

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA



A SIMULTANEOUS SYSTEM OF
LINEAR AND NONLINEAR MIXED EFFECTS MODELS FOR
FOREST GROWTH AND YIELD PREDICTION

by

ZIXING FANG

B.S., Huazhong Agricultural University, China, 1988

M.S. Beijing Forestry University, China, 1991

M.S. The University of Georgia, 1998

A Dissertation Submitted to the Graduate Faculty
of The University of Georgia in Partial Fulfillment
of the
Requirement for the Degree
DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

1999

UMI Number: 9949495



UMI Microform 9949495

Copyright 2000 by Bell & Howell Information and Learning Company.

All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.

Bell & Howell Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

A SIMULTANEOUS SYSTEM OF
LINEAR AND NONLINEAR MIXED EFFECTS MODELS FOR
FOREST GROWTH AND YIELD PREDICTION

by

ZIXING FANG

Approved:

Robert L. Bailey
Major Professor

July 21, 1999
Date

Approved:

Gordhan L. Patel

Dean of the Graduate School

July 22, 1999
Date

ZIXING FANG

A Simultaneous System of Linear and Nonlinear Mixed Effects Models for Forest
Growth and Yield Prediction
(Under the direction of ROBERT L. BAILEY)

A nonlinear mixed-effects model and a linear multi-level mixed-effects model were developed for slash pine (*Pinus elliottii* Engelm.) dominant height and basal area growth in conjunction with different silvicultural treatments. The mixed-effects model is extended to a simultaneous growth model system including the components of dominant height, basal area and total volume. Prediction for new cases (i.e. plots) is one of the main focuses of the study. The advantages of a mixed-effects model and the simultaneous system in prediction are illustrated with detailed formulas and specific examples. As a comparison, a Bayesian model formulation with a Gibbs sampling estimation process are presented for the nonlinear mixed-effects dominant height growth model.

The general techniques illustrated and applied include: ① Model building for a nonlinear mixed-effects model and for a linear multi-level mixed-effects model. ② Model building in a Bayesian framework with a Gibbs sampler for estimation, ③ Predictions of variable values and determination of variance estimates in different situations, and ④ Development of a multi-variate simultaneous model system with fixed and random effects.

INDEX WORDS: Longitudinal data, Repeated measurements, Split-plot design, Silvicultural treatment, Nonlinear mixed effects model, Multi-level mixed effects model, Prediction for new observations, Variance function, Contemporary correlation, Gibbs sampling

ACKNOWLEDGMENTS

I would like to thank my major professor, Dr. Robert L. Bailey, for his patience, valuable inspiration, encouragement and invaluable assistance throughout my graduate studies at The University of Georgia. Without him I would probably never have been able to start this interesting work on forest growth and yield modeling. I deeply appreciate his time, effort, ingenuity and precision, while training me to be a professional biometristian.

Thanks also to all the rest of my committee who never gave up on me. Robert L. Taylor, Bruce E. Borders, Tim Harrington, Daniel Hall and Chris Cieszewski were always encouraging and helpful throughout my graduate tenure. I sincerely appreciate the time these very busy people spent in evaluating and editing my dissertation.

I would also like to thank W. Michael Harrison and Stacey Martin for their invaluable help on the data preparation of my current work. I am also indebted to Ms. Kimberly Joyner for her generous help on the final report of my dissertation.

My final heartfelt thanks go to my wife for her understanding, encouragement, full support, and sacrificing love during my years of study at The University of Georgia. I dedicated this dissertation to my wife, Gang Hua, daughter, Yuhua, and son Jack.

TABLE OF CONTENTS

	Page
ACKNOWLEDGMENTS.....	iii
LIST OF TABLES.....	vii
LIST OF FIGURES.....	ix
CHAPTER	
1 INTRODUCTION.....	1
Background and Literature Review.....	3
Objectives.....	9
2 NONLINEAR MIXED EFFECTS MODELS FOR THE ANALYSIS OF REPEATED MEASUREMENTS FROM PERMANENT PLOTS IN A DESIGNED EXPERIMENT: A STUDY OF SLASH PINE DOMINANT HEIGHT GROWTH FOLLOWING INTENSIVE SILVICULTURAL TREATMENTS.....	10
Introduction.....	10
Data	13
Method	17
Model Specification.....	20
Parameter Estimation and Model Diagnosis.....	39
Interpretation of Model Parameters.....	44
Individual Prediction for New Observations.....	48

Discussion.....	55
3 LINEAR MULTILEVEL MIXED EFFECTS MODELS FOR THE ANALYSIS OF REPEATED MEASUREMENTS FROM PERMANENT PLOTS IN A DESIGNED EXPERIMENT: A STUDY OF SLASH PINE BASAL AREA GROWTH FOLLOWING INTENSIVE SILVICULTURAL TREATMENTS.....	58
Introduction.....	58
Data.....	59
Generalized Linear Multilevel Mixed Effects Model.....	62
Likelihood Reference for Linear Multilevel Mixed Model.....	64
Model Specification.....	67
Predictions.....	88
Discussion.....	100
4 A NONLINEAR MIXED MODEL FOR SLASH PINE DOMINANT HEIGHT GROWTH: A GIBBS SAMPLER APPROACH.....	103
Introduction	103
Bayesian Model Formulation for A Nonlinear Mixed Effects Model....	104
Gibbs Sampling.....	108
Convergence of Gibbs Sampler.....	113
Posterior Distributions.....	115
Model Diagnostics and Comparison.....	120
Discussion	124
5 A MULTIVARIATE SIMULTANEOUS PREDICTION SYSTEM FOR	

STAND GROWTH AND YIELD WITH FIXED AND RANDOM EFFECTS.....	127
Introduction.....	127
Generalized Simultaneous Stand Growth and Yield Model System.....	128
The Data.....	130
The Simultaneous Growth and Yield Model.....	131
Prediction.....	140
Examples.....	151
Discussion.....	163
6 SUMMARY AND CONCLUSIONS.....	166
LITERATURE CITED.....	170

LIST OF TABLES

	Page
Table 1. The prediction of stand characteristics based on measurements and predetermined exogenous variables at different time points.....	6
Table 2. Basic statistics for dominant height (in meters) by individual factor levels for slash pine based on permanent experimental plots from the Plantation Management Research Cooperative (PMRC) of The University of Georgia.....	14
Table 3. Likelihood ratio tests (LRT) for nonlinear mixed- effects models of slash pine dominant height with different fixed and random effects components.....	25
Table 4. Comparison of mixed-effects model performance for slash pine dominant height growth data with different variance functions.....	28
Table 5. Comparison of mixed effects model performance for slash pine dominant height growth data with different within-plot correlation structures and no inter-plot covariates.....	31
Table 6. Comparison of mixed effects model performance for slash pine dominant height growth data with between-plot covariates (Eq. 2.12) and different within-plot variance function and correlation structures.....	38
Table 7. The estimated fixed effects and corresponding variance covariance matrix for the slash pine nonlinear mixed-effects dominant height growth model (Model 10).....	39
Table 8. The estimated parameters of random effects (\hat{D}), variance function ($\hat{\alpha}$), correlation structure ($\hat{\theta}_1, \hat{\theta}_2$), and the error of the model ($\hat{\sigma}$) for the slash pine nonlinear mixed-effects dominant height growth model (Model 10).....	40
Table 9. The estimated parameters for “typical” dominant height growth (random effects input as 0) with modified Richard’s equation by silvicultural treatment groups.....	46

Table 10.	The significant terms (at $\alpha = 0.05$) resulting from a stepwise regression procedure.....	70
Table 11.	Comparison of mixed-effects model performance for slash pine basal area growth data with different variance functions.....	77
Table 12.	Comparison of mixed-effects model performance for slash pine basal area growth data with different within-plot correlation structures.....	78
Table 13.	Comparison of mixed-effects model performance for slash pine basal area growth data with different fixed and random effects terms.....	80
Table 14.	Approximate 95% confidence intervals for the estimated parameters of Model 9 (with silvicultural treatment and soil type covariates, 3-level nesting of random effects, heteroscedasticity (exponential function) and correlated [AR(1)] within group errors.).....	82
Table 15.	Best linear unbiased predictors (<i>BLUP</i> 's) for random effects at installation level.....	83
Table 16.	The estimated parameters and quantiles for the slash pine nonlinear mixed-effects dominant height growth model fitted with a Gibbs sampling approach.....	117
Table 17.	Comparisons on the basic statistics of residuals, absolute values of the residuals and squares of residual respectively from the Gibbs sampler with mean and median of the samples and the conventional nonlinear mixed effects approach.....	122
Table 18.	The estimated parameters with the Gibbs sampler for dominant height growth with modified Richard's equation by silvicultural treatments groups.....	124
Table 19.	The significant terms (at $\alpha = 0.05$) result from stepwise regression procedure for choosing the fixed effects in the total volume component of the simultaneous system.....	132
Table 20.	The estimated parameters for the fixed effects in the dominant height, basal area and total volume simultaneous model system.....	140
Table 21.	Prediction values and confidence intervals with the simultaneous system based on 8 examples.....	160

LIST OF FIGURES

	Page
Figure 1. Profile graph of slash dominant height growth on 191 plots with 5 repeated measures by 11 silvicultural treatment groups.....	15
Figure 2. Profile graph of slash dominant height growth on 191 plots with 5 repeated measures by 4 soil types.....	16
Figure 3. The mean responses of slash pine dominant height growth by silvicultural treatments and soil types.....	16
Figure 4. The residual plot by the modified Richards dominant height growth equation with ordinary least square (OLS) estimations.....	27
Figure 5. The residual plot by the modified Richards dominant height growth equation with random between plot effects and independent homogenous within-plot errors.....	27
Figure 6. Empirical autocorrelation function corresponding to the standardized residuals of the modified Richards dominant height growth model with random effects and <i>iid</i> within-plot errors.....	30
Figure 7. Scatter plot of estimated random effects for the modified Richards dominant height growth model with random effects and <i>iid</i> within-plot errors.....	33
Figure 8. Comparisons of estimated random effects by silvicultural treatments for the modified Richards dominant height growth model with random effects and <i>iid</i> within-plot errors.....	34
Figure 9. Comparisons of estimated random effects by soil types for the modified Richards dominant height growth model with random effects and <i>niid</i> within-plot errors.....	35
Figure 10. The residual plot for the modified Richards dominant height growth equation with random between-plot effects and <i>MA(2)</i> within-plot autocorrelation and power variance function [Model (2.10)].....	41

Figure 11.	Normal plot of residuals for the nonlinear mixed effects dominant height growth model with $MA(2)$ autocorrelation and power variance function [Model (2.10)].....	42
Figure 12.	Observed dominant height versus predicted dominant height with the nonlinear-mixed effect dominant height growth model [Model (2.10)].....	43
Figure 13.	Normal plot of random effects for the nonlinear mixed-effects dominant height growth model.....	43
Figure 14.	Scatter plot of estimated random effects for the nonlinear mixed-effects dominant height growth model by soil type groups (spodosol or nonspodosol).....	44
Figure 15.	A comparison of the mean responses (typical responses) of dominant height growth by 6 different silvicultural treatments (<i>CNTL</i> , <i>FCHP</i> , <i>FCHB</i> , <i>FCBB</i> , <i>FCBH</i> , <i>FBHB</i>) with the nonlinear mixed-effects dominant height growth model.....	46
Figure 16.	A comparison of individual dominant height growth predictions by typical response and subject-specified response with the nonlinear mixed-effects dominant height growth model.....	47
Figure 17.	A comparison of the prediction and corresponding confidence interval for a new observation by typical response and subject specified response with the nonlinear mixed effects dominant height growth model.....	54
Figure 18.	Profile graph of slash pine basal area growth on 191 plots with 5 repeated measures conditioning on 11 silvicultural treatments.....	60
Figure 19.	Scatter plots of response [$\ln(BA)$] and covariates [1/age, $\ln(HD)$ and $\ln(TPH)$].....	61
Figure 20.	The nesting relationships of the original design for the slash pine basal area growth data.....	61
Figure 21.	The logarithm curve.....	83
Figure 22.	The standardized residual plot of the multilevel linear mixed -effects basal area model [Model (3.9)] for the slash pine data.....	84
Figure 23.	The standardized residual plot by silvicultural treatments with the multilevel linear mixed-effects basal area model.....	85

Figure 24.	The standardized residual plot by soil type groups (spodosol or nonspodosol) with the multilevel linear mixed-effects basal area model.....	85
Figure 25.	Normal plot of residuals for the multilevel linear mixed-effects basal area growth model with $AR(1)$ autocorrelation and exponential variance function [Model (3.9)].....	86
Figure 26.	Normal plot of random effects at installation level for the multilevel linear mixed-effects basal area growth model based on the slash pine data.....	86
Figure 27.	Normal plot of random effects at treatment within installation level for the multilevel linear mixed-effects basal area growth model based on the slash pine data.....	87
Figure 28.	Normal plot of random effects at individual plot level for the multilevel linear mixed effects basal area growth model based on the slash pine data.....	87
Figure 29.	The observed response [$\ln(BA+1)$] versus predicted with the multilevel linear mixed effect basal area growth model.....	88
Figure 30.	A comparison of the predictions and corresponding confidence intervals for a new observation at four nesting levels with the linear multilevel mixed-effects basal area growth model.....	99
Figure 31.	A graphical representing of the nonlinear mixed effects dominant height growth model with a Bayesian reformulation.....	110
Figure 32.	Convergence diagnostics of the Gibbs sampler by modified Gelman-Rubin statistics for the fixed effects of $UCHP$ and the parameters in dominant height growth equation on Plot 1. Values $a[2,1]$, $a[2,2]$ and $a[2,3]$ denote the fixed effects of silvicultural treatment $UCHP$. Values $\beta[1,1]$, $\beta[1,2]$ and $\beta[1,3]$ denote the three parameters in dominant height growth equation for plot 1. The lower line is the average width of the 80% intervals within the individual runs (the intervals are normalized to have maximum value 1, the same as follows), the middle line is the width of the central 80% interval of the pooled runs and the upper line is the ratio (pooled/within).....	115
Figure 33.	Posterior probability distributions of the variance (precision) parameters. Σ_0 , σ_1 , σ_2 and σ_C are the standard errors respectively for fixed effects of silvicultural treatment ($a_{j,0}$, $a_{j,1}$ and $a_{j,2}, j=1, \dots, 11$) and dominant height model.....	116

Figure 34.	Posterior probability distributions of for the three mixed effects parameters (beta's in the graph) on treatment Plot 4 and 5.....	118
Figure 35.	The 95% empirical confidence intervals for fixed silvicultural treatment effects by <i>UCHB</i> ($a[4,1], a[4,2]$ and $a[4,3]$) and <i>FCBH</i> ($a[5,1]$, $a[5,2]$ and $a[5,3]$).....	119
Figure 36.	Autocorrelation plots for the samples of variance (precision) parameters.....	119
Figure 37.	Residual plots for the modified Richards dominant height growth equation with mixed-effects fitted with the Gibbs sampler approach respectively with the mean of samples as estimation (upper) and with the median of the samples as estimation(lower).....	121
Figure 38.	The observed values versus the fitted using the estimated parameters from the median of the Gibbs samples.....	122
Figure 39.	A comparison of “Bound on Error” for different prediction situations (8 examples) in the simultaneous growth and yield model system. “Bound on Error” is defined as half the approximate 95% confidence interval divided by the predicted value.....	161
Figure 40.	A comparison of the predictions and corresponding confidence intervals for slash pine dominant height in the simultaneous growth and yield model system with or without the information on repeated measurements.....	161
Figure 41.	A comparison of the predictions and corresponding confidence intervals for slash pine basal area (in logarithm) in the simultaneous growth and yield model system with or without the contemporary dominant height measured.....	162
Figure 42.	A comparison of the predictions and corresponding confidence intervals for slash pine total stem volume (in logarithm) in the simultaneous growth and yield model system based on 8 distinct situations with different information available.....	162

CHAPTER 1

INTRODUCTION

Reliable prediction of future values of forest stand variables, such as dominant height, basal area and total tree volume, is required for good forest management planning. The traditional way to accomplish this is by fitting growth curves describing the course of stand variables and stand characteristics as functions of time. Predictions are made based on the current or most recent observation of the "predictors" on the right-hand-side (RHS) of the equation. One of the drawbacks of this regression-type prediction is that only part of the observation history is used for the prediction and the underlying individual variability is ignored. It was noted as early as 1961 that observations from a particular plot may reflect an underlying pattern that is particular to that plot and is probably a function of time as well (Clutter, J. L. 1961, "The development of compatible analytic models for growth and yield of Loblolly pine". Duke university, PhD dissertation). Thus, there may well be losses in prediction precision by ignoring these underlying individual variations in the prediction process.

Forestry prediction models are usually based on repeated measurement data that are made in time or space on the same subject or experimental unit, such as a tree or plot. The special properties of such observations include (Lindsey, page 3, 1993): ① the same variable is measured on the same observational unit more than once, thus the responses

are not independent as in the usual regression analysis and ② more than one observational unit is involved, thus the responses do not form a simple time series. In the case of forestry repeated measurement data (usually from monumented plots), there is another property, namely that the measurements of an individual plot (tree) on one occasion are multivariate and these observations are interdependent. These properties present a large challenge to forest growth and yield modelers. For example, traditional regression assumptions (iid assumption) are often violated and new approaches should take these violations appropriately into account in growth and yield modeling.

The philosophy of tree and stand growth usually results in an interdependent system of stand characteristics. As Borders and Bailey (1986) pointed out "*It is apparent that forest growth and yield modelers deal with an interdependent system of underlying growth processes*". The challenge is that the processes may not necessarily be time stationary and sectionally identical. Several people have discussed simultaneous growth and yield systems in the forestry literature, but primary interests have been given to parameter estimation and regression type prediction (Clutter 1963, Furnival and Wilson 1971, Borders and Bailey 1986, Borders 1989, LeMay 1990). Predicting growth and yield for subsequent periods based on all past and current measurements, however, is one of the most important topics in forestry, especially in an inventory updating context. This type of prediction was first addressed in conjunction with dominant height growth by Lappi and Bailey (1988), and further defined and explored with linear and nonlinear mixed models in conjunction with forestry engineering by Liski and Nummi (1995,1996). But, no one to date has investigated this problem in conjunction with a simultaneous stand growth and yield system. This perhaps is partially due to the

complexity of this model system. For example, the compatibility requirement (Clutter 1963), simultaneous biases (Bailey and Borders 1986, Borders 1989), contemporaneous correlation among the responses (LeMay 1990), individually serial correlation, and heteroscedasity (Gregoire ,1987, 1989,1995) are all possible challenges for one to face in stand simultaneous growth and yield modeling.

On the other hand, the special properties of forest growth and yield data also offer some great opportunities to improve the progress of the prediction task in forest management by appropriately modeling the variation and correlation of the multivariate responses in the system. Developments in this area seem more tractable now due to the many contributions made in techniques for the analysis of repeated measurements, including linear and nonlinear mixed models, in the 1990's (Crowder and Hand,1990; Godambe,1991; Jones, 1993, Lindsey,1993; Diggle, Liang and Zeger,1994; Lindsey , 1997; Lindstrom and Bates,1990; Vonesh and Carter, 1992; Davidian and Giltinan , 1995; Mentre and Gomeni, 1995; Wolfinger, 1993; Wakefield et al., 1994; Pinheiro and Bates, 1995; Vonesh ,1996; Diggle and Verbyla, 1998; Vonesh and Chinchilli, 1997), and development in prediction theory (Geisser, 1993, Rao, 1987, Liski and Nummi,1990, 1996, Bondeson 1990, Bondeson and Lanke 1990).

Background and Literature Review

a. *The Nature of the Problem*

The prediction of future observations based on past and current sample values with stand growth and yield models is the focus in this work. This focus is different from

the prediction concept in regression (more often used in forest management) in that prediction is only based on the regressors in that type of model.

