WHOLE STAND GROWTH AND YIELD MODELING OF SOUTHERN HARDWOODS

by

ANIL KOIRALA

(Under the Direction of Cristian R. Montes)

ABSTRACT

Majority of the forest management decisions rely on the future projection of the forest stands. Thus, it is important to have a reliable growth and yield system for effective stand management. However, predicting the future growth of natural stands is challenging because of the different growth requirements of multiple species. First, this dissertation presents a comprehensive review of the past studies focusing on growth and yield model of southern hardwoods and explains the historical trend in the model development process. The stand-level growth and yield system developed in this study includes a novel site index model fitted through an extended Kalman filtering approach, compatible prediction-projection basal area models with assimilated variance, a dynamic stand density model with localized initial stand density, and total stand volume and green weight equations. This study utilizes data from natural hardwood stands established across the southern states by the Hardwood Research Cooperative (HRC) of the North Carolina State University, which includes nine different site types representing the overall condition of the natural forests in the region. The models demonstrated in this study are new advances for the southern hardwood modeling and were displayed to be improvements over prevailing systems.

INDEX WORDS: Growth and Yield; Uncertainty Estimation; Southern Hardwoods

WHOLE STAND GROWTH AND YIELD MODELING OF SOUTHERN HARDWOODS

by

ANIL KOIRALA

B.Sc., Tribhuvan University, Nepal, 2015

M.S., University of Maine, 2017

M.S., University of Georgia, 2021

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2022

WHOLE STAND GROWTH AND YIELD MODELING OF SOUTHERN HARDWOODS

by

ANIL KOIRALA

Major Professor:	Cristian R. Montes
Committee:	Bronson P. Bullock
	Pete Bettinger
	JohnPaul McTague
	Michael B. Kane

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
August 2022

DEDICATION

I want to dedicate this PhD dissertation to my late father Mr. Basanta Raj Koirala who passed away while I was pursuing this degree.

ACKNOWLEDGEMENTS

First of all, I would like to thank my major advisor Dr. Cristian Montes for introducing me to the new topic about growth and yield modeling of forest stands, as well as providing me with invaluable assistance at each step throughout my PhD journey. He made himself constantly available to help with issues big and small. I have no doubt my career would be taking a very different path were it not for him.

I would also like to thank my committee members (Drs. Bronson Bullock, Pete Bettinger, Michael Kane and John Paul McTague) for their guidance through the course of my doctoral degree program. I am especially thankful to Dr. Pete Bettinger for providing financial support for my graduate assistantship. I would like to acknowledge the Hardwood Research Cooperative (HRC) of the North Carolina State University for permission to use the hardwood stand data for my research. I will forever be in debt to Dr. John Paul McTague for allowing me to use the HRC data.

Finally, I would also like to acknowledge the Plantation Management Research Cooperative and its members for providing financial as well as technical support for my PhD program at the University of Georgia.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Overview of Natural Forests in the US South

Forest areas together cover more than one-third of the US land area and contain approximately 1 trillion cubic feet of wood volume. Timberlands – forests that are legally and actively harvested for timbers – occupy about 67% of total forestlands (Nelson et al., 2020). The majority (87%) of timberlands in the US are naturally regenerated or simply natural forests. Natural forests are those that are regenerated by seed fall and not by direct seeding or planting. The US South has the highest forest cover in the nation and sometimes referred to as the "wood basket" of the country because of its high harvesting and plantation rates. Although, the plantation acreage in the South is increasing by double digits over the years, the natural forests still occupy more than 75% of southern forests (Oswalt et al., 2019). The most common natural forest communities in the US south are oak/hickory, loblolly/shortleaf pine, oak/pine, and oak/gum/cypress. Southern hardwood forests can be generally classified into upland and bottomland hardwoods based on site and species characteristics. Oak species are the most important species in both forest types. The majority of southern hardwoods are owned by nonindustrial private landowners while a little portion is owned and managed by the governmental agencies as national and state forests. These private owners rarely engage in intensive forest management as in pine plantation but supply more than half of the hardwood lumbers in the region (Hicks et al., 2004). Thus, these forests are vital for regional timber supply and local economy. As compared to plantation forests, natural forests could deliver more benefits in terms of fresh water, clean air, carbon sequestration, aesthetics/leisure activities and habitats for wildlife (Siry et al., 2005). The natural forest ecosystem is complex in the sense it consists of many tree species, flora, fauna, microbes, insects that rely on each other for food, growth and survival. Thus, conservation and responsible management of natural forests is crucial for protecting the environment and sustaining these valuable resources to our future generations.

Growth and Yield Modeling

The number of scientific research on growth and yield modeling (GYM) in the US increased with some important forest biometrics milestones such as size-class distribution models (Bailey and Dell, 1973; Burkhart, 1971), individual tree models (Biging and Dobbertin, 1995; West, 1987), and process models (Landsberg and Waring, 1997). However, the major shift in their development occurred after the invention of computer technology, which dramatically increased the computational capability of researchers, thus increasing number of scientific studies in the 1980s (Weiskittel et al., 2011). To date, thousands of growth and yield models have been developed for many conifer and hardwood tree species in the U.S. and around the world.

The GYM studies in the southern US has principally focused on monoculture plantation of loblolly or shortleaf or longleaf or slash pine (Mctague et al., 2008). This is further justified by increase in research and development works with excessive increase in the pine plantation acreage in the region. Until 1950s, only about 2 million acres of pine plantation existed in the region, which has dramatically increased twenty times in just 50 years period (Fox et al., 2004). However, the naturally regenerated forests still occupy larger land area in the south. The information generated from growth and yield provides the basis for many forest management decisions. However, there is a paucity of information on growth and yield of southern natural forests, thus affecting their sustained management. One major reason is the complexity of the species and age distributions in natural forests. This creates difficulties in finding appropriate analysis that ensures accuracy over a large area. Relatively lower number of growth modeling studies have been carried out for hardwood in this region. Very few comprehensive systems of whole stand level growth models have been developed for southern hardwoods. Only one study has tested the accuracy and efficacy of available hardwood growth models in the US South (Rauscher et al., 2000), but studies reviewing and summarizing these literatures are few (Schlaegel, 1982). In the following subsection, I present a comprehensive review of the growth and yield studies developed for southern hardwoods in chronological order.

Literature Review

Growth and yield models developed from year 1921 to 1960 were categorized as first phase, as this stage was just focused on yield table construction and site index derivations through guide curves. Since empirical modeling gained its momentum from 1960, I grouped studies from 1961 to 1980 as second phase. The computational capabilities of researchers improved after home computers became accessible after 1980s. Thus, I grouped studies from 1981 to 2000 as third stage. Finally, the studies from 2001 to 2020 were grouped as final stage. This period has seen adoption of advance computing techniques in forest growth modeling such as non-parametric, stochastic, Bayesian and machine learning models. All the equations presented in succeeding subsections are based on original notations from the publications in order to minimize ambiguity.

First Phase (1921-1960)

Initial growth and yield studies for southern hardwoods were based on yield tables. Sterrett (1921) developed a generalized yield table for even-aged stands of upland hardwoods. This yield table was considerably different than normal yield table that were constructed mainly for single species. To our knowledge, this study is first of its kind for southern hardwoods and consists growth and yield information of now extinct American chestnut.

The study carried out by Schnur (1937) developed yield, stand and volume tables for even-aged upland oak forests. The data for this study came from vast area of upland oak region ranging from New York to Mississippi. Site index curves were created using following generalized equation (with original notations).

$$I = H_A + \sigma_A \left(\frac{H - H_a}{\sigma_a} \right)$$

where, I *= site index;* H *= average dominant height of the stand;* H_A *= average height at any reference age A;* H_a*= average height at any age* a*;* σ_A *= standard deviation of height about the average at any reference age A; and* σ_a *= standard deviation of height about the average at any age* a*.*

Apart from site index, this study also presented calculations for stand mortality, stand basal area and stand volumes. The upper limit of site index curves in this study was only 85 feet, which did not represent productive sites of the southern region. Olson (1959) extended Schnur's site index curves by adding upland oak sites from Appalachian and Piedmont mountains, while Doolittle (1957) attempted to relate site index of scarlet and black oak with soil and topography of southern Appalachians.

Second Phase (1961 to 1980)

Most of the pioneer growth and yield studies for southern hardwoods are from this phase. This phase witnessed modeling efforts to predict growth and yield of single-species as well as mixed-species hardwood stands. Studies were carried out for upland and bottomland hardwoods covering vast area of southern hardwood growth region.

Yellow poplar: The most important achievement of this phase is the work done for yellow poplar stand in the southern mountains. Starting from 1962, a series of reports on yellow poplar growth and yield were published by US Forest Service southern research station. In the beginning, Beck (1962) developed yellow poplar site index curves for the southern Appalachian Mountains and Piedmont Plateau. With the index age of 50 years, the average site index for Appalachian and Piedmont plots were 87 and 82 feet, respectively. Based on this site index, Beck (1963) constructed a cubic-foot volume tables for yellow poplar in the southern Appalachians. McGee and Della-Bianca (1967) presented a method to approximate the diameter distribution of pure, unthinned stands of yellow poplar. Beck and Della-Bianca (1970) used this diameter-distribution approach to predict yield for unthinned stands. Initially, they converted basal area per diameter class into number of trees per class using following equation.

$$N_i = P_i \frac{BA}{B_i}$$

where, P_i = proportion of total basal area per acre that lies in the ith diameter class; N_i = number of trees in the ith diameter class; BA= total basal area per acre; H_a= average height at any age α; σ_A = standard deviation of height about the average at any reference age A; and B_i = basal area per tree for the mid-point tree in the ith diameter class.

The volume of individual tree with information of diameter and height was calculated using standard combined-variable equation (i.e. D^2H). These volumes for individual trees were then applied to the number of trees in each diameter class for different combinations of age, site index, and number of trees to produce yield tables.

After developing models for unthinned stands, Beck and Della-Bianca (1972) developed yield equations for thinned even-aged stands of yellow poplar. They applied a simultaneous growth and yield model developed by Sullivan and Clutter (1972) for their yield projection. Following equation gives final model and parameter values.

$$lnY_2 = 5.36437 - 101.16296(S^{-1}) - 22.00048(A_2^{-1}) + 0.97116\left(\frac{A_1}{A_2}\right)(lnB_1) + 3.71796\left(1 - \frac{A_1}{A_2}\right) + 0.01619\,(S)\left(1 - \frac{A_1}{A_2}\right)$$

where, Y_2 *= total stand volume of wood and bark at some future age* A_2*;* A_1 *= present stand age;* A_2*= stand age at the end of projection period; S = site index;* B_1 *=present stand basal area; and* ln *= natural logarithm.*

As a continuation of thinned yellow-poplar study, Beck and Della-Bianca (1975) developed equations and tables for estimating board-foot growth and yield, and residual quadratic mean stand diameter growth. The board foot yield equation is given as:

$$\frac{BFV}{B_1} = -545.33701 + 222.63551\left(D^{\frac{1}{2}}\right) - 18.18270(D) + 0.35306(H * D^{\frac{1}{2}})$$

where, BFV = international stand volume per acre of all trees 11.0 inches dbh and over, B_1 *= residual stand basal area in square feet per acre; H = average stand total height in feet of dominant and codominant; and D = residual quadratic mean stand diameter in inches.*

Apart from aforementioned studies, there are two other yellow-poplar growth and yield studies from West Virginia. The growth and yield study by Schlaegel and Kulow (1969) was based on compatible stand models proposed by Clutter (1963). With this compatible method, the growth model could be summed to obtain predictions from the yield model. Based on this equation, empirical yield tables were created for same study area (Schlaegel et al., 1969).

<u>Upland Oaks</u>: In this period, there was only one notable growth and yield study for upland oaks. This study published by Dale (1972) developed a system of equations to estimate growth and yield of upland oaks after 10 years of initial thinning. The growth and yield were presented in terms of basal area, total cubic-foot volume, cordwood volume, and board-foot volume over a broad range of site, age, and residual stand density classes. Although data were from Kentucky, Ohio, Missouri and Iowa, they were mostly representative of upland oaks present in the southern US. The site index of upland oaks was in the range of 55 to 89 feet. The models that were selected for basal area growth and cubic-foot volume is presented as:

For basal area, $Y_1 = -BA^{-0.8} * lnB + 3.68521 * BA^{-0.75} + 0.011383 * B * S * A^{-1.05}$

For cubic-foot volume, $Y_2 = 3.09094 + 0.009302 * S + 1.03909 * lnB - 20.11035 * A^{-1}$

where, Y_1 *= net annual basal area growth per acre in square feet for all trees 2.6 inches or larger in dbh,* Y_2 *= natural logarithm of total cubic-foot volume per acre for all trees 2.6 inches or larger in dbh; B = basal area in square feet per acre of all living trees 2.6 inches or larger in dbh; S = site index in feet; and A = average stand age in years.*

<u>Mixed hardwoods</u>: While many studies were focused on single species hardwood stands during this period, Smith et al. (1975) developed yield equations for mixed species hardwood stands in the US south. This study included large area of southern hardwood sites ranging from Virginia to the north to Florida to the south and Arkansas to the west. It also covered broad site classification that are typical to those of southern hardwoods such as muck swap, peat swamp, wet flat, red river bottom, black river bottom, bottomland, coves, gulfs and lower slopes, and upland slopes and ridges. This study was based on the data collected by Hardwood Research Cooperative of North Carolina State University. A large number of yield equations was developed for a variety of

merchantability classes. This study still remains as one of the largest southern hardwood studies. The general model to predict yield variable of interest is given as:

$$log_{10}yield = a + b\left(\frac{1}{age}\right) + c\left[\frac{log_{10}total\ height}{age}\right] + d(log_{10}basal\ area)$$

where, a, b, c and d are coefficients to be estimated.

The general models to predict height and basal area over time are given as:

For height, $log_{10}height = a + b(age) + d(STATE) + \ c_i(\%\ of\ \ species\ i)$

For basal area, $log_{10}basal\ area = a + b\left(\frac{1}{agfe}\right) + \ c\left[\frac{log_{10}total\ height}{age}\right] + \ e(STATE)$

$$+ f\left[\frac{log_{10}total\ height}{age^2}\right] + \ d_i(\%\ of species\ i)$$

where, i is the species and STATE = 0 (if North Carolina, Tennessee, or Virginia) or 1 (if elsewhere).

The STATE variable changes the intercept of the equation to account for the geographic locations. The coefficients for different species and geographic locations are available in the original publication.

During this phase, some studies attempted to relate soil properties to site index for mixed southern hardwoods (Broadfoot, 1969; Della-Bianca and Olson, 1961). These studies suggested that it was challenging to accurately evaluate site quality for southern hardwoods with equations derived over broad areas and complex land pattern. The relationships between soil properties and height growth was hard to quantify. Widely varying responses of different species to a single set of conditions increased the confounding effects. The procedure, however, worked well for some upland hardwoods, but it was discouraging for the complex alluvial soils of the midsouth most of the times.

Third Phase (1981 to 2000)

This phase also witnessed continuation of research on upland as well as mixed hardwoods. Use of computer technology resulted in complex models in this era than previous. The G-HAT model (growth and yield of Appalachian mixed hardwoods after thinning) developed by Harrison et al. (1986a) was the most prominent and modern modeling system of this period. This work was a joint effort of Virginia Tech and U.S. Forest Service. G-HAT was a system of computer programs used for growth and yield predictions. This system used individual tree equations to predict tree basal area increment and total height for the residual stand. The program was available in both interactive program or as a code library in FOTRAN. Data for this study came from Blue Ridge physiographic region of Virginia, North Carolina, Tennessee, and Georgia. Site index was calculated using oak site index equation developed by Olson (1959) (Note: This is described in Phase 1). Individual tree periodic basal area increment and total height equations were based on work of Harrison et al. (1986b) of same project. All the species-specific coefficient value is provided in the original publication.

For basal area, $G = a + bX_1 + cX_2 + dX_3$

For total height, $H = 4.5 + \; H_d\,[1 + a * \exp(bH_d)]\left[1 - \exp\left(\frac{cD}{H_d}\right)\right]$

Where, G = periodic annual basal area increment over five years (square inches, if G < 0, then G should be set to 0); X_1 = original tree basal area (square inches); X_2 = stand basal area after thinning (square feet/acre); X_3 = stand basal area before thinning (square feet/acre); H = total height (feet); H_d = height of dominants and codominant oaks (feet), D = dbh (inches); and a, b, c = species-specific coefficients.

The individual tree total volume prediction model with coefficients is given as:

$$V = \left(\frac{0.1104435}{-ab}\right)\left(\frac{D^2}{H}\right)\left[\frac{(H-4.5)}{4.5}\right]^{b+1} * exp\left\{-a\left[\frac{4.5}{H-4.5}\right]^{b}\right\}$$

Where, V = individual tree total volume (cubic feet, outside bark); other notations are defined previously.

Knoebel et al. (1986) enhanced Beck and Della-Bianca (1972) thinned yellow-poplar model by adding information about diameter and product distributions to better evaluate the effects of different thinning options. The data for this study was from the same permanent plots used in previous yellow-poplar studies. The parameter recovery technique proposed by Hyink (1983) was utilized to obtain parameter estimates of Weibull probability density function. The main reason behind selecting recovery method was its ability to provide compatibility for both whole stand and diameter distribution estimates. In another study, the findings from G-HAT models were improved using diameter distribution models and parameter recovery methods by Bowling et al. (1989). This phase also saw development of taper and volume equations for yellow-poplar (Knoebel et al., 1984). Data from 336 destructively sampled yellow-poplar trees were utilized to develop compatible equations for stem volume and taper.

Two studies carried out by Alabama Agricultural Experiment Station, Auburn University utilized data from US Forest Service, FIA program to develop individual-tree growth and yield models for southeastern hardwoods and Georgia hardwoods, respectively (Bolton and Meldahl, 1990, 1989). These two studies (SE-TWIGS and GA-TWIGS) were based on the TWIGS modeling framework, originally developed by US Forest Service North Central Forest Experiment Station, Indiana for central hardwood species (Belcher, 1982).

The study conducted by Mengel and Roise (1990) developed diameter-class matrix model for coastal plain bottomland hardwood stands. The diameter distribution projection of species groups was based on diameter class, stand basal area, and trees per acre. The authors argued that the matrix approach was easier to implement and it did not require information on individual tree competition, which are essential for individual-tree model.

Graney and Murphy (1994) developed a system growth and yield model for oak stands in the Boston Mountains of Arkansas. This system consisted a series of projection equations for several stand variables such as total basal area, quadratic mean diameter, total and merchantable volume, sawtimber volume and board-foot volume. Later, Murphy and Graney (1998) developed models to predict individual-tree basal area growth, survival and total heights from same data. The authors used a basal area growth equation (Murphy and Shelton, 1996) developed for uneven aged loblolly pine as their model. For total tree heights, the model developed by Harrison et al. (1986) was selected as the best model. One novel contribution of this study was the individual tree survival model, which is presented as:

$$P = \left[1 + exp\left\{-\left(c_0 + c_1\frac{D}{DQ} + c_2 BAL + c_3\frac{1}{N} + c_4\frac{1}{D} + c_5 SI + c_6 SBA + c_7\frac{1}{A}\right)\right\}\right]^{-1}$$

Where, P = probability that a tree will survive for 5 years; D = tree dbh (in.); DQ = average quadratic mean dbh of the stand (in.); BAL = basal area in trees whose dbh's are equal to or larger than the subject tree (ft^2/ac); N = trees/ac; SI = site index (ft); SBA = stand basal area (ft^2/ac); A = stand age (yr); and c_i's = coefficients to be estimated.

At the end of this phase, Rauscher et al. (2000) performed an accuracy test on 10 publicly available hardwood growth and yield models for southern hardwoods. Out of these ten models, five were specifically developed for southern hardwoods. The results of this study showed that the G-HAT model by Harrison et al. (1986) and the mixed hardwood model by Smith et al. (1975) performed well for upland hardwoods. Although, publicly available bottomland hardwood growth and yield models were not available that time, the authors suggested SETWIGS by Bolton and Medldahl (1989) could perform fairly well in that region.

Fourth Phase (2001 to 2020)

As studies from previous phase pointed out the lack of growth and yield studies from bottomland hardwoods, this phase witnessed increased number of studies on those species. Out of nine studies complied from this era, six were purely based on bottomland hardwoods, two were for overall south hardwoods (including bottomland hardwoods) and one was for upland yellow poplar. First, we will describe smaller scale studies that are focused only on certain aspects of growth and yield such as total height, basal area, site index, mortality and crown size. Later, we will focus on two comprehensive studies that have included all growth and yield components in one system.

Zhao et al. (2004) developed individual tree diameter growth and mortality models for bottomland mixed-species hardwood stands in the lower Mississippi alluvial valley. The potential predictor variables in this study were based on ecological importance of tree growth and mortality rather than only on fitting statistics. The general form of the basal area increment model for each species group is given as:

$$\ln(DDS) = \beta_0 + \beta_1 \frac{1}{D} + \beta_2 D + \beta_3 D^2 + \beta_4 \frac{D}{\overline{D}} + \beta_5 \frac{D^2}{\overline{D}} + \beta_6 BA + \beta_7 RBA + \beta_8 S + \sum_{i=1}^{4} \beta_{9i} C_i$$

where, $ln(DDS)$ = natural logarithm of 5-year change in squared diameter; βs = coefficients to estimated; $\frac{D}{\overline{D}}$ = relative diameter i.e. ration of tree dbh (D) to a mean stand diameter ($\overline{D}$); BA = basal area, RBA = relative basal area; S = dummy variable for habitat type (S = 0 for type A; S = 1 for type B); C_i = dummy variable identifying specific crown classes.

As a continuation of this study, Zhao et al. (2005) developed a density-dependent matrix model for the growth of bottomland mixed-species stands. This model was an extension of an earlier model developed by Mengel and Roise (1990). With current diameter distribution information, this model could be used to project stand diameter distribution for 5 years. As projection period was relatively short, authors suggested to implement this model mainly for short-term inventory updating.

Jiang et al. (2005) developed a compatible taper and volume equations for yellow poplar in West Virginia. A simultaneous fitting procedure was adopted to fit both taper and volume equations. Lockhart et al. (2005) estimated the relationships between mean crown radius (MCR) and dbh for six bottomland hardwood species. Parameter estimation was carried out using a simple linear model with dbh as the independent variable and MCR as the dependent variable and vice-versa. The results showed that site conditions or stand history can influence the dbh/MCR ration for a given species. Schuler et al. (2013) developed equations to predict maximum and minimum crown sizes from open-grown black willow and natural even aged black willow, respectively.

SOHARC Model System: The SOHARC model system for growth and yield of southern hardwoods was also a research product of the Hardwood Research Cooperative, NC State University (Mctague et al., 2008). This study presented a system of stand-level and individual tree growth and yield for even-aged hardwoods from permanent plots established across 13 southern states. The permanent plots sued in this system were similar to that of Smith et al. (1975). Earlier study in the series developed site index equations for studied regions (Mctague et al., 2006). The final site index model for SOHARC system with base age 25 years was:

$$S = 245 \left[\frac{H}{245}\right]^{\left(\frac{A}{25}\right)^{0.4051}}$$

where, S= site index; H = predominant mean height (40 tallest trees per acre irrespective of species), and A = total stand age.

Other stand level models in the SOHARC system include merchantable stand survival, stand basal area, number of merchantable trees including ingrowth, and volume. The total stand volume is expressed as:

$$V = a_0 B^{a_1} H^{a_2} e^{\left[a_3\left(\frac{smba}{B}\right)100\right]}$$

where, V = total stand volume (ft^3/ac) for all trees (dbh > 1.5 in.) to the tip of the stem; B = total stand basal area for all trees (ft^2/ac); smba = sub merchantable basal area for all trees (ft^2/ac); a_i = parameters to be estimated; and H is previously explained.

For individual-tree model computations, the variety of mixed hardwood species were grouped into four broad species groups. The individual tree models in the system include mortality, diameter growth and height and diameter projection models. This is the most comprehensive hardwood growth and yield system ever published and we recommend readers to consult original publication for equations and coefficients values.

