    if
        cc > 3
        cc <= 50
        lofrc4 <= 111
        sprrc3 > 20
    then
        Tba = -6.97 + 0.1 cc + 0.12 lofc3 + 0.08 lofrc5 - 0.07 sprc3
              - 0.05 sprrc5 - 0.14 lonrc2 - 0.06 sprc1 + 0.07 lonrc3
              + 0.04 lonrc6 - 0.03 lonrc5 + 0.1 sprrc1

  Rule 8: [1013 cases, mean 35.254, range 0 to 331.25, est err 37.740]

    if
        cc > 3
        imp <= 5
        sprc1 > 62
        lonrc6 > 26
    then
        Tba = 135.25 - 4.75 sprc3 - 4.18 sprc1 - 2.23 sprrc5 + 2.78
lofc1
              + 2.61 lonc1 + 1.74 lonc3 - 1.53 lofrc4 + 2.66 sprrc3
              - 1.09 lonrc4 - 1.01 lofrc5 + 2.1 sprc2 - 2.94 lonrc2
              + 3.05 sprrc2 + 1 lonrc6 - 2.71 lofrc2 - 1.09 sprrc6 + 3
sprrc1
              + 0.65 sprrc4 + 1.1 lofrc3 + 0.37 lonrc5 - 0.69 lonrc3 -
0.85 imp
              + 0.2 cc + 0.39 lofc2 + 0.14 ewslp2 - 0.19 nsslp2 + 0.09
posidx2

  Rule 9: [25 cases, mean 35.380, range 0 to 195.87, est err 27.082]

    if
        cc > 3
        sprc1 <= 62
        lofrc5 > 73
        lonrc6 > 26
        dem2 <= 216
    then
        Tba = -950.3 + 34.5 sprrc1 - 6.52 sprc1 - 14.87 sprrc2 + 4.2
lofrc5
              + 3.39 lonrc6 - 0.19 lofc3 - 0.17 lofrc6 + 0.12 lofc1 -
0.09 lofc2
              + 0.02 cc
```

```
  Rule 10: [27 cases, mean 40.924, range 0 to 178.74, est err 44.072]

    if
        lofc2 > 132
        lonc1 <= 92
        lofrc4 <= 105
        lonrc2 <= 35
        lonrc6 <= 26
        sprrc3 > 20
        sprrc4 > 63
        ewslp2 > 468
        curv2 > 500
    then
        Tba = 0

  Rule 11: [184 cases, mean 46.981, range 0 to 228.55, est err 44.211]

    if
        cc > 3
        sprc1 <= 62
        lofrc4 > 75
        lofrc5 <= 73
        lonrc1 <= 41
        lonrc6 > 26
        dem2 <= 216
    then
        Tba = 2489.46 - 14.39 lofc3 - 12.47 lofrc6 - 11.42 sprc1 -
8.12 lofrc5
              - 12.06 lofc2 + 8.16 lofrc4 + 5.22 sprc3 + 3.65 sprrc5
              + 4.3 sprrc4 + 4.87 sprrc6 + 4.69 sprrc2 + 1.49 lonrc6
              - 3.54 lonrc2 - 2.6 lofrc3 + 0.47 cc

  Rule 12: [58 cases, mean 56.154, range 0 to 215.14, est err 41.918]

    if
        cc > 50
        lofrc4 <= 83
        lonrc2 <= 35
        sprrc3 > 20
        sprrc4 > 63
        curv2 > 500
        dem2 <= 260
        nsslp2 <= 121
    then
        Tba = -885.91 + 11.49 lonc2 - 8.42 lonrc4 + 1.95 lonc3 - 1.19
lofrc5
              + 1.09 lonrc5 - 1.16 sprrc4 - 1.11 lofc3 + 1.02 sprc1 -
1.3 lonrc3
              + 0.87 lonc1 + 0.55 sprc2 + 0.19 curv2 + 0.044 dem2

  Rule 13: [99 cases, mean 57.603, range 0 to 236.21, est err 53.021]

    if
        imp <= 14
        sprc1 <= 62
        sprc2 > 96
        lonrc6 > 26
        dem2 > 216
    then
        Tba = 1510.044 - 9.23 lofrc5 - 11.35 lofc3 + 6.49 lofc1 + 3.96
sprrc5
              - 4.99 lofrc6 - 4.14 sprc1 + 3.76 sprc3 + 3.26 lonc3 -
2.36 lonc2
              + 2.89 lonrc6 - 7.2 lofrc1 + 1.9 sprrc6 + 1.48 sprrc4 -
1.86 lofc2
              + 1.81 lonc1 - 4.2 lonrc1 + 0.63 cc - 1.98 sprrc3 - 0.96
lofrc4
              - 1.26 sprc2 + 1.04 lofrc3 + 0.36 lonrc3 + 0.12 ewslp2 -
0.3 imp
              + 0.013 dem2

  Rule 14: [222 cases, mean 60.235, range 0 to 270.45, est err 51.782]

    if
        imp <= 14
        sprc1 <= 62
        lofrc2 <= 28
        lofrc5 <= 73
        lonrc6 > 26
        dem2 <= 216
```

```
    then
        Tba = 1594.77 - 14.19 lofrc6 - 11.33 lofc3 + 2.12 cc - 3.14
lofrc5
              + 3.43 lonrc6 - 1.93 sprc1 - 2.01 lofc2 + 1.36 lofrc4 +
0.98 sprc3
              + 0.68 sprrc5 + 0.8 sprrc4 + 0.91 sprrc6 + 0.88 sprrc2
              - 0.66 lonrc2 - 0.49 lofrc3

  Rule 15: [85 cases, mean 61.076, range 0 to 255.93, est err 44.790]

    if
        cc > 3
        imp <= 14
        sprc1 <= 62
        lonrc1 > 41
        lonrc6 > 26
        dem2 <= 216
    then
        Tba = 138.68 + 23.97 sprc1 - 11.88 sprrc4 - 17.85 sprrc3 -
6.68 sprrc6
              + 2.86 lonrc3 - 1.18 lofc3 - 0.85 lofrc5 + 2.9 lonrc1
              - 0.85 lofrc6 + 0.72 lofc1 + 0.58 sprc3 + 0.44 sprrc5 -
0.45 lofc2
              + 0.09 cc - 0.2 sprc2 + 0.1 lonrc6 - 0.07 lonc2 - 0.2
lofrc1

  Rule 16: [255 cases, mean 66.163, range 0 to 326.15, est err 50.839]

    if
        imp <= 6
        lonc1 <= 88
        lofrc4 <= 111
        lonrc5 > 56
        lonrc6 <= 26
        sprrc3 > 20
        curv2 <= 500
        dem2 > 55
        nsslp2 > 105
    then
        Tba = -61.275 - 6.87 sprc1 - 5.18 lonrc6 + 5.7 sprc2 + 4.47
sprrc3
              - 2.49 sprc3 + 8.2 sprrc1 + 0.57 cc - 0.73 lofrc5 - 0.77
sprrc4
              - 0.85 lofc2 + 0.42 sprrc6 - 1.3 lofrc1 + 0.23 sprrc5 -
0.23 lofc3
              + 0.37 lofrc3 + 0.18 lofc1 + 0.16 lofrc4 + 0.1 lonc2 +
0.019 dem2
              - 0.1 lonc1 + 0.2 lonrc1