Suppose that n individuals (sample units, i.e. plots) are selected and p responses (stand characteristics, such as mean diameter, mean dominant height, basal area per acre, etc.) in each plot are measured on m occasions. Furthermore, q dimensional exogenous variables (such as silvicultural treatments, age of the stand) in each plot are known. Let $y_i(t_k) = (y_{i1}(t_k), \dots, y_{ip}(t_k))'$ denote a $p \times 1$ vector of observed responses of stand characteristics from 1 to p in plot i at time point t_k ($k=1, \dots, m$), $X_i(t_k) = (x_{i1}(t_k), \dots, x_{iq}(t_k))'$ denote a $q \times p$ matrix of predetermined exogenous variables corresponding to p responses in stand i at time point t_k , where $i = 1, \dots, n$. Now, primary interest is given to the prediction of the responses (any or all of the components) of a new sample unit, say the $(n+1)^{th}$ plot, at some time points t_k under a chosen stand growth and yield system. The t_k can be smaller or larger than t_m , the largest age of the stands whose values were observed, but it is larger than the time point t_0 when the last occasion measurements occurred on individual $n+1$; t_0 can be zero, which means there have been no data previously collected for this individual. Thus, with those different situations considered, the prediction problem may be addressed with following cases:

Case 1: Complete prediction.

There are no previous observations for the plot on which prediction of the stand characteristics is required. I call this complete prediction, which means the prediction is totally dependent on population parameters estimated from prior data on the general population being sampled without any data being available for the specific plot or stand.

Case 2: Conditional prediction.

In this case, observations of responses on the plot are partially obtained.

There are some other sub-cases in each case of prediction. In forest practice, it is quite common that on some occasions both endogenous variables and exogenous variables are measured, but on most occasions only exogenous variables are measured. The results of prediction based on these two kinds of pre-observed information will usually be different. This will be discussed in following chapters.

When discussing prediction of future observations, some workers (e.g. Lee, 1988) distinguish two cases:

Case 1 $t_k \leq t_m$, namely, time points at which predictions are required are within the age ranges of stands with response observations and

Case 2 $t_k > t_m$, time points at which predictions are required are beyond (right side) the age ranges of stands with response observations.

The first case is called “conditional prediction” and the second “time serial prediction.”

In the forestry context there will be no differences between these two cases, so I do not distinguish between them.

Table 1 displays the above prediction situations. In general the number of responses for each plot can be different and the ages of different stands (plots) are not necessarily the same. So, missing values should not present any problem for this framework. For example, prediction may be required for a response on an occasion at which all the other exogenous and endogenous variables are measured. This kind of prediction reduces to the ordinary prediction in regression. Another general situation in the above framework is that when $p=1$, i.e., the simultaneous response model system is

reduced to an ordinary univariate model. All of these situations will be addressed in a subsequent model system later in this work.

Even though the formulation outlined in Table 1 shows that only one plot's observation is predicted, prediction of observations for more than one plot is straightforward. For example, measurements are taken on say n_k plots and for the other $n - n_k$ plots one or more observations can be simultaneously predicted. The number of observations to be predicted can also vary from stand to stand.

Table 1. -- The prediction of stand characteristics based on measurements and predetermined exogenous variables at different time points

		Time Points			
Plots	responses	t_1	...	t_m	$t_k > t_m$
1	1	y_{111}, x_{111}	...	y_{11m}, x_{11m}	
:	:	:	:	:	:
1	p	y_{1p1}, x_{1p1}	...	y_{1pm}, x_{1pm}	
:	:	:	:	:	:
:	:	:	:	:	:
n	1	y_{n11}, x_{n11}	...	y_{n1m}, x_{n1m}	
:	:	:	:	:	:
n	p	y_{np1}, x_{np1}	...	y_{npm}, x_{npm}	
<i>case 1:</i>					
$n+1$	1	? , x_{n+111}	...	? , x_{n+11}	? , x_{n+11}
:	:			:	:
$n+1$	p	? , x_{n+1p1}	...	? , x_{n+1p}	? , x_{n+1p}
<i>case 2:</i>					
$n+1$	1	y_{n11}, x_{n11}	...	? , x_{n+11}	...
:	:	:	:	:	:
$n+1$	p	y_{np1}, x_{np1}	...	? , x_{n+1p}	...

Note: "?" denotes the possible point where prediction is required.

It is useful to note that the above prediction problem addressed in a simultaneous model system is not only different from the traditional regression prediction and time

series prediction, but also different from the prediction problem for partially observed vectors so intensively discussed in statistical literature (Lee and Geisser, 1975; Geisser, 1981, 1993; Lee 1988; Rao, 1976, 1985, 1987, Liski and Nummi, 1990, 1996; Valliant 1985). The difference lies in the fact that not only should the sectional effects (inter-individual variation, for example, stand effects, treatment effects) and intra-individual variation (time series effects, for example, heterogeneity of within individual variance, within-individual correlation) be properly addressed in prediction, but also the contemporaneous correlation among the equations in the system should play some key role in the prediction.

b. *Mixed Effects Model Approach*

In the forestry literature, work on the above simultaneous equation system primarily has been addressed on parameter estimation techniques. To date, no one has formulated models from the standpoint of predicting future observations. Since the endogenous variables are used on both LHS and RHS of equations, "simultaneous bias" may result due to the correlation of RHS endogenous variables and the error term of LHS endogenous variable if OLS is directly applied. Also the errors of different equations in the system are likely to be correlated. Some special parameter estimation techniques have been used in forestry literature, such as two-stage and three-stage least squares (Borders and Bailey 1986), modified three-stage least squares (Borders 1989), a simple squared error loss function (Reed 1982, Burkhart and Sprinz 1984), multistage least squares (LeMay 1990), and full information maximum likelihood (Fang and Bailey 1999, Fang, Borders and Bailey 1999).

For prediction purposes, and based on the biological philosophy of stand growth and yield, I will use a mixed effects model approach to the simultaneous equations system. Mixed effects models allow parameters in the model to be varied systematically or randomly from group to group. The unit within a group can be an individual experiment unit (such as plot) or some other nested or crossed group factor (such as soil types, silvicultural treatments, time factors, etc.). This makes mixed effects models very flexible for handling complicated variance-covariance structures which commonly result with longitudinal growth and yield data (Diggle, et al 1994). Depending on the feature of the responses, linear or nonlinear mixed effects models may be used for the components of the responses in the system. For example, dominant height is often modeled as a nonlinear function of stand age and basal area in a log-linear form (e.g Borders 1989). In such cases, a nonlinear mixed effects model may be required for the dominant height component and a linear mixed effects model for a basal area component. Multilevel random effects models [see Goldstein (1986), Longford (1993), Karlsson and Sheiner (1994), Gilks et al. (1993), and Hall and Bailey (1997)] may be more appropriate for data from a multi-level nesting design, for example, a split-plot design.

In subsequent chapters I will discuss a general strategy of model building for nonlinear mixed-effects models and linear multi-level mixed-effects models in conjunction with longitudinal data from permanent plots within a designed experiment to study intensive forest management practices. New response predictions will be the main focus for the established model (system). Individual components (stand characteristics) will be discussed first and then a simultaneous model system combining all the components will be presented.

Objectives

In summary, the **objectives** of this study are:

- To advance the use of nonlinear mixed effects and linear multi-level mixed effects modeling techniques in forest growth and yield modeling.
- To build a forest growth and yield model system in conjunction with intensive management practices such as herbicide and fertilization.
- To present basic univariate and multivariate prediction theory that is applicable to the forestry inventory prediction problem.
- To represent a nonlinear mixed effects model in the Bayesian framework with a Gibbs sampling approach.
- To justify the advantages in predicting an unknown response value with a mixed effects model with a univariate or simultaneous multivariate model system by showing examples in detail.

CHAPTER 2

NONLINEAR MIXED EFFECTS MODELS FOR THE ANALYSIS OF REPEATED MEASUREMENTS FROM PERMANENT PLOTS IN A DESIGNED EXPERIMENT: A STUDY OF SLASH PINE DOMINANT HEIGHT GROWTH FOLLOWING INTENSIVE SILVICULTURAL TREATMENTS

Introduction

Permanent sample plots, measured repeatedly over a fixed length of time, are often used for evaluating forest growth and yield and are especially effective as a sampling method to evaluate changes in forest conditions (Avery and Burkhart, 1994). In the practice of intensive forest management, some new attributes are added into the usual permanent plot sampling method. For example, to accelerate stand growth and development and to increase financial returns, silvicultural treatments such as mechanical and chemical site preparation, herbicide or fertilization applications are very common. To monitor and predict the changes of forest stands with different silvicultural treatments or factors, the permanent plots are usually established in a more careful way: they are actually built up from some standard experimental design. For example, a split-plot design is often used in the Southeastern United States. In one such designed study, soil type serves as the whole-plot factor and silvicultural treatment as the split-plot factor (Shiver et al 1994). A split-plot design with repeated measurements naturally forms a

split-split plot design with the time factor as the within-plot effect (Gumpertz and Brownie 1993). Data from a permanent plot study with split-plot design include the following two basic properties: ① There is usually more than one source of error. The whole plot, split plot, and time factors are all likely to introduce errors into the model. ② Samples from the same plot (including whole plot, split-plot and time) may not be independent. For forest biometrists, when the growth and yield curves are the focus of the analyses, another four characteristics that are common in growth curves (Lindsey page 69, 1997) should also be taken into account: ① The growth profile will generally be a nonlinear function of time, often with an asymptote. ② Growth is not necessarily stationary. ③ Heteroscedasty is very likely, i.e. random variability will generally increase with size. ④ Different individuals may have different growth rates or profiles, either inherently or due to environmental effects.

In the forestry literature, a common approach to modeling such split-plot repeated measurement data from permanent plots is to only “adjust” the fixed part of the model with some additional fixed terms that partially explain the “gains” from the silvicultural treatments(Pienaar and Rheney 1995, Martin and Bailey 1998, Castleberry 1998). The advantage of this approach is its simplicity. It allows any treatment effects on forest growth and yield to be explicitly expressed in the model. However, the special properties of split-plot repeated-measurement data as discussed above are partially or totally ignored by this approach. For example, the variability of individual plots or within-plot correlations are rarely taken into account in such an approach. Consequently, different plots, no matter how different their plot attributes, will obtain exactly the same “gains” as long as they are treated as the same silvicultural treatments. This is, however,

usually not realistic. Also, any information contained in past observations for a plot or stand is only partially applicable in this approach when predictions are desired for expected future values for a specified plot.

Gregoire et al (1995) recently presented a linear mixed effects model based on data from permanent plots and justified the necessity of an appropriate variance-covariance for the modeling of such data. While linear models can be made robust to assumptions about the correlation, especially when the number of observations per plot are small relative to the number of plots (Diggle et al. ch. 4, 1994), this is not true for the nonlinear models. The expectation of an individual mean response usually does not coincide with the marginal mean of the population for a nonlinear model (Zeger et al 1988, Vonesh and Chinchilli, page 295, 1997). Thus it is obviously inappropriate to use a population mean response as the prediction for an individual whose former information is available.

Data from the same plot are very likely to be correlated and observations from a particular plot may reflect an underlying pattern that is particular to that plot. This pattern may be a function of time as well (Clutter, 1961). Thus, a longitudinal data analysis approach (Diggle et al 1994), such as with a mixed effects model, is appropriate for the analysis of forest growth and yield data from permanent plots with any special experimental design.

There are two basic purposes for establishing permanent plots with a special experimental design. One is to evaluate the general effects of different silvicultural treatments on forest growth and yield, and the other is to predict some future forest growth and yield values in conjunction with different silvicultural treatments. In this

paper, I use repeated measures of slash pine (*Pinus elliottii* Engelm.) dominant height from the a permanent plot study with a split-plot design to show how these two objectives can be reached by using a nonlinear mixed effects model. I also present he techniques of model building and diagnosis for nonlinear mixed effects models with consideration for the special nature of the data.

Data

The data used in this study come from a designed study for slash pine (*Pinus elliottii* Engelm.) installed by the Plantation Management Research Cooperative (PMRC) of the University of Georgia. The purpose of this study is to evaluate growth, yield and stand structure of slash pine plantations with different site preparation treatments alone and in combination with fertilization and vegetation control (Pienaar, Shiver and Harrison 1998). The original design used in the study is a split-plot with installation locations as whole-plots and the permanent plots as the sub-plots. The whole-plot factor is the soil type and the sub-plot factor is the silvicultural treatment. Eleven (11) treatment plots were established on four types of soil, each of which was assigned randomly to 20 installation locations, across the flatwoods of Georgia to north Florida. Some common stand characteristics such as dominant height and basal area have been measured up to 5 times on each treatment plot from age 5 to age 17 with 3-year intervals since establishment in 1979. Seventeen years after installation, 16 of the original installations remain for the analysis. These comprise seven nonspodosol and nine spodosol soil groups. The total number of treatment plots is 191. Details of the data description were given by Pienaar, Shiver and Harrison (1998). In the current analysis, I only focus on

the response of dominant/codominant height. The basic statistics of the data are summarized in Table 2. Dominant height growth profiles by different treatments and soil groups are presented in Figures 1 and Figure 2 respectively. The mean dominant heights are compared with different design factors (Figure 3).

Table 2. -- Basic statistics for dominant height (in meters) by individual factor levels for slash pine based on permanent experimental plots from the Plantation Management Research Cooperative (PMRC) of The University of Georgia.

Age		5 yrs.		8 yrs.		11 yrs.		14 yrs.		17 yrs.		
<u>Treatments¹</u>		Dominant height (m)										
Trt.	n	Mean	std.	Mean	std.	Mean	std.	Mean	std.	Mean	std.	
CNTL	16	2.57	0.77	5.31	1.32	7.87	1.65	10.84	1.95	13.13	2.14	
FBHB	18	5.17	0.43	8.87	0.57	11.81	0.73	14.93	0.98	17.29	1.20	
FCBB	17	4.01	0.51	7.39	0.87	10.35	1.13	13.67	1.33	16.07	1.78	
FCHB	16	4.48	0.39	8.23	0.51	11.26	0.78	14.41	1.12	16.89	1.10	
FCHP	18	3.40	0.80	6.55	1.15	9.53	1.50	12.76	1.69	15.69	1.91	
UBHB	18	4.82	0.61	8.45	0.70	11.50	0.86	14.40	1.12	16.46	1.40	
UCBB	19	3.32	0.68	6.65	1.04	9.53	1.32	12.48	1.46	14.82	1.86	
UCBH	17	4.24	0.75	7.83	0.91	10.84	1.13	13.88	1.51	15.95	1.80	
UCHB	18	2.62	0.70	5.68	1.13	8.51	1.53	11.62	1.85	13.87	2.30	
UCHP	17	2.74	0.60	5.71	0.91	8.38	1.15	11.40	1.53	13.58	1.73	
<u>Soil groups²</u>												
SoilGp	n	Mean	std.	Mean	std.	Mean	std.	Mean	std.	Mean	std.	
A	36	3.68	0.87	7.08	1.21	10.18	1.29	13.35	1.60	15.60	2.08	
B	48	3.88	1.00	7.25	1.51	10.16	1.86	13.27	2.04	15.73	2.27	
C	47	3.53	1.22	6.82	1.69	9.65	1.94	12.78	2.03	15.21	2.11	
D	60	3.68	1.08	6.91	1.44	9.72	1.68	12.74	1.89	15.03	2.05	
<u>Overall</u>		191	3.69	1.06	7.01	1.48	9.90	1.73	13.00	1.92	15.36	2.13

¹Treatments: CNTL:Control (harvest and plant, no site preparation); UCHP:Chop (single pass with a rolling drum chopper); FCHP Chop, fertilize; UCHB: Chop, burn (chop followed by a broadcast burn) ; FCHB: Chop, burn, fertilize; UCBB: Chop, burn, bed; FCBB: Chop, burn, bed ,fertilize; UCBH: Chop, burn, herbicide; FCBH: Chop, burn, herbicide, fertilize; UBHB: Chop, burn, bed, herbicide; FBHB: Chop, burn, bed, herbicide, fertilize.

² Soil type: A: Poorly drained, nonspodosol; B: somewhat poorly to moderately well drained, nonspodosol; C: Poorly to moderately well drained spodosol with an underlying argillic horizon; D: Poorly to moderately well drained spodosol without an underlying argillic horizon.

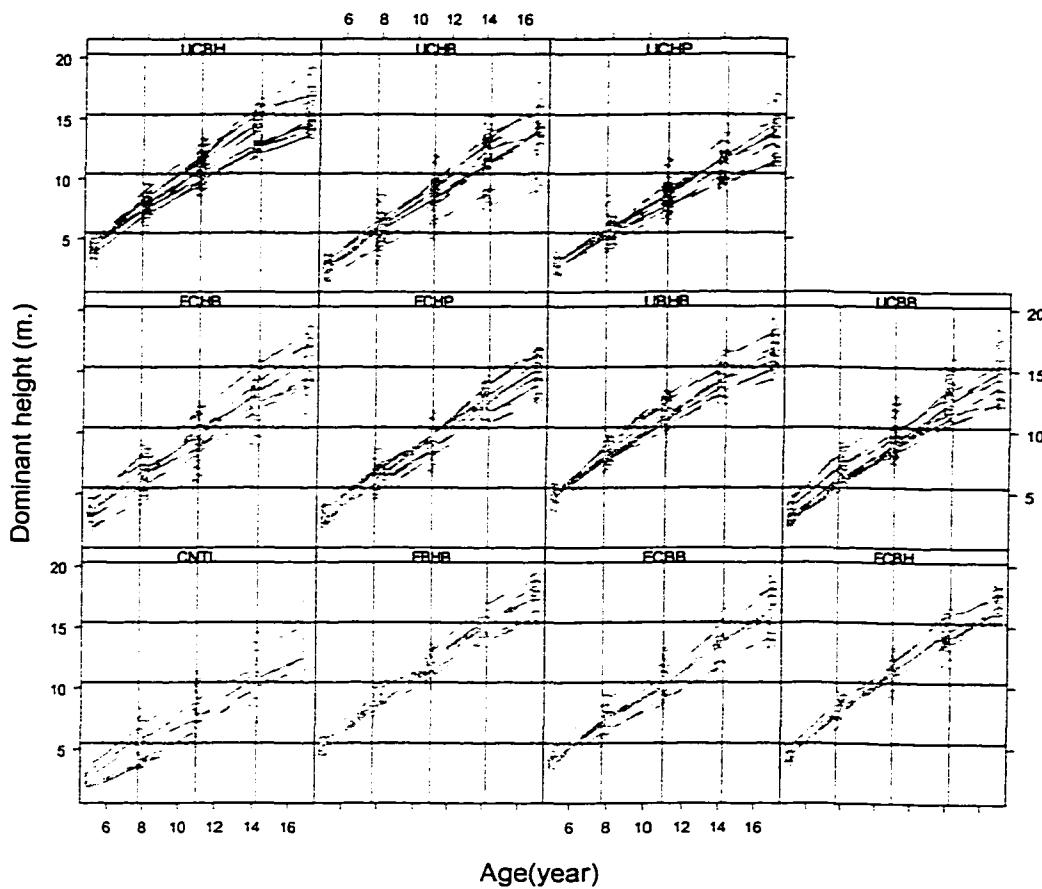


Figure 1: Profile graph of slash dominant height growth on 191 plots with 5 repeated measures by 11 silvicultural treatment groups.

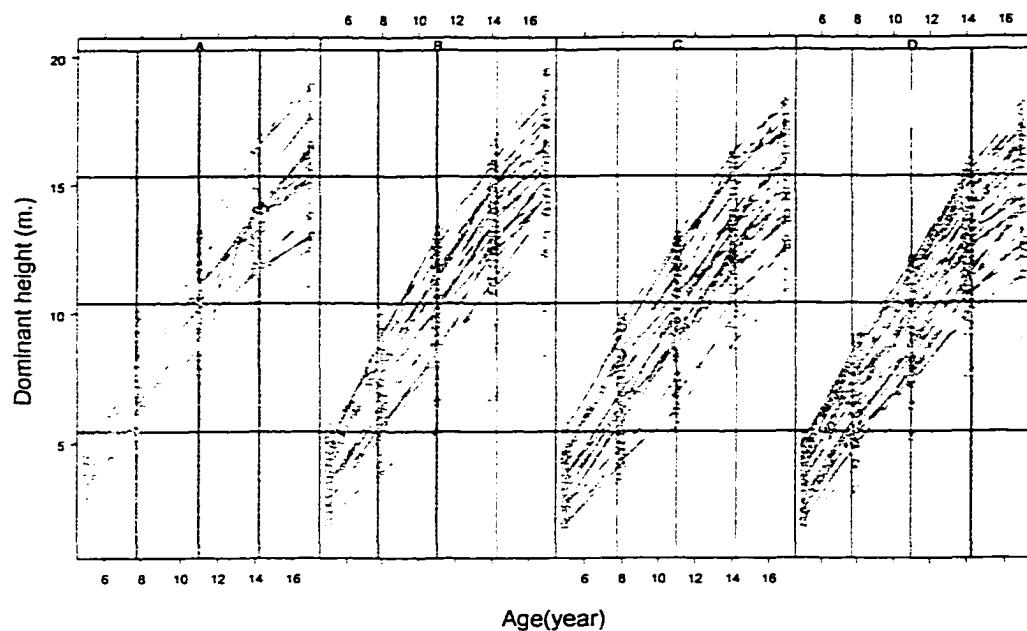


Figure 2: Profile graph of slash dominant height growth on 191 plots with 5 repeated measures by 4 soil types.