<u>Red oak-sweetgum growth and yield system</u>: Schultz et al. (2010) developed a stand-level growth and yield models system for red oak-sweetgum forests on mid-south minor stream bottoms (a part of bottomland hardwood region). Tree species from the permanent plots were categorized into six species groups. For site index calculation, a Chapman-Richards type anamorphic equation with base age 50 year was used:

$$SI = HD\left(\frac{1- e^{b*BaseAge}}{1-e^{b*Age}}\right)^{c}$$

where, Si = site index of red oaks in feet; HD = average height of the dominant and codominanat red oaks (feet); Age = age of dominant and codominant red oaks; a, b, c = parameters to be estimated.

This study also created stand-level species group models (six each) for stand density, arithmetic mean diameter and quadratic mean diameter. The total merchantable stand volume model was constructed for combined species and unadjusted species group levels.

$$\ln(Vol) = a + bln(TPA) + cln(QD) + dln(Age) + \frac{eln(TPA)}{Age} + \frac{fln(QD)}{Age} + \frac{gln(HD)}{Age}$$

where, Vol = total merchantable volume (ft^3/ac); TPA = trees per acre of all trees for combined species volume or for a specific species group volume; QD = quadratic mean dbh; and a, b, c, d, e, f, g = parameters to be estimated.

Using similar permanent plots, Banzhaf et al. (2016) developed a log-grade volume models for red-oak sweetgum stands. Authors used 2,149 professionally graded trees to construct their prediction models. The original volume model, however, was similar to Schultz et al. (2010).

Research gaps and Future Outlook

Despite excessive increase in plantation pine acreage, the mixed hardwood forest is the most common forest type in the US south. The mixture of tree communities within different geographical range is so high in the US south that the southern forests are ranked among the most biologically diverse temperate forests in the world (Trani 2002). More than half of the region's forest area is dominated by mixed hardwood forests; most of such forests are under private ownership, especially, non-industrial private forest (NIPF) landowners. Income generation through periodic harvesting is one of the main motivations behind owning and managing timberland for landowners. Reliable information on the growth and yield pattern of the forests is therefore extremely important not only to estimate potential harvest volume but to preserve growing stock for future use as well. Forest management plans for all forest ownership types are basically based on the predictions derived from growth and yield models.

Since the beginning of forest modeling research works in the US, many researchers have carried out commendable studies to quantify southern hardwood growth and yield. However, the number of growth and yield studies on southern hardwoods is plummeting. Most of the studies are carried out in the period between 1960 to 2000. Except few comprehensive growth and yield systems like GHAT, SETWIGS, SOHARC and Red oak-sweetgum model, most of the past literature are focused on specific components of growth and yield models. About half of these studies are published as reports or conference papers, which are difficult to access for public use. This review has attempted to compile most, if not all, growth and yield studies for southern hardwood in one place. Even after 100 years of forest modeling works, it is evident from the review of available models that much work remains in the area southern hardwood growth and yield. Despite technological advances in the field of computation and statistics, why the number of studies is in decreasing trend? Some of the main challenges and research gaps identified from this review are mentioned henceforth.

Lack of growth modeling data was identified as the prime challenge in southern hardwood growth and yield modeling research at current time. Permanent plots are regarded as the main data source for growth and yield studies. They are more preferred than temporary plots because of their continuous measurements and management (Burkhart and Tome, 2012; Weiskittel et al., 2011). Many permanent plots of southern hardwoods that were used in the past studies no longer exist at present time. Efforts to create new plots for future research are almost none. As growth and yield research are fundamentally long-term studies, only few researchers are prepared to devote many years, especially for natural forests with long rotation. In addition, high uncertainty and variability associated with mixed species stands does not help the cause. However, there is a room for improvement. Past data can still be utilized and modeled with modern statistical approaches. The FIA periodic inventory data along with geographic information can supplement historic data. This also allows validation of old models for current forest conditions. Another attractive source of data for future research could come from NIPF landowners. As industrial as well as non-industrial landowners own most of the southern hardwood forests, they will be more benefitted with new and updated models. Future research should focus more on acquiring data from existing hardwood stands from private landowners through correct partnership channels.

Another problem with mixed hardwood forests is the grouping of multiple tree species. The reviewed models showed inconsistency in grouping, for example in GHAT (Harrison et al., 1986) all species were modeled separately, in SETWIGS (Bolton and Meldahl, 1989) grouping was based on physiographic regions, while in Schultz et al. (2010) grouping was based on commercial importance and frequency. Modern statistical classification technique such as decision trees, nearest neighbor or neural networks along with biological characteristics of tree species could be used to create consistency in species grouping.

Results from literature also lack validation with independent or test datasets. The applications of most forest growth models are restricted to the areas covered by calibration datasets. Out of several applications, two most important applications of forest models are prediction and knowledge sharing. Growth predictions are used to update forest inventories and assess alternative silvicultural regimes. Likewise, models are also used to inform landowners and researchers about the influence of disturbing agents like disease and pests. Therefore, statistical validation of models is crucial and if not done correctly could sometime leads to unfortunate consequences. Statistical technique of cross-validation could be a key to legitimize old models in present day. With the knowledge of current southern hardwood conditions, we can simulate mixed hardwood stands growth and development using old plots data.

Some aspects of forest growth and yield modeling is still untouched or has received less attention in previous hardwood growth and yield studies in the US south. Leaf area index (LAI) model is a popular component of growth models nowadays (Yan et al., 2019). There is an ample opportunity for researchers to work on this aspect as freely available high-quality remote sensing data could also be used for LAI modeling. Other important topics with less attention are taper, biomass and green weight equations. Only two studies have developed taper or stem profile equation for southern hardwood species (Knoebel et a., 1984; Jiang et al., 2005). Hardwood tree species such as sweetgum is viewed as a promising bioenergy crop in the US south (Merkle and Cunningham, 2011). Modeling total biomass and green weight for such forest stands could prove important for researchers and landowners in near future. There is a possibility that those types of modeling for southern hardwood will continue to get attention in coming days.

Lastly, advance technologies in data acquisition could highly benefit hardwood growth and yield research in near future. Spatial technologies such as terrestrial LiDAR, drone-based inventory, structure-for-motion as well as high resolution satellite imagery are already in use from some years now. Spatial data will also enable researcher to utilize ecophysiological information for the study area and stands. This will further improve the prediction of growth and yield model while addressing environmental issues such as global warming and climate change. Modern statistical analysis techniques such as machine learning can be used for both spatial as well as field based data. Models correctly tested with machine learning approach could be utilized for various stand and site conditions in future.

Motivation of the Dissertation

There is a paucity of information on growth and yield of southern natural forests, thus affecting their sustained management. One major reason is the complexity of the species distributions in natural forests. This creates difficulties in finding appropriate analysis that ensures accuracy over a large area. Relatively lower number of growth modeling studies have been carried out for hardwood in this region. Very few comprehensive systems of whole stand level growth models have been developed for southern hardwoods. Only one study has tested the accuracy and efficacy of available hardwood growth models in the US South (Rauscher et al., 2000), while studies reviewing and summarizing these literatures are very scarce (Schlaegel, 1982).

Site Index (SI) is a universally accepted indirect method of assessing forest site quality that is based on the relationship between tree height and their age at a given site for a desired species. Most of the growth and yield systems in the US is based on SI derivation. Typically for natural hardwoods, there are limited number of studies on this crucial topic and developed models have not considered quantifying uncertainty in future projections. Chapter 2 intends to develop new site index and dominant height equations for southern hardwoods using advanced filtering process.

Among several important components of forest growth and yield systems, mortality models exhibit a unique and influential role in forest management planning. Estimation of future number of trees in a given forest stand not only drives silvicultural decisions but also impacts the amount and volume of different products derived from the forest over future time horizons (Davis et al., 2001). Despite its importance, mortality still remains one of the least understood components of the growth and yield system (Hamilton, 1986). Because of the complexity arising from interaction of plant growth with environment and other bio-physical factors, mortality has always been a difficult subject to model (Thapa and Burkhart, 2015). Unlike plantations, acquiring information on initial stand density of natural forests could be a dauting task. Thus, in chapter 3, we introduce a novel concept of localizing initial stand density for projecting future number of trees in a stand.

Stand basal area is an essential component of the forest growth model system because it is highly correlated with important economic variables of the forest stands such as total volume and quadratic mean diameter. In addition, basal area can be used as a constrain or a transition function to form a link between low resolution (individual-tree and size-class) and high resolution (whole-stand) models (Cao, 2021; Castedo Dorado et al., 2006; Rodríguez et al., 2010). Chapter 4 aims to develop a new whole stand basal area prediction and projection equations for mixed southern hardwood forests incorporating uncertainty.

The final chapter summarizes the key findings from all chapters and presents potential topics that could be useful for future studies.

Objectives and Organization of the Dissertation

The overall purpose of this dissertation is to develop, test and evaluate a new growth and yield system for even-aged mixed species southern hardwood stands. This dissertation attempts to achieve two major goals. The first one is to evaluate existing growth and yield models for southern hardwoods and provide insight to the limitation and future outlook. The second goal is to develop a full system of stand level growth and yield models, which is further extended into four specific objectives focusing on site index, basal area, mortality and volume of the stands.

Chapters 1 achieves the first goal, while chapters 2 – 5 achieve the second goal of this dissertation. Chapters 2 – 4 are formatted as three independent journal style manuscripts, each having own abstract, introduction, methodologies, results and discussions. Since this study did not conduct destructive sampling of trees, the measured total and merchantable volumes were not available. As a consequence, predicted total and merchantable individual-tree volume equations were used to summarize into total and merchantable stand volumes. For this reason, it was decided a separate journal article style chapter for volume models was not possible. Instead, the methods and results of volume modeling are included in Appendix of this dissertation.

References

Adams, H.D., Park Williams, A., Xu, C., Rauscher, S.A., Jiang, X., McDowell, N.G., 2013. Empirical and process-based approaches to climate-induced forest mortality models. Front. Plant Sci. 4, 1–5. https://doi.org/10.3389/fpls.2013.00438

Bailey, R.L., Dell, T.R., 1973. Quantifying Diameter Distributions with the Weibull Function. For. Sci. 19, 97–104. https://doi.org/10.1093/forestscience/19.2.97

Banzhaf, G.M., Matney, T.G., Schultz, E.B., Meadows, J.S., Jeffreys, J.P., Booth, W.C., Li, G., Ezell, A.W., Leininger, T.D., 2016. Log-Grade Volume Distribution Prediction Models for Tree Species in Red Oak-Sweetgum Stands on US Mid-South Minor Stream Bottoms. For. Sci. 62, 671–678. https://doi.org/https://doi.org/10.5849/forsci.15-138

Beck, D.E., 1963. Cubic-foot volume tables for yellow-poplar in the southern Appalachians. USDA For. Serv. Southeast. For. Exp. Station. Res. Note SE-16 4.

Beck, D.E., 1962. Yellow-poplar site index curves.

Beck, D.E., Della-Bianca, J.O., 1970. Yield of Unthinned Yellow-Poplar. USDA For. Serv. Southeast. For. Exp. Station. Res. Pap. SE-58 20.

Beck, D.E., Della-Bianca, L., 1975. Board-Foot and Díameter Growth of Yellow-Poplar After Thinning. USDA For. Serv. Southeast. For. Exp. Station. Res. Pap. SE-123 20.

Beck, D.E., Della-Bianca, L., 1972. Growth and yield of thinned yellow-poplar. USDA For. Serv. Southeast. For. Exp. Station. Res. Pap. SE-101 20.

Belcher, D.M., 1982. TWIGS: The Woodsman's ideal growth projection system, in: Microcomputers, a New Tool for Foresters. Proceedings of a Conference Sponsored by

Purdue University Department of Forestry and Natural Resources and the S.A.F. Systems Analysis and Inventory Working Groups. West Lafayette, Indiana, p. 155.

Biging, G.S., Dobbertin, M., 1995. Evaluation of Competition Indices in Individual Tree Growth Models. For. Sci. 41, 360–377. https://doi.org/10.1093/forestscience/41.2.360

Bolton, R.K., Meldahl, R.S., 1990. User's guide to a multipurpose forest projection system for southern forests. Alabama Agric. Exp. Station. Bull. No. 604 15.

Bolton, R.K., Meldahl, R.S., 1989. Summary of modeling efforts for southeastern TWIGS. Sch. For. Auburn Univ. Intern. Rep. 216.

Bowling, E.H., Burkhart, H.E., Burk, T.E., Beck, D.E., 1989. A Stand-level Multispecies Growth Model for Appalachian Hardwoods. Can. J. For. Res. 19, 405–412. https://doi.org/https://doi.org/10.1139/x89-064

Broadfoot, W.M., 1969. Problems in Relating Soil to Site Index for Southern Hardwoods. For. Sci. 15, 354–364. https://doi.org/https://doi.org/10.1093/forestscience/15.4.354

Buckman, R.E., 1962. Growth and yield of red pine in Minnesota. USDA Tech. Bull. No. 1272 50.

Burkhart, H.E., 1979. Growth and yield of southern pines - State of the art, in: Southern Forest Economist Workshop 1979. Chapel Hill, North Carolina, p. 11.

Burkhart, H.E., 1971. Slash Pine Plantation Yield Estimates Based on Diameter Distribution: An Evaluation. For. Sci. 17, 452–453. https://doi.org/10.1093/forestscience/17.4.452

Burkhart, H.E., Tomé, M., 2012. Modeling Forest Trees and Stands, First. ed. Springer.

Cao, S., Sun, G., Zhang, Z., Chen, L., Feng, Q., Fu, B., McNulty, S., Shankman, D., Tang, J., Wang, Y., Wei, X., 2011. Greening China naturally. Ambio 40, 828–831. https://doi.org/10.1007/s13280-011-0150-8

Clutter, J.L., 1963. Compatible Growth and Yield Models for Loblolly Pine. For. Sci. 9, 354–371. https://doi.org/10.1093/forestscience/9.3.354

Dale, M.E., 1972. Growth and Yield Predictions for Upland Oak Stands: 10 Years After Initial Thinning. USDA For. Serv. Northeast. For. Exp. Station. Res. Pap. NE-241 21.

Della-Bianca, L., Olson, D.F., 1961. Soil-Site Studies in Piedmont Hardwood And Pine-Hardwood Upland Forests. For. Sci. 7, 320–329. https://doi.org/10.1093/forestscience/7.4.320

Doolittle, W.T., 1957. Site Index of Scarlet and Black Oak in Relation to Southern Appalachian Soil and Topography. For. Sci. 3, 114–124. https://doi.org/10.1093/forestscience/3.2.114

Fox, T.R., Jokela, E.J., Allen, H.L., 2004. The evolution of pine plantation silviculture in the Southern United States, in: Rauscher, H.M., Johnsen, K. (Eds.), Southern Forest Science: Past, Present, and Future. Forest Service, Southern Research Station, Asheville, NC, pp. 63–81.

Graney, D.L., Murphy, P.A., 1994. Growth and Yield of Thinned Upland Oak Stands in the Boston Mountains of Arkansas. South. J. Appl. For. 18, 10–14. https://doi.org/10.1093/sjaf/18.1.10

Harrison, Wade C, Burk, T.E., Beck, D.E., 1986. Individual Tree Basal Area Increment and Total Height Equations for Appalachian Mixed Hardwoods After Thinning. South. J. Appl. For. 10, 99–104. https://doi.org/10.1093/sjaf/10.2.99

Harrison, Wade C., Burkhart, H.E., Burk, T.E., Beck, D.E., 1986. Growth and yield of Appalachian mixed hardwoods after thinning. Sch. For. Wildl. Resour. Virginia Ploytechnic Inst. State Univ. Publ. No. FWS-1-86 48.

Hicks, R.R.J., Conner, W.H., Kellison, R.C., Van Lear, D., 2004. Silviculture and Management Strategies Applicable to Southern Upland Hardwoods, in: General Technical Report SRS-75. Asheville, NC. US Department of Agriculture, Forest Service, Southern Research Station, Chapter 7, pp. 51–62. https://doi.org/10.1017/CBO9781107415324.004

Hyink, D.M., Moser, J.W., 1983. A Generalized Framework for Projecting Forest Yield and Stand Structure Using Diameter Distributions. For. Sci. 29, 85–95. https://doi.org/10.1093/forestscience/29.1.85

Jiang, L., Brooks, J.R., Wang, J., 2005. Compatible taper and volume equations for yellow-poplar in West Virginia. For. Ecol. Manage. 213, 399–409. https://doi.org/10.1016/j.foreco.2005.04.006

Keyser, T.L., Arthur, M., Loftis, D.L., 2017. Repeated burning alters the structure and composition of hardwood regeneration in oak-dominated forests of eastern Kentucky, USA. For. Ecol. Manage. 393, 1–11. https://doi.org/10.1016/j.foreco.2017.03.015

Knoebel, B.R., Burkhart, H.E., Beck, D.E., 1986. A Growth and Yield Model for Thinned Stands of Yellow-Poplar. For. Sci. 32, a0001-z0002. https://doi.org/10.1093/forestscience/32.s2.a0001

Knoebel, B.R., Burkhart, H.E., Beck, D.E., 1984. Stem Volume and Taper Functions for Yellow-Poplar in the Southern Appalachians. South. J. Appl. For. 8, 185–188. https://doi.org/10.1093/sjaf/8.4.185

Korzukhin, M.D., Ter-Mikaelian, M.T., Wagner, R.G., 1996. Process versus empirical models: which approach for forest ecosystem management? Can. J. For. Res. 26, 879–887. https://doi.org/10.1139/x26-096

Landsberg, J.J., Waring, R.H., 1997. A generalised model of forest productivity using simplified concepts of radiation-use efficiency, carbon balance and partitioning. For. Ecol. Manage. 95, 209–228. https://doi.org/10.1016/S0378-1127(97)00026-1

Lockhart, B.R., Weih, R.C., Smith, K.M., 2005. Crown Radius and Diameter at Breast Height Relationships for Six Bottomland Hardwood Species. J. Ark. Acad. Sci. 59, 110–115.

MacKinney, A.L., Chaiken, L.E., 1939. Volume, yield, and growth of loblloly pine in the Mid-Atlantic Coastal Region (No. U.S. Forest Service Technical Note No. 33).

MacCleery, D.W., 2011. American forests: ahistory of resiliency and recovery. Revised Ed. Forest History Society, Durham, North Carolina, USA.

McGee, C.E., Della-Bianca, L., 1967. Diameter distribution in natural yello-poplar stands. U.S. For. Serv. Res. Pap. SE-25 10.

Mctague, J.P., Loughlin, D.O., Roise, J.P., Robison, D.J., Kellison, R.C., 2008. The SOHARC Model System for Growth and Yield of Southern Hardwoods. South. J. Appl. For. 32, 173–183. https://doi.org/https://doi.org/10.1093/sjaf/32.4.173

Mctague, J.P., Robison, D.J., Loughlin, D.O., Roise, J., Kellison, R., 2006. Development of Growth and Yield Models for Southern Hardwoods: Site Index Determination, in: Connor, K.F. (Ed.), Proceedings of the 13th Biennial Southern Silvicultural Research Conference. pp. 319–321.

Mengel, D.L., Roise, J.P., 1990. A Diameter-Class Matrix Model for Southeastern U.S. Coastal Plain Bottomland Hardwood Stands. South. J. Appl. For. 14, 189–195. https://doi.org/https://doi.org/10.1093/sjaf/14.4.189

Merkle, S., Cunningham, M., 2011. Southern Hardwood Varietal Forestry: A New Approach to Short-Rotation Woody Crops for Biomass Energy. J. For. 109, 7–14. https://doi.org/10.1093/jof/109.1.7

Murphy, P.A., Graney, D.L., 1998. Individual-Tree Basal Area Growth, Survival, and Total Height Models for Upland Hardwoods in the Boston Mountains of Arkansas. South. J. Appl. For. 22, 184–192. https://doi.org/10.1093/sjaf/22.3.184

Murphy, P.A., Shelton, M.G., 1996. An individual-tree basal area growth model for loblolly pine stands. Can. J. For. Res. 26, 327–331.

Nelson, M., Ritters, K., Coulston, J., Domke, G., Greenfield, E., Langner, L., Nowak, D., O’Dea, C., Oswalt, S., Reeves, M., Wear, D., 2020. Defining the United States Land Base. USDA For. Serv. Gen. Tech. Rep. NRS-191 70.

Olson, D.J., 1959. Site index curves for upland oak in the southeast. USDA For. Serv. Southeast. For. Exp. Station. Res. Pap. No. 125 2.

Oswalt, S.N., Smith, W.B., Miles, P.D., Pugh, S.A., 2019. Forest resources of the United States, 2017, US Forest Service, Gen. Tech. Report WO-97, 223 p. https://doi.org/10.5962/bhl.title.101492

Pretzsch, H., Forrester, D.I., Rötzer, T., 2015. Representation of species mixing in forest growth models: A review and perspective. Ecol. Modell. 313, 276–292.

https://doi.org/10.1016/j.ecolmodel.2015.06.044

Rauscher, H.M., Young, M.J., Webb, C.D., Rohison, D.J., 2000. Testing the Accuracy of Growth and Yield Models for Southern Hardwood Forests. South. J. Appl. For. 24, 176–185. https://doi.org/https://doi.org/10.1093/sjaf/24.3.176

Schlaegel, B.E., 1982. Predicting growth and yield in southern hardwood stands: A summary, in: Hotvedt, J.E., Jackson, B.D. (Eds.), 31st Abnnual Forestry Symposium: Predicting Growth and Yield in the Midsouth. Louisiana State University, Baton Rouge, pp. 101–117.

Schlaegel, B.E., Kulow, D.L., 1969. Compatible growth and yield equations of West Virginia yellow-poplar. West Virginia Univ. Agric. Station. Bull. No. 573T 20.

Schlaegel, B.E., Kulow, D.L., Baughman, R.N., 1969. Empirical yield tables for West Virginia yellow-poplar. West Virginia Univ. Agric. Station. Bull. No. 574T 24.

Schnur, G.L., 1937. Yield Stand and Volume Tables for Even-Aged Upland Oak Forests. USDA Tech. Bull. No. 560 1–87.

Schuler, J.L., Woods, B., Adams, J., Souter, R., 2013. Crown Size Relationships for Black Willow in the Lower Mississippi Alluvial Valley, in: Holley, G.A., Connor, K.F., Haywood, J.D. (Eds.), Proceedings of the 17th Biennial Southern Silvicultural Research Conference. Shreveport, Louisiana, pp. 364–366.

Siry, J.P., Cubbage, F.W., Ahmed, M.R., 2005. Sustainable forest management: global trends and opportunities. Forest policy and Economics. 7(4), 551-61.

Smith, H.D., Hafley, W.L., Holley, D.L., Kellison, R.C., 1975. Yields of mixed hardwood stands occuring naturally on a variety of sites in the southern United States. Sch. For. Resour.

North Carolina State Univ. Tech. Rep. No. 55 36.

Sterrett, W.D., 1921. A Generalized Yield Table for Even-Aged, Well-Stocked Stands of Southern Upland Hardwoods. J. For. 19, 382–389. https://doi.org/https://doi.org/10.1093/jof/19.4.382

Sullivan, A.D., Clutter, J.L., 1972. A Simultaneous Growth and Yield Model for Loblolly Pine. For. Sci. 18, 76–86. https://doi.org/https://doi.org/10.1093/forestscience/18.1.76

Weiskittel, A.R., Hann, D.W., Kershaw Jr, J.A., Vanclay, J.K., 2011. Forest growth and yield modeling. John Wiley & Sons.