  Rule 17: [160 cases, mean 69.756, range 0 to 326.15, est err 49.348]

    if
        lonc1 <= 92
        lonrc5 > 56
        lonrc6 <= 26
        curv2 <= 500
        dem2 > 55
        posidx2 <= 19
        nsslp2 > 105
    then
        Tba = -1193.3 - 8.17 sprc1 + 2.78 lonc2 + 6.03 sprc2 + 6.32
sprrc3
              - 3.35 lofrc4 + 3.71 lofc1 - 2.91 lonc1 + 2.35 sprrc6 +
6.7 sprrc1
              - 1.66 sprc3 + 4.5 lonrc1 + 1.65 lofc2 + 1.2 lofc3 +
0.68 lonrc4
              + 0.38 cc + 0.69 lofrc3 - 0.13 lofrc5 - 0.11 lonrc6 -
0.08 lonc3

  Rule 18: [53 cases, mean 75.885, range 0 to 192.37, est err 71.586]

    if
        cc > 50
        lonc1 <= 92
        lofrc4 > 105
        lonrc2 <= 35
        lonrc6 <= 26
        sprrc3 > 20
        sprrc4 > 63
        ewslp2 > 468
        curv2 > 500
```

```
        curv2 <= 568
    then
        Tba = 2123.94 - 19.44 lofrc4 + 1.86 sprrc4 - 3.03 sprrc3 -
0.23 lonrc4
              + 0.2 lonc2 + 0.24 sprc1 - 0.21 lonrc6 - 0.14 lofrc5 +
0.21 lonc1
              - 0.16 lofc3 + 0.2 sprc2
```

Rule 19: [669 cases, mean 79.666, range 0 to 332.1, est err 45.456]

```
    if
        cc > 3
        lonc1 > 92
        lofrc4 <= 111
        lonrc6 <= 26
        sprrc3 > 20
        dem2 > 55
    then
        Tba = 1573.52 - 7.25 sprrc5 - 8.6 sprc3 + 5.61 sprc1 - 3.63
sprrc6
              - 6.96 lonrc2 - 1.62 lonrc5 + 1.03 cc + 5.7 lonrc1 -
3.05 sprc2
              + 1.25 lonc2 - 2.07 lofc2 - 2.43 sprrc3 - 1.42 lofrc6 +
1 lofrc4
              - 0.57 lonrc4 + 0.65 lofc3 + 0.43 lofrc5 + 0.065 dem2 +
0.7 sprrc1
              - 0.2 lofc1 - 0.14 lonc3 + 0.23 lonrc3 - 0.29 imp + 0.26
lofrc3
              - 0.4 lofrc1 - 0.13 lonc1 + 0.25 lofrc2 + 0.05 lonrc6
              + 0.04 sprrc4
```

Rule 20: [91 cases, mean 80.366, range 0 to 232.76, est err 42.656]

```
    if
        cc > 3
        lofrc4 <= 111
        lonrc6 <= 26
        sprrc3 > 20
        curv2 <= 500
        dem2 > 55
        nsslp2 <= 105
    then
        Tba = -447 + 3.5 lonc2 + 1.2 cc - 1.39 lonrc4 - 2.15 lonc1 +
4.1 lonrc1
              + 2.16 lofrc3 + 0.84 lofc3 - 0.48 lonrc6 - 0.34 lonc3 -
0.1 lofrc4
              + 0.07 lofrc5
```

Rule 21: [136 cases, mean 82.048, range 0 to 259.06, est err 50.103]

```
    if
        cc > 50
        imp <= 6
        lofrc4 <= 105
        lonrc6 <= 26
        sprrc3 > 20
        ewslp2 > 468
        curv2 > 500
        curv2 <= 568
        dem2 > 260
    then
        Tba = 617.82 - 20 sprrc2 + 9.82 sprrc3 - 2.97 sprc1 - 1.1
lonrc6
              - 1.37 lonc1 - 0.65 sprrc4 + 0.56 sprrc6 + 0.62 sprc2
              - 0.17 sprrc5 + 0.09 lofc3 + 0.05 curv2 + 0.06 lofrc5 +
0.05 lonc2
              + 0.04 ewslp2 - 0.04 sprc3 + 0.02 cc
```

Rule 22: [40 cases, mean 82.741, range 0 to 326.38, est err 57.343]

```
    if
        cc > 3
        dem2 <= 55
    then
        Tba = -289.042 + 9.434 dem2 - 12.09 lonrc6 + 4.61 lofrc6 +
0.37 lonc2
              - 0.19 lonrc4 + 0.11 cc + 0.2 lofc3 + 0.11 lofrc5 + 0.4
lonrc1
              - 0.15 lonc1 + 0.18 lofrc3 - 0.06 sprrc5 - 0.07 sprc3 -
0.05 lonc3
              + 0.2 sprrc1 - 0.06 lofc1 - 0.03 lonrc5 - 0.09 lonrc2 -
```

```
0.07 imp
  Rule 23: [235 cases, mean 86.979, range 0 to 286.89, est err 43.562]

    if
        cc > 50
        lofc2 <= 132
        lofc3 <= 122
        lonc1 <= 92
        lofrc4 > 83
        lonrc2 <= 35
        lonrc6 <= 26
        sprrc3 > 20
        sprrc4 > 63
        ewslp2 > 468
        curv2 > 500
        curv2 <= 568
        dem2 <= 260
    then
        Tba = -76.782 - 7.54 lofc3 - 5.44 lofrc5 - 5.95 lonrc6 - 6.22
sprrc4
              + 6.18 sprc2 + 2.15 sprrc6 + 2.36 lofc2 + 1.06 curv2 +
3.06 sprrc3
              + 0.242 dem2 + 0.84 lonc3 + 0.97 sprc1 - 1.12 lonrc3 -
0.59 lonrc4
              + 0.54 lonrc5 + 0.4 lonc2 + 0.56 lonc1 - 0.34 lofc1 +
0.3 lofrc4
              - 0.24 sprc3 - 0.15 sprrc5 + 0.31 lonrc2 + 0.4 sprrc1 +
0.05 cc
              - 0.3 lofrc1

  Rule 24: [64 cases, mean 87.802, range 0 to 267.39, est err 50.680]

    if
        cc > 50
        lonc1 <= 92
        lofrc4 <= 111
        sprrc3 > 20
        curv2 > 568
    then
        Tba = -28.63 + 1.54 lonc2 + 0.99 cc - 2.23 lonc1 - 1.29 lonrc6
              - 0.91 lonc3 + 2.7 lonrc1 + 1.19 lofrc3 - 0.39 lofrc5

  Rule 25: [91 cases, mean 90.137, range 0 to 270.22, est err 51.960]

    if
        cc > 50
        imp <= 6
        lonrc2 > 35
        lonrc6 <= 26
        sprrc3 > 20
        curv2 > 500
        dem2 > 55
    then
        Tba = 538.65 - 10.18 lonrc6 - 18.74 lonrc2 + 5.33 lofrc6 +
10.1 lonrc1
              - 6.02 sprrc3 + 0.55 sprc1 - 0.51 sprrc4 + 0.33 lofc3
              + 0.23 lofrc5 + 0.3 lofrc4 - 0.36 lofc2 - 0.2 lonc3 +
0.32 sprc2
              - 0.2 lofc1 - 0.5 lofrc1 + 0.11 sprc3 + 0.05 cc - 0.06
lonrc3