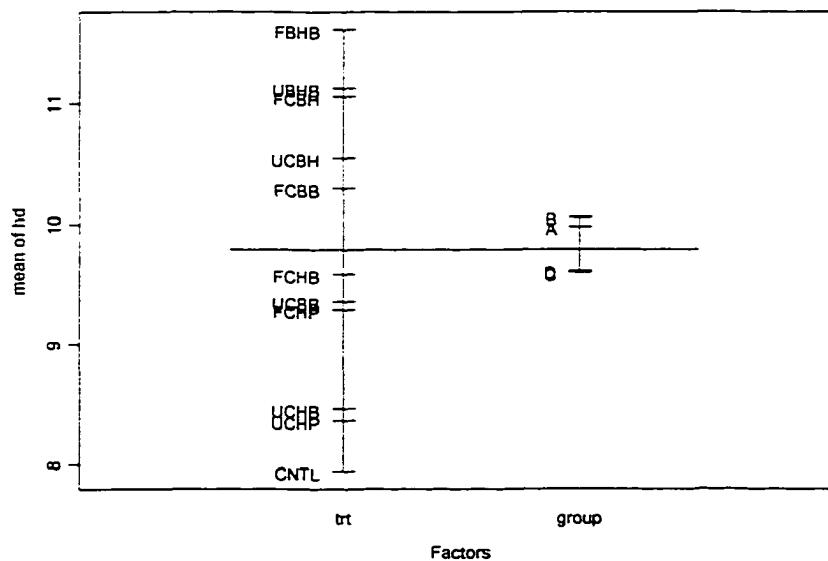


Figure 3: The mean responses of slash pine dominant height growth by silvicultural treatments and soil types.

Method

The general dominant height growth equation

The well-known three-parameter Richards' equation (1959) serves as the basic dominant height growth model,

$$f(t) = A(1 - e^{-\beta_1 t})^{\beta_2} \quad (2.1)$$

where t is stand age, $f(t)$ is the mean response function of dominant height, A is the asymptote parameter which denotes the asymptotic value of the dominant height, and β_1 is the rate parameter. Since A is the most unstable parameter to estimate with our data, it was replaced by an expected-value parameter, β_0 , corresponding to the expected height when $t = t_0$. This allows equation (2.1) to be represented as:

$$f(t, \beta) = \beta_0 \left(\frac{1 - e^{-\beta_1 t}}{1 - e^{-\beta_1 t_0}} \right)^{\beta_2} \quad (2.2)$$

Equation (2.2) has exactly the same form as an algebraic difference project model (Borders and Bailey 1984), but with one big difference in interpretation. In (2.2), β_0 is treated as an unknown parameter (rather than a measured height) and t_0 is treated as a given age which may not necessarily be included in the data. There is no limit on t_0 as long as it is positive, although a reasonable t_0 should be chosen within the range of the data. The properties of equation (2.2) can be summarized as follows:

- ① β_0 is the expected dominant/co-dominant height at given age t_0 , thus, if t_0 is taken as the site index reference age (25 yrs. for slash pine) then β_0 corresponds to the site index; if t_0 is taken as a reasonably large value, then β_0 corresponds to the asymptote and (2.2) has the same form as (2.1).

- ② For a given data set, the estimated values of β_1 and β_2 will be unchanged for different choices of t_0 , i.e. changing t_0 only affects the estimate of β_0 ;
- ③ The asymptotic height is $\beta_0/[1 - \exp(-\beta_1 t_0)]^{\beta_2}$.

The advantages of models with expected-value parameter(s) were discussed by Schnute (1981), Ratkowsky (1990), Cieszewski and Bella(1989), and Cieszewski and Bailey (1999).

Nonlinear mixed effects model

Equation (2.2) is a mathematical model that has been proposed deterministically. To apply it to the real data, various statistical models will result in different statistical assumptions. Let y_{ij} denote a dominant height measurement at occasion j for the i th plot ($i=1, \dots, n; j = 1, \dots, m_i$). m_i is the total number of measurements on plot i . The stand age for plot i at occasion j is t_{ij} . The function f (Model 2.2) is common to all plots, but the parameters may vary across plots. In vector form, let $y_i = (y_{i1}, \dots, y_{im_i})^T$, $e_i = (e_{i1}, \dots, e_{im_i})^T$, $t_i = (t_{i1}, \dots, t_{im_i})^T$ and $\beta_i = (\beta_{i0}, \beta_{i1}, \beta_{i2})^T$. There are some stand characteristics corresponding to an individual plot, such as installation number of the plot, the soil group, and the silvicultural treatment, which have not been explicitly expressed thus far, but they will be addressed in the section on model specification. I adopt the two-stage model formulation (Lindstrom and Bates 1990, Davidian and Giltinan 1995) in order to write the general nonlinear model as:

Stage 1 within-plot variation

$$\left. \begin{array}{l} y_i = f(t_i, \beta_i) + e_i \\ e_i \mid \beta_i \sim iid N[0, R_i(\beta_i, \alpha, \theta)] \\ \mu_i = E[y_i \mid \beta_i] = f(t_i, \beta_i) \\ f(t_i, \beta_i) = \beta_{0i} \left(\frac{1 - e^{-\beta_{1i} t_i}}{1 - e^{-\beta_{1i} t_0}} \right)^{\beta_{2i}} \end{array} \right\} \quad (2.3)$$

In (2.3), $R_i(\beta_i, \alpha, \theta)$ is a $m_i \times m_i$ within-plot variance-covariance matrix, which may depend on parameter β_i (via the mean function) and some other covariance parameters θ . From Davidian and Giltinan (1995), one may formalize the within-individual variance-covariance as:

$$Cov(e_i \mid \beta_i) = R_i(\beta_i, \alpha, \theta) = \sigma^2 G_i^{1/2}(\beta_i, \alpha) \Gamma_i(\theta) G_i^{1/2}(\beta_i, \alpha) \quad (2.4)$$

where the $m_i \times m_i$ diagonal matrix $G_i^{1/2}(\beta_i, \alpha)$ specifies within-plot variance and the $m_i \times m_i$ matrix $\Gamma_i(\theta)$ describes the correlation pattern within-plot i . Both heteroscedasticity and correlation of the intra-plot error are considered by this formulation.

Stage 2 inter-plot variation

$$\left. \begin{array}{l} \beta_i = A_i \beta + B_i b_i \\ b_i \sim N(0, D_i) \end{array} \right\} \quad (2.5)$$

In (2.5), A_i is a $3 \times p$ design matrix for $p \times 1$ fixed effects β , b_i is the $p_i \times 1$ vector of random effects associated with the i^{th} plot, and B_i is the corresponding design matrix. D_i is a $(p_i \times p_i)$ covariance matrix for the random effects, which is assumed to be common across all individuals except for the dimension. Missing values and unbalanced data are allowed in this formulation. The design matrix A_i and matrix B_i are determined by the stand characteristics, such as silvicultural treatments and soil types, which are usually, but not necessarily, matrices containing only zeroes and ones as elements.

Equations (2.3) to (2.5) represent the general formulation of a nonlinear mixed effects model based on the deterministic dominant height growth equation (2.2).

Different models can be obtained with distinct variance-covariance specifications. In the following section, I discuss the advantages of this two-stage hierarchical formulation in the process of model specification.

Model specification

To apply the model represented by (2.3)-(2.5) with a real set of data, one needs to:

- ① Specify the nature of the three parameters $(\beta_0, \beta_1, \beta_2)^T$, as fixed and random effects or purely fixed effects;
- ② Determine an appropriate variance-covariance structure for the individual plot [the structure and components of $R_i(\beta_i, \xi)$];
- ③ Choose appropriate covariates to explain inter-plot parameter variability (i.e., determine the dimension and components of the fixed effects β and random effects and the corresponding design matrix A_i and B_i).

The first two issues above arise naturally in the first phase of model formulation. But, the last, which is of most concern by a model user, should be answered before addressing the questions presented by the original research, such as silvicultural treatment effects tests, prediction problems, etc. Corresponding to the above two-stage hierarchical formulation, I address these three problems in the following three steps.

Step 1: Determine parameter effects

Which effects should be considered as mixed (both fixed and random) and which should be considered as purely fixed in modeling, are generally data dependent. For example, Lappi and Bailey (1988) chose the rate parameters β_1 as fixed while the asymptote parameter and β_2 were considered as random when they used Richard's equation to model dominant height growth. They did so simply because it is impossible to obtain reasonable estimates with all parameters random due to collinearity problems in their data. Intuitively, one simple approach to this question is to obtain separate fits for each plot and assess the variability of estimated parameters across plots by considering the individual confidence intervals for the parameters. The parameters with high variability and less overlap in confidence intervals across plots should be considered as mixed effects. This approach requires sufficient observations on each plot to give meaningful parameter estimates by separate fitting. However, in forestry, repeated measurements from permanent plots usually do not cover a long enough time span or contain enough degrees of freedom to produce stable individual parameter estimates. In our case there are only 5 observations on each plot and 3 parameters to be estimated. In this situation, the separate-fittings approach is not likely to be helpful in judging the nature of the parameters.

If no prior information about the random effects variance-covariance structure is available and convergence is possible, Pinheiro and Bates (1998) suggested that all parameters in the model should first be considered mixed (both random and fixed). After the initial estimates are available, the eigenvalues of the estimated covariance matrix of the effects (D matrix) may be studied to see if one or more are close to zero.

The associated eigenvector(s) for such eigenvalues will then give an indication of the linear combinations of parameters that could be taken as fixed. Since small eigenvalues may arise when the relative magnitudes of the scales of the parameters in the model are quite different, the “normalized” variance-covariance matrix i.e. the *coefficient of variation (CV)* matrix D_{cv} was recommended instead of the variance-covariance matrix (Pinheiro and Bates (1998).

$$[D_{cv}]_{ij} = \frac{[D]_{ij}}{|\beta_{k(i)}\beta_{k(j)}|} \quad (2.6)$$

In (2.6), β_k is the k^{th} fixed effect and $k(i)$, $k(j)$ are the fixed effects associated with the i^{th} and the j^{th} random effects. This approach is still an ad hoc method, for there is no decision rule to indicate how close an eigenvalue may be to zero and not indicate rank deficiency.

Alternatively, once the larger model (for example, with all the parameters as mixed) is fitted, one actually can evaluate it with some smaller (reduced) model using a *likelihood ratio test (LRT)* or information criterion statistics, such as *Akaike Information Criterion (AIC)*. Suppose L_0 is the likelihood of the more general model (e.g. the model with all parameters mixed) and L_1 is the likelihood of the restricted model (e.g. some parameters are purely fixed), and the total number of estimated parameters of these two models are respectively k_0 and k_1 . Then the likelihood ratio test statistic is defined as:

$$LRT = 2\log(L_0 / L_1) = 2 [\log(L_0) - \log(L_1)] \quad (2.7)$$

Under the null hypothesis that the restricted model is adequate, the asymptotic (large sample) distribution of LRT is a χ^2 with $k_0 - k_1$ degree of freedom. If $LRT > \chi^2(k_0 - k_1, \alpha)$ then one rejects the null hypothesis with $(1-\alpha)\%$ confidence that the full model

should not be reduced to the restricted model. Here α is some specified percentile of the Chi-squared distribution.

The *Akaike Information Criterion (AIC)* is defined as:

$$AIC = -2 \log(L) + 2k \quad (2.8)$$

Where, L is the likelihood value of the model with k parameters. The *AIC* is often used to compare models with alternative sets of fixed-effects and covariance parameters, especially non-nested models for which the likelihood ratio test is inappropriate (Sakamoto et al. 1986, Gregoire et. al. 1995).

Following Maitre et al(1991), in this first step I didn't consider any covariates on the mixed effects. Suppose

$$\beta_i = \beta + b_i \quad (2.9)$$

where β_i is a 3×1 vector of the mixed effects in the nonlinear dominant height Model (2.3), i.e. $\beta_i = (\beta_{i0}, \beta_{i1}, \beta_{i2})^T$, β and b_i are respectively the corresponding vector of fixed and random effects for plot i ($i = 1, \dots, n$, indexing individual treatment plot). $b_i \sim N(0_{3 \times 1}, D_{3 \times 3})$.

It is worth noting that the inter-plot variation Model (2.9) is just a special case of the general formulation of inter-individual variation [see (2.5)], in which no covariate was put into the random effect. In other words, it is simply a random coefficient model and both design matrices A_i and B_i in (2.5) are 3×3 identity matrices in this case. When all three parameters are mixed, the estimated coefficient of variation matrix D_{cv} is³:

³ Maximum likelihood method was used in the parameter estimation.(ref. Pinheiro and Bates, 1999). t_0 in (2.3) was taken as $t_0 = 25$ year all through this study.

$$\hat{D}_{cv} = \begin{pmatrix} \hat{\sigma}_0^2 / \hat{\beta}_0^2 & \hat{\sigma}_{01} / |\hat{\beta}_0 \hat{\beta}_1| & \hat{\sigma}_{02} / |\hat{\beta}_0 \hat{\beta}_2| \\ \hat{\sigma}_{10} / |\hat{\beta}_1 \hat{\beta}_0| & \hat{\sigma}_1^2 / \hat{\beta}_1^2 & \hat{\sigma}_{12} / |\hat{\beta}_1 \hat{\beta}_2| \\ \hat{\sigma}_{20} / |\hat{\beta}_2 \hat{\beta}_0| & \hat{\sigma}_{21} / |\hat{\beta}_2 \hat{\beta}_1| & \hat{\sigma}_2^2 / \hat{\beta}_2^2 \end{pmatrix} = \begin{pmatrix} 0.017378 & -0.00732 & -0.00427 \\ -0.00732 & 0.00970 & -0.00709 \\ -0.00427 & -0.00709 & 0.01302 \end{pmatrix}$$

The eigenvalues of \hat{D}_{cv} are 0.02181, 0.018286 and 5.2014×10^{-6} . There is some evidence of rank deficiency. But the evidence is not so strong. The eigenvector corresponding to the smallest eigenvalue, converted back to the original scale of the random effects and normalized, is -0.002193, -0.99948 and -0.03201. Even though both the first and the third component of this eigenvector are very small, it is hard to conclude that they are close to each other because one is more than ten times the other in magnitude. So a formal likelihood ratio test is necessary in this case to determine the effects of the parameters. I first took all the three parameters as mixed effects without considering any covariates (this is referred as the full model), and then picked each of the three parameters as purely fixed in turn. This resulted in three reduced models. Since the reduced models are nested within the full model, likelihood ratio tests can be carried out to check if the reduction in parameters caused any significant changes in model performance. Replacing any mixed effects parameter with a purely fixed effect significantly lowered the likelihood values (Table 3). Therefore, preference must be given to all three parameters being considered as mixed effects. Of course, results may be different with different covariates for the fixed effects and variance-covariance structure for the random effects, but the above procedure should be the same. It is interesting to note that changing β_0 parameters to fixed effects has a much bigger impact on the likelihood than doing so with either of the other two parameter sets.

Table 3 Likelihood ratio tests (LRT) for nonlinear mixed-effects models of slash pine dominant height with different fixed and random effects components

Model mixed effects	Number [*] fixed only	AIC	log-likelihood ^{**}	LRT	p-value
$\beta_0, \beta_1, \beta_2$	none	10	-855.593		
β_0, β_1	β_2	7	1731.185	-885.992	60.799 <.0001
β_1, β_2	β_0	7	2160.942	-1056.455	401.72 <.0001
β_0, β_2	β_1	7	1739.007	-862.504	13.822 <.0032

The number of parameters here includes both the parameters in the mean function (3 in this case) and variance-covariance (6 for the full model and 3 for the reduced model) and add another deviance parameters σ .

^{**} Maximum likelihood method was applied in parameter estimation.

Step 2: determine within plot variance-covariance structure

To specify the within plot variance-covariance structure, as implied in (2.4), two components should be addressed, one is the heteroscedasticity and the other is autocorrelation structure. Forest growth and yield data from permanent plots usually exhibit both autocorrelation and heteroscedasticity (Gregoire 1995).

1. variance function

The variances of errors around growth and yield models are often found to be dependent on the means, larger means usually have larger variance. This trend can be clearly observed from the residuals plotted after an OLS fitting of Model (2.2) to the slash pine dominant height growth data (Figure 4). Some frequently used variance functions for growth modeling include (Davidian and Giltinan 1995) the power function model,

$$g(\mu_{ij}, \alpha) = \mu_{ij}^\alpha , \quad (2.10)$$

and the exponential function model,

$$g(\mu_{ij}, \alpha) = \exp(\alpha \mu_{ij}) , \quad (2.11)$$

where μ_{ij} is the mean function which is defined in (2.3).

Note that both of the above functions imply that variance of responses depends on the regression parameters through the means. Random effects in the mean function may remove some heterogeneity in variance. This is not hard to understand. Since the variances depend on the mean response through the parameters, random effects in parameters will definitely affect the underlying distribution of error of the model and thus the variance. Compared to Figure 4, the residual plot of the full model [all three parameters are taken as mixed effects, the mixed effects are defined in (2.9)] is much more homogeneous (Figure 5). Are the residuals of the full model homogeneous enough? One can make a formal test based on the residuals from the fitting. Alternatively, we can still use a likelihood ratio test to answer this question. We may try a different variance function such as the power or exponential function [see (2.10) and (2.11)], and compare the likelihood values of different approaches by LRT (if the alternative models are nested) or AIC (if not nested).

The fitting comparison of a full mixed-effects model with both exponential and power variance functions indicates that the LRT tests are significant for both of these variance functions (Table 4). So even with random effects in the parameters, heteroscedasticity still exists in the mixed-effects dominant height model. With a power function to model the variance, a smaller AIC value results than with an exponential function. Since both have the same number of parameters in the model, the power function is judged superior in this case.

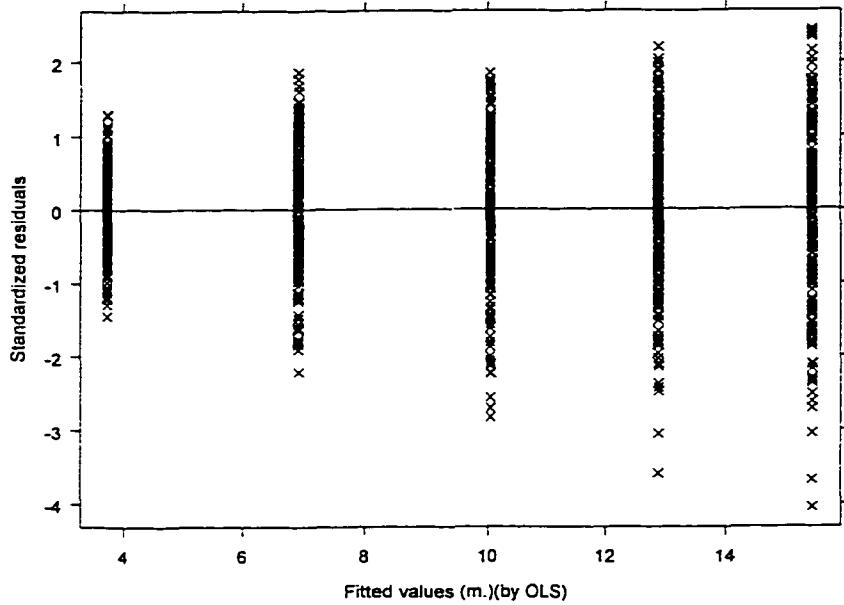


Figure 4: The residual plot by the modified Richards dominant height growth equation with ordinary least square (OLS) estimations.