WEST, P.W., 1987. A Model for Biomass Growth of Individual Trees in Forest Monoculture. Ann. Bot. 60, 571–577. https://doi.org/10.1093/oxfordjournals.aob.a087480

Yan, G., Hu, R., Luo, J., Weiss, M., Jiang, H., Mu, X., Xie, D., Zhang, W., 2019. Review of indirect optical measurements of leaf area index: Recent advances, challenges, and perspectives. Agric. For. Meteorol. 265, 390–411. https://doi.org/https://doi.org/10.1016/j.agrformet.2018.11.033

Zhao, D., Borders, B., Wilson, M., 2005. A density-dependent matrix model for bottomland hardwood stands in the Lower Mississippi Alluvial Valley. Ecol. Modell. 184, 381–395. https://doi.org/10.1016/j.ecolmodel.2004.11.002

Zhao, D., Borders, B., Wilson, M., 2004. Individual-tree diameter growth and mortality models for bottomland mixed-species hardwood stands in the lower Mississippi alluvial valley. For. Ecol. Manage. 199, 307–322. https://doi.org/10.1016/j.foreco.2004.05.043

CHAPTER 2

AN APPLICATION OF EXTENDED KALMAN FILTER TO DEVELOP HEIGHT GROWTH MODELS FOR NATURAL HARDWOOD FORESTS OF THE US SOUTH[1]

[1] Koirala, A. and C.R. Montes. Submitted to *Forest Ecology and Management*.

Abstract

Naturally regenerated hardwood forests are extremely important to preserve regional biodiversity as well as to support regional timber supplies. Decision support tools, such as growth and yield models, are needed to better manage these forests. However, comprehensive growth and yield models for hardwoods in the southern US are still scarce. Predicting growth and yield of hardwoods is difficult because of the multi-species, leading to growth rates that largely differ between individuals. Because of which, these systems are also subject to uncertainty in their predictions. We developed data fitting methodology to calibrate a dominant height equation using a data assimilation approach that filters out much of the variation. The concept of data assimilation is based on the assumption that neither the model nor measurements can perfectly describe a system, but an analysis that combines both model and data could provide a better estimate of system dynamics. This approach is well suited for site index data that includes height at a current age. For this study, we used permanent sampling plot data from even-aged mixed species stands measured all over the southern US. Our results indicated that site index models developed with this approach was able to reduce uncertainty in dominant height projection for mixed hardwood forests in the US South.

Introduction

Forest areas in the US cover more than one-third of the land area and contain about 1 trillion cubic feet of wood volume. Timberlands – forests that are legally and actively harvested for timbers – occupy about 67% of total forestlands (Nelson et al., 2020). The majority (87%) of timberlands in the US are naturally regenerated or simply natural forests. Natural forests are those that are regenerated by seed fall and not by direct seeding or planting. The US South has the highest forest cover in the nation and sometimes referred to as the "wood basket" of the country because of its high forest products harvesting and supply chain (Oswalt et al., 2019). The most common natural forest communities in the US South are oak/hickory, loblolly/shortleaf pine, oak/pine, and oak/gum/cypress. Southern hardwood forests are generally categorized as upland and bottomland forests. Their geographic range encompasses the region between Virginia and Florida from north to south; and Missouri and the Atlantic Ocean from west to east (Hicks et al., 2004). Predicting growth and yield of hardwood stands is challenging because of inconsistent growth rates and different management requirements for different species. The early development of growth and yield modeling (GYM, hereafter) systems was primarily based on single species plantations (Westfall et al., 2017). Therefore, there has been a relatively higher number of GYM studies on them, and information from these systems is better understood than for mixed-species natural stands (Mctague et al., 2008). Natural hardwoods and pines mixed forest types are very typical to that owned by small and medium-sized private landowners in the southern US. Therefore, there is a considerable interest among landowners and forest sector for accurate and reliable GYM for mixed-hardwood forest stands.

In the whole stand level forest management, forest site quality is considered as an influential metrics that guides overall forest management decisions ranging from developing growth and yield projections to evaluating effectiveness and economics of silvicultural treatments (McNab and Keyser, 2020; Westfall et al., 2017). Site Index (SI) is a universally accepted indirect method of assessing forest site quality that is based on the relationship between tree height and their age at a given site for a desired species. As reliance of whole stand growth models on SI increased over time, a large number of scientific studies has been conducted on SI derivation. Traditional efforts to derive site index curves involve proportional scaling of height/age curves for different site qualities (Bruce, 1926). Such curves are anamorphic in nature and their shapes are identical for all sites in the stand. Next stage of SI equation development witnessed more advance approach that could allow flexibility in the curves by addressing site-specific variability. Curves derived such way could be polymorphic in nature whose shapes could vary among stands (Bailey and Clutter, 1974; Stage, 1963). After that, more sophisticated techniques involving mixed and fixed effects models (Cieszewski and Bailey, 2000; Wang et al., 2008); stochastic differential equations (Garcia, 1983; García, 2019; Orrego et al., 2021), penalized regressions (Koirala et al., 2021) have been utilized for site index derivation over the years.

While the predictor variable or stand age in the SI equation is assumed to be continuous, the response variable like height is generally measured at discrete point in the time. Hence, a time-discrete model may prove to be more logical for modeling site index than other traditional approach. A Kalman filter-based time-discrete model can be utilized in this situation (Kalman and Bucy, 1961). The original approach of Kalman filtering is only applicable for linear models. As some sort of non-linearity persists in height and age relationship of forest stands, we utilize a generalization of the Kalman filter that allows the process to be governed by a nonlinear stochastic differential equation. This generalized version is known as extended Kalman filter (EKF). It has been successfully implemented in various fields, such as positioning systems, robot navigation and economics. The Kalman filter is a set of equations that provides an efficient computational technique to estimating the state of a process, in a way that minimizes the covariance of the estimation error. The filter is very powerful in several aspects: it supports estimations of past, present, and even future states, and it can do so even when the precise nature of the modeled system is unknown. Unlike most of the classical parameter estimation methods, the Kalman filter is a recursive estimator. At each time step the filter refines the previous estimate by incorporating in it new information from the model and from the output. More details on derivation of site index equation using EKF approach is presented in materials and methods section.

For southern hardwoods, foresters still rely on the site index equations derived by Olson (1959), which was one of the most comprehensive site index study for upland oak species. Our study has attempted to develop new site index and dominant height equations for southern hardwoods using advanced filtering process. This approach has proven ability to minimize uncertainties in both process and measurements in forest inventory (Ehlers et al., 2013; Nyström et al., 2015).

Methods

Data and Measurements

Data utilized in this research came from the Hardwood Research Cooperative (HRC) of the North Carolina State University (Mctague et al., 2008). More than 600 plots were established across the US South in cooperation with member forest companies of the HRC. Since natural hardwood forests cover a wide range of geographical regions within the US South, the first task was to identify the forest site types. The cooperative identified nine forest site types which could be classified as separate forest operating units (Table 1). Among 600 plots, 187 plots were permanently maintained for repeated measurements and as a data source for growth and yield prediction analysis. The forest areas in the established plots were even-aged and fully stocked stands. Majority of the new growth natural forests in the US South is even-aged, thus the study ignored uneven-aged and high graded forest stands. There were five age groups in the data collection: ranging from 20 to 60 years at 10-year intervals. Table 1 presents the distribution of plots by age class and site types.

Field crew installed circular 1⁄5-acre (~ 0.081 hectare) plots, and measured total height, merchantable height and dbh of all trees greater than 14 cm dbh. In addition, crew also recorded tree species and age of individual trees. Several considerations were made to measure trees with some sort of defects. For trees with butt swell or other malformations at breast height, crew measured the diameter of the tree 0.3 m above the irregularity. For trees with forks below 1.37 m, crew recorded them as individual trees. The merchantable height of the tree with at least 2.45 m long bole was measured until the diameter of tree reaches 10 cm at the top.

Base Models

In order to develop the height prediction model, we selected two functions as our base models: Chapman-Richards and Lundqvist-Korf functions. The Chapman-Richards function (equation 3.1), also known as the Bertalanffy-Richards growth model, is a popular exponential type function, which is frequently used in site-index modeling (Chapman, 1961; Diéguez-Aranda et al., 2006; Koirala et al., 2021; Richards, 1959). It is a monomolecular type function which assumes that the absolute growth rate is proportional to the difference between maximum yield and current yield of the forest.

$$\text{Chapman} - \text{Richards (CR)}: \quad H = b_1(1 - e^{b_2 t})^{b_3} \tag{3.1}$$

where, H is the dominant height (m) of the forest stand, t is the age of the stand and b_1-b_3 are estimation parameters. The asymptote parameter b_1 is the maximum possible height an individual tree or the forest stand can gain or the maximum potential of the forest to grow, which is also sometimes refer to as H_{max}. The parameter b_2 determines the absolute growth rate of a tree or forest over time t. The parameter b_3 represents the catabolism mechanism, which is said to be proportional to the organism's mass (Pommerening and Muszta, 2015). Hence, some authors have restricted its value to a constant like 3 for biological reasons (Pommerening and Muszta, 2015).

The Lundqvist-Korf function (equation 3.2) can be regarded as a family of functions in which the absolute growth rate has a linear relationship with some power of time. The popular Schumacher function is one type of Korf family function. The Korf function is considered very suitable for height modeling and preferred by foresters in Europe and North America (Bailey and Clutter, 1974; Lundqvist, 1957).

$$\text{Lundqvist} - \text{Korf (LK)}: \quad H = b_1 \, e^{-b_2 \frac{1}{t^{b_3}}} \tag{3.2}$$

where, H, t and b_1 and b_2 are previously defined. The parameter b_3, however, represents the shape of the curve or function. It influences the age at which the inflection point occurs. Inflection point in this case refers to the age or time when the growth rate starts to stabilize or decline.

<u>Development of Projection Models</u>

In order to develop a model that can project future height of the forest stands based on the current information of the forests, Bailey and Clutter (1974) formulated an approach known as algebraic difference approach (ADA). In the original paper, the authors used the same LK model as their base model during ADA formulation. To formulate ADA models from our base models, first we need to change the expressions of our base models.

For time *t* and *t+1* with height H_t and H_{t+1}, our base models become

$$\text{CR}: \quad H_t = b_1(1 - e^{b_2 t})^{b_3} \quad \text{and} \quad H_{t+1} = b_1\left(1 - e^{b_2(t+1)}\right)^{b_3} \tag{3.3}$$

$$\text{LK}: \quad H_t = b_1\, e^{-\,b_2 \frac{1}{t^{b_3}}} \quad \text{and} \quad H_{t+1} = b_1\, e^{-\,b_2 \frac{1}{(t+1)^{b_3}}} \tag{3.4}$$

Since we have common parameters in both equations for two time periods (*t* and *t+1*), we can reformulate equation (3.3) and (3.4) to make anamorphic and polymorphic height projection equations. For example, if we solve b_1 parameter, we will get anamorphic equations (equations 3.5 and 3.6).

$$M1 \text{ (CR Anamorphic)}: \quad H_{t+1} = H_t * \frac{\left(1-e^{b_2(t+1)}\right)^{b_3}}{\left(1-e^{b_2 t}\right)^{b_3}} \tag{3.5}$$

$$M2 \text{ (LK Anamorphic)}: \quad H_{t+1} = H_t * \frac{e^{b_2 * \frac{1}{(t+1)^{b_3}}}}{e^{b_2 * \frac{1}{t^{b_3}}}} \tag{3.6}$$

Likewise, if we solve b_2 parameter, we will get polymorphic equations (equations 7 and 8).

$$M3 \text{ (CR Polymorphic)}: \quad H_{t+1} = b_1 \left\{1 - \left[1 - \left(\frac{H_t}{b_1}\right)^{\frac{1}{b_3}}\right]^{\frac{t+1}{t}}\right\}^{b_3} \tag{3.7}$$

$$M4 \text{ (LK Polymorphic)}: \quad H_{t+1} = b_1 * \left(\frac{H_t}{b_1}\right)^{\left(\frac{t}{t+1}\right)^{b_3}} \tag{3.8}$$

Model Parametrization with Extended Kalman Filter

In the first step, we estimated the parameters of anamorphic and polymorphic ADA models (*M1 – M4*) with traditional nonlinear least squares. The initial values to run nonlinear models were decided with the help of previous literatures (Diéguez-Aranda et al., 2006; Sharma et al., 2011). In the next step, we applied Kalman Filtering approach to the problem of forest height growth forecasting. The Kalman Filtering approach is a popular particle filtering method that is well-known in the field of astronomy and navigation. Here, we implemented the Extended Kalman filter in two steps: the prediction step, where the next state of the system is predicted given the previous measurement data, and the update step, where the current state of the system is estimated given the measurement at that time. Before moving into derivation of EKF, let us explain some mathematical notations and their dimensions

x_k $n \times 1$ –– State vector

w_k $n \times 1$ –– Process noise vector

z_k $m \times 1$ –– Observation vector

v_k $m \times 1$ –– Measurement noise vector

$f\,(.)$ $n \times 1$ –– Process nonlinear vector function

$h\,(.)$ $m \times 1$ –– Observation nonlinear vector function

Q_k $n \times n$ –– Process noise covariance matrix

R_k $m \times m$ –– Measurement noise covariance matrix

Let us consider the following nonlinear system, described by the difference equation and the observation model with additive noise.

$$x_k = f(x_{k-1}) + w_{k-1}$$

$$z_k = h(x_k) + v_k \quad (3.9)$$

The initial state x_0 is a random vector with known mean μ_0 and covariance $\boldsymbol{P}_0$.

Thus, the initialization:

$$x_0^a = \mu_0 \text{ with error covariance } \boldsymbol{P}_0$$

Model Forecast Step / Prediction:

$$x_k^f \approx f(x_{k-1}^a)$$

$$P_k^f = J_f(x_{k-1}^a)P_{k-1} J_f^T(x_{k-1}^a) + Q_{k-1} \quad (3.10)$$

Data Assimilation Step / Correction:

$$x_k^a \approx x_k^f + f(x_{k-1}^a)K_k\left(z_k - h\left(x_k^f\right)\right)$$

$$K_k = P_k^f J_h^T\left(x_k^f\right)\left(J_h\left(x_k^f\right)P_k^f J_h^T\left(x_k^f\right) + R_k\right)^{-1}$$

$$P_k = \left(I - K_k J_h\left(x_k^f\right)\right) P_k^f \quad (3.11)$$

In the following equations (equations 3.12 to 3.15), we define our dynamic and measurement models for our previously defined four projection or ADA functions (equations 5 to 8), in order to run the extended Kalman filter steps.

$$M5 \text{ (CR Anamorphic)}: \ H_{t+i,k} \sim N\left(H_{t,k} * \frac{\left(1-e^{b_2(t+i)}\right)^{b_3}}{\left(1-e^{b_2 t}\right)^{b_3}}, i\sigma_{proc}^2\right)$$

$$H_{t,obs,k} \sim \left(H_{t,k}\,,\sigma_{obs}^{2}\right)$$

$$H_{0,k} = \theta_{0,k} \qquad (3.12)$$

$$M6\ (\text{LK Anamorphic}):\ H_{t+i,k} \sim N\left(H_{t,k} * \frac{e^{b_2 * \frac{1}{(t+i)^{b_3}}}}{e^{b_2 * \frac{1}{t^{b_3}}}}, i\sigma_{proc}^{2}\right)$$

$$H_{t,obs,k} \sim N\left(H_{t,k}\,,\sigma_{obs}^{2}\right)$$

$$H_{0,k} = \theta_{0,k} \qquad (3.13)$$

$$M7\ (\text{CR Polymorphic}):\ H_{t+i,k} \sim N\left(b_1 \left\{1-\left[1-\left(\frac{H_{t,k}}{b_1}\right)^{\frac{1}{b_3}}\right]^{\frac{t+i}{t}}\right\}^{b_3}, i\sigma_{proc}^{2}\right)$$

$$H_{t,obs,k} \sim N\left(H_{t,k}\,,\sigma_{obs}^{2}\right)$$

$$H_{0,k} = \theta_{0,k} \qquad (3.14)$$

$$M8\ (\text{LK Polymorphic}):\ H_{t+i,k} \sim N\left(b_1 * \left(\frac{H_{t,k}}{b_1}\right)^{\left(\frac{t}{t+i}\right)^{b_3}}, i\sigma_{proc}^{2}\right)$$

$$H_{t,obs,k} \sim N\left(H_{t,k}\,,\sigma_{obs}^{2}\right)$$

$$H_{0,k} = \theta_{0,k} \qquad (3.15)$$

where, H_{t+i} is the forest stand height at future time $(t + i)$ and H_t is the height at current time t; σ^2_{proc} and σ^2_{obs} are the process and observed variances, respectively; b_1 , b_2 and b_3 are parameters to be estimated; k is the number of unique plots in the data; and θ is the starting value for the plot k.

Cross-validation

Rather than dividing data into fit and validation sets as done in many contemporary height-diameter modeling studies (Calama and Montero, 2004; Zang *et al.*, 2016; Ogana and Gorgoso-Varela, 2020), a more robust five-fold cross-validation approach was used during data fitting process (Hastie *et al.*, 2009). This approach involves randomly dividing the set of observations into *K* groups or folds (5 groups in our case), of approximately equal size. The first group of the data is held out as a validation set and the remaining folds i.e. (*K-1*) are used to train the model. Parameter estimates for the candidate models, including both ADA and EKF models, are obtained from the training data, and later used to predict the dependent variable in the validation set. The process continues for five times as each sample is given the opportunity to be utilized as the hold out for one time while remaining *K-1* samples act as training set. At the end of a complete five-fold cross-validation run for each model, five sets of predicted values and respective goodness-of-fit statistics are generated. In order to test the model performance, an average of these five values are reported and compared (Figure 2). Comparison of eight candidate models were based on three major goodness-of-fit statistics: root mean square error (RMSE), bias, and mean absolute error (MAE) and adjusted coefficient of determination (adj. R^2). Lower values of RMSE, bias and MAE and higher value adj. R^2 are preferred during model selection.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n}} \qquad (3.16)$$

$$Bias = \frac{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)}{n} \quad (3.17)$$

$$MAE = \sum_{i=1}^{n} \left|\frac{Y_i - \hat{Y}_i}{Y_i}\right| \times \frac{1}{n} \quad (3.18)$$

$$adj.R^2 = 1 - \left(\frac{(n-1)\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{(n-p)\sum_{i=1}^{n}(Y_i - \bar{Y}_i)^2}\right) \quad (3.19)$$

where, Y_i is the observed height of the tree; $\hat{Y}_i$ is the predicted height; $\bar{Y}_i$ is the mean value of Y_i; n is the total number of samples and p is the number of parameters in a model.

K-medoid Cluster Analysis for Tree Species Grouping

As we can see in the data (Table 1), there are several hardwood site types and some sites can be more productive than others. The productivity of each forest stand may vary within a given site as well. The SI is a common method to quantify the productive capacity of a forest stand, which is, however, a species – specific measure. Since natural hardwood forests are generally mixed species, it is difficult to justify fitting individual species – specific site index curves. But with available site/stand information and tree species characteristics, one can perform cluster analysis to group similar species in several groups. In this study, we utilized k – medoid cluster analysis technique to group southern hardwoods (Kaufman and Rousseeuw, 1990; Park and Jun, 2009). In order to perform the cluster analysis, four different site/stand and tree species' functional attributes were considered such as quadratic mean diameter of the forest stand, coefficient of variation of the tree diameter, basal area proportion of each species in the given forest stand, and shade tolerance property of each species. Out of these four attributes, the shade tolerance of the species is a character variable and can have five different classes such as 1 = "very tolerant", 2 = "tolerant", 3 = "intermediate", 4 = "intolerant", 5 = "very intolerant" (Zhao et al., 2004). The cluster analysis was performed using PAM (partitioning around medoids) algorithm and Gower distance using two packages in R statistical software "cluster" and "daisy". After dividing the dataset into different clusters, parameters and corresponding standard errors for each cluster were estimated with the best model selected from previous steps.

Results

Performance of ADA Models

We fitted four ADA type dominant height models to measure the site productivity of southern hardwoods in the US South. The parameterization of all four models were carried out utilizing nonlinear least squares approach with five-fold cross-validations. Table 2 illustrates the parameters estimates, standard errors and cross-validation statistics for each model. All the parameters from the models were significant at 1% level. The polymorphic model *M4* generated a higher value of asymptote parameter (50.13 m) than another polymorphic model *M3* (35.43 m). Out of four models, *M4* had least values for RMSE, Bias and MAE. This model also showed the highest value of adjusted R2. Overall, the RMSE values for all four models were lower between 1.5 to 2 m. The bias values for all models were very low between - 0.003 to 0.21 m. In a five-fold cross-validation process, a positive bias on one validation set can offset a negative bias on another validation dataset. Therefore, the bias values are generally low in projection models. This implies bias alone is not enough to evaluate the precision of our models, however, highly bias models could indicate that something is already wrong in forecasting. The mean absolute error (MAE) values were also lower for all models; the lowest value of 0.95 m was for model *M4*. The adj. R^2 values for all models were higher between 85% to 90%, indicating that models were able to explain about 90% variability. Overall, it is clear from the results that the polymorphic form of equations had lower values for both error and bias. Thus evaluated the residuals of these two models graphically (Figure 2). Residual plots from both models were quite similar. However, the residual plot for model *M4* depicted a better distribution of residuals around zero line than of model *M3*.

<u>Performance of EKF Models</u>

Since polymorphic models performed better in our nonlinear least squares analysis, we decided to fit these two polymorphic models with EKF approach (model *M7* and model *M8*). This also reduced our time consumption in EKF model fitting since each run of an EKF took about 8 hours to run in R software. Table 3 shows the results from the EKF model fitting and cross-validation. The EKF results showed similar pattern as of ADA models. The cross-validation statistics of LK polymorphic model showed lower error and bias. As we can see in the table, there are two new columns indicating process and observed variances from EKF model fitting, accounting for both model and measurement errors. Both process and observed variances were slightly lower in LK polymorphic as compared to CR model. Like ADA model, the EKF model also requires prior information about the height-age relationship in the form of an estimate of the coefficient vector and its covariance matrix. However, unlike ADA, the initial equation in EKF is updated with new information such as repeated plot-level measurements and new estimation of forest stand height is provided, which is also known as data assimilation step. This "update step" also deals with assimilating variances for new assimilated height estimations.

The values of RMSE, bias and MAE were slightly lower in model *M8* when compared to model *M4* (Table 2 and 3). This suggests that EKF fitting procedure performed a better job in terms of validation statistics because it utilized the assimilated measurements and variances. The comparison of residual plots between model *M4* and *M8* (Figure 3) showed clear advantage of EKF procedure over the ADA. The residuals of model *M8* showed randomness and more constant variance than that of *M4*. Finally, we compared the predicted height growth trajectories for both model *M4* and *M8* for different site index values (Figure 4). The dashed line representing model *M4* showed underprediction of dominant height after surpassing base age of 50 years. The figure also revealed that the dominant height curves for different SI values for EKF model passed through original data more precisely than *M4*.

<u>Models for Hardwood Species Groups or Clusters</u>

Initially, we need to determine the optimal number of clusters that we can find from our dataset with k – medoids clustering. Figure 5 illustrates the silhouette values for the candidate number of the clusters. Theoretically, number of clusters with higher silhouette width are preferred in initial clustering process. The figure showed that 5, 6 or 7 number of clusters could be appropriate as initial candidates for our datasets. In the next stage, we performed PAM cluster analysis and found that 5 clusters were appropriate for our hardwood dataset. Figure 6 shows the perfect allocation of species into 5 clusters with the cluster analysis. Table 4 shows the name of the species divided in five different clusters and associated parameters, standard errors, and observed and process variances estimated from our best model *M8*, i.e., LK Polymorphic model fitted with EKF technique. Parameters for all species groups were significant at 1% level. These parameter values along with observed and process variances can be utilized to calculate dominant height and site index for similar species or species groups in the US South. In order to calculate site index, we can reformulate our dominant height equation and utilize appropriate coefficients for species group.

$$SI = b_1 * \left(\frac{H}{b_1}\right)^{\left(\frac{A}{50}\right)^{b_3}} \tag{3.20}$$

where, SI is site index in meters, H is dominant height of the given stand in meters, A is the age of the stand in years, and b_1 and b_3 are parameter estimates corresponding to each species group given in Table 4. The choice of 50 years as the base age is common for hardwood forests in the US.