  Rule 26: [118 cases, mean 91.801, range 0 to 285.68, est err 54.338]

    if
        di > 52
        imp <= 3
        lonc3 <= 119
        lofrc4 > 51
        lonrc5 <= 56
        lonrc6 <= 26
        sprrc3 > 20
        curv2 <= 500
        dem2 > 65
        nsslp2 > 105
    then
        Tba = -485.312 - 18 imp + 5.6 sprrc5 + 5.47 sprc3 + 3.95
lonrc5
              + 4.18 sprrc4 - 7.54 sprrc3 - 3.42 sprc1 + 2.12 lonc2 -
4.58 sprc2
              - 2.26 lonc3 + 4.3 lofrc3 - 2.99 lonc1 - 2.13 lonrc6 +
```

```
2.13 lofc1
          + 1.73 lofc3 - 2.46 lonrc3 - 1.52 lofrc4 + 3.3 lonrc1 -
0.8 nsslp2
          + 0.084 dem2 - 0.52 di

  Rule 27: [136 cases, mean 92.644, range 0 to 441.12, est err 65.591]

    if
        cc > 3
        imp <= 14
        sprc1 <= 62
        sprc2 <= 96
        lonrc6 > 26
        dem2 > 216
    then
        Tba = 1077.8 - 14.27 lonrc2 - 4.7 sprrc4 - 3.43 lonc2 + 3.64
sprc1
              - 2.19 lofrc5 - 2.51 lofc3 + 1.35 lofc1 + 0.74 ewslp2
              + 0.46 sprrc5 - 1.7 lonrc1 - 0.63 lofrc6 + 0.49 lonc3 +
0.65 lonc1
              - 0.39 lofrc4 + 0.1 cc

  Rule 28: [388 cases, mean 93.967, range 0 to 375.37, est err 47.488]

    if
        cc > 50
        imp <= 6
        lonrc2 <= 35
        lonrc6 <= 26
        sprrc3 > 20
        sprrc4 <= 63
        curv2 > 500
        curv2 <= 568
    then
        Tba = 259.666 - 6.69 sprrc4 + 6.22 sprc1 + 5.39 lofc3 + 4.13
lofrc5
              - 4.46 lonc3 - 4.43 lonrc6 - 4.15 sprc3 - 2.84 sprrc5 -
3.54 lofc1
              + 3.16 lofrc4 - 3.14 sprrc6 - 3.37 lofc2 + 1.28 cc +
3.91 sprc2
              - 1.79 lonrc5 - 2.94 lonrc3 + 3.31 lonrc2 + 4.4 sprrc1
              - 3.4 lofrc1 + 0.9 sprrc3 + 0.24 curv2 + 0.35 lonc1 +
0.039 dem2
              + 0.021 lofdy + 0.1 ewslp2

  Rule 29: [174 cases, mean 94.815, range 0 to 297.19, est err 48.170]

    if
        cc > 3
        lofrc4 > 111
        lonrc6 <= 26
        sprrc3 > 20
    then
        Tba = -445.53 + 2.45 lonc3 + 1.86 lonc2 + 0.75 cc - 1.32 sprc3
              - 1.48 sprrc6 + 1.08 lofc3 + 0.67 lofrc5 - 0.46 lonrc6
              - 0.38 sprrc5 - 0.54 lonc1 + 1.2 lonrc1 - 0.41 lofc1 -
0.26 lonrc4
              + 0.55 lofrc3 + 1 sprrc1 - 0.63 lonrc2 - 0.46 imp - 0.18
lonrc5
              + 0.22 lofrc6 + 0.27 sprc2 + 0.14 lofrc4 - 0.5 lofrc1
              + 0.32 lofrc2 - 0.11 sprrc4 - 0.27 sprrc2 - 0.13 lofc2
              - 0.15 sprrc3 + 0.1 lonrc3 + 0.01 dem2 + 0.04 sprc1

  Rule 30: [57 cases, mean 96.701, range 0 to 441.12, est err 65.987]

    if
        cc > 3
        cc <= 41
        imp <= 14
        sprc1 <= 62
        lonrc6 > 26
        dem2 > 216
    then
        Tba = 1029.938 - 6.92 lofrc5 - 6.52 lofc3 + 4.87 lofc1 + 3.06
sprrc5
              - 3.9 lofrc6 - 3.25 sprc1 + 2.94 sprc3 - 1.62 lonc2 +
1.36 sprrc4
              + 1.41 sprrc6 - 1.49 lofc2 - 3 lonrc1 + 0.51 cc + 0.85
lonc3
              - 1.64 sprrc3 + 1.14 lonc1 - 1.18 sprc2 - 0.68 lofrc4
              + 0.51 lonrc3 + 0.26 lonrc6 - 1 lofrc1 + 0.1 ewslp2 +
```

```
0.018 dem2

  Rule 31: [124 cases, mean 104.335, range 0 to 330.1, est err 46.170]

    if
        cc > 50
        imp <= 6
        lonc1 <= 92
        lofrc4 <= 111
        sprrc3 > 20
        ewslp2 <= 468
        curv2 > 500
        curv2 <= 568
    then
        Tba = 9.35 - 5.78 sprrc4 - 5.32 lonrc6 + 4.75 sprc1 + 3.86
lofrc4
              - 4.79 lofc2 + 3.46 lofc3 - 2.78 lonc3 + 4.82 sprc2 +
2.14 lofrc5
              + 1.03 cc - 2.22 lofc1 - 6.4 lofrc1 + 3.7 sprrc1 + 0.36
sprc3
              - 0.09 lonrc3

  Rule 32: [259 cases, mean 104.431, range 0 to 264.15, est err
46.799]

    if
        di <= 52
        imp <= 3
        lonc1 <= 92
        lonc3 <= 119
        sprc1 <= 56
        lofrc4 > 51
        lofrc4 <= 111
        lonrc5 <= 56
        sprrc3 > 20
        curv2 <= 500
        nsslp2 > 105
    then
        Tba = -383.381 + 5.42 sprrc5 - 13.21 imp + 5.36 sprrc4 - 9.39
sprrc3
              + 5.04 sprc3 - 6.65 sprc2 + 2.9 lonrc5 - 3.39 sprc1 +
2.94 lofc1
              + 1.83 lonc2 - 1.98 lofrc4 - 2.19 lonc1 + 1.27 lofc3 +
2.31 lofrc3
              - 1.81 lonrc3 + 2.4 lonrc1 - 0.85 nsslp2 - 0.6 lofrc6
              - 0.47 sprrc6 + 0.093 dem2 - 0.38 di

  Rule 33: [133 cases, mean 104.437, range 0 to 319.41, est err
41.132]

    if
        cc > 50
        imp <= 6
        lofc3 <= 122
        lofrc4 <= 83
        lonrc2 <= 35
        lonrc6 <= 26
        sprrc3 > 20
        sprrc4 > 63
        curv2 > 500
        curv2 <= 568
        dem2 <= 260
        nsslp2 > 121
    then
        Tba = -394.106 + 5.53 sprrc6 - 2.55 sprrc5 + 1.69 lonc3 + 1.37
lofc3
              - 1.03 lofrc5 + 0.95 lonrc5 - 1 sprrc4 - 0.79 lonrc4 +
0.88 sprc1
              - 1.13 lonrc3 + 0.53 lonc2 + 0.75 lonc1 + 0.47 sprc2 +
0.16 curv2
              + 0.038 dem2