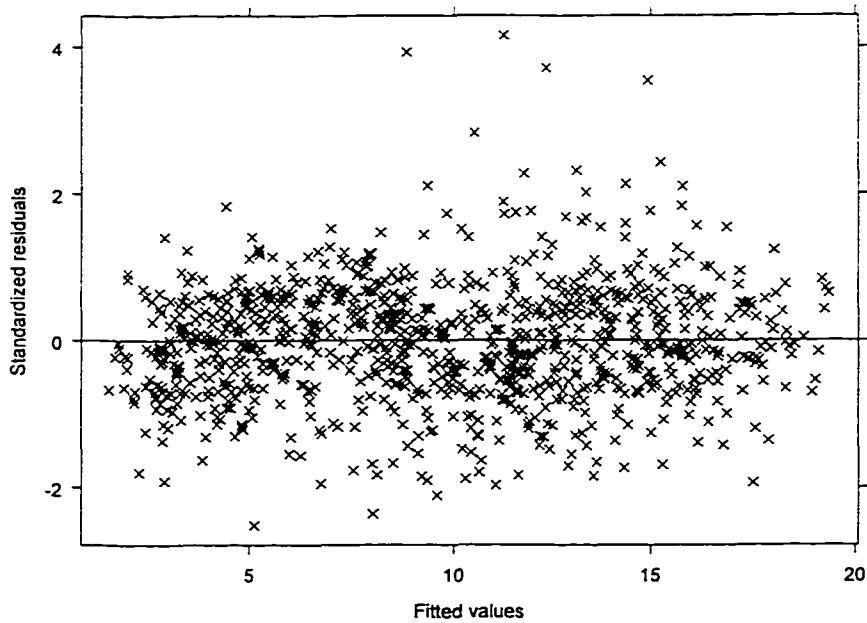


Figure 5: The residual plot by the modified Richards dominant height growth equation with random between-plot effects and independent homogenous within-plot errors.

Table 4 Comparison of mixed-effects model performance for slash pine dominant height growth data with different variance functions

Model	Variance Function	Number of parameters	AIC	log-likelihood	LRT*	p-value
1	1 [†]	10	1731.185	-855.593		
2	Exponential	11	1703.053	-840.527	30.132	<.0001
3	Power	11	1697.299	-837.650	35.886	<.0001

[†] Variance function 1 means that the variances are homogeneous

** Likelihood ratio is calculated with respect to Model 1.

2. Serial correlation structure

Forest growth and yield data from permanent plots may be correlated and, thus, some adjustment in regression techniques might be necessary (Leak 1966, Gregoire 1987, Omule and MacDonald 1991, Flewelling and Jong 1994, Lappi and Malinen 1994). However, correlation problems in such data have received little attention in the forestry literature. Gregoire et. al. (1995) summarized two main reasons for this. Since the main purpose of forest growth and yield modeling is for prediction and it is well known that OLS is inefficient but still consistent under correlated errors, the predicted response using OLS is still valid. The other reason that modeling the longitudinal aspect of permanent plot measurements is less popular in forestry is the complexity of recommended approaches. Gregoire et al.(1995) present strategies for permanent plot measurements with linear models. The complexities of the model with serial correlation arise mainly from two aspects. One is that the correlation parameters are not so easy to estimate. For instance, for data with a large number of individuals but only a small number of repeated occasions, as is often the case in forestry, stable estimation of the correlation parameters is almost unattainable (e.g. Diggle et al. 1994).

Many factors may affect the within-individual correlation structure. In the hierarchical Model (2.3-2.5), for example, the correlation among observations on a given

individual at the marginal level will depend on both inter- and intra-individual covariance matrix (Davidian and Giltinan 1995). Consequently, within-individual correlation may be affected by some other factors. Without additional information or assumptions, it may not be possible to determine them.

My attitude toward the with-in plot correlation is that checking the correlation structure is always informative. If some correlation structure other than independence gave a better fit to the data, we should keep this structure in the model.

The empirical auto-correlation function (Box et. al. 1994) can be used to investigate serial correlation. Let r_{ij} denote the standardized residuals from a fitted mixed model for individual i ($i = 1, \dots, n$) at occasion j ($j = 1, \dots, n_i$). The empirical auto-correlation at lag l is defined as (also see Bates and Pinheiro, 1999):

$$\hat{\rho}(l) = \frac{\frac{1}{N(l)} \sum_{i=1}^n \sum_{j=1}^{n_i-l} r_{ij} r_{i(j+l)}}{\frac{1}{N} \sum_{i=1}^n \sum_{j=1}^{n_i} r_{ij}^2}, \quad (2.12)$$

where $N(l)$ is the number of residual pairs used in the calculation for $\hat{\rho}(l)$ and

$N = \sum_{i=1}^n n_i$ is the total number of the observations.

An approximate two-sided critical value for auto-correlation $\hat{\rho}(l)$ at significance level α is given by $z_{(1-\alpha/2)} / \sqrt{N(l)}$, where $z_{(1-\alpha/2)}$ is the standard normal quantile at percentile $1 - \alpha/2$ and $N(l)$ is defined in (2.12).

The estimated empirical correlation for Model 3 in Table 4 is

$$\hat{\rho} = [\hat{\rho}(1), \hat{\rho}(2), \hat{\rho}(3), \hat{\rho}(4)]^T = [-0.4895, -0.1859, -0.1731, 0.4014]^T$$

A plot of estimated auto-correlation against lags with critical values shows that the auto-correlations are significant even at 4 lags (Figure 6). But the autocorrelation

looks “strange” in that the first three lags are all negatively correlated and the fourth lag is significantly positively correlated. This is not consistent with our intuition, because one usually expects that observations spaced closely in time should be positively associated. However, such a counter-intuitive autocorrelation is not uncommon in the real world (e.g. Davidian and Giltinan 1995). This is an excellent example to show the complexity of within-plot auto correlation. Knowing that the empirical auto-correlation is significant, it is necessary to modify Model 3 (in Table 4) to reselect this within-plot auto-correlation.

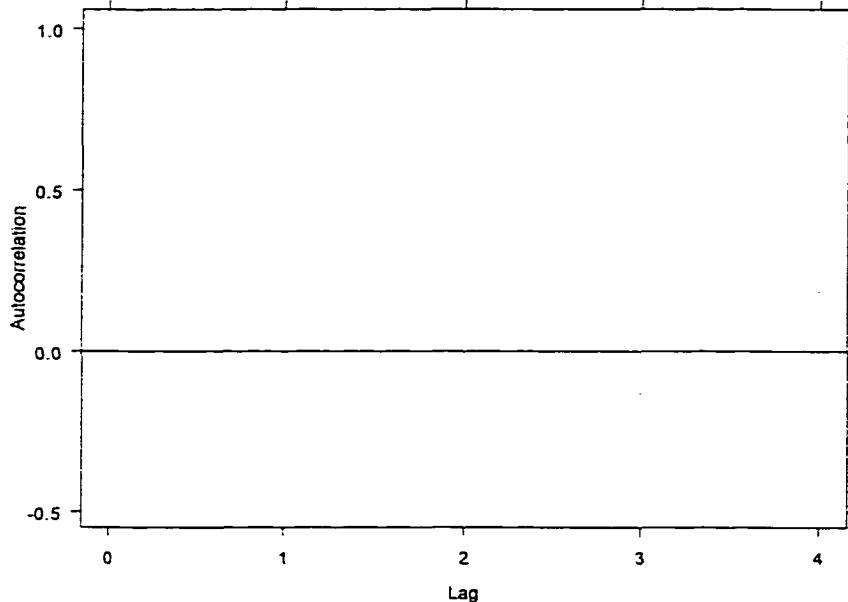


Figure 6: Empirical autocorrelation function corresponding to the standardized residuals of the modified Richards dominant height growth model with random effects and *iid* within-plot errors.

One of the most commonly used within-individual auto-correlation structures with repeated measurement data is $AR(1)$ (i.e., autoregressive model with order 1)(Jones 1993, Gregoire et al. 1995). However, there was no significant improvement when I

included an *AR(1)* auto-correlation structure with mixed effects Model 2.2 (see Table 5). Therefore, an inappropriate correlation structure can result in no improvement in fitting even though there is some evidence of correlation in the error. This may partially explain why several authors in the forestry literature concluded that modeling the within-plot autocorrelation produced no improvements in forest growth and yield prediction (see for example, Sullivan and Clutter 1972). After trying several models, I found that a moving average correlation model was the best for improvement in the mixed effects version of Model 3 (see Table 5).

Table 5 Comparison of mixed-effects model performance for slash pine dominant height growth data with different within-plot correlation structures and no inter-plot covariates

Model	Correlation [*] Structure	Number of parameters	AIC	log- likelihood	LRT ^{**}	p- value
3	independent	11	1697.299	-837.650		
	<i>AR(1)</i>	12	1697.092	-836.544	2.2078	0.1373
	<i>ARMA(1,1)</i>	13	1695.872	-834.936	5.4278	0.0663
4	<i>MA(2)</i>	13	1694.808	-834.404	6.4919	0.0389

^{*} refer to Pinheiro and Bates 1999;

^{**} Likelihood ratio is calculated with respect to Model 3.

Moving average correlation models assume that the current observation is a linear function of independent and identically distributed noise terms (Box et al. 1994).

If e_t is the current error in the model, then $MA(q)$, i.e. a moving average correlation model with order q , is given by:

$$e_t = \theta_1 e_{t-1} + \dots + \theta_q e_{t-q} + \varepsilon_t , \quad (2.13)$$

where ε_i ($i = t-q, \dots, t$) are $q+1$ white noise terms, i.e., $\varepsilon_i \sim N(0,1)$. $\theta = (\theta_1, \dots, \theta_q)^T$ is the q parameters in $MA(q)$ correlation structure.

The correlation function for a $MA(q)$ model is:

$$\Gamma(k, \theta) = \begin{cases} \frac{\theta_k + \theta_1\theta_{k-1} + \dots + \theta_{k-q}\theta_q}{1 + \theta_1^2 + \dots + \theta_q^2}, & k = 1, \dots, q \\ 0 & k = q+1, q+2, \dots \end{cases} \quad (2.14)$$

Observations more than q time units apart are deemed as uncorrelated in a $MA(q)$ correlation structure.

For current data, a moving average correlation structure with order 2 was the best of the candidate correlation structures investigated with the likelihood ratio test and AIC values (Table 5). I call this Model 4 (i.e., all parameters are mixed, with no covariate considered, power function as the variance function and $MA(2)$ as the within-plot correlation structure). Thus I have finished one cycle of model specification in Step 1 and Step 2. Model 4 is “the best model” specified thus far. However, after considering covariates for the mixed effects (Step 3), the appropriate within-plot variance function and correlation structure may be different, but the general procedure described above is still valid.

Step 3: Specify inter-plot variation: Covariate modeling

After determining the nature of the parameters in the model (random or fixed), an immediate question is how to trace the random effects parameters. In this step, I will address the question of which variables in the survey or experiment are potentially useful in explaining random effects variation and how the random effects are explained by those variables [i.e., determine the design matrix in (2.5)]. This is a similar procedure to variable selection in ordinary regression analysis and should be primarily determined by the design of the experiment and the variables in the survey.

It will be informative as a first look to check the correlation matrix of the random effects from the final model in Step 2. High correlations among the random effects usually

indicate that some similar patterns exist among the design components which may be explained by some other covariates. A scatter plot for the three random effects from Model 4 (see Step 2) is given in Figure 7.

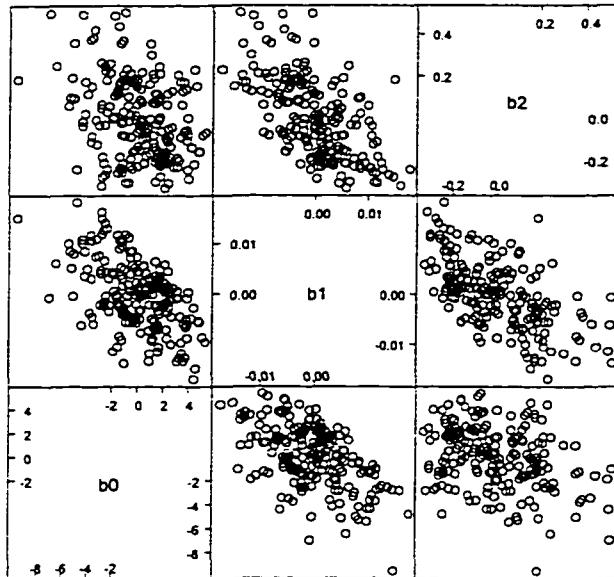


Figure 7: Scatter plot of estimated random effects for the modified Richards dominant height growth model with random effects and *iid* within-plot errors.

The correlation matrix of the random effects in Model 4 is:

$$\begin{matrix} \hat{b}_0 & \hat{b}_1 & \hat{b}_2 \\ \hat{b}_0 & 1 & & \\ \hat{b}_1 & -0.544 & 1 & \\ \hat{b}_2 & -0.468 & 0.710 & 1 \end{matrix}$$

All correlations among the three random effects are moderately high giving us a strong message that a covariate model for the mixed effects is necessary.

Another useful graphical device for investigating possible relationships between mixed effects and individual plot attributes (such as silvicultural treatments, soil type etc.) is to plot estimates of the random effects against potential covariates. The empirical Bayes estimates of random effects in Model 4 from Step 2 are plotted conditionally on

silvicultural treatments (see Figure 8) and soil type (see Figure 9). Visually, silvicultural treatments are more likely to be the significant factor for the random effects than the soil types. A formal split plot analysis of variance (e.g. Lentner and Bishop, 1986, Ch. 11) to the three random effects shows that silvicultural treatment is the only significant term to the random effects. Soil type and the interactions are not significant. This conclusion is also consistent with reports by other researchers using different approaches with these same data (e.g. Castleberry 1998).

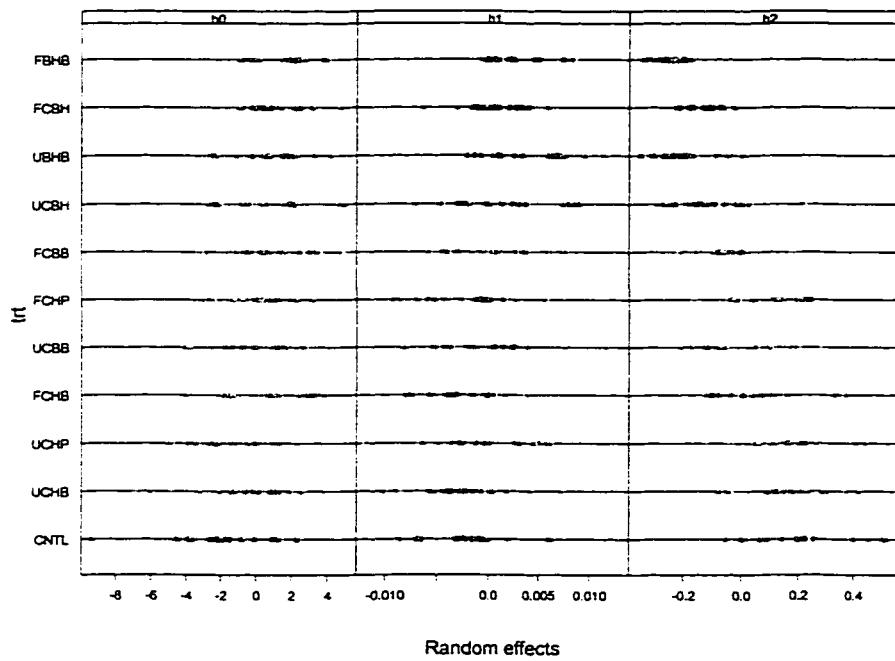


Figure 8: Comparisons of estimated random effects by silvicultural treatments for the modified Richards dominant height growth model with random effects and *n iid* within-plot errors.

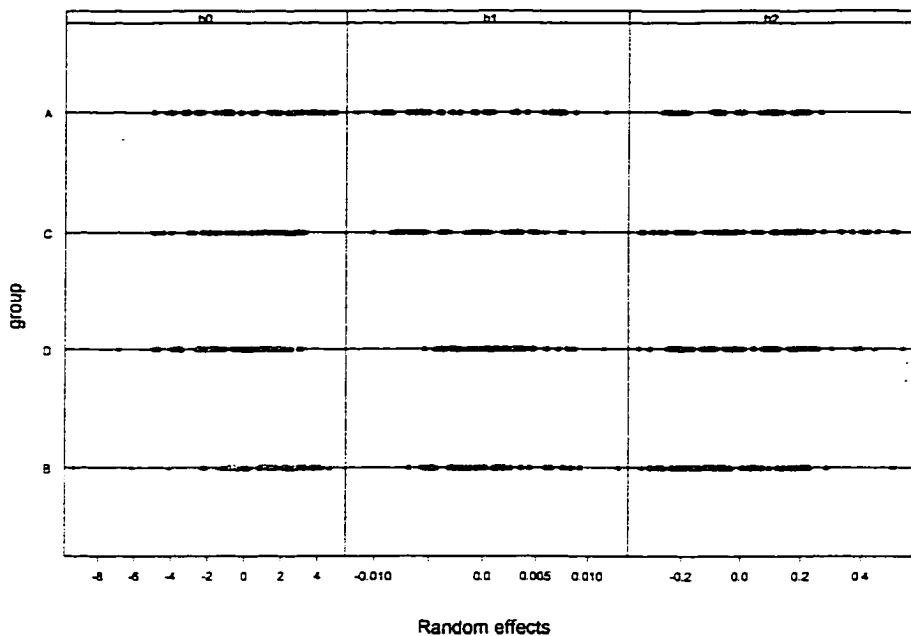


Figure 9: Comparisons of estimated random effects by soil types for the modified Richards dominant height growth model with random effects and *n iid* within-plot errors.

Even though the general silvicultural treatments are significant for all the mixed effects, it may well be an overparameterization to consider every silvicultural treatment as a fixed effect. There are 11 levels of treatments (including the control) and the differences between response curves for some of them are very small. One way around this is to take the significant levels in the treatments as fixed effects and the others as random. But some problems still remain because terms significant in ANOVA are not necessarily significant when they are put into the model as fixed effects. One of the attributes of the original study is that the silvicultural treatments are arranged in an additive way. Five distinct silvicultural activities (chop, burn, bed, fertilizer, and herbicide) are applied to individual treatment plots individually or in combination. For example, treatment FCHB is the combination of chopping, burning and fertilization. By assuming that the effects of the silvicultural treatments on the response (dominant height

growth) are additive (e.g. Pienaar et al. 1998) some dimension reduction can be obtained. Only 5 fixed effects will be involved instead of 11. However, it is still unrealistic to take all 5 silvicultural treatments as fixed in the three mixed effects ($\beta_{i,0}$, $\beta_{i,1}$ and $\beta_{i,2}$). After initial efforts, effects of burning and fertilizer were taken as fixed to $\beta_{i,0}$, effects of fertilizer and herbicide fixed to $\beta_{i,1}$, and effects of fertilizer, bed and herbicide fixed to $\beta_{i,2}$. For the random effects terms, the intercepts of the first two parameters (β_0 and β_1) are chosen as random, which may explain the combined random effects such as soil types, silvicultural treatment combinations, installations, or individual plot. Chopping has a random effect on parameter β_2 . Let $zchop_i$, $zburn_i$, zf_i , $zbed_i$ and zh_i be respective dummy variables indicating if chopping, burning, fertilizer, bed or herbicide treatment were applied on a specified treatment plot i. The inter-plot formulation for the slash pine dominant height growth may be represented as:

$$\beta_i = \begin{pmatrix} \beta_{i,0} \\ \beta_{i,1} \\ \beta_{i,2} \end{pmatrix} = \begin{pmatrix} 1 & zburn_i & zf_i & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & zf_i & zh_i & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & zf_i & zbed_i & zh_i \end{pmatrix} \begin{pmatrix} \beta_{00} \\ \beta_{01} \\ \beta_{02} \\ \beta_{10} \\ \beta_{11} \\ \beta_{12} \\ \beta_{20} \\ \beta_{21} \\ \beta_{22} \\ \beta_{23} \end{pmatrix} + \begin{pmatrix} b_{i,0} \\ b_{i,1} \\ zchop_i \times b_{i,2} \end{pmatrix} \quad (2.15)$$

where $\beta_{00}, \dots, \beta_{23}$ denote the fixed effects. Parameters $\beta_{00}, \beta_{10}, \beta_{20}$ are respectively the intercepts for fixed effects. Parameter vectors (β_{01}, β_{02}) , (β_{11}, β_{12}) , $(\beta_{21}, \beta_{22}, \beta_{23})$ are the

fixed “gains” on parameters $\beta_{i,0}$, $\beta_{i,1}$, $\beta_{i,2}$ respectively due to burning and fertilization, fertilization and herbicide, fertilizer, bed and herbicide.