Discussion

Natural forests provide a plethora of goods and services to the people and help sustain the forest-based economies. They are also important in terms of biodiversity conservation and carbon sequestration. Several studies have shown that natural mixed-species forests store more carbon than monoculture plantations (Lewis et al., 2019; Osuri et al., 2020). Recently, in Glasgow COP26, world leaders acknowledged rapid depletion of natural forests and committed to invest in restorations of these forests in coming decades to alleviate global warming (Einhorn and Buckley, 2021). More than two-third of the forested areas in the US South is covered by natural hardwood forests (Oswalt et al. 2019). These forests are commonly found in many industrial and nonindustrial private landowners' properties across the region. SI is used as a proxy to the productive capacity of a forest site for a given tree species, which is based on the relationship between dominant height of the forest stand at base age. Most of the growth and yield systems in the US is based on SI derivation. Typically for natural hardwoods, there are limited number of studies on this crucial topic and developed models have not considered error in height measurements so far. The main objective of this study was to develop dominant height – age curves and SI for mixed hardwood forests of the US South that can account for uncertainty in future projections.

In the first step of this study, we compared four difference equations (ADA models) derived from two base models Chapman-Richards (CR) and Lundqvist-Korf (LK). The polymorphic form of the LK model outperformed CR models (Table 2). The base-age invariant polymorphic models are dynamic in nature and their parameter estimates are always same regardless of the choice of base age for site index (Bailey and Clutter, 1974; Cieszewski and Bailey, 2000). The functional form of our best ADA model, *M4* was similar to that of SOHARC model system, which utilized same data for growth and yield development (McTague et al., 2008). Our study is novel in the sense it improves our understanding of site index modeling of natural hardwood forests of the US South in two ways. First, the study successfully applied Kalman filter technique to account for both measurement as well as process errors. The basic assumption of the least square regression is that the independent variables are free of errors. In other words, no measurement error persists in the data; and the errors are only related to the model or the process. In forestry, measuring tree heights is considered as a difficult and error prone task. Thus, the application of Kalman filter technique in this study has improved the projection of dominant height in future periods. This technique not only allow us to utilize previous observations as starting values, but it also creates a system in which new information can be automatically assimilated for future projection (Montes, 2012; Walters et al., 1991). The comparison between models *M4* and *M8* showed that EKF technique was able to minimize the error and bias in model prediction (Table 3), which further reinforced the added advantage of Kalman filter approach.

Second, this study performed cluster analysis to divide tree species into different clusters based on site, species characteristics and field measurement data. This has allowed us to develop a separate dominant height and site index equations for separate species groups. Natural hardwood forest of the South is home to numerous tree species that have extremely variable growth rates and requirements. Since these forests are almost always found in mixture of different species, a productive forest stand for one species may not be so productive for another tree species. For example, a highly productive site for red oaks in Piedmont, Georgia may not be as productive for cotton wood trees. Thus, it is important to have site index models that can represent the productive capacity of similar species. We utilized k – medoids clustering approach instead of a well-known k – means clustering technique to perform non-hierarchical clustering. One major disadvantage of k – means clustering is its sensitivity to the distance of each object from the centroids. This approach behaves poorly when data has numerous outliers (Park and Jun, 2009). K – medoids clustering on the other hand are not sensitive to the outliers and are based on the representative objects called medoids and their distances. Instead of directly dealing with similarities between the objects or individuals, k – medoids algorithm minimizes average dissimilarity of all objects of the dataset to the nearest medoid (Kaufman and Rousseeuw, 1990). The five clusters, reveled from cluster analysis, were consistent with the species groups developed by US Forest Service FIA database (citation). In addition, species in each cluster exhibits somewhat similar characteristics.

Conclusion

We evaluated two well-known dominant height and site index functions – Chapman-Richards and Lundqvist-Korf – to describe the relationships between dominant height and stand age of southern hardwood forest stands. The model fitting and validation was carried out in two different ways (i) algebraic difference approach (ADA) and (ii) extended Kalman filter (EKF) technique. The results from the cross-validation showed that the polymorphic from of Lundqvist-Korf model outperformed anamorphic as well as polymorphic forms of Chapman-Richards models. Furthermore, the EKF approach proved to be better than ADA approach based on model validation statistics and graphical analysis of residuals. The main advantage of EKF over ADA was the ability of EKF to separate observation error from the process error. The EKF methodology can be a sound technique to calculate dominant height when uncertainty estimation is a goal. In addition, the ability of EKF to assimilate new information in the system without requiring a comprehensive forest inventory could be crucial in reducing sampling costs. The cluster analysis of the data revealed five different species groups for southern hardwood. Our analysis provides parameter values for each hardwood species groups for the region, which could fill the gap of knowledge and confusion in hardwood site index calculation.

Tables and Figures

Table 3.1. Number of hardwood growth and yield plots by site type and age class across the southern United States. The stand ages in the plots were determined using past records and increment cores taken from dominant and codominant trees.

Site type	Age classes (year)					
	20	30	40	50	60	Total
Muck swamp	88	11	8	99	13	49
Peat swamp	1	5	2	11	2	11
Wet flat	7	35	19	12	3	76
Red river bottom	8	18	15	11	6	58
Black river bottom	5	14	24	2	7	52
Branch bottom	12	33	24	18	5	92
Bottomland	36	59	40	23	7	165
Coves, gulfs, lower slopes	5	8	17	7	4	41
Upper slopes and ridges	5	27	34	18	13	97
Total	87	210	183	101	60	641

Table 3.2. Parameter estimates with their standard errors from the entire dataset, and their corresponding validation statistics of four ADA models from five-fold cross validation. The values inside parentheses are the standard error values of each parameter.

Model	b_1	b_2	b_3	**RMSE (m)**	**Bias (m)**	**MAE (m)**	**Adj. R^2**
CR: Anamorphic (M1)		- 0.035 (0.009)	0.972 (0.118)	1.89	0.21	1.43	0.85
LK: Anamorphic (M2)		- 4.363 (0.021)	0.402 (0.064)	1.87	0.06	1.39	0.86
CR: Polymorphic (M3)	35.433 (1.221)		0.916 (0.894)	1.66	0.02	1.25	0.89
LK: Polymorphic (M4)	50.130 (0.835)		0.646 (0.227)	**1.62**	**- 0.003**	**0.95**	**0.90**

Table 3.3. Parameter estimates with their standard errors from the entire dataset, and their corresponding validation statistics of two EKF models from five-fold cross validation. The values inside parentheses are the standard error values of each parameter.

Model	b_1	b_3	**RMSE (m)**	**Bias (m)**	**MAE (m)**	**Process Variance**	**Observed Variance**
CR: Polymorphic (M7)	36.289 (0.812)	0.937 (0.226)	1.63	- 0.03	1.01	1.714	0.296
LK: Polymorphic (M8)	65.252 (0.501)	0.511 (0.038)	**1.59**	**0.001**	**0.92**	**1.697**	**0.211**

Table 3.4. Parameter estimates and variance values for five species groups of southern hardwoods.

Cluster	Tree species in each cluster	b_1	b_3	Process Variance	Observed Variance
1	Sweetgum; Blackgum; Cottonwood; Tupelo	66.163	0.612	1.741	0.642
2	Red Maple; Sweet Bay	65.431	0.503	1.715	0.706
3	Yellow Poplar, Sycamore, Birch	66.433	0.588	1.627	0.372
4	Entire Red Oak, Cutleaf Red Oak, White Oak, Hickory, Elm, Hackberry, Ash	65.896	0.525	1.638	0.513
5	Back cherry, Others	67.282	0.671	1.977	0.625

Figure 3.1. Schematic representation of five-fold cross-validation approach. Each horizontal bar is divided into five sections, out of which four are training folds and one section is the validation fold. There are altogether five horizontal bars representing five iterations or folds. Goodness-of-fit statistics (GOF) are calculated for each iteration and the final GOF statistics for a model is calculated as the average of five GOF statistics from five iterations.

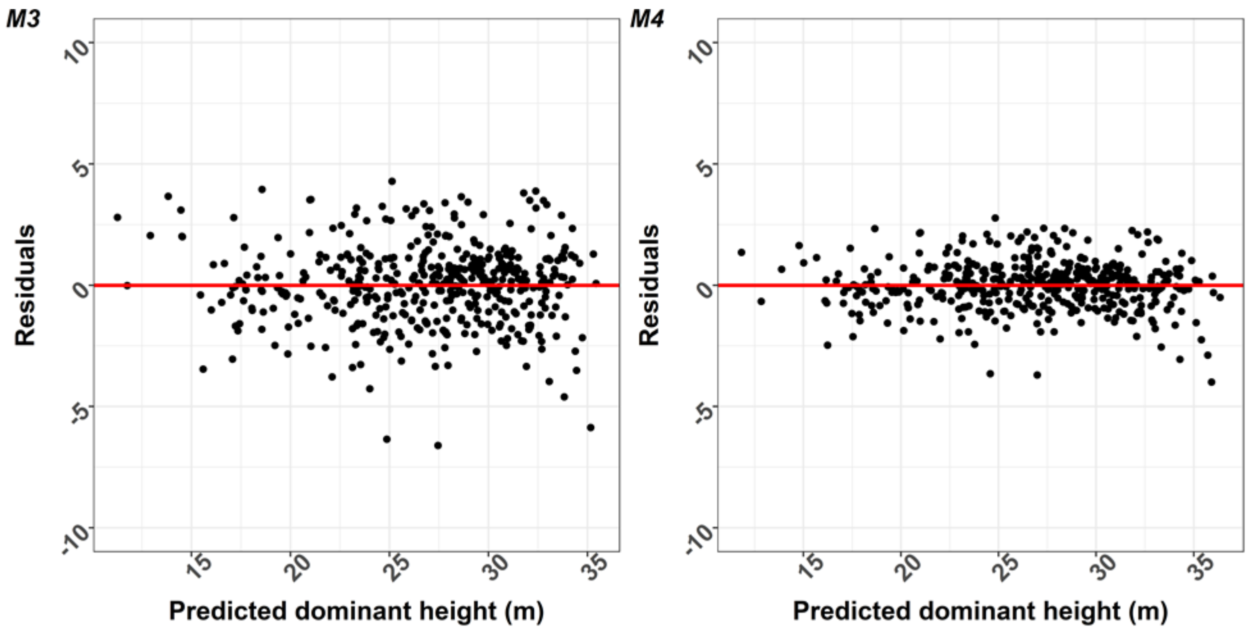

Figure 3.2. Residual plots for two polymorphic ADA models model *M3* and *M4*. It is clear from the plots that the distribution of residuals is better in model *M4* than *M3*.

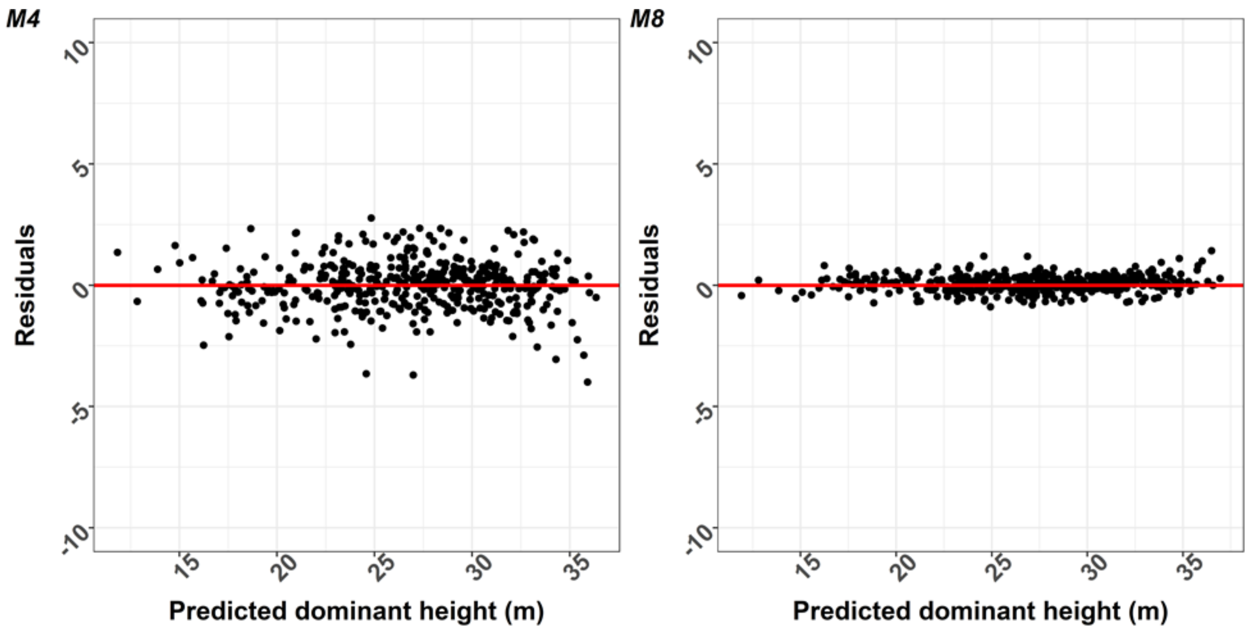

Figure 3.3. Residual plots for polymorphic ADA model (*M4*) and polymorphic EKF model (*M8*). As evidenced in the residual plot, the EKF model proved to be effective in minimizing the error in predictions.

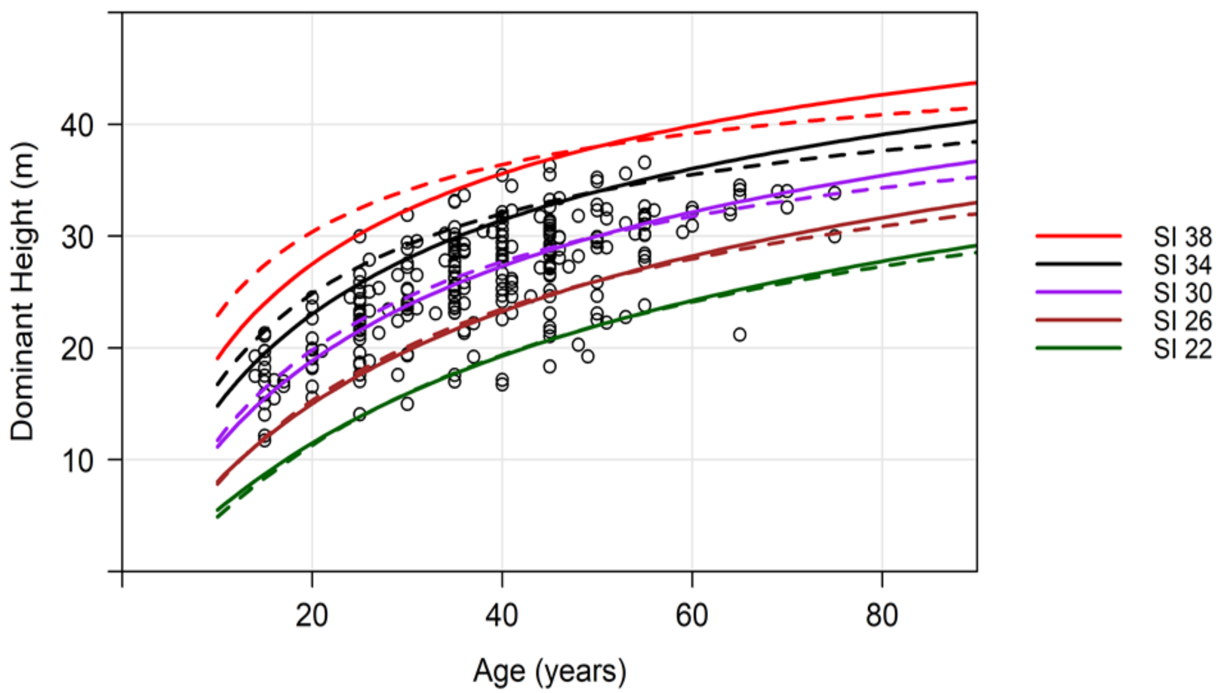

Figure 3.4. Dominant height curves constructed from two models (*M4* with the dashed line and *M8* with the bold line) for five site index values (22 m, 26 m, 30 m, 34, and 38 m) overlaid with the observed dominant height values shown in black circles on the background.

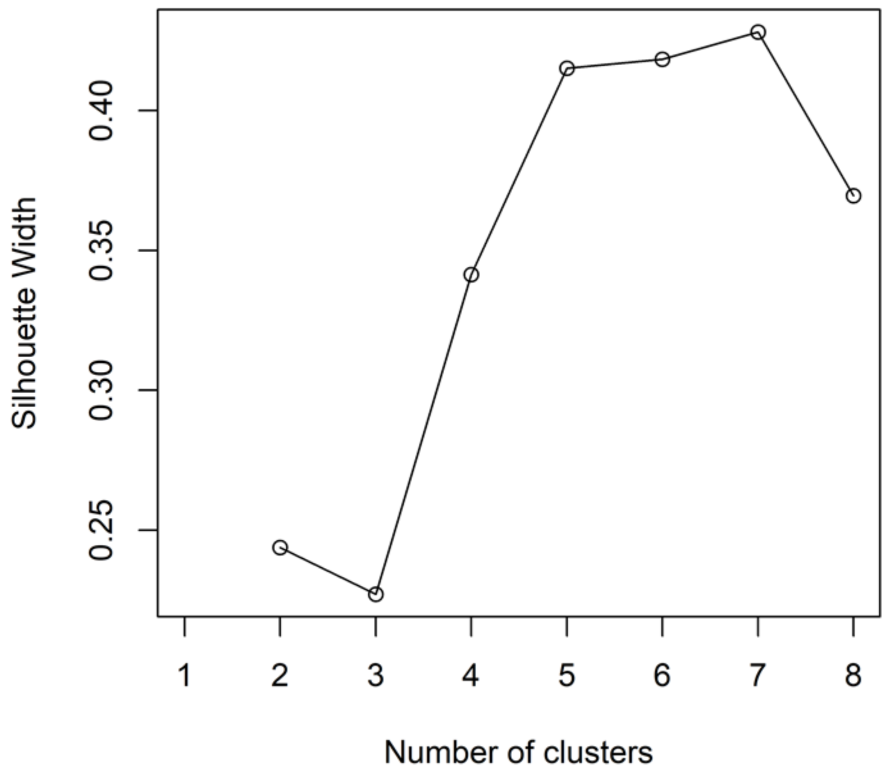

Figure 3.5. Number of potential clusters and Silhouette width resulted from the cluster analysis.

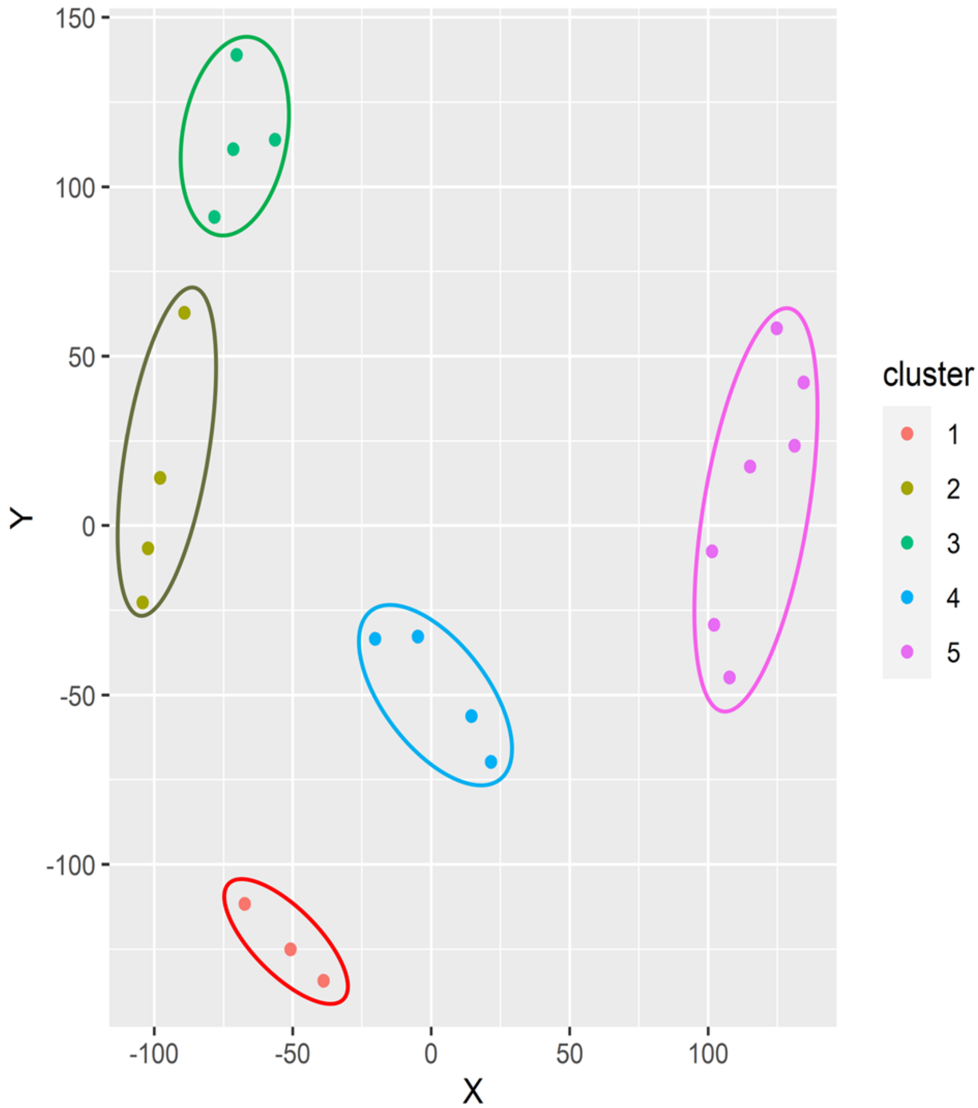

Figure 3.6. A cluster plot showing the division of hardwood tree species in five different clusters resulted from the cluster analysis.

References

Bailey, R.L., Clutter, J.L., 1974. Base-age invariant polymorphic site curves. For. Sci. 20, 155–159. https://doi.org/10.1093/forestscience/20.2.155

Bruce, D., 1926. A method of preparing timber-yield tables. J. Agric. Res. 32, 543–557.

Calama, R., Montero, G., 2004. Interregional nonlinear height-diameter model with random coefficients for stone pine in Spain. Can. J. For. Res. 34, 150–163. https://doi.org/10.1139/x03-199

Chapman, D.G., 1961. Statistical problems in population dynamics, in: Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability. University of California California, pp. 153–186.

Cieszewski, C.J., Bailey, R.L., 2000. Generalized algebraic difference approach: Theory based derivation of dynamic site equations with polymorphism and variable asymptotes. For. Sci. 46, 116–126. https://doi.org/https://doi.org/10.1093/forestscience/46.1.116

Diéguez-Aranda, U., Burkhart, H.E., Amateis, R.L., 2006. Dynamic Site Model for Loblolly Pine (Pinus taeda L.) Plantations in the United States, Forest Science. Oxford Academic. https://doi.org/10.1093/FORESTSCIENCE/52.3.262

Ehlers, S., Grafström, A., Nyström, K., Olsson, H., Ståhl, G., 2013. Data assimilation in stand-level forest inventories. Can. J. For. Res. 43, 1104–1113. https://doi.org/10.1139/CJFR-2013-0250/ASSET/IMAGES/CJFR-2013-0250TAB5.GIF

Einhorn, C., Buckley, C., 2021. Global Leaders Pledge to End Deforestation by 2030. New York Times.

Garcia, O., 1983. A Stochastic Differential Equation Model for the Height Growth of Forest Stands. Biometrics 39, 1059. https://doi.org/10.2307/2531339

García, O., 2019. Estimating reducible stochastic differential equations by conversion to a least-squares problem. Comput. Stat. 34, 23–46. https://doi.org/10.1007/S00180-018-0837-4/FIGURES/1

Hastie, T., Tibshirani, R., Friedman, J., 2009. The elements of statistical learning: data mining, inference, and prediction, Second. ed. Springer New York.