  Rule 34: [41 cases, mean 105.995, range 0 to 253.51, est err 65.469]

    if
        cc > 50
        imp <= 6
        lofc2 <= 132
        lofc3 > 122
        lonrc2 <= 35
        lonrc6 <= 26
```

```
        sprrc3 > 20
        sprrc4 > 63
        curv2 > 500
        dem2 <= 260
    then
        Tba = -30.634 + 4.27 lofc3 + 3.04 lofrc5 - 2.81 lonrc6 - 2.52
lonc3
              - 1.68 sprrc5 - 2.11 sprc3 - 2.02 sprrc4 - 2.19 lofc1
              - 1.25 lonrc5 + 1.61 lofrc4 + 0.76 cc + 2.26 sprc2 - 1.8
lofc2
              + 4 sprrc1 + 0.96 sprc1 - 2.9 lofrc1 - 1.72 sprrc2 -
0.68 lonrc3
              + 0.31 lonc2 - 0.74 imp + 0.69 lofrc3 + 1 lonrc1 + 0.29
lofrc6
              + 0.7 lofrc2 - 0.32 lonc1 + 0.26 sprrc6 + 0.13 curv2 +
0.039 dem2
              + 0.11 ewslp2 + 0.26 sprrc3 + 0.016 lofdy - 0.07 lonrc4
              - 0.09 nsslp2 - 0.12 lonrc2 - 0.08 di

  Rule 35: [182 cases, mean 108.837, range 0 to 281.65, est err
47.562]

    if
        lonrc4 > 113
        lonrc6 <= 26
        sprrc3 <= 20
    then
        Tba = -1021.81 + 8.58 lofc3 + 6.12 lofrc5 - 2.77 lonrc5 - 4.23
lofc2
              + 11.3 sprrc1 + 2.79 lofrc4 - 2.21 sprc1 + 0.44 lonc2
              - 0.28 sprrc5 + 0.18 cc - 0.31 lonrc6 - 0.32 sprc3 -
0.24 lonrc4
              - 0.34 lofc1 - 0.26 lonc3 + 0.7 lonrc1 - 0.34 imp - 0.39
lonrc2
              + 0.18 lofrc6 + 0.27 lofrc3 + 0.22 sprc2 - 0.17 lonc1
              - 0.13 sprrc4 - 0.4 lofrc1 - 0.24 sprrc2 + 0.22 lofrc2
              - 0.12 sprrc3

  Rule 36: [24 cases, mean 109.975, range 0 to 250.87, est err 69.480]

    if
        cc > 3
        sprc1 <= 62
        lofrc2 > 28
        lofrc4 <= 75
        lofrc5 <= 73
        lonrc1 <= 41
        lonrc6 > 26
    then
        Tba = -36.81 + 1.09 cc - 2.11 sprc1 + 1.7 lofc3 - 3.55 lonrc2
              + 1.16 lonrc6 + 0.95 lofrc5 + 1.83 lonrc3 + 0.97 sprrc4
              - 0.7 lonrc5 + 3.3 sprrc1 - 0.83 sprc3 - 0.63 sprrc5 +
1.31 lofrc3
              - 1.38 imp - 2.6 lofrc1 + 1.8 lonrc1 + 0.67 sprrc6 -
0.49 lonc3
              - 0.7 lofc2 + 1.38 lofrc2 - 0.84 sprrc3 - 0.58 lonc1 +
0.3 lonc2
              - 0.43 lofrc6 + 0.14 ewslp2 + 0.037 dem2 - 0.17 nsslp2 -
0.15 di
              + 0.09 lofc1 + 0.011 lofdy + 0.12 sprc2

  Rule 37: [26 cases, mean 111.056, range 0 to 314.1, est err 65.778]

    if
        lonc1 <= 92
        lofrc4 > 51
        lonrc5 <= 56
        sprrc3 > 20
        curv2 <= 500
        dem2 > 55
        dem2 <= 65
    then
        Tba = 2586.39 - 32.66 lonrc6 - 7.33 cc - 9.7 lonrc1 - 3.23
lofc2
              - 3.39 nsslp2


  Rule 38: [180 cases, mean 111.491, range 0 to 315.44, est err
```

```
50.594]

    if
        di <= 52
        imp <= 3
        lonc3 <= 119
        sprc1 > 56
        lofrc4 <= 111
        lonrc5 <= 56
        lonrc6 <= 26
        curv2 <= 500
    then
        Tba = -1012.859 + 12.17 lofc1 - 9.55 lofrc4 + 3.46 lonc3 +
2.95 sprrc5
              + 2.82 lonrc5 + 4.32 lofc2 + 2.88 sprrc4 - 4.47 sprrc3
              + 2.48 sprc3 + 1.74 lonc2 - 3.52 sprc2 - 1.57 lofrc5 -
1.91 sprc1
              - 3.16 imp - 0.93 lofrc6 - 0.72 sprrc6 - 0.53 lonc1 -
0.53 nsslp2
              + 0.55 lofrc3 + 0.3 lofc3 - 0.43 lonrc3 + 0.063 dem2 +
0.6 lonrc1
              - 0.09 di

  Rule 39: [122 cases, mean 112.121, range 0 to 482.18, est err
61.290]

    if
        cc > 3
        imp <= 6
        lonc1 > 88
        lonc1 <= 92
        lofrc4 <= 111
        lonrc5 > 56
        lonrc6 <= 26
        curv2 <= 500
        dem2 > 55
        posidx2 > 19
        nsslp2 > 105
    then
        Tba = -501.37 - 26.17 lonc1 + 10.08 lonc2 + 7.41 lonrc6 + 17.6
lonrc1
              - 5.36 sprc1 + 3.95 sprc2 + 4.15 sprrc3 + 2.44 lofc1 -
2.19 lofrc4
              + 1.54 sprrc6 + 4.4 sprrc1 - 1.09 sprc3 + 1.08 lofc2 +
0.71 lofc3

  Rule 40: [201 cases, mean 114.875, range 0 to 446.42, est err
52.626]

    if
        cc > 3
        imp <= 3
        lonc1 <= 92
        lonc3 > 119
        lofrc4 > 51
        sprrc3 > 20
        curv2 <= 500
        dem2 > 65
    then
        Tba = -121.61 + 1.66 cc + 0.89 sprrc5 + 2.05 lofrc3 + 0.97
sprc3
              - 1.46 sprrc3 + 0.7 sprrc4 - 0.99 sprc2 - 0.38 lonrc4 -
0.47 sprc1
              + 0.55 lonc1 - 0.33 lofrc5 + 0.19 lonc2 + 0.19 lofrc2

  Rule 41: [21 cases, mean 129.163, range 0 to 325.61, est err 61.650]

    if
        lonc2 > 139
        lofrc1 > 35
        lonrc4 <= 113
        sprrc3 <= 20
    then
        Tba = 1082.61 - 16.72 lonrc6 - 4.87 lonc2 + 3.69 lonrc4 - 8.3
lofrc1
              - 1.6 sprrc5 + 2.03 lofrc6 + 0.65 cc - 3 lonrc1 - 0.15
posidx2