Parameters $b_{i,0}$, $b_{i,1}$, $b_{i,2}$ are the random effects, in which $b_{i,0}$ and $b_{i,1}$ are the random intercept effects on $\beta_{i,0}$ and $\beta_{i,1}$ respectively. Parameter $b_{i,2}$ is the random effect in conjunction with site preparation chopping.

Assume $(b_{i,0}, b_{i,1} b_{i,2})^T \sim N(0_{3 \times 1}, D_{3 \times 3})$, D is 3×3 positive symmetric covariance matrix for the random effects.

Therefore, matrices A_i and B_i in (2.5) are specified as:

$$A_i = \begin{pmatrix} 1 & zburn_i & zf_i & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & zf_i & zh_i & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & zf_i & zh_i \end{pmatrix}, \quad B_i = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & zchop_i \end{pmatrix}$$

Once appropriate covariate models (design matrices A_i, B_i) are determined, one still needs to go back to Step 2, to choose an appropriate within-plot variance function and correlation structure. Usually, Step 3 and Step 2 need to be repeated until a satisfactory model is obtained. Just as before, the likelihood ratio test and the AIC values indicate that both heterogeneity and correlation are evident even after including covariates in the model. For example, both the exponential and the power functions significantly improve model fitting. Moreover, correlation structures $AR(1)$, $ARMA(1,1)$ and $MA(2)$ are all significant compared to the model with an independent structure. A power variance function combined with the moving average correlation with $q=2$ denote the best among these models (Table 6).

Table 6 Comparison of mixed effects model performance for slash pine dominant height growth data with between-plot covariates (Eq. 2.12) and different within-plot variance function and correlation structures

Model	Variance Function	Correlation Structure	Number of parameters	<i>AIC</i>	<i>log-likelihood</i>	<i>LRT</i> **	<i>p-value</i>
5	1*	independent	17	1524.069	-745.035		
6	Exp.	independent	18	1499.589	-731.795	26.480	<0.0001
7	power	independent	18	1493.780	-728.890	32.289	<0.0001
8	power	<i>AR</i> (1)	19	1489.443	-725.722	6.337	0.0118
9	power	<i>ARMA</i> (1,1)	20	1489.831	-724.916	7.949	0.0188
10	power	<i>MA</i> (2)	20	1487.933	-723.966	9.848	0.0073

* Variance function 1 means that the variances are homogeneous

** For Model 6 and 7, the likelihood ratios are calculated with respect to Model 5; for Model 8, 9, and 10, the likelihood ratios are calculated with respect to Model 7.

Thus the final specified model has the form

$$y_i = f(t_i, \beta_i) + e_i$$

$$\mu_i = E[y_i | \beta_i] = f(t_i, \beta_i) \text{ i.e.}$$

$$\mu_i = (\beta_{00} + \beta_{01} zburn_i + \beta_{02} zf_i + b_{i,0}) \left(\frac{1 - e^{-(\beta_{10} + \beta_{11} zf_i + \beta_{12} zh_i + b_{i,1}) t_i}}{1 - e^{-(\beta_{10} + \beta_{11} zf_i + \beta_{12} zh_i + b_{i,1}) t_0}} \right)^{(\beta_{20} + \beta_{21} zf_i + \beta_{12} zbed_i + \beta_{13} zh_i + b_{i,2} zchop_i)}$$

$$e_i \sim N[0, R_i(\beta_i, \alpha, \theta)]$$

$$Cov(e_i | \beta_i) = R_i(\beta_i, \xi) = \sigma^2 G_i^{1/2}(\beta_i, \alpha) \Gamma_i(\theta) G_i^{1/2}(\beta_i, \alpha)$$

$$G_i(\beta_i, \alpha) = \mu_i^\alpha$$

$$\Gamma_i(\theta) = MA(2)$$

$$(b_{i,0}, b_{i,1}, b_{i,2})^T \sim N(0_{3 \times 1}, D_{3 \times 3})$$

where *MA*(2) denotes a moving average correlation model with two parameters as defined in (2.13) and (2.14). There are 20 parameters (including the parameters in the variance-covariance matrices) to estimate in the above mixed effects model. Individual

random effects can be predicted by first-order approximation (Lindstrom and Bates, 1990)

Parameter estimation and model diagnosis

I used maximum likelihood estimation as a model specification procedure.

Detailed descriptions for this method in a nonlinear mixed effects model context can be found in Davidian and Giltinan (1995). S-plus software implementations are available from the *NLME* library by Bates and Pinheiro (1999). Model 10 fitted well to the data and all 10 fixed-effect parameters were significant with generally low covariances between pairs of parameters. The estimated fixed effects and random effects are presented in Table 7 and Table 8 respectively.

Table 7 The estimated fixed-effects and corresponding variance covariance matrix for the slash pine nonlinear mixed effects dominant height growth model (Model 10)

Parameter	Estimates	Std.Error	DF	t-value	p-value
β_{00}	18.57601	0.3768783	755	49.28914	<.0001
β_{01}	1.10252	0.3865288	755	2.85236	0.0045
β_{02}	2.52956	0.4076575	755	6.20511	<.0001
β_{10}	0.07860	0.0031822	755	24.69924	<.0001
β_{11}	-0.02428	0.0044125	755	-5.50155	<.0001
β_{12}	0.01446	0.0039667	755	3.64638	0.0003
β_{20}	1.96768	0.0364024	755	54.05373	<.0001
β_{21}	-0.29170	0.0246961	755	-6.25931	<.0001
β_{22}	-0.15458	0.0451633	755	-6.45872	<.0001
β_{23}	-0.17516	0.0440942	755	-3.97245	0.0001
----- The estimated covariance matrix ($\hat{\Omega}$) for the fixed effects -----					
β_{00}	β_{01}	β_{02}	β_{10}	β_{11}	β_{12}
0.14055	-0.10191	-0.05919	-0.0005	0.000389	0.000252
β_{01}	0.14784	-0.01599	0.000118	3.23E-05	-0.000365
β_{02}	-0.05919	0.164445	0.000398	-0.000988	0.000041
β_{10}	-0.0005	0.000118	0.000398	1E-05	-8.14E-06
β_{11}	0.00039	3.23E-05	-0.00099	-8.1E-06	1.93E-05
β_{12}	0.00025	-0.00036	0.000041	-4.9E-06	-9.9E-07
β_{20}	-0.00356	0.000696	0.002849	9.49E-05	-7.44E-05
β_{21}	0.00287	0.000124	-0.00667	-7.5E-05	0.000166
β_{22}	-0.00048	0.000816	1.86E-06	-5.2E-07	-9.2E-06
β_{23}	0.00201	-0.00266	0.000302	-4.8E-05	-8.98E-06

Table 8 The estimated parameters of random effects (\hat{D}), variance function ($\hat{\alpha}$), correlation structure ($\hat{\theta}_1, \hat{\theta}_2$), and the error of the model ($\hat{\sigma}$) for the slash pine nonlinear mixed-effects dominant height growth model (Model 10)

$\hat{D} = \begin{pmatrix} 5.7594 & -0.01498 & -0.05384 \\ -0.01498 & 0.0002574 & 0.001797 \\ -0.05384 & 0.001797 & 0.03327 \end{pmatrix}$
$\hat{\sigma} = 0.13265324$
$\hat{\alpha} = 0.3932779$ (the estimated power for variance function)
$\hat{\theta}_1 = 0.1852754$ $\hat{\theta}_2 = 0.3385666$ (correlation structure $MA(2)$)
----- with-plot correlation matrix $\Gamma(\hat{\theta}_1, \hat{\theta}_2)$ [by (2.14)] -----
$\Gamma(\hat{\theta}_1, \hat{\theta}_2) = \begin{pmatrix} 1 & 0.1613 & 0.3246 & 0 & 0 \\ 0.1613 & 1 & 0.1613 & 0.3246 & 0 \\ 0.3246 & 0.1613 & 1 & 0.1613 & 0.3246 \\ 0 & 0.3246 & 0.1613 & 1 & 0.1613 \\ 0 & 0 & 0.3246 & 0.1613 & 1 \end{pmatrix}$

Comparing Model 10 with Model 4 (see Tables 5 and 6), one can see that the model performance has remarkably improved by including the inter-plot covariates (silvicultural treatments). For example, the log-likelihood value increased from -834.404 to -723.966 . This gives a LRT of 220.876 with $20-13=7$ degree of freedom, which is significant at any common significance level. A closer diagnosis of the final model can be obtained by checking plots of the residuals, the fitted values, and the estimated random effects. Because the within-plot errors are assumed to be *n iid*, given the appropriate within-plot variance-covariance [R matrix in (2.4)] the residual plot (residuals or standardized residuals against the fitted values) should not reveal any observable pattern and the Q-Q plot should be approximately linear. The residual and Q-Q plots of Model 10 do not indicate serious departures of randomness and normality (Figures 10 and 11). The numbers identified on the residual plots are the observations which are potential

outliers in that the corresponding standardized residuals are outside of the range 2.5 to 97.5 standard normal quantiles. Overall, the standardized residuals are small, thus suggesting the nonlinear mixed effects Model 10 was successful in explaining the dominant height growth data. This is better seen by looking at the plot of observed dominant height versus the fitted values (Figure 12). The fitted values are in close agreement with the observed dominant heights since the plot closely follows a straight line with slope 1 and a zero intercept (Figure 12). Since the random effects are also assumed to be normally distributed with zero means, we may use a q-q plot to check the normality of the random effects (Figure 13). There is some indication of a heavy tail for the random intercept for β_1 (see the plot labeled "b1.Intercpt", Figure 13). Considering the large sample size ($n=191$), I concluded that the normality assumption for the random effects is justified.

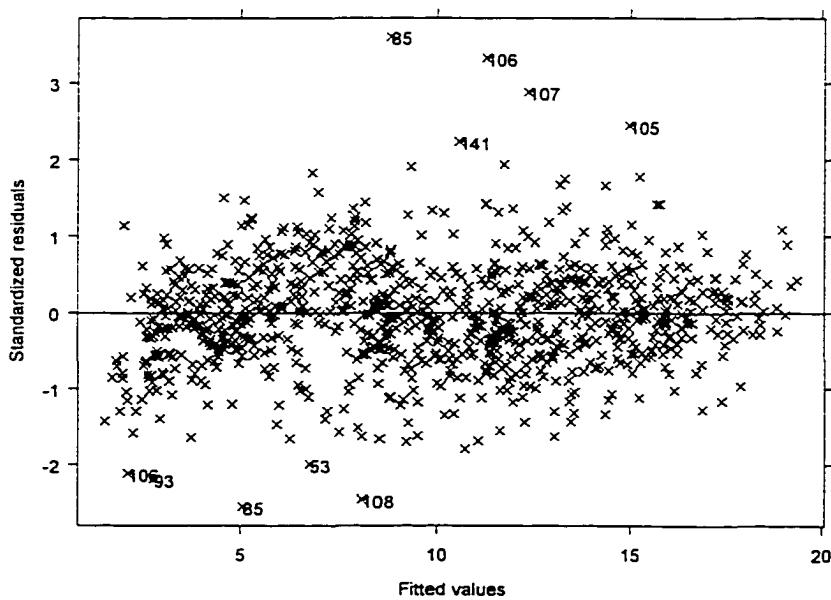


Figure 10: The residual plot for the modified Richards dominant height growth equation with random between-plot effects and $MA(2)$ within-plot autocorrelation and power variance function [Model (2.10)].

Another useful diagnostic plot to assess the assumptions on the random effects is the scatter plot matrix of the estimated random effects in that it can identify outliers and check the assumption of homogeneity of the random effects (Bates and Pinheiro 1999). For example, the scatter plots of the random effect estimates by soil type (nonspodosol or spodosol) do not indicate obvious heterogeneity in random effects between the two type of soil groups (Figure 14). This confirms a conjecture that soil types are not significant should not be included in the nonlinear mixed effects model.

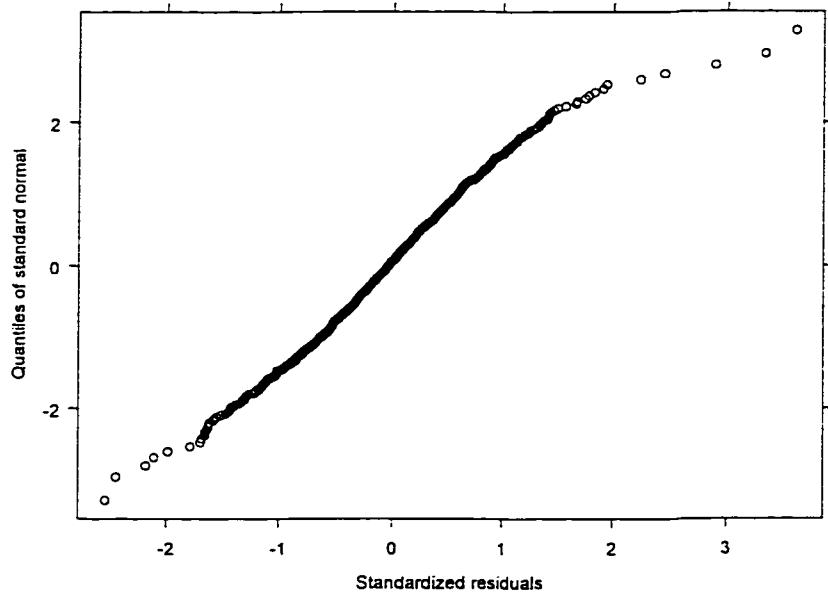


Figure 11: Normal plot of residuals for the nonlinear mixed effects dominant height growth model with $MA(2)$ autocorrelation and power variance function [Model (2.10)].

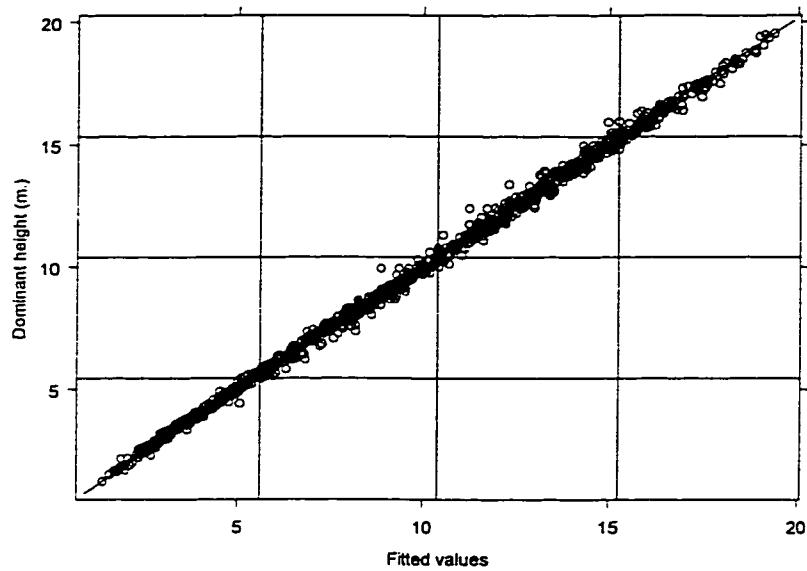


Figure 12: Observed dominant height versus predicted dominant height with the nonlinear mixed-effect dominant height growth model [Model (2.10)].

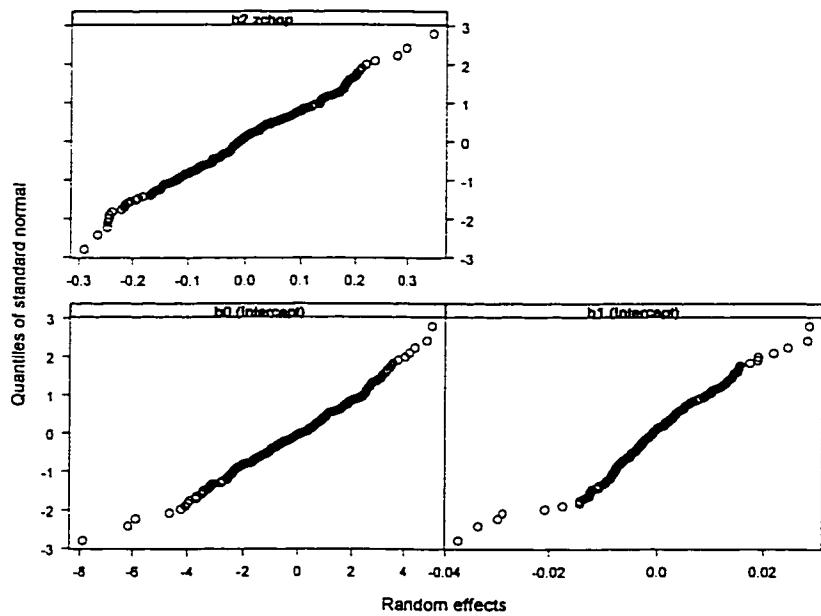


Figure 13: Normal plot of random effects for the non-linear mixed-effects dominant height growth model.

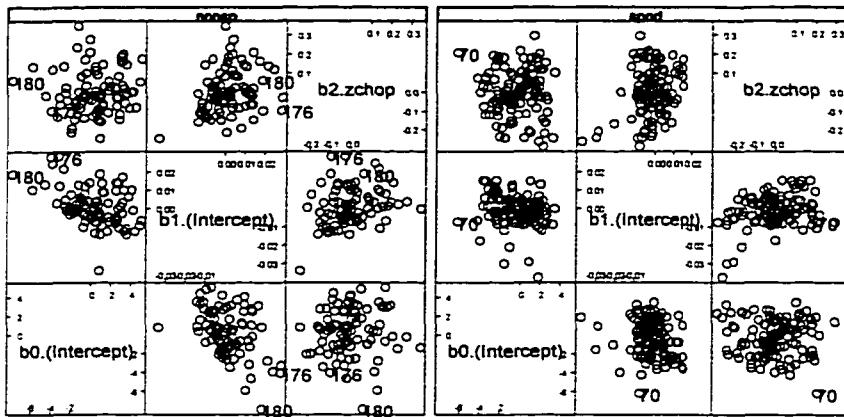


Figure 14: Scatter plot of estimated random effects for the nonlinear mixed-effects dominant height growth model by soil type groups (spodosol or nonspodosol)

Interpretation of model parameters

1. “Typical” responses for different silvicultural treatments

Parameter interpretation for a nonlinear mixed effects model will usually be quite different depending on whether one chooses to model the response y_i at the individual level or to consider population behavior and the marginal moments directly. Zeger et al. (1988) refer to the case of the hierarchical model for an individual as a *subject-specified* model and that of a model for marginal moments as *population -averaged* model (also see Diggle et. al. 1994). I have chosen a subject-specified approach thus far. A higher precision for individual prediction usually results with this approach. However, it is still possible in this approach to assess the “typical” response for an individual or some grouping of individuals. For example, we assume the random individual treatment plot effects to be 0 and then this “typical” dominant height growth for a given individual (or

group) can be obtained by just substituting corresponding fixed effects into the final model (Model 10). I obtained parameters in the modified Richard's equation for 11 silvicultural treatment groups with random effects set to zero (Table 9). One may use the parameters in Table 9 to obtain the "typical" dominant height growth curve for each individual silvicultural treatment. For example, I constructed a comparison of the "typical" dominant height growth for six silvicultural treatments: *FCHP* (chop, fertilize), *FCHB* (chop, burn, fertilize), *FCBB* (chop, burn, bed, fertilize), *FCBH* (chop, burn, herbicide, fertilize), *FBHB* (chop, burn, bed, herbicide, fertilize), and *CNTL* (no treatment). Treatment *FBHB* has highest growth response (Figure 15). It is also the most comprehensive treatment combination. *CNTL* has lowest growth response and has no silvicultural treatment included. It might be interesting to carry out a detailed contrast for the "typical" dominant height growth responses by different silvicultural treatment combinations, but it is beyond the focus of my current study. While this model is still informative to assess the "typical" or means responses for individual prediction, information on random effects is always helpful. As an example, I graphed a comparison of predicted curves with and without random effects for an experimental plot in the original data (a treatment plot in Installation 7, soil type *C*) with silvicultural treatment *FCBH* (chop, burn, herbicide and fertilize, Figure 16). From this it is clear that one obtains much more precise prediction with estimated random effects included in the model (Figure 16).