Hicks, R.R.J., Conner, W.H., Kellison, R.C., Van Lear, D., 2004. Silviculture and Management Strategies Applicable to Southern Upland Hardwoods, in: General Technical Report SRS-75. Asheville, NC. US Department of Agriculture, Forest Service, Southern Research Station, Chapter 7, pp. 51–62. https://doi.org/10.1017/CBO9781107415324.004

Kalman, R.E., Bucy, R.S., 1961. New Results in Linear Filtering and Prediction Theory. J. Basic Eng. 83, 95–108. https://doi.org/10.1115/1.3658902

Kaufman, L., Rousseeuw, P., 1990. Finding groups in data: An introduction to cluster analysis. Wiley, New York , NY.

Koirala, A., Montes, C.R., Bullock, B.P., 2021. Modeling dominant height using stand and water balance variables for loblolly pine in the Western Gulf, US. For. Ecol. Manage. 479, 118610. https://doi.org/10.1016/j.foreco.2020.118610

Lewis, S.L., Wheeler, C.E., Mitchard, E.T.A., Koch, A., 2019. Restoring natural forests is the best way to remove atmospheric carbon. Nat. 2021 5687750 568, 25–28. https://doi.org/10.1038/d41586-019-01026-8

Lundqvist, B., 1957. On the height growth in cultivated stands of pine and spruce in Northern Sweden. Medd Fran Statens Skogforsk 47, 1–64.

McNab, W.H., Keyser, T.L., 2020. A vegetative index of stand productivity based on tree inventory for predicting oak site index in the central Hardwood region. Can. J. For. Res. 50, 760–773. https://doi.org/10.1139/CJFR-2019-0412/ASSET/IMAGES/LARGE/CJFR-2019-0412F6.JPEG

Mctague, J.P., Loughlin, D.O., Roise, J.P., Robison, D.J., Kellison, R.C., 2008. The SOHARC Model System for Growth and Yield of Southern Hardwoods. South. J. Appl. For. 32, 173–183. https://doi.org/https://doi.org/10.1093/sjaf/32.4.173

Montes, C.R., 2012. A Resource Driven Growth and Yield Model for Loblolly Pine Plantations. North Carolina State University.

Nelson, M., Ritters, K., Coulston, J., Domke, G., Greenfield, E., Langner, L., Nowak, D., O'Dea, C., Oswalt, S., Reeves, M., Wear, D., 2020. Defining the United States Land Base. USDA For. Serv. Gen. Tech. Rep. NRS-191 70.

Nyström, M., Lindgren, N., Wallerman, J., Grafström, A., Muszta, A., Nyström, K., Bohlin, J., Willén, E., Fransson, J.E.S., Ehlers, S., Olsson, H., Ståhl, G., 2015. Data assimilation in forest inventory: First empirical results. Forests 6, 4540–4557. https://doi.org/10.3390/f6124384

Ogana, F.N., Gorgoso-Varela, J.J., 2020. A nonlinear mixed-effects tree height prediction model: Application to Pinus pinaster Ait in Northwest Spain. Trees, For. People 1, 100003. https://doi.org/10.1016/j.tfp.2020.100003

Olson, D.J., 1959. Site index curves for upland oak in the southeast. USDA For. Serv. Southeast. For. Exp. Station. Res. Pap. No. 125 2.

Orrego, S., Montes, C., Restrepo, H.I., Bullock, B.P., Zapata, M., 2021. Modeling height growth for teak plantations in Colombia using the reducible stochastic differential equation approach. J. For. Res. 32, 1035–1045. https://doi.org/10.1007/S11676-020-01174-Y/TABLES/4

Osuri, A.M., Gopal, A., Raman, T.R.S., Defries, R., Cook-Patton, S.C., Naeem, S., 2020. Greater stability of carbon capture in species-rich natural forests compared to species-poor plantations. Environ. Res. Lett. 15, 034011. https://doi.org/10.1088/1748-9326/AB5F75

Oswalt, S.N., Smith, W.B., Miles, P.D., Pugh, S.A., 2019. Forest resources of the United States, 2017, US Forest Service, Gen. Tech. Report WO-97, 223 p. https://doi.org/10.5962/bhl.title.101492

Park, H.S., Jun, C.H., 2009. A simple and fast algorithm for K-medoids clustering. Expert Syst. Appl. 36, 3336–3341. https://doi.org/10.1016/J.ESWA.2008.01.039

Pommerening, A., Muszta, A., 2015. Methods of modelling relative growth rate. For. Ecosyst. 2, 1–9. https://doi.org/10.1186/S40663-015-0029-4/FIGURES/7

Richards, F.J., 1959. A flexible growth function for empirical use. J. Exp. Bot. 10, 290–301.

Sharma, R.P., Brunner, A., Eid, T., Øyen, B.-H., 2011. Modelling dominant height growth from national forest inventory individual tree data with short time series and large age errors. For. Ecol. Manage. 262, 2162–2175. https://doi.org/10.1016/j.foreco.2011.07.037

Stage, A.R., 1963. A Mathematical Approach to Polymorphic Site Index Curves for Grand Fir.

For. Sci. 9, 167–180. https://doi.org/10.1093/FORESTSCIENCE/9.2.167

Walters, D.K., Burkhart, H.E., Reynolds, M.R., Gregoire, T.G., 1991. A Kalman Filter Approach to Localizing Height-Age Equations. For. Sci. 37, 1526–1537. https://doi.org/10.1093/FORESTSCIENCE/37.6.1526

Wang, M., Borders, B.E., Zhao, D., 2008. An empirical comparison of two subject-specific approaches to dominant heights modeling: The dummy variable method and the mixed model method. For. Ecol. Manage. 255, 2659–2669. https://doi.org/10.1016/J.FORECO.2008.01.030

Westfall, J.A., Hatfield, M.A., Sowers, P.A., Connell, B.M.O., Service, F., June, F.S., 2017. Site index models for tree species in the Northeastern United States 63, 283–290. https://doi.org/https://doi.org/10.5849/FS-2016-090

Zang, H., Lei, X., Zeng, W., 2016. Height–diameter equations for larch plantations in northern and northeastern China: a comparison of the mixed-effects, quantile regression and generalized additive models. Forestry 89, 434–445. https://doi.org/10.1093/forestry/cpw022

Zhao, D., Borders, B., Wilson, M., 2004. Individual-tree diameter growth and mortality models for bottomland mixed-species hardwood stands in the lower Mississippi alluvial valley. For. Ecol. Manage. 199, 307–322. https://doi.org/10.1016/j.foreco.2004.05.043

CHAPTER 3

MODELING WHOLE STAND MORTALITY OF EVEN-AGED SOUTHERN HARDWOOD FORESTS USING LOCALIZED INITIAL STAND DENSITY[2]

---

[2] Koirala, A., C.R. Montes, and B.P. Bullock. To be submitted to *Forest Science*.

Abstract

Whole stand mortality equation is an important component of stand growth and yield models. However, modeling future number of trees in a forest is relatively challenging than modeling growth components like height and volume. The main objective of this research is to develop whole stand mortality equation to predict future stand density of southern hardwood forests. We used permanent sampling plot data from even-aged mixed species hardwood stands measured all over the US South. The data consists of almost all major types of natural hardwood forest in the region. Models were constructed using differential equation approach, in which the underlying assumption for estimating tree number reduction was based on the relationships in which the relative rate of instantaneous mortality is related to stand age, site index, and stand densities. We introduced a localized parameter in the model to account for stand-specific initial density. Although past studies on loblolly pine reported direct effect of site index on mortality, our results from hardwood dataset showed site index was not significant in predicting mortality of hardwood forests.

Introduction

There is an unspoken consensus among statisticians and modelers that no perfect model exists in the world; statistical models are just an abstraction of the real world that seeks to build a logical relationship of different elements in a system (Weiskittel et al., 2011). In forestry, all forest management decisions are driven by the present and future conditions of the forest resources. As forests are continually changing over time, it is necessary to have information about the future forest dynamics for better forest management planning (Bettinger et al., 2009). In order to predict future forest conditions such as growth, mortality and other quantities, foresters rely on indirect method of predictions using prediction tables, statistical equations and computer simulation models (Burkhart et al., 2019). These techniques are collectively referred to as forest growth and yield models, and they are exclusively based on the statistical and conceptual relationship between several growth variables of trees and forest stands.

Among several important components of forest growth and yield systems, mortality models exhibit a unique and influential role in forest management planning (Álvarez-González et al., 2010; Amateis et al., 1997). Estimation of future number of trees in a given forest stand not only drives silvicultural decisions but also impacts the amount and volume of different products derived from the forest over future time horizons (Davis et al., 2001). For example, in an industrial forestry scenario, stand density at rotation age regulates the harvested volume of sawtimber and pulpwood. While in a carbon forestry scenario, it determines the quantity of forest carbon sequestered by the stand (Mildrexler et al., 2020). Despite its importance, mortality still remains one of the least understood components of the growth and yield system (Hamilton, 1986). Because of the complexity arising from interaction of plant growth with environment and other bio-physical factors, mortality has always been a difficult subject to model (Thapa and Burkhart, 2015). Many early growth and yield studies did not attempt to model mortality because of lack of large remeasurement data. Several approaches to model even-aged stand mortality were introduced since 1970s. First out of two major approaches involves predicting number of future trees directly from all available data using current information like age, stand density and/or site index (Scolforo et al., 2019; Thapa and Burkhart, 2015; Zhao et al., 2007). This approach is based on algebraic difference approach (ADA) proposed by Bailey and Clutter (1974). The functional form of these types of equation can be represented as $N_2 = f(N_1, A_2, A_1, SI)$, where SI is the site index of the stand and N_i is the number of trees per unit area at age A_i such that $A_1 \leq A_2$. The second approach also known as "two-step mortality modeling" is carried out in two modeling phases (González et al., 2004; Woollons, 1998; Zhao et al., 2007). In the first step, the probability of death or non-death occurrence of a tree is predicted using all available data. And, in the second step, mortality model is utilized to estimate the reduction of tree numbers in the stand using only plot where mortality had occurred. Finally, using the probability the adjusted number of trees in the future is calculated. Most of the mortality models developed for plantation pines in the US South have employed either one or combination of both aforementioned approaches. Unfortunately, a limited number of studies on modeling mortality of southern hardwood exists in the literature.

In terms of natural southern hardwoods, the G-HAT model system incorporated survival equation in their growth and yield system (Harrison et al., 1986). The authors fitted a single nonlinear regression to predict the five-year probability of survival for a tree of any species using annual basal area increment as an independent variable. The total number of surviving trees would then be calculated by multiplying the survival probability to the initial density for each age-class. Bowling et al. (1989) was the first study to include ADA type mortality equations in a system for southern upland hardwood forests. The authors fitted and evaluated two mortality models to their

Appalachian hardwood data, which in fact were previously published for plantation forests (Clutter and Jones, 1980; Pienaar and Shiver, 1981). Later, McTague et al. (2008) utilized Clutter and Jones (1980) model to develop a new merchantable stand survival model for southern hardwoods. For bottomland hardwoods, Schultz et al. (2010) developed a Chapman-Richards type stand density prediction model for red oak – sweetgum forests of mid-south minor stream bottoms. In lieu of predicting future stand density from current stand density, this model directly predicted future stand density from current stand age and site index. These examples clearly depict the lack of mortality prediction models for natural hardwoods in the US South. With the advancement of computational skills at current time, we believe a more robust mortality equations for southern hardwoods could be developed. The main objective of this research is to develop whole stand mortality model for even-aged mixed species southern hardwoods. For this, we started with selecting four commonly used ADA type mortality models. We then introduced a novel concept of localizing initial stand density in those models. Unlike plantations, acquiring information on initial stand density of natural forests is a challenging task. Thus, instead of utilizing observed initial density, we proposed to utilize predicted initial density during model fitting.

Methods

Data

Data utilized in this research came from the Hardwood Research Cooperative (HRC) of the North Carolina State University (McTague et al., 2008). More than 600 plots were established across the US South in cooperation with member forest companies of the HRC. Since natural hardwood forests cover a wide range of geographical regions within the US South, the first task was to identify the forest site types. The cooperative identified nine forest site types which could be classified as separate forest operating units (Table 1). Among 600 plots, 187 plots were permanently maintained for repeated measurements and as a data source for growth and yield prediction analysis. The forest areas in the established plots were even aged and fully stocked stands. Majority of the new growth natural forests in the US South is even-aged, thus the study ignored uneven-aged and high graded forest stands. There were five age groups in the data collection: ranging from 20 to 60 years at 10-year intervals. Table 1 presents the distribution of plots by age class and site types.

Model Formulation

We utilized differential equation approach to develop mortality functions for predicting tree number reduction. The underlying assumption for estimating tree number reduction was based on the relationships in which the relative rate of instantaneous mortality is related to stand age, site index, and stand densities. For the first model, our assumption was the relative rate of change in number of trees is related to the stand initial density and age (Eq. 5.1).

$$\frac{1}{N}\frac{dN}{dA} = \alpha + \frac{\delta}{A} \qquad (5.1)$$

where *N* is the number of trees per acre at age *A, dN/dA* is the instantaneous mortality rate at age A and α, δ are unknown parameters. We integrated Eq. (5.1) with initial condition that $N = N_0$ when $A = A_0$, which resulted in our first tree number reduction model for predicting $N = N_t$ at remeasurement age $A = A_t$ (Eq. 5.2).

$$N_t = N_0\, exp\big(b_1(A_t - A_0)\big)\left(\frac{A_t}{A_0}\right)^{b_2} \qquad (5.2)$$

where, $b_1$and b_2 are parameters to be estimated in the model. For the second model, our assumption of rate of relative change in tree numbers remains the same. However, we utilized a different differential equation by modifying the effect of the stand age in predicting the change in tree number (Eq. 5.3).

$$\frac{1}{N}\frac{dN}{dA} = \alpha N^{\beta}\delta^{A} \qquad (5.3)$$

Integration of Eq. (3) with previously described initial and final conditions resulted in our second tree number reduction model (Eq. 5.4).

$$N_t = \left[N_0^{b_0} + b_1\left({b_2}^{A_t} - {b_2}^{A_0}\right)\right]^{\left(\frac{1}{b_0}\right)} \qquad (5.4)$$

where, b_0 , $b_1$and b_2 are parameters to be estimated in the model. We introduced the effect of site index on instantaneous rate of mortality for our third model (Eq. 5.5). Results from several studies showed that site index could be an important variable in predicting mortality, as mortality generally increased with higher site index (Bailey et al., 1985; Stankova and Diéguez-Aranda, 2014; Zhao et al., 2007).

$$\frac{1}{N}\frac{dN}{dA} = \alpha SI^{\gamma} \qquad (5.5)$$

Integration of Eq. (5.5) with previously described initial and final conditions resulted in our third tree number reduction model (Eq. 5.6).

$$N_t = N_0 exp\left(b_1 SI^{b_2}(A_t - A_0)\right) \qquad (5.6)$$

Where, *SI* is site index of the stand. Finally, for our fourth and last model, in addition to site index, we modified the effect of stand density on the instantaneous rate of mortality (Eq. 5.7). Integration of which with aforementioned initial and final stand conditions gave our fourth mortality model (Eq. 5.8).

$$\frac{1}{N}\frac{dN}{dA} = \alpha N^{\beta} S^{\gamma} \qquad (5.7)$$

$$N_t = \left[N_0^{b_0} + b_1 SI^{b_2}(A_t - A_0)\right]^{\left(\frac{1}{b_0}\right)} \qquad (5.8)$$

Localizing Initial Stand Density

In this section, we describe the process of localizing initial stand density for each stand in our dataset. Instead of utilizing observed initial density, we proposed to utilize predicted initial density during model fitting. In our four mortality models (Eqs. 5.2, 5.4, 5.6 and 5.8), we substituted observed initial stand density N_0 with a local parameter χ_j. This local parameter was assigned to individual stand and was estimated as part of the optimization process. In order to assign this parameter in our models, we created a dummy matrix using individual stand ID. This method generated a unique numerical value for each stand, which can also be termed as site-specific parameter. Since we changed the initial stand density with a local parameter, we assumed that the initial stand age A_0 ($A_0 : A_0 \leq A_t$) should be fixed for each stand. For this, the age of the stand during first inventory measurement was used as initial stand age. This is different than traditional ADA approach, in which the initial stand density is not fixed. With this change, our ADA type mortality models (Eqs. 5.2, 5.4, 5.6 and 5.8) transformed into Eq. (5.9), Eq. (5.10), Eq. (5.11) and Eq. (5.12), respectively.

Model M1: $$N_{t,j} = \chi_j\, exp\big(b_1(A_t - A_0)\big)\left(\frac{A_t}{A_0}\right)^{b_2} \qquad (5.9)$$

Model M2: $$N_{t,j} = \left[\chi_j^{b_0} + b_1\left({b_2}^{A_t} - {b_2}^{A_0}\right)\right]^{\left(\frac{1}{b_0}\right)} \qquad (5.10)$$

Model M3: $N_{t,j} = \chi_j exp\left(b_1 SI^{b_2}(A_t - A_0)\right)$ (5.11)

Model M4: $N_{t,j} = \left[\chi_j^{b_0} + b_1 SI^{b_2}(A_t - A_0)\right]^{\left(\frac{1}{b_0}\right)}$ (5.12)

where, χ_j a site-specific parameter for plot j, representing estimated initial stand density at stand age A_0.

Parameter Estimation and Cross-validation

Before parameter estimation, we conducted Ljung-Box test (Ljung and Box, 1978) using **`Box.test`** function of **`stats`** package in R software to check autocorrelation in the dataset. The test revealed no significant autocorrelation between the measurements of same independent variables across different time period. We utilized a maximum likelihood (ML) approach to solve for parameters of the mortality models. The objective function to minimize in ML approach assumes a normal distribution for the expected values (Eq. 5.13). Using **`optimr`** package in R version 4.1.0, we obtained the parameter estimates of the model. In order to calculate standard error of each parameter, we took square root of the diagonal of the inverse Hessian matrix, also known as the delta method.

$$argmin\, f(x,\beta) \sim \sum_{i=1}^{n} log\left(\frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{(f(x,\beta) - Y_i)^2}{2\sigma^2}\right)\right) \quad (5.13)$$

where, function $f(x,\beta)$ represents our yield models with independent variables x and parameters β; Y_i represents observed stand density; and σ^2 represents the variance. In order to control the heteroscedasticity of the model, we modeled the residual variance as a power function of the fitted values (Eq. 5.14).

$$Var(\varepsilon_t) = \sigma_t^2 = (r^2)\, \widehat{Y_t^{\gamma}} \quad (5.14)$$

where, r = residual standard error; $\hat{Y}$= predicted densities; t is time and γ is variance parameter to be estimated.

For model validation, we utilized a robust five-fold cross-validation approach during data fitting process (Hastie et al., 2009). This approach involves randomly dividing the set of observations into K groups or folds (5 groups in our case), of approximately equal size. The first group of the data is held out as a validation set and the remaining folds i.e. (K-1) are used to train the model (Figure 1). Model fits and cross-validation results of difference from candidate models were evaluated using three model evaluation measures: root mean square error (RMSE), mean absolute error (MAE) and Akaike information criterion (AIC).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n}}$$

$$MAE = \frac{\sum_{i=1}^{n}|Y_i - \hat{Y}_i|}{n}$$

$$AIC = -2\ loglik + 2k$$

where, k is the number of parameters in the model, Y_i is the observed trees per acre, and $\hat{Y}_i$ is the predicted trees per hectare.

Results

We introduced a novel concept of localizing initial stand density in ADA type whole-stand mortality models for southern hardwoods. We utilized maximum likelihood approach to obtain parameter estimates from the models and performed five-fold cross validation to select the best model. Table 2 illustrates the parameter estimates and standard errors for each model. All the parameter estimates of the models M1, M2 and M4 were significant at the 1% level. However, in model M3, parameter b_2 was not significant at 1% level with a p-value of 0.0453. This parameter represents the effect of site index in the mortality of hardwood stand. On the other hand, the p-value for b_2 parameter was extremely lower in model M4, indicating the significant effect of site index on predicting tree number reduction.

The RMSE values for all models ranged from 104.36 to 108.27 trees per hectare (Table 3). Model M2 depicted lowest RMSE values in both fit (104.36 TPH) and cross validation statistics (101.29 TPH) followed by model M4 (fit: 107.61 TPH; cross-validation: 105.44 TPH). Model M2 also had the lowest MAE (or absolute bias) value of 84.11 for fitting dataset, which was closely followed by model M4 with 84.70 TPH. Models M2 and M4 had the lowest values of MAE from five cross-validation process as well. In a cross-validation process, a positive bias on one validation set can offset a negative bias on another validation dataset. Therefore, the bias values from traditional approach are generally lower. To avoid that, we utilized absolute values of the bias. The AIC value of M4 was lowest among all four models. Overall, model M3 had the poorest fit with highest RMSE, MAE and AIC values, while model M2 had lowest RMSE and MAE values, and Model M4 had lowest AIC value. Therefore, based on the fit and validation statistics, model M2 was selected to be the best model. However, all models were further compared in terms of distribution of residuals and predictive performance.

The graphical evaluation of residual plots for four models showed that models M3 and M4 displayed some degree of bias and heteroscedasticity (Figure 2). Previously, based on fit and validation statistics, model M4 was performing second best after model M2. However, the residual analysis clearly demonstrated its shortcoming. The residuals of models M1 and M2 seemed to be symmetrically distributed around zero line indicating less bias and constant variance of residuals. In addition to the residual plots, we also evaluated the projected future trees per hectare of the stand using 6 different initial densities, 500, 700, 900, 1100, 1300 and 1500 TPH (Figure 3). The initial stand age was 10 years, and the projection horizon was 80 years. For model M3 and M4, we assumed site index to be 25m. Projection lines from model M2 depicted biological property of forest stand dynamics as hardwoods are slow growing species the resource driven mortality starts at later age. Projections from other three models suggested steady decline in stand density over time.

Discussion and Conclusion

Although not explicitly grouped, natural mortality in a forest stand can be classified as catastrophic and non-catastrophic (Thapa and Burkhart, 2015; Vanclay, 1995). The catastrophic mortality are results of sudden and random disturbances such as wildfire, insect outbreaks, tornado, snowstorm etc. The non-catastrophic mortality (or regular mortality) results from the competition for scarce resources among trees within a stand. A typical even-aged growth and yield system includes mortality model to predict non-catastrophic mortality rather than catastrophic one (Monserud, 1976; Monserud and Sterba, 1999). This does not mean there are fewer studies on catastrophic mortality. In fact, a review of the effect of natural disturbances on forest ecosystems showed many studies on catastrophic mortality over the period (Seidl et al., 2011). However, to our knowledge, the growth and yield system utilized in the US South have not yet incorporated models predicting mortality of catastrophic nature. Majority of the non-catastrophic mortality models developed in the US South are for monoculture softwood species such as loblolly or slash pines. There are comparatively lower number of studies on southern hardwoods' survival and mortality. Lack of permanent research trials for hardwood forest could have created difficulty in obtaining repeated measurements data that are utilized for growth and yield analysis.

This paper developed a whole-stand survival/mortality equation for even-aged unthinned southern hardwood forests. For modeling purpose, we utilized direct estimation approach of predicting number of trees in the future stand. Another popular technique to model mortality is a two-step approach. Several past studies have reported better results from direct estimation approach as compared to the two-step approach (Dieguéz-Aranda et al. 2005; Scolforo et al. 2019; Thapa and Burkhart 2015). Dieguéz-Aranda et al. (2005) reported that direct estimation technique could be a better approach for unthinned fully stocked forests such as in our study. In addition, this technique also warrants the path-invariance property desired in projection equations. Thapa and Burkhart (2015) evaluated both approaches to estimate stand mortality or survival for loblolly pine plantations in the US South. Although similar level of accuracy was reported from both approaches, the authors recommended the direct estimation as their preferred approach because of simplicity in the modeling process. Based on these examples, we did not include two-step modeling approach in our analysis.