  Rule 42: [99 cases, mean 140.807, range 0 to 302.1, est err 50.824]
```

```
    if
        cc > 3
        lonc2 <= 139
        lofrc1 > 35
        lonrc4 <= 113
        lonrc6 <= 26
        sprrc3 <= 20
    then
        Tba = 634.66 + 5.69 lonrc4 - 5.36 sprrc5 - 4.74 lonc2 + 5.93
lofrc6
              + 4.21 lonc3 - 12.08 lofrc2 - 13.2 lofrc1 + 1.46 cc -
0.83 lonrc6
              - 0.56 posidx2 - 0.133 dem2 - 0.49 sprc3 - 0.8 lonrc1
              - 0.18 lonrc5 + 0.21 sprc1

  Rule 43: [81 cases, mean 156.180, range 0 to 318.37, est err 43.603]

    if
        cc > 3
        lofrc1 <= 35
        lonrc4 <= 113
        lonrc6 <= 26
        sprrc3 <= 20
    then
        Tba = -170.87 - 5.62 lonrc5 + 5.35 sprc1 + 3.41 lonc2 - 2.79
sprrc5
              + 3.46 lofc3 - 2.65 sprc3 - 1.63 lofrc4 - 3.38 lofrc2 +
0.3 cc
              + 0.39 lonrc4


Evaluation on training data (7750 cases):

    Average  |error|              30.758
    Relative |error|                0.52
    Correlation coefficient         0.68


        Attribute usage:
          Conds  Model

           86%    71%    cc
           60%    69%    lonrc6
           45%    55%    sprrc3
           41%    58%    lofrc4
           35%    39%    imp
           33%    77%    sprc1
           29%    27%    dem2
           25%     9%    curv2
           22%    59%    lonc1
           11%    27%    nsslp2
           10%    48%    lonrc5
            9%    56%    lonrc2
            9%    61%    sprrc4
            7%    52%    lonc3
            5%    30%    ewslp2
            5%     6%    di
            4%    72%    lofrc5
            4%    55%    lofc3
            4%    43%    lonrc4
            3%    54%    lofc2
            3%    29%    lonrc1
            3%    22%    posidx2
            2%    36%    lofrc2
            2%    58%    sprc2
            2%    25%    lofrc1
            1%    37%    lonc2
                  74%    sprc3
                  72%    sprrc5
                  60%    lonrc3
                  57%    sprrc1
                  55%    lofc1
                  54%    sprrc6
                  50%    lofrc3
                  30%    sprrc2
                  26%    lofrc6
                   4%    lofdy
```

Appendix 18. Cubist 2.05 total basal area model for USGS zone 57 (mountain region) with National Elevation Dataset DEM

```
Cubist [Release 2.05]  Wed May 28 10:06:34 2008

    Target attribute `Tba'

Read 3820 cases (59 attributes) from z57tba.data

Model:

  Rule 1: [438 cases, mean 7.276, range 0 to 241.01, est err 9.238]

    if
        cc <= 25
    then
        Tba = -88.09 + 1.03 lofc3 + 0.46 lofrc5 - 0.38 sprc1 + 0.19
sprrc4
              - 0.33 lofc2 - 0.9 lonrc2 + 0.15 cc + 0.31 lofrc6 + 0.13
lofrc4
              + 0.16 lofc1 + 0.4 lonrc3 - 0.7 lofrc1 + 0.13 sprrc6 +
0.2 sprrc2
              + 0.05 lonrc4

  Rule 2: [298 cases, mean 56.227, range 0 to 379.81, est err 53.431]

    if
        cc > 25
        lonrc3 > 22
    then
        Tba = -229.97 - 4.83 lofrc4 + 6.78 lofc2 - 2.93 sprc3 - 2.22
sprrc5
              + 8.5 lofrc3 - 1.55 lonrc5 + 0.85 lonrc4 + 0.66 lofc3
              + 0.27 lofrc5 + 0.16 cc + 0.22 lofc1 + 0.11 sprc1 - 0.5
lonrc2
              + 0.12 sprrc6 - 0.21 sprrc3 + 0.09 lofrc6 - 0.04 sprrc4
              + 0.1 lonrc3

  Rule 3: [3084 cases, mean 110.699, range 0 to 489.6, est err 52.316]

    if
        cc > 25
        lonrc3 <= 22
    then
        Tba = -275.038 + 6.17 lofc3 - 5.08 lofc2 + 3.82 lofc1 + 2.03
cc
              + 2.41 lofrc5 + 1.53 sprrc4 + 1.48 lofrc4 + 1.34 sprc3
              - 2.67 lonc1 - 2.86 sprrc3 + 1.11 lonrc4 - 2.28 sprc2
              + 1.24 sprrc6 - 4.9 lonrc2 - 2.6 lofrc3 + 1.13 lonrc5 -
3.9 lofrc1
              + 0.29 sprc1 + 0.053 ewslp - 0.09 sprrc5


Evaluation on training data (3820 cases):

    Average  |error|              46.446
    Relative |error|                0.75
    Correlation coefficient         0.55


        Attribute usage:
          Conds  Model

          100%   100%    cc
           89%    19%    lonrc3
                 100%    lofc1
                 100%    lofc2
                 100%    lofc3
                 100%    sprc1
                 100%    lofrc4
                 100%    lofrc5
                 100%    lonrc2
                 100%    lonrc4
                 100%    sprrc4
                 100%    sprrc6
                  92%    lofrc1
                  89%    sprc3
```

```
89%     lofrc3
89%     lonrc5
89%     sprrc3
89%     sprrc5
81%     ewslp
81%     lonc1
81%     sprc2
19%     lofrc6
11%     sprrc2
```

Appendix 19. Cubist 2.05 total basal area model for USGS zone 57 (mountain region) with LiDAR DEM

```
Cubist [Release 2.05]  Wed May 28 10:08:01 2008

    Target attribute `Tba'

Read 3820 cases (59 attributes) from z57tba.data

Model:

  Rule 1: [438 cases, mean 7.276, range 0 to 241.01, est err 10.105]

    if
        cc <= 25
    then
        Tba = -135.152 + 1.56 lofc3 + 0.76 lofrc5 - 0.49 sprc1 + 0.42
lofc1
              - 0.52 lofc2 - 1.3 lonrc2 + 0.23 sprrc4 + 0.18 cc + 0.19
lonrc4
              + 0.3 lofrc6 - 0.19 lonc2 + 0.2 sprrc6 + 0.5 lonrc3 -
0.9 lofrc1
              + 0.12 lofrc4 + 0.3 sprrc2 + 0.15 lonc3 - 0.2 lofrc3 -
0.04 sprrc5
              + 0.012 ewslp2 - 0.01 sprdy + 0.04 posidx2 + 0.06 lonrc5
              + 0.02 slp2 - 0.01 nsslp2

  Rule 2: [298 cases, mean 56.227, range 0 to 379.81, est err 52.732]

    if
        cc > 25
        lonrc3 > 22
    then
        Tba = -604.33 - 4.25 lofrc4 + 5.44 lofc2 + 7.6 lofrc3 + 1.34
lofc3
              + 0.89 lonrc4 - 1.29 lonrc5 + 0.58 lofrc5 + 0.62 lofc1
              - 0.26 sprc1 + 0.29 sprrc6 + 0.19 cc + 0.21 sprrc4 - 1.1
lonrc2
              - 0.23 lonc2 - 0.4 lonrc6 - 0.38 sprrc3 - 0.1 sprrc5 -
0.8 lofrc1
              + 0.25 lonc1 + 0.4 sprrc1 + 0.006 dem2 + 0.016 ewslp2 -
0.11 sprc2
              - 0.02 nsslp2 + 0.04 posidx2

