

MIXED-EFFECTS MODELING OF SHORTLEAF
PINE (*PINUS ECHINATA* MILL.)
GROWTH DATA

By

CHAKRA BAHADUR BUDHATHOKI

Bachelor of Science
Tribhuvan University, Nepal
1992

Diploma in Statistics
The University of Reading, England
1996

Master of Science in Biometry
The University of Reading, England
1997

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
July, 2006

MIXED-EFFECTS MODELING OF SHORTLEAF
PINE (*PINUS ECHINATA* MILL.)
GROWTH DATA

Dissertation Approved:

Dr. Thomas B. Lynch

Thesis Advisor

Dr. David K. Lewis

Dr. Robert F. Wittwer

Dr. Mark E. Payton

Dr. A. Gordon Emslie

Dean of the Graduate College

PREFACE

The objectives of this dissertation are to review mixed-effects models in forestry literature, and to use these analysis tools for modeling shortleaf pine (*Pinus echinata* Mill.) growth variables. Data were available from over 200 permanent plots established in naturally-regenerated shortleaf pine stands of eastern Oklahoma and western Arkansas. It is evident from the review of literature that the mixed-effects models have been increasingly used in forestry since the 1990's. However, these tools have never previously been used in shortleaf pine growth and yield modeling.

This dissertation project was successful in expanding two of the major components of growth and yield modeling for shortleaf pine. The distance-independent individual-tree model for annual basal area growth model of Lynch et al. (1999) was improved to incorporate random-effects for plots in a potential-modifier framework with stand level and tree level explanatory variables. Furthermore, the individual-tree model for total height of Lynch et al. (1999) was also expanded to have plot-specific random parameters for the relationship between total height and diameter at breast height in which dominant height is also a predictor. The fitted mixed-effects models for annual basal area growth and total height were found to fit the data, and to predict the responses better than the previous models reported by Lynch et al. (1999). Possible correlated and/or heterogeneous within-plot errors were also investigated. It was found that within-

plot errors did not appear to be significantly correlated in the presence of plot random-effects; however, there was some evidence of heterogeneous errors. These model estimates can be utilized in the Shortleaf Pine Stand Simulator developed at Oklahoma State University after fitting other components such as mortality functions in mixed modeling framework. Due to voluminous data, complexity of the models attempted and some degree of correlations among model parameters, computing limitations were experienced at times, resulting in model convergence issues.

ACKNOWLEDGMENTS

First and the foremost, I am grateful to my major advisor Dr. Tom Lynch for his guidance, encouragement and cooperation throughout my graduate study, both coursework and research at Oklahoma State University. Thank you Dr. Lynch!

Thanks are also due to Drs. Mark Payton, David Lewis and Robert Wittwer for helping and encouraging me as advisory committee members. I also would like to thank the Department of Forestry at OSU and its faculty and staff for providing all the facilities, including a Graduate Research Assistantship and Dr. Michel Afanasiev Graduate Fellowship that made my study possible. The assistance of the USDA Forest Service Southern Research Station and the Ozark and Ouachita National Forests in data collection and financial support for the study is much appreciated. I also appreciate the friendship of all graduate students in the Department of Forestry and Environmental Science graduate program at OSU.

I cannot say enough in words how much love and help I have received from my wife Reena and son Saubhagya. Thanks for understanding me as a husband and as a father. I am thankful to my parents and brother Milan for their understanding of my study pursuit for such a long period of about four years. Our stay in Stillwater would not have been so comfortable, had there not been a small but coherent community of Nepalese friends and families. Thank you all for your time and help that we received.

TABLE OF CONTENTS

Chapter	Page
I. INTRODUCTION.....	1
Forest Growth and Yield Modeling.....	3
Objectives.....	6
II. REVIEW OF LITERATURE.....	7
Background.....	7
Growth Data.....	7
Modeling Biological Growth.....	8
Types of Forest Growth Models.....	13
Density-Free Stand-Level Models.....	13
Variable-Density Stand-Level Models.....	13
Diameter Class Models.....	17
Individual-Tree Models.....	18
Distance-Dependent Individual-Tree Model.....	19
Distance-Independent Individual-Tree Model.....	21
Uneven-Aged Stand Modeling.....	30
Forest Growth and Yield Simulation Systems.....	32
Mixed Models.....	34
Fixed vs. Random-Effects.....	34
Longitudinal Data/Repeated Measures.....	35
Correlated and/or Heterogeneous Errors.....	38
Application of Mixed Models in Forestry.....	39
Estimation/Computational Methods.....	49
III. DATA AND DATA MANAGEMENT.....	50
Study Area and Establishment of Plots.....	50
Measurements/Observations.....	54
Repeated Measurements.....	55
Data Management.....	57

IV. METHODS FOR DATA ANALYSIS AND MODELING	64
Mixed-Effects Models	65
Linear Mixed Model	67
Nonlinear Mixed Model.....	68
Extended Mixed Models.....	70
Estimation and Computing	71
Model Comparison Criteria	74
Error Modeling.....	75
Correlated Errors.....	75
Spatial Correlation Structure for Within-Group Errors	76
Heterogeneous Errors.....	78
Model Checking.....	80
An Approach to Mixed Model Development	80
Exploratory Data Analysis.....	82
Fitting the Growth Models.....	83
Basal Area Growth Model	83
Total Height Model.....	86
Fitting the Extended Models.....	87
Calibration and Validation.....	90
Summary Statistics for Calibration Data Set	90
Summary Statistics for Validation Data Set	92
V. RESULTS	94
Basal Area Growth Model	94
Calibration Results.....	94
Extended Models with Period and Plot Random-Effects	98
Extended Models with Power Variance Function	99
Extended Models with Spatial Correlation Function for Within-Plot Errors	104
Extended Models with Variance and Spatial Correlation Functions	105
Validation Results.....	105
Residuals vs. Predicted Values	106
Residuals vs. Tree Basal Area	109
Residuals by Design Criteria	111
Total Height Model.....	119
Calibration Results.....	119
Validation Results.....	120
VI. DISCUSSION AND CONCLUSIONS	125
Basal Area Growth Model	125
Extended Model with Period and Plot Random-Effects	128

Extended Model Accounting for Variance Heterogeneity.....	129
Extended Model with Random-Effects for Growth Period and Plots Accounting for Variance Heterogeneity	130
Residual Analysis for Basal Area Growth Models.....	131
Total Height Model.....	136
Residual Analysis for Height Model II.....	137
Discussion.....	139
Conclusions.....	145
 LITERATURE CITED.....	 147
 APPENDIX I	 163
 APPENDIX II.....	 167

LIST OF TABLES

Table	Page
2.1 Some selected individual tree growth and yield simulation models/systems available for the United States (Adapted from Davis et al. 2001)	33
3.1 Study design for naturally regenerated even-aged shortleaf plots installed from 1985 to 1987 with thinning and herbicide treatment in eastern Oklahoma and western Arkansas (Lynch et al. 1999)	52
3.2 Summary statistics for 208 plots (Studies 48 and 58 combined) for plot/stand level variables	61
3.3 Summary statistics for 208 plots (Studies 48 and 58 combined) for individual tree variables	62
4.1 List of fitted basal area growth models	89
4.2 Summary statistics of plot level variables for calibration data set with 139 plots	91
4.3 Summary statistics of tree level variables for calibration data set with 139 plots	91
4.4 Summary statistics of plot level variables for validation data set with 69 plots	92
4.5 Summary statistics of tree level variables for validation data set with 69 plots	93
5.1 Parameter estimates and other associated statistics for BAG Model I for calibration dataset (total degrees of freedom (df) = 10268, residual standard deviation (SD) = 0.006839469, residual df = 10261)	95
5.2 Summary statistics for fitted basal area growth models from calibration dataset	96

5.3	Parameter estimates and other associated statistics for BAG Model II for calibration dataset (number of plots = 139, residual df = 10124, and total observations = 10269).....	97
5.4	Fixed-effect parameter estimates and related statistics for BAG Model VI for calibration dataset (number of growth period=2, plots within period = 278, residual df = 9986, and total observations = 10269).....	98
5.5	Parameter estimates and other associated statistics for BAG Model VII for calibration dataset (number of plots = 139, residual df = 10124, and total observations = 10269)	101
5.6	Parameter estimates and other associated statistics for BAG Model VIII for calibration dataset (number of plots = 139, residual df = 10124, and total observations = 10269).....	102
5.7	95% confidence intervals for the parameters in BAG Model VIII.....	103
5.8	Parameter estimates and other associated statistics for BAG Model X for calibration dataset (number of periods = 2, plots within period = 278, residual df = 9986, and total observations = 10269)	103
5.9	Summary statistics for candidate basal area growth models using validation dataset	106
5.10	Summary of results for Height Model I without random-effects for plots for calibration dataset (total observations = 5968, residual df = 5964, residual variance=22.8269)	119
5.11	Summary of results for Height Model II with random-effects for plots for calibration dataset (total observations = 5968, group df = 138).....	120
6.1	Summary statistics for fitted basal area growth models using complete dataset (total observations = 15,669, and number of plots = 208).....	125
6.2	Parameter estimates and other associated statistics for BAG Model I for complete dataset (number of plots = 208, total observations=15,669, and residual df = 15,661).....	126
6.3	Parameter estimates and other associated statistics for BAG Model II for complete dataset (number of plots = 208, total observations = 15,669, and residual df = 15,455).....	127

6.4	Parameter estimates and other associated statistics for BAG Model VI for complete dataset (number of periods = 2, plots within period = 416, residual df = 15248, and total observations = 15,669).....	128
6.5	Parameter estimates and other associated statistics for BAG Model VII for complete dataset (number of plots = 208, total observations = 15,669, and residual df = 15,455).....	129
6.6	Parameter estimates and other associated statistics for BAG Model X for complete dataset (number of plots = 208, total observations = 15,669, and residual df = 15,248).....	130
6.7	Parameter estimates from SAS PROC NLIN for complete dataset (total observations = 8964, and residual variance = 21.5688).....	136
6.8	Parameter estimates from PROC NLMIXED for complete dataset (total observations = 8964, and number of groups = 208).....	136

LIST OF FIGURES

Figure	Page
1.1 General distribution of shortleaf pine the US	2
3.1 Shortleaf pine growth study locations in eastern Oklahoma and western Arkansas (Lynch et al. 1991).....	51
3.2 A representation of 0.2 acres measurement plot with two 5-milacre plots (USDA Forest Service 1995)	53
3.3 Histogram of mid-DBH for the first growth period.....	63
3.4 Histogram of mid-DBH for the second growth period	63
5.1 Plot of residuals vs. tree basal area for BAG Model II.....	100
5.2 Plot of standardized residuals vs. fitted values for BAG Model II.....	107
5.3 Plot of standardized residuals vs. fitted values for BAG Model VI	107
5.4 Plot of standardized residuals vs. fitted values for BAG Model VII.....	108
5.5 Plot of standardized residuals vs. fitted values for BAG Model X.....	108
5.6 Plot of standardized residuals vs. tree basal area for BAG Model II.....	109
5.7 Plot of standardized residuals vs. tree basal area for BAG Model VI	109
5.8 Plot of standardized residuals vs. tree basal area for BAG Model VII.....	110
5.9 Plot of standardized residuals vs. tree basal area for BAG Model X	110
5.10 Boxplots of standardized residuals by age class for BAG Model II.....	111
5.11 Boxplots of standardized residuals by age class for BAG Model VI	112
5.12 Boxplots of standardized residuals by age class for BAG Model VII.....	112
5.13 Boxplots of standardized residuals by age class for BAG Model X.....	113

5.14	Boxplots of standardized residuals by site index class for BAG Model II.....	113
5.15	Boxplots of standardized residuals by site index class for BAG Model VI.....	114
5.16	Boxplots of standardized residuals by site index class for BAG Model VII....	114
5.17	Boxplots of standardized residuals by site index class for BAG Model X.....	115
5.18	Boxplots of standardized residuals by basal area class for BAG Model II.....	116
5.19	Boxplots of standardized residuals by basal area class for BAG Model VI.....	116
5.20	Boxplots of standardized residuals by basal area class for BAG Model VII....	117
5.21	Boxplots of standardized residuals by basal area class for BAG Model X.....	117
5.22	Plot of standardized residuals vs. predicted values for total height model.....	121
5.23	Plot of standardized residuals vs. diameter at breast height for total height model.....	121
5.24	Plot of standardized residuals vs. dominant height for total height model.....	122
5.25	Boxplots of standardized residuals for Height Model II by age classes.....	122
5.26	Boxplots of standardized residuals for Height Model II by site index classes.....	123
5.27	Boxplots of standardized residuals for Height Model II by stand basal area classes.....	124
6.1	Standardized residuals vs. fitted values for BAG Model II (complete dataset).....	131
6.2	Standardized residuals vs. fitted values for BAG Model VI (complete dataset).....	132
6.3	Standardized residuals vs. fitted values for BAG Model VII (complete dataset).....	132
6.4	Standardized residuals vs. fitted values for BAG Model X (complete dataset).....	133
6.5	Boxplots of standardized residuals for BAG Model II (complete dataset).....	133

6.6	Boxplots of standardized residuals for BAG Model VI (complete dataset)	134
6.7	Boxplots of standardized residuals for BAG Model VII (complete dataset)	135
6.8	Boxplots of standardized residuals for BAG Model X (complete dataset)	135
6.9	Plot of standardized residuals vs. predicted values for total height model from complete dataset	138
6.10	Plot of standardized residuals vs. DBH for total height model from complete dataset	138

CHAPTER I

INTRODUCTION

Forests have been managed since time immemorial for direct benefit of human kind, although some forests are kept untouched for several indirect (or even unknown) reasons. Forest management involves one or more than one objective such as timber production, biodiversity and water supply. With increasing understanding of basic sciences and increasing demand for resources, forest management is becoming more and more important (Davis et al. 2001). Progressive forest management requires a good understanding of how individual trees and stands grow over time under different environmental conditions. Therefore, forest growth and yield modeling is an integral part of forest management, at least in developed countries.

Shortleaf pine (*Pinus echinata* Mill.) is second to loblolly pine (*Pinus taeda* L.) in terms of volume of the southern pines in the United States. It has the widest range of any southern yellow pine in the US. It grows in 22 states over more than 440,000 square miles (1,139,600 km²), ranging from southeastern New York to eastern Texas. Despite its importance, there has been relatively little research and management effort compared to other southern pines (Willet 1986). The overall distribution of shortleaf pine in the US is presented in Figure 1.1.



Figure 1.1. General distribution of shortleaf pine in the US

(Source: Western North Carolina Nature Center,

http://wildwnc.org/trees/Pinus_echinata.html March 30, 2006)

Shortleaf pine is adapted to a range of geography, soils, topography and habitats; however, individual trees grow best on deep well-drained soils of the Upper Coastal Plain, and the most prominent shortleaf communities are found in the Ouachita Highlands (Guldin 1986). Natural stands of shortleaf pine occur from almost sea level to 3,300 feet (1,006 m). Favorable average temperature for growth ranges from 48^o to 70^oF, with minimums of -22^oF and maximums of 102^oF. A rainfall range of 40 to 55 inches per year is common in areas where shortleaf stands occur naturally (Williston and Balmer 1980).

It is believed that the Ouachita Mountains of Arkansas and Oklahoma originally contained the largest shortleaf pine forest in the world. According to an estimate, shortleaf and shortleaf-hardwood stands must have covered approximately 5,000 square

miles, which is approximately 50% of the Ouachita Mountains, until the middle of the twentieth century (Smith 1986). Shortleaf pine is common on non-industrial private ownerships in Oklahoma and Arkansas, but is important in industrial ownerships also. It is native to 20 counties of Oklahoma. Shortleaf pine is also the state tree of Arkansas (ODAFS 2000).

It is a common observation that loblolly pine is preferred over shortleaf pine for commercial reasons due to higher growth rates at younger ages. However, shortleaf pine continues to be an important pine species in the south, especially in naturally regenerated forests (McWilliams et al. 1986).

Forest Growth and Yield Modeling

Forest resource management goals can be achieved only after proper assessment of available resources. Forest resource assessment is becoming increasingly quantitative, and more reliable estimate of resources scattered in a wider area can be made. The estimation procedure can be improved both in the field during data collection and in the office during data management and analysis. This can be achieved by following rigorous scientific techniques and methods.

For resource assessment, standard methods in existence can be followed, although it is a subject of continuous research. With the advent of technologies in other areas (e.g. computing) research methods in forestry are also changing over time. Researchers are using techniques like remote sensing and geographic information systems for forest resource assessments at a larger scale.

We can recognize the following two extremes in forest growth modeling in terms of the scale of unit of sampling and observation.

1. Models based on individual plant level information for biological processes that attempt to generalize to growth models that could be used for economically important practical applications. These are generally called physiological process models.
2. Models based on large scale geographic information systems such as models utilizing satellite imagery data that attempt to estimate or predict stand or forest level information that could be used for practical applications. A complete issue (number 3) of Volume 49 of *Forest Science* (2003) has been devoted to research publication for this aspect, although a wide array of literature are available from elsewhere.

Forest growth models in general have often been regression-type statistical models utilizing site index, tree and stand measurements. These traditionally popular growth models often ignore environmental variables such as climatic factors. Such typical forest growth models can possibly be considered to fall somewhere in between the two extremes mentioned above. Forest growth data are typically longitudinal in nature resulting from repeated measurements in permanent plots or approximately longitudinal in nature resulting from cross-sectional data from tree measurements in stands of different ages.

Traditional forest growth and yield models have been used to predict the present as well as future states of forest conditions. It has been common to use data from temporary

plots, however permanent plot studies are becoming more popular. Modeling techniques are also being refined to make use of new developments in the areas of statistics and computing.

Forest growth and yield models can be developed either for natural stands or for plantations. The models for natural stands could be either for even-aged or uneven-aged stands. Similarly, the models for plantations could be either for thinned or unthinned stands (Clutter et al. 1983). According to Davis et al. (2001), forest growth and yield models can be classified as follows.

- (1) Whole stand models: (a) density-free, and (b) variable-density
- (2) Diameter class models
- (3) Individual tree models: (a) distance-dependent, and (b) distance-independent

These models are reviewed, and a short description is presented in Chapter II (Review of Literature). This dissertation presents the work of analysis of shortleaf pine growth measurements data from eastern Oklahoma and western Arkansas. There have been considerable growth studies in the past on shortleaf pine including Murphy (1982, 1986), Lynch et al. (1991, 1999), Murphy et al. (1992), and Lynch and Murphy (1995). Moreover, this project attempts to make an improvement over the previous work and to revise the parameter estimates, especially on the work reported by Lynch et al. (1999). Previous growth model parameters for shortleaf pine have generally been fitted using ordinary least squares methods. Shortleaf pine individual-tree models fitted by ordinary least squares have not accounted for plot-level grouping of individual observations. Mixed-effects models can use random plot effects to account for grouping in this data structure. Thus, it is expected that better parameter estimates can be obtained using more

advanced statistical tools of mixed modeling. Furthermore, this work incorporates a third measurement of permanent plots that was not used by Lynch et al. (1999). It is expected that the resulting estimates could be used for better prediction of the shortleaf pine growth for Oklahoma and Arkansas areas.

Objectives

This dissertation concerns development of individual-tree mixed-effects models for even-aged shortleaf pine. It improves on what has been done in the past on shortleaf pine growth and yield modeling. Specifically, the following are the objectives of this project:

- (1) Review the past work in shortleaf pine growth and yield modeling, and mixed-effects modeling work in forestry in general
- (2) Improve on past work of individual-tree growth model for even-aged shortleaf pine by developing a nonlinear mixed-effects growth model with plot-level random-effects that takes into account possible spatially correlated and heterogeneous errors using a sample (calibration) of complete data set, and with minimal explanatory variables that are easy to measure in the field
- (3) Make further assessments of the model and verification using an independently selected (validation) data set
- (4) Carry out model checking and diagnostics, and arrive with final estimates from the complete data set
- (5) Write and present the work with recommendations in a format that is common for the scientific audience in the field of forestry

CHAPTER II

REVIEW OF LITERATURE

Background

The general objective of this chapter is to review aspects of forest growth and yield modeling relevant to development of shortleaf pine growth models. Moreover, specific objectives are to review shortleaf pine modeling in general, and mixed-effects modeling in forestry. The following areas will be reviewed, but emphasis will remain in application of mixed models in forestry.

- Forest growth and yield models in general
- Shortleaf pine growth and yield modeling
- Mixed-effects models in general
- Application of mixed-effects models in forestry

Growth Data

In the terminology of Moser and Hall (1969), forestry data can be classified as: (1) real growth series, (2) abstract growth series, and (3) approximated real growth series. When a complete chronological data series from establishment (e.g. planting) to harvest for a tree or stand is available, then it is called a “real growth series.” On the other hand,

when data are available from different temporary plots representing different age groups and sites, then the data are called an “abstract growth series.” Ideally, one would prefer data from real growth series, but due to time and expense factors, real growth series data are not common. On the other hand, abstract growth series data may not be of high quality. Therefore, it is common to have an “approximated real growth series.” An approximated real growth series is in between real and abstract growth series in which remeasurements are utilized from several age groups and sites. In this method, data are collected from a wide range (in terms of age and site) of permanent plots, but a particular plot would not have data from a complete life history of trees. Data for this study are of approximated real growth series nature, which will be discussed in Chapter III. Repeated measurements from permanent plots have different sources of variation. The error component in a linear model from such measurements can be regarded to consist of plot, time, and residual random variations (Gregoire 1987).

Modeling Biological Growth

A model is a description of a system. A system can be defined as any collection of interrelated objects or units in which observations are made. Therefore, a description is a signal that can be interpreted by humans. That is, a system is anything humans wish to understand and models are one tool that facilitates the understanding (Haefner 1996). Mathematical models not only should fit the real-life data well, but also should provide some practically useful interpretation. Irrespective of the nature of the model, whether empirical or theoretical, the modeler should be careful in selecting explanatory variables to provide realistic and robust predictions (Vanclay 1994). Typically, nonlinear

regression models of biological growth should have parameters that can be interpreted in terms of biological significance. In growth modeling, one would be mostly interested at the rates of growth over time, and at the time point when growth is stopped. It is also important to be able to understand the process under which different factors (genetic and environmental) influence the growth rate. This information might help in practice to manipulate the growth artificially to some extent by altering the growth rate.

As biological growth is achieved due to cell division, it is in theory considered an exponential process (Zeide 1989), which is constrained by several factors. Biological systems inherently have two common components of gain and loss in metabolism commonly called anabolism and catabolism. Such processes have been considered in modeling biological systems, such as early work by Von Bertalanffy (Pienaar and Turnbull 1973). According to Pienaar and Turnbull (1973), an allometric model can be presented as:

$$P = cQ^a \tag{1}$$

where

P = one dimension of an organism, e.g. length of the femur

Q = another dimension of the same organism, e.g. width of the skull

a = allometric constant, which characterizes the particular kind of organism and environment

c = parameter depending on initial conditions

This model assumes that the specific growth rate of P is constant proportional to the specific growth rate of Q , i.e.

$$\frac{dP}{Pdt} = a \left(\frac{dQ}{Qdt} \right) \quad (2)$$

Von Bertalanffy's model can be written as (Pienaar and Turnbull 1973):

$$\frac{dY}{dt} = \alpha Y^{2/3} - \gamma Y \quad (3)$$

where

Y = size of the organism or population

t = time

α, γ = constants ($\alpha > 0, \gamma > 0$), and $2/3$ is the allometric constant.

The Chapman-Richards function is a flexible model which conceptualizes the growth rate of an organism or a population as a resultant of an anabolic growth rate, which is a positive term, and a catabolic growth rate, which is a negative term (Clutter et al. 1983). This function introduces additional parameter in Bertalanffy's function replacing the $2/3$ allometric constant. A standard Chapman-Richards function is written as:

$$\frac{dY}{dt} = \alpha Y^\beta - \gamma Y \quad (4)$$

where

Y = size of the organism or population

t = time

α, β, γ = constants ($\alpha > 0, 0 < \beta < 1, \gamma > 0$)

This model attempts to explain how an organism or population grows over time in presence of environmental stress or pressure that tends to reduce growth.

The Chapman-Richards function is popular in forest growth and yield modeling, which is empirical in nature. The model is derived as a generalization of the Bertalanffy's growth model (Pienaar and Turnbull 1973, Yang et al. 1978). According to Yang et al. (1978), Weibull function can be modified to model different biological growth processes due to flexible nature of the function. Zeide (1989) described growth equations as a combination of power and exponential functions by taking diameter growth as an example. He compared (tested) five different forms of diameter growth models (Chapman-Richards, Gompertz, Logistic, Power decline and Weibull) in integral form, and found that power decline model, which he derived, performed better in predicting growth for various species and site qualities. He presented an elementary power function called power decline 1 as:

$$\frac{dY}{Ydt} = at^{-b} \quad (5)$$

where

Y = a measure of plant size

$t = \text{plant age}$

$a, b = \text{constants } (a, b > 0)$

The parameters a and b are interpreted as the initial relative growth rate and the rate of aging, respectively.

Many biological systems are modeled using standard regression-type techniques as described in Draper and Smith (1998), and Neter et al. (1996). Nonlinear regression techniques are more popular than standard linear regression to model real-life biological data (Bates and Watts 1988, Ratkowsky 1983). Interpretability, parsimony and validity beyond the observed range of the data are the main reasons why nonlinear models are often more appropriate than linear models for biological data (Pinheiro and Bates 2000). Haefner (1996) is an example of how complex biological systems can be modeled through computer simulation, instead of using empirical methods. Simulation is a popular method in systems modeling. However, many forest growth and yield modeling exercises are not systems modeling, so statistical models supported by measurements are still popular methods. However, process modeling is also gaining popularity in forestry (Baldwin et al. 2001, Johnsen et al. 2001), especially in organizations where strong multi-disciplinary teams of forest biometricians, computer scientists and physiologists exist.

There has been considerable work in forest growth and yield modeling including southern pines, but there is much less work in shortleaf pine growth studies despite its relative importance. Murphy (1986) is a good source on shortleaf pine growth studies prior to 1986. Later work will be described in the following sections.

Types of Forest Growth Models

Density-Free Stand-Level Models

These models are based on full stocking assumptions. Full stocking, often called “normal”, is the density thought to maximize standing tree volume. However, the concept of normal density is subjective, and it is rarely observed in practice. “Normal” yield tables are developed from temporary plot measurements. For example as cited by Rose (1998), Sylvester (1938) developed yield tables based on data from 240 plots of loblolly pine in Louisiana and Arkansas, and compared the results with the yield tables of Miscellaneous Publication 50 (USDA Forest Service 1929). It was concluded that the yield tables of Miscellaneous Publication 50 were not very correct. But the conclusions drew comments from others such as F.X. Schumacher (1939), again as cited by Rose (1998) that the normality concept was subjective, which led to the discrepancy. The yield tables developed from average density instead of maximum density are called empirical yield tables. Another example of density-free yield table is the work of Schumacher and Coile (1960), which is quoted by Lynch et al. (1999) for shortleaf pine. Schumacher and Coile (1960) developed “normal” yield tables based on data from 74 temporary plots in the Piedmont region of North Carolina.

Variable-Density Stand-Level Models

In this category of growth and yield models, stand density is used as an explicit explanatory variable. These are considered an improvement over density-free models, as growth and yield information can be obtained at varying levels of stand density.

According to Clutter et al. (1983), Schumacher (1939) developed the following variable-density model:

$$\ln(V) = \beta_0 + \beta_1 A^{-1} + \beta_2 f(S) + \beta_3 g(D_s) \quad (6)$$

where

V = per unit area volume/yield,

A = stand age,

$f(S)$ = function of site index,

$g(D_s)$ = function of stand density, and

$\beta_0, \beta_1, \beta_2, \beta_3$ are model parameters

This framework can be used to fit a stand-level variable-density model. Clutter (1963) developed a compatible growth and yield model for loblolly pine. Compatible growth and yield models are developed with mathematical properties that permit yield models to be derived by integration of growth models (Davis et al. 2001). Murphy and Beltz (1981) developed the first variable-density models for natural even-aged shortleaf pine. They used permanent plot measurements from Arkansas, Louisiana, Oklahoma and Texas. Murphy (1982) used the same set of data and basal area equation to predict sawtimber volume.

Lynch et al. (1991) developed variable-density stand volume equations for natural even-aged shortleaf pine of eastern Oklahoma and western Arkansas. They used the Schumacher-type yield model with data from 191 permanent plots. The plots were established by the USDA Forest Service and Oklahoma State University in the period

between 1985 and 1987. The models are available for estimating merchantable cubic-foot, sawtimber cubic-foot, and boardfoot volumes per acre for natural even-aged shortleaf pine. The volume equations have the following general form:

$$V = \beta_0 B^{\beta_1} H^{\beta_2} \quad (7)$$

where

V = volume per unit area,

B = basal area per unit area,

H = average total height of dominant and codominant trees, and

$\beta_0, \beta_1, \beta_2$ are model parameters

This equation predicts current volume. Future volume can be predicted by first estimating the future basal area and total height, and then by using the projected basal area and total height in the volume equation. The volume growth prediction is obtained by using the basal area growth projection equation along with the stand volume equation. The future basal area growth per unit area can be projected as a function of stand density and age (Lynch et al. 1991).

The following shortleaf pine site index relationship for Ouachita Mountains has been developed by Graney and Burkhart (1973):

$$H = [a_0 + a_1 SI] [1 - \exp(-(a_2 + a_3 SI) \times A)]^{a_4} \quad (8)$$

where

H = average total height (ft) of dominant and codominant trees,

SI = site index (ft at base age 50 years),

A = stand age (years), and

a_0, a_1, a_2, a_3, a_4 are model parameters

Since the parameters in this equation are intrinsically nonlinear, they must be fitted by nonlinear regression methods. A numerical method as explained in Gerald and Wheatley (1994) must be used to find a solution for site index (SI). Similarly, some other site index (height prediction) equations have been suggested for southern pines, including shortleaf, by Farrar (1973). Alternative methods of site index estimation for shortleaf as well as other southern pines are mentioned in Miscellaneous Publication 50 (USDA Forest Service 1929), as cited by Lynch et al. (1999).

Zeide (2002) also developed techniques for growth modeling for even-aged stands in which stand density (number of trees per unit area and their average diameter) is explicitly included in the model. The modeling framework combines two sets of processes; one that models the growth of trees and the other that makes adjustments to the growth process as a result of competition among trees. Stand density is an important variable that is part of tree dynamics in a stand, which can be used to manage the stands with a specific objective.

Diameter Class Models

Diameter class models provide growth and yield information by specified diameter classes, so these are more detailed than whole-stand models. Stand volume can be computed by summing the volumes over diameter classes. Diameter class models often use a probability density function to allocate individual trees to diameter classes. The Weibull distribution is the probability distribution most widely used to model diameter distributions, because it is a flexible distribution that can accommodate a range of shapes and scales of distribution. Bailey and Dell (1973) first used the Weibull distribution to model tree diameter distribution. They found it to be very useful in modeling diameter distribution, although other distributions such as beta and exponential functions have also been used. The Weibull probability density can be presented as:

$$f(X) = \frac{c}{b} \left(\frac{X-a}{b} \right)^{c-1} \exp \left[- \left(\frac{X-a}{b} \right)^c \right] \quad (a \leq X < \infty) \quad (9)$$

= 0, otherwise

where

X = random variable such as diameter at breast height

a , b , and c are parameters of the distribution

Parameter a is called the location parameter, which is non-negative ($a \geq 0$) for diameter distribution, although it can take any real value theoretically. Parameters b and c are respectively called scale and shape parameters, and must always be positive ($b, c > 0$).

The Weibull is a very flexible probability distribution. Certain other distributions can be represented as special cases of the Weibull. For example, if $a=0$ and $c=1$, it is actually an exponential distribution. These results are found in standard statistical references such as Mood et al. (1974) and Wackerly et al. (2002). Use of the Weibull distribution in forestry has been well documented by Clutter et al. (1983) and Bailey and Dell (1973). To use the Weibull distribution for diameter class modeling, the model parameters need to be estimated, which is typically achieved by using maximum-likelihood method. Then the number of trees in a stand can be multiplied by appropriate diameter class probability to assign number of trees in diameter classes accordingly. Smalley and Bailey (1974) used the Weibull distribution to develop a diameter class model for shortleaf pine plantations. They used tree measurements from 104 shortleaf pine plantation plots in Tennessee, Alabama and Georgia.

Individual-Tree Models

Models which use individual trees as the basic unit to predict stand growth and yield are called individual-tree models. Such models can be either distance-independent or distance-dependent, depending on whether or not the spatial arrangement of individual trees is taken into account. Distance-independent models predict tree growth either individually or by size classes. Growth prediction is made as a function of present size and stand-level variables such as stand age, site index, and stand basal area (Avery and Burkhart 2002).

Individual tree data are aggregated after the model grows each tree under individual-tree model, whereas the stand level model aggregates individual tree data into

stand variables before applying the growth model (Davis et al. 2001). So growth rate estimates are on an individual tree basis and stand level basis, respectively. That is, no individual tree growth estimate can be made from a stand level model, whereas stand level growth can be estimated from an individual-tree level model. Therefore, diameter class and stand level information can be derived from individual-tree models.

Distance-Dependent Individual-Tree Model

This model takes the spatial relationship between trees into account using a coordinate system. Initial stand conditions are either user-specified or computer-generated, and the growth of each tree is simulated as a function of its attributes, the site quality, and a measure of competition from neighbors as per the assigned tree locations (coordinate system). Yield is estimated by summing the individual-tree values (either from tree volume or a taper equation), and by adjusting for some expansion factors (Avery and Burkhart 2002). This modeling approach requires extensive computer programming.

PTAEDA2 is an example of distance-dependent individual-tree model developed for loblolly pine. This simulation system has two main sub-systems: (1) one dealing with the generation of an initial pre-competitive stand, and (2) another with the growth and dynamics of that stand. Various management scenarios for competition, fertilization and thinning options have been incorporated in the system (Burkhart et al. 1987, Avery and Burkhart 2002). Different diameter growth and height growth models were evaluated by Martin and Ek (1984) for red pine plantations. They observed some benefit of incorporating a competition index in the model for diameter growth model, but not in

height growth model. Similarly, Tome and Burkhart (1989) compared different measures of competition for *Eucalyptus* plantations under a spacing study in Portugal.

Spurr (1962) developed a method of quantifying competition as point density by using a sampling angle gauge (prism). Another measure of point density as the potentially available area was developed by Brown (1965), as cited by Rose (1998). Crown competition factor (CCF) as defined by Krajicek et al. (1961) is a popular competition index (Avery and Burkhart 2002). Bullock and Burkhart (2005) evaluated spatial dependence in juvenile loblolly pine plantations using a simultaneous autoregressive model, and they found a significant dependence among neighboring stems.

Distance-dependent models are based on the concept of spatial variation. The literature of spatial statistics is huge (e.g. Cressie 1993, Ripley 1981, Webster and Oliver 1990). Spatial statistical methods are popular in geography, geology and ecology, but have started to be used in other fields. Units that are physically close tend to be similar, and correlation between the units decreases as the distance between the units increases. However, the plants or animals too close together might also compete for resources. For example, Mead (1967) has developed a mathematical model for estimating inter-plant competition. This kind of inter-plant competition might lead more dissimilar plants that are physically close. Therefore, the utilization of concept of spatial variability or competition measures depends on purpose and the scale of consideration.

In summary, distance-dependent models are more expensive and computationally intensive than distance-independent models. Distance-independent models are popular among forest professionals (Tome and Burkhart 1989, Davis et al. 2001).

Distance-Independent Individual-Tree Model

Distance-independent models do not require data on spatial location for each tree. That is, they ignore actual spatial distribution of trees to estimate individual tree growth, and assume that the trees are uniformly distributed over space. Typically, a distance-independent model consists of the following components (Avery and Burkhart 2002):

- (1) a model for tree diameter (or basal area) growth
- (2) a model for tree height growth (or height-diameter relationship to predict height as diameter is easily measured)
- (3) a tree mortality estimate

Combining the three model components into a simulation system would provide different output scenarios with different input or management variables such as thinning/density regimes and site quality. These outputs might be useful to a forest manager in decision making.

Distance-independent models have comparatively less data demand than distance-dependent models. In distance-independent modeling, measures of competition often relate individual tree size to average tree size or to the size distribution. Two most commonly used competition measures are: (1) ratio of the quadratic mean diameter to diameter of an individual tree at breast height, and (2) cumulative basal area of trees larger than the subject tree. Basal area growth is a common response variable modeled in practice, because it is less sensitive to competition indices than diameter growth (Bella 1971). West (1980) compared different models using both basal area growth and diameter growth as response variables, and found that basal area growth was more

strongly related with explanatory variables than diameter growth. Depending on the type of model, different stand level as well as individual-tree level variables can be used as explanatory variables. For example, Lynch et al. (1999) used two stand level variables (stand age and stand basal area), and two tree level variables (ratio of quadratic mean diameter to DBH of individual tree and tree basal area) to model annual basal area growth of natural even-aged shortleaf pine in Oklahoma and Arkansas.

Distance-independent models can be either (1) composite models of tree growth as a function of tree measurements, site properties and stand characteristics or (2) potential-modifier growth functions. The potential-modifier function models tree growth as a theoretically possible maximum growth, which is expressed as a function of tree characteristics. The potential is then adjusted for the modifier, which is expressed as a function of a stand and tree characteristics including competition (Rose 1998). The Forest Vegetation Simulator (FVS), previously called PROGNOSIS, is a simulation framework for composite distance-independent modeling (Davis et al. 2001, see also www.fs.fed.us/fmnc/fvs). The composite growth model is appropriate when it is difficult to obtain dominant-age and site relationships for mixed-species stands. In such cases, a potential-modifier growth function is difficult to apply (Wykoff 1990). FVS can be used to simulate growth for either basal area or diameter from such models, and modeling could be either composite or potential-modifier type in nature.

The following composite model was fitted by Wykoff (1990) to inventory and regeneration study data sets for 11 conifer species in the northern Rocky Mountains.

$$\ln(ds) = b_0 + b_1 \ln(dbh) + b_2 dbh^2 + b_3 SL[\cos(ASP)] + b_4 SL[\sin(ASP)] + b_5 SL + b_6 SL^2 + b_7 EL + b_8 EL^2 + b_9 CR + b_{10} CR^2 + b_{11} BL / \ln(dbh + 1) + b_{12} CCF \quad (10)$$

where

$\ln(ds)$ = natural logarithm of 10-year periodic change in squared diameter
(inches)

dbh = tree diameter outside bark at breast height (in.)

BL = total basal area in trees with larger dbh than the subject tree (ft^2/ac)

CR = ratio of live crown length to total tree height

CCF = crown competition factor

SL = average slope percent for the stand

ASP = average aspect for the stand (radians)

EL = average elevation for the stand (100 ft)

b_i = estimated regression coefficients; b_0 dependent on habitat type and location,

b_2 dependent on location, and b_{12} dependent on habitat type

This model incorporated the site characteristics (EL , SL and ASP for each combination of habitat and forest type), competition indices (CR , CCF and BL) and tree diameter. The model is specific to a habitat and forest type (location) as quantified by a separate intercept for each combination of habitat and forest type. The model was also validated using data based on permanent-plot measurements.

On the other hand, a potential-modifier function is suitable when we can easily obtain dominant-age and site relationships, for example even-aged stand of a single species, in which site-index is often used. This is one of the reasons why even-aged natural shortleaf pine has previously been modeled based on this approach. There are two components in this framework, a potential and a modifier. The potential quantifies the

theoretical maximum growth (diameter or basal area) of a tree which is grown without any competition for resources, e.g. an open-grown tree. The modifier component makes adjustment to the potential in light of the competition factors that exist in the field, e.g. competition for light based on crown size, spacing etc. Since a potential-modifier model has two components, parameter estimation can be achieved in one of the two ways: (1) fit potential function and then estimate modifier by fixing the potential to a constant, or (2) fit both potential and modifier functions simultaneously. The second approach may be a better one as its estimation procedure considers both the functions together, which is what happens in tree growth. Tree growth is dynamic interplay between a potential to grow and slowing down of the growth due to environmental restrictions.

Shifley and Brand (1984) provide a good example of how a potential function can be constrained to follow a biological principle. They used a modified Chapman-Richards function (Pienaar and Turnbull 1973) to restrict the tree growth to a biologically possible total size.

When a standard Chapman-Richards equation is set to zero, and solved for organism/tree size (to find the maximum as the Chapman-Richards equation is the first derivative of size with respect to time), the resulting tree size would be a maximum biologically potential size (Shifley and Brand 1984).

$$\text{Potential/Asymptote } (A) = \left(\frac{\alpha}{\lambda} \right)^{\frac{1}{1-\beta}} \quad (11)$$

Any measure of tree size, such as basal area, could be used. The derivative form of Chapman-Richards function presents the variable in growth form, so if tree size is represented by basal area, then the response variable would be basal area growth.

The modified Chapman-Richards model obtained by Shifley and Brand (1984) was of the following form:

$$\frac{dY}{dt} = \alpha Y^\beta - \alpha Y / A^{(1-\beta)} \quad (12)$$

One needs to insert an appropriate value for A (e.g. maximum possible tree basal area) to revise the estimates of α and β . According to Shifley and Brand (1984), such a value for A can be obtained from the National Register of Big Trees (American Forestry Association 1982 as cited by Rose 1998).

Murphy and Shelton (1996) also used a Chapman-Richards function for potential of the form

$$\text{Potential} = \beta_1 [1 - \exp(\beta_2 B_t)] \quad (13)$$

where

B_t = average tree basal area during the growth period

β_i = parameters, $i = 1,2$

They experienced a severe convergence problem with the nonlinear least-squares estimation. Some other examples of potential functions are found in Amateis et al. (1989) and Smith et al. (1992). Amateis et al. (1989) proposed the idea of estimating potential

growth from open grown trees without any competition effects. This idea motivated Smith et al. (1992) to estimate maximum potential growth for shortleaf, longleaf and loblolly pines.

A modifier function takes the growing conditions into account to adjust the potential to estimate net growth. Several examples of modifier functions are available in forestry literature. One that is most relevant to shortleaf pine growth modeling is that of Hitch (1994). He considered two modifier functions: (1) A variant of Shifley's (1987) functions for TWIGS and STEMS for the Central States, and (2) A modified logistic function developed by Murphy and Shelton (1996). The logistic function is a useful modifier, since it is restricted between 0 and 1, and can include several independent variables. The logistic modifier has a form:

$$\text{Modifier} = \frac{1}{1 + \exp(\beta_3 B_l + \beta_4 B_s + \beta_5 D_q)} \quad (14)$$

where

B_l = basal area in trees larger than or equal to the diameter of the subject tree

B_s = stand basal area

D_q = quadratic mean diameter

β_i = parameters, $i = 3,4,5$

One advantage of the logistic modifier function is that we can add more explanatory variables to improve the fit index of a model, but still it is within the range of 0 and 1. So a complete growth model would take the form

Predicted growth = Potential growth × Modifier

That is, the Murphy and Shelton (1996) model would be:

$$\frac{\beta_1 [1 - \exp(\beta_2 B_t)]}{1 + \exp(\beta_3 B_1 + \beta_4 B_s + \beta_5 D_q)} \quad (15)$$

The model terms are defined as before. They used this equation to fit an individual-tree basal area growth model for loblolly pine

A distance-independent individual-tree model for shortleaf pine for the Ouachita Mountains was first developed by Hitch (1994) using the potential-modifier function approach. He used a data set for natural even-aged shortleaf pine in Oklahoma and Arkansas from a cooperative study between OSU Department of Forestry and USDA Forest Service.

Bitoki et al. (1997) used data from a study for uneven-aged continuous forest inventory (CFI) plots for shortleaf pine from the same Oklahoma and Arkansas region to develop a basal area growth model. They also used a potential-modifier function approach. They first estimated the potential growth separately by excluding a parameter from Hitch's (1994) equation. Then the parameters in the modifier function were estimated by keeping the potential constant, so all parameters were not estimated simultaneously. A simultaneous estimation technique for estimating the parameters in a potential-modifier form has been used by Murphy and Shelton (1996) for loblolly pine.

Hasenauer et al. (1998) used simultaneous regression methods to analyze individual tree data for Norway spruce from Australia. They found strong cross-equation

correlations, and the simultaneous estimation method was more efficient than separate estimation by ordinary least squares. Similar estimation method in a nonlinear system was examined for white spruce data by Huang and Titus (1999). An example of simultaneous estimation for height-diameter model fitting is Omule and MacDonald (1991). Rose (1998), and Rose and Lynch (2001) used a seemingly unrelated regression technique to fit a basal area growth model for shortleaf pine as the technique would help model correlated errors. Lynch and Murphy (1995) also used seemingly unrelated regression to fit a height-diameter model for natural even-aged shortleaf pine.

Lynch and Murphy (1995) used the following general framework for modeling height growth in natural, even-aged shortleaf pine stands:

$$\hat{h}_i = f(A_i, \hat{H}_{D_i}, D_i, S_{D_i}, D_{L_i}, \beta) \quad (16)$$

where

\hat{h}_i = predicted individual tree height at time i

A_i = stand age at time i

\hat{H}_{D_i} = predicted average height of dominants and codominants at time i (could be obtained from site index curves in even-aged stands)

D_i = individual tree dbh at time i

D_{L_i} = dbh distribution location identifier at time i such as maximum dbh or quadratic mean dbh

S_{D_i} = an expression of stand density at time i

β = vector of parameters

Lynch and Murphy (1995) also provide a comprehensive review of models relating tree height to DBH and age (or time). The Hitch's (1994) model was later slightly modified by Lynch et al. (1999) by including additional data available from later plot measurements. They developed a complete growth and yield prediction system including models for basal area growth, height-diameter relationship, crown characteristics and mortality functions. The estimates are now being used in the shortleaf pine stand simulator, SLPSS (Huebschmann et al. 1998).

They fitted the following model:

$$G_i = \frac{b_1 B_i^{b_2} - (b_1 B_i / B_M^{1-b_2})}{1 + \exp(b_3 + b_4 B_s + b_5 A + b_6 R_i + b_7 B_i)} + \varepsilon_i \quad (17)$$

where

G_i = annual basal area growth (ft²) of tree i

B_i = basal area (ft²) of tree i

A = stand age (years)

R_i = ratio of quadratic mean stand diameter to the dbh of tree i

B_s = stand basal area (ft²/ac)

$B_M = 7.068384$ ft² (the maximum expected basal area for a shortleaf pine in managed stands)

b_1, b_2, \dots, b_7 = parameters

ε_i = random error

Lynch et al. (1999) used the following nonlinear model for individual tree height given by Lynch and Murphy (1995):

$$(H_i - 4.5) = \beta_0 (H_D - 4.5)^{\beta_1} \exp(-\beta_2 D_i^{-\beta_3}) + \varepsilon_i \quad (18)$$

where

H_i = total height (ft) of tree i

D_i = dbh (in) of tree i

H_D = dominant height as per Graney and Burkhart (1973)

$\beta_0, \beta_1, \beta_2, \beta_3$ = model parameters

ε_i = random error

Beaumont et al. (1999) evaluated the generalized method of moments for modeling dominant height through site index in black spruce. They found that the new method of simultaneous estimation to be better than traditional two-equation approach. In the traditional method, site index is first estimated as the average height of dominant trees at base age, and then dominant height is predicted as a function of age and site index as done by Graney and Burkhart (1973).

Uneven-Aged Stand Modeling

Several basic ideas of even-aged stand modeling are relevant to uneven-aged stand modeling also. Furthermore, this dissertation deals with data from even-aged shortleaf stands only. Therefore, only a brief review indicating the differences is presented.

The most important difference in uneven-aged stands modeling is that the concept of site index and stand age cannot be easily used since there are different age-classes in these stands. And at least in principle, modeling techniques for uneven-aged stands can also be applied to even-aged stands (Avery and Burkhart 2002). Key ideas on uneven-aged stands modeling are presented in a classic work by Moser and Hall (1969), and Moser (1972). They developed a modeling framework in which yield is expressed as a differential function of elapsed time from a given initial condition. A Markov Chain approach for predicting diameter distributions in uneven-aged stands is provided by Bruner and Moser (1973). In such a framework, future diameter distributions, number of surviving trees, number of dead trees, and number of harvested trees are predicted using inventory data under a Markov process, which is a stochastic process for modeling an uncertain event. Lynch and Moser (1986) developed a technique of predicting stand tables for two species groups based on a differential equations approach.

A matrix model approach to modeling uneven-aged forest management has been reported by Buongiorno and Michie (1980). In this approach, a matrix model of forest growth similar to the Markov model of Brunner and Moser (1973) is combined with linear programming techniques to answer some economic questions. Following the idea of Buongiorno and Michie (1980), Schulte and Buongiorno (2004) developed a growth and yield model for naturally-regenerated shortleaf pine forests including hardwoods from southern US. Their work was based on repeated measurements on forestry inventory and analysis plots.

According to Smith (1986), industrial forestry organizations and USDA Forest Service began a basic shift from uneven- to even-aged stands in about 1970. Many

commercial activities involve even-aged stand management, but some argue against the ideas of even-aged management because of ecological and environmental reasons. Murphy and Farrar (1988) developed a stand-level growth and yield model for uneven-aged loblolly-shortleaf pine using inventory data. Guldin and Baker (1988) carried out yield comparisons for both even-aged and uneven-aged loblolly-shortleaf pine stands. They found that total merchantable cubic-foot yields were highest for conventionally managed even-aged plantations. Atta-Boateng and Moser (2000) developed a compatible growth and yield modeling system for managed mixed forests of the tropical rain forest region. An individual-tree growth and yield model for uneven-aged shortleaf pine stands was developed by Huebschmann et al. (2000).

Forest Growth and Yield Simulation Systems

Different models/systems have been reported in the literature for simulating forest growth and yield. The USDA Forest Service has developed a comprehensive system of vegetation simulation called Forest Vegetation Simulator (FVS). The FVS system simulates vegetation change based on forest inventory data or stand examination data about the forest, stand, and trees (USDA 2002). A brief summary of some simulation systems is presented in Table 2.1.

Table 2.1. Some selected individual tree growth and yield simulation models/systems available for the United States (Adapted from Davis et al. 2001)

Simulation System	Description
The Forest Vegetation Simulator (FVS)	Previously called PROGNOSIS, developed by USDA Forest Service for nationally supported framework for forest growth and yield modeling, more than 20 variants nationwide, many models, developed by universities and research institutions, such as TWIGS, STEMS, or GENGEM are now part of FVS (www.fs.fed.us/fmnc/fvs)
PTAEDA2	Distance-dependent system capable of modeling the growth of loblolly pine plantations, has also been linked with MAESTRO, a process model, developed and supported by Virginia Tech
ORGANON	Developed for western Oregon conifer and hardwood types, developed through a public and industry cooperative and is supported by Oregon State University
CRYPTOS, CACTOS	CRYPTOS developed for the California coastal redwood and fir forests, and CACTOS covers the California mixed conifer forests, both developed through a public and industry cooperative and are supported by University of California-Berkeley
SLPSS	Shortleaf Pine Stand Simulator, which simulates the growth of natural shortleaf pine stands in western Arkansas and eastern Oklahoma, developed from a cooperative study by Oklahoma State University (OSU), the US Forest Service Southern Station, and the Ouachita and Ozark National Forests, supported by the Department of Forestry at OSU.

The following sections present a review of mixed modeling techniques that form the foundation of this dissertation work.

Mixed Models

Fixed vs. Random-Effects

When we are interested only in the levels or classes observed in the study sample, then the estimates are fixed-effects. On the other hand, if the classes or levels observed are a sample from a population, and we are not primarily interested in the particular levels observed, then this prompts the random-effects approach. In this approach, we are interested in the variability pattern in the population itself from which the sample was observed (Snedecor and Cochran 1980, Laird and Ware 1982). For example, when a forest growth model is developed from a fixed-effects approach for the stands observed only, then the model is fixed-effects model, and the results are applicable only to the stands observed in the sample. If the forest growth model needs to be generalized to a forest from which a sample of stands was observed, plot random-effects should be considered. Basically, we are not interested in the particular stands. Therefore, a growth model in which stands/plots have random-effects would be desirable. However, many traditional/regular forest growth models have been based on a fixed-effects approach.

Observations classified/grouped/clustered by a certain categorical variable are often analyzed by using analysis of variance (ANOVA) and analysis of covariance (ANCOVA) techniques. However, these techniques are relevant only when the number of categories is relatively small, and we are interested in comparisons of the observed categories only. This represents a case of fixed-effects. In many cases, we are interested in considering the categories as a representative random sample from a population. A fixed-effects analysis would not be appropriate in such cases, instead a random-effects

approach should be used that estimates the variance components for different levels of a population (Longford 1993). In random coefficient models, the coefficients for each class (intercept and slope for each combination of different levels) quantify deviation from a defined population regression model (Tao 2002).

A mixed model is an extension of a random-coefficient regression model in which fixed-effect parameters are also included. In other words, when a model has both fixed and random-effects, then the model is called a mixed model. A mixed model typically has more than one error level, so such a model is also called a multi-level model. If classification factors form a hierarchy (nested structure), then the model is also called a hierarchical model.

Gumpertz and Pantula (1989) provide a simple way of making inference based on mean of the individual coefficients in random coefficient regression models. They interestingly contrast the longitudinal data in biological science studies with those of economic studies data. Their appreciation of agricultural and biomedical studies with a small number of repeated measures on large number of experimental units/subjects is important in the context of our project. This is different in nature compared to economic and meteorological data in which there are often multiple measurements/observations for a long period of time (time series) on relatively small number of subjects. The mathematical details for mixed-modeling will be addressed in Chapter IV.

Longitudinal Data/Repeated Measures

Data resulting from repeated observations/measurements on the same unit/subject over time are often called longitudinal data. Such data are common in practice, especially

in medicine, biology and economics (Diggle et al. 1994, Gregoire et al. 1995, Davis 2002). Mixed models are an increasingly popular method for analysis of longitudinal data because of their flexibility in handling different data structures (irregularity and unbalancedness) and in satisfying model assumptions. There are other alternative analysis approaches such as a multivariate ANOVA (MANOVA), which are not as flexible and powerful as the mixed modeling approach. MANOVA can not handle irregular and unbalanced data properly, and unbalanced data are very common in practice. Another alternative is to follow a split-plot in time approach (Kuehl 2000). The split-plot approach to repeated measures analysis would assume that observations/treatments within a unit are randomized. That is, different observations over time within the same subject/unit would be assumed random, which is not a very realistic assumption (Diggle et al. 1994). Instead, we would prefer to model within-unit correlation (serial correlation in repeated measurements, and spatial correlation among trees within a plot). Therefore a model that recognizes a proper covariance structure for errors would be a better approach to analyze such data (Gregoire et al. 1995).

Laird and Ware (1982) and Ware (1985) developed a random-effects model approach for analyzing longitudinal data that would model serial correlation on the same subject. A two-level model was developed in which grouping of observations would be made by subjects. Such models could be fitted using either maximum likelihood or residual maximum likelihood (empirical Bayes) methods. Linear mixed-effects models for longitudinal data are described by Chi and Reinsel (1989), and Verbeke and Molenberghs (2000). They assessed maximum likelihood and restricted maximum likelihood methods of estimation. They also used a first-order autoregressive model to

account for within-individual errors resulting from longitudinal measurements on the same individuals.

Crowder and Hand (1990) is a standard reference for repeated measures analysis, including responses of categorical nature. They deal with standard split-plot analysis, MANOVA and ante-dependence analysis to account for consecutive measurements on the same unit. Diggle et al. (1994) described principles and methods for analysis of longitudinal data. They dealt with both quantitative as well as qualitative data types, and provided a significant amount of information on random-effects models. Major theoretical and computational contributions to the nonlinear mixed-modeling approach were made by Vonesh and Carter (1992), and Davidian and Giltinan (1995). Similarly, researchers such as Lindstrom and Bates (1990), and Pinheiro and Bates (2000) developed algorithms and software that allowed practical usage of the models. Reiczigel (1999) reviewed methods for analyzing repeated measurements from designed experiments in which the main objective is to compare treatments as precisely as possible. VanLeeuwen et al. (1996) presented a mixed model that incorporated random trends through time, and also allowed correlations among observations at the same time.

Pinheiro and Bates (2000) developed the `nlme` library in S-Plus that has made a substantial impact, especially on use of nonlinear mixed models in practice. They followed the approach of Laird and Ware (1982) to develop these computational tools. SAS PROC NLMIXED has been considered limited for nonlinear mixed modeling (Tao 2002, Littell et al. 1996), although a random-effect with limited options for covariance structure can be used. Littell (2002) compared analysis of variance (ANOVA) versus residual maximum likelihood (REML)/generalized least squares (GLS) methods for

analyzing unbalanced mixed model data, and found that ANOVA was popular method prior to the early 1990s because of its simplicity and lack of easily available computing tools for likelihood based methods. However, more powerful likelihood based methods such as REML and GLS have started gaining popularity due to the increasing availability of software since the early 1990s. Davidian and Giltinan (2003) provide an overview of nonlinear models for repeated measurements data.

Correlated and/or Heterogeneous Errors

Correlation among successive observations over time on the same unit is discussed in a framework of time series analysis by Diggle (1990). Similarly, spatial correlation among observations is dealt with by Cressie (1993), Ripley (1981), and Webster and Oliver (1990). However, these references do not work in the context of mixed modeling. Both inter- and intra-individual variations in nonlinear mixed modeling for repeatedly measured observations are discussed and reviewed by Davidian and Giltinan (1995, 2003).

Wolfinger (1996) reviewed and evaluated different covariance structures for modeling heterogeneity in repeated measures data. The variance-covariance structures included heterogeneous versions of the compound symmetry and first-order autoregressive structures, the Huynh-Feldt structure, the independent-increments structure, correlated random coefficients models, the first-order antedependence model, and a simplified factor-analytic construction. Use of an appropriate variance-covariance structure would avoid data transformation allowing parameter interpretability and more accurate inference. Lin et al. (1997) discuss linear mixed models with heterogeneous

within-cluster variances. Methods are shown to predict cluster-specific random effects variances.

Application of Mixed Models in Forestry

Ferguson and Leech (1978) and West et al. (1984) were among the earlier workers to recognize the problem of temporal correlation in forestry data due to multiple measurements from individual sampling units. They mentioned the difficulty in regression analysis with regular assumptions for such data, and have discussed and recommended alternative approaches for data analysis. They discussed the availability of statistical theory to solve these problems. However, these discussions were mostly from a designed experiments perspective, so their direct application was not appropriate to many forest growth modeling datasets. Ferguson and Leech (1978) developed two-stage yield models. In the first stage, they used ordinary least squares, whereas a generalized least squares method was used in the second stage to improve the estimation since the errors in the second stage violated the assumptions of ordinary least squares. West et al. (1984) also found a two-stage regression model to be the most suitable method to analyze repeated measurements data common in forestry problems. They demonstrated the problem with a study on *Eucalyptus* in which plot was considered a sampling unit. In this study, a sample of trees was selected from each plot, and measurements such as diameter at breast height were taken.

With motivation from the work of Ferguson and Leech (1978), Gregoire (1987) evaluated four different covariance structures for basal area yield models. He compared

the following covariance structures: (a) uncorrelated and homoscedastic plot and time effects; (b) autoregressive time effects; (c) uncorrelated and heteroscedastic plot effects and autoregressive time effects; and (d) correlated and heteroscedastic plot effects and autoregressive time effects. It was found that method of ordinary least squares was nearly always best; so further work was recommended.

A random stand and tree parameters approach was used to model height in slash pine (*Pinus elliottii* Engelm.) by Lappi and Bailey (1988). This was presented as a good alternative to the conventional site index approach to height prediction. They fitted the following model:

$$h_{ki}(t) = \mu(t) + b_k(t) + e_{ki}(t) \quad (19)$$

where

$h_{ki}(t)$ = dominant height for tree i in stand k at age t

$\mu(t)$ = population mean height at age t over all stands

$b_k(t)$ = random stand effect at age t

$e_{ki}(t)$ = random tree effect and is uncorrelated with $b_k(t)$

They used the well known Richards' (1959) equation to model the mean height, and then estimated the parameters of the covariance structure using observed residuals as if they were the true residuals.

Gregoire et al. (1995) is an excellent reference on mixed modeling issues in forestry. They deal mostly with linear mixed models, but discuss the issues of importance of nonlinear mixed models in forestry research. They have fitted stand-level mixed-

effects models with two data sets consisting of eastern white pine and Douglas-fir with a mixed model approach having plot random-effects. Due to the availability of repeated measurements, random-effects for stands were fitted even though the response was a stand level variable. Errors that are temporally and/or spatially correlated have also been modeled. This project will follow basic ideas from this publication. Lack of easily available and user-friendly software was mentioned by Gregoire et al. (1995) to be an important reason that there was still not much application of nonlinear mixed-effects models in forestry.

Gregoire and Schabenberger (1996a) used a nonlinear mixed-effects approach for modeling individual-tree cumulative bole volume of sweetgum from east Texas. They later modeled cumulative bole volume by taking spatial correlation between sections of a bole into account (Gregoire and Schabenberger 1996b). The objective was to express cumulative bole volume as a smooth curve while allowing for fluctuations within and among trees as much as possible, yet allowing the curve's applicability to all trees of the species with similar morphological characteristics. They used a nonlinear mixed-effects model by comparing restricted maximum likelihood (REML) and generalized estimating equations (GEE), and found that GEE to be simpler (computationally less intensive) than REML. REML is a normal-distribution based estimation method, whereas GEE is a semi-parametric approach for multiple measurements per subject (repeated measurements). They used SAS PROC IML and MACRO for the modeling work.

Candy (1997) estimated parameters in a sigmoidal forest yield model using composite link functions with random plot effects in a generalized linear mixed-effects

model framework. A generalized linear mixed-effects model can be considered as a special case of nonlinear mixed-effects models (Tao 2002).

Lappi (1997) analyzed two jack pine data sets, one from plantations and the other from naturally regenerated stands. Both the data sets were longitudinal in nature, and they were analyzed using the same model structure. The height/diameter curve parameters were partitioned into an age-dependent trend (population mean), a random stand effect, and a random time effect. A good practical use of such a model would be calibration with which the random stand and time effects could be predicted given some additional measurements without needing to make detailed observations as would normally be done for new stands. Tasissa and Burkhart (1998) used stem analysis data from permanent plots for loblolly pine to evaluate thinning effects on form exponent, a measure of stem form. They modeled form exponent in response to thinning and by accounting for correlation among within-tree observations with the first-order autoregressive method. Apiolaza et al. (2000) used variance components estimation to analyze genetic data for *Pinus radiata* from a progeny test.

Fang and Bailey (2001) modeled dominant height growth of slash pine from a study with several silvicultural treatments installed in Georgia and north Florida. They reparameterized the three-parameter Richards model to predict dominant height growth in presence of silvicultural treatments, such as chopping, fertilization and burning, using a nonlinear mixed-effects model approach. The study design was a split-plot, with soil type as main-plot factor and silvicultural treatments as subplot factor. They claim that this would be a better approach to predict dominant height growth rather than using the site index approach. Fang et al. (2001) used a simultaneous equation system to develop stand-

level mixed models for growth and yield of slash pine with two components: basal area model and total volume model.

Another important example of mixed modeling in forestry is Hall and Bailey (2001). They modeled forest growth variables using multi-level nonlinear mixed models with data from a loblolly pine spacing study in Georgia. They used a linearization approach of estimating equation rather than maximum likelihood or restricted maximum likelihood method. They included random-effects for both plots and trees. They found that the multi-level nonlinear mixed model approach provided several advantages over traditional forest growth modeling methods.

Guilley et al. (2004) modeled averaged ring density with individual tree random effects for sessile oak in France. It was concluded that there was no evident effect on wood density due to changing environment and forest management type when ring width and cambial age were constant. Garber and Maguire (2003), and Leites and Robinson (2004) have applied the mixed modeling method to develop taper equations. A nonlinear mixed model with autoregressive error structures was used to model stem taper of ponderosa pine (*Pinus ponderosa* Dougl. ex Laws.), lodgepole pine (*Pinus contorta* Dougl. ex Loud.), and grand fir (*Abies grandis* Dougl. ex D. Don) by Garber and Maguire (2003). Leites and Robinson (2004) improved the Max and Burkhart's (1976) taper equation with crown dimensions and individual tree random-effects for loblolly pine.

Mehtätalo (2004) used a mixed model with height and diameter data of longitudinal nature for Norway spruce (*Picea abies* (L.) Karst.). The Korf growth curve was used as a basic growth function for the relationship. The model could be calibrated for a new stand with substantially less observations than one would require for fitting a

completely new model. It was concluded that the growth pattern of a stand was dependent on mean tree size in the stand but not on stand age. This may or may not be true in other species.

Zhang and Borders (2004) used a mixed-model to estimate tree component biomass for managed loblolly pines with an allometric approach. They found that the percent of stem biomass increased with age while the opposite was the case for foliage and branches. It was also found that cultural treatments affected the proportional allocation among various tree compartments. Calama and Montero (2004) used a mixed model approach to model individual-tree height-diameter relationship for stone pine (*Pinus pinea* L.) in Spain.

Lynch et al. (2005) used a random-parameter (mixed model) approach to analyze height-DBH data for cherrybark oak (*Quercus pagoda* Raf.) from east Texas. They fitted a similar model as reported by Lappi (1991). That is, natural logarithm of difference between total height and breast height was modeled. They fitted a linear model with random-effects for stands, which could be used for calibration with minimum observations from a new stand. The fitted model is as follows:

$$\ln(H_{ki} - bh) = \beta_0 + \beta_1 D_{ki}^{-0.9} + \alpha_{0k} + \alpha_{1k} D_{ki}^{-0.9} + e_{ki} \quad (20)$$

where

H_{ki} = total height of tree i in stand k

bh = breast height

D_{ki} = dbh of tree i in stand k

β_0, β_1 = fixed effect parameters

α_{0k}, α_{1k} = random effect parameters with 0 mean for stand k

e_{ki} = random residual error for tree i in stand k

They also present an example of how this model could be calibrated for new stands. Uzoh and Oliver (2006) used a composite approach (as described by Wykoff 1990) for height increment modeling for managed even-aged stands of ponderosa pine (*Pinus ponderosa* Dougl.). They used permanent-plot measurements from the western US. Random effects for locations, plots and trees were used in the model, and an autoregressive covariance structure was used to model the repeated measurements. They used a linear form of mixed model after transforming the periodic annual height increment to a logarithmic scale. They selected the following model:

$$E[\ln(\text{PAIH})] = b_0 + b_1 \ln(\text{DBH}) + b_2(\text{DBH})^2 + b_3\text{SIM} + b_4\text{SL}[\cos(\text{ASP})] + b_5\text{ELEVA} + b_6\text{SDI} + b_7\text{BAL} + \hat{h}_l + e_{j(l)} + e_{ik(jl)} \quad (21)$$

where

$E[\ln(\text{PAIH})]$ = expectation of the natural logarithm of 5-year periodic annual height increment (m)

DBH = initial diameter at breast height (cm)

SIM = Meyer's site index (m)

SL = average slope for the stand (%)

ASP = average aspect for the stand (radians)

ELEVA = elevation for the stand (m)

SDI = stand density index (trees/ha)

BAL = basal area in larger trees (m^2/ha) divided by dbh of subject tree

\hat{h}_l = fixed-effect of the l th location

$e_{j(l)}$ = random error for plot j within location l (assumed to have mean 0 and variance σ_l^2)

$e_{ik(jl)}$ = random error for measurement k on tree i within plot j and location l (assumed to have mean 0 and variance c^2)

b_i = parameter estimates

They found that the height increment model displayed a unimodal and positively skewed shape for tree growth process, which was as per expectation. Site index (SIM) was found to have more effect on height growth than other variables.

Vázquez and Pereira (2005) used a linear mixed model to explain variation in cork weight of cork oak (*Quercus suber* L.) from Portugal. They decomposed the random effects in regional, plot and tree-level that would allow better estimates for fixed effects.

Budhathoki et al. (2005) also used a linear mixed model approach for preliminary analysis of basal area growth of shortleaf pine using a portion of data reported by Lynch et al. (1999). The natural logarithm of annual basal area growth was modeled with a dataset consisting of plot measurements at three points in time. In the analysis, they used a compound symmetry covariance structure and plot random-effects. Their fixed-effects parameter estimates were similar in magnitude to those reported by Lynch et al. (1999), but the standard errors were reduced as a result of using mixed model.

The linear mixed model approach discussed above for Lynch et al. (2005), Uzoh and Oliver (2005), Budhathoki et al. (2005), and others was partly for computational

convenience that would be available in SAS PROC MIXED, although the inherent nature of such responses would be nonlinear. With the increasing access to user-friendly software such as S-Plus `n1me` library (Pinheiro and Bates 2000), nonlinear mixed-modeling is gaining popularity. Hall and Bailey (2001) and Fang and Bailey (2001) as indicated above also used nonlinear mixed modeling.

A nonlinear mixed-effects model was developed for height growth for *Eucalyptus* plantations in Brazil by Calegario et al. (2005). They used the `n1me` library available in S-Plus to model the dominant height. They modeled the dominant height as a logistic function of age and with plot random-effects. The fitted logistic function was:

$$HD_{ij} = \frac{\phi_{1i}}{1 + \exp[-(a_{ij} - \phi_{2i}) / \phi_{3i}]} + \varepsilon_{ij} \quad (22)$$

where

HD_{ij} = dominant height (m) for i th plot on time j

a_{ij} = age (years) for i th plot on time j

ε_{ij} = random error; $\varepsilon_{ij} \sim N(0, \sigma^2)$

$$\boldsymbol{\Phi}_i = \begin{bmatrix} \phi_{1i} \\ \phi_{2i} \\ \phi_{3i} \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} b_{1i} \\ b_{2i} \\ b_{3i} \end{bmatrix} = \boldsymbol{\beta} + \mathbf{b}_i \quad (23)$$

$$\mathbf{b}_i \sim N(\mathbf{0}, \boldsymbol{\psi})$$

They used repeated measurements from 115 permanent plots measured three to 10 times between 1992 and 2001. The three parameter logistic function along with plot random-effect was found to fit the data well.

A multilevel nonlinear mixed-effects model was fitted for modeling stand volume growth on loblolly pine, and for modeling four silvicultural treatments by Zhao et al. (2005). They found that vegetation control and fertilization resulted in the largest volume growth. The nonlinear mixed model approach was reported to be useful to handle the unbalanced and incomplete repeated measurements. A similar methodology was used by Jordan et al. (2005) to analyze earlywood and latewood microfibril angle data on loblolly pine. Rose et al. (2006) used a multilevel approach to estimate individual tree survival on loblolly pine using complementary log-log link function. They estimated random-effects at both plot and tree level, and compared effects of four cultural treatments on tree survival.

Overall, if we fit a model for each plot separately, it would often result in over-parameterization, whereas ignoring the grouping by plot we would lose much of the variability. A mixed model provides a balance between these two extremes (Pineiro and Bates 2000, Calegario et al. 2005). This review shows that mixed models have been increasingly used in forest growth and yield modeling. However, no peer-reviewed published work on shortleaf pine growth modeling applying mixed models has yet been done. This provides an opportunity to apply this new technique in shortleaf pine, and to determine whether there is an advantage over the modeling techniques that are currently in use.

Estimation/Computational Methods

Corbeil and Searle (1976) describe the foundation of estimation with restricted or residual maximum likelihood (REML) procedure for mixed modeling. They also developed an algorithm to implement a mixed model. Although mixed modeling literature in statistics existed earlier than Corbeil and Searle (1976) (for example they cite Hartley and Rao (1967)), application in forestry appears to have begun only in late 1980's and early 1990's.

Generalized estimating equations (GEE) were used by Gregoire and Schabenberger (1996b), and Hall and Bailey (2001). In S-Plus `nlme` library, the default estimation method for nonlinear mixed-effects modeling is maximum likelihood (ML), whereas the default method for linear mixed-effects modeling is REML. The library enables to fit mixed models with correlated and/or heterogeneous errors (Pinheiro and Bates 2000).

CHAPTER III

DATA AND DATA MANAGEMENT

This chapter describes the data used for this dissertation. The details on plot establishment, sampling, tree measurements, data management and summary statistics for important variables are presented.

Study Area and Establishment of Plots

The Department of Forestry at Oklahoma State University cooperated with the USDA Forest Service Southern Research Station, and the Ouachita and Ozark National Forests to establish the study plots for shortleaf pine. The objective was to study growth and yield of shortleaf pine in even-aged natural stands. The study was called Study 48. The plots were established from fall 1985 to fall 1987 in eastern Oklahoma and western Arkansas. The study area is presented in Figure 3.1.

To quantify the species distribution within a stand in terms of proportionate coverage of shortleaf pine and other species, basal area was assessed using a 10-factor prism. Only five shortleaf trees that were classified either dominant or codominant as per the definition of Avery and Burkhart (2002) were selected for height and age

determination. This information was used to adopt the design criteria of age-density-site index combinations as given in Table 3.1 for selecting plots (Lynch et al. 1999).

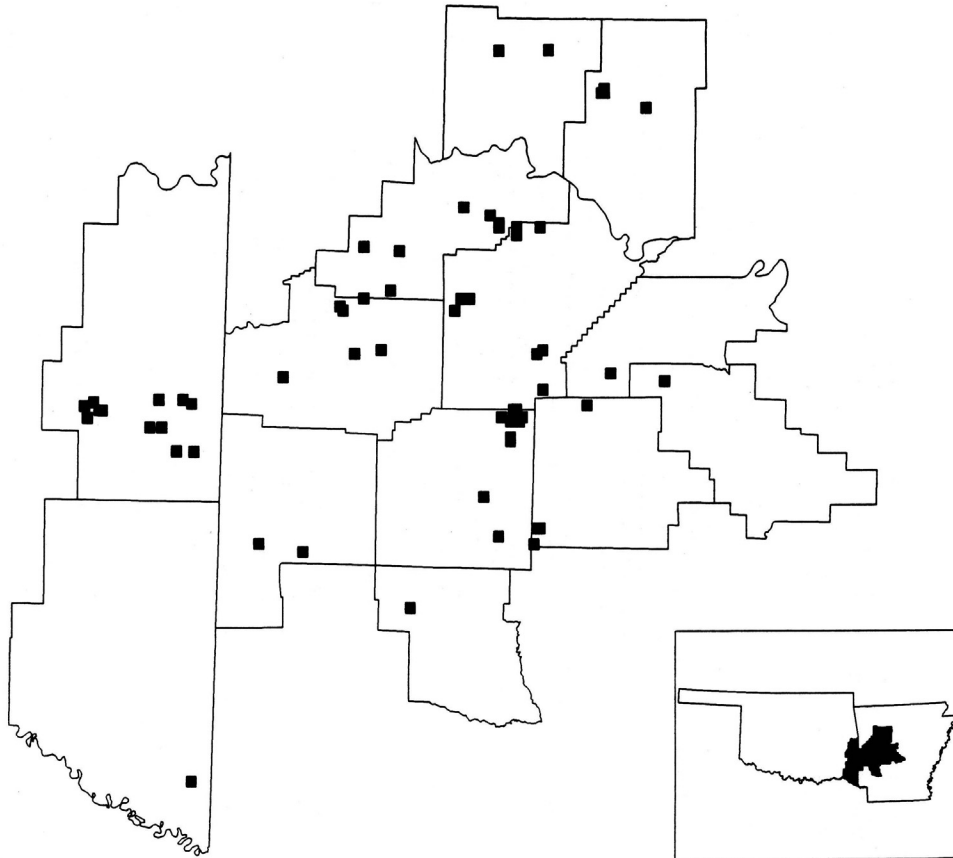


Figure 3.1. Shortleaf pine growth study locations in eastern Oklahoma and western Arkansas (Lynch et al. 1991)

The original plots were established based on the design criteria, and the following stand properties (Rose 1998):

- (1) naturally regenerated stand with at least 70% shortleaf pine basal area with trees having DBH 0.6 inches and larger,

- (2) stand with dominant and codominant trees having maximum age range of 10 years or less,
- (3) stand having site index variation of less than 10 feet,
- (4) even-aged stand with no clumping and no more than two age classes per plot, and
- (5) reasonably free from insect, disease or fire damage and no harvesting in the last five years.

Table 3.1. Study design for naturally regenerated even-aged shortleaf plots installed from 1985 to 1987 with thinning and herbicide treatment in eastern Oklahoma and western Arkansas (Lynch et al. 1999)

Variable	Class range	Class midpoint
Basal area (ft ² /ac)	16-45	30
	46-75	60
	76-105	90
	106-135	120
Site index (ft at base age 50)	<56	50
	56-65	60
	66-75	70
	>75	80
Stand age (yr)	<31	20
	31-50	40
	51-70	60
	71-90	80

Circular fixed-area plots of size 0.2 acres (52.7-foot radius) were established. The plots were surrounded by a 33-foot isolation boundary, which was painted white. A buffer boundary of 33-foot was marked that surrounded each contiguous group of plots. The buffer boundary was painted blue. The study plots were selected using aerial photographs. A typical net plot with two 5-milacre sub-plots is given in Figure 3.2. The 5-milacre sub-plots were used for understory measurements. Plot centers were posted

with an 18-inch orange plastic surveyor's stake or steel reinforcing rod. Each plot number was permanently marked on the center stake. During establishment, each plot was thinned to a predetermined basal area as per the study design. Hardwood control was achieved by using a chemical herbicide. The buffer boundary also received the same treatments for thinning and hardwood control as the study plot. A total of 192 plots were planned with three plots in each combination of design variables (basal area, site index and age), but only 191 were established. Of the established plots, eight were either damaged due to windstorms or not thinned as per the design, so only 183 plots and 7740 trees were available for analysis for the first growth period from Study 48 (Lynch et al. 1999).

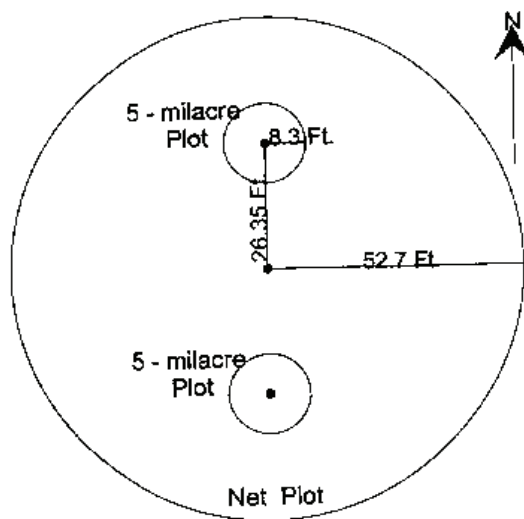


Figure 3.2. A representation of 0.2 acres measurement plot with two 5-milacre plots (USDA Forest Service 1995)

Study 58 was established in 1963–1964 by Frank Freese to study thinning effects in naturally occurring shortleaf pine stands. Study 58 was modified in 1988 to conform to the same design criteria used for Study 48 (Lynch et al. 1999). Additional plots from Study 58 were also available for this dissertation. Only the relevant data for the two growth periods from 1985 to 1996 have been used, although more data exist prior to 1985 for Study 58. Although Freese’s study was established with 35 plots, only 25 plots remained after 1987. Study 58 has only 664 trees available for analysis. Therefore, measurements for 208 plots and 8404 trees from the two studies through three time points are available for the purpose of this dissertation.

Measurements/Observations

Initial tree measurements were taken at the time of plot establishment. The azimuth and distance from plot center were recorded for each tree in a plot. Each tree was also labeled with a tree number and diameter measurement mark. Measurements of diameter at breast height and assignment of tree crown class were made on each tree. Tree diameter was measured at breast height (4.5 feet) to nearest 0.1-inch for each tree in the plot. Tree status and crown characteristics were also measured on all the trees. Crown class was recorded as (1) Dominant, (2) Codominant, (3) Intermediate, or (4) Suppressed following the definitions of Avery and Burkhart (2002). Trees with crowns extending above the general level of the crowns and being in a position to receive full light from above and partially from sides were classified as dominants. The trees classified as codominants were at the general crown level, which were in a position to receive full

light from above but little from the sides. Trees that were shorter than dominants and codominants receiving little direct light from above and none from the sides were classified as intermediates. Trees classified as suppressed were completely below the general level of the crown receiving no direct light. Total tree height, crown height and crown length were recorded to nearest foot for representative dominant and codominant trees. Ring count data of the representative dominant and codominant trees was recorded to determine tree age. The ring count data were collected by means of increment cores taken at 4.0 feet height on the uphill side of the tree for the sample trees. Five years were added to annual ring count to determine the total age in years.

At least four trees were measured for total height and crown height from plots having the biggest trees, whereas a maximum of 39 height measurement trees were selected from plots having the smallest trees.

Repeated Measurements

Second measurements (first remeasurements) were made between fall 1990 and fall 1992. Third measurements (second remeasurements) were taken between fall 1995 and fall 1997. Remeasurements were made at an interval of either four or five years. For remeasurements, the plots were located with the help of aerial photos, which indicated the location of the plots. If the plot center stake was missing during remeasurements, it was located by taping from previously known tree numbers. If a tree number was not found during remeasurement, it was located with the help of azimuth and distance records of previous measurements. If previously recorded diameter appeared to be incorrect, it

was checked with the help of increment core taken just below breast height. For remeasurement, only new height-sample trees were used for increment cores since ring count data for previous height-sample trees were already available, although not all height-sample trees were bored for the ring count.

A significant change was made in status codes for third measurements. The details on field instructions including the changes can be found in the following field guides.

- (1) Field Instructions FS-SO-4106-48 Growth and Yield of Thinned Natural Shortleaf Stands in the Ozark and Ouachita National Forests (August 1991) for first and second measurements
- (2) Field Instructions Shortleaf Growth and Yield Study FS-SO-4106-48 (July 1995) for third measurement

Each tree was assigned a two-digit status code for third measurement. However, the status code for the first two measurements was only a single digit. Status codes for Study 58 were slightly different. The details of status codes for both the studies are provided in Appendix I.

No damage code was recorded for first measurement in Study 48, because all healthy and sound trees were selected. On the other hand, damage code was available from all the three measurements for Study 58, because it was a thinning study previously established in 1963–64. A slightly different status code, i.e. double-digit code was used for Study 58 (Appendix I)

Data Management

This section presents the issues relating to quality check, corrections and intermediate calculations carried out for preliminary analysis as well as for preparing data for more formal analysis.

There were some plots with the same identification number across studies 48 and 58, e.g. plots 9 and 10; so the duplicate plot numbers were resolved by adding 48000 and 58000 respectively to make the plot identification unique so that random plot-effects could be estimated properly.

The records with DBH growth over the period outside the range of 0 and 2 inches were listed. Similarly the trees with total height growth outside the range of 0 and 9 feet were also listed. Original plot sheets for such cases were thoroughly checked. Necessary corrections and adjustments were made where appropriate. Data entry errors were corrected accordingly, e.g. wrongly entered status codes, DBH values etc. Moreover, the cases in which DBH data were correctly entered in computer from plot sheets, but growth values appearing unusually large were compared to the expected maximum growth of an open grown tree (Smith et al. 1992). Almost all of such cases were found reasonable based on the analysis. Thus, such cases were left unaltered as there was no basis to make adjustments otherwise.

Missing and suspicious values for variables such as DBH, height and status code were checked in plot sheets, and corresponding corrections were made as appropriate. Diameter at breast height was estimated for those trees which were in the plot but were missed for measurements. The estimate was used to calculate stand basal area only, but

such a record would not be included in modeling. Interpolation was used if middle DBH value was missing. Similarly, constant rate of growth was assumed for two periods when either first or third DBH was missing unless the third missing observation was due to mortality. In general, it was likely that first missing measurement could be due to the field worker missing the tree, but the third missing measurement could be due to either mortality or observation error by the recorders. For example, the values of DBH for a tree in Study 58 for different years are as follows:

<u>Year</u>	<u>1987</u>	<u>1992</u>	<u>1997</u>
DBH (in)	12.4 (D ₁)	13.0 (D ₂)	? (D ₃)

When the tree was alive in 1997, then the DBH was estimated by assuming the same growth rate during the two periods. That is:

$$\text{Estimate of } D_3 = D_2 + \text{growth rate} \times D_2$$

$$\text{where growth rate (for period 1)} = \left(\frac{D_2 - D_1}{D_1} \right)$$

$$\text{So, estimate of } D_3 = 13.0 + \left(\frac{13.0 - 12.4}{12.4} \right) \times 13.0$$

$$= 13.63 \text{ (which rounds to 13.6 with 0.1 inch rounding)}$$

Site index was calculated based on all three measurements using the method of Graney and Burkhart (1973). Since their site index equation could not be solved analytically, a numerical procedure called the secant method as described in Gerald and Wheatley (1994) was used. Trees with possible height measurement errors were not

included in site index calculation. The ratio of quadratic mean diameter to individual tree DBH was calculated. Stand level variables such as basal area per acre and stand age were computed from individual tree measurements. Stand basal area was calculated by summing the individual tree basal areas in the plot, and by adjusting to a per acre value. Stand age was assumed to be the average age of the representative dominant and codominant trees in the plot.

Study 48 data for the first two measurements (first period) were available as a SAS ver. 7 data file from the work of Lynch et al. (1999), which was converted to a SAS ver. 9 data file. Moreover, if DBH correction from increment core data during the third measurement was available, the master file was updated by running the programs again. A separate Excel file for the third measurement for Study 48 was available, and it was thoroughly checked for such errors as discussed before. Since the file was only for the third measurement, records for dead trees were not entered in the Excel file. But it was necessary to keep track of what happened to a specific tree in previous measurements. So, an appropriate death code, including appropriate code to distinguish if a tree was a height sample tree or not, was entered for such records. This would allow the users to keep track of individual trees during the fourth measurement. For Study 48, growth calculations for the first period were performed by Hitch (1994) and Lynch et al. (1999). These same computations were done for the second period. The previous SAS programs were revised for the second period, and new programs were also written to handle the added complexity due to additional growth period. For example, site index calculation is now based on all three measurements. Growth calculations were also completed for both the

periods for Study 58. Thus, a master file was created with both the studies and two growth periods.

Conversion of azimuth and distance data to Cartesian coordinate system was achieved in three steps:

- (1) Transformation (translation) of azimuth angle by subtracting 90^0 from azimuth in quadrants II, III and IV, but adding 270^0 to azimuth in quadrant I. The resulting angle is similar to that used for polar coordinates but measured in a clockwise direction
- (2) Subtracting the value of step (1) above from 360^0 to obtain the correct angle for the polar coordinates system
- (3) Calculation of Cartesian (x-y) coordinates from polar coordinates by using the following formulas:

$$\text{x-coordinate} = r \cos(\theta) \quad (r = \text{distance}, \theta = \text{angle for polar coordinate})$$

$$\text{y-coordinate} = r \sin(\theta).$$

This was modified to our situation to calculate in SAS using appropriate variable names. SAS uses radian measure but azimuth was in degrees. The Cartesian coordinates will be used for analysis of possible spatial dependence of trees within a plot.

Plot establishment, measurements and the type of data used in growth and yield modeling from the first two measurements are given in Hitch (1994), Lynch et al. (1999) and Rose (1998). Data from fourth measurements made between fall 2000 and fall 2001 will not be utilized in this project because of significant ice-storm damage.

The data for this project is “approximated real growth series” type in the terminology of Moser and Hall (1969). The results from analysis of these data should apply to naturally-regenerated shortleaf pine stands of eastern Oklahoma and western Arkansas. The summary statistics of stand or plot variables as well as tree variables are presented in Tables 3.2 and 3.3.

Table 3.2. Summary statistics for 208 plots (Studies 48 and 58 combined) for plot/stand level variables

Variable	Mean	Standard Deviation	Minimum	Maximum
Basal area (ft ² /ac)				
First measurement	92.9	29.1	27.3	129.0
Mid (first)	102.6	30.2	22.5	142.5
Second measurement	112.3	32.2	14.9	156.5
Mid (second)	121.3	34.1	16.1	178.4
Third measurement	130.7	36.8	17.3	200.4
Stand age (yr)				
First measurement	41.8	19.7	18.0	93.0
Second measurement	46.5	19.7	23.0	98.0
Third measurement	51.6	19.7	28.0	103.0
Site index (ft at age 50 yr)	57.4	9.5	39.9	87.4
Quadratic mean diameter, QMD (in)	8.3	3.5	3.3	19.9
Dominant height (ft)	54	19	27	106

Overall, a negative tree growth value is not impossible in calculations due to measurement/observation error or, more importantly, due to loss of some bark thickness, especially with old trees. Therefore, analyses including negative growth as well as setting negative growth to zero will be discussed.

Table 3.3. Summary statistics for 208 plots (Studies 48 and 58 combined) for individual tree variables

Variable	Mean	Standard Deviation	Minimum	Maximum
Diameter at breast height (in)				
First measurement	7.4	3.9	1.1	24.4
Second measurement	8.2	3.9	1.2	25.4
Third measurement	9.1	4.0	1.5	26.6
Tree basal area (ft ²)				
First measurement	0.3825	0.3835	0.0066	3.2472
Second measurement	0.4469	0.4128	0.0079	3.5188
Third measurement	0.5397	0.4526	0.0123	3.8591
Annual average basal area growth (ft ² /tree)				
Overall	0.01378	0.01127	-0.01669	0.09964
First period	0.01303	0.01044	-0.01669	0.07176
Second period	0.01459	0.01205	-0.01396	0.09965
Ratio of QMD to DBH (<i>R</i>)	1.1397	0.4423	0.3559	7.3621
Total height (ft)				
First measurement	57	22	10	112
Second measurement	61	21	10	113
Third measurement	65	20	13	119

Figures 3.3 and 3.4 provide some idea about how DBH distribution changed slightly over the periods. The mid value of DBH for two measurements was computed for use as an independent variable in modeling tree basal area growth.

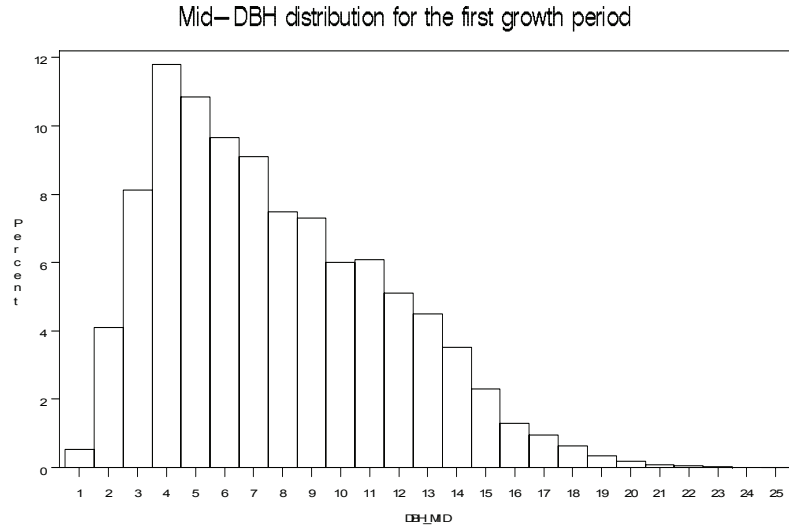


Figure 3.3. Histogram of mid-DBH for the first growth period

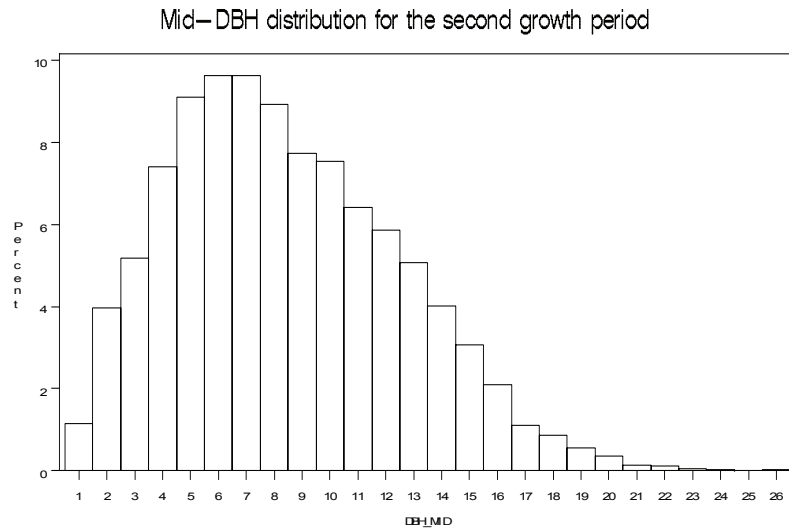


Figure 3.4. Histogram of mid-DBH for the second growth period

CHAPTER IV

METHODS FOR DATA ANALYSIS AND MODELING

This chapter presents some background on statistical methods used in data analysis, and their application to the shortleaf pine growth study data set. Basic statistics (summary statistics and graphical techniques) were used for exploratory analysis. Two major tree attributes were considered for modeling: average annual basal area growth, and total height. For both of these attributes, models were first fitted with a calibration data set, and validation was carried out with a different data set, although the validation data set in this study would not be completely independent of the calibration data set.

Ideally, a model would have good properties both statistically and biologically. On statistical grounds, some models might fit the data well, but they might not necessarily have good biological properties. For example, a high degree polynomial model may fit sample data very well, but such a model often does not have much biological significance. Thus an empirical modeling approach can be useful, but interpreting individual coefficients in the resulting model may be difficult. However, overall data fit and good predictability of response for a given value of explanatory variable(s) are important, and an empirical model may be able to provide these characteristics.

Since biological growth is considered intrinsically nonlinear (Zeide 1993), a nonlinear statistical model is naturally a good choice for forest modelers. Thus, attempts were made to fit different nonlinear models for shortleaf pine growth data. This approach is consistent with typical forest growth and yield modeling practices (e.g. Wykoff 1990, Vanclay 1994). A variety of nonlinear regression models for shortleaf pine growth variables have been reported by Murphy (1982, 1986), Lynch et al. (1991, 1999), Murphy et al. (1992), and Lynch and Murphy (1995). None of the shortleaf pine growth researchers took the random plot-effects and possible dependence of within-plot errors into account. These factors could influence the resulting parameter estimates and their precision, although tools of mixed modeling that could handle such an analysis had already appeared in forestry literature since late 1980's (see Review of Literature). Therefore, this project carried out analyses that would make improvements on the two major components of shortleaf pine growth models developed by Lynch et al. (1999): basal area growth and total height models. Random-effects for plots (or stands assuming a plot typically represents a stand) were incorporated in a modeling framework to obtain better parameter estimates. Possible correlated and heterogeneous within-plot errors were also evaluated. Furthermore, this modeling exercise utilized additional data resulting from the third measurement on these permanently established plots.

Mixed-Effects Models

As briefly discussed in Chapter II, when all the levels of a factor are present in the data for analysis, then the factor has a fixed-effect. That is, the whole population of the

factor levels is represented. The analyst is basically interested in comparing the effects of the factors on the response variable for those levels included in the study only. A model containing only fixed effects is called a fixed-effects model. On the other hand, a factor may have a large number of levels in some cases, and the direct testing or representation of all the levels are not often economical, practical, or necessary. So only a random sample of the factor levels is tested or studied. Thus inference from such a study can be applied to the population of levels from which the sample was drawn. Such effects are called random-effects. For example, use of random plot effects in a model would allow valid generalization of the results to a population of plots from which the study plots were sampled. A model in which all the effects are random is called a random-effects model. Variances associated with random-effects are known as variance components (Graybill 1976, Snedecor and Cochran 1980). Data analysts are interested in the variance components rather than in the individual random coefficients themselves.

Models in which some effects are fixed and the others are random are called mixed models. Each random-effect is associated with a particular fixed-effect, but all the fixed-effects in the model may or may not have associated random-effects. In the mixed-effects approach also, the objectives of fixed effects testing are as in the fixed-effects modeling. On the other hand, one would test if the variance components associated with the random effects equal zero rather than testing or interpreting the individual random coefficients themselves, although best linear unbiased predictors (BLUP) of the random-effects are also estimated or in the commonly used terminology, predicted. In fact, random-effects help in accounting for sources of variation, which enables accurate estimation and testing of fixed-effects (Tao 2002). Mixed-models are also defined to be

models with errors at more than one level. So these models are also called multi-level models or hierarchical models. In mixed models, effects can be either nested or crossed. Depending on how a response is modeled as a function of explanatory variables, typically mixed-effects models are classified as:

(a) Linear mixed models

(b) Nonlinear mixed models

Linear Mixed Model

A basic linear mixed-effects model, as described by authors such as Laird and Ware (1982), Gregoire et al. (1995), and Pinheiro and Bates (2000) is given below. In matrix notation the model can be written as:

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i, i=1,2, \dots, M, \text{ i.e. } M = \text{number of groups} \quad (24)$$

where,

\mathbf{y}_i is the n_i -dimensional response vector for the i^{th} group (the number of observations could be different in different groups)

$\boldsymbol{\beta}$ is the p -dimensional parameter vector for fixed-effects

\mathbf{b}_i is the q -dimensional group specific vector of parameters for random-effects (scalar if only one random effect)

\mathbf{X}_i is the $n_i \times p$ design matrix for fixed-effects variables

\mathbf{Z}_i is the $n_i \times q$ design matrix for random-effect variables (random-effect of i^{th} group) that are often chosen as a subset of fixed-effects variables

$\boldsymbol{\varepsilon}_i$ is the n_i -dimensional within-group (or group-specific) error vector

It is assumed that $\mathbf{b}_i \sim N(\mathbf{0}, \boldsymbol{\Psi})$, and $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$, where $\boldsymbol{\Psi}$ is a variance-covariance matrix for random effects and \mathbf{I} is an identity matrix with random within-group error variance σ^2 . Moreover, it is also assumed that $\text{cov}(\mathbf{b}_i, \boldsymbol{\varepsilon}_i) = \mathbf{0}$, i.e. random-effects and within-group errors are assumed to be independent for different groups, and to be independent of each other in the same group. The columns of \mathbf{Z}_i are often a subset of the columns of \mathbf{X}_i .

Nonlinear Mixed Model

A basic nonlinear mixed-effects model can be written as follows (Pineiro and Bates 2000).

$$y_{ij} = f(\boldsymbol{\varphi}_{ij}, \mathbf{v}_{ij}) + \varepsilon_{ij}, \quad i=1, \dots, M, j=1, \dots, n_i, \quad (25)$$

where

M = the number of groups (e.g. number of plots if data are grouped by plots)

n_i = the number of observations on the i th group (e.g. number of trees in a plot at one time point)

f = a general, real-valued, differentiable function, which is nonlinear in at least one component of the group-specific parameter vector

$\boldsymbol{\varphi}_{ij}$ = a group-specific parameter vector

\mathbf{v}_{ij} = a covariate vector

ε_{ij} = a normally distributed within-group error term independent of \mathbf{b}_i (random effects), i.e. $\varepsilon_{ij} \sim N(0, \sigma^2)$

Moreover, $\boldsymbol{\varphi}_{ij}$ is modeled as

$$\boldsymbol{\varphi}_{ij} = \mathbf{A}_{ij}\boldsymbol{\beta} + \mathbf{B}_{ij}\mathbf{b}_i$$

where

$$\mathbf{b}_i \sim N(0, \boldsymbol{\Psi})$$

$\boldsymbol{\beta}$ = p -dimensional vector of fixed-effects parameters

\mathbf{b}_i = q -dimensional vector of random-effects parameters associated with the i th group, and $\mathbf{b}_i \sim N(\mathbf{0}, \boldsymbol{\Psi})$. Note that $\boldsymbol{\Psi}$ is a scalar, if there is only one random-effect.

\mathbf{A}_{ij} = design matrix for fixed-effects parameters

\mathbf{B}_{ij} = design matrix for random-effects parameters

It is assumed that random-effects and within-group errors are independent. All of the above for the nonlinear mixed model can be summarized in matrix-notation as:

$$\mathbf{y}_i = \mathbf{f}_i(\boldsymbol{\varphi}_i, \mathbf{v}_i) + \boldsymbol{\varepsilon}_i, i=1, \dots, M \quad (26)$$

where $\boldsymbol{\varphi}_i$ is modeled as $\boldsymbol{\varphi}_i = \mathbf{A}_i\boldsymbol{\beta} + \mathbf{B}_i\mathbf{b}_i$, and

$$\mathbf{y}_i = \begin{bmatrix} y_{i1} \\ \vdots \\ y_{in_i} \end{bmatrix}, \quad \mathbf{f}_i(\boldsymbol{\varphi}_i, \mathbf{v}_i) = \begin{bmatrix} f(\phi_{i1}, v_{i1}) \\ \vdots \\ f(\phi_{in_i}, v_{in_i}) \end{bmatrix}, \quad \boldsymbol{\varphi}_i = \begin{bmatrix} \phi_{i1} \\ \vdots \\ \phi_{in_i} \end{bmatrix}, \quad \mathbf{v}_i = \begin{bmatrix} v_{i1} \\ \vdots \\ v_{in_i} \end{bmatrix}$$

$$\boldsymbol{\varepsilon}_i = \begin{bmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{in_i} \end{bmatrix}, \quad \mathbf{A}_i = \begin{bmatrix} \mathbf{A}_{i1} \\ \vdots \\ \mathbf{A}_{in_i} \end{bmatrix}, \quad \mathbf{B}_i = \begin{bmatrix} \mathbf{B}_{i1} \\ \vdots \\ \mathbf{B}_{in_i} \end{bmatrix}$$

For this work, only nonlinear mixed-effects models are used, and are therefore the focus of all discussions.

Extended Mixed Models

The mixed models considered in this section have two aspects, viz. extension to more than one level of random-effects, and incorporation of correlated and/or heterogeneous errors.

The basic models can be extended to more than one level. A typical two-level nonlinear mixed model can be written as given below.

$$y_{ijk} = f(\boldsymbol{\varphi}_{ijk}, \boldsymbol{\nu}_{ijk}) + \boldsymbol{\varepsilon}_{ijk}, \quad i=1, \dots, M, j=1, \dots, M_i, k=1, \dots, n_{ij} \quad (27)$$

where

$$\boldsymbol{\varphi}_{ijk} = \boldsymbol{A}_{ijk} \boldsymbol{\beta} + \boldsymbol{B}_{i,jk} \boldsymbol{b}_i + \boldsymbol{B}_{ijk} \boldsymbol{b}_{ij}$$

$$\boldsymbol{b}_i \sim N(\mathbf{0}, \boldsymbol{\Psi}_1), \quad \boldsymbol{b}_{ij} \sim N(\mathbf{0}, \boldsymbol{\Psi}_2)$$

Mixed models with any level of random-effects can be further extended to incorporate correlated and/or heterogeneous errors, although the complexity increases quickly since a regular mixed-effects model in itself is computationally very intensive. It is not always necessary to assume that the errors are normally and independently distributed. Thus in the single-level models above, we can assume that $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \sigma^2 \boldsymbol{A}_i)$, where \boldsymbol{A}_i are positive-definite matrices. The within-group errors are again assumed to be independent for different groups indexed by i , and independent of random effects \boldsymbol{b}_i . The structure of \boldsymbol{A}_i can accommodate correlated and/or heterogeneous within-group error structures. That is, the variance-covariance structure of the within-group errors can be partitioned into two components: a correlation structure and a variance structure. The error distribution would extend to $\boldsymbol{\varepsilon}_{ik} \sim N(\mathbf{0}, \sigma^2 \boldsymbol{A}_{ik})$ for two-level model (Pineiro and

Bates 2000). Details on correlated and/or heterogeneous errors are presented in a separate section called Error Modeling.

Estimation and Computing

Two major parameter estimation methods are maximum likelihood (ML) and restricted/residual maximum likelihood (REML). The theory of estimation for these methods can be found in Pinheiro and Bates (2000).

The basic mixed-effects models described above can be fit by popular software packages such as SAS and S-Plus. A linear mixed model or a linearized form of nonlinear model can be fit using either SAS PROC MIXED or S-Plus `lme` command. A nonlinear mixed model can be fit by PROC NLMIXED or S-Plus `n1me` command. Both `lme` and `n1me` commands in S-Plus and PROC MIXED have options for both ML and REML methods, whereas PROC NLMIXED has only ML option available since the REML method for nonlinear mixed model is a very complicated estimation procedure. Therefore, S-Plus `n1me` and PROC NLMIXED have ML as default estimation method, whereas S-Plus `lme` and PROC MIXED have REML method as default method (Pinheiro and Bates 2000, Tao 2002).

The extended linear mixed models can be fit by both PROC MIXED and S-Plus `lme`. However S-Plus `n1me` could accommodate more complicated error structures for correlated and heterogeneous errors, but PROC NLMIXED can fit only the basic nonlinear mixed-effects model assuming independently and normally distributed errors. S-Plus `gls` and `gnls` commands can be used to fit equivalent basic as well as extended

linear and nonlinear models without random-effects using generalized least-squares methods. Overall, S-Plus has more facilities for handling the extended models, especially for nonlinear mixed models (Pinheiro and Bates 2000, Tao 2002).

Since intrinsically nonlinear mixed models are considered in this dissertation, typical syntax and command structure for S-Plus and SAS procedures are given below.

S-Plus `nlme` Syntax (Pinheiro and Bates 2000, Venables and Ripley 2002):

```
nlme(model, data, fixed, random, groups, start, method)
```

where `model` is a two sided nonlinear formula separating response on the left and an expression involving parameters and explanatory variables on the right; `data` specifies the name of data file; `fixed` and `random` specify fixed and random components of the model; the `groups` option declares the grouping structure of the data if the data file is not already grouped; the `start` option is used to provide starting values for the estimates unless a self-starting function is used; and `method` can be used to choose between ML and REML methods for parameter estimation. The output from this command can be stored in an appropriate file that could be specified as: `outfile<-nlme(.)`, where `outfile` is a name given to an output file.

If correlated and heterogeneous errors were to be modeled, then the syntax would expand to:

```
nlme(model, data, fixed, random, groups, start, method,  
correlation, weights)
```

where `correlation` specifies correlation structure (temporal, spatial or both); and `weights` allows specification of a variance function that accounts for variance heterogeneity in within-group errors. S-Plus `gnls` command can be used to fit a similar

nonlinear model without random-effects, but with or without correlated and/or heterogeneous errors (Pinehiro and Bates 2000).

SAS `proc nlmixed` Syntax (Tao 2002):

```
proc nlmixed options (e.g. data file);  
  parms parameters and starting values;  
  model dependent variable ~ distribution;  
  random random-effects ~ distribution subject=variable;  
  estimate 'label' expression;  
  predict expression;  
  program statements (if any);  
  title ''; (global option)  
run;
```

where `parms` is used to specify parameters in the model and their starting values; `model` specifies the distribution of the response variable given explanatory variables; the `random` statement declares what random-effects are included with a specific distribution, and it also specifies the level of random-effects with `subject` option; `estimate` allows computation of estimates of the parameters if needed; `predict` can be used to calculate fitted or predicted values from the model for an expression of all observations.

Model Comparison Criteria

The following statistics have been reported to be useful measures to evaluate goodness-of-fit of different mixed models (Gregoire et al. 1995, Pinheiro and Bates 2000, Tao 2002).

1. Akaike Information Criterion (AIC)

$$\text{AIC} = -2 (\text{Log-likelihood}) + 2 (\text{No. of parameters}) \quad (28)$$

AIC is similar to the Cp-statistic and adjusted R^2 in ordinary regression modeling in the sense that AIC takes the number of parameters into account (Gregoire et al. 1995, Draper and Smith 1998).

2. Bayesian Information Criterion (BIC)

$$\text{BIC} = -2 (\text{Log-likelihood}) + (\text{No. of parameters}) \times \log(N) \quad (29)$$

where N = total number of observations used to fit the model

BIC is also called Schwarz's Bayesian Criterion (SBC).

The S-Plus `n1me` library treats variance components as parameters for computing AIC and BIC. However, residual degrees of freedom are calculated by considering group-specific random coefficients as parameters (Pinheiro and Bates 2000). Models with smaller values for these criteria than other competing models are considered better in fitting the data. One could also use $-2(\text{residual log-likelihood})$ to compare the models, and again a model with smaller value for this criterion is better than other models with higher values for the same. However, one can always make this statistic smaller by adding extra parameters to the model similar to increasing ordinary R^2 in regular

regression models. Equivalently, a likelihood ratio test (LRT) can also be used to compare two nested mixed models. The LRT is based on Chi-squared distribution under the null hypothesis. Nested models are such that the parameters in one model are a subset (special case) of parameters in the other model. Thus either AIC or BIC, or both of them, can be used (Pinheiro and Bates 2000, Tao 2002), but AIC appears to be the most commonly used criterion.

Error Modeling

One can extend the basic linear and nonlinear mixed-effects models by accounting for the dependence and heterogeneity of within-group errors, which are likely to be observed in practice.

Correlated Errors

Basically, there are two types of correlation for within-group errors.

1. Spatial correlation: Trees could be correlated due to their spatial distribution within a plot.
2. Temporal correlation: Repeated measurements on the same tree over different time points might be correlated.

For the purpose of basal area growth analysis, three repeated measurements are reduced to only two growth values. Since the growth is calculated as a first difference in the terminology of time series (Diggle 1990, Chatfield 1989), part of the temporal correlation is already taken into account, although some temporal correlation may still exist. With only two time (period) points, there will not be enough data for temporal

modeling. However, we would be able to make evaluation of spatial correlation between trees as coordinates for each tree are available.

Spatial Correlation Structure for Within-Group Errors

It is logical to explore whether the tree measurements within a plot are correlated since the trees are physically (spatially) together as per the concept of spatial variation (Cressie 1993, Ripley 1981, Webster and Oliver 1990). Thus isotropic spatial correlation could be estimated in which correlations depend only on distance, i.e. distance between trees. Spatial correlation is often studied with the help of semi-variogram. The semi-variogram is a standard measure of spatial variability, which is a function of the distance between observations (Cressie 1993). Before defining semi-variogram, one needs to specify or select a distance measure such as *Euclidean* and *Manhattan*. Denote the observation taken at position $\mathbf{x} = (x_1, \dots, x_r)^T$ by ε_x , so the *Euclidean* distance would

be calculated as $d(\varepsilon_x, \varepsilon_y) = \sqrt{\sum_{i=1}^r (x_i - y_i)^2}$. With an appropriate distance measure, the

semi-variogram can be defined as (Pinheiro and Bates 2000):

$$\gamma[d(\varepsilon_x, \varepsilon_y)] = \frac{1}{2} \text{var}(\varepsilon_x - \varepsilon_y) \quad (30)$$

Alternatively, it can be written in ordinary notation as (Webster and Oliver 1990, Tao 2002):

$$\gamma(h) = \frac{1}{2m} \sum (y_i - y_j)^2 \quad (31)$$

where m is the number of pairs of observations with distance h between observations (y 's). Three terms are important in a semi-variogram.

1. Nugget: It is the intercept, often denoted by σ_1^2 . This is a variance component to account for abrupt changes over a relatively small distance. The nugget effect can only vary between 0 and 1.
2. Sill: It is defined as the value of the semi-variogram at the plateau reached for larger h . The corresponding variance component notation is $\sigma^2 + \sigma_1^2$.
3. Range: Range is defined as the value of h at which the semi-variogram reaches the sill. This corresponds to the parameter ρ in the correlation function. The range only takes positive values.

According to Pinheiro and Bates (2000), semi-variograms increase monotonically with distance, and their values range from 0 to 1. This corresponds to non-negative correlation functions that decrease monotonically with distance. That is:

$$\gamma(s, \rho) = 1 - h(s, \rho) \quad (32)$$

where $\gamma(s, \rho)$ is semi-variogram function with distance s and correlation coefficient ρ

$h(s, \rho)$ is correlation function with the same parameters

The following isotropic (invariant with respect to direction) variogram models for spatial correlation structures can be used from S-Plus `nlme` library (Pinheiro and Bates 2000).

1. Linear: $\gamma(s, \rho) = 1 - (1 - s / \rho)I(s < \rho)$ (33)

where $I(s < \rho)$ is a binary variable: 1 when $s < \rho$, 0 otherwise

2. Exponential: $\gamma(s, \rho) = 1 - \exp(-s / \rho)$ (34)

3. Gaussian: $\gamma(s, \rho) = 1 - \exp[-(s / \rho)^2]$ (35)

The functions in S-Plus `nlme` library that can be used for modeling spatially correlated errors are: `corLin` (linear), `corExp` (exponential), and `corGaus` (Gaussian), among others. These options were used for data analysis since Cartesian coordinates for each tree were available.

Heterogeneous Errors

It is also likely that the within-group errors have heterogeneous variances, which could be taken into account by expressing variance as a function of covariates. A framework for variance function is summarized by Pinheiro and Bates (2000) following the work of Davidian and Giltinan (1995) as:

$$\text{var}(\varepsilon_{ij} | \mathbf{b}_i) = \sigma^2 g^2(\mu_{ij}, \mathbf{v}_{ij}, \boldsymbol{\delta}), i = 1, \dots, M, \text{ and } j = 1, \dots, n_i \quad (36)$$

where

$$\mu_{ij} = E(y_{ij} | \mathbf{b}_i)$$

\mathbf{v}_{ij} = vector of covariates for variance

$\boldsymbol{\delta}$ = vector of variance parameters, and

$g(\cdot)$ is the variance function assumed continuous in $\boldsymbol{\delta}$.

The following variance functions among others in S-Plus `nlme` library can be used for modeling heterogeneous errors.

`varFixed`: fixed variance, variance function with no parameters and a single variance covariate – within-group variance is known up to a proportionality constant

`varPower`: power of covariate, can model the cases with increasing or decreasing variance with the absolute value of the variance covariate

`varExp`: exponential function of the covariate variable

Due to flexibility, the `varPower` is a common variance function for modeling monotonic heterogeneity. In this type, error variance is modeled as $\text{var}(\varepsilon_{ij}) = \sigma^2 |v_{ij}|^{2\delta}$ with one covariate, where variance function is $g(v_{ij}, \delta) = |v_{ij}|^\delta$ (Pinheiro and Bates 2000). A covariate can be chosen from a list of explanatory variables in the model that appears to explain the variance heterogeneity better than others. Tree basal area was selected for this project as a covariate in variance function for the reasons to be explained in Chapter V.

Model Checking

Although models are selected based on criteria such as AIC, the selected models are further evaluated based on analysis of residuals. The plots of residuals against predicted values, and also against individual explanatory variables could reveal any poor fit of the data to the model or presence of outliers, if any. However possible outliers were identified at data management stage also.

An Approach to Mixed Model Development

Typically, a mixed modeling exercise is started with individual regression for each group, e.g. for each plot. If the parameter estimates appear to be different for different groups (plots), then it is considered that mixed modeling could help in which the regression parameters would be random (random effects). The parameter estimates from such a model are often called shrinkage estimates. On the other extreme, one could fit an overall model by ignoring groups. In fact, mixed model would provide a compromise (shrinkage) between these two extremes. This model fitting strategy has been called an “inside-out” approach by Pinheiro and Bates (2000). However, we could not fit individual regression for each plot with our data, because some of the explanatory variables in the model of Lynch et al. (1999) are plot-level variables such as stand age and dominant height. Thus mixed models were fitted directly without taking an inside-out approach.

The models reported in Lynch et al. (1999) were taken as a starting point for developing nonlinear mixed-effects models for average annual basal area growth and total height for this project. The following approach was taken to select a model.

(a) Split the whole data set into calibration set and validation set. The calibration set and validation set have two-thirds and one-third of the total number of study plots, respectively. The plots were randomly assigned to one of the datasets, but the plots were properly distributed over the design criteria.

(b) Start with the calibration data set.

- Estimate the parameters for Lynch et al. (1999) models again with additional data, which would possibly provide starting values for further model fitting work. Equivalently, the estimates in Lynch et al. (1999) itself could be another set of starting values.
- Fit nonlinear mixed-effects model by adding random-effects for plots under the assumption of independently and normally distributed errors. Start with one random coefficient.
- Try to add more random effect coefficients, and drop some coefficients if needed following similar strategy as in stepwise regression as described in Draper and Smith (1998).
- Attempt to fit two-level models also.
- Impose some structure to within-group errors first starting with spatial correlation, and then with heterogeneous errors. Depending on their significance, one could probably incorporate both correlated and heterogeneous errors.

- Compare competing models with criteria like AIC, and possibly use a likelihood ratio test to compare two nested models.
- (c) Perform residuals analysis for the fitted models using validation data set, and see if the fitted models appear adequate.
- (d) Obtain final estimates with the combined data set.
- (e) If the results show poor model fit, then further analysis might be necessary such as checking data distribution over calibration set and validation set, although a random sampling of plots based on design criteria was performed.
- (f) Perform model checking for the final candidate models by means of error analysis such as plots of residuals vs. predicted values or residuals vs. explanatory variables. If there appear to be any outliers, we would be able to check the raw data by identifying the observation, although substantial data quality check was performed at data management stage.

Exploratory Data Analysis

Descriptive statistics and graphical techniques were used to develop preliminary ideas about the data. Summary statistics for different variables separately for calibration and validation data set would help provide insights on the data.

Fitting the Growth Models

Two major components of shortleaf pine growth and yield models were fitted, viz. annual basal area growth model and DBH-total height relationship model.

Both the models were fitted first with a calibration data set and then validated with a different data set. The results of this analysis are presented in Chapter V (Results). Based on the results from these two data sets, overall estimation was carried out with the complete dataset. The results from the combined data set are presented and discussed in Chapter VI (Discussion and Conclusions).

Basal Area Growth Model

The basal area growth model developed by Lynch et al. (1999) was called BAG Model I, which was modified to accommodate plot-specific random coefficients. This was called BAG Model II. Moreover, errors can be modeled to account for possible correlation within a plot, and also to account for likely heterogeneity in errors. The following model for shortleaf pine basal area growth with plot random effects for some regression coefficients would be desirable for the reasons described before. A nonlinear mixed model for annual basal area growth in which random-effect (b_7) for only one associated fixed-effect (β_7) can be written in regular notation as:

$$y_{ij} = \frac{\beta_1 B_{ij}^{\beta_2} - (\beta_1 B_{ij} / B^{1-\beta_2})}{1 + \exp(\beta_3 + \beta_4 B_{si} + \beta_5 A_i + \beta_6 R_{ij} + (\beta_7 + b_{7i}) B_{ij})} + \varepsilon_{ij} \quad (37)$$

where

y_{ij} = annual basal area growth (ft²/year) of tree j in plot i

B_{ij} = basal area (ft²) of tree j in plot i

A_i = stand age for plot i

R_{ij} = ratio of quadratic mean stand diameter to the DBH of tree j in plot i

B_{si} = stand basal area (ft²/ac) for plot i

$B = 7.068384$ ft² (the maximum expected basal area for a shortleaf pine in managed stands)

β_1, \dots, β_7 = fixed-effects parameters

b_{7i} = random parameter specific to i^{th} plot associated with mid-tree basal area fixed-effect coefficient β_7 that appears in the modifier (denominator)

ε_{ij} = within-plot error (i.e. residual for tree j in plot i)

It is assumed that $b_{7i} \sim N(0, \sigma_b^2)$, $\varepsilon_{ij} \sim N(0, \sigma^2)$, and $\text{cov}(b_{7i}, \varepsilon_{ij}) = 0$. It is clear from above that there are as many random-effect parameters as the number of plots along with fixed-effect parameters in the model. So the parameter estimation is a very computationally demanding exercise. The expected value for B (7.068384 ft²) was taken as reported by Hitch (1994), which was derived by assuming maximum DBH of 36 inches.

A more complicated model given below, which was called BAG Model III, was also fitted.

$$y_{ij} = \frac{(\beta_1 + b_{1i})B_{ij}^{\beta_2} - ((\beta_1 + b_{1i})B_{ij} / B^{1-\beta_2})}{1 + \exp(\beta_3 + \beta_4 B_{si} + \beta_5 A_i + \beta_6 R_{ij} + (\beta_7 + b_{7i})B_{ij})} + \varepsilon_{ij} \quad (38)$$

where b_{li} is random parameter specific to i^{th} plot associated with mid-tree basal area fixed-effect coefficient β_1 that appears twice in the potential (numerator), and other parameters are described before. Alternatively in matrix notation, $\mathbf{b}_i = \begin{bmatrix} b_{li} \\ b_{7i} \end{bmatrix}$, $E(\mathbf{b}_i) = \mathbf{0}$,

$$\text{and } \text{var}(\mathbf{b}_i) = \boldsymbol{\Psi} = \begin{bmatrix} \text{var}(b_{li}) & \text{cov}(b_{li}, b_{7i}) \\ \text{cov}(b_{li}, b_{7i}) & \text{var}(b_{7i}) \end{bmatrix}.$$

It would be logical to try to fit another random-effect associated with fixed-effect for R (ratio of quadratic mean diameter to DBH), but it may be difficult to estimate these coefficients due to a possible convergence problem since R and tree basal area are correlated variables. This is because tree basal area is directly calculated using DBH, and DBH is the denominator in R . This model was called BAG Model IV.

$$y_{ij} = \frac{(\beta_1 + b_{li})B_{ij}^{\beta_2} - ((\beta_1 + b_{li})B_{ij} / B^{1-\beta_2})}{1 + \exp(\beta_3 + \beta_4 B_{si} + \beta_5 A_i + (\beta_6 + b_{6i})R_{ij} + (\beta_7 + b_{7i})B_{ij})} + \varepsilon_{ij} \quad (39)$$

where b_{6i} is random parameter specific to i^{th} plot associated with R fixed-effect coefficient β_6 that appears in the modifier, and other parameters are as described before.

In matrix notation, the random parameters and their variance components can be written as follows.

$$\mathbf{b}_i = \begin{bmatrix} b_{li} \\ b_{6i} \\ b_{7i} \end{bmatrix}, \text{ and } \text{var}(\mathbf{b}_i) = \boldsymbol{\Psi} = \begin{bmatrix} \text{var}(b_{li}) & & \\ \text{cov}(b_{li}, b_{6i}) & \text{var}(b_{6i}) & \\ \text{cov}(b_{li}, b_{7i}) & \text{cov}(b_{6i}, b_{7i}) & \text{var}(b_{7i}) \end{bmatrix}.$$

The variance-covariance matrix for random-effects can be specified to have a particular pattern such as identity, diagonal or compound symmetry with S-Plus `pdMat` classes. This model would have random parameters three times the number of plots.

The dataset also contains some negative values for individual tree basal area growth, which are not expected in terms of actual biological growth. However, such values are possible in practical data analysis because of slight error in repeated measurements or loss of some bark thickness over the years. So the modeling exercise considered both the scenarios: with negative values for growth; and negative values set to zero.

Total Height Model

A total height prediction model of Lynch et al. (1999) was called Height Model I, and it was modified to include random-effects for plots, which is given below. This new model was called Height Model II.

$$(H_{ij} - 4.5) = \beta_0 (H_{D_i} - 4.5)^{\beta_1} \exp(-(\beta_2 + b_{2i}) D_{ij}^{-\beta_3}) + \epsilon_{ij} \quad (40)$$

where

H_{ij} = total height (ft) of tree j in plot i

D_{ij} = DBH (in) of tree j in plot i

H_{D_i} = dominant height for plot i as per Graney and Burkhart (1973)

$\beta_0, \beta_1, \beta_2, \beta_3$ = fixed-effects parameters

b_{2i} = random parameter, associated with β_2 (DBH), specific to i^{th} plot

ε_{ij} = within-plot error (residual for tree j in plot i)

It is again assumed that $b_{2i} \sim N(0, \sigma_b^2)$, $\varepsilon_{ij} \sim N(0, \sigma^2)$, and $\text{cov}(b_{2i}, \varepsilon_{ij}) = 0$. We would be mostly interested in an estimate of $\text{var}(b_{2i})$, i.e. $\hat{\sigma}_b^2$, a variance component describing the spread of the random coefficients, rather than in the individual coefficients themselves. This model also makes it possible to use calibration techniques to estimate random parameter b_2 using measurements from a particular forest of interest. Although total height can be modeled as an explicit function of tree age (Curtis 1967, Lappi and Bailey 1988, Meng et al. 1997), it can also be modeled using dominant height and DBH as predictors where dominant height is a function of tree age and site index (Lynch and Murphy 1995, Lynch et al. 1999). Since the dominant height is a plot level variable, plot random effects was fitted only for fixed effect coefficient associated with DBH, a tree level variable. Although we now have three repeated measurements for dominant height, the dominant height calculation is based on site index value averaging over all the measurements into account. Thus a random-effect associated with fixed coefficient for dominant height was not included in the model.

Fitting the Extended Models

The basic basal area growth model was also extended to model possible spatially correlated and heterogeneous errors. The extended models were fitted using appropriate correlation and variance functions in S-Plus. Their significance was judged based on the model evaluation criteria as described before.

The dataset could be grouped in two different ways: by plots only; and by plots within periods for basal area growth data. Modeling spatial correlation structure would require grouping by both periods and plots (two-level model) for the reasons to be described in the final paragraph of this section. For plot random-effects we needed to group by plots only, but if two-level random effects, i.e. periods and plots within periods, were desired then we needed to group by periods, and plots within periods. That is, this grouping structure would provide random coefficients for plots separately within each period, which might or might not have practical significance. For example, the two-level form for BAG Model II can be written as:

$$y_{ijk} = \frac{\beta_1 B_{ijk}^{\beta_2} - (\beta_1 B_{ijk} / B^{1-\beta_2})}{1 + \exp(\beta_3 + \beta_4 B_{sik} + \beta_5 A_{ik} + \beta_6 R_{ijk} + (\beta_7 + b_{7ik}) B_{ijk})} + \varepsilon_{ijk} \quad (41)$$

where $k=1,2$; and the variables and parameters are just extended with another index k for period. This model will be called BAG Model V. These models were fitted with and without correlated and heterogeneous errors.

The dataset structure for total height modeling is slightly different from basal area growth modeling. That is, data were indexed by time points for total height dataset instead of growth period since total height itself was an analysis variable unlike basal area growth. The two-level form for height model can be written as:

$$(H_{ijk} - 4.5) = \beta_0 (H_{D_{ki}} - 4.5)^{\beta_1} \exp(-(\beta_2 + b_{2ik}) D_{ijk}^{-\beta_3}) + \varepsilon_{ijk} \quad (42)$$

where $k = 1,2,3$; and the variables and parameters are just extended with another index k for time points. This model was called Height Model III.

The spatial correlation functions would work only in two-level model when data are grouped by both plots and periods together, i.e. by plots within periods for basal area growth. This is because the coordinates for each tree are the same in both the periods or in all three time points; and tree-to-tree distance computation would result in zero distance for the same tree, if grouped by plots only. However, variance functions could be used in either single or two-level model. All the fitted models for annual basal area growth are summarized in Table 4.1.

Table 4.1 List of fitted basal area growth models

S. No.	Model	Description
Fitted basic BAG model and models with plot random-effects		
1.	BAG Model I	Annual basal area growth model of Lynch et al. (1999)
2.	BAG Model II	BAG Model I + Plot random-effect for b_7
3.	BAG Model III	BAG Model II + Plot random-effect for b_1 also
4.	BAG Model IV	BAG Model III + Plot random-effect for b_6 also
Fitted extended BAG models with period and plot random-effects		
5.	BAG Model V	BAG Model II + Period and plot random-effects for b_7
6.	BAG Model VI	Dropping stand age (β_5) from BAG Model V
Fitted extended BAG models accounting for variance heterogeneity in errors		
7.	BAG Model VII	BAG Model II + Power variance function
8.	BAG Model VIII	BAG Model III + Power variance function
9.	BAG Model IX	BAG Model IV + Power variance function
10.	BAG Model X	BAG Model VI + Power variance function
Fitted extended BAG models with period and plot random-effects accounting for spatial correlation for within-plot errors		
11.	BAG Model XI	BAG Model VI + Linear correlation function
12.	BAG Model XII	BAG Model VI + Exponential correlation function
13.	BAG Model XIII	BAG Model VI + Gaussian correlation function
Fitted extended BAG models with period and plot random-effects accounting for heterogeneous and spatially correlated errors		
14.	BAG Model XIV	BAG Model XI + Power variance function
15.	BAG Model XV	BAG Model XII + Power variance function
16.	BAG Model XVI	BAG Model XIII + Power variance function

Calibration and Validation

The term calibration has a slightly different meaning in forest biometrics literature compared to statistical literature. In statistics, calibration is generally considered an inverse prediction, i.e. prediction of a value of explanatory covariate given a value of response variable (e.g. Neter et al. 1996), whereas in forestry literature calibration means parameter estimation using a dataset that would be validated using another, preferably independent, data set before getting final estimates from the overall combined data set (e.g. Wykoff 1990). The results from calibration and validation data sets will be presented in Chapter V (Results); and those from combined data set are discussed in Chapter VI (Discussion and Conclusions).

The complete data set was split into two subsets: a calibration set with two-thirds, and a validation set with one-third of the total number of plots. The selection procedure used a random selection of plots, but balance was maintained with respect to the design criteria, viz. site index, stand density and stand age.

The number of trees per plot ranged from 4 (plots 48153, 48227 and 48261) to 491 (plot 48120) in calibration data set, and from 4 (plots 48130 and 48258) to 218 (plot 48106) in the validation data set. The summary statistics of data used in fitting the models are presented in Tables 4.2 to 4.5.

Summary Statistics for Calibration Data Set

The summary statistics for plot level variables and individual tree variables for the calibration data set are presented in Tables 4.2 and 4.3, respectively.

Table 4.2. Summary statistics of plot level variables for calibration data set with 139 plots

Variable	Mean	Standard Deviation	Minimum	Maximum
Basal area (ft ² /ac)				
First measurement	94.1	29.9	27.3	129.0
Mid (first)	103.4	31.0	22.5	142.5
Second measurement	112.6	32.9	14.9	156.5
Mid (second)	121.6	34.9	16.1	178.4
Third measurement	130.9	37.7	17.3	200.4
Stand age (yr)				
First measurement	42.1	19.7	18.0	93.0
Second measurement	46.9	19.7	23.0	98.0
Third measurement	51.9	19.8	28.0	103.0
Site index (ft at age 50 yr)	57.2	9.9	41.0	87.4
Quadratic mean diameter (in)	8.4	3.6	3.3	19.5
Dominant height (ft)	54	19	27	106

Table 4.3. Summary statistics of tree level variables for calibration data set with 139 plots

Variable	Mean	Standard Deviation	Minimum	Maximum
Diameter at breast height (in)				
First measurement	7.5	4.0	1.1	24.4
Second measurement	8.2	4.0	1.2	25.4
Third measurement	9.2	4.0	1.5	26.6
Tree basal area (ft ²)				
First measurement	0.3904	0.3929	0.0066	3.2472
Second measurement	0.4554	0.4217	0.0079	3.5188
Third measurement	0.5520	0.4615	0.0123	3.8591
Annual average basal area growth (ft ² /tree)				
Overall	0.01376	0.01120	-0.01669	0.08918
First period	0.01288	0.01040	-0.01669	0.07176
Second period	0.01470	0.01192	-0.01396	0.08918
Ratio of QMD to DBH	1.1446	0.4549	0.4384	7.3621
Total height (ft)				
First measurement	58	22	11	109
Second measurement	62	21	10	113
Third measurement	66	20	13	119

Summary Statistics for Validation Data Set

Similarly, the summary statistics for plot level variables and individual tree variables for the validation data set are presented in Tables 4.4 and 4.5, respectively.

Table 4.4. Summary statistics of plot level variables for validation data set with 69 plots

Variable	Mean	Standard Deviation	Minimum	Maximum
Basal area (ft ² /ac)				
First measurement	90.5	27.5	28.1	128.3
Mid (first)	101.1	28.5	29.2	133.1
Second measurement	111.6	30.9	28.1	147.2
Mid (second)	120.8	32.6	29.7	161.5
Third measurement	130.2	35.0	27.9	177.4
Stand age (yr)				
First measurement	41.1	19.7	19.0	89.0
Second measurement	45.9	19.7	24.0	93.0
Third measurement	51.0	19.6	29.0	98.0
Site index (ft at age 50 yr)	57.8	8.8	39.9	83.8
Quadratic mean diameter (in)	8.3	3.4	3.7	19.9
Dominant height (ft)	53	18	29	105

Table 4.5. Summary statistics of tree level variables for validation data set with 69 plots

Variable	Mean	Standard Deviation	Minimum	Maximum
Diameter at breast height (in)				
First measurement	7.3	3.8	1.1	22.8
Second measurement	8.0	3.8	1.5	23.5
Third measurement	8.9	3.9	1.9	25.2
Tree basal area (ft ²)				
First measurement	0.3673	0.3644	0.0066	2.8352
Second measurement	0.4308	0.3947	0.0123	3.0121
Third measurement	0.5164	0.4345	0.0197	3.4636
Annual average basal area growth (ft ² /tree)				
Overall	0.01382	0.01141	-0.00641	0.09965
First period	0.01330	0.01050	-0.00362	0.06920
Second period	0.01437	0.01228	-0.00641	0.09965
Ratio of QMD to DBH	1.1301	0.4169	0.3559	4.3910
Total height (ft)				
First measurement	56	22	10	112
Second measurement	60	21	14	112
Third measurement	64	20	16	109

CHAPTER V

RESULTS

This chapter presents results obtained from the calibration data set and model diagnostics using a validation data set. The modeling results for annual basal area growth are followed by those from modeling total height and DBH relationships. Overall results and related discussions based on the complete data set for both the basal area growth model and total height model are presented in Chapter VI.

Basal Area Growth Model

Calibration Results

The parameter estimates for BAG Model I using the calibration data set are presented in Table 5.1. These estimates from three measurements are similar to those obtained by Lynch et al. (1999) based on the first two measurements, although the current result is based only on calibration data set.

Table 5.1. Parameter estimates and other associated statistics for BAG Model I for calibration dataset (total degrees of freedom (df) = 10268, residual standard deviation (SD) = 0.006839469, residual df = 10261)

Parameter	Estimate	Standard error	t-value	P-value
β_1	0.111161	0.0066690	16.67	<0.0001
β_2	0.566732	0.0234667	24.12	<0.0001
β_3	-1.958148	0.1808504	-10.83	<0.0001
β_4	0.010741	0.0004282	25.09	<0.0001
β_5	0.022658	0.0011546	19.62	<0.0001
β_6	0.930197	0.0521353	17.84	<0.0001
β_7	-0.884331	0.0509437	-17.36	<0.0001

BAG Model I for annual basal area growth was fitted using S-Plus `gnls` command, which would provide statistics such as AIC that could be used for making comparisons with the mixed models. The `nlme` command uses either ML or REML method, whereas `gnls` uses generalized least-squares. However, there is no direct way of comparing two models with and without random effects using S-Plus. Therefore, an approximate assessment of significance of plot random effects was carried out by comparing BAG Models I and II. The statistics required for such comparison are presented in Table 5.2. An example output for a basic S-Plus `nlme` function is presented in Appendix II.

Table 5.2. Summary statistics for fitted basal area growth models from calibration dataset

BAG Model	AIC	BIC	Log-likelihood	Residual df	Residual SD
I	-73231.70	-73173.81	36623.85	10261	0.00683947
II	-75919.55	-75854.41	37968.77	10124	0.00583212
VI	-76521.73	-76456.60	38269.87	9986	0.00554681
VII	-80884.29	-80811.92	40452.14	10124	0.00910395
VIII	-81431.71	-81344.87	40727.86	10124	0.00889254
X	-81219.71	-81147.34	40619.85	9986	0.00865303
XI	-76504.36	-76432.00	38262.18	9983	0.00554476
XII	-76504.38	-76432.01	38262.19	9983	0.00554474
XIII	-76504.41	-76432.05	38262.21	9983	0.00554475
XIV	-81202.18	-81122.58	40612.09	9983	0.00864828
XV	-81201.96	-81122.36	40611.98	9983	0.00864813
XVI	-81201.99	-81122.39	40611.99	9983	0.00864816

BAG Model II is better than BAG Model I due to substantial reduction in AIC, BIC and residual SD, although degrees of freedom equal to the number of plots are sacrificed for this estimation. This shows that estimation of random-effects for plots improves the fit of a growth model. The summary statistics for BAG Model II are presented in Table 5.3. These estimates are comparable to those for BAG Model I. Further exploration of random-effects was carried out. Several other mixed models for basal area growth as described in Chapter IV were also fitted, and these results are also summarized in Table 5.2.

Table 5.3. Parameter estimates and other associated statistics for BAG Model II for calibration dataset (number of plots = 139, residual df = 10124, and total observations = 10269)

Parameter	Estimate	Standard error	t-value	P-value
β_1	0.127698	0.0094784	13.47	<0.0001
β_2	0.577809	0.0406986	14.20	<0.0001
β_3	-1.296957	0.1775033	-7.31	<0.0001
β_4	0.008657	0.0004271	20.27	<0.0001
β_5	0.009476	0.0014489	6.54	<0.0001
β_6	1.180113	0.0764147	15.44	<0.0001
β_7	-0.653616	0.0908105	-7.20	<0.0001

Variance component estimate for:

$$b_7 (\hat{\sigma}_{b_7}^2) = (0.6641947)^2 = 0.4411546$$

$$\text{Residual } (\hat{\sigma}_e^2) = (0.00583212)^2 = 0.000034014$$

A 95% confidence interval for true variance component standard deviation (SD) for b_7 random-effect is [0.5633475, 0.783095], which indicates variance component SD to be significantly different from zero. No convergence for parameter estimates was achieved for BAG Models III and IV. This might be because R and DBH are highly correlated, and separate fixed effects for these variables are already in the models. Separate random-effects for plots for these, when both b_7 and b_6 are in a single model, could not be estimated. Moreover, b_1 and b_7 random-effects are associated with the same variable, tree basal area.

Extended Models with Period and Plot Random-Effects

There was again a convergence problem for BAG Model V when grouping was done by both growth period and plots with an attempt to estimate the random-effect for b_7 . As the model complexity increased considerably, stand age and stand basal area, both plot-level variables, could not be retained in a single model. Therefore, stand age was dropped from the model, which was called BAG Model VI. Although not containing stand age, this model appears to have better fit statistics than BAG Model II due to period-specific and plot-specific random coefficients for b_7 . That is, this model would estimate random-effects for plots separately for each growth period. Since one is not basically interested in the value of period-specific coefficient, this model may or may not be interesting for practical applications. The parameter estimates and other statistics for BAG Model VI are presented in Table 5.4.

Table 5.4. Fixed-effect parameter estimates and related statistics for BAG Model VI for calibration dataset (number of growth period =2, plots within period = 278, residual df = 9986, and total observations = 10269)

Parameter	Estimate	Standard error	t-value	P-value
β_1	0.166694	0.0151721	10.99	<0.0001
β_2	0.429624	0.0246202	17.45	<0.0001
β_3	-1.191876	0.1613941	-7.38	<0.0001
β_4	0.011323	0.0003818	29.66	<0.0001
β_6	1.690621	0.0587827	28.76	<0.0001
β_7	-0.051761	0.1330071	-0.39	0.6972

Variance components for:

$$b_7 \text{ for period } (\hat{\sigma}_{Tb_7}^2) = (0.1640622)^2 = 0.0269164$$

$$b_7 \text{ for plots within period } (\hat{\sigma}_{T(p)b_7}^2) = (0.54797)^2 = 0.30027112$$

$$\text{Residual } (\hat{\sigma}_e^2) = (0.005546808)^2 = 0.00003076708$$

A 95% confidence interval for the variance component SD of plots within period for b_7 is [0.491575, 0.6108348]. Similarly, a 95% confidence interval for the variance component SD of growth period for b_7 is [0.06121673, 0.4396902]. Both the interval estimations show that the variance components for b_7 random-effect are significantly different from zero. Therefore, the fixed-effect coefficient β_7 is retained in the model, although it is not significant ($P = 0.6972$), because random-effect without an associated fixed-effect is not normally estimated.

Extended Models with Power Variance Function

A scatter plot of residuals vs. tree basal area is presented in Figure 5.1 for BAG Model II. The figure shows that there is some evidence of error variance heterogeneity, i.e. residual variance appears to increase with tree basal area. Therefore, a power variance function (Pinheiro and Bates 2000 p 210) was included in BAG Models VII, VIII, IX and X to see whether such a variance modeling exercise would improve the model fit.

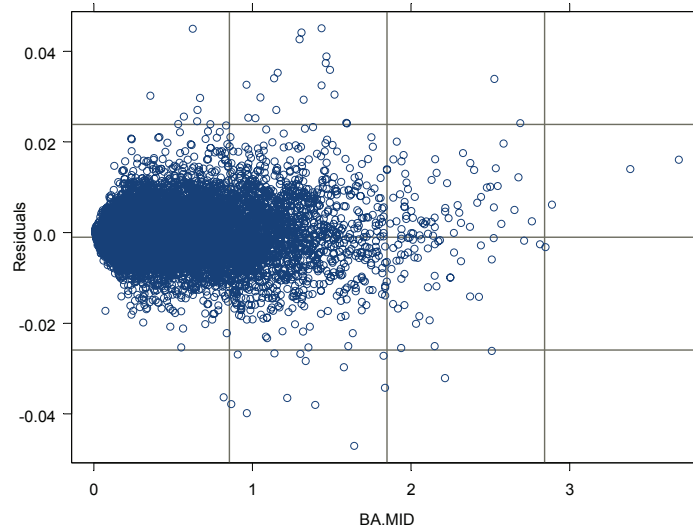


Figure 5.1. Plot of residuals vs. tree basal area for BAG Model II

The statistics in Table 5.2 show that inclusion of variance function as a power of tree basal area substantially improves the fit (BAG Model VII) compared to BAG Model II in which a constant variance for errors is assumed. It is also evident that random-effects for both b_1 and b_7 can be estimated when a power variance function is included (BAG Model VIII). These random-effects could not be estimated together with a constant variance assumption in BAG Model III. However, there was no convergence for BAG Model IX, which shows that b_6 random-effect cannot be estimated once there is random-effect for either b_1 , b_7 or both, even with variance modeling. As discussed earlier, R and tree basal area are functionally related, since both of these variables are calculated using DBH. The parameter estimates for BAG Model VII are presented in Table 5.5.

Table 5.5. Parameter estimates and other associated statistics for BAG Model VII for calibration dataset (number of plots = 139, residual df = 10124, and total observations = 10269)

Parameter	Estimate	Standard error	t-value	P-value
β_1	0.076648	0.0039253	19.53	<0.0001
β_2	0.535239	0.0273428	19.58	<0.0001
β_3	-2.789757	0.1637983	-17.03	<0.0001
β_4	0.017108	0.0005430	31.51	<0.0001
β_5	0.001977	0.0015009	1.32	0.1878
β_6	1.696839	0.0710717	23.88	<0.0001
β_7	-1.100494	0.1738845	-6.33	<0.0001

Variance component estimate for:

$$b_7 (\hat{\sigma}_{b_7}^2) = (1.546647)^2 = 2.392117$$

$$\text{Residual } (\hat{\sigma}_e^2) = (0.00910395)^2 = 0.000082882$$

Power estimate ($\hat{\delta}$) = 0.5676372

A 95% confidence interval for b_7 variance component for plot random-effect is [1.318518, 1.814246]. Similarly, a 95% confidence interval for power coefficient is [0.5524855, 0.5827889]. Table 5.5 shows that fixed-effect of stand age (β_5) is non-significant (P = 0.1878) for which there is no associated random-effect. It appears that stand age could be dropped from this model in presence of variance function. However further evaluation was carried out. Random-effects for b_1 were also included in BAG Model VIII, and the results are presented in Table 5.6.

Table 5.6. Parameter estimates and other associated statistics for BAG Model VIII for calibration dataset (number of plots = 139, residual df = 10124, and total observations = 10269)

Parameter	Estimate	Standard error	t-value	P-value
β_1	0.054198	0.0032559	16.65	<0.0001
β_2	0.444292	0.0429761	10.34	<0.0001
β_3	-2.287280	0.1987150	-11.51	<0.0001
β_4	0.015521	0.0007423	20.91	<0.0001
β_5	0.010391	0.0023658	4.39	<0.0001
β_6	1.449964	0.0738843	19.62	<0.0001
β_7	-3.604463	0.3090490	-11.66	<0.0001

Variance component estimate for:

$$b_1 (\hat{\sigma}_{b_1}^2) = (0.024518736)^2 = 0.000601168$$

$$b_7 (\hat{\sigma}_{b_7}^2) = (2.966580699)^2 = 8.800601$$

$$\text{Residual } (\hat{\sigma}_e^2) = (0.008892544)^2 = 0.000079077$$

Estimated correlation between b_1 and b_7 ($\hat{\rho}_{b_1, b_7}$) = 0.576

Power estimate ($\hat{\delta}$) = 0.5818104

The estimates indicate that variance component for b_7 random-effect is much larger than that for b_1 . It is evident from Table 5.6 that stand age fixed-effect is now significant (P < 0.0001). BAG Model VIII shows that random-effects for both b_1 and b_7 can be estimated once power variance function is modeled. BAG Models VII and VIII have the same number of residual degrees of freedom. This might be because S-Plus `n1me` library would operate on the number of variables to calculate degrees of freedom, and both b_1 and b_7 random-effects are associated with a single variable: tree basal area.

However, it is not completely clear from the text of Pinheiro and Bates (2000). Table 5.7 shows 95% confidence interval estimates for some of the parameters of BAG Model VIII.

Table 5.7. 95% confidence intervals for the parameters in BAG Model VIII

Parameter	Lower estimate	Upper estimate
Variance component for b_1 random-effect	0.01966971	0.03056315
Variance component for b_7 random-effect	2.53054857	3.47774437
Power coefficient	0.566585	0.5970357
Correlation coefficient (b_1, b_7)	0.4301	0.6932

Like BAG Model IV, BAG Model IX also suffered from convergence problems. In fact, tree basal area and R are highly correlated, and random-effects for b_7 (with or without b_1) for tree basal area and for b_6 for R could not be estimated in a single model. BAG Model X was developed by including a power variance function in a basic model with random-effects for period and plot. The summary statistics of parameter estimation and testing are presented in Table 5.8.

Table 5.8. Parameter estimates and other associated statistics for BAG Model X for calibration dataset (number of periods = 2, plots within period = 278, residual df = 9986, and total observations = 10269)

Parameter	Estimate	Standard error	t-value	P-value
β_1	0.093661	0.0047786	19.60	<0.0001
β_2	0.550648	0.0249314	22.09	<0.0001
β_3	-3.601262	0.1563829	-23.03	<0.0001
β_4	0.018748	0.0005195	36.09	<0.0001
β_6	2.197537	0.0663635	33.11	<0.0001
β_7	0.021967	0.2178313	0.10	0.9197

Variance component estimate for:

$$b_7 \text{ for period } (\hat{\sigma}_{Tb_7}^2) = (0.2613539)^2 = 0.068305861$$

$$b_7 \text{ for plots within period } (\hat{\sigma}_{T(p)b_7}^2) = (1.046919)^2 = 1.096039393$$

$$\text{Residual } (\hat{\sigma}_e^2) = (0.008653032)^2 = 0.000074875$$

Power parameter estimate ($\hat{\delta}$) = 0.5526116

Table 5.8 shows that fixed-effect for β_7 (the coefficient associated with tree basal in the modifier) is not significant ($P = 0.9197$) once the data are grouped by period and plots with a power variance function. The variance components could not be tested due non-positive definite approximate variance-covariance matrix.

Extended Models with Spatial Correlation Function for Within-Plot Errors

When data are grouped by period and plot, each tree can be properly located by its coordinates. Therefore, a spatial correlation function accounting for tree-to-tree distance within a plot was estimated. It was expected that such a function would account for spatial dependence between trees within a plot. Linear, exponential and Gaussian correlation functions were estimated for BAG Model VI, which were called BAG Models XI, XII and XIII, respectively. It is clear from Table 5.2 that these three models are not much different from each other, or from BAG Model VI. This indicates that the spatial correlation functions would not add much once plot and period specific random coefficients are included. Moreover, the spatial correlation functions available in S-Plus estimate only average functions over all the groups (combination of periods and plots). Since the plots are quite variable in terms of design criteria, this averaging did not

provide much benefit, because most of the variability was accounted for by plot random-effects. This is a very reasonable result for even-aged plots, although plot-specific spatial correlation estimation could have possibly provided more explanation. However, this could not be done with the available S-Plus functions.

Extended Models with Variance and Spatial Correlation Functions

Table 5.2 also shows that BAG Models XIV, XV and XVI are very similar to each other. Also, they are not much different from BAG Model X. This clearly shows that there is no benefit of inclusion of average spatial correlation function in such models even in presence of variance function. On the other hand, there is evidence of benefit when a variance function is included in a basal area growth model. Therefore, the spatial correlation issue will not be considered for validation.

The calibration results show that there are some good candidate models for further evaluation. In particular, BAG Models II, VI, VII, VIII and X are interesting models for validation with a different dataset.

Validation Results

This section presents results based on residual analysis of candidate basal area growth models using a validation dataset. Some of the models that were fitted with calibration dataset could not be fitted with the validation dataset due to convergence problems. Summary statistics for basal area growth models that could be fitted using the validation dataset are presented in Table 5.9.

Table 5.9. Summary statistics for candidate basal area growth models using validation dataset

BAG Model	AIC	BIC	Log-likelihood	Residual df	Residual SD
II	-40671.98	-40612.63	20344.99	5325	0.00543559
VI	-40767.17	-40707.83	20392.59	5257	0.00527382
VII	-43364.47	-43298.53	21692.23	5325	0.00888164
X	-43331.97	-43266.03	21675.98	5257	0.00865876

Parameters could not be estimated for BAG Model VIII with a validation data set. Only random-effect for b_7 could be estimated, and b_1 and b_7 random-effects could not be estimated together even after dropping stand age from the model. In fact, these coefficients are based on the same variable tree basal area, although one appears in the potential and the other in the modifier part of the model. Only the four promising models shown in Table 5.9 will be validated with data from about one-third of total number of plots, i.e. 69 plots and 5400 records over two growth periods. The residuals for the candidate models are analyzed by means of plots against fitted values, independent variables, or by design criteria.

Residuals vs. Predicted Values

The plots of standardized residuals vs. predicted basal area growth values are presented in Figures 5.2 to 5.5.

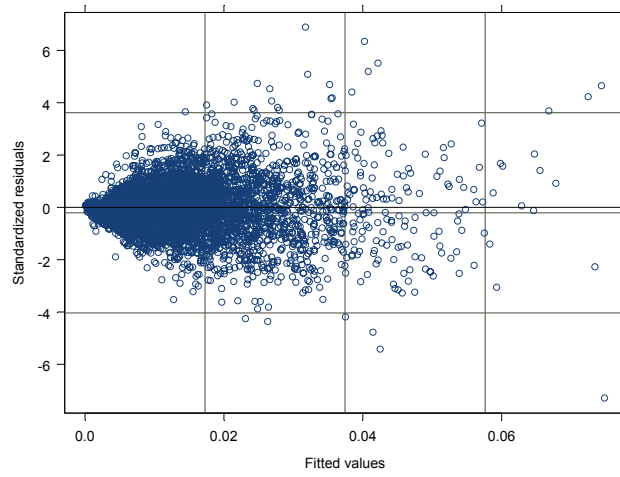


Figure 5.2. Plot of standardized residuals vs. fitted values for BAG Model II

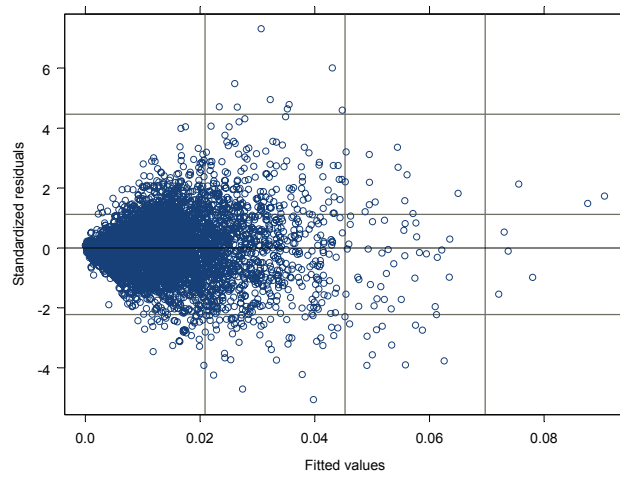


Figure 5.3. Plot of standardized residuals vs. fitted values for BAG Model VI

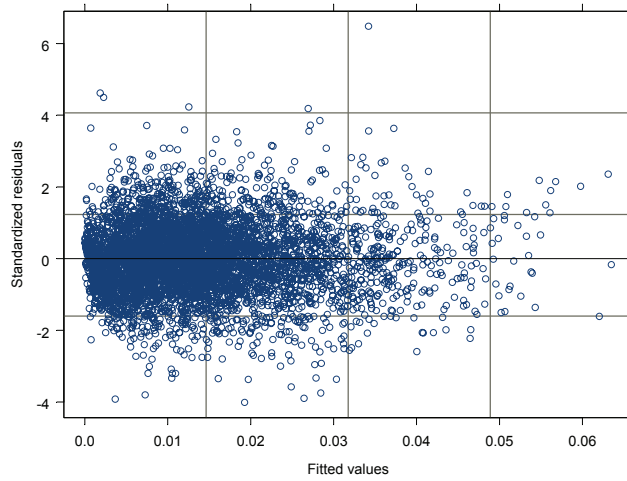


Figure 5.4. Plot of standardized residuals vs. fitted values for BAG Model VII

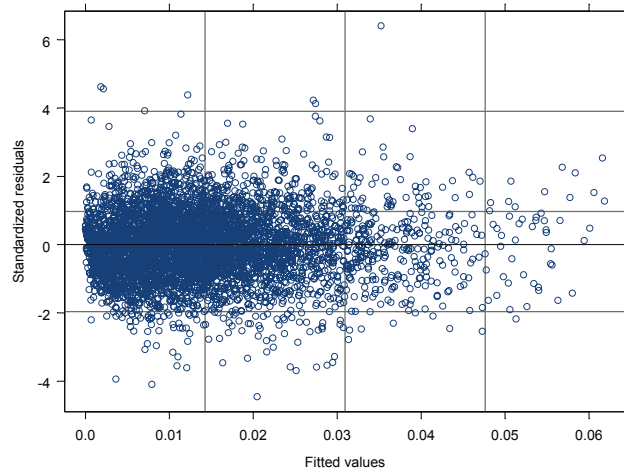


Figure 5.5. Plot of standardized residuals vs. fitted values for BAG Model X

The residual plots show no clear patterns, however Figures 5.2 and 5.3 suggest some increasing residual variation with fitted values. Using a power function for variance modeling has helped in improving the fit as shown by Figures 5.4 and 5.5.

Residuals vs. Tree Basal Area

Standardized residuals were plotted against tree basal area, and they are presented in Figures 5.6 to 5.9.

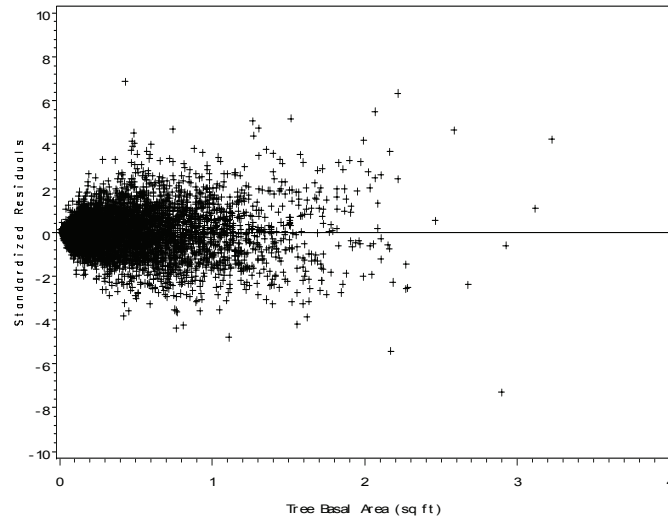


Figure 5.6. Plot of standardized residuals vs. tree basal area for BAG Model II

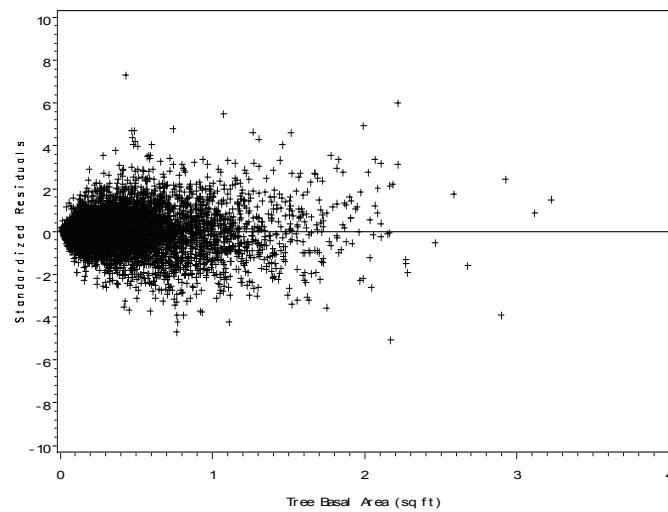


Figure 5.7. Plot of standardized residuals vs. tree basal area for BAG Model VI

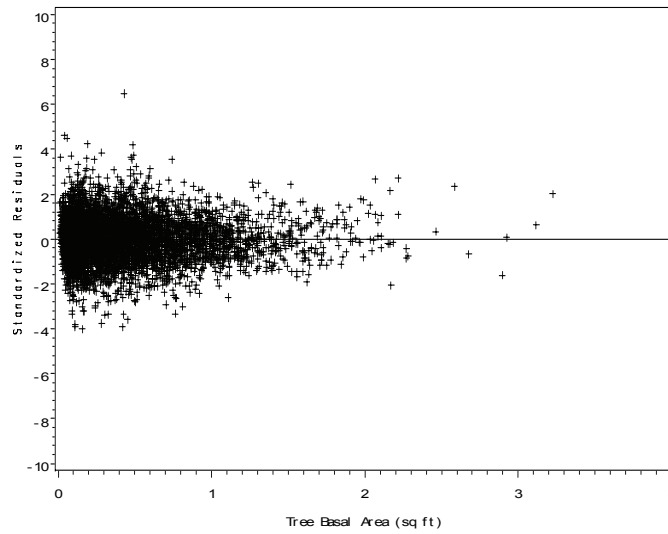


Figure 5.8. Plot of standardized residuals vs. tree basal area for BAG Model VII

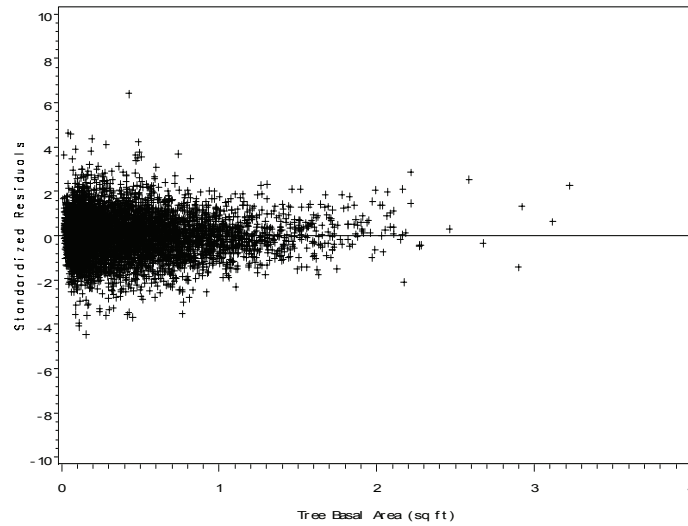


Figure 5.9. Plot of standardized residuals vs. tree basal area for BAG Model X

Based on the residual plots above, it appears that there is no clear evidence of variance heterogeneity. Also, residuals do not show any clear pattern with tree basal area, which indicates that the assumption of independent errors may be justified.

Residuals by Design Criteria

Standardized residuals for the fitted models are displayed using boxplots by design criteria. The boxplots of standardized residuals by stand age class are presented in Figures 5.10 to 5.13.

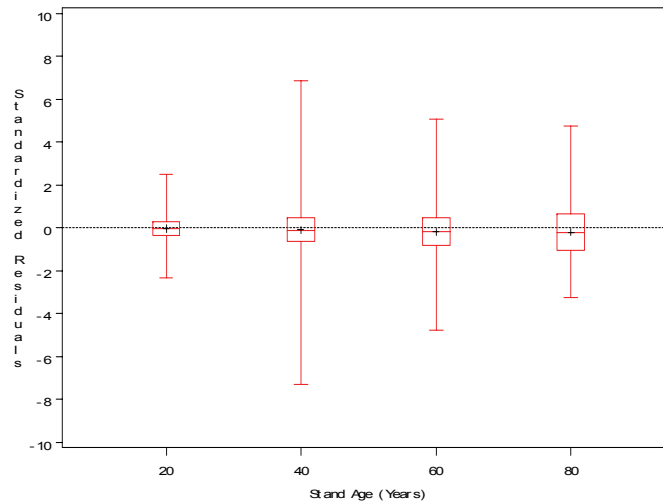


Figure 5.10. Boxplots of standardized residuals by age class for BAG Model II

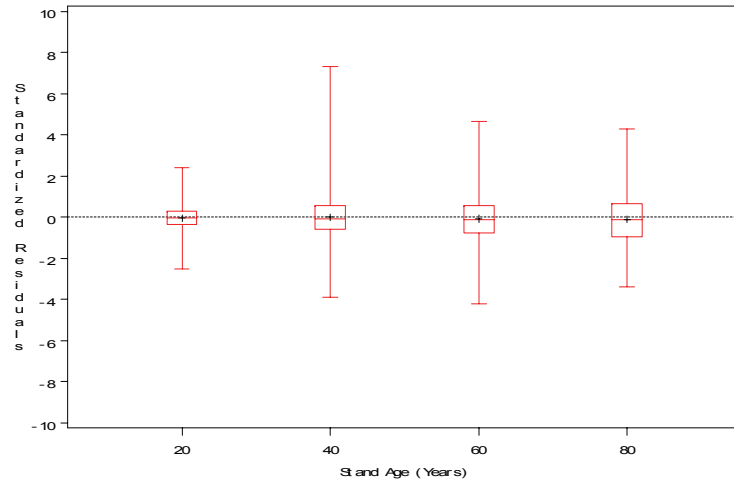


Figure 5.11. Boxplots of standardized residuals by age class for BAG Model VI

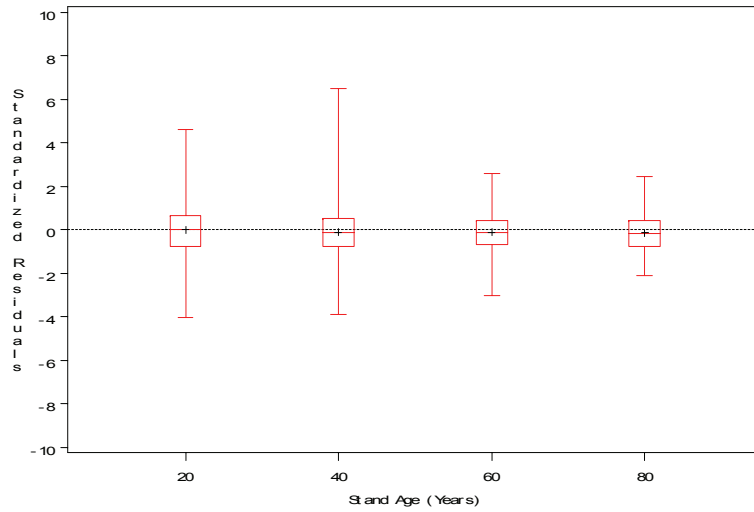


Figure 5.12. Boxplots of standardized residuals by age class for BAG Model VII

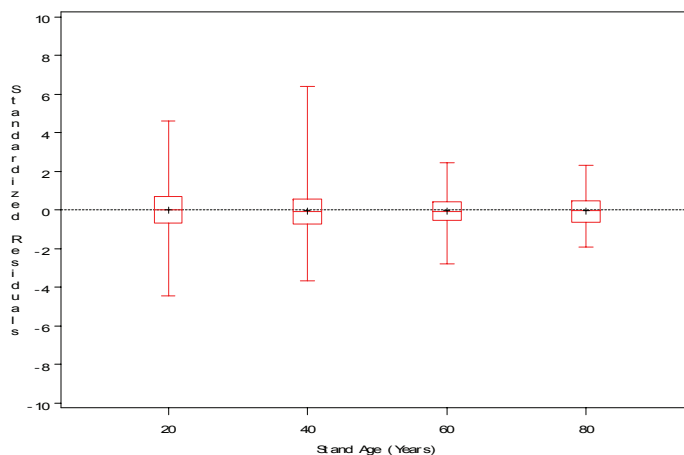


Figure 5.13. Boxplots of standardized residuals by age class for BAG Model X

The 40-year stand age class has shown more variability than other age classes for the first two models, while age classes 20 and 40 have similarity in residual variation when a variance function is included. The BAG Models II and VII appear to slightly over-predict at higher stand ages, particularly at 60 and 80 years.

The boxplots of standardized residuals by site index class (ft at base age 50 years) are presented in Figures 5.14 to 5.17.

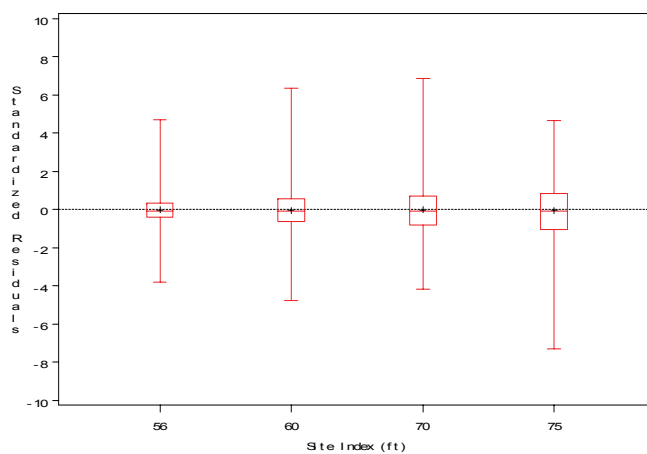


Figure 5.14. Boxplots of standardized residuals by site index class for BAG Model II

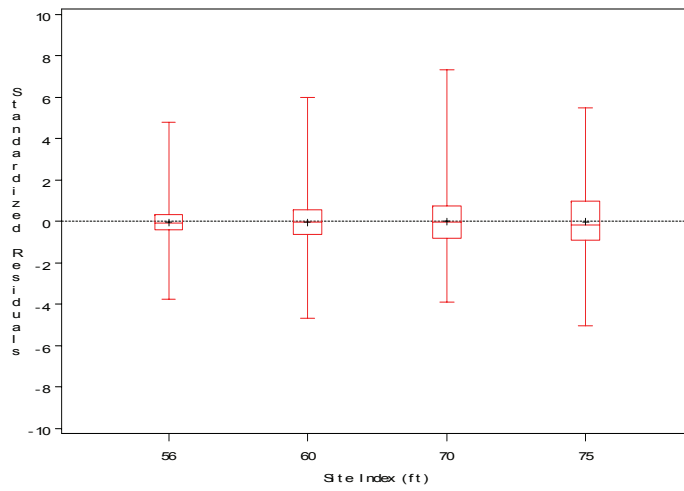


Figure 5.15. Boxplots of standardized residuals by site index class for BAG Model VI

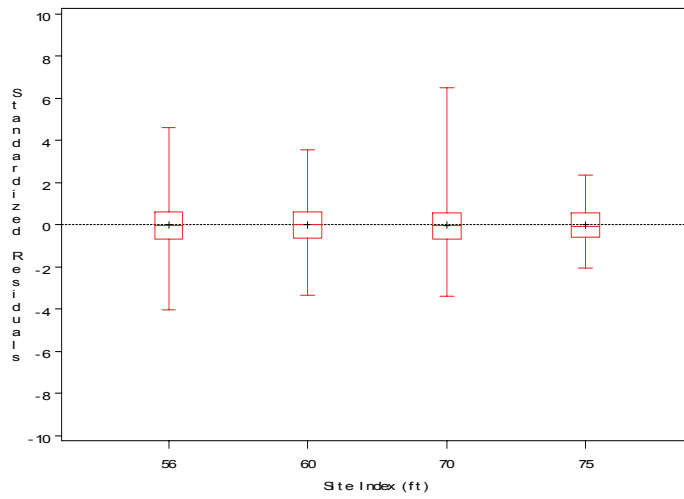


Figure 5.16. Boxplots of standardized residuals by site index class for BAG Model VII

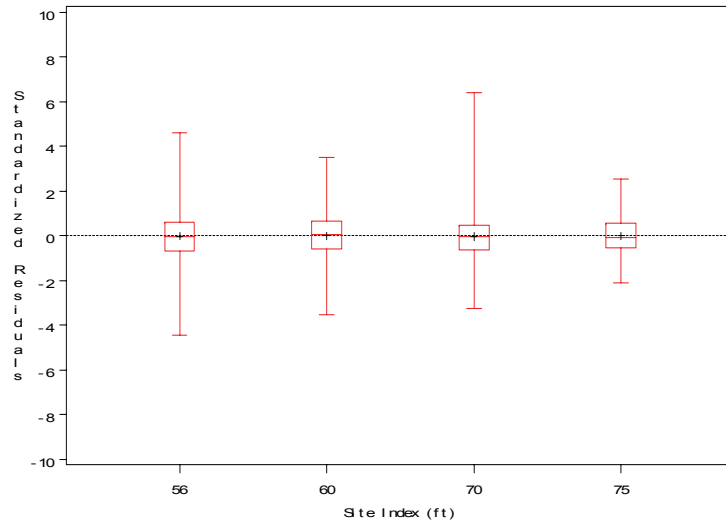


Figure 5.17. Boxplots of standardized residuals by site index class for BAG Model X

Residual variation appears similar over all site index classes for the first two models, whereas site index classes <56 and 70 appear to have higher residual variation than site index classes 60 and >75. All the models appear to make reasonable predictions over all site index classes.

The boxplots of standardized residuals by basal area class are presented in Figures 5.18 to 5.21.

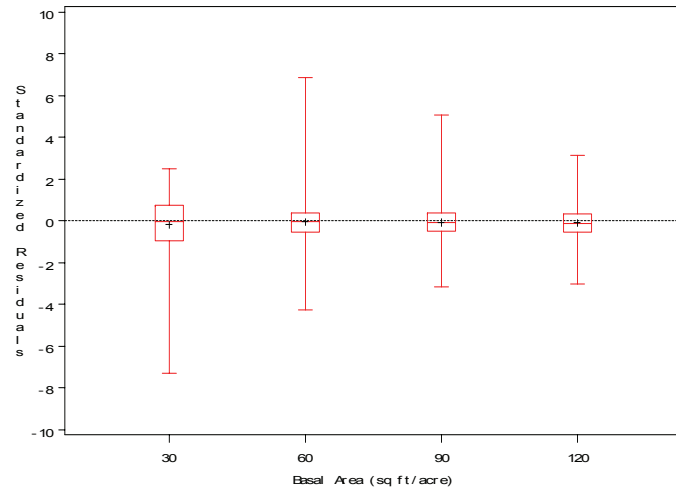


Figure 5.18. Boxplots of standardized residuals by basal area class for BAG Model II

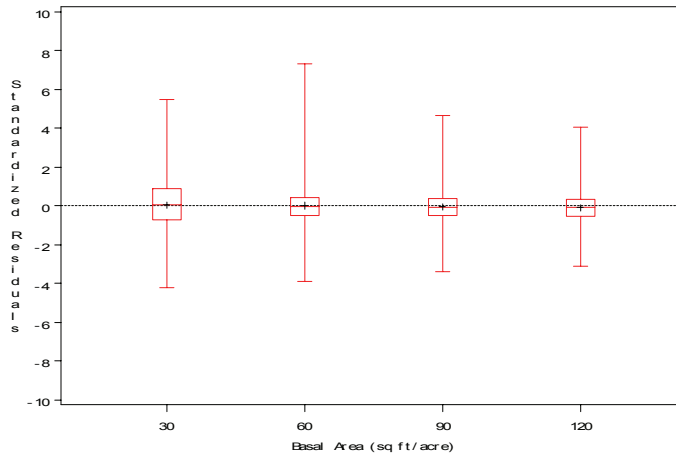


Figure 5.19. Boxplots of standardized residuals by basal area class for BAG Model VI

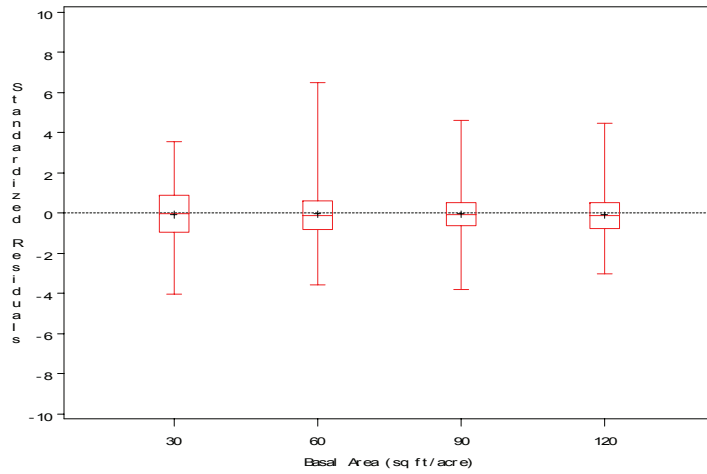


Figure 5.20. Boxplots of standardized residuals by basal area class for BAG Model VII

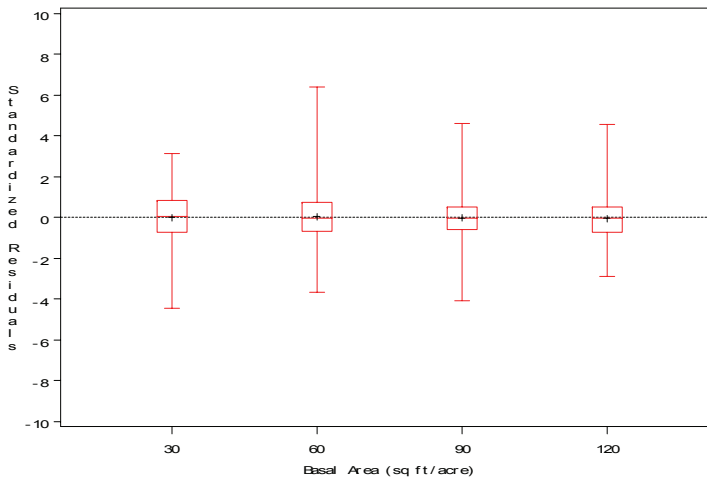


Figure 5.21. Boxplots of standardized residuals by basal area class for BAG Model X

Residual variation appears to decrease with stand basal area classes in basal area growth models II and VI in which no variance function was included. On the other hand, variability appears similar over all basal area classes for basal area growth models VII and X, although basal area class 60 appears to have more variability than other classes in

all the models. These plots show that inclusion of variance function has helped to stabilize the error variance. Overall, all the models appear to make good predictions over all basal area classes analyzed.

In summary of validation results, all the four basal area growth models are comparable to each other despite some differences resulting from variance modeling. Therefore, all the models were refitted using the complete data set. The results from this modeling exercise, and discussion of merits and demerits of these models based on the additional evidence are presented in Chapter VI.

Total Height Model

Calibration Results

The total height models were fitted using SAS PROC NLMIXED, although limited options were available in the procedure. The S-Plus `n1me` command failed to converge even after numerous attempts with different starting values and options for iterations and tolerance. This must be because of a difference in algorithms of PROC NLMIXED and S-Plus `n1me`. However, a basic nonlinear mixed-model for the total height with the assumption of independent and normal errors could be fitted for both calibration and validation data sets using PROC NLMIXED. This section presents results for the calibration data set.

The estimates from PROC NLIN for Height Model I with all three measurements are presented in Table 5.10, which could be compared with the estimates obtained in mixed models, although directly comparable fit statistics will not be available.

Table 5.10. Summary of results for Height Model I without random-effects for plots for calibration dataset (total observations=5968, residual df=5964, residual variance=22.8269)

Parameter	Estimate	Standard error	t-value	P-value
β_0	3.8028	0.1719	22.12	<0.0001
β_1	0.7950	0.0055	144.55	<0.0001
β_2	-2.9393	0.1055	-27.86	<0.0001
β_3	-1.0834	0.0319	-33.96	<0.0001

The parameter estimates and tests from PROC NLMIXED for Height Model II are presented in Table 5.11.

Table 5.11. Summary of results for Height Model II with random-effects for plots for calibration dataset (total observations = 5968, group df = 138)

Parameter	Estimate	Standard error	t-value	P-value
β_0	2.2560	0.1028	21.95	<0.0001
β_1	0.8456	0.0097	87.24	<0.0001
β_2	-5.7145	0.4200	-13.60	<0.0001
β_3	-1.4278	0.0463	-30.86	<0.0001
σ_b^2	1.5208	0.4071	3.74	0.0003
σ_e^2	16.2675	0.3020	53.86	<0.0001

Table 5.11 shows that all the parameters of Height Model II are significant. This model also has random coefficient for total height and DBH relationship, for which the variance component is significant (P = 0.0003). The estimates are similar for both the models with slight differences in the magnitude of the coefficients (Tables 5.10 and 5.11), although PROC NLIN and NLMIXED fit statistics are not directly comparable. Height Model III could not be fitted since PROC NLMIXED could not fit a two-level model.

Validation Results

The residual plots for validation data sets are presented in Figures 5.22 to 5.24. The plots show that there is random scatter of residuals in all the plots, and there is no clear pattern of residuals. Therefore, there is no clear evidence of violation of assumptions of normal and independent errors.

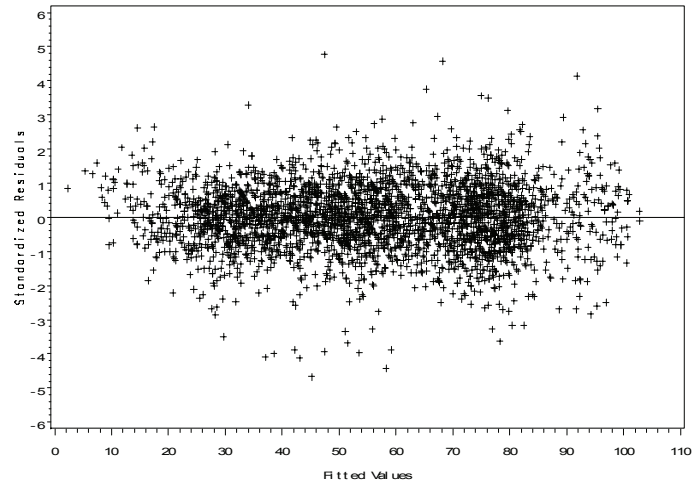


Figure 5.22. Plot of standardized residuals vs. predicted values for total height model

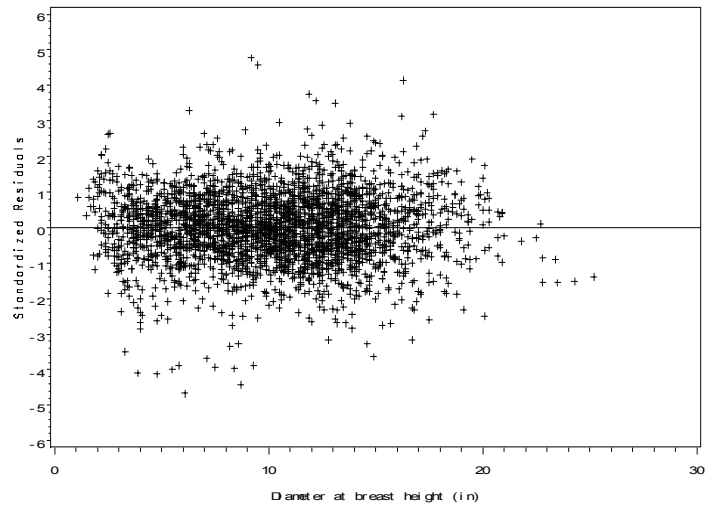


Figure 5.23. Plot of standardized residuals vs. diameter at breast height for total height model

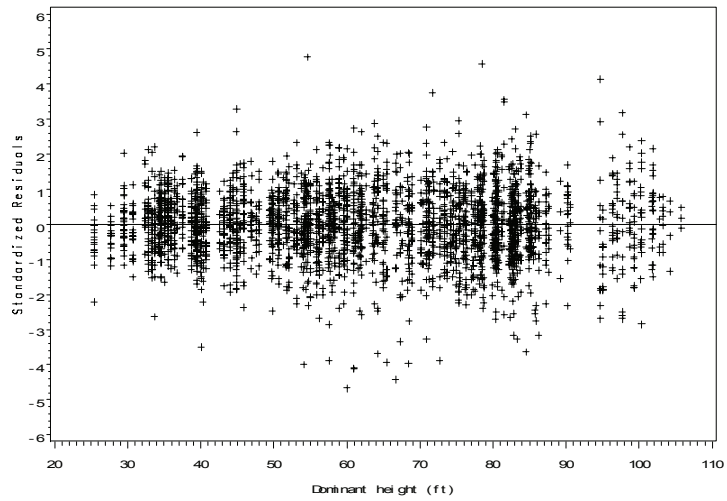


Figure 5.24. Plot of standardized residuals vs. dominant height for total height model

Boxplots of standardized residuals by design criteria are presented in Figures 5.25 to 5.27.

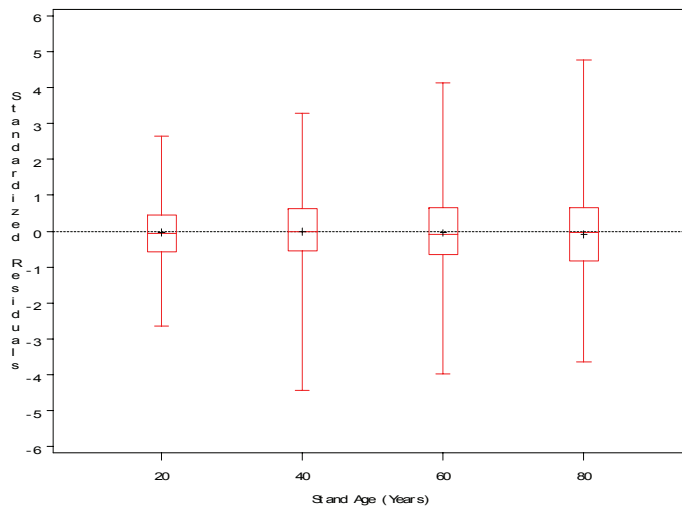


Figure 5.25 Boxplots of standardized residuals for Height Model II by age classes

Figure 5.25 shows that there is some evidence that residuals increase with stand age, however more than 75% of the standardized residuals are within ± 1 SD. It is as expected that old trees or large trees would have more variability than young trees in terms of height variation.

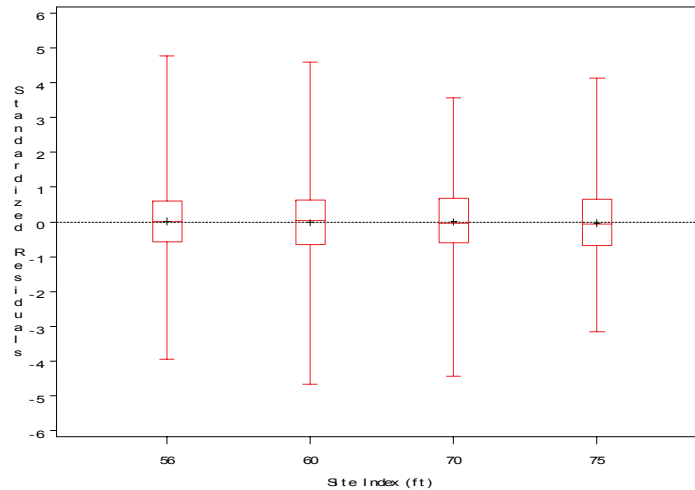


Figure 5.26 Boxplots of standardized residuals for Height Model II by site index classes

Figure 5.26 suggests that error variance is more or less constant over site index classes.

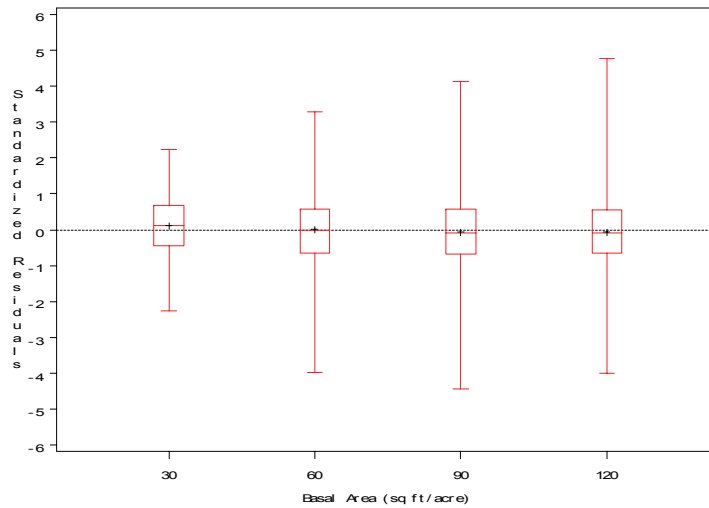


Figure 5.27. Boxplots of standardized residuals for Height Model II by stand basal area classes

Similar to Figure 5.25, Figure 5.27 also indicates that large basal area classes have more variability than small basal area classes.

Overall, it was found that random-plot effect could be incorporated in total height model of Lynch et al. (1999). The residual plots with predicted values and independent variables indicate no problem with assumptions of independent and normal errors. However, residuals appear to increase slightly with age, site index and basal area classes, as expected. The random-effect model was refitted with the complete data set, and the results and discussion are presented in Chapter VI.

CHAPTER VI

DISCUSSION AND CONCLUSIONS

This chapter presents the summary of results from fitting model parameters using the complete data set. These results are discussed for drawing conclusions, also considering the calibration results and model validation as presented in Chapter V.

Basal Area Growth Model

The four most promising models for basal area growth along with BAG Model I were refitted with the complete data from all the 208 plots. The summary statistics for the fitted models are presented in Table 6.1.

Table 6.1. Summary statistics for fitted basal area growth models using complete dataset (total observations = 15,669, and number of plots = 208)

BAG Model	AIC	BIC	Log-likelihood	Residual df	Residual SD
I	-112108.5	-112047.2	56062.25	15661	0.006760735
II	-116554.6	-116485.6	58286.29	15455	0.005702120
VI	-117290.9	-117221.9	58654.44	15248	0.005454462
VII	-124169.2	-124092.6	62094.62	15455	0.009030833
X	-124508.5	-124431.9	62264.25	15248	0.008648375

Since there was no contribution of spatial correlation function as shown by fitted models for calibration data set, the models with spatial correlation function only and the models with both spatial correlation and variance functions were not considered for complete data set.

Parameter estimates and other related statistics for BAG Model I are presented in Table 6.2.

Table 6.2. Parameter estimates and other associated statistics for BAG Model I for complete dataset (number of plots = 208, total observations = 15,669, and residual df = 15,661)

Parameter	Estimate	Standard error	t-value	P-value
β_1	0.157678	0.0103335	15.26	<0.0001
β_2	0.648118	0.0187773	34.52	<0.0001
β_3	-1.583036	0.1542148	-10.27	<0.0001
β_4	0.009764	0.0003173	30.77	<0.0001
β_5	0.022652	0.0008645	26.20	<0.0001
β_6	0.800802	0.0391721	20.44	<0.0001
β_7	-0.730702	0.0343997	-21.24	<0.0001

Similarly, parameter estimates and other statistics for BAG Model II are presented in Table 6.3.

Table 6.3. Parameter estimates and associated statistics for BAG Model II for complete dataset (number of plots = 208, total observations = 15,669, and residual df = 15,455)

Parameter	Estimate	Standard error	t-value	P-value
β_1	0.134241	0.0072800	18.44	<0.0001
β_2	0.600847	0.0309380	19.42	<0.0001
β_3	-1.393511	0.1330752	-10.47	<0.0001
β_4	0.008939	0.0003403	26.26	<0.0001
β_5	0.011221	0.0011279	9.95	<0.0001
β_6	1.162256	0.0588411	19.75	<0.0001
β_7	-0.683569	0.0707261	-9.66	<0.0001

Variance components for:

$$b_7 (\hat{\sigma}_{b_7}^2) = (0.6882522)^2 = 0.47369109$$

$$\text{Residual } (\hat{\sigma}_e^2) = (0.00570212)^2 = 0.0000325142$$

A 95% confidence interval for b_7 variance component SD is [0.6057073, 0.7820461], which shows that the component is significantly different from zero. The fixed-effect parameter estimates are slightly changed compared to those of Lynch et al. (1999) in terms of the magnitude; but signs of the coefficients remain unchanged. Although mixed modeling is a powerful tool (Gregoire et al. 1995), the standard errors have remained similar between BAG Models I and II. BAG Model II is better than BAG Model I due to much smaller AIC and BIC values in BAG Model II. Furthermore, BAG Model II has larger log-likelihood and smaller residual SD than BAG Model I. These statistics clearly show that addition of plot random-effects improves the model fit.

Extended Model with Period and Plot Random-Effects

The estimates for parameters and variance components for BAG Model VI are given below (Table 6.4).