Unlike plantations, acquiring information on initial stand density of natural forests could be a dauting task. Thus, we introduced a novel concept of localizing initial stand density. For this, we substituted observed initial stand density N_0 with a local parameter χ_j for each plot in the data. Results showed that model M2 outperformed other candidate models for predicting future number of trees in the stand. The model utilized future stand age and initial stand density as the main predictor variables for predicting future stand density. This result contradicts with studies carried out for plantation pines in the southern US (Thapa and Burkhart, 2015; Zhao et al., 2007). In those studies, models including site index and age were reported to perform best for pine plantations. Our results coincide with that of McTague et al. (2008), where authors reported model without site index performed better for southern hardwoods. However, the mathematical formulation of

McTague et al (2008) and model M2 is different. Scolforo et al (2019) also reported that model without site index performed better and thus included environmental variable to account for site variability. However, Thapa and Burkhart (2015) conclude that inclusion of climatic variables in mortality model provided only a little advantage. Notably, in their Timber management textbook, Clutter et al. (1983) also acknowledged that several growth and yield analyses failed to demonstrate any effect of site index on tree and stand mortality. We evaluated two mortality models (M3 and M4) that had site index as predictor variable. In model M3, the effect of site index was not significant at 1% level confidence, but it was significant in model M4. Overall, model M4 showed second best results and closely followed model M2 in fit and cross-validation statistics. We also evaluated residual plots of all four models, in which once again model M2 outperformed other models. Therefore, model M2 was recommended as the best model to predict future number of trees for southern hardwood stands. We observed higher RMSE values for all models that ranged between 104 and 108 trees per hectare The higher RMSE value in tree number prediction through direct estimation approach was reported in previous studies as well (Dieguéz-Aranda et al. 2005; Thapa and Burkhart 2015). This is a clear indication that in general mortality models have higher error values because of high variability in tree mortality among different site types. In this regard, to account for some sort of site level variability, we will also recommend parameter estimates of model M4 for future density prediction despite it showed non-constant variance of residuals. Forest managers have the flexibility to use either of the models based on their data availability.

In conclusion, our evaluation showed that model M2 with initial stand density and age as predictor variables performed best in terms of model evaluation statistics and performance of the residuals. Since this model accounts for stand-specific initial density as a parameter, inclusion of site index in the model does not seem necessary. Further evaluation with soil and microclimatic information combined with permanent plot data could shed light on site specific mortality of hardwood forests.

Tables and Figures

Table 5.1. Number of hardwood growth and yield plots by site type and age class across the southern United States. The stand ages in the plots were determined using past records and increment cores taken from dominant and codominant trees.

Site type	Age classes (year)					
	20	**30**	**40**	**50**	**60**	**Total**
Muck swamp	88	11	8	99	13	49
Peat swamp	1	5	2	11	2	11
Wet flat	7	35	19	12	3	76
Red river bottom	8	18	15	11	6	58
Black river bottom	5	14	24	2	7	52
Branch bottom	12	33	24	18	5	92
Bottomland	36	59	40	23	7	165
Coves, gulfs, lower slopes	5	8	17	7	4	41
Upper slopes and ridges	5	27	34	18	13	97
Total	87	210	183	101	60	641

Table 5.2. Parameter estimates and standard error values for four mortality models.

Model	Parameter	Estimate	Standard error	P-Value
M1	b_0	-	-	-
	b_1	-0.0591	0.0411	<.0001
	b_2	0.5325	0.0536	<.0001
M2	b_0	-0.3513	0.1082	<.0001
	b_1	0.0028	0.0124	<.0001
	b_2	1.0715	0.2859	<.0001
M3	b_0	-	-	-
	b_1	-0.0034	0.0247	<.0001
	b_2	0.4529	0.2918	0.0453
M4	b_0	0.2252	0.2705	<.0001
	b_1	-0.0867	0.0371	<.0001
	b_2	-0.4113	0.2303	<.0001

Table 5.3. Fit and cross validation statistics (RMSE, MAE and AIC) for four mortality models.

Model	Fit statistics			Cross-validation statistics		
	RMSE	**MAE**	**AIC**	**RMSE**	**MAE**	**AIC**
Model M1	108.25	85.45	5014.11	110.43	84.41	5011.82
Model M2	**104.36**	**84.11**	5008.41	**101.29**	**82.16**	5007.94
Model M3	108.27	86.56	5018.92	107.35	86.17	5014.74
Model M4	107.61	84.70	**5006.59**	105.44	83.34	**5005.88**

Figure 5.1. Schematic representation of five-fold cross-validation approach. Each horizontal bar is divided into five sections, out of which four are training folds and one section is the validation fold. There are altogether five horizontal bars representing five iterations or folds. Goodness-of-fit statistics (GOF) are calculated for each iteration and the final GOF statistics for a model is calculated as the average of five GOF statistics from five iterations.

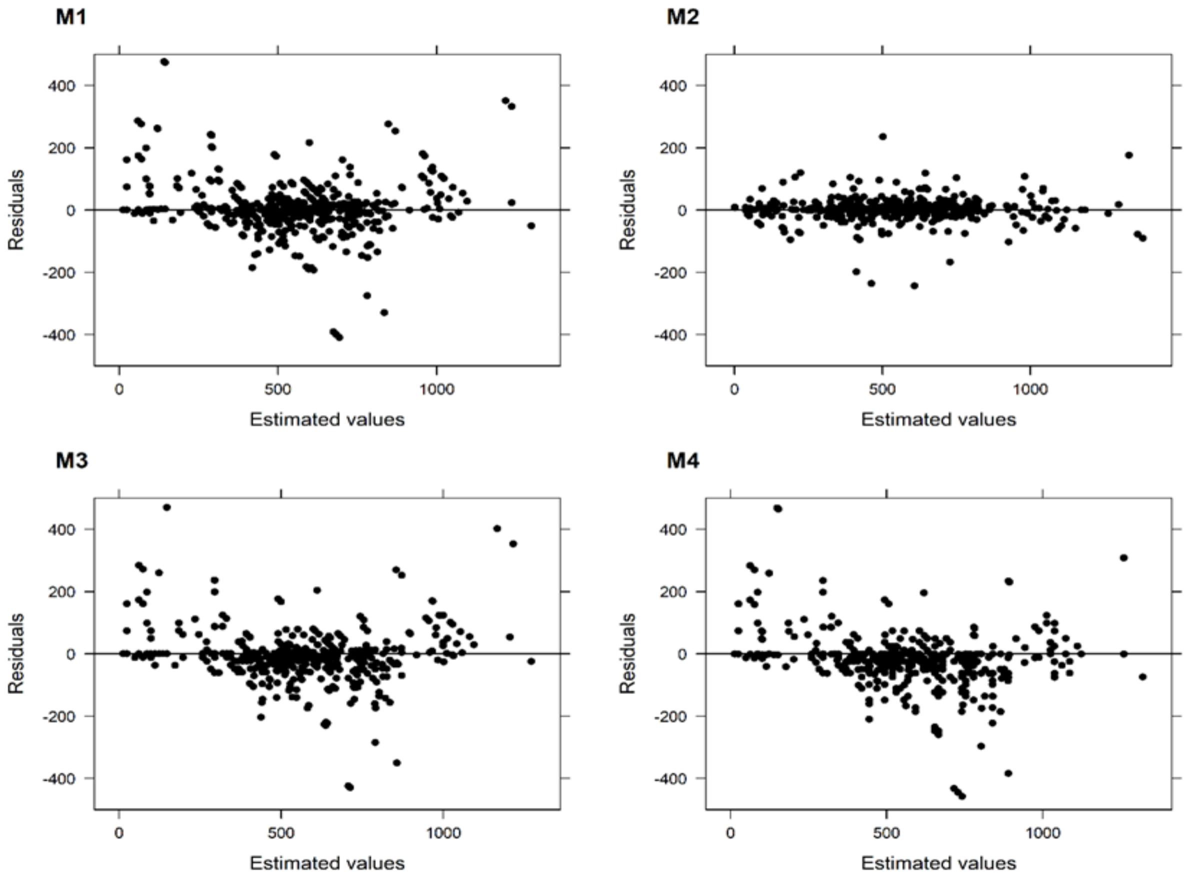

Figure 5.2. Residual plots for all four models.

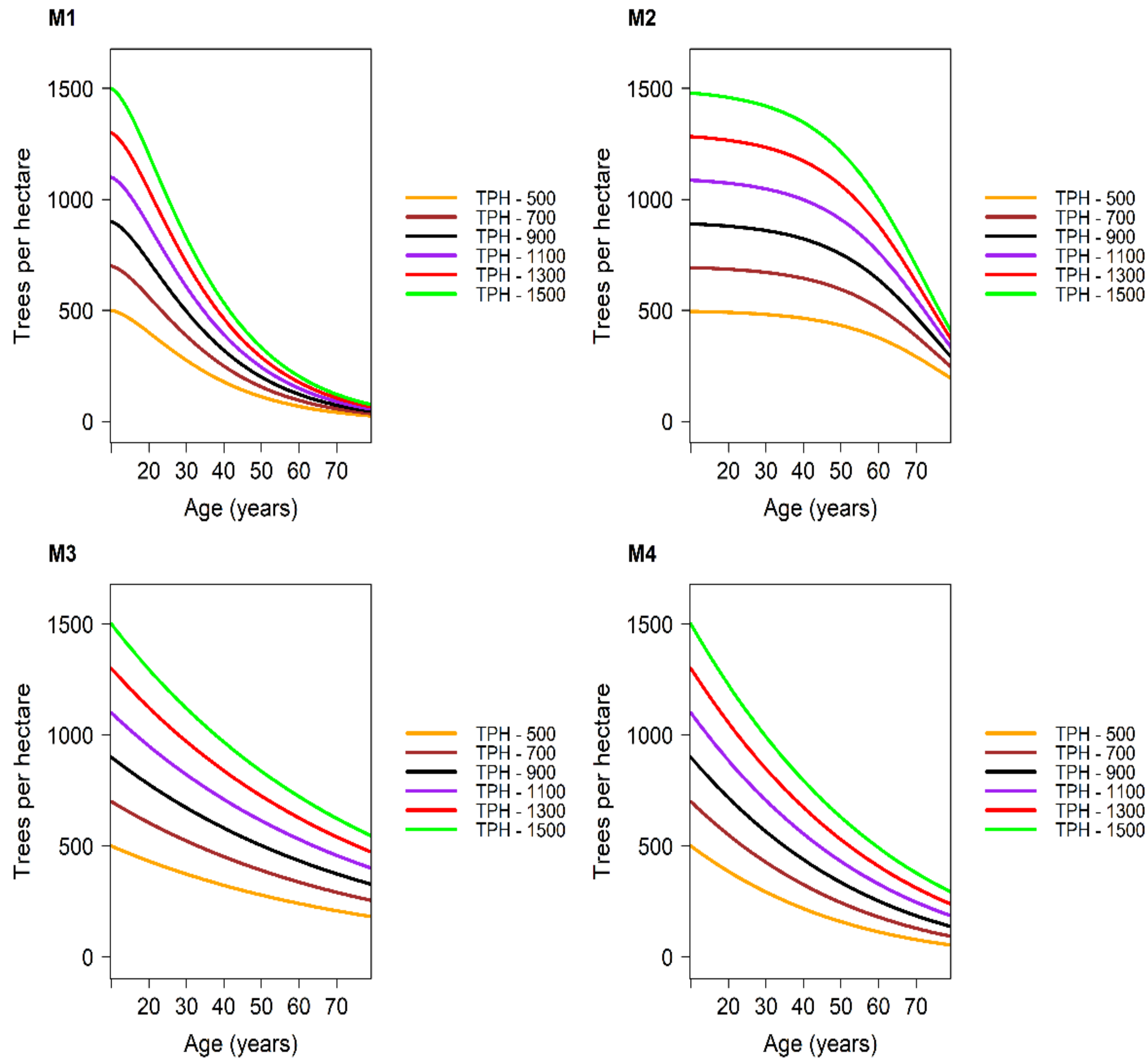

Figure 5.3. Plots showing projection of future stand densities of hardwood stands with given initial condition. Six initial stand densities of 500 (orange), 700 (brown), 900 (black), 1100 (purple), 1300 (red) and 1500 (green) trees per hectare were used for projection. A site index of 25 m was used for models with site index as prediction variable (M3 and M4).

References

Álvarez-González, J.G., Zingg, A., Gadow, K. V, 2010. Estimating growth in beech forests: a study based on long term experiments in Switzerland. Ann. For. Sci. 67, 307.

Amateis, R.L., Burkhart, H.E., Liu, J., 1997. Modeling survival in juvenile and mature loblolly pine plantations. For. Ecol. Manage. 90, 51–58.

Bailey, R.L., Borders, B.E., Ware, K.D., Jones, E.P., 1985. A compatible model relating slash pine plantation survival to density, age, site index, and type and intensity of thinning. For. Sci. 31, 180–189.

Bailey, R.L., Clutter, J.L., 1974. Base-age invariant polymorphic site curves. For. Sci. 20, 155–159. https://doi.org/10.1093/forestscience/20.2.155

Bettinger, P., Boston, K., Siry, J., Grebner, D., 2009. Forest management and planning. Elsevier, Burlington, MA.

Bowling, E.H., Burkhart, H.E., Burk, T.E., Beck, D.E., 1989. A Stand-level Multispecies Growth Model for Appalachian Hardwoods. Can. J. For. Res. 19, 405–412. https://doi.org/https://doi.org/10.1139/x89-064

Burkhart, H.E., Avery, T.E., Bullock, B.P., 2019. Forest Measurements, Sixth. ed. Waveland Pr Inc.

Clutter, J.L., Jones, E.P., 1980. Prediction of growth after thinning in old-field slash pine plantations. Department of Agriculture, Forest Service, Southeastern Forest Experiment ….

Davis, L.S., Johnson, K.N., Bettinger, P., Howard, T.E., 2001. Forest management: to sustain ecological, economic, and social values . 2001, First. ed. McGraw-Hill, New York , NY.

González, J.G.Á., Dorado, F.C., González, A.D.R., Sánchez, C.A.L., von Gadow, K., 2004. A two-step mortality model for even-aged stands of Pinus radiata D. Don in Galicia (Northwestern Spain). Ann. For. Sci. 61, 439–448.

Hamilton, D.A., 1986. A logistic model of mortality in thinned and unthinned mixed conifer stands of northern Idaho. For. Sci. 32, 989–1000.

Harrison, W.C., Burkhart, H.E., Burk, T.E., Beck, D.E., 1986. Growth and yield of Appalachian mixed hardwoods after thinning. Sch. For. Wildl. Resour. Virginia Ploytechnic Inst. State Univ. Publ. No. FWS-1-86 48.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The elements of statistical learning: data mining, inference, and prediction, Second. ed. Springer New York.

Mctague, J.P., Loughlin, D.O., Roise, J.P., Robison, D.J., Kellison, R.C., 2008. The SOHARC Model System for Growth and Yield of Southern Hardwoods. South. J. Appl. For. 32, 173–183. https://doi.org/https://doi.org/10.1093/sjaf/32.4.173

Mildrexler, D.J., Berner, L.T., Law, B.E., Birdsey, R.A., Moomaw, W.R., 2020. Large Trees Dominate Carbon Storage in Forests East of the Cascade Crest in the United States Pacific Northwest . Front. For. Glob. Chang. .

Monserud, R.A., 1976. Simulation of forest tree mortality. For. Sci. 22, 438–444.

Monserud, R.A., Sterba, H., 1999. Modeling individual tree mortality for Austrian forest species. For. Ecol. Manage. 113, 109–123.

Pienaar, L. V, Shiver, B.D., 1981. Survival functions for site-prepared slash pine plantations in the flatwoods of Georgia and northern Florida. South. J. Appl. For. 5, 59–62.

Schultz, E.B., Iles, J.C., Matney, T.G., Ezell, A.W., Meadows, J.S., Leininger, T.D., Booth, W.C., Jeffreys, J.P., 2010. Stand-Level Growth and Yield Component Models for Red Oak–Sweetgum Forests on Mid-South Minor Stream Bottoms. South. J. Appl. For. 34, 161–175. https://doi.org/https://doi.org/10.1093/sjaf/34.4.161

Scolforo, H.F., McTague, J.P., Burkhart, H., Roise, J., Alvares, C.A., Stape, J.L., 2019. Modeling whole-stand survival in clonal eucalypt stands in Brazil as a function of water availability. For. Ecol. Manage. 432, 1002–1012. https://doi.org/https://doi.org/10.1016/j.foreco.2018.10.044

Seidl, R., Fernandes, P.M., Fonseca, T.F., Gillet, F., Jönsson, A.M., Merganičová, K., Netherer, S., Arpaci, A., Bontemps, J.-D., Bugmann, H., 2011. Modelling natural disturbances in forest ecosystems: a review. Ecol. Modell. 222, 903–924.

Stankova, T.V., Diéguez-Aranda, U., 2014. Derivation and analysis of new stand-level mortality models based on existing growth equations. Ecol. Res. 29, 319–330.

Thapa, R., Burkhart, H.E., 2015. Modeling Stand-Level Mortality of Loblolly Pine (Pinus taeda L.) Using Stand, Climate, and Soil Variables. For. Sci. 61, 834–846. https://doi.org/10.5849/forsci.14-125

Vanclay, J.K., 1995. Synthesis: Growth models for tropical forests: A synthesis of models and methods. For. Sci. 41, 7–42.

Weiskittel, A.R., Hann, D.W., Kershaw Jr, J.A., Vanclay, J.K., 2011. Forest growth and yield modeling. John Wiley & Sons.

Woollons, R.C., 1998. Even-aged stand mortality estimation through a two-step regression process. For. Ecol. Manage. 105, 189–195.

Zhao, D., Borders, B., Wang, M., Kane, M., 2007. Modeling mortality of second-rotation loblolly pine plantations in the Piedmont / Upper Coastal Plain and Lower Coastal Plain of the southern United States 252, 132–143. https://doi.org/10.1016/j.foreco.2007.06.030

CHAPTER 4

SIMULTANEOUS PREDICTION – PROJECTION BASAL AREA EQUATIONS FOR MIXED SOUTHERN HARDWOOD FOREST STANDS INCORPORATING UNCERTAINTY ESTIMATORS[3]

[3] Koirala, A., C.R. Montes, and B.P. Bullock. To be submitted to *Forests*.

Abstract

Unlike plantations, the stand structure of natural forests is complex and composed of mixed tree species. Predicting the growth and yield of hardwood stands is challenging because of inconsistent growth rates and different management requirements for different species. Consequently, there are limited number of studies on natural hardwood stand-level growth and yield models. Basal area is an important measure of tree growth in stand level forest management. In order to estimate the basal area of a stand at a certain age, a prediction model with different stand-level variables is required. However, to estimate future basal area, a projection model based on the current basal area combined with the associated stand-level variables is essential. Generally, information from the prediction model is used to fit the projection model and parameter estimates can be obtained either one at a time or simultaneously for prediction and projection. The main aim of this study is to fit simultaneous prediction and projection equations of basal area growth using traditional Pienaar basal area model (1986) and compare it with the novel technique that incorporates weighted factor assimilating two sources of errors, prediction and projection. The differences on the estimated coefficients and the goodness-of-fit statistics between these two techniques are also evaluated. Our results showed that the introduction of weighting factor was able to improve estimation of future stand basal area while reducing bias and increasing precision of the estimates.

Introduction

In a typical production forestry scenario, forest managers continuously monitor the growth of the stand and prescribe appropriate silvicultural treatments at increasing stand ages based on the merchantability of the trees and market demands. Most of these management prescriptions rely on the future projection of the forest stands, utilizing some sort of empirical models that are fed with past and current stand information. Thus, it is important to have a reliable and accurate growth and yield model system for effective stand management.

Stand basal area is an essential component of the forest growth model system because it is highly correlated with important economic variables of the forest stands such as total volume and quadratic mean diameter (Barrio Anta et al., 2006; Zhao and Li, 2013). In addition, basal area can be used as a constrain or a transition function to form a link between low resolution (individual-tree and size-class) and high resolution (whole-stand) models (Cao, 2021; Castedo Dorado et al., 2006; Rodríguez et al., 2010). The basal area growth describes the biological property of forest stand dynamics over time including self-thinning of suppressed trees and carrying capacity of the stand (Anton-Fernandez et al., 2012; Lam and Guan, 2020). Thus, it is widely used in forest management for its ability to update inventories, predict future yield and to explore management alternatives. Basal area in combination with stand density also plays a decisive role in dictating the intensity and timing of thinning, intermediate cut and final harvesting (Chikumbo et al., 1999; Sun et al., 2007). The popularity of basal area in forest models stems from the fact that its measurement on the ground is relatively easier and more accurate than height measurement.

The evolution of stand-level models gained momentum in the US South after an important contribution from Clutter (1963). In this paper, the author derived yield prediction equations of forest stands by integrating the growth rate equations, ensuring compatibility between growth and yield. Since then, many authors developed stand basal area models using a similar methodology. Clutter and Jones (1980) developed a single equation that can project future basal area based on stand age and basal area at the start of the projection period. Pienaar and Shiver (1986) derived two stand basal area equations (prediction and projection equations) using the Schumacher function as their base model. The prediction equation was able to predict stand basal area using information such as stand age, stand density and dominant height. The projection equation, on the other hand, was able to project forward the future stand basal area based on current basal area, age, density and dominant height information. In order to achieve consistent estimates, the projection equation should hold three desirable properties, biologically sound, simple (parsimonious) and path invariance (Barrio Anta et al., 2006; Castedo-Dorado et al., 2007). The first property "biologically sound" means the stand basal area function must possess asymptotic value (maximum value of the basal area in a forest stand) when stand age increases to infinity. The second property "simple" means the model must be parsimonious and should avoid unnecessary interaction between the variables because such models could make predictions unstable. The third desirable property of a projection function is "path invariance". Path invariance concept explains that the resulting stand basal area from either two-step projection (first step: projecting basal area from time t0 to t1; second step: projecting the basal area from time t1 to t2) or one-step projection (directly projecting basal area from time t0 to t2) must be equal. The path invariance property ensures that the estimated projections do not suffer from compounded errors in the fitting process (Clutter et al., 1983; Gyawali and Burkhart, 2015). The algebraic difference approach (ADA) suggested by Bailey and Clutter (1974) can be used to adjust prediction equation into projection equation. This also helps achieve path invariance property in the stand basal area modeling.

While most of the aforementioned studies on stand basal area are based on pine plantations, very few studies are published for natural hardwoods in the US South (Mctague et al., 2008). Natural forests in the US South cover more than half of the total forests in the region (Oswalt et al., 2019). However, the research on natural forests including growth and yield modeling studies is limited due to several complexities. There are numerous permanent plots with monospecific plantations in the US South managed by many cooperatives, industries and private landowners. Information on the growth and yield of these plantations are well-recorded, thus the stand level model systems for these forests are well developed (Koirala et al., 2021). Unlike plantations, the stand structure of natural forests is complex and composed of mixed tree species. Predicting the growth and yield of hardwood stands is challenging because of inconsistent growth rates and different management requirements for different species. Brooks et al. (2008) developed a whole stand basal area projection model for Appalachian hardwoods using data from West Virginia. McTague et al. (2008) developed a whole stand model for mixed southern hardwoods including both upland and bottomland tree species. In this study, the authors utilized a simultaneous fitting approach for modeling both basal area prediction and projection equations.

The simultaneous fitting ensures compatibility between the parameters in both prediction and projection models, which is a desirable characteristic since projection equations are a mere modification of prediction equations. Several studies have shown simultaneous fitting with two-stage or three-stage least squares as a theoretically sound estimation procedure than separate least square fitting (Borders, 1989; McTague and Bailey, 1987; Scolforo et al., 2019). However, in simultaneous fitting, two or more equations share a common error structure (non-zero covariances), which may possess bias in basal area estimation. One way to deal with this problem is to develop a framework to evaluate uncertainty in the simultaneous system. This includes assigning two separate types of weights to prediction and projection equations. The main objective of this study is to develop a whole stand basal area prediction and projection equations for mixed southern hardwood forests incorporating uncertainty. In this study, we utilized the simultaneous fitting technique to fit our equations and compare (i) simultaneous fitting with common error structure and (ii) separate fitting of prediction and projection models. In the end, we presented separate basal area equations (both prediction and projection) for separate species groups that we derived from cluster analysis.