  Rule 3: [2770 cases, mean 109.300, range 0 to 489.6, est err 51.478]

    if
        lc in 11, 21, 41, 81, 90
        cc > 25
        lonrc3 <= 22
    then
        Tba = -138.402 + 5.17 lofc3 - 4.09 lofc2 + 2.67 lofc1 + 1.67
cc
              + 1.85 sprrc4 + 1.91 lofrc5 + 1.4 lofrc4 - 2.58 sprrc3
              - 5.3 lonrc2 - 2.24 sprc2 + 0.8 lonrc4 + 0.92 sprrc6 +
0.66 sprc3
              - 1.14 lonc1 - 1.5 lofrc3 - 0.48 sprc1 - 3 lofrc1 + 0.61
lonrc5
              - 0.25 sprrc5 - 0.33 lonc2 - 0.62 lonrc6 + 0.051 ewslp2
              + 0.01 dem2 + 0.6 sprrc1 + 0.06 posidx2 - 0.02 nsslp2

  Rule 4: [314 cases, mean 123.036, range 0 to 418.08, est err 55.610]

    if
        lc in 42, 43, 52, 71
        cc > 25
        lonrc3 <= 22
    then
        Tba = -6.72 + 3.07 lofc3 - 2.84 lofc2 + 1.98 lofc1 + 1.17 cc
              + 1.06 sprrc4 + 0.95 lofrc4 + 1.06 lofrc5 - 1.72 lonc1
              + 0.74 sprc3 + 0.7 lonrc4 - 1.53 sprrc3 - 1.48 sprc2 -
1.12 lofrc6
              - 2.8 lonrc2 + 0.55 sprrc6 - 1.4 lofrc3 + 0.68 lonrc5 -
2.2 lofrc1
              + 0.032 ewslp2
```

```
Evaluation on training data (3820 cases):

    Average  |error|              46.289
    Relative |error|                0.74
    Correlation coefficient         0.56


        Attribute usage:
          Conds  Model

          100%   100%    cc
           89%    11%    lonrc3
           81%           lc
                 100%    lofc1
                 100%    lofc2
                 100%    lofc3
                 100%    lofrc1
                 100%    lofrc3
                 100%    lofrc4
                 100%    lofrc5
                 100%    lonrc2
                 100%    lonrc4
                 100%    lonrc5
                 100%    sprrc4
                 100%    sprrc6
                 100%    ewslp2
                  92%    lonc2
                  92%    sprc1
                  92%    sprrc5
                  92%    posidx2
                  92%    nsslp2
                  89%    lonc1
                  89%    sprc2
                  89%    sprrc3
                  81%    sprc3
                  80%    lonrc6
                  80%    sprrc1
                  80%    dem2
                  20%    lofrc6
                  11%    sprdy
                  11%    lonc3
                  11%    sprrc2
                  11%    slp2
```

Appendix 20. Cubist 2.05 white pine basal area model for USGS zone 57 (mountain region) with National Elevation Dataset DEM

```
Cubist [Release 2.05]  Thu May 29 08:47:45 2008

    Target attribute `B129'

Read 3820 cases (59 attributes) from z57b129.data

Model:

  Rule 1: [97 cases, mean 0.535, range 0 to 34.13, est err 1.165]

    if
        lc in 21, 42, 43
        cc <= 66
        sprrc6 > 33
    then
        B129 = 13.008 - 0.012 sprdy + 0.024 cc

  Rule 2: [1880 cases, mean 1.381, range 0 to 337.57, est err 1.382]

    if
        sprc2 <= 94
    then
        B129 = 0

  Rule 3: [159 cases, mean 1.590, range 0 to 67.88, est err 1.804]

    if
        nsslp > 229
        lonc3 <= 114
        sprc2 <= 94
    then
        B129 = -7.15 + 0.05 sprc3 + 0.02 sprrc5

  Rule 4: [3304 cases, mean 3.110, range 0 to 276.92, est err 3.112]

    if
        lc in 11, 22, 24, 41, 52, 71, 81, 82, 90
    then
        B129 = 0

  Rule 5: [3209 cases, mean 3.726, range 0 to 337.57, est err 3.729]

    if
        lonc3 <= 114
    then
        B129 = 0

  Rule 6: [249 cases, mean 5.461, range 0 to 337.57, est err 5.505]

    if
        cc <= 81
        dem <= 926
        sprrc6 <= 33
    then
        B129 = 0

  Rule 7: [55 cases, mean 13.975, range 0 to 137, est err 14.637]

    if
        lc in 21, 42, 43
        cc > 66
        sprrc6 > 33
    then
        B129 = 8.614 - 0.012 sprdy - 0.038 lofrc4 + 0.05 lofc1 + 0.025
cc
               + 0.03 lofc2 - 0.01 lofrc5

  Rule 8: [91 cases, mean 15.000, range 0 to 224.13, est err 16.540]

    if
        lc in 21, 42, 43
        sprdy > 757
        nsslp <= 104
        sprrc6 <= 33
    then
```

```
        B129 = -52.477 - 1.961 lofrc4 + 2.37 lofc1 + 1.64 sprc1 - 1.2
sprrc6
               + 1.37 lofc2 - 0.546 sprrc4 - 0.59 lofrc5 - 1.27 sprrc3
               - 0.33 lofc3

  Rule 9: [31 cases, mean 17.674, range 0 to 80.09, est err 19.700]

    if
        lc in 21, 42, 43
        sprdy > 757
        cc > 81
        dem <= 926
        nsslp > 104
        lonc2 > 158
        sprc2 <= 121
        sprrc6 <= 33
    then
        B129 = -137.325 - 3.44 lofrc6 - 1.827 sprrc4 + 1.81 lofrc5 +
1.8 sprc1
               + 1.64 sprc2 - 0.49 sprrc6 - 0.161 lofrc4 - 0.39 sprrc3
               + 0.19 lofc1 - 0.0116 dem + 0.18 lofc2

  Rule 10: [32 cases, mean 23.567, range 0 to 137, est err 25.573]

    if
        lc in 21, 42, 43
        cc > 66
        lonrc4 <= 143
        sprrc6 > 33
    then
        B129 = -90.97 + 1.087 sprrc5 - 0.813 lofrc4 + 0.98 lofc1 +
0.56 lofc2
               - 0.25 lofrc5 - 0.19 sprrc6 + 0.14 sprc1 - 0.14 lofc3
               - 0.19 sprrc3

  Rule 11: [41 cases, mean 35.142, range 0 to 300.36, est err 43.743]

    if
        lc in 21, 42, 43
        dem > 926
        nsslp > 104
        sprc2 <= 121
        sprrc6 <= 33
    then
        B129 = 384.947 - 2.257 cc - 2.45 sprrc6 - 0.0919 dem + 0.04
sprc1
               - 0.07 sprrc3 + 0.014 sprrc4