Table 6.4. Parameter estimates and other associated statistics for BAG Model VI for complete dataset (number of periods=2, plots within period = 416, residual df = 15248, and total observations = 15,669)

Parameter	Estimate	Standard error	t-value	P-value
β_1	0.157922	0.0075157	21.01	<0.0001
β_2	0.408310	0.0185716	21.99	<0.0001
β_3	-1.192032	0.1059112	-11.26	<0.0001
β_4	0.011538	0.0003001	38.44	<0.0001
β_6	1.684489	0.0444872	37.86	<0.0001
β_7	-0.138805	0.1231828	-1.13	0.2598

Variance component estimates for:

$$b_7 \text{ for period } (\hat{\sigma}_{Tb_7}^2) = (0.1583321)^2 = 0.025069053$$

$$b_7 \text{ for plots within period } (\hat{\sigma}_{T(p)b_7}^2) = (0.5755995)^2 = 0.331314784$$

$$\text{Residual } (\hat{\sigma}_e^2) = (0.005454462)^2 = 0.0000297512$$

All the fixed-effects coefficients are significantly different from zero except β_7 (Table 6.4). A 95% confidence interval for b_7 random-effect SD for period is [0.05241957, 0.4782387], and that for plots within period is [0.5304568, 0.6245839]. Since the variance components for b_7 are statistically significant, its non-significant fixed-effect would still be included in the model since a random-effect without its

associated fixed-effect is not typically estimated (Pinheiro and Bates 2000 p 355).

Moreover, this model does not have stand age as an independent variable.

Extended Model Accounting for Variance Heterogeneity

BAG Model VII includes a variance function as power of tree basal area to improve upon BAG Model II. Parameter estimates and test statistics for BAG Model VII are presented in Table 6.5.

Table 6.5. Parameter estimates and other associated statistics for BAG Model VII for complete dataset (number of plots = 208, total observations = 15,669, and residual df = 15,455)

Parameter	Estimate	Standard error	t-value	P-value
β_1	0.092850	0.0045174	20.55	<0.0001
β_2	0.595209	0.0227244	26.19	<0.0001
β_3	-2.886070	0.1285173	-22.46	<0.0001
β_4	0.016052	0.0004206	38.16	<0.0001
β_5	0.006191	0.0011752	5.27	<0.0001
β_6	1.684587	0.0531012	31.72	<0.0001
β_7	-0.873136	0.1239771	-7.04	<0.0001

Variance components for:

$$b_7 (\hat{\sigma}_{b_7}^2) = (1.348321)^2 = 1.8179695$$

$$\text{Residual } (\hat{\sigma}_e^2) = (0.009030833)^2 = 0.0000815559$$

$$\text{Power estimate } (\hat{\delta}) = 0.5727463$$

A 95% confidence interval for b_7 variance component SD is [1.178826, 1.542188], which shows that the component is significantly different from zero.

Extended Model with Random-Effects for Growth Period and Plots Accounting for
Variance Heterogeneity

BAG Model X shows improvement over BAG Model VI because of the addition of a power variance function. Parameter estimates and significance testing information are presented in Table 6.6.

Table 6.6. Parameter estimates and other associated statistics for BAG Model X for complete dataset (number of plots = 208, total observations = 15,669, and residual df = 15,248)

Parameter	Estimate	Standard error	t-value	P-value
β_1	0.097151	0.0037167	26.14	<0.0001
β_2	0.521164	0.0195808	26.62	<0.0001
β_3	-3.226228	0.1132377	-28.49	<0.0001
β_4	0.017791	0.0003867	46.01	<0.0001
β_6	2.096922	0.0477478	43.92	<0.0001
β_7	-0.036040	0.1949761	-0.18	0.8534

Variance components for:

$$b_7 \text{ for period } (\hat{\sigma}_{Tb_7}^2) = (0.2469393)^2 = 0.060979017$$

$$b_7 \text{ for plots within period } (\hat{\sigma}_{T(p)b_7}^2) = (0.9969519)^2 = 0.99391309$$

$$\text{Residual } (\hat{\sigma}_e^2) = (0.008648375)^2 = 0.0000747944$$

$$\text{Power estimate } (\hat{\delta}) = 0.5598145$$

The criteria such as AIC and BIC in Table 6.1 indicate that BAG Model X is better than BAG Model VI resulting from addition of variance function to account for

possible heterogeneous errors. All the coefficients are significantly different from zero except β_7 (Table 6.6). This model could not have stand age as an independent variable in presence of stand basal area.

Residual Analysis for Basal Area Growth Models

Standardized residuals from the fitted models are plotted against the predicted values, and plots are presented in Figures 6.1 to 6.4.

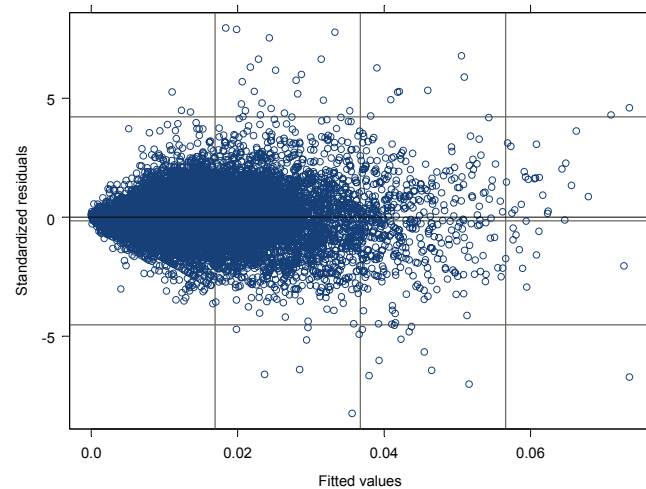


Figure 6.1. Standardized residuals vs. fitted values for BAG Model II (complete dataset)

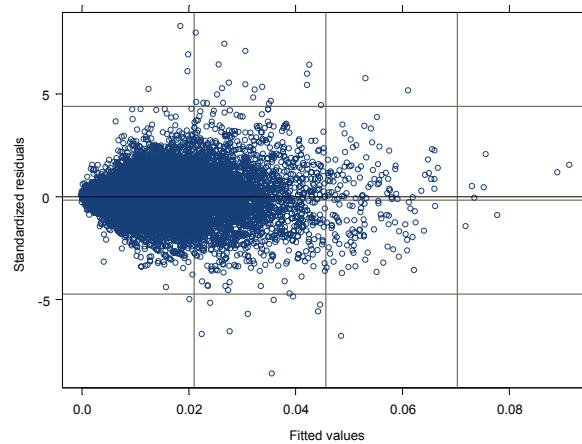


Figure 6.2. Standardized residuals vs. fitted values for BAG Model VI (complete dataset)

Although it is not very evident, Figures 6.1 and 6.2 indicate that residual variability has increased slightly with predicted values. Once power variance function is included in these two models, the residuals show more random scatter than before, as shown by Figures 6.3 and 6.4. Therefore, the variance modeling was found useful.

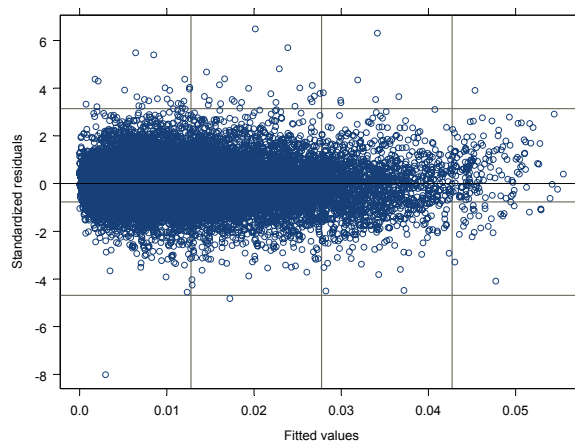


Figure 6.3. Standardized residuals vs. fitted values for BAG Model VII (complete dataset)

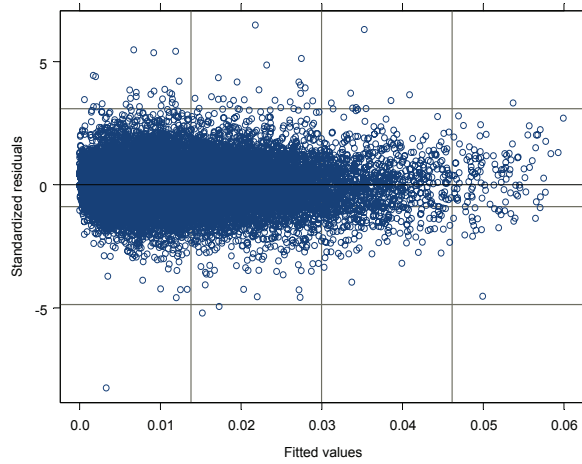


Figure 6.4. Standardized residuals vs. fitted values for BAG Model X (complete dataset)

Another common interest among model users is to see how error variance is distributed over diameter classes. Thus distribution of residuals over 2-inch diameter classes was derived for all the four candidate models. The variability is represented by boxplots, which are presented in Figures 6.5 to 6.8.

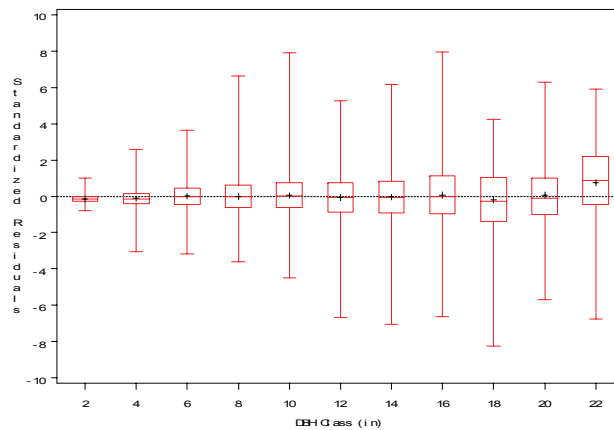


Figure 6.5. Boxplots of standardized residuals for BAG Model II (complete dataset)

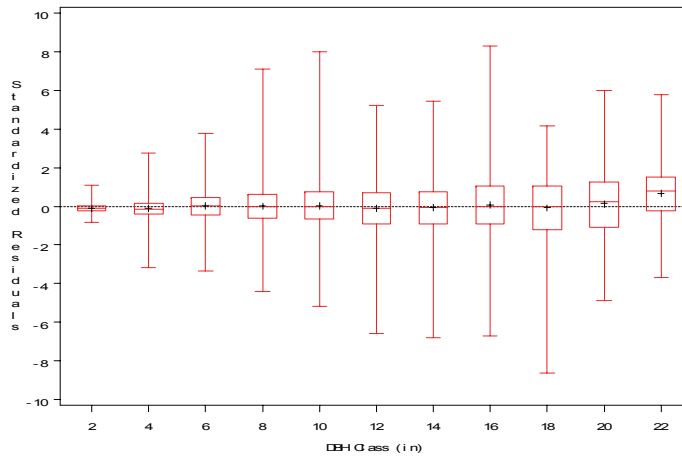


Figure 6.6. Boxplots of standardized residuals for BAG Model VI (complete dataset)

Figure 6.5 shows that variance is increasing with diameter size class, which is as expected because large trees would have more variability than small trees. On the other hand, Figure 6.6 indicates increasing error variance up to 16-in class, then it appears to decrease or to remain constant. However, the variance heterogeneity was not found so serious when the residuals were plotted against predicted values, in which prediction is contributed by other independent variables as well, not only tree basal area, which is in fact calculated from DBH. On the other hand, Figures 6.7 and 6.8 show that inclusion of variance somewhat reverses the scenario. The boxplots show that all the fitted basal area growth models appear to make slight under-prediction above 20-inch diameter classes.

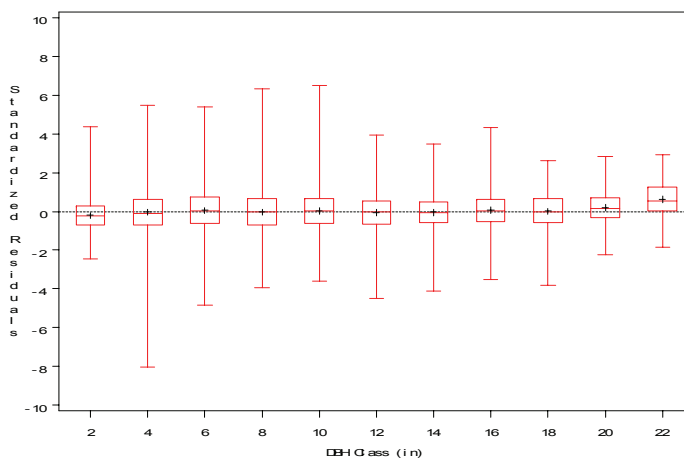


Figure 6.7. Boxplots of standardized residuals for BAG Model VII (complete dataset)

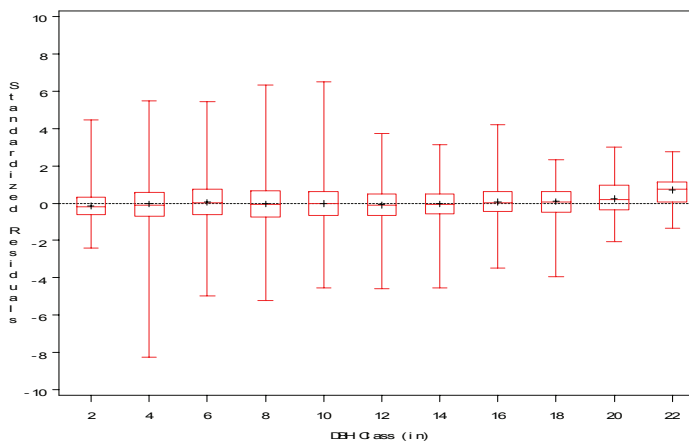


Figure 6.8. Boxplots of standardized residuals for BAG Model X (complete dataset)

Total Height Model

The results obtained by fitting Height Models I and II to the complete data set are presented in this section. Parameter estimates for Height Model I are given in Table 6.7.

Table 6.7. Parameter estimates from SAS PROC NLIN for complete dataset (total observations = 8964, and residual variance = 21.5688)

Parameter	Estimate	Standard error	t-value	P-value
β_0	3.9170	0.1445	27.11	<0.0001
β_1	0.7926	0.00436	181.79	<0.0001
β_2	-2.7474	0.0742	-37.03	<0.0001
β_3	-1.0354	0.0251	-41.25	<0.0001

Summary statistics for Height Model II from SAS PROC NLMIXED are as:

AIC = 51117, BIC = 51143, and $-2 \times \text{Log-likelihood} = 51,105$

Similarly, parameter estimates and testing information from the same model are presented in Table 6.8.

Table 6.8. Parameter estimates from PROC NLMIXED for complete dataset (total observations = 8964, and number of groups = 208)

Parameter	Estimate	Standard error	t-value	P-value
β_0	2.2289	0.08303	26.84	<0.0001
β_1	0.8487	0.007932	106.99	<0.0001
β_2	-5.4830	0.3237	-16.94	<0.0001
β_3	-1.4098	0.03729	-37.81	<0.0001
σ_b^2	1.3630	0.3004	4.54	<0.0001
σ_e^2	15.5597	0.2358	65.98	<0.0001

All the model coefficients are significantly different from zero for both the models (Tables 6.7 and 6.8). However, Height Model II also includes variance component for plot random-effect associated with variable DBH, i.e. parameter β_2 , and this variance component is also significant ($P < 0.0001$). These estimates are similar to those of the model reported by Lynch et al. (1999) with two measurements, or to those of Height Model I with three measurements. However, magnitude of the estimates is slightly changed. These two models are not directly comparable in terms of fit statistics, because one is a nonlinear mixed model and the other is regular nonlinear regression model. However, these two models can be compared in terms of error variance, and residual variance is decreased in Height Model II due to estimation of plot random-effects. Therefore, the total height model with plot random-effects is a better model that could be used to predict the height growth in shortleaf pine natural stands.

Residual Analysis for Height Model II

The standardized residuals from the total height model with plot random-effects were plotted against the predicted values (Figure 6.9), and also against DBH (Figure 6.10).

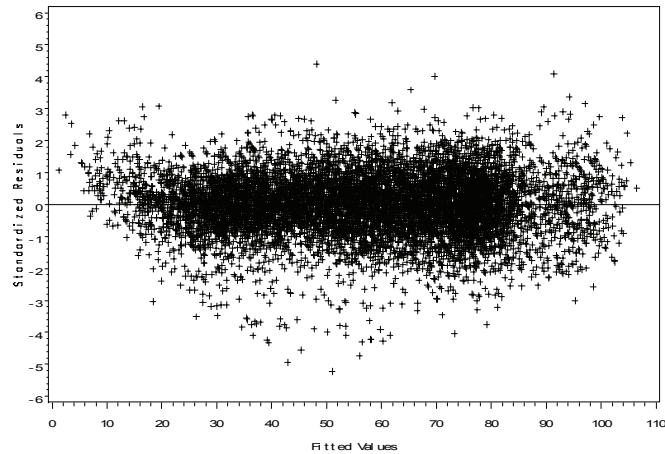


Figure 6.9. Plot of standardized residuals vs. predicted values for total height model from complete dataset

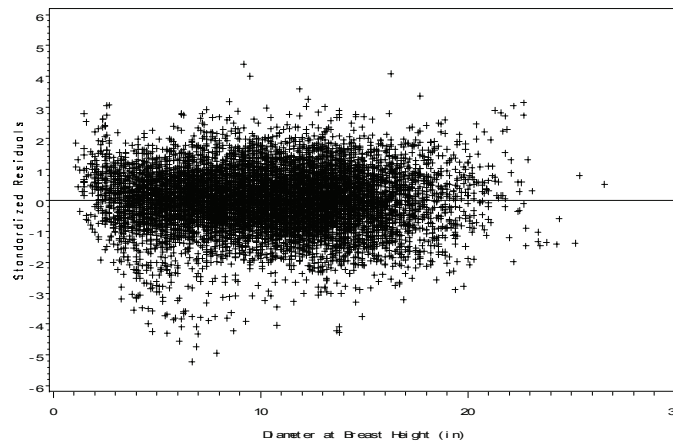


Figure 6.10. Plot of standardized residuals vs. DBH for total height model from complete dataset

These two plots do not reveal any systematic patterns, which indicate that the model fits the dataset well. Height Model II appears to make very good predictions over the range of design variables with a minimum bias, although there appears to be slight over-prediction for small trees. As we considered only one mixed-effects model for total

height, it is clearly a better alternative to the model reported in Lynch et al. (1999). This model is similar in idea to those reported by Lappi (1991), and Lappi and Bailey (1988). The random plot parameters in the total height model would facilitate the prediction of total height in other plots, which are part of the same population. Such models can also be used to estimate random-effects for new stands with minimal measurements. For example, Lynch et al. (2005) used a random-parameter model and calibration for cherrybark oak data from Texas. Similarly, Mehtätalo (2004) used a mixed model with height and diameter data for Norway spruce, and used calibration techniques for a new stand.

Discussion

No shortleaf pine growth models involving random-effects for plots have been published for the Oklahoma and Arkansas region. There has been relatively little work in mixed modeling for other tree species in the region. Thus, comparison of present findings to other work is difficult, but it is evident that the models with plot random-effects are clearly superior to those fitted with ordinary least squares methods due to their statistical properties such as smaller residuals and standard errors. These mixed-models are also more attractive for the reasons of interpretation and applicability of the results. BAG Model I ignores grouping of trees by plots, and the mixed models account for the real data structure. In terms of presentation of results, estimates for fixed-effect parameters and variance components are given for reasonably good models for discussion. However, the main objective of the fitted models is prediction of the response variables rather than

interpretation of individual fixed-effect coefficients. An attempt to interpret individual coefficients would be an unreasonable exercise in such an empirical modeling, although some other nonlinear models could provide options of interpreting parameters that are biologically meaningful. For random-effects, inference on variance components is of interest rather than interpretation of group-specific individual coefficients.

As in validation results, all four basal area growth models were found to make reasonable predictions over the range of design criteria. However, BAG Models II and VII appear to slightly over-predict at higher stand ages. Overall, residual variance increased with tree size classes, which could be corrected with variance modeling. Thus, either BAG Model II or VII would be a good selection. Despite improved statistical properties, BAG Models VI and X are not practically very appealing because they do not have stand age in the model, and also they have period-specific coefficients. The following likelihood ratio test shows that BAG Model VII is significantly better than BAG Model II.

```
> anova(fm1.abagComp.nlme, fmVar1.abagComp.nlme)
              Model df      AIC      BIC  logLik
fm1.abagComp.nlme    1   9 -116554.6 -116485.6 58286.29
fmVar1.abagComp.nlme  2  10 -124169.2 -124092.6 62094.62

              Test L.Ratio p-value
fm1.abagComp.nlme
fmVar1.abagComp.nlme 1 vs 2 7616.65 <.0001
```

where `fm1.abagComp.nlme` and `fmVar1.abagComp.nlme` are output files from complete data set for BAG Models II and VII, respectively. Similarly the height model with plot

random-effects, i.e. Height Model II, is better than Height Model I based on statistical criteria such as reduced residual variance, although no direct comparison could be made since one is a mixed model and the other is not. They often have different fit statistics, making comparison difficult. Mixed models take the grouped data structure into account, e.g. data grouped by plots, which are preferred over the models that ignore grouping structure.

Convergence problems in model parameter estimation were experienced at times due to voluminous data, complexity of the fitted models and correlations among model parameters. Correlation among parameters is well recognized for the Chapman-Richards function, as reported by Murphy and Shelton (1996) for loblolly pine.

The two components of growth models considered in this dissertation are only a part of a complex growth and yield modeling system in which mortality and crown characteristics, and possibly some aspects of environmental variables, would also be taken into account. For example, a form of crown ratio model used by Dyer and Burkhart (1987) has also been used for shortleaf pine modeling (Lynch et al. 1999). However, basal area growth and height-DBH relationship models are two very important components in the growth and yield modeling system. Since these two attributes are modeled with the powerful tool of mixed modeling for the first time in shortleaf pine, this work should make a valuable contribution to the shortleaf pine growth and yield modeling effort for the Oklahoma and Arkansas region.

Since this dissertation involves only two aspects of shortleaf pine growth modeling, a number of other things should be considered to develop an extensive shortleaf growth modeling framework. Input variables in the simulator should be

carefully chosen, and output should be interpreted accordingly. Growth models are just a tool that aids in forest management decisions. There are several other factors that forest managers take into account for better decision making. The factors could be some or several of the biological, economic, ecological and social issues. Avery and Burkhart (2002), Adlard (1995) and several others caution on the proper use of forest growth models.

As indicated before in Chapter IV, modeling was also carried out by setting negative growth values to zero for basal area growth modeling. There were only about 1.3% negative values in the complete dataset. These negative values could have resulted from genuine measurements or some measurement errors. The analysis showed that there was not much difference in results when negative values were set to zero. However, the model fit was improved slightly as indicated by the fit criteria. Therefore, all the results and discussion in this dissertation are based on analysis of the dataset including negative growth values. There has been some work in correcting the measurement error, such as by Canavan and Hann (2004) in which they propose a two-stage method for measurement error characterization. This study did not consider the measurement error characterization; however, errors were modeled to consider for spatial correlation and heterogeneity.

Underutilization of such a powerful tool as mixed-effects modeling has been apparent in forestry (Gregoire et al. 1995). However, the availability of computing facilities such as `n.lme` library in S-Plus has facilitated increasing use of such methods. This view is supported by the fact that articles with application of mixed-modeling

approach have increasingly appeared in professional journals such as *Forest Science*, *Canadian Journal of Forest Research*, and *Forest Ecology and Management*.

This study utilized data from permanently established growth research plots; however, forest inventory data have also been used for growth and yield modeling such as Lessard et al. (2001). Due to the sparse plots and mixed species, models from inventory data are sometimes considered inferior to the models from specifically designed studies like this shortleaf pine growth study. This study used a population of trees within a plot, not a sample of independent individual trees.

It is common to use calibration and validation techniques in forest growth and yield modeling (Wykoff 1990). Similar procedures were used for this dissertation also. But instead of using parameter estimates from calibration set and making predictions for the validation dataset, the promising models were again refitted using validation set. The residuals from validation set were then analyzed. This approach was reasonable because use of plot-specific random-effect coefficients from calibration set would not be appropriate for prediction using validation dataset.

There is also an opinion that validation may not be needed once models are developed from complete dataset (Kozak and Kozak 2003). They argue that the cross-validation does not necessarily add extra information when fit and lack of fit criteria are properly used from the complete dataset.

Moreover, this study used the averaging method to calculate annual basal area growth assuming linear growth for the period of either four or five years. Some researchers such as Cao et al. (2002) have suggested an alternative interpolation method to calculate annual growth.

One important reason why spatial correlation within plot does not appear important is that a correlation coefficient is estimated as an average over all the plots within S-Plus software instead of getting a separate correlation coefficient for each plot. Even if a separate correlation coefficient could be computed for each plot, it would be more or less equivalent to plot random-effect for the variables.

This study is still limited in terms of the number of repeated measurements. However, data gathering from the plots over the years will continue. Due to limited observations over time (repeated measurements), temporal correlations among measurements could not be addressed through modeling of within-subject covariance matrix over time. However, part of temporal correlation was taken into account by taking difference between observations for two time points, at least for basal area growth modeling.

The idea of mixed modeling can be utilized for other components of shortleaf growth models such as crown models and mortality models. When all the components can capture the random nature of plot-specific parameters, the whole growth and yield modeling system should be more efficient than the existing system of Lynch et al. (1999).

With the availability of additional data, the mixed models could include temporal correlations also. Temporal correlation modeling could be captured using the idea of longitudinal data analysis (Fitzmaurice et al. 2004). Although we are not interested in the tree-level random-effect, the idea of subject-specific random-effect modeling might be useful. Spatial correlation for within-plot error could possibly be expanded to capture plot-specific correlation patterns, although this would be largely captured by plot random-effects (Gregoire et al. 1995).

Conclusions

Since the models of Lynch et al. (1999) are intrinsically nonlinear and good computing tools for such models are already available, nonlinear mixed-models were fitted. BAG Model VII is a good model that could be used to predict annual basal area growth for shortleaf pine natural stands in Oklahoma and Arkansas. The residual-plots indicated that the model makes reliable predictions for the basal area growth. Although BAG Models VI and X appear to be statistically sound, their practical application is doubtful due to two main reasons: forest managers will not be as interested in a period-specific coefficient, and these models do not have stand age as an explicit independent variable, which is a practically useful variable used in forest growth analysis. BAG Model VII might be a very good model since variance modeling improved the model fit in addition to estimation of plot random-effects. Therefore, BAG Model VII can be selected for annual basal area growth prediction in shortleaf pine natural stands for the Oklahoma and Arkansas region.

Height Model II is an attractive alternative to the model reported in Lynch et al. (1999) due to estimation of random-effects for plots. This approach would allow generalization of the results over the region from which the data were collected. Also, this model could be used to estimate a random coefficient of a new sample stand or plot with limited measurements as done by Lynch et al. (2005) for cherrybark oak.

The results from this modeling exercise for basal area growth and total height can be used in the Shortleaf Pine Stand Simulator. A complete forest growth and yield modeling system would also include mortality functions and models to account for crown

attributes, in addition to basal area growth and total height growth models. The remaining components of the growth models can also be fitted under the framework of mixed modeling. It is expected that the mixed-modeling framework would provide more precise estimates for use by the simulator.

LITERATURE CITED

- Adlard, P.G. 1995. Myth and Reality in Growth Estimation. *For. Ecol. Manage.* 71:171–176.
- Amateis, R.L., H.E. Burkhart, and T.A. Walsh. 1989. Diameter Increment and Survival Equations for Loblolly Pine Trees Growing in Thinned and Unthinned Plantations on Cutover, Site-Prepared Lands. *South. J. Appl. For.* 13:170–174.
- American Forestry Association. 1982. National Register of Big Trees. *American Forests.* 88(4):18–47.
- Apiolaza, L.A., A.R. Gilmour, and D.J. Garrick. 2000. Variance Modelling of Longitudinal Height Data from a *Pinus radiata* Progeny Test. *Can. J. For. Res.* 30:645–654.
- Atta-Boateng, J., and J.W. Moser, Jr. 2000. A Compatible Growth and Yield Model for the Management of Mixed Tropical Rain Forest. *Can. J. For. Res.* 30:311–323.
- Avery, T.E., and H.E. Burkhart. 2002. *Forest Measurements*, 5th Edition. McGraw-Hill, New York. 456 p.
- Bailey, R.L., and T.R. Dell. 1973. Quantifying Diameter Distributions with the Weibull Function. *For. Sci.* 19:97–104.
- Baldwin, Jr., V.C., H.E. Burkhart, J.A. Westfall, and K.D. Peterson. 2001. Linking Growth and Yield and Process Models to Estimate Impact of Environmental Changes on Growth of Loblolly Pine. *For. Sci.* 47:77–82.

- Bates, D.M., and D.G. Watts. 1988. *Nonlinear Regression Analysis and Its Applications*. John Wiley & Sons, Inc., New York. 365 p.
- Beaumont, J.F., C.H. Ung, and M. Bernier-Cardou. 1999. Relating Site Index to Ecological Factors in Black Spruce Stands: Tests of Hypotheses. *For. Sci.* 45:484–491.
- Bella, I.E. 1971. A New Competition Model for Individual Trees. *For. Sci.* 17:364–372.
- Bitoki, O., L.R. Gering, T.B. Lynch, and P.A. Murphy. 1997. An Individual Tree Basal Area Growth Model for Uneven-Aged Stands for Shortleaf Pine (*Pinus echinata* Mill.) in the Ouachita Mountains in Arkansas and Oklahoma. In: *Proceedings of the 9th Biennial Southern Silvicultural Research Conference*, Clemson, SC; February 25–27, 1997.
- Brown, G.S. 1965. Point Density in Stems per Acre. *New Zealand For. Res. Note No.* 38.
- Bruner, H.D., and J.W. Moser, Jr. 1973. A Markov Chain Approach to the Prediction of Diameter Distributions in Uneven-Aged Forest Stands. *Can. J. For. Res.* 3:409–417.
- Budhathoki, C.B., T.B. Lynch, and J.M. Guldin. 2005. Individual Tree Growth Models for Natural Even-Aged Shortleaf Pine (*Pinus echinata* Mill.). In: *Proceedings of the 13th Biennial Southern Silvicultural Conference*, Memphis, TN; February 28–March 4, 2005.
- Bullock, B.P., and H.E. Burkhart. 2005. An Evaluation of Spatial Dependency in Juvenile Loblolly Pine Stands Using Stem Diameter. *For. Sci.* 51:102–108.
- Buongiorno, J., and B. Michie. 1980. A Matrix Model of Even-Aged Forest Management. *For. Sci.* 26:609–625.

- Burkhart, H.E., K.D. Farrar, R.L. Amateis, and R.F. Daniels. 1987. Simulation of Individual Tree Growth and Stand Development in Loblolly Pine Plantations on Cutover, Site-Prepared Areas. Virginia Poly. Inst. and State Univ., Pub. FWS-1-87. 47 p.
- Calama, R., and G. Montero. 2004. Interregional Nonlinear Height-Diameter Model with Random Coefficients for Stone Pine in Spain. *Can. J. For. Res.* 34:150-163.
- Calegario, N., R.F. Daniels, R. Maestri, and R. Neiva. 2005. Modeling Dominant Height Growth Based on Nonlinear Mixed-Effects Model: a Clonal *Eucalyptus* Plantation Case Study. *For. Ecol. Manage.* 204:11-20.
- Canavan, S.J., and D.W. Hann. 2004. The Two-Stage Method for Measurement Error Characterization. *For. Sci.* 50:743-756.
- Candy, S.G. 1997. Estimation in Forest Yield Models Using Composite Link Functions with Random Effects. *Biometrics* 53:146-160.
- Cao, Q.V., S. Li, and M.E. McDill. 2002. Developing a System of Annual Tree Growth Equations for the Loblolly Pine – Shortleaf Pine Type in Louisiana. *Can. J. For. Res.* 32:2051-2059.
- Chatfield, C. 1989. *The Analysis of Time Series: An Introduction*, 5th Edition. Chapman and Hall, London.
- Chi, E.M., and G.C. Reinsel. 1989. Models for Longitudinal Data with Random Effects and AR(1) Errors. *J. Amer. Stat. Assoc.* 84:452-459.
- Clutter, J.L. 1963. Compatible Growth and Yield Models for Loblolly Pine. *For. Sci.* 9:354-371.

- Clutter, J.L., J.C. Fortson, L.V. Pienaar, G.H. Brister, and R.L. Bailey. 1983. Timber Management: A Quantitative Approach. John Wiley & Sons Inc., New York. 333p.
- Corbeil, R.R., and S.R. Searle. 1976. Restricted Maximum Likelihood (REML) Estimation of Variance Components in the Mixed Model. *Technometrics*. 18:31–38.
- Cressie, N.A.C. 1993. *Statistics for Spatial Data*. Wiley, New York. 900 p.
- Crowder, M.J., and D.J. Hand. 1990. *Analysis of Repeated Measures*. Chapman and Hall, London. 257 p.
- Curtis, R.O. 1967. Height-Diameter and Height-Diameter-Age Equations for Second Growth Douglas-Fir. *For. Sci.* 13:365–375.
- Davidian, M., and D.M. Giltinan. 2003. Nonlinear Models for Repeated Measurement Data: An Overview and Update. *J. Agric. Biol. Environ. Stat.* 8:387–419.
- Davidian, M., and D.M. Giltinan. 1995. *Nonlinear Models for Repeated Measurement Data*, 1st Edition. Chapman and Hall, London. 359 p.
- Davis, C.S. 2002. *Statistical Methods for the Analysis of Repeated Measurements*. Springer-Verlag, New York. 415 p.
- Davis, L.S., K.N. Johnson, P.S. Bettinger and T.E. Howard. 2001. *Forest Management*, 4th Edition. McGraw-Hill, New York. 804 p.
- Diggle, P.J. 1990. *Time Series: A Biostatistical Introduction*. Clarendon Press, Oxford, UK. 257 p.
- Diggle, P.J., K.Y. Liang, and S.L. Zeger. 1994. *Analysis of Longitudinal Data*. Clarendon Press, Oxford, UK. 253 p.

- Draper, N.R., and H. Smith. 1998. *Applied Regression Analysis*. 3rd Edition. Wiley, New York.
- Dyer, M.E., and H.E. Burkhart. 1987. Compatible Crown Ratio and Crown Height Models. *Can. J. For. Res.* 17:572–574.
- Fang, Z., and R.L. Bailey. 2001. Nonlinear Mixed Effects Modeling for Slash Pine Dominant Height Growth Following Intensive Silvicultural Treatments. *For. Sci.* 47:287–300.
- Fang, Z., R.L. Bailey, and B.D. Shiver. 2001. A Multivariate Simultaneous Prediction System for Stand Growth and Yield with Fixed and Random Effects. *For. Sci.* 47:550–562.
- Farrar, R.M., Jr. 1973. Southern Pine Site Index Equations. *J. For.* 71:696–697.
- Ferguson, I.S., and Leech, J.W. 1978. Generalized Least Squares Estimation of Yield Functions. *For. Sci.* 24:27–42.
- Fitzmaurice, G.M., N.M. Laird, and J.H. Ware. 2004. *Applied Longitudinal Analysis*. John Wiley & Sons, Inc. 506 p.
- Garber, S.M., and D.A. Maguire. 2003. Modeling Stem Taper of Three Central Oregon Species Using Nonlinear Mixed Effects Models and Autoregressive Error Structures. *For. Ecol. Manage.* 179:507–522.
- Gerald, C.F., and P.O. Wheatley. 1994. *Applied Numerical Analysis*. 5th Edition. Addison-Wesley Publishing Company Inc., New York. 748 p.
- Graney, D.L., and H.E. Burkhart. 1973. Polymorphic Site Index Curves for Shortleaf Pine in the Ouachita Mountains. *USDA For. Ser. Res. Pap.* SO-85. 12 p.

- Graybill, F. 1976. Theory and Application of the Linear Model. Wadsworth Publishing Company Inc., Belmont, CA. 704 p.
- Gregoire, T.G. 1987. Generalized Error Structure for Forestry Yield Models. *For. Sci.* 33:423–444.
- Gregoire, T.G. and O. Schabenberger. 1996a. A Non-linear Mixed-Effects Model to Predict Cumulative Bole Volume of Standing Trees. *J. Appl. Stat.* 23:257–271.
- Gregoire, T.G. and O. Schabenberger. 1996b. Nonlinear Mixed-Effects Modeling of Cumulative Bole Volume with Spatially Correlated Within-Tree Data. *J. Agric. Biol. Environ. Stat.* 1:107–119.
- Gregoire, T.G., O. Schabenberger and J. Barrett. 1995. Linear Modelling of Irregularly Spaced, Unbalanced, Longitudinal Data from Permanent-Plot Measurements. *Can. J. For. Res.* 25:137–156.
- Guilley, E., J.C. Hervé, and G. Nepveu. 2004. The Influence of Site Quality, Silviculture and Region on Wood Density Mixed Model in *Quercus petraea* Liebl. *For. Ecol. Manage.* 189:111–121.
- Guldin, J.M. 1986. Ecology of Shortleaf Pine. In: Proc. Symp. on the Shortleaf Pine Ecosystem. Arkansas Coop. Ext. Ser., University of Arkansas, AR, 25 – 40 p.
- Guldin, J.M., and J.B. Baker. 1988. Yield Comparisons from Even-Aged and Uneven-Aged Loblolly-Shortleaf Pine Stands. *South. J. Appl. For.* 12:107–114.
- Gumpertz, M.L., and S.G. Pantula. 1989. A Simple Approach to Inference in Random Coefficient Models. *The American Statistician* 43:203–210.
- Haefner, J.W. 1996. Modeling Biological Systems: Principles and Applications. Chapman and Hall, New York. 473 p.

- Hall, D.B. and R.L. Bailey. 2001. Modeling and Prediction of Forest Growth Variables Based on Multilevel Nonlinear Mixed Models. *For. Sci.* 47:311–321.
- Hartley, H.O., and J.N.K. Rao. 1967. Maximum Likelihood Estimation for the Mixed Analysis of Variance Model. *Biometrika* 54:93–108.
- Hasenauer, H., R.A. Monserud, and T.G. Gregoire. 1998. Using Simultaneous Regression Techniques with Individual-Tree Growth Models. *For. Sci.* 44:87–95.
- Hitch, K.L. 1994. A Distance-Independent Individual Tree Basal Area Growth Model for Natural Even-Aged Stands of Shortleaf Pine in Eastern Oklahoma and Western Arkansas. M.S. Thesis. Oklahoma State University, Stillwater, OK. 78 p.
- Huang, S., and S.J. Titus. 1999. Estimating a System of Nonlinear Simultaneous Individual Tree Models for White Spruce in Boreal Mixed-Species Stands. *Can. J. For. Res.* 29:1805–1811.
- Huebschmann, M.M, T.B. Lynch, and P.A. Murphy. 1998. Shortleaf Pine Stand Simulator: An Even-aged Natural Shortleaf Pine Growth and Yield Model User's Manual. Res. Rep. P- 967, Okla. State Univ. Agric. Exp. Sta. 25 p.
- Huebschmann, M.M., L.R. Gering, T.B. Lynch, O. Bitoki and P.A. Murphy. 2000. An Individual-Tree Growth and Yield Prediction System for Uneven-Aged Shortleaf Pine Stands. *South. J. Appl. For.* 24:112–120.
- Johnsen, K., L. Samuelson, R. Teskey, S. McNulty, and T. Fox. 2001. Process Models as Tools in Forestry Research and Management. *For. Sci.* 47:2–8.
- Jordan, L., R.F. Daniels, A. Clark III, and R. He. 2005. Multilevel Nonlinear Mixed-Effects Models for the Modeling of Earlywood and Latewood Microfibril Angle. *For. Sci.* 51:357–371.

- Kozak, A., and R. Kozak. 2003. Does Cross Validation Provide Additional Information in the Evaluation of Regression Models? *Can. J. For. Res.* 33:976–987.
- Krajicek, J.E., K.A. Brinkman, and S.F. Gingrich. 1961. Crown Competition – A Measure of Density. *For. Sci.* 7:35–42.
- Kuehl, R.O. 2000. *Design of Experiments: Statistical Principles of Research Design and Analysis*. 2nd Edition. Duxbury Press, NY. 666 p.
- Laird, N.M., and J.H. Ware. 1982. Random-Effects Models for Longitudinal Data. *Biometrics* 38:963–974.
- Lappi, J. 1991. Calibration of Height and Volume Equations with Random Parameters. *For. Sci.* 37:781–801.
- Lappi, J. 1997. A Longitudinal Analysis of Height/Diameter Curves. *For. Sci.* 43:555–570.
- Lappi, J., and R.L. Bailey. 1988. A Height Prediction Model with Random Stand and Tree Parameters: An Alternative to Traditional Site Index Methods. *For. Sci.* 34:907–927.
- Leites, L.P., and A.P. Robinson. 2004. Improving Taper Equations of Loblolly Pine with Crown Dimensions in a Mixed-Effects Modeling Framework. *For. Sci.* 50:204–212.
- Lessard, V.C., R.E. McRoberts, and M.R. Holdaway. 2001. Diameter Growth Models Using Minnesota Forest Inventory and Analysis Data. *For. Sci.* 47:301–310.
- Lin, X., J. Raz, and S.D. Harlow. 1997. Linear Mixed Models with Heterogeneous within-Cluster Variances. *Biometrics* 53:910–923.

- Lindstrom, M.J., and D.M. Bates. 1990. Nonlinear Mixed-Effects Models for Repeated Measures Data. *Biometrics* 46:673–687.
- Littell, R.C. 2002. Analysis of Unbalanced Mixed Model Data: A Case Study Comparison of ANOVA versus REML/GLS. *J. Agric. Biol. Environ. Stat.* 7:472–490.
- Littell, R.C., G.A. Milliken, W.W. Stroup, and R.D. Wolfinger. 1996. *SAS System for Mixed Models*. SAS Institute Inc., Cary, NC. 633 p.
- Longford, N.T. 1993. *Random Coefficient Models*. Clarendon Press, Oxford, UK. 270 p.
- Lynch, T.B., and J.W. Moser, Jr. 1986. A Growth Model for Mixed Species Stands. *For. Sci.* 32:697–706.
- Lynch, T.B., and P.A. Murphy. 1995. A Compatible Height Prediction and Projection System for Individual Trees in Natural Even-Aged Shortleaf Pine Stands. *For. Sci.* 41:194–209.
- Lynch, T.B., A.G. Holley, and D.J. Stevenson. 2005. A Random-Parameter Height-Dbh Model for Cherrybark Oak. *South. J. Appl. For.* 29:22–26.
- Lynch, T.B., P.A. Murphy and E.R. Lawson. 1991. Stand Volume Equations for Managed Natural Even-Aged Shortleaf Pine in Eastern Oklahoma and Western Arkansas. Res. Rep. P-921. Okla. State Univ. Agric. Exp. Sta. 12p.
- Lynch, T.B., K.L. Hitch, M.M. Huebschmann and P.A. Murphy. 1999. An Individual-Tree Growth and Yield Prediction System for Even-Aged Natural Shortleaf Pine Forests. *South. J. Appl. For.* 23:203–211.

- Martin, G.L., and A.L. Ek. 1984. A Comparison of Competitive Measures and Growth Models for Predicting Plantation Red Pine Diameter and Height Growth. *For. Sci.* 30:731–743.
- Max, T.A., and H.E. Burkhart. 1976. Segmented Polynomial Regression Applied to Taper Equations. *For. Sci.* 22:283–289.
- McWilliams, W.H., R.M. Sheffield, M.H. Hansen, and T.W. Birch. 1986. The Shortleaf Resource. In: *Proc. Symp. on the Shortleaf Pine Ecosystem*. Arkansas Coop. Ext. Ser., University of Arkansas, AR. 9–24 p.
- Mead, R. 1967. A Mathematical Model for the Estimation of Inter-Plant Competition. *Biometrics* 23:189–205.
- Mehtätalo, L. 2004. A Longitudinal Height-Diameter Model for Norway Spruce in Finland. *Can. J. For. Res.* 34:131–140.
- Meng, F.R., C.H. Meng, S. Tang, and P.A. Arp. 1997. A New Height Growth Model for Dominant and Codominant Trees. *For. Sci.* 43:348–354.
- Mood, A.M., F.A. Graybill, and D.C. Boes. 1974. *Introduction to the Theory of Statistics*. McGraw-Hill.
- Moser, J.W., Jr. 1972. Dynamics of an Uneven-Aged Forest Stand. *For. Sci.* 18:184–191.
- Moser, J.W. Jr., and O.F. Hall. 1969. Deriving Growth and Yield Functions for Uneven-Aged Forest Stands. *For. Sci.* 15:183–188.
- Murphy, P.A. 1986. Growth and Yield of Shortleaf Pine. In: *Proc. Symp. on the Shortleaf Pine Ecosystem*. Arkansas Coop. Ext. Ser., University of Arkansas, AR. 159–177p.

- Murphy, P.A. 1982. Sawtimber Growth and Yield for Natural Even-Aged Stands of Shortleaf Pine in the West Gulf. USDA For. Ser. Res. Pap. SO-181. 13 p.
- Murphy, P.A., and R.C. Beltz. 1981. Growth and Yield of Shortleaf Pine in the West Gulf Region. USDA For. Ser. Res. Pap. SO-169. 15 p.
- Murphy, P.A., and R.M. Farrar, Jr. 1988. A Framework for Stand Structure Projection of Uneven-Aged Loblolly-Shortleaf Pine Stands. For. Sci. 34:321–332.
- Murphy, P.A., and M.G. Shelton. 1996. An Individual-Tree Basal Area Growth Model for Loblolly Pine Stands. Can. J. For. Res. 26:327–331.
- Murphy, P.A., E.R. Lawson, and T.B. Lynch. 1992. Basal Area and Volume Development of Natural Even-Aged Shortleaf Pine Stands in the Ouachita Mountains. South. J. Appl. For. 16:30–34.
- Neter, J., M.H. Kutner, C.J. Nachtsheim, and W. Washerman. 1996. Applied Linear Regression Models. 3rd Edition. Irwin. 720 p.
- ODAFS. 2000. Forest Trees of Oklahoma. Oklahoma Department of Agriculture Forestry Services, Oklahoma City. 205 p.
- Omule, S.A.Y., and MacDonald, R.N. 1991. Simultaneous Curve Fitting for Repeated Height-Diameter Measurements. Can. J. For. Res. 21:1418–1422.
- Pienaar, L.V., and K.J. Turnbull. 1973. The Chapman-Richards Generalization of Von Bertalanffy's Growth Model for Basal Area Growth and Yield in Even-Aged Stands. For. Sci. 19:2–22.
- Pinheiro, J. and D. Bates. 2000. Mixed-Effects Models in S and S-Plus. Springer-Verlag, New York. 528 p.

- Ratkowsky, D.A. 1983. *Nonlinear Regression Modeling: A Unified Practical Approach*. Marcel Dekker, Inc., New York. 276 p.
- Reiczigel, J. 1999. Analysis of Experimental Data with Repeated Measurements. *Biometrics*. 55:1059–1063.
- Richards, F.J. 1959. A Flexible Growth Function for Empirical Use. *J. Exp. Bot.* 10:290–300.
- Ripley, B.D. 1981. *Spatial Statistics*. John Wiley & Sons, New York. 241 p.
- Rose, C. E., Jr. 1998. Parameter Estimation for a Shortleaf Pine (*Pinus echinata* Mill.) Basal Area Growth Model. M.S. Thesis. Oklahoma State Univ., Stillwater, OK. 100 p.
- Rose, C.E., Jr., and T.B. Lynch. 2001. Estimating Parameters for Tree Basal Area Growth with a System of Equations and Seemingly Unrelated Regressions. *For. Ecol. Manage.* 148:51–61.
- Rose, C. E., Jr., D.B. Hall, B.D. Shiver, M.L. Clutter and B. Borders. 2006. A Multilevel Approach to Individual Tree Survival Prediction. *For. Sci.* 52:31–43.
- Schulte, B.J., and J. Buongiorno. 2004. A Growth and Yield Model for Naturally-Regenerated Mixed Shortleaf Pine Forests in the Southern United States of America. *International Forestry Review* 6:19–29.
- Schumacher, F.X. 1939. A New Growth Curve and its Application to Timber-Yield Studies. *J. For.* 37:819–820.
- Schumacher, F.X., and T.S. Coile. 1960. *Growth and Yield of Natural Stands of the Southern Pines*. T.S. Coile, Inc., Durham, NC. 115 p.

- Shifley, S.R. 1987. A Generalized System of Models Forecasting Central States Tree Growth. USDA For. Res. Pap. NC-279. 10 p.
- Shifley, S.R., and G.J. Brand. 1984. Chapman-Richards Growth Function Constrained for Maximum Tree Size. For. Sci. 30:1066–1070.
- Smalley, G.W., and R.L. Bailey. 1974. Yield Tables and Stand Structure for Shortleaf Pine Plantations in Tennessee, Alabama, and Georgia Highlands. Southern Forest Experiment Station. USDA For. Ser. Res. Pap. SO-97. 58 p.
- Smith, K.L. 1986. Historical Perspective. In: Proc. Symp. on the Shortleaf Pine Ecosystem. Arkansas Coop. Ext. Ser., University of Arkansas, AR, 1–8 p.
- Smith, W.R., R.M. Farrar, Jr., P.A. Murphy, J.L. Yeiser, R.S. Meldahl, and J.S. Kush. 1992. Crown and Basal Area Relationships of Open-Grown Southern Pines for Modeling Competition and Growth. Can. J. For. Res. 22:341–347.
- Snedecor, G.W., and W.G. Cochran. 1980. Statistical Methods. 7th Edition, Iowa State University Press, Ames, IA.
- Spurr, S.H. 1962. A Measure of Point Density. For. Sci. 8:85–96.
- Sylvester, W.A. 1938. A Comparison of Two Methods of Yield Table Construction. J. For. 38: 681–686.
- Tao, J. 2002. Mixed Models Analyses Using the SAS System: Course Notes. SAS Institute, Inc., Cary, NC. 493 p.
- Tasissa, G., and H.E. Burkhart. 1998. An Application of Mixed Effects Analysis to Modeling Thinning Effects on Stem Profile of Loblolly Pine. For. Ecol. Manage. 103:87–101.

- Tome, M., and H.E. Burkhart. 1989. Distance-Dependent Competition Measures for Predicting Growth of Individual Trees. *For. Sci.* 35:816–831.
- USDA. 2002. Forest Vegetation Simulator (FVS). U.S. Department of Agriculture Forest Service, Forest Management Service Center, Ft. Collins, CO.
(www.fs.fed.us/fmhc/fvs)
- USDA For. Serv. 1929 (rev. 1976). Volume, Yield and Stand Tables for Second-Growth Southern Pines. USDA Misc. Publ. 50. 202 p.
- USDA For. Serv. 1995. Field Instructions Shortleaf Growth and Yield Study. FS-SO-4106-48. Southern Research Station, USDA Forest Service, and Univ. Ark., Monticello.
- Uzoh, F.C.C., and W.W. Oliver. 2006. Individual Tree Height Increment Model for Managed Even-Aged Stands of Ponderosa Pine throughout the western United States using Linear Mixed Effects Models. *For. Ecol. Manage.* 221:147–154.
- Vanclay, J.K. 1994. Modelling Forest Growth and Yield: Applications to Mixed Tropical Forests. CAB International, Wallingford, UK.
- VanLeeuwen, D.M., L.W. Murray, and N.S. Urquhart. 1996. A Mixed Model with Both Fixed and Random Trend Components across Time. *J. Agric. Biol. Environ. Stat.* 1:435–453.
- Vázquez, J., and H. Pereira. 2005. Mixed Models to Estimate Tree Oven-Dried Cork Weight in Central and Southern Portugal. *For. Ecol. Manage.* 213:117–132.
- Venables, W.N., and B.D. Ripley. 2002. Modern Applied Statistics with S, 4th Edition. Springer-Verlag Inc., New York. 495 p.

- Verbeke, G., and G. Molenberghs. 2000. *Linear Mixed Models for Longitudinal Data*. Springer-Verlag Inc., New York. 568 p.
- Vonesh, E.F., and R.L. Carter. 1992. Mixed-Effects Nonlinear Regression for Unbalanced Repeated Measures. *Biometrics* 48:1–18.
- Wackerly, D, W. Mendenhall, and R.L. Scheaffer. 2002. *Mathematical Statistics with Applications*. Duxbury Press.
- Ware, J.H. 1985. Linear Models for the Analysis of Longitudinal Studies. *The American Statistician* 39:95–101.
- Webster, R., and M.A. Oliver. 1990. *Statistical Methods in Soil and Land Resource Survey*. Oxford University Press, New York. 316 p.
- West, P.W. 1980. Use of Diameter Increment and Basal Area Increment in Tree Growth Studies. *Can. J. For. Res.* 10:71–77.
- West, P.W., Ratkowsky, D.A., and Davis, A.W. 1984. Problems of Hypothesis Testing of Regressions with Multiple Measurements from Individual Sampling Units. *For. Ecol. Manage.* 7:207–224.
- Willet, R.L. 1986. Foreword. In: *Proc. Symp. on the Shortleaf Pine Ecosystem*. Arkansas Coop. Ext. Ser., University of Arkansas, AR. iii – iv p.
- Williston, H.L., and W.E. Balmer. 1980. *Shortleaf Pine Management*. USDA For. Ser. Forestry Report SA-FR6. 9 p.
- Wolfinger, R.D. 1996. Heterogeneous Variance: Covariance Structures for Repeated Measures. *J. Agric. Biol. Environ. Stat.* 1:205–230.
- Wykoff, W.R. 1990. A Basal Area Increment Model for Individual Conifers in the Northern Rocky Mountains. *For. Sci.* 36:1077–1099.

- Yang, R.C., A. Kozak, and J.H.G. Smith. 1978. The Potential of Weibull-Type Functions as Flexible Growth Curves. *Can. J. For. Res.* 8:424–431.
- Zeide, B. 2002. Density and Growth of Even-Aged Stands. *For. Sci.* 48:743–754.
- Zeide, B. 1993. Analysis of Growth Equations. *For. Sci.* 39:594–616.
- Zeide, B. 1989. Accuracy of Equations Describing Diameter Growth. *Can. J. For. Res.* 19:1283–1286.
- Zhang, Y., and B.E. Borders. 2004. Using a System Mixed-Effects Modeling Method to Estimate Tree Compartment Biomass for Intensively Managed Loblolly Pines—An Allometric Approach. *For. Ecol. Manage.* 194:145–157.
- Zhao, D., Wilson, M., and B.E. Borders. 2005. Modeling Response Curves and Testing Treatment Effects in Repeated Measures Experiments: A Multilevel Nonlinear Mixed-Effects Model Approach. *Can. J. For. Res.* 35:122–132.

APPENDIX I: Details of Status and Damage Codes

Status code for first two measurements (some of them are not appropriate for first measurement)

- 0 – a live, and height-sample tree
- 1 – a live, and non-height-sample tree
- 2 – a tree that has died since the last inventory
- 3 – a tree missed at the last inventory and not selected for height sample
- 4 – a tree missed at the last inventory and selected for height sample
- 5 – not suitable for growth calculations

Status code for the third measurements

First digit (note there is no 2)

- 0 – tree alive at both last and present inventory
- 1 – tree alive at last inventory but is now dead
- 3 – live tree and marked for thinning
- 4 – live tree not suitable for growth calculations
- 5 – live tree not suitable for growth calculations and marked for thinning
- 6 – maverick tree (problem to be described)
- 7 – maverick tree marked for thinning

Second digit

0 – height-sample tree

1 – non-height-sample tree

Tree damage, if any was recorded using a four-digit system.

First digit (location of damage, defect or injury on the tree)

0 – no damage or no information

1 – tip or leader

2 – foliage

3 – branches

4 – crown, overall damage to all crown portions of the tree (including foliage, branches, tip or leader, and crown pole)

5 – crown bole

6 – main bole (below crown)

7 – base

8 – roots

9 – whole tree

Second digit (severity of damage)

0 – none or unspecified

1 – minor

2 – moderate

3 – severe

4 – lethal

Third digit (nature of damage or defect)

0 – none or unspecified

1 – lean

2 – broken or dead

3 – crook

4 – sweep

5 – loss/skinned

6 – forked

7 – canker

8 – rot

9 – unhealthy appearance, malformed

Fourth digit (Cause of damage or defect)

0 – none or unspecified

1 – logging, felling or other

2 – logging, skidding

3 – insects

4 – weather, other

5 – lightning

6 – animals

7 – competition

8 – disease

9 – fire

A slightly different status code, i.e. double digit, was used for Study 58 Freese's study.

00 – live, height measurement

01 – live, no height measurement

10 – dead (live at last measurement), height measurement

11 – dead (live at last measurement), no height measurement

20 – live, marked for thinning, height measurement, second measurement

21 – live, marked for thinning, no height measurement, second measurement

30 – live, marked for thinning, height measurement, third measurement

31 – live, marked for thinning, no height measurement, third measurement

APPENDIX II: Example S-Plus nlme Output for a Model Using Calibration Data Set

Basic S-Plus nlme Command

```
> fm.abag.nlme<-nlme(ABAG.PT~(b1*BA.MID^(b2) - (b1*BA.MID/(7.068384)^(1-  
b2)))/(1+exp(b3+b4*BA.ACRMD+b5*PLOTAG.M+b6*R+b7*BA.MID)),data=ANALYSIS,  
fixed=(b1+b2+b3+b4+b5+b6+b7~1), random=(b7~1|PLOTNEW),  
start=c(b1=0.1,b2=0.5,b3=-3,b4=0.01,b5=0.05,b6=1.5,b7=-  
1.5),na.action=na.exclude)
```

Basic S-Plus nlme Output

```
> summary(fm.abag.nlme)  
Nonlinear mixed-effects model fit by maximum likelihood  
Model: ABAG.PT ~ (b1 * BA.MID^(b2) - ((b1 * BA.MID)/(7.068384  
)^(1 - b2)))/(1 + exp(b3 + b4 * BA.ACRMD + b5 * PLOTAG.M +  
b6 * R + b7 * BA.MID))  
Data: ANALYSIS  
AIC BIC logLik  
-75919.55 -75854.41 37968.77  
Random effects:  
Formula: b7 ~ 1 | PLOTNEW  
b7 Residual  
StdDev: 0.6641947 0.00583212
```

Fixed effects: list((, b1 + b2 + b3 + b4 + b5 + b6 + b7 ~ 1)

	Value	Std.Error	DF	t-value	p-value
b1	0.127698	0.0094784	10124	13.47257	<.0001
b2	0.577809	0.0406986	10124	14.19726	<.0001
b3	-1.296957	0.1775033	10124	-7.30666	<.0001
b4	0.008657	0.0004271	10124	20.27069	<.0001
b5	0.009476	0.0014489	10124	6.54037	<.0001
b6	1.180113	0.0764147	10124	15.44353	<.0001
b7	-0.653616	0.0908105	10124	-7.19758	<.0001

Standardized Within-Group Residuals:

Min	Q1	Med	Q3	Max
-8.072766	-0.53156	-0.07765724	0.4768549	7.727945

Number of Observations: 10269

Number of Groups: 139

VITA

Chakra B. Budhathoki

Candidate for the Degree of

Doctor of Philosophy

Thesis: MIXED-EFFECTS MODELING OF SHORTLEAF PINE (*PINUS ECHINATA* MILL.) GROWTH DATA

Major Field: Environmental Science, Minor Fields: Forest Resources, and Statistics

Biographical:

Personal Data: Born in Myagdi, Nepal on December 05, 1967. Parents are Mr. Jit Bahadur Budhathoki and Ms. Lilamati Budhathoki.

Education: Received Bachelor of Science in Agriculture from Tribhuvan University, Nepal in July 1992. Received Diploma in Statistics from The University of Reading, England in July 1996. Received Master of Science in Biometry from The University of Reading, England in September 1997. Completed the requirements for the Doctor of Philosophy degree with a major in Environmental Science with option in Forest Resources at Oklahoma State University in July, 2006.

Experience: Employed as an Assistant Agronomist at Lumle Agricultural Research Centre (LARC), Nepal from October 1992 to November 1997 (including study leave). Employed as a Biometrician at LARC, Nepal from December 1997 to June 1998. Worked as a Lecturer at Institute of Agriculture and Animal Sciences (IAAS), Tribhuvan University, Nepal from July 1998 to October 2000. Employed as an Agricultural Statistician at International Centre for Integrated Mountain Development (ICIMOD), Kathmandu, Nepal from November 2000 to October 2001. Worked as a Biometrician at South Asia Regional Office of International Maize and Wheat Improvement Center (CIMMYT) from November 2001 to June 2002. Employed as a Graduate Research Assistant in Department of Forestry at Oklahoma State University from August 2002 to May 2006.

Academic Achievements: Student Stipend for certificate and B.S. studies, Tribhuvan University, Nepal. The British Council Scholarship for M.S. and Diploma studies in England (1995-1997). Dr. Michel Afanasiev Distinguished Graduate Fellowship, Oklahoma State University (2002-2006).

Professional Memberships and Honors: Sigma Xi, Gamma Sigma Delta, Xi Sigma Pi, American Statistical Association, International Biometric Society, International Environmetrics Society, Life member of Indian Society of Agricultural Statistics.

Name: Chakra B. Budhathoki

Date of Degree: July, 2006

Institution: Oklahoma State University

Location: Stillwater, Oklahoma

Title of Study: MIXED-EFFECTS MODELING OF SHORTLEAF PINE
(*PINUS ECHINATA* MILL.) GROWTH DATA

Pages in Study: 168

Candidate for the Degree of Doctor of Philosophy

Major: Environmental Science with Forest Resources Option

Minor: Statistics

Scope and Methods of Study: The objective of this study was to develop individual-tree mixed-effects models for basal area growth and the diameter-height relationship of shortleaf pine (*Pinus echinata* Mill.). Repeated measurements for attributes including diameter at breast height and total height from over 200 permanent plots were available from eastern Oklahoma and western Arkansas. Models with plot random-effects were fitted using the S-Plus `nlme` library and SAS PROC NL MIXED utilizing a calibration dataset. Models with independently and normally distributed errors were fitted first. Then possible spatially correlated and/or heterogeneous within-plot errors were modeled for basal area growth. The most promising models were tested using an independently selected dataset from the same study.

Results and Conclusions: Though increasingly popular in forestry, mixed-effects modeling technique has never previously been used in shortleaf pine growth modeling. Nonlinear mixed models with plot random-effects were found to fit the data better than the models fitted with a complete random sample assumption (the ordinary least-squares method) as reported in Lynch et al. (1999) for both a basal area growth model and a model for diameter-height relationship. Because data were grouped by plots, a mixed-effects model with plot-level random-effects was a more realistic representation of the data structure than ordinary least squares. Spatial correlation among tree measurements within a plot did not appear to be important in presence of plot random-effects. However, variance modeling using a variance function with tree basal area as a covariate accounted for heterogeneity of within-plot errors better than the modeling approach in which constant variance was assumed.

Advisor's Approval: Dr. Thomas B. Lynch