Methods

Data and Measurements

Data utilized in this research came from the Hardwood Research Cooperative (HRC) of the North Carolina State University (McTague et al., 2008). More than 600 plots were established across the US South in cooperation with member forest companies of the HRC. Since natural hardwood forests cover a wide range of geographical regions within the US South, the first task was to identify the forest site types. The cooperative identified nine forest site types which could be classified as separate forest operating units (Table 1). Among 600 plots, 187 plots were permanently maintained for repeated measurements and as a data source for growth and yield prediction analysis. The forest areas in the established plots were even-aged and fully stocked stands. Majority of the new growth natural forests in the US South is even-aged, thus the study ignored uneven-aged and high graded forest stands. There were five age groups in the data collection: ranging from 20 to 60 years at 10-year intervals. Table 1 presents the distribution of plots by age class and site types.

Field crew installed circular 1⁄5-acre (~ 0.081 hectare) plots, and measured total height, merchantable height and dbh of all trees greater than 14 cm dbh. In addition, crew also recorded tree species and age of individual trees. Several considerations were made to measure trees with some sort of defects. For trees with butt swell or other malformations at breast height, crew measured the diameter of the tree 0.3 m above the irregularity. For trees with forks below 1.37 m, crew recorded them as individual trees. The merchantable height of the tree with at least 2.45 m long bole was measured until the diameter of tree reaches 10 cm at the top. The basal area for each tree was calculated using standard basal area formula of the circular objects. The total basal area for a stand was calculated by adding individual basal area of sampled trees in a stand, which was then converted into per acre basis by multiplying the conversion factor.

Model Formulation

We started with selecting Schumacher type function as our base model for basal area prediction. Using similar base model, Pienaar and Shiver (1986) proposed the whole stand-level prediction and projection equations for both unthinned and thinned slash pine plantations. Since then, this model has been utilized by several researchers to develop stand-basal area models including for central and southern hardwoods by Brooks et al. (2008) and McTague et al. (2008), respectively. Following the work of Pienaar and Shiver (1986), the system of prediction and projection equations can be formulated with the following prediction model form:

$$BA = \exp\left(\beta_0 + \frac{\beta_1}{A}\right) HD^{\beta_2} N^{\beta_3} \qquad (4.1)$$

where, A is the age of the stand in years, BA the basal area in m^2 per hectare, HD the dominant height in m and N the stand density, whereas β_0, β_1, β_2, and β_3 are the parameters to be estimated. Equation (4.1) is a nonlinear form of equation that can be transformed into linear model using natural logarithm on both sides as:

$$ln(BA) = \beta_0 + \beta_1 \frac{1}{A} + \beta_2\, ln(HD) + \beta_3\, ln(N) \qquad (4.2)$$

In order to derive projection equation, we can assume two states of basal area prediction, previous state at time (*t-1*) (eq. 4.3) and future state at time *t* (eq. 4.4).

For previous state,

$$\ln(BA_{t-1}) = \beta_0 + \beta_1 \frac{1}{A_{t-1}} + \beta_2\, ln(HD_{t-1}) + \beta_3\, ln(N_{t-1}) \qquad (4.3)$$

For future state,

$$\ln(BA_t) = \beta_0 + \beta_1 \frac{1}{A_t} + \beta_2\, ln(HD_t) + \beta_3\, ln(N_t) \qquad (4.4)$$

With equations (4.3) & (4.4), it is just a matter of algebraic transformations to obtain the basal area projection equation. However, the form of final projection equation to be either anamorphic or polymorphic depends on which parameter we isolate for transformation. For example, if we isolate the asymptote parameter (β_0), the resulting projection equation will be anamorphic in form. In contrast, when we isolate other remaining parameters, the resulting projection equation will be polymorphic in form. We selected polymorphic projection model because it performs better than anamorphic model for long-term projection. We isolated β_1 parameter from both equations (4.3) & (4.4) and with algebraic manipulation formulated a basal area projection model as equation (4.5).

$$\ln(BA_t) = \frac{A_{t-1}}{A_t}\ln(BA_{t-1}) + \beta_0\left(1 - \frac{A_{t-1}}{A_t}\right) - \beta_2\left(ln(HD_t) - \frac{A_{t-1}}{A_t}\, ln(HD_{t-1})\right) -$$

$$\beta_3\left(\ln(N_t) - \frac{A_{t-1}}{A_t}\ln(N_{t-1})\right)$$

(4.5)

Equation (4.2) and equation (4.5) are the prediction and projection basal area models with common parameters such as β_0, β_2, and β_3. One desirable property of stand-level growth and yield system is compatibility among different models used for estimating different stand dynamics. Since basal area projection model depends on prediction model, fitting both models simultaneously can ensure compatibility between them. We set up a system of linear equations using equations (4.2) and (4.5), which was solved simultaneously using matrix algebra that uses values from each time step. In the following, we present an example of simultaneous fit for equation (4.2) and (4.5) using first plot Plot "5502002" of the dataset with 4-time steps (Scolforo et al., 2019). The corresponding design matrix is also presented.

$$Y = \beta_0 X_1 + \beta_1 X_2 + \beta_2 X_3 + \beta_3 X_4 \tag{4.6}$$

$$\begin{vmatrix} \boldsymbol{Y} \\ ln(BA_{11}) \\ ln(BA_{12}) \\ ln(BA_{13}) \\ ln(BA_{14}) \\ ln(BA_{11}) \\ ln(BA_{12}) - \frac{A_1}{A_2} ln(BA_{11}) \\ ln(BA_{13}) - \frac{A_2}{A_3} ln(BA_{12}) \\ ln(BA_{14}) - \frac{A_3}{A_4} ln(BA_{13}) \end{vmatrix} = \begin{vmatrix} \boldsymbol{X_1} & \boldsymbol{X_2} & \boldsymbol{X_3} & \boldsymbol{X_4} \\ 1 & \frac{1}{A_1} & ln(HD_{11}) & ln(N_{11}) \\ 1 & \frac{1}{A_2} & ln(HD_{12}) & ln(N_{12}) \\ 1 & \frac{1}{A_3} & ln(HD_{13}) & ln(N_{13}) \\ 1 & \frac{1}{A_4} & ln(HD_{14}) & ln(N_{14}) \\ 1 & 0 & ln(HD_{11}) & ln(N_{11}) \\ 1 - \frac{1}{2} & 0 & ln(HD_{12}) - \frac{A_1}{A_2} ln(HD_{11}) & ln(N_{12}) - \frac{A_{t-1}}{A_t} ln(N_{11}) \\ 1 - \frac{2}{3} & 0 & ln(HD_{13}) - \frac{A_2}{A_3} ln(HD_{12}) & ln(N_{13}) - \frac{A_{t-1}}{A_t} ln(N_{12}) \\ 1 - \frac{3}{4} & 0 & ln(HD_{14}) - \frac{A_3}{A_4} ln(HD_{13}) & ln(N_{14}) - \frac{A_{t-1}}{A_t} ln(N_{13}) \end{vmatrix} \begin{vmatrix} \beta_0 \\ \beta_1 \\ \beta_3 \\ \beta_4 \end{vmatrix}$$

where, indices *11*, *12*, *13*, and *14* refers to plot 1 and the times of the plot measurement.

Uncertainty Estimation

The aforementioned system is a classic linear least square approximations for simultaneous fitting of two equations using matrix solutions. One major caveat of this system is an assumption of error free observations. In other words, the system assumes that measurements collected at time *t-n* are error free, which might yield inconsistency in estimators. Thus, we applied a more complex co-variance matrix structure with generalized least squares (GLS). The expression for parameter estimation with generalized least squares is shown in equation (4.7).

$$\hat{\beta} = (X^T \widehat{W} X)^{-1} X^T \widehat{W} Y \tag{4.7}$$

where, $\hat{\beta}$ is the estimated parameters of the GLS system; X^T is the transpose of design matrix X of independent variables; $\widehat{W}$ is the weighted covariance matrix in the linear system; and Y is the design matrix of dependent variable. If no correlation exists among measurements, the $\widehat{W}$ matrix turns out to be $\sigma^2 I$, where I is the identity matrix and σ^2 is the regression variance. In that case, the system behaves as if linear least squares. However, in our case, we are dealing with two sources of error from two different models, prediction and projection. Let us assume that the errors arising from prediction and projection models are ε and $\breve{\varepsilon}$, respectively. The prediction error ε depends on the allometric relation between the independent variables and the dependent variables at the time the relation was evaluated. Likewise, it is also evident that on projection model, the error term $\breve{\varepsilon}$ depends on the variance between the time *t* and *t-1* and the difference between both measurement periods (Δt). Thus, our weight matrix $\widehat{W}$ in the GLS system considered both error structures as in equation (4.8).

$$w_{t\,;\,t-1} = \begin{cases} \frac{1}{\sigma_t^2} & BA = f(\hat{\beta}; t) \\ \\ \frac{1}{\sigma_t^2 + \sigma_{t-1}^2 \, \Delta t} & \breve{BA} = f(\hat{\beta}; t, t-1) \end{cases} \tag{4.8}$$

This technique will allow us to fit both prediction (BA) and projection ($\breve{BA}$) basal area models simultaneously while dealing with the uncertainty in each phase with the help of weighting factor $w_{t\,;\,t-1}$.

Model Fit and Validation

We simultaneously fitted prediction (equation 2) and projection (equation 5) models using two different error structures: (A) homoscedastic errors with a single variance parameter, and (B) assimilated error structure with weighting factor $w_{t\,;\,t-1}$, as presented in equation (8). In addition, we also fitted prediction and projection models independently to compare the differences in parameter estimates and errors between independent fit and simultaneous fit. For model validation, we utilized a robust five-fold cross-validation approach during data fitting process (Hastie *et al.*, 2009). This approach involves randomly dividing the set of observations into *K* groups or folds (5 groups in our case), of approximately equal size. The first group of the data is held out as a validation set and the remaining folds i.e. (*K-1*) are used to train the model. The process continues for five times as each sample is given the opportunity to be utilized as the hold out for one time while remaining *K-1* samples act as training set. At the end of a complete five-fold cross-validation run for each model, five sets of predicted values and respective goodness-of-fit statistics are generated. In order to test the model performance, an average of these five values are reported and compared (Figure 1). Comparison of these two systems of models were based on four major goodness-of-fit statistics: root mean square error (RMSE), Bias, Akaike's Information Criterion (AIC) and Bayesian Information Criterion (AIC). Lower values of RMSE, Bias, AIC, and BIC are preferred during model selection.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n}} \qquad (4.9)$$

$$Bias = \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)}{n} \qquad (4.10)$$

$$AIC = n\ln(SSE) - n\ln(n) + 2(k+1) \qquad (4.11)$$

$$BIC = n\,ln(SSE) - n\,ln(n) + (k+1)\,ln\,(n) \quad (4.12)$$

where, Y_i is the observed basal area of the plot; $\hat{Y}_i$ is the predicted basal area; $\bar{Y}_i$ is the mean value of Y_i; SSE is the sums of square of error; n is the total number of samples and k is the number of parameters in a model.

Results

We utilized two types of error structures to estimate basal area growth of the hardwood stand at current and future periods. First, the prediction (equation 2) and projection (equation 5) models were simultaneously fitted assuming homoscedastic error as in linear least squares. Later, a novel assimilated error with weighting factor was assigned to the simultaneous system that behaves similar to the generalized least squares with weights. Table 2 shows the parameter estimates and standard errors of different combination of models and modeling techniques. All the parameter estimates of the models were significant at the 5% level. The parameter estimates for independent fitting were similar for both techniques A and B. However, for simultaneous fit, the β_1 parameter was drastically different between technique A and B. The simultaneous fitting procedure resulted in lower standard error values of the parameter estimates as compared to the independent fit, which was due to the compatibility constraint assigned in the simultaneous fitting.

The results from the validation statistics showed that the weighting factor (technique B) was able to improve the accuracy of the basal area estimates (Table 3). The fitting technique B was superior to technique A in all validation statistics for both independent and simultaneous model formulations. The inclusion of weighting factor $w_{t\,;\,t-1}$ decreased the RMSE values from 8.69 m^2ha^{-1} to 6.72 m^2ha^{-1} for prediction model with independent fitting. It also decreased the bias value from 0.95 m^2ha^{-1} to 0.84 m^2ha^{-1} in prediction model indicating that expected basal area values are not far from true values. Similar to prediction model, the error and bias values decreased after assigning weighting factor to the projection model. As expected, the simultaneous models showed a slightly higher error and bias values in both techniques A and B. The AIC and BIC values generated from technique B were lower than technique A with homogenous error structure. Higher drop in AIC value was observed in technique B for independent prediction model, whereas a slight drop was noticed for independent projection model. The BIC value also showed similar trend between technique A and B. Overall, the weighting factor results in an increase accuracy and precision.

In the next step, we evaluated the performance of both techniques by projecting forward the 10 years old stand up to 70 years using three hypothetical initial stand basal area values of 20 ft^2ac^{-1} (4.6 m^2ha^{-1}), 30 ft^2ac^{-1} (6.9 m^2ha^{-1}), and 40 ft^2ac^{-1} (9.2 m^2ha^{-1}). We assumed that the stand had site index of 60 feet (18.29 m) at base age 50 years, and initial stand density of 500 trees per acre (1235.5 trees per hectare). The projection of the basal area over 60 years horizon showed that the projection from technique A (with dashed lines) was over-projecting the basal area (Figure 2). As previously mentioned, technique A assumes homoscedastic error in fitting process and has a single variance value. When we compare the projection basal areas between technique A and B, we found that the difference in projected basal area was more prominent when the age of the stand moves ahead of base age 50 years. The projected lines from technique B (with solid lines) accounted for two types of errors: prediction and projection. The weighting factor assigned in the process penalized the model for over-projecting.

We also examined the residual plots for both techniques using simultaneous projection values calculated from equation 5 (Figure 3). The comparison of residual plots showed clear advantage of the weighting procedure over the traditional least squares. Although not perfect, the residuals from technique B showed randomness and more constant variance than that of technique A.

Discussion

About two-third of the woodland areas in the southeast US is covered by natural hardwood forests (Oswalt et al., 2019). These forests are commonly found in many industrial and nonindustrial private landowners' properties across the region. For a management perspective, basal area is an important measure of competition dynamics among individual trees in a forest stand (Hasenauer et al., 1997; Pszwaro et al., 2016). Basal area estimation is essential for planning silvicultural regimes of the forest as it is directly related to some important stand variables such as quadratic mean diameter, volume and stand density. However, there are limited studies on basal area modeling of natural forests due to complexity in modeling mixed species stands and lack of permanent sample plots for repeated measurements. The main objective of this study was to develop basal area prediction and projection equations for even-aged mixed hardwood forests of the US South that can account for uncertainty in future projections.

This paper highlights the use of novel uncertainty estimation technique in modeling whole stand basal area of natural hardwood forests in the US South. We introduced a technique to assimilate variances of prediction and projection basal area models in a simultaneous fitting system that can reduce bias in the final estimation. The major benefit of this procedure is its ability to evaluate uncertainty for long-term projection period while accounting for errors induced during field measurements. Therefore, instead of testing and evaluating different types of basal area models found in literature, we rather focused on a simple, yet effective Schumacher type function proposed by Pienaar and Shiver (1986) to implement our uncertainty projection technique. In

Pienaar and Shiver (1986) paper, authors developed an anamorphic basal area projection model based on the prediction equation. In our study, we developed polymorphic model instead of anamorphic. Borders et al. (2004) has shown that the anamorphic formulation could lead to large absolute differences in basal area projection. Polymorphic models are well-suited for longer term projections because they consider the fact that in long-term projection, the effects of initial stand conditions (basal area, density, dominant height) gradually decline and eventually, the basal area projected from the projection equation becomes similar to that obtained from prediction equation (Mctague et al., 2008; Von Gadow and Hui, 1999).

When prediction and projection models were independently fitted, both techniques A and B yielded similar parameter estimates (Table 2). However, for simultaneous fitting, technique B gave noticeably different estimates for parameter β_1. The effect of weighting factor was seen in β_1 parameter as it was isolated during polymorphic model formulation. The effect of simultaneous fitting was also apparent in model validation statistics, as the RMSE and bias values were higher than independent fitting procedure (Table 3). However, the assurance of compatibility from simultaneous fitting could weigh higher than slight increase in error and bias in the estimates. Previous studies agreed on this trend of reduction in fit precision due to the application of simultaneous system, as this procedure tends to have higher number of parameters resulting from more than one equation (LeMay, 1990). The projection lines plot (Figure 2) showed high difference in projected basal area values between two techniques. This was because of the overfitting resulted from assumption of homoscedastic error structure in technique A. Our new assimilated error structure was able to compensate this overfitting by incorporating two sources of errors. Thus, it reduces the effect of the projection values as the elapsed time between two consecutive measurements increases, giving more weight to the observation instead of the projection. The effect of inclusion of weighting factor was also observed in the residual plots analysis as technique B provide more randomness and constant variance of the residuals. The basic assumption of the least square regression (technique A in our case) is that the independent variables are free of errors. In other words, no measurement error persists in the data; and the errors are only related to the model or the process. Field measurements in forestry is always considered a difficult task, which is full of error and approximations. Thus, the application of assimilated error structure in this study has improved the projection of stand basal area in future periods.

Conclusions

In this study, we developed a whole stand basal area prediction and projection equations for mixed southern hardwood forests incorporating uncertainty. In order to achieve our objective, we formulated a simultaneous system of equations to ensure compatibility between prediction and projection models. We tested two types of error structures (A) a homoscedastic error structure that is similar to least squares problem, and (B) a weighting factor calculated as an assimilated error structure that accounts for the uncertainty during basal area projection. The results from our cross-validation analysis showed that technique B outperformed technique A in terms of accuracy and precision. In addition, technique B was able to solve the overfitting problem experienced by technique A while projecting basal area over a large span of time. Based on the information available, researchers can apply the parameter estimate values from technique B to project natural hardwood stand basal area in the US South. The main advantage of utilizing weighting factor (technique B) over the traditional approach (technique A) is its ability to account for uncertainty arising from prediction as well as projection stages of basal area estimation.

Tables and Figures

Table 4.1. Number of hardwood growth and yield plots by site type and age class across the southern United States. The stand ages in the plots were determined using past records and increment cores taken from dominant and codominant trees.

Site type	**Age classes (year)**					
	20	**30**	**40**	**50**	**60**	**Total**
Muck swamp	88	11	8	99	13	49
Peat swamp	1	5	2	11	2	11
Wet flat	7	35	19	12	3	76
Red river bottom	8	18	15	11	6	58
Black river bottom	5	14	24	2	7	52
Branch bottom	12	33	24	18	5	92
Bottomland	36	59	40	23	7	165
Coves, gulfs, lower slopes	5	8	17	7	4	41
Upper slopes and ridges	5	27	34	18	13	97
Total	87	210	183	101	60	641

Table 4.2. Parameter estimation and standard errors from the independent and simultaneous fitting of the prediction and projection models using technique A (Homoscedastic error) and technique B (Weighting factor, $w_{t\,;\,t-1}$).

Fit technique	**Estimated parameters (Standard errors)**				
	Function	β_0	β_1	β_2	β_3
A Homoscedastic error	Pred.	- 4.808 (0.325)	- 13.637 (1.744)	1.302 (0.079)	0.676 (0.028)
	Proj.	- 4.151 (0.483)		0.613 (0.096)	0.846 (0.032)
	Sim.	- 5.161 (0.287)	- 3.710 (1.430)	1.134 (0.070)	0.777 (0.025)
B Weighting factor $w_{t\,;\,t-1}$	Pred.	- 4.874 (0.318)	- 13.136 (1.705)	1.323 (0.078)	0.674 (0.028)
	Proj.	- 4.105 (0.494)		0.610 (0.097)	0.842 (0.032)
	Sim.	- 5.181 (0.306)	- 0.236 (1.401)	1.046 (0.075)	0.810 (0.025)

Pred. = prediction model; Proj. = Projection model; Sim. = simultaneous fitting

Table 4.3. Cross-validation statistics obtained from the independent and simultaneous fitting of the prediction and projection models using technique A (Homoscedastic error) and technique B (Weighting factor, $w_{t\,;\,t-1}$).

Fit technique	**Cross-validation statistics for independent and simultaneous fitting**								
	Function	**RMSE**		**Bias**		**AIC**		**BIC**	
		Ind.	Sim.	Ind.	Sim.	Ind.	Sim.	Ind.	Sim.
A	Prediction	8.69	9.41	0.95	2.15	1448.52	1498.12	1463.47	1513.06
	Projection	3.47	4.81	1.39	2.11	780.53	982.28	795.44	997.23
B	Prediction	**6.72**	**7.93**	**0.84**	**1.64**	**1351.01**	**1302.38**	**1365.96**	**1327.30**
	Projection	**2.98**	**3.23**	**1.21**	**1.99**	**774.64**	**935.26**	**790.59**	**866.43**

Ind. = Independent or separate fitting; Sim. = simultaneous fitting

Figure 4.1. Schematic representation of five-fold cross-validation approach. Each horizontal bar is divided into five sections, out of which four are training folds and one section is the validation fold. There are altogether five horizontal bars representing five iterations or folds. Goodness-of-fit statistics (GOF) are calculated for each iteration and the final GOF statistics for a model is calculated as the average of five GOF statistics from five iterations.

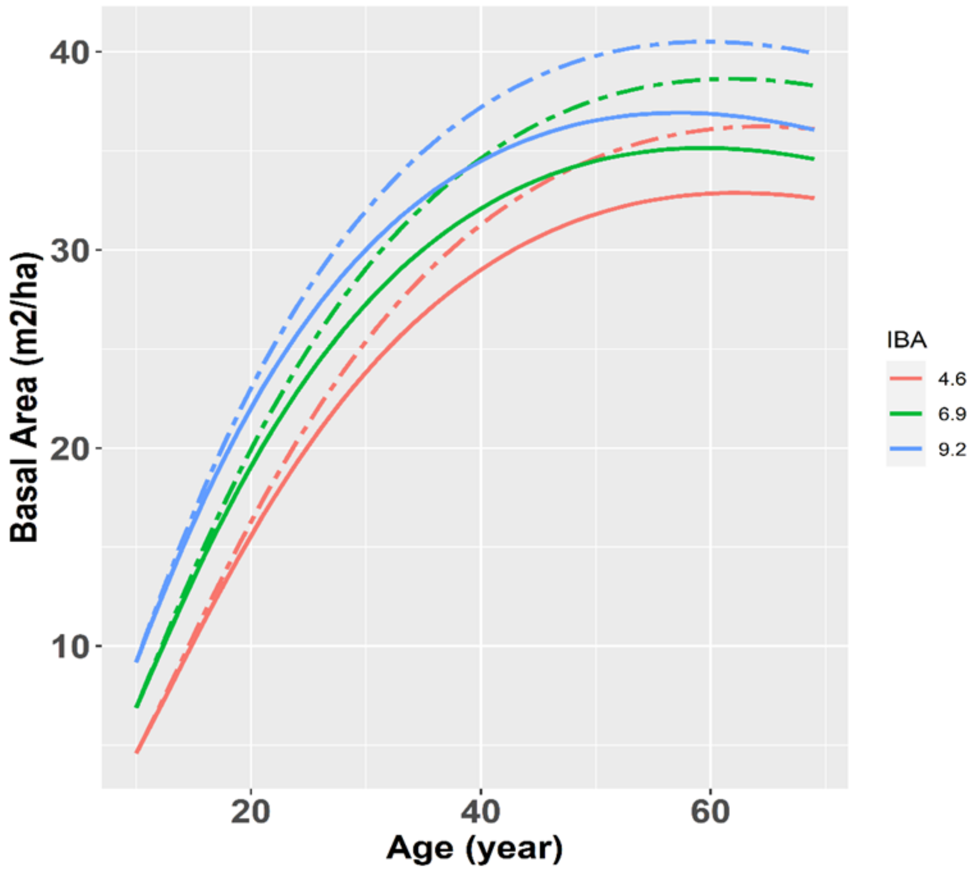

Figure 4.2. Basal area projection of hardwood stands with three initial basal areas of 4.6 (red), 6.9 (green), and 9.2 m^2ha^{-1} (blue). The dashed lines represent basal area projection using technique A, while the solid lines indicate basal area projection from technique B. All three curves had a site index of 60 feet (18.29 m) at base age 50 years, and initial stand density of 500 trees per acre (1235.5 trees per hectare).