  Rule 12: [66 cases, mean 36.094, range 0 to 169.78, est err 37.824]

    if
        lc in 21, 42, 43
        sprdy > 757
        cc > 81
        dem <= 926
        nsslp > 104
        lonc2 <= 158
        sprrc4 <= 66
    then
        B129 = -480.727 + 5.23 lonc2 - 4.72 sprrc6 - 3.43 lonrc4 +
3.94 lofrc6
               - 0.1572 dem + 2.55 sprc2 - 0.138 sprrc4 + 0.1 sprc1

  Rule 13: [32 cases, mean 59.962, range 0 to 245.74, est err 50.690]

    if
        sprdy > 757
        cc > 81
        dem <= 926
        nsslp > 104
        lonc2 <= 158
        sprc2 <= 121
        sprrc4 > 66
        sprrc6 <= 33
    then
        B129 = -5664.879 - 32.486 sprrc4 + 40.88 sprc2 + 20.1 sprrc6
               + 18.42 lonc3 + 14.73 lonrc5 + 1.726 sprdy - 6.18 lonc2
               - 7.08 lofc3 - 0.46 lonrc4 - 0.0183 dem
```

```
Rule 14: [35 cases, mean 67.699, range 0 to 337.57, est err 66.333]

    if
        lc in 21, 42, 43
        sprdy <= 757
        sprc2 <= 121
        sprrc6 <= 33
    then
        B129 = 41.59

  Rule 15: [30 cases, mean 97.254, range 0 to 250.19, est err 66.486]

    if
        lc in 21, 42, 43
        sprc2 > 121
        sprrc6 <= 33
    then
        B129 = 413.147 - 8.752 sprrc5 + 7.28 sprrc6 - 0.2126 dem +
1.67 posidx
               - 0.204 lofrc4 + 0.24 lofc1 + 0.23 lofc2


Evaluation on training data (3820 cases):

    Average  |error|              5.150
    Relative |error|               0.48
    Correlation coefficient        0.59


        Attribute usage:
          Conds  Model

           41%           lc
           36%           lonc3
           24%     1%    sprc2
            7%     3%    sprrc6
            6%     2%    cc
            5%           nsslp
            5%     2%    dem
            3%     2%    sprdy
            1%     1%    lonc2
            1%     3%    sprrc4
                   1%    lonrc4
                   3%    sprc1
                   3%    lofc1
                   3%    lofc2
                   3%    lofrc4
                   2%    sprrc5
                   2%    lofrc5
                   2%    sprrc3
                   2%    sprc3
                   2%    lofc3
                   1%    lofrc6
```

Appendix 21. Cubist 2.05 white pine basal area model for USGS zone 57 (mountain region) with LiDAR DEM

```
Cubist [Release 2.05]  Thu May 29 08:46:34 2008

    Target attribute `B129'

Read 3820 cases (59 attributes) from z57b129.data

Model:

  Rule 1: [97 cases, mean 0.535, range 0 to 34.13, est err 1.165]

    if
        lc in 21, 42, 43
        cc <= 66
        sprrc6 > 33
    then
        B129 = 13.008 - 0.012 sprdy + 0.024 cc

  Rule 2: [3304 cases, mean 3.110, range 0 to 276.92, est err 3.329]

    if
        lc in 11, 22, 24, 41, 52, 71, 81, 82, 90
    then
        B129 = -0.174 + 0.057 sprrc4 - 0.06 sprc1

  Rule 3: [823 cases, mean 3.307, range 0 to 337.57, est err 3.316]

    if
        cc <= 81
        dem2 <= 927
    then
        B129 = 0

  Rule 4: [55 cases, mean 13.975, range 0 to 137, est err 14.652]

    if
        lc in 21, 42, 43
        cc > 66
        sprrc6 > 33
    then
        B129 = 11.329 - 0.012 sprdy + 0.025 cc

  Rule 5: [91 cases, mean 15.000, range 0 to 224.13, est err 16.502]

    if
        lc in 21, 42, 43
        sprdy > 757
        sprrc6 <= 33
        nsslp2 <= 104
    then
        B129 = -1236.985 + 7.14 sprc3 + 4.782 sprrc5 - 1.989 sprrc4
               + 2.54 sprrc6 + 2.55 sprc2

  Rule 6: [32 cases, mean 23.567, range 0 to 137, est err 25.462]

    if
        lc in 21, 42, 43
        cc > 66
        lonrc4 <= 143
        sprrc6 > 33
    then
        B129 = -473.528 + 2.521 sprrc5 + 2.31 sprc3 + 1.05 sprrc6 -
0.645 sprrc4
               + 0.88 sprc2

  Rule 7: [39 cases, mean 24.778, range 0 to 119.06, est err 23.302]

    if
        lc in 21, 42, 43
        sprdy > 757
        cc > 86
        sprc2 <= 121
        curv2 <= 529
        dem2 <= 927
        nsslp2 > 104
    then
```

```
        B129 = 192.927 - 0.236 ewslp2 - 2.73 lofrc3 - 0.46 sprrc6 +
0.138 sprrc5
               - 0.0118 dem2 + 0.02 sprc3

  Rule 8: [26 cases, mean 31.959, range 0 to 149.45, est err 31.252]

    if
        lc in 21, 42, 43
        sprdy > 757
        cc > 81
        cc <= 86
        sprrc6 <= 33
        curv2 <= 529
        dem2 <= 927
        nsslp2 > 104
    then
        B129 = 315.427 - 8.586 cc + 1.093 slp2 + 0.355 ewslp2 - 0.557
nsslp2
               + 0.768 sprrc5 + 2.17 sprc2 + 0.91 sprc3 - 0.0751 dem2
               - 0.285 sprrc4 + 0.02 sprrc6

  Rule 9: [41 cases, mean 35.142, range 0 to 300.36, est err 45.728]

    if
        lc in 21, 42, 43
        sprc2 <= 121
        sprrc6 <= 33
        dem2 > 927
        nsslp2 > 104
    then
        B129 = 483.207 - 6.74 sprrc6 - 3.44 sprc3 - 2.672 cc + 2.238
sprrc5
               + 2.51 lofc3 - 1.071 sprrc4 - 0.1011 dem2 + 1.39 sprc2

  Rule 10: [62 cases, mean 47.774, range 0 to 245.74, est err 43.059]

    if
        lc in 21, 42, 43
        sprdy > 757
        cc > 81
        sprc2 <= 121
        sprrc6 <= 33
        curv2 > 529
        dem2 <= 927
        nsslp2 > 104
    then
        B129 = -2315.974 + 12.5 sprc3 + 8.411 sprrc5 - 3.906 sprrc4
               + 5.08 sprrc6 + 6.75 sprc2 - 0.47 lonrc4

  Rule 11: [35 cases, mean 67.699, range 0 to 337.57, est err 68.744]

    if
        lc in 21, 42, 43
        sprdy <= 757
        sprc2 <= 121
        sprrc6 <= 33
    then
        B129 = 414.284 - 15.8 lonrc6 - 5.413 cc + 9.38 lonrc5 - 4.74
sprrc6
               + 0.28 sprc3 + 0.2 sprrc5 - 0.058 sprrc4 + 0.09 sprc2

  Rule 12: [30 cases, mean 97.254, range 0 to 250.19, est err 73.543]

    if
        lc in 21, 42, 43
        sprc2 > 121
        sprrc6 <= 33
    then
        B129 = 363.111 - 8.29 sprrc6 - 0.1523 dem2


Evaluation on training data (3820 cases):