Table 9 The estimated parameters for “typical” dominant height growth (random effects input as 0) with modified Richard’s equation by silvicultural treatment groups

Silvicultural Treatments	Estimated parameters		
	β_0	β_1	β_2
CNTL	18.57601	0.07859702	1.967684
UCHP	18.57601	0.07859702	1.967684
FCHP	21.10557	0.05432134	1.675987
UCHB	19.67853	0.07859702	1.967684
FCHB	22.20809	0.05432134	1.675987
UCBB	19.67853	0.07859702	1.813103
FCBB	22.20809	0.05432134	1.521406
UCBH	19.67853	0.09306127	1.792522
FCBH	22.20809	0.06878559	1.500825
UBHB	19.67853	0.09306127	1.637942
FBHB	22.20809	0.06878559	1.346245

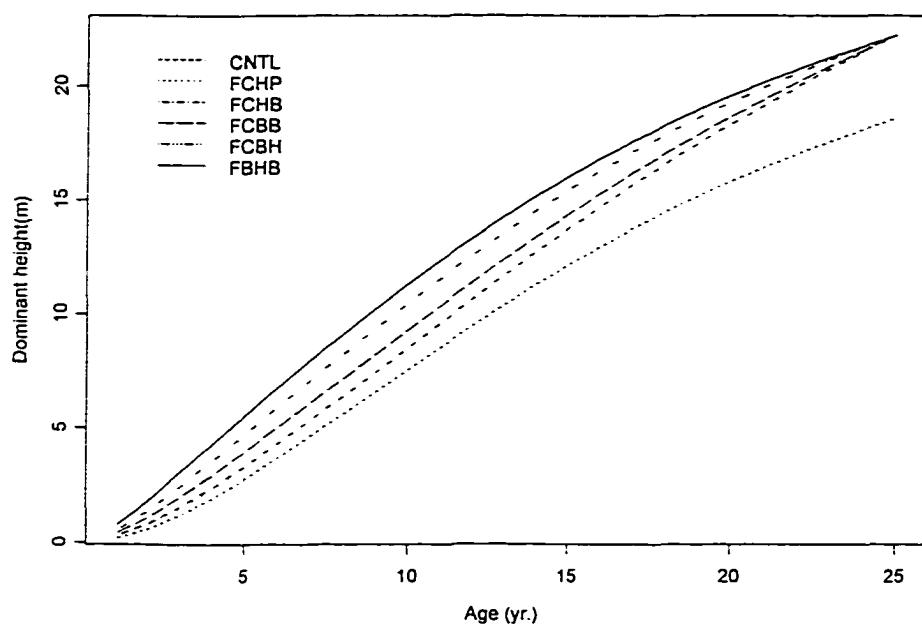


Figure 15: A comparison of the mean responses (typical responses) of dominant height growth by 6 different silvicultural treatments (CNTL, FCHP, FCHB, FCBB, FCBH, FBHB) with the nonlinear mixed-effects dominant height growth model.

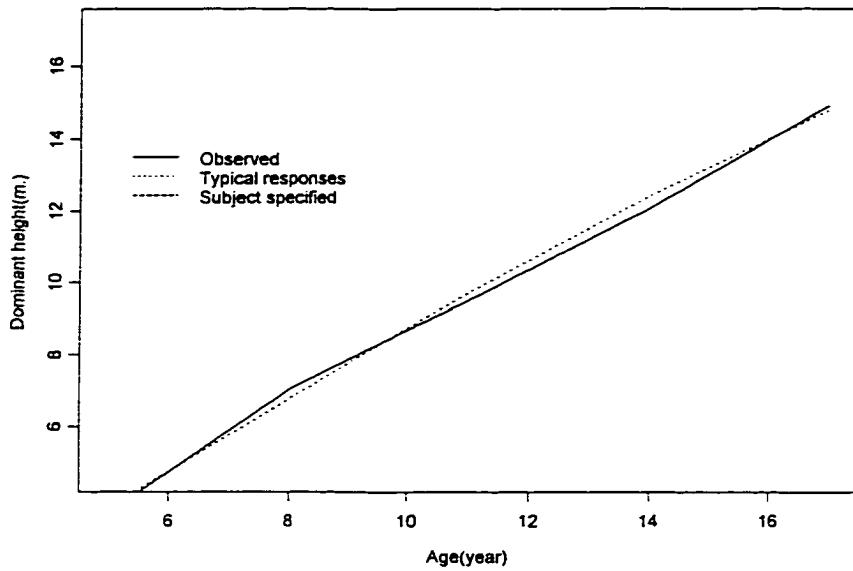


Figure 16: A comparison of individual dominant height growth predictions by typical response and subject-specified response with the nonlinear mixed-effects dominant height growth model.

2. Localization of site index curves

Age t_0 in the modified dominant height equation [equation (2.2)] was chosen to be 25, which is often used as the reference age for slash pine. The estimated parameter $\hat{\beta}_{i,0}$ is naturally interpreted to be the estimated site index for plot i . So the dominant height growth and the site index prediction models are combined into one model in current approach. For example, to obtain the site index curve, one may solve for β_0 directly from the current dominant height growth model or estimate a series of β_0 's by changing t_0 serially. With the second approach, one can not only obtain a site index curve for an individual plot, but also estimate upper and lower bounds on this curve. This

naturally provides a method to label site index curves for slash pine plantations with silvicultural treatments.

It is easy to show that the nonlinear mixed effects model based on Richard's equation (Model 10) is polymorphic in that varying rates of height growth on different sites are reflected. Model 10 is also base age invariant. That is, predictions and curves are invariant with respect to the changes in the base age for site index. Other examples of base age invariant models include Bailey (1974), Borders et al (1984), Cieszewski and Bella (1989), Cao et al (1993). It is also interesting to note that site index is affected by silvicultural treatments. This has been demonstrated not only because some silvicultural treatments (e.g. burning and fertilization) are significant for parameter β_0 , but also because the dominant height growth curve is affected by silvicultural treatments (see Figure 15).

Individual prediction for new observations

Suppose we are interested in the prediction of a single observation from a new experimental plot. This plot may or may not contain prior observations. Traditional regression type prediction does not distinguish these two situations. The nonlinear mixed effects model will distinguish these two situations and prior observations for a specified individual will improve the precision of the individual prediction.

Case 1 Prediction is required for a new individual with no prior observations

Suppose we are interested in the prediction of the dominant height on plot k at age t_{kh} and there is no prior observation on this plot. Assume all stand attributes are

given, so the design matrices of the fixed (A_k) and random effects (B_k) for this plot are also known. However, since there are no prior observations on this plot, it is impossible to determine the random effects (b_k) corresponding to this plot. Prediction and reference can be obtained by replacing b_k with 0's in Model 10 in this case:

$$\hat{y}_{kh} = f(\hat{\beta}_k, t_{kh}) = \hat{\beta}_{k0} \left(\frac{1 - e^{-\hat{\beta}_{k1} t_{kh}}}{1 - e^{-\hat{\beta}_{k1} t_0}} \right)^{\hat{\beta}_{k2}} \quad (2.16)$$

with prediction variance:

$$\text{var}(y_{kh} - \hat{y}_{kh}) = F_k^T (\hat{\beta}_k) (A_k \hat{\Omega} A_k^T + B_k \hat{D} B_k^T) F_k (\hat{\beta}_k) + \hat{\sigma}^2 R_{kh} \quad (2.17)$$

Where

$$\hat{\beta}_k = \begin{pmatrix} \hat{\beta}_{k0} \\ \hat{\beta}_{k1} \\ \hat{\beta}_{k2} \end{pmatrix} = A_k \begin{pmatrix} \hat{\beta}_{00} & \hat{\beta}_{01} & \hat{\beta}_{02} & \hat{\beta}_{10} & \hat{\beta}_{11} & \hat{\beta}_{12} & \hat{\beta}_{20} & \hat{\beta}_{21} & \hat{\beta}_{22} & \hat{\beta}_{23} \end{pmatrix}^T$$

$$F_k^T (\hat{\beta}_k) = \left. \frac{\partial f(t_{kh}, \beta_k)}{\partial \beta_k^T} \right|_{\beta_k = \hat{\beta}_k}$$

$R_{kh} = f^{\hat{\alpha}}(\hat{\beta}_k, t_{kh})$, $t_0 = 25$, The $\hat{\beta}_k$'s, $\hat{\Omega}$, \hat{D} , $\hat{\sigma}$ and $\hat{\alpha}$ are the estimated parameters given in Table 7. All other terms are defined above.

So an approximate $100(1-\alpha)\%$ confidence interval for this prediction is:

$$\hat{y}_{kh} \pm t(n-q, 1-\alpha/2) \sqrt{\text{var}(\hat{y}_{kh})} \quad (2.18)$$

where $n-q$ is the degrees of freedom associated with the fixed effects. $n-q=955-10=945$. For example, if we want to predict the slash pine dominant height at age 20 on a new plot with silvicultural treatments $FCBH$ (chop, burn, herbicide, fertilize) and there is no other information on this plot available. We have

$$A_k = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \end{pmatrix}, \quad B_k = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Thus

$$\hat{\beta}_k = \begin{pmatrix} \hat{\beta}_{k0} \\ \hat{\beta}_{k1} \\ \hat{\beta}_{k2} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{pmatrix}^{-1} \begin{pmatrix} 18.57601 \\ 1.10252 \\ 2.52956 \\ 0.07860 \\ -0.02428 \\ 0.01446 \\ 1.96768 \\ -0.29170 \\ -0.15458 \\ -0.17516 \end{pmatrix} = \begin{pmatrix} 22.20809 \\ 0.068786 \\ 1.500825 \end{pmatrix}$$

Substitute the $\hat{\beta}_k$'s into Model 10, then the estimated dominant height at age 20 is

$$\hat{y}_{kh} = 19.29 \text{ m.}$$

$F_k^T = (0.8686262 \ 37.81246 \ -1.810287)$. Substitute $\hat{\Omega}$, \hat{D} , $\hat{\sigma}$, $\hat{\alpha}$ and design matices into (2.16) to obtain a variance of this prediction: $\text{var}(\hat{y}_{kh}) = 3.8758$

The 95% confidence interval for the prediction is [15.43 m, 23.15 m]

Case 2 Prediction is required for a new individual with prior observations

An alternate, and perhaps more practical, situation in forestry occurs when a prediction is required for a known individual for which a collection of formerly observed data is available. One may still use the procedure in Case 1 to obtain the prediction and corresponding prediction limits just as with traditional regression. However, this approach would ignore any information contained in the prior.

observations of the specified individual. Instead, we may use the information contained in former observations to estimate the subject-specific random effects for this individual and incorporate this directly into our prediction.

Suppose the information available for individual plot k is $\{y_k, t_k, A_k, B_k\}$, where y_k is a vector of former observed dominant heights at corresponding age vector t_k . We wish to predict a future dominant height y_{kh} at a given age t_{kh} . We first estimate the random effect b_k . We can take a first order expansion of $f(t_k, A_k \beta_k + B_k b_k)$ about $b_k = 0$. From the residual, $e_k = y_k - f(t_k, A_k \beta_k)$, one can estimate b_k by the post mean and variance of $(b_k | y_k)$, β_k, ξ_k , where ξ_k is the vector of variance covariance parameters for individual k [see the general model form (2.4)]. For details refer to Vonesh and Chinchilli (1997, Ch. 7).

$$\hat{b}_k = \hat{D} \hat{Z}_k^T (\hat{Z}_k \hat{D} \hat{Z}_k^T + \hat{R}_k)^{-1} \hat{e}_k \quad (2.19)$$

where $\hat{Z}_k = \hat{F}_k^T B_k$,

$$\hat{e}_k = y_k - f(t_k, A_k \hat{\beta}_k)$$

With variance

$$V_k V_k = \text{var}(b_k - \hat{b}_k) = \hat{D} - \hat{D} \hat{Z}_k^T \Sigma_k^{-1} (\hat{\beta}, \hat{\xi}) \hat{Z}_k \hat{D} + \hat{D} \hat{Z}_k^T \hat{F}_k \Omega(\hat{\beta}) \hat{F}_k^T \hat{Z}_k \hat{D} \quad (2.20)$$

where $\Sigma_k(\hat{\beta}, \hat{\xi}) = \hat{Z}_k \hat{D} \hat{Z}_k^T + \hat{R}_k$

$$\Omega(\hat{\beta}) = A_k \hat{\Omega} A_k^T$$

After estimating b_k , we may take a first-order expansion of $f(t_k, A_k \beta_k + B_k b_k)$ about $\beta = \hat{\beta}$

and $b_k = \hat{b}_k$, with a little algebra, the prediction for y_{kh} can be realized as:

$$\hat{y}_{kh} = f(t_{kh}, A_k \hat{\beta}_k + B_k \hat{b}_k) \quad (2.21)$$

with prediction variance:

$$\text{var}(y_{kh} - \hat{y}_{kh}) = \hat{F}_{kh}^T \hat{\Omega}(\hat{\beta}) \hat{F}_{kh} + \hat{z}_{kh}^T \hat{V}_k \hat{z}_{kh} + \hat{F}_{kh}^T \hat{C}_k \hat{z}_{kh} + \hat{z}_{kh}^T \hat{C}_k^T \hat{F}_{kh} + \hat{R}_{kh} \quad (2.22)$$

where $C_k = \text{Cov}[(\hat{\beta} - \beta), (\hat{b}_k - b_k)^T] \equiv -\Omega(\hat{\beta}) \hat{F}_k^T \Sigma_k^{-1}(\hat{\beta}, \hat{\xi}) \hat{Z}_k \hat{D}$ and \hat{F}_{kh} , \hat{z}_{kh} are respectively the component of a column vector in matrix \hat{F}_k , \hat{z}_k corresponding to the h^{th} occasion (on which prediction is required).

Hence, the approximate $100(1-\alpha)\%$ prediction confidence interval is given as:

$$\hat{y}_{kh} \pm t(n-q, 1-\alpha/2) \sqrt{\text{var}(\hat{y}_{kh})} \quad (2.23)$$

\hat{y}_{kh} and $\text{var}(\hat{y}_{kh})$ are calculated by (2.20) and (2.21) respectively.

For a specific example, suppose we know the dominant heights at age 11, 14, 17 to be 12m, 16m and 18m for a plot with the *FCBH* treatment. We want to predict the dominant height of this plot at age 20. Thus,

$$\hat{e}_k = y_k - f(t_k, A_k \beta_k) = \begin{pmatrix} 12 \\ 16 \\ 18 \end{pmatrix} - \begin{pmatrix} 11.54 \\ 14.51 \\ 17.09 \end{pmatrix} = \begin{pmatrix} 0.46 \\ 1.49 \\ 0.91 \end{pmatrix}$$

Suppose the within-plot correlation still follow *MA(2)* structure, then

$$R_k = \text{diag}(\hat{\mu}_k^{\hat{\alpha}/2}) \Gamma_{3 \times 3} \text{diag}(\hat{\mu}_k^{\hat{\alpha}/2}) = \begin{pmatrix} 2.6168 & 0.4415 & 0.9176 \\ 0.4415 & 2.8635 & 0.47697 \\ 0.9176 & 0.47697 & 3.0537 \end{pmatrix}$$

$$\hat{Z}_k = \hat{F}_k^T B_k = \begin{pmatrix} 0.5197 & 0.6535 & 0.7696 \\ 73.9598 & 69.4579 & 56.4948 \\ -5.0331 & -4.1139 & -2.9823 \end{pmatrix}^T \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0.5197 & 73.9598 & -5.0331 \\ 0.6535 & 69.4579 & -4.1139 \\ 0.7696 & 56.4948 & -2.9823 \end{pmatrix}.$$

Hence, by (2.19),

$$\hat{b}_k = \begin{pmatrix} 1.3199720976 \\ -0.0005567815 \\ -0.0165628772 \end{pmatrix}$$

With the random effects considered, the predicted dominant heights for this specific plot at ages 11,14,17 are respectively 12.27m, 15.41m,18.13m, which are closer to the observed values (12m,16m,18m) than those given by the means response function (11.54m, 14.51m, 17.09m). This is not a surprising result, because the random effect estimation process has used all observed information whereas the mean response prediction does not.

The estimated variance-covariance matrix for the random effects is [by (2.20)]:

$$\hat{V}_k = \text{var}(b_k - \hat{b}_k) = \begin{pmatrix} 3.52872532 & -0.0132174726 & -0.031320308 \\ -0.01321747 & 0.0002521917 & 0.001825877 \\ -0.03132031 & 0.0018258767 & 0.032479971 \end{pmatrix}$$

Therefore, with the random effects considered, the predicted dominant height at age 20 [by (2.21)] is: $\hat{y}_{kh} = 20.45$ m, the estimated variance is [by (2.22)]:
 $\text{var}(\hat{y}_{kh}) = 2.097$ m, thus the 95% confidence interval for the prediction is :
[17.61m, 23.68 m].

Comparing with case 1, we can see that former observations for a specified individual can improve the precision of the future prediction for that individual with the mixed effects model approach. In this example, the standard error of the prediction decreased by 26.5% from 1.969 m to 1.448 m A graphical comparison of these two prediction approaches is given in Figure 17. Generally, the more special an individual is (i.e., the more different from the “typical” response with b_i ’s of zero), the more “gain” will result by using the mixed effects prediction.

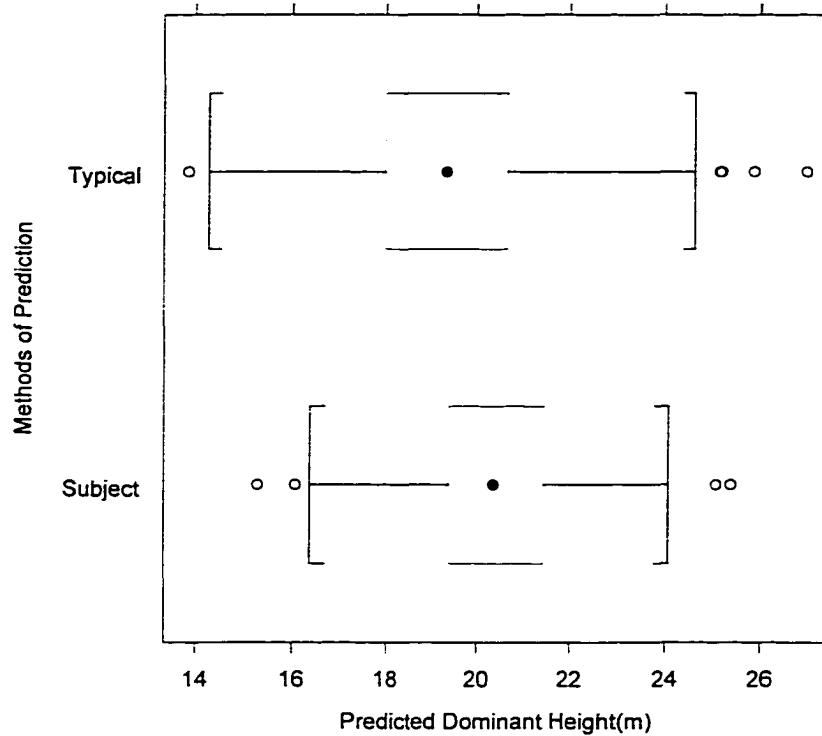


Figure 17: A comparison of the prediction and corresponding confidence interval for a new observation by typical response and subject specified response with the nonlinear mixed-effects dominant height growth model.

Discussion

A mixed effects model is appealing for the analysis of repeated measurement data from a given experiment design because of its flexibility in representing the covariance structures that are not readily identified by the traditional regression approach. A modified Richards equation with nonlinear mixed effects is recommended for modeling slash pine dominant height growth in conjunction with different silvicultural treatments. The analysis results show that a random effects model always provides better model fitting and more precise prediction compared to a similar model without random effects. This is not hard to understand. Mixed effects reasonably explain the individual random variation that is quite common for repeated measurement data from permanent plots with an experimental design. If prior observations are available, the random effects corresponding to this specified individual can be estimated, thus more precise prediction for future observations can be obtained. This makes intuitive sense since more information is applied in the prediction by using the random effects approach. Random effects may relieve such problem as non-homogeneous variance and within-plot auto-correlation error in modeling forest growth and yield repeated measurement data. But heterogeneity and correlation still occur for slash pine dominant height growth even after considering the appropriate random effects. Heteroscedasticity is usually easier to handle. Once the trend of the variance (variance function) is identified, an appropriate individual weight (inverse of the variance) can be applied to individual observations to justify the process of parameter estimation. Comparatively, within-plot auto-correlation is not so easy to justify. Because many other factors in the model (or data) affect the auto-correlation, a reasonable structure may not be obvious for a given collection of data and

model. No benefits will accrue from an inappropriate structure when auto-correlation is evident in the data. So, the model for auto-correlation structure should usually be as simple as possible when an appropriate one is not available(Davidian and Giltinan 1995). Moving average correlation with two parameters was used to model the within-plot correlation for the slash pine data. The estimated θ_1 is smaller than that of θ_2 (0.1853 vs. 0.3386). This seems counter-intuitive. One would expect observations spaced closer in time to exhibit higher correlation. Actually, this is also demonstrated on the empirical auto-correlation function (Figure 4) in that negative correlation was estimated for the first three lags. This counter-intuitive phenomenon itself is good evidence of the complexity in correlation structure modeling . Such is not uncommon in the real world of data (see Davidian and Giltinan 1995, pages 132-133).

The well known Richard's equation was modified to include an expected-value parameter which is naturally interpreted as site index with t_0 as the base age. In the recommended final structure , dominant height growth and site index are explicitly combined into one model form. Incompatibility between the height growth and site index model (Curtis et al 1974, Cieszewski and Bailey 1999) can be avoided in the recommended model form. By changing t_0 serially, one can obtain a method of labeling site index curves that does not really involve the concept of site index. As Northway (1985) pointed out, site index of a stand should be a label that refers to a curve which most closely follows the height development of the stand. Even though traditionally the label is made so that the curve passes through the point (reference age, site index), site index itself is not the primary value of interest. This is clearly demonstrated by the procedure suggested by Clutter et al (1992, p. 58) in which the concept of index age is

completely avoided . The above recommended model form can express these ideas about site index calibration. It is easy to show that the modified Richard's equation with mixed effects naturally forms a base age invariant polymorphic site index model (Bailey 1974).

Silvicultural treatment is the only significant term in the original split-plot design. Since the different effects on slash pine dominant height growth are trivial among some silvicultural treatments, some treatment effects are included as fixed and some as random in the model. For the random effects, we have not considered the level of factors of the original design. Even though the current model (Model 10) fitted the data quite well, a generally more appealing approach to the data with a split-plot design is the multi-level mixed effects model (Goldstein 1995). This would allow random effects as multi-sources of errors. I was not able to apply such an approach in this modeling effort because I could not figure out a reasonable and tractable variance covariance structure for the multilevel random effects in these data.

CHAPTER 3

LINEAR MULTILEVEL MIXED EFFECTS MODELS FOR THE ANALYSIS OF REPEATED MEASUREMENTS FROM PERMANENT PLOTS IN A DESIGNED EXPERIMENT: A STUDY OF SLASH PINE BASAL AREA GROWTH FOLLOWING INTENSIVE SILVICULTURAL TREATMENTS

Introduction

Forest growth and yield data often form a hierarchical structure, in that samplings are obtained in multilevel nested units. A common example is remeasured permanent plot data with individual tree measurements. Such cases naturally form a three-level nesting: the remeasured characteristic as Level 1 are nested within individual tree and individual trees as Level 2 are nested within the individual fixed-area plots (Level 3). This type of nesting was discussed in the context of mixed effects models in forestry by Lappi and Bailey (1988) and Hall and Bailey (1997). Multiple levels of nesting in data for stand level modeling is also very common, especial in intensive forest management studies. For example, the permanent plot data from a split-plot design typically include at least three-levels of nesting: repeated measurements are nested in sub-plots and sub-plots are nested in whole plot. With block effects considered, repeated measures from a split-plot design can form more than three-level nesting. The slash pine plantation data collected by The University of Georgia *PMRC* (described in Chapter 2), for example,

may be viewed as a 5-level nesting: repeated observations within individual plot, plot within silvicultural treatment group, treatment group within soil group, and soil group within installation region. This data set will be explored again latter in this chapter.

Since the data from multiple levels of nesting may have multiple sources of error in modeling, a multi-level mixed effects model (Goldstein 1995) is an obviously appealing approach. In this chapter, as an example, we will show how a multi-level mixed effects model can be applied to model basal area growth in the *PMRC* slash pine site preparation study introduced in Chapter 2.

Data

The basal area growth data also come from the designed study for slash pine (*Pinus elliottii Engelm.*) by the Plantation Management Research Cooperative (PMRC) of the University of Georgia (see Chapter 2). Basal area was measured 5 times from age 5 to age 17 with 3-year intervals on 191 silvicultural treatment plots. Dominant height and trees per hectare are also available at these same ages simultaneously with basal area. The basal area growth profiles grouped by silvicultural treatments are shown in Figure 18. The correlation matrix plot of log-basal area (logarithm of basal area), age inverse (1/age), log-dominant height (logarithm of dominant height) and log-tree per hectare (logarithm of tree per hectare) is shown in Figure 19.

The basal area growth data can be viewed as a 5-level nesting (Figure 20). The repeated observations occur at Level 1 and are nested within the Level 2 units, the individual plots. Plots are nested within silvicultural treatments (Level 3) and silvicultural treatments are nested in the Level 4 units, soil types which are themselves

nested within installations (Level 5). Since Level 1(repeated factor) is usually modeled by the error term in a model, at most only 4 levels of nesting in random effects can be considered in a mixed effects model context. Instead of 5-level nesting, it is often called 4-level nesting (Pinheiro and Bates, 1998). Our data are unbalanced. For example, in each installation, two plots were established for one of the 11 silvicultural treatments and only one plot was established for each of the others. Additionally, because the original design is incomplete for treatments within soil group, the nesting is incomplete.

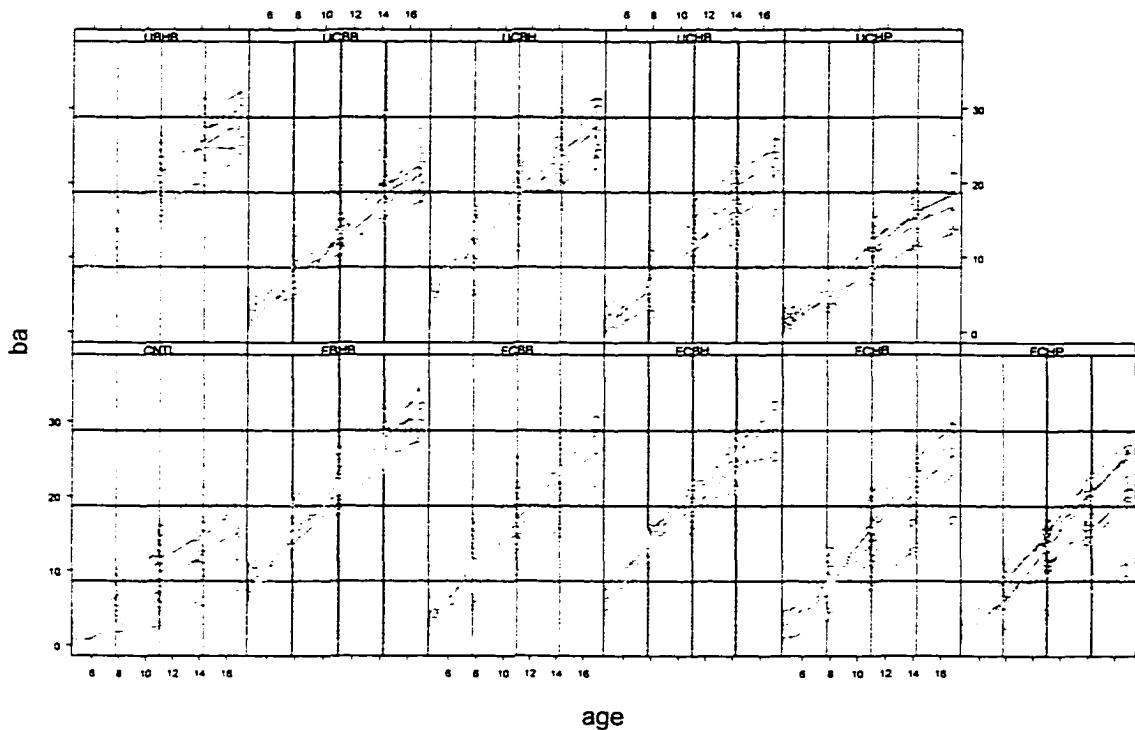


Figure 18: Profile graph of slash pine basal area growth on 191 plots with 5 repeated measures conditioning on 11 silvicultural treatments.

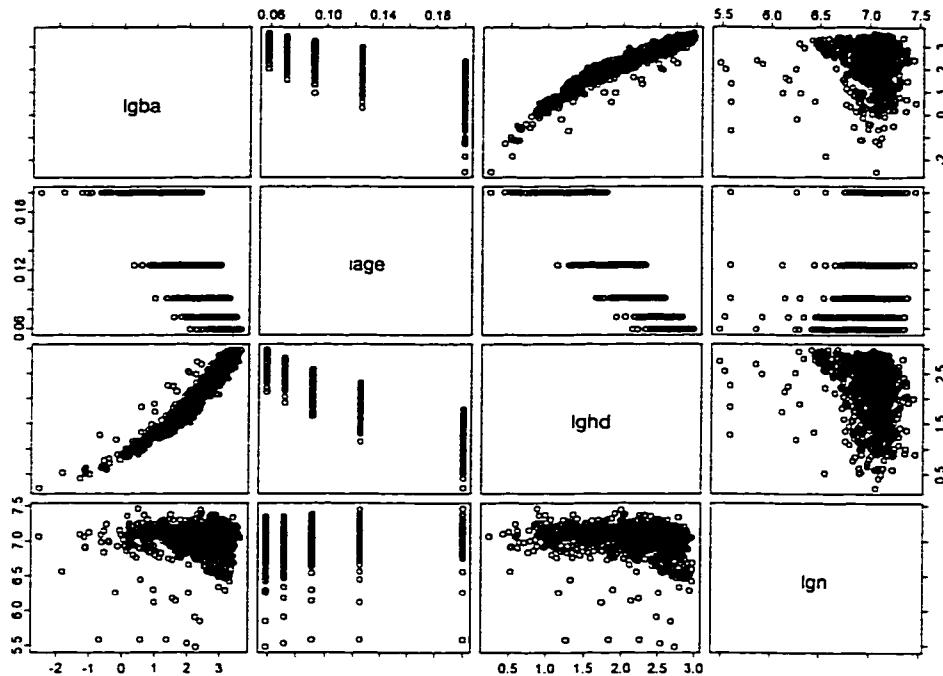


Figure 19: Scatter plots of response [$\ln(\text{BA})$] and covariates [$1/\text{age}$, $\ln(\text{HD})$ and $\ln(\text{TPH})$].

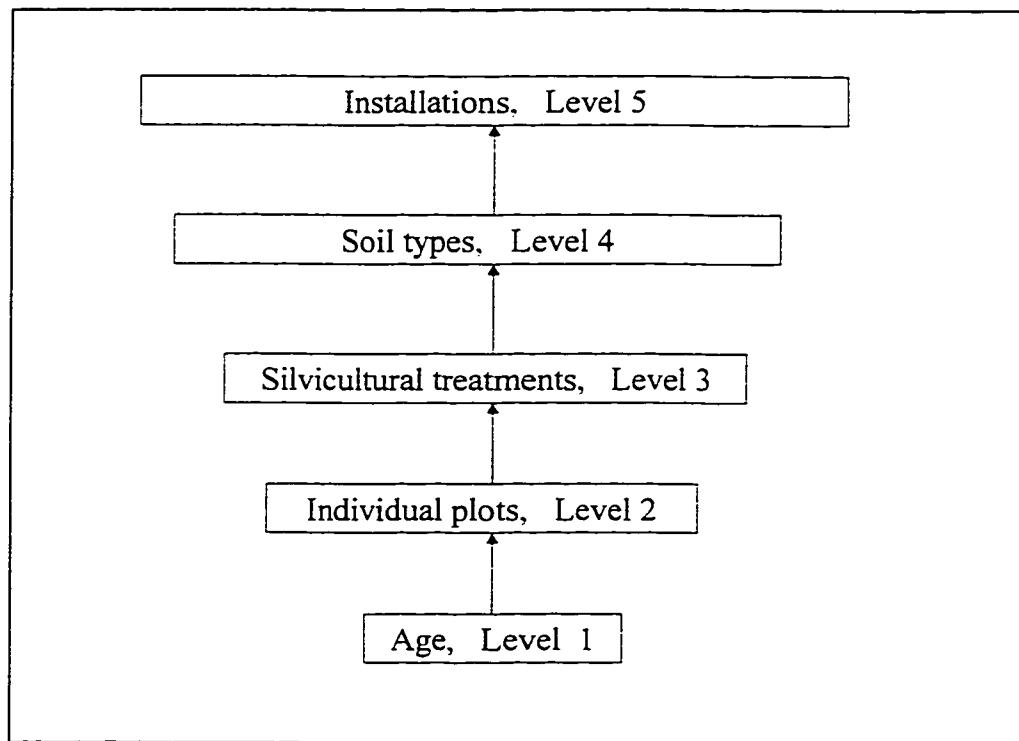


Figure 20: The nesting relationships of the original design for the slash pine basal area growth data.

Generalized linear multilevel mixed-effects models

Let y_i be a $n_i \times 1$ vector of repeated measurements on individual i (plot i) ($i = 1, \dots, N$). Suppose that there are M ($M \geq 1$) sources of heterogeneity contributing to y_i . These sources of heterogeneity are introduced into the model through m factors, the j^{th} factor taking level $k = 1, \dots, M_j$ ($j = 1, \dots, M$). If the m factors are in a nesting structure and a model considers all m sources of variation, it is called multilevel modeling (Bryk and Raudenbush, 1992, Goldstein 1995, Bates and Pinheiro 1998). The count of *levels of nesting* can be different, which depends on whether the error term in a model is counted in or not. In a mixed-effects modeling context, we prefer the terminology from the experimental design literature and count the number of “levels” as the number of nested levels of random effects. So instead of *m-level nesting* (Goldstein, 1995), the convention is to call it *(m-1)-level nesting* (Bates and Pinheiro, 1998) with m sources of variation. For example, Hall and Bailey (1997) consider three levels of variation in the context of site index modeling. They considered repeated observations (occasions) as nested within individual trees and trees nested in stands (plots). So, $M = 2$ in that case.

Similar to Gilks et al (1993), we denote the level of factor j for individual i by $K(i,j)$ and write the general multilevel linear mixed model as:

$$y_i | \beta, b_{K(i,j)} = X_i \beta + \sum_j z_{j,K(i,j)} b_{K(i,j)} + \varepsilon_i \quad (3.1a)$$

Here y_i and ε_i are the $(n_i \times 1)$ vectors of responses and random intra-individual errors for individual i . Parameter β is a $p \times 1$ vector of fixed effects, the $b_{K(i,j)}$ are a $q_{ij} \times 1$

vector of random effects on individual i due to level (factor) j . Matrices X_i and z_{ij} are design matrices corresponding to fixed and random effects respectively. Suppose that:

$$E[\alpha_i | \beta_i] = 0 \quad (3.1b)$$

$$\text{and} \quad \text{Cov}(\alpha_i | \beta_i, b_i) = R_i \quad (3.1c)$$

where R_i is the $(n_i \times n_i)$ intra-individual covariance matrix. Let

$$[b_{K(i,j)} | \theta] \sim N(0, D_j(\theta)) \quad (3.1d)$$

where $D_j(\theta)$ is a $m_j \times m_j$ variance covariance matrix for the random effects due to factor j .

In summary, the generalized linear multilevel mixed effects model can be written in hierarchical form (Davidian and Giltinan 1995, Ch.3) as:

$$E[y_i | b_{K(i,j)}] = x_i^T \beta + \sum_{j=1}^M z_{j,k(i,j)} b_{K(i,j)} \quad (3.2a)$$

$$\text{Cov}(y_i | b_{K(i,j)}) = R_i \quad (3.2b)$$

$$b_{K(i,j)} \sim N[0, D_j(\theta)] \quad (3.2c)$$

$$E[y_i] = x_i^T \beta \quad (3.2d)$$

$$Var(y_i) = \Sigma_i = \sum_{j=1}^M z_{j,k(i,j)} D_j(\theta) z_{j,k(i,j)}^T + R_i \quad (3.2e)$$

The advantage of this notation is that an unbalanced data structure is allowed in this framework, thus missing values present no problems. A similar model was discussed by Gilks et al (1993) using Gibbs sampling in a Bayesian framework. As they pointed out, there are several special cases of the model expressed by equations (3.2). For example, if the individual is the only source of heterogeneity, then (3.2) is reduced to the random-effects model proposed by Laird and Ware (1982). If, as another special case, several nested sources of heterogeneity are present Model (3.2) is reduced to the nested-

random-effects model of Goldstein (1986), Longford (1993), Karlsson and Sheiner (1994) and Hall and Bailey (1997).

Likelihood reference for linear multilevel mixed model

The marginal likelihood function for the above linear multilevel mixed effects model is:

$$\begin{aligned} L(\beta, \theta, \sigma^2 | y) &= \prod_i^N p(y_i | \beta, \theta, \sigma^2) \\ &= \prod_i^N \prod_j^M \int_{\mathbb{R}^{q_y}} p(y_i | b_{K(i,j)}, \beta, \theta, \sigma^2) p(b_{K(i,j)} | \theta, \sigma^2) db_{K(i,j)} \end{aligned} \quad (3.3)$$

where the conditional density of y_i is multivariate normal:

$$p(y_i | b_{K(i,j)}, \beta, \theta, \sigma^2) = \frac{\exp[-(y_i - X_i \beta - \sum_j z_{j,i} b_{K(i,j)})^T R_i^{-1} (y_i - X_i \beta - \sum_j z_{j,i} b_{K(i,j)})]}{(2\pi)^{n/2} \sqrt{|R_i|}}$$

and the marginal density of $b_{K(i,j)}$ is also multivariate normal:

$$p(b_{K(i,j)} | \theta, \sigma^2) = \frac{\exp[-b_{K(i,j)}^T D_{ij}^{-1} b_{K(i,j)}]}{(2\pi)^{q_y/2} \sqrt{|D_{ij}(\theta)|}} \quad (3.4)$$

For given D_j and R_i (i.e. θ and σ^2 are known), the maximum likelihood estimation, and generalized least squares estimation in this case, of β , b_{ij} and θ are respectively (see Hocking 1996, p622 and Davidian and Giltinan 1995, p78):

$$\hat{\beta} = \left(\sum_{i=1}^N x_i^T \Sigma_i^{-1} x_i^T \right)^{-1} \sum_{i=1}^N x_i^T \Sigma_i^{-1} y_i \quad (3.5)$$

$$\hat{b}_{K(i,j)} = D_j z_{j,i}^T \Sigma_i^{-1} (y_i - x_i^T \hat{\beta}) \quad (3.6)$$

where

$$\Sigma_i = \sum_{j=1}^M z_{j,i} D_j(\theta) z_{j,i}^T + R_i(\theta) \quad (3.7)$$

But in practice D_j and R_i are usually unknown. A common alternative is to use point estimates of $\hat{\theta}$ and $\hat{\sigma}^2$, thus \hat{D} and \hat{R} replace the corresponding terms in (3.5) and (3.6).

Maximum likelihood (*ML*) and restricted (or residual) maximum likelihood (*REML*) are often used to estimate the covariance parameters θ . *ML* involves maximization of marginal loglikelihood with respect to the components of θ . The log-likelihood may be written

$$\log L = -\frac{1}{2} N n_i \log(2\pi) - \frac{1}{2} \sum_i^N [\log|\Sigma_i| + (y_i - X_i \beta)^T \Sigma_i^{-1} (y_i - X_i \beta)] \quad (3.8)$$

Joint maximization of (3.8) with respect to fixed effects β and θ results in the generalized least squares (*GLS*) estimates of β that is given in (3.6).

Restricted maximum likelihood estimation (*REML*) adjusts for loss of degrees of freedom due to estimating fixed effects (e.g. β) in the model. *REML* can be obtained by estimating variance components based on residuals from fitting the fixed effects only.

Algebraically, this corresponds to replacing $\log L$ in (3.8) by

$$\log L_R = \log L + \frac{1}{2} p \log(2\pi) - \frac{1}{2} \sum_i^N [\log|X_i^T \Sigma_i X_i|] , \quad (3.9)$$

where $\log L$ is the marginal log-likelihood evaluated at $\hat{\beta}$.

Linear mixed effects model with four-level nesting

After some form of transformation of basal area so that a linear model is appropriate, let y_{ijkl} denote repeated measurements of a $n_{ijkl} \times 1$ response vector on the i^{th} plot, with j^{th} silvicultural treatment, on k^{th} type of soil group, in l^{th} installation. With a

four-level nesting (5-level nesting according to Goldstein 1995), the general multilevel linear mixed effects representation (1a) can be specified as:

$$y_{ijkl} = X_{ijkl}^T \beta + z_{1,ijkl}^T b_{1,ijkl} + z_{2,ijkl}^T b_{2,ijkl} + z_{3,ijkl}^T b_{3,ijkl} + z_{4,ijkl}^T b_{4,J} + \varepsilon_{ijkl} \quad (3.10)$$

where X_{ijkl} is a $p \times n_{ijkl}$ matrix of covariates associated with the fixed effects β . The variables $z_{1,ijkl}, z_{2,ijkl}, z_{3,ijkl}, z_{4,k}$ are respectively $q_m \times n_{ijkl}$ ($m = 1, 2, 3, 4$) matrices of covariates associated with random effects due to distinct nesting levels. Since the random effects are usually assumed to have mean 0 and any non-zero mean for a term in the random effects must be expressed as part of the fixed effects terms, the covariates for random effects are usually taken as partial of those for the fixed effects. The random effects in different nesting levels are assumed to be independent. The random effects are all assumed to be normally distributed with mean 0 and covariance matrix D_m ($m = 1, 2, 3, 4$). If q_m , the dimension of the random effects at level m , is larger than 1, then D_m is assumed to be a symmetric positive-definite matrix. The random error term ε_{ijkl} is also assumed to be normally distributed with mean 0 and covariance matrix R_{ijkl} . Matrix R_{ijkl} is common to all individuals except for the dimension, which ensures the above notation works in case missing values occur. Matrix R_{ijkl} is a $n_{ijkl} \times n_{ijkl}$ positive symmetric matrix, which may result from two components: one is the individual heterogeneity and the other is the auto-correlation (Davidian and Giltinan 1995).

Model (3.10) is the general form of a multilevel mixed linear model based on the original design of our data. The dimensions of the fixed and random effects, corresponding covariates and variance covariance structure will be determined by the original data in the next section.

Model specification

We chose a widely used (Borders and Bailey 1986, Zhang and Borders 1999) log-linear basal area model as the basic model:

$$\ln(BA) = \beta_0 + \beta_1/A + \beta_2 \ln(HD) + \beta_3 \ln(TPH) + \varepsilon \quad (3.11)$$

where BA is basal area per hectare (m^2/ha), A is stand age, HD is the mean dominant-codominant height (m) of the stand, TPH is the number of trees per hectare, ε is the random error term and the β_i are coefficients.

In a multilevel mixed effects model context we will view the basic model (3.11) in the following ways:

1. The four parameters in (3.11) (β_0 to β_3) may vary on different level of factors systematically. For example, different silvicultural treatments (or treatment groups) and different soil types (or soil groups) may associate with different values of a coefficient in the model.
2. The responses [$\ln(BA)$] in the same factor level can be correlated and heterogeneity may exist among different levels of factors. This implies that random variation exists among the different levels of nesting in the data. For instance, the parameters in (3.11) may vary randomly from plot to plot so that an individual plot has its own basal area growth pattern. However, the randomness of the parameters for plots with the same silvicultural treatment may have some kind of association.
3. Prediction of basal area for a new individual (plot) can be obtained at different nesting levels and depends on the information on hand. Generally, with more information available, more precise prediction results.

To specify the basic model (3.11) in conjunction with multilevel nesting of our original data, we need to determine the dimension of the fixed and random effects and corresponding covariates in the model. Here attention will be given to the effects on the parameters in (3.11) due to the silvicultural treatments and soil groups (nonspodosol or spodosol). We take the view that the growth pattern of basal area can basically be described in model form (3.11), but silvicultural treatment, soil group differences, and some other factors (e.g. installations, individual plot) may systematically or randomly affect the parameters in the model. Firstly, we determine the fixed effect terms. This can be done by a stepwise-selection procedure with all factor levels and related interactions terms considered*. Because the silvicultural treatments (except for *CNTL*) are formed by combinations of five distinct silvicultural activities (chop, burn, bed, fertilizer and herbicide), we represent the 11 original silvicultural treatments by combinations of five dummy variables: *zchop*, *zburn*, *zf*, *zbed* and *zh*, which each have value 1 if the respective activity was applied in the original treatment, otherwise a zero value results. For example, treatment *UCBB* represents the combination of chopping, burning and bedding, so *zchop*, *zburn* and *zbed* have value 1 and *zf*, *zh* have value 0. There are two basic advantages with this approach. They are dimension reduction and flexibility in model application. The 11 silvicultural treatment levels are reduced to 5 dimensions with the dummy variable technique. Since the differences in response among some original silvicultural treatments may be trivial, this dimension reduction can be important. Contributions in the response due to silvicultural treatment can be explicitly traced

* The terms considered in stepwise procedure include: $1/A$, $\ln(HD)$, $\ln(TPH)$, *zchop*, *zburn*, *zbed*, *zf*, *zh*, *zsoil* (=1 if nonspodosol, 0 otherwise), and the respective interaction terms of silvicultural treatments and $1/A$, $\ln(HD)$, $\ln(TPH)$, *zsoil*. Totally, 34 terms are put in this stepwise variable choosing procedure.

separately to distinct activities (e.g. fertilization or herbicide) using this approach thus making it very easy to use in conjunction with different silvicultural treatments. For example, after fitting the model, one may even predict the potential response due to some new activity combination that is not available in the original design (i.e. not included in the original 11 treatments). This is very convenient in the application of the model. However, one should keep in mind that this dummy variable approach is based on the assumption that the effects of the silvicultural treatments on the response (basal area growth in logarithm form) are additive (e.g. Pienaar et al. 1998). There may, in fact, be no biological justification for this assumption. Hopefully, this shortcoming will be partially relieved by including the random effects terms in the model. A different approach using Gibbs sampling to overcome this deficiency is addressed in Chapter 4.

After stepwise screening ($\alpha=.05$) the following 11 terms remained in the model: $1/A$, $\ln(HD)$, $\ln(TPH)$, zf and interaction terms of zh and $1/A$, $zburn$ and $zsoil$, $zchop$ and $\ln(HD)$, $zchop$ and A , $zchop$ and $1/A$, $zbed$ and $\ln(TPH)$, $zburn$ and $1/A$. (Table 10). Thus, the fixed effects in (3.11) can be expressed as:

$$\left. \begin{aligned} \beta_0 &= \beta_{00} + \beta_{01}zf + \beta_{02}zburn \times zsoil + \beta_{03}zchop \times A \\ \beta_1 &= \beta_{10} + \beta_{11}zh + \beta_{12}zburn + \beta_{13}zchop \\ \beta_2 &= \beta_{20} + \beta_{21}zchop \\ \beta_3 &= \beta_{30} + \beta_{31}zbed \end{aligned} \right\} \quad (3.12)$$

Thus far, we haven't considered the random effects in the model. So the stepwise results are purely based on the fixed effects. It should be no surprise that the significant terms will change when random effects are also considered in the model. Ideally, the above stepwise procedure should be repeatedly applied with both fixed and random

effects in the model. Alternatively, we consider random effects based on these fixed-effect terms in the model and justify the final model with some diagnosis tools.

Table 10 The significant terms (at $\alpha = 0.05$) resulting from a stepwise regression procedure

Step	Variable (Term)		In	R ²	Partial R ²	Model C(p)	F	Prob>F
	Entered	Removed						
1	ln(HD)		1	0.9320	0.9320	4641.0280	13057.3784	0.0001
2	ln(TPH)		2	0.0321	0.9641	2001.1241	852.5379	0.0001
3	zh/A		3	0.0199	0.9840	367.5026	1183.3192	0.0001
4	1/A		4	0.0015	0.9855	244.5859	99.7581	0.0001
5	zburn*zsoil		5	0.0007	0.9862	192.2181	45.4494	0.0001
6	zf*ln(HD)		6	0.0003	0.9865	165.6686	24.4562	0.0001
7	zf*A		7	0.0003	0.9868	144.5851	20.1739	0.0001
8	zf/A		8	0.0002	0.9870	129.3454	15.2941	0.0001
9	zf		9	0.0002	0.9873	111.5526	17.8722	0.0001
10		zf*ln(HD)	8	0.0000	0.9873	109.8039	0.2269	0.6339
11	zchop*ln(HD)		9	0.0001	0.9873	106.7624	4.5732	0.0327
12	zchop*A		10	0.0009	0.9882	37.9411	68.8562	0.0001
13		zf*A	9	0.0000	0.9882	36.0078	0.0648	0.7992
14	zchop/A		10	0.0002	0.9884	19.8113	18.0282	0.0001
15		zf/A	9	0.0000	0.9884	20.3640	2.5291	0.1121
16	zbed*ln(TPH)		10	0.0001	0.9885	15.2802	7.0518	0.0081
17	zburn/A		11	0.0001	0.9885	12.4275	4.8504	0.0279

To determine which terms in (3.12) should be considered further as mixed (both fixed and random) and which should be considered as purely fixed, our first step is to choose all 4 terms in the basal area growth Model (3.11) to be mixed. However, this approach was stymied due to a singularity problem. Because the mortality is relatively low in the period of the observed data, the variation in TPA is small among the repeated measurements for an individual plot, so one should not consider both the intercept and

$\ln(TPH)$ to be mixed effects. By comparison with AIC , the intercept only was chosen to be mixed. Likelihood ratio test and AIC value comparison (see Chapter 2) are the main criterion used in evaluating the models with different random effect terms and random variance-covariance structures.

For comparison, the following models have been evaluated:

Model 1 the basic basal area growth model (3.11):

$$\left. \begin{aligned} \ln(BA) &= \beta_0 + \beta_1 / A + \beta_2 \ln(HD) + \beta_3 \ln(TPH) + \varepsilon \\ \varepsilon &\sim N(0, \sigma^2) \end{aligned} \right\} \quad (3.13)$$

In terms of the the general multilevel mixed effects model (2a –2e), for plot i , with 5 occasions of repeated measurments:

$$\mathbf{X}_i = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1/5 & 1/8 & 1/11 & 1/14 & 1/17 \\ \ln(HD_{i,1}) & \ln(HD_{i,2}) & \ln(HD_{i,3}) & \ln(HD_{i,4}) & \ln(HD_{i,5}) \\ \ln(TPH_{i,1}) & \ln(TPH_{i,2}) & \ln(TPH_{i,3}) & \ln(TPH_{i,4}) & \ln(TPH_{i,5}) \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix},$$

and $R_i = \sigma^2$. Neither random effects nor other covariates (silvicultural treatments or soil groups) have been considered in Model 1, and error terms are *niid*.

Model 2 Basic basal area growth model in conjunction with silvicultural treatments and soil groups based on equation (3.12):

$$\left. \begin{aligned} \ln(BA) &= \beta_{00} + \beta_{01}zf + \beta_{02}zburn \times zsoil + \beta_{03}zchop \times A \\ &+ (\beta_{10} + \beta_{11}zh + \beta_{12}zburn + \beta_{13}zchop) / A \\ &+ (\beta_{20} + \beta_{21}zchop) \ln(HD) \\ &+ (\beta_{30} + \beta_{31}zbed) \ln(TPH) + \varepsilon \\ \varepsilon &\sim N(0, \sigma^2) \end{aligned} \right\} \quad (3.14)$$

In this case,

$$\mathbf{X}_i = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ zf_i & zf_i & zf_i & zf_i & zf_i \\ zburn_i \times zsoil_i & zburn_i \times zsoil_i & zburn_i \times zsoil_i & zburn_i \times zsoil_i & zburn_i \times zsoil_i \\ 5 \times zchop_i & 8 \times zchop_i & 11 \times zchop_i & 14 \times zchop_i & 17 \times zchop_i \\ 1/5 & 1/8 & 1/11 & 1/14 & 1/17 \\ zh_i/5 & zh_i/8 & zh_i/11 & zh_i/14 & zh_i/17 \\ zburn_i/5 & zburn_i/8 & zburn_i/11 & zburn_i/14 & zburn_i/17 \\ zchop_i/5 & zchop_i/8 & zchop_i/11 & zchop_i/14 & zchop_i/17 \\ \ln(HD_{i,1}) & \ln(HD_{i,2}) & \ln(HD_{i,3}) & \ln(HD_{i,4}) & \ln(HD_{i,5}) \\ zchop_i \ln(HD_{i,1}) & zchop_i \ln(HD_{i,2}) & zchop_i \ln(HD_{i,3}) & zchop_i \ln(HD_{i,4}) & zchop_i \ln(HD_{i,5}) \\ \ln(TPH_{i,1}) & \ln(TPH_{i,2}) & \ln(TPH_{i,3}) & \ln(TPH_{i,4}) & \ln(TPH_{i,5}) \\ zbed_i \ln(TPH_{i,1}) & zbed_i \ln(TPH_{i,2}) & zbed_i \ln(TPH_{i,3}) & zbed_i \ln(TPH_{i,4}) & zbed_i \ln(TPH_{i,5}) \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_{00} \\ \beta_{01} \\ \beta_{02} \\ \beta_{03} \\ \beta_{10} \\ \beta_{11} \\ \beta_{12} \\ \beta_{13} \\ \beta_{20} \\ \beta_{21} \\ \beta_{30} \\ \beta_{31} \end{pmatrix},$$

and $R_i = \sigma^2$, without random effects, heterogeneity and correlation.

Model 3 In conjunction with heterogeneity and correlation structure with Model 2

(considering effects of the repeated measurements).

The model form and X-matrix are the same as Model 2, but heterogeneity and correlation are allowed in the error term. For 5 repeated measurements, the most parameterized form is an unstructured variance-covariance for an individual error term, i.e.,

$$\mathbf{R}_i = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} & \sigma_{15} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} & \sigma_{24} & \sigma_{25} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 & \sigma_{34} & \sigma_{35} \\ \sigma_{14} & \sigma_{24} & \sigma_{34} & \sigma_4^2 & \sigma_{45} \\ \sigma_{15} & \sigma_{25} & \sigma_{35} & \sigma_{45} & \sigma_5^2 \end{pmatrix}$$

Model 4 Mixed effects model with random effects of one-level random effects nesting and independent homogeneous error terms.

In this case, the fixed effects and error terms are the same as Model 2, but parameters β_{00} , β_{10} and β_{20} in (3.14) are allowed to vary from plot to plot. In terms of the

general form of (3a-3e), the X matrix and β parameters are the same as for Model 2.

With $j=1$, the individual plots are the only sources of random effects. Thus,

$$\mathbf{z}_{1,i} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1/5 & 1/8 & 1/11 & 1/14 & 1/17 \\ \ln(HD_{i,1}) & \ln(HD_{i,2}) & \ln(HD_{i,3}) & \ln(HD_{i,4}) & \ln(HD_{i,5}) \end{pmatrix}, \quad \mathbf{b}_{1,i} = \begin{pmatrix} b_{i,00} \\ b_{i,10} \\ b_{i,20} \end{pmatrix}$$

$$\mathbf{b}_{1,i} \sim N(\mathbf{0}, \mathbf{D}_1), \quad \mathbf{D}_1 = \begin{pmatrix} \varphi_1^2 & \varphi_{12} & \varphi_{13} \\ \varphi_{12} & \varphi_2^2 & \varphi_{23} \\ \varphi_{13} & \varphi_{23} & \varphi_3^2 \end{pmatrix}$$

This mixed-effects model is explored here because it is the most common modeling approach in a mixed-effect modelling context. A mixed-effects model is often presented with experimental units (individual plots here) as the level on which the random effects are considered.

Model 5 Mixed effects model with random effects of 4-level nesting and independent homogeneous error terms.

This is the general multilevel mixed linear model based on the original design (equation 3.10), but the simplest structure (*iid*) for within-plot error in the model that has been considered thus far.

In this case, the fixed effects and the error terms are the same as Model 2. Let individual plot, silvicultural treatment, soil type and installation be, respectively, at levels 1, 2, 3 and 4 in nesting. On nested Level 1, the general intercept in (3.14) (β_{00}) has been considered as random. Parameters β_{00} , β_{10} are considered to be random on Level 2 (silvicultural treatments). Again, β_{00} has been considered to be random on Level 3 (soil group). Parameter β_{00} , β_{10} and β_{20} in (3.14) are considered to be random at Level 4

(installation). We choose such terms as random effects basically after exploratory work with the data. But, there are some general guidelines (see e.g. Verbeke 1997). The random effects for time-invariant covariates (for example, silvicultural treatments and soil group) can be interpreted as subject-specific corrections to the overall mean structure. This makes them hard to distinguish from random intercepts and one often includes random intercepts and random effects only for time-varying covariates.

In the context of the general multilevel mixed effects linear model, there are 4 levels of nesting in random effects, $q_1 = 1$, $q_2 = 2$, $q_3 = 1$, and $q_4 = 3$, the dimension of the nested four levels are respectively 1,2,1, and 3. Using the same notation as for Model (3.10),

$$\mathbf{z}_{1,ijkl} = (1 \ 1 \ 1 \ 1 \ 1), \quad b_{1,ijkl} = (b_{ijkl,00}), \quad b_{1,ijkl} \sim N(0, D_1) \text{ and } D_1 = \varphi_{1,1}^2$$

$$\mathbf{z}_{2,ijkl} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1/5 & 1/8 & 1/11 & 1/14 & 1/17 \end{pmatrix}, \quad \mathbf{b}_{2,ijkl} = \begin{pmatrix} b_{ijkl,00} \\ b_{ijkl,10} \end{pmatrix}$$

$$\mathbf{b}_{2,ijkl} \sim N(\mathbf{0}, \mathbf{D}_2), \quad \mathbf{D}_2 = \begin{pmatrix} \varphi_{2,1}^2 & \varphi_{2,12} \\ \varphi_{2,12} & \varphi_{2,2}^2 \end{pmatrix}$$

$$\mathbf{z}_{3,ijkl} = (1 \ 1 \ 1 \ 1 \ 1), \quad b_{3,kl} = (b_{kl,00}), \quad b_{3,kl} \sim N(0, D_3) \text{ and } D_3 = \varphi_{3,1}^2$$

$$\mathbf{z}_{4,ijkl} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1/5 & 1/8 & 1/11 & 1/14 & 1/17 \\ \ln(HD_{ijkl,1}) & \ln(HD_{ijkl,2}) & \ln(HD_{ijkl,3}) & \ln(HD_{ijkl,4}) & \ln(HD_{ijkl,5}) \end{pmatrix},$$

$$\mathbf{b}_{4,1} = \begin{pmatrix} b_{l,00} \\ b_{l,10} \\ b_{l,20} \end{pmatrix}, \quad \mathbf{b}_{4,1} \sim N(\mathbf{0}, \mathbf{D}_4), \quad \mathbf{D}_4 = \begin{pmatrix} \varphi_{4,1}^2 & \varphi_{4,12} & \varphi_{4,13} \\ \varphi_{4,12} & \varphi_{4,2}^2 & \varphi_{4,23} \\ \varphi_{4,13} & \varphi_{4,23} & \varphi_{4,3}^2 \end{pmatrix}$$

Model 6 Mixed effects model with random effects of 3-level nesting and independent homogeneous error terms.

As mentioned above, the silvicultural treatments are nested in soil groups in an incomplete way in that partial treatments are nested in one level of the soil type (e.g. nonspodosol) and the others are nested in another level of soil type (e.g, spodosol) within the same installation. One may wonder if there is any evidence of deficiency to reduce Model 5 to a 3-level nesting by withdrawing the factor of soil group. So, in this case, every term is the same as Model 5 except that only three-level nesting (without the factor of soil type) is considered in the random effects. Let i, j, k index factor of plot, treatment and installation respectively, then,