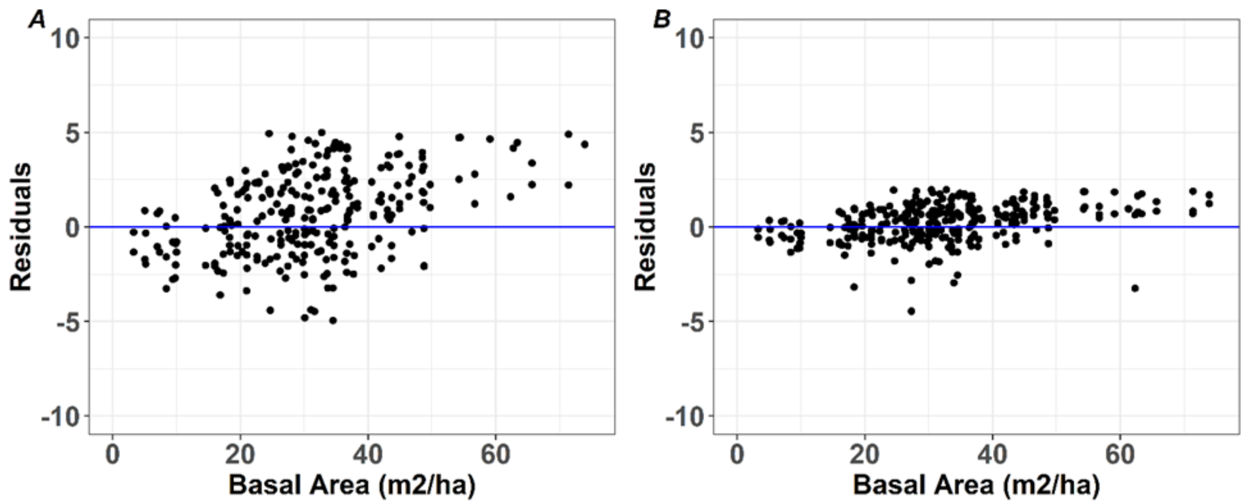

Figure 4.3. Residual plots for two model fitting techniques (A and B).

References

Anton-Fernandez, C., Burkhart, H.E., Amateis, R.L., 2012. Modeling the Effects of Initial Spacing on Stand Basal Area Development of Loblolly Pine. For. Sci. 58, 95–105. https://doi.org/10.5849/FORSCI.10-074

Bailey, R.L., Clutter, J.L., 1974. Base-age invariant polymorphic site curves. For. Sci. 20, 155–159. https://doi.org/10.1093/forestscience/20.2.155

Barrio Anta, M., Castedo Dorado, F., Diéguez-Aranda, U., Álvarez González, J.G., Parresol, B.R., Rodríguez Soalleiro, R., 2006. Development of a basal area growth system for maritime pine in northwestern Spain using the generalized algebraic difference approach. Can. J. For. Res. 36, 1461–1474. https://doi.org/10.1139/X06-028

Borders, B., Harrison, W.M., Zhang, Y., Shiver, B.D., Clutter, M., Cieszewski, C.J., Daniels, R.F., 2004. Growth and yield models for second rotation loblolly pine plantations in the Piedmont/Upper Coastal Plain and Lower Coastal Plain of the southeastern U.S, University of Georgia, PMRC Technical Report 2004-4, 63p.

Borders, B.E., 1989. Systems of Equations in Forest Stand Modeling. For. Sci. 35, 548–556. https://doi.org/10.1093/forestscience/35.2.548

Brooks, J.R., Jiang, L., Perkowski, M., Sharma, B., 2008. A Whole Stand Basal Area Projection Model for Appalachian Hardwoods. IIn Jacobs, Douglass F.; Michler, Charles H., eds. 2008. Proceedings, 16th Cent. Hardwood For. Conf. 2008 April 8-9; West Lafayette, IN. Gen. Tech. Rep. NRS-P-24. Newt. Square, PA US Dep. Agric. For. Serv. North. Res 324–330.

Cao, Q. V., 2021. A unified system for tree- and stand-level predictions. For. Ecol. Manage. 481, 118713. https://doi.org/10.1016/j.foreco.2020.118713

Castedo-Dorado, F., Diéguez-Aranda, U., Barrio-Anta, M., Álvarez-Gonzàlez, J.G., 2007. Modelling stand basal area growth for radiata pine plantations in Northwestern Spain using the GADA. Ann. For. Sci. 64, 609–619. https://doi.org/10.1051/forest:2007039

Castedo Dorado, F., Diéguez-Aranda, U., Barrio Anta, M., Sánchez Rodríguez, M., von Gadow, K., 2006. A generalized height-diameter model including random components for radiata pine plantations in northwestern Spain. For. Ecol. Manage. 229, 202–213. https://doi.org/10.1016/j.foreco.2006.04.028

Chikumbo, O., Mareels, I.M.Y., Turner, B.J., 1999. Predicting stand basal area in thined stands using a dynamical model. For. Ecol. Manage. 116, 175–187. https://doi.org/10.1016/S0378-1127(98)00449-6

Clutter, J.L., 1963. Compatible Growth and Yield Models for Loblolly Pine. For. Sci. 9, 354–371. https://doi.org/10.1093/forestscience/9.3.354

Clutter, J.L., Fortson, J.C., Pienaar, L. V., Brister, G.H., Bailey, R.L., 1983. Timber Management: A Quantitaive Approach. John Wiley & Sons Inc, New York , NY.

Clutter, J.L., Jones, E.P., 1980. Prediction of growth after thinning in old-field slash pine plantations. Department of Agriculture, Forest Service, Southeastern Forest Experiment ….

Gyawali, N., Burkhart, H.E., 2015. General response functions to silvicultural treatments in loblolly pine plantations. Can. J. For. Res. 45, 252–265. https://doi.org/10.1139/cjfr-2014-0172

Hasenauer, H., Burkhart, H.E., Amateis, R.L., 1997. Basal area development in thinned and unthinned loblolly pine plantations. Can. J. For. Res. 27, 265–271. https://doi.org/10.1139/x96-163

Hastie, T., Tibshirani, R., Friedman, J., 2009. The elements of statistical learning: data mining, inference, and prediction, Second. ed. Springer New York.

Koirala, A., Montes, C.R., Bullock, B.P., 2021. Modeling dominant height using stand and water balance variables for loblolly pine in the Western Gulf, US. For. Ecol. Manage. 479, 118610. https://doi.org/10.1016/j.foreco.2020.118610

Lam, T.Y., Guan, B.T., 2020. Modeling stand basal area growth of Cryptomeria japonica D. Don under different planting densities in Taiwan. J. For. Res. 25, 174–182. https://doi.org/10.1080/13416979.2020.1733171

LeMay, V.M., 1990. MSLS: a linear least squares technique for fitting a simultaneous system of equations with a generalized error structure. Can. J. For. Res. 20, 1830–1839. https://doi.org/10.1139/x90-246

McTague, J.P., Bailey, R.L., 1987. Simultaneous total and merchantable volume equations and a compatible taper function for loblolly pine. Can. J. For. Res. 17, 87–92. https://doi.org/10.1139/x87-015

Mctague, J.P., Loughlin, D.O., Roise, J.P., Robison, D.J., Kellison, R.C., 2008. The SOHARC Model System for Growth and Yield of Southern Hardwoods. South. J. Appl. For. 32, 173–183. https://doi.org/https://doi.org/10.1093/sjaf/32.4.173

Oswalt, S.N., Smith, W.B., Miles, P.D., Pugh, S.A., 2019. Forest resources of the United States, 2017, US Forest Service, Gen. Tech. Report WO-97, 223 p.

https://doi.org/10.5962/bhl.title.101492

Pienaar, L. V., Shiver, B.D., 1986. Basal Area Prediction and Projection Equations for Pine Plantations. For. Sci. 32, 626–633. https://doi.org/10.1093/FORESTSCIENCE/32.3.626

Pszwaro, J.L., D'Amato, A.W., Burk, T.E., Russell, M.B., Palik, B.J., Strong, T.F., 2016. Analysis of stand basal area development of thinned and unthinned Acer rubrum forests in the upper Great Lakes region, USA. Can. J. For. Res. 46, 645–655. https://doi.org/10.1139/cjfr-2015-0411

Rodríguez, F., Pemán, J., Aunós, Á., 2010. A reduced growth model based on stand basal area. A case for hybrid poplar plantations in northeast Spain. For. Ecol. Manage. 259, 2093–2102. https://doi.org/10.1016/j.foreco.2010.02.021

Scolforo, H.F., McTague, J.P., Burkhart, H., Roise, J., McCarter, J., Alvares, C.A., Stape, J.L., 2019. Stand-level growth and yield model system for clonal eucalypt plantations in Brazil that accounts for water availability. For. Ecol. Manage. 448, 22–33. https://doi.org/https://doi.org/10.1016/j.foreco.2019.06.006

Sun, H.G., Zhang, J.G., Duan, A.G., He, C.Y., 2007. A review of stand basal area growth models. For. Stud. China 9, 85–94. https://doi.org/10.1007/s11632-007-0014-2

Von Gadow, K., Hui, G., 1999. Modelling forest development. Springer Science & Business Media.

Zhao, L.F., Li, C.M., 2013. Stand basal area model for Cunninghamia lanceolata (Lamb.) Hook. plantations based on a multilevel nonlinear mixed-effect model across south-eastern China. South. For. 75, 41–50. https://doi.org/10.2989/20702620.2013.769750

CHAPTER 5

OVERALL CONCLUSIONS

Whether it is Appalachian region with forests dominated by upland oak – hickory or Mississippi flood plain with species like cottonwood, sweetgum or black willow, both regions are important source of valuable hardwood timber in the US. The current condition of these unique and productive ecosystems is largely dictated by past stand treatments. Stands witnessing favorable stand treatments in the past are still well stocked with high-value trees. But stands that witnessed unfavorable treatments such as high-grading, overgrazing or land conversion are in bad condition with very few high-value trees. As current silvicultural practices are slowly shifting towards intensive management, the available quantitative tools should be upgraded to become more reliable and sophisticated. There is a critical need to develop growth and yield models that could describe the nature of southern hardwood as accurately as possible. Improvement of existing site productivity (site index) models and development of other stand models are also necessary for efficient management of southern hardwoods.

The first chapter of this dissertation presents the overall trend of scientific studies in growth and yield model development for southern hardwoods. The through literature search reveals that the latest scientific article solely focusing on developing stand-level GYM for southern hardwood was published in 2010 for bottomland hardwoods (Schultz et al., 2010). Much of the work in southern hardwoods GYM was carried out in 1970s and 1990s. The literature review points out some major research gaps in the field such as lack of permanent sample plots for continuous forest measurements. The remeasurement data are the foundation of stand-level models; therefore, lack of permanent plots in current time is hindering the capability of quantitative forest scientists to conduct new research.

In the second chapter of this dissertation, we develop a new dominant height and site index models for southern hardwoods. Since the concept of site index are species-specific, we attempt to group hardwood species with similar properties and stand characteristics. Through cluster analysis we propose five species groups for southern hardwoods based on the available data. To model future dominant height trajectory, we implement an advanced nonlinear filtering technique, extended Kalman Filter (EKF) that has been successfully tested with loblolly pine data before (Montes et al., 2018). Our results indicates that site index models developed with this approach is able to reduce uncertainty in dominant height projection for mixed hardwood forests in the US South. The EKF methodology can be a sound technique to calculate dominant height when uncertainty estimation is a goal. In addition, the ability of EKF to assimilate new information in the system without requiring a comprehensive forest inventory could be crucial in reducing sampling costs. Our analysis also provides parameter values for each hardwood species groups for the region, which could fill the gap of knowledge and confusion in hardwood site index calculation.

We develop a stand level mortality model for southern hardwood in third chapter. We construct models with differential equation approach, with the underlying assumption that the relative rate of instantaneous mortality is related to stand age, site index, and stand densities. We introduce a localized parameter in the model to account for stand-specific initial density. Our evaluation shows that model with initial stand density and age as predictor variables performed best in terms of model evaluation statistics and performance of the residuals. Since this model accounts for stand-specific initial density as a parameter, inclusion of site index in the model does not seem necessary. Furthermore, models including site index do not improve the prediction accuracy in future tree number prediction. This approach could be a valuable tool for predicting stand density of natural forests where number of trees during stand formation are unknown.

The fourth chapter focuses on modeling stand basal area of southern hardwoods. We formulate a simultaneous system of equations to ensure compatibility between prediction and projection models. This chapter highlights the use of novel uncertainty estimation technique in modeling whole stand basal area. We introduce a technique to assimilate variances of prediction and projection basal area models in a simultaneous fitting system that can reduce bias in the final estimation. The major benefit of this procedure is its ability to evaluate uncertainty for long-term projection period while accounting for errors induced during field measurements. The results from our cross-validation analysis shows that our approach outperformed traditional technique in terms of accuracy and precision. Based on the information available, researchers can apply the parameter estimate values to project natural hardwood stand basal area in the US South. The main advantage of utilizing weighting factor over the traditional approach is its ability to account for uncertainty arising from prediction as well as projection stages of basal area estimation.

The Appendix section includes methodological description and results of whole stand volume modeling for southern hardwoods. The individual model components such as site index, stand density and basal area, derived from previous chapters can be utilized to calculate the whole stand volumes for southern hardwoods.

APPENDIX: VOLUME EQUATIONS

Total Volume and Green Weight Equations

Methods

The tree level dataset in this study contains dbh and total height measurements of all trees greater than 5.5-inch dbh in the 1/5 acre plots. In order to calculate individual tree total volume and green weight, we utilized Clark et al. (1985, 1986a, 1986b) equations. The individual tree green weight equation includes combined prediction of green weight of wood, bark and foliage of a tree. These studies developed volume and green weight equations for hardwood tree species growing in three different regions of the southern US, Gulf and Atlantic Coastal Plain, Piedmont and Upland-South. The volume and green weight equations in these studies use dbh and total tree height as independent variables. We converted individual tree level information to the stand level volume and green weight by summarizing the plot level information based on the 1/5 acre plot size.

We tested two stand level yield models that utilized both volume and green weight per hectare yield as the response variable for parameter estimation for volume and green weight predictions. The first model (Eq. 1) is a variable density Schumacher-type yield function that incorporates dominant height, stand age, basal area and density as independent variables (Gallagher et al., 2019). The second model (Eq. 2) is also a Schumacher-type yield function which incorporates exponential multiplicative term (Allen II et al., 2020).

$$Y = \beta_0 \, HD^{\left(\beta_1 + \frac{\beta_2}{Age}\right)} \, BA^{\left(\beta_3 + \frac{\beta_4}{Age}\right)} \, TPH^{\left(\beta_5 + \frac{\beta_6}{Age}\right)} + \varepsilon \qquad (1)$$

$$Y = \beta_0 \, HD^{\beta_1} \, BA^{\beta_2} \, e^{\left(\frac{\beta_3}{Age}\right)} + \varepsilon \qquad (2)$$

where, Y = total volume of the stand (cubic-meter/ha) or green weight of the stand (metric ton/ha), β_i = parameters to be estimated; HD = dominant height (meter); BA = basal area (square-meter/ha); TPH = trees per hectare and Age = stand age (years).

Before parameter estimation, we conducted Ljung-Box test (Ljung and Box, 1978) using **`Box.test`** function of **`stats`** package in R software to check autocorrelation in the dataset. The test revealed no significant autocorrelation between the measurements of same independent variables across different time period. We utilized a maximum likelihood (ML) approach to solve for parameters of two yield models. The objective function to minimize in ML approach assumes a normal distribution for the expected values (Eq. 3). Using **`optimr`** package in R version 4.1.0, we obtained the parameter estimates of the model. In order to calculate standard error of each parameter, we took square root of the diagonal of the inverse Hessian matrix, also known as the delta method.

$$argmin\, f(x,\beta) \sim \sum_{i=1}^{n} log\left(\frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{(f(x,\beta) - Y_i)^2}{2\sigma^2}\right)\right) \qquad (3)$$

where, function $f(x,\beta)$ represents our yield models with independent variables x and parameters β; Y_i represents observed yield values; and σ^2 represents the variance. In order to control the heteroscedasticity of the model, we modeled the residual variance as a power function of the fitted values (Eq. 4).

$$Var(\varepsilon_t) = \sigma_t^2 = (r^2)\, \widehat{Y_t^{\gamma}} \qquad (4)$$

where, r = residual standard error; $\hat{Y}$= predicted values from Eqs. 1 and 2; t is time and γ is variance parameter to be estimated.

We compared model fits using three model evaluation measures: root mean square error (RMSE), mean absolute error (MAE) and Akaike information criterion (AIC).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n}}$$

$$MAE = \frac{\sum_{i=1}^{n} |Y_i - \hat{Y}_i|}{n}$$

$$AIC = -2loglik + 2k$$

where, k is the number of parameters in the model, Y_i is the measured yield of the stand, and $\hat{Y}_i$ is the predicted yield.

Results: Final Model Selection

Table 1 depicts the parameter estimates, standard errors and p-values for the fitted models. For Equation 1, we can notice that the parameters representing stand density (β_5 and β_6) were not significant in both volume and green weight predictions at significance level of 0.05. In the second model (Equation 2), all parameters were significant for both volume and green weight.

Table 1. Parameter estimates and standard errors values for volume and green weight models.

Equation	Parameter	Estimate		Std Error		P-Value	
		Volume	GWT	Volume	GWT	Volume	GWT
Equation 1	β_0	1.4932	0.0001	0.3479	0.0001	<.0001	<.0001
	β_0	0.4341	0.4561	0.0824	0.0817	<.0001	<.0001
	β_0	10.7461	10.3499	2.4211	2.4190	<.0001	<.0001
	β_0	1.1561	1.2071	0.0282	0.0282	<.0001	<.0001
	β_0	-8.7729	-8.7476	0.9623	0.9720	<.0001	<.0001
	β_0	-0.0057	-0.0661	0.0377	0.0372	0.881	<.0001
	β_0	-0.9651	-0.7683	1.3548	1.3460	0.477	0.5683
Equation 2	β_0	0.8936	0.0001	0.1514	0.0001	<.0001	<.0001
	β_0	0.7942	0.8650	0.0484	0.0483	<.0001	<.0001
	β_0	0.9417	0.9529	0.0146	0.0144	<.0001	<.0001
	β_0	0.2858	-0.3781	0.9624	0.9739	<.0001	<.0001

The results from model fitting showed that the first model (Equation 1) outperformed the second model (Equation 2) for predicting whole stand volume. However, our analysis revealed exact opposite result for green weight prediction i.e., equation 2 outperformed equation 1. Table 2 shows the model evaluation statistics for whole stand volume and green weight models. We can see that equation 1 had lower error, bias and AIC values. However, one could notice that the difference in error and bias between two models are not that large. This can be attributed to the fact that both models are Schumacher-type functions, and both have utilized similar independent variables. In case of green weight predictions, equation 2 had lower values for all model validation statistics.

Table 2. Model validation statistics (RMSE, MAE and AIC) for volume and green weight models.

Equation	Volume			GWT		
	RMSE	**MAE**	**AIC**	**RMSE**	**MAE**	**AIC**
Model M1	**37.46**	**23.33**	**3272.65**	4.25	1.98	2174
Model M2	39.97	25.92	3349.41	**3.57**	**0.92**	**2110**

Values in bold font suggest the lowest validation statistics values.

We did not notice obvious bias while comparing observed vs predicted values for each component (volume and green weight) in both equations (Figure 1). Since stand density was not significant in predicting stand volume, we refitted the volume model excluding stand density from Equation 1. Based on the final evaluation, we propose Equation 5 for stand volume prediction and Equation 6 for stand green weight prediction for southern hardwoods.

$$Volume = 1.2137 * HD^{\left(0.5020+\frac{9.2284}{Age}\right)} * BA^{\left(1.1439+\frac{-9.2433}{Age}\right)} \quad (5)$$

$$Green\ Weight = 0.1382 * HD^{0.9293} * BA^{0.9694}\ e^{\left(\frac{-1.3532}{Age}\right)} \quad (6)$$

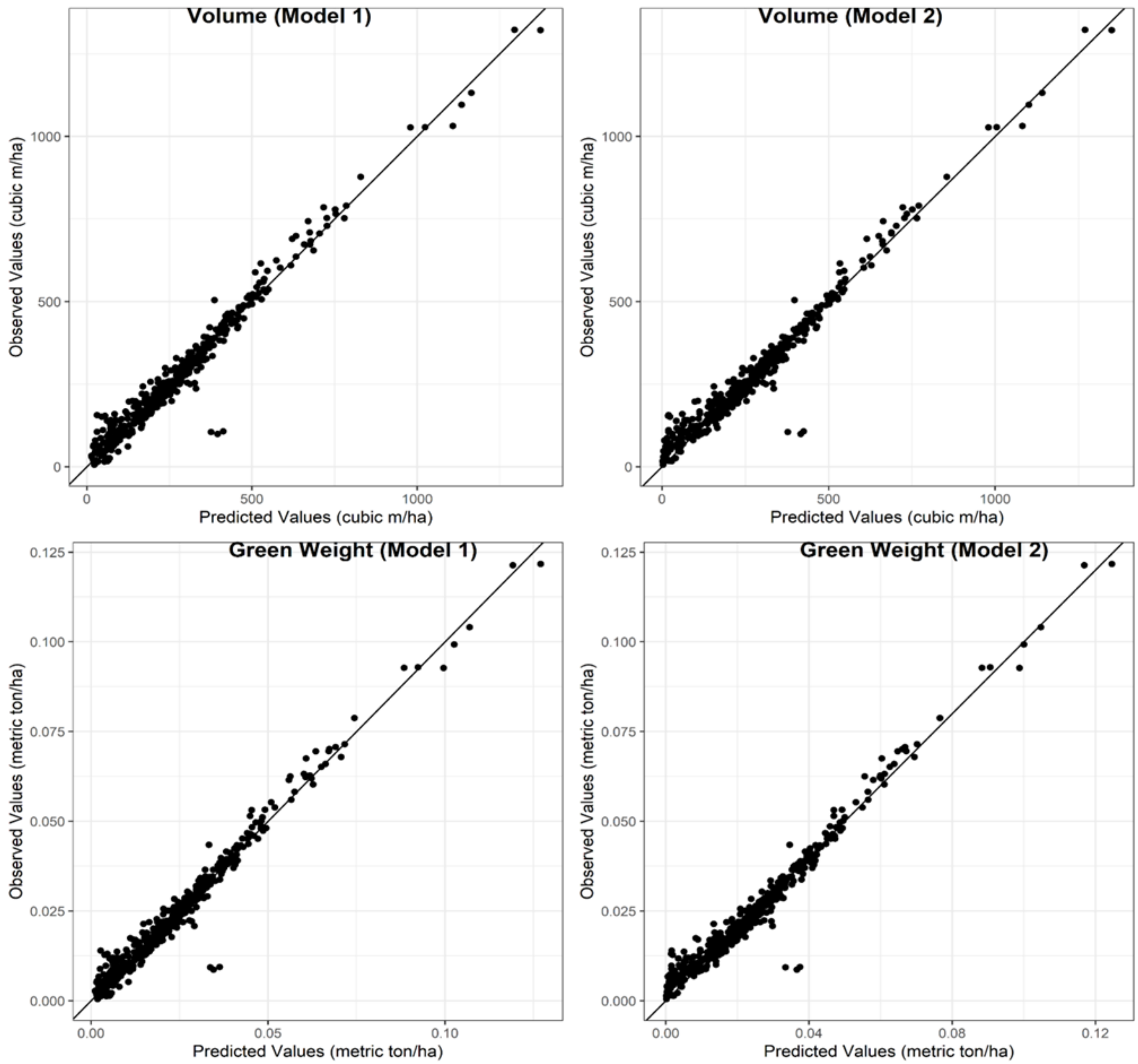

Figure 1. Predicted vs observed plots for total volume and green weight models.