    Average  |error|              5.333
    Relative |error|               0.50
    Correlation coefficient        0.60
```

```
Attribute usage:
  Conds  Model

   82%           lc
   24%      5%    cc
   21%      3%    dem2
   10%      8%    sprrc6
    6%            nsslp2
    5%      3%    sprdy
    4%      6%    sprc2
    3%            curv2
            1%    lonrc4
           77%    sprrc4
           71%    sprc1
            7%    sprc3
            7%    sprrc5
            1%    ewslp2
```

Appendix 22. Cubist 2.05 yellow birch basal area model for USGS zone 57 (mountain region) with National Elevation Dataset DEM

```
Cubist [Release 2.05]  Thu May 29 08:44:53 2008

    Target attribute `B371'

Read 3820 cases (59 attributes) from z57b371.data

Model:

  Rule 1: [1973 cases, mean 0.008, range 0 to 8.65, est err 0.008]

    if
        dem <= 840
    then
        B371 = 0

  Rule 2: [1934 cases, mean 0.099, range 0 to 43.79, est err 0.099]

    if
        lofrc4 > 78
    then
        B371 = 0

  Rule 3: [3786 cases, mean 0.863, range 0 to 229.59, est err 0.864]

    if
        dem <= 1616
    then
        B371 = 0

  Rule 4: [43 cases, mean 1.622, range 0 to 16.73, est err 2.289]

    if
        dem > 1151
        dem <= 1616
        lofrc4 <= 78
        lonrc5 <= 55
    then
        B371 = 4.256 - 0.076 lofrc4

  Rule 5: [118 cases, mean 2.240, range 0 to 119.55, est err 2.397]

    if
        dem > 840
        nsslp <= 134
        ewslp <= 216
        lofc2 <= 109
        lofrc4 <= 78
        lonrc5 > 55
    then
        B371 = -0.113 - 0.036 lonrc4 + 0.072 lonc1 - 0.08 lonrc3 +
0.0006 dem
               - 0.008 lofc1 + 0.009 lofrc6

  Rule 6: [36 cases, mean 3.118, range 0 to 71.65, est err 3.785]

    if
        dem > 840
        slp > 95
        nsslp > 134
        nsslp <= 243
        lofc2 <= 107
        lonrc5 > 55
        sprrc1 <= 36
    then
        B371 = -21.026 - 0.071 lofrc4 + 0.078 lofrc5 + 0.0049 dem +
0.109 lonc3
               + 0.074 lonrc5 - 0.022 lofc1 + 0.026 lofrc6

  Rule 7: [800 cases, mean 3.773, range 0 to 229.59, est err 3.828]

    if
        dem > 840
        dem <= 1616
        lofrc4 <= 78
        lonrc5 > 55
```

```
    then
        B371 = 1.106 + 0.054 lonc1 - 0.021 lonrc4 - 0.07 lonrc3 -
0.009 lonc2

  Rule 8: [49 cases, mean 4.384, range 0 to 43.01, est err 4.884]

    if
        dem > 840
        nsslp > 134
        lofc2 <= 107
        lofrc4 <= 78
        lonrc4 > 174
        sprrc1 > 36
    then
        B371 = -3.233 + 0.0029 dem

  Rule 9: [72 cases, mean 9.865, range 0 to 229.59, est err 10.804]

    if
        dem > 840
        nsslp > 243
        lonrc5 > 55
    then
        B371 = -22.544 + 0.64 lonrc1 - 0.017 lofc1 + 0.001 dem + 0.019
lofrc6

  Rule 10: [51 cases, mean 10.034, range 0 to 119.55, est err 10.902]

    if
        dem > 840
        dem <= 1616
        nsslp <= 243
        lofc2 > 107
        lofc2 <= 109
        lofrc4 <= 78
        lonrc5 > 55
    then
        B371 = -4.413 + 0.0044 dem

  Rule 11: [24 cases, mean 54.391, range 0 to 156.08, est err 30.842]

    if
        dem > 1616
        lofrc4 <= 78
    then
        B371 = -8.314 - 1.236 lofrc4 + 1.786 lonc1


Evaluation on training data (3820 cases):

    Average  |error|              1.013
    Relative |error|               0.44
    Correlation coefficient        0.53


        Attribute usage:
          Conds  Model

           78%     4%    dem
           34%     1%    lofrc4
           13%           lonrc5
            4%           nsslp
            3%           lofc2
            1%           ewslp
                  10%    lonrc4
                  11%    lonc1
                  10%    lonrc3
                   9%    lonc2
                   3%    lofc1
                   3%    lofrc6
```

Appendix 23. Cubist 2.05 yellow birch basal area model for USGS zone 57 (mountain region) with LiDAR DEM

```
Cubist [Release 2.05]  Thu May 29 08:43:28 2008

    Target attribute `B371'

Read 3820 cases (59 attributes) from z57b371.data

Model:

  Rule 1: [1976 cases, mean 0.008, range 0 to 8.65, est err 0.008]

    if
        dem2 <= 840
    then
        B371 = 0

  Rule 2: [1934 cases, mean 0.099, range 0 to 43.79, est err 0.099]

    if
        lofrc4 > 78
    then
        B371 = 0

  Rule 3: [3774 cases, mean 0.819, range 0 to 229.59, est err 0.820]

    if
        dem2 <= 1565
    then
        B371 = 0

  Rule 4: [43 cases, mean 1.622, range 0 to 16.73, est err 2.285]

    if
        lofrc4 <= 78
        lonrc5 <= 55
        dem2 > 1146
        dem2 <= 1565
    then
        B371 = 4.536 - 0.081 lofrc4

  Rule 5: [719 cases, mean 3.011, range 0 to 119.55, est err 3.074]

    if
        lofrc4 <= 78
        lonrc5 > 55
        dem2 > 840
        dem2 <= 1565
        nsslp2 <= 243
    then
        B371 = 5.119 - 0.027 lofrc5 - 0.034 lofc3

  Rule 6: [607 cases, mean 3.529, range 0 to 119.55, est err 3.916]

    if
        lofc2 <= 109
        lofrc4 <= 78
        lonrc5 > 55
        dem2 > 840
        dem2 <= 1565
        nsslp2 <= 243
    then
        B371 = -3.35 + 0.0033 dem2 - 0.028 lofrc5 + 0.044 lofrc6

  Rule 7: [43 cases, mean 16.519, range 0 to 229.59, est err 17.296]

    if
        lonrc1 > 34
        lonrc5 > 55
        dem2 > 840
        nsslp2 > 243
    then
        B371 = -4.701 + 0.0014 dem2 + 0.09 lonrc1
```

```
  Rule 8: [32 cases, mean 45.276, range 0 to 156.08, est err 30.152]

    if
        lofrc4 <= 78
        dem2 > 1565
    then
        B371 = 1607.618 - 0.7192 dem2 - 3.149 lofrc4 + 2.587 sprc1
               - 12.76 lofrc1 + 0.2187 ewslp2


Evaluation on training data (3820 cases):

    Average  |error|               1.003
    Relative |error|                0.43
    Correlation coefficient         0.55


        Attribute usage:
          Conds  Model

           79%     7%    dem2
           37%           lofrc4
           15%           lonrc5
           15%           nsslp2
            7%           lofc2
                  15%    lofrc5
                   8%    lofc3
                   7%    lofrc6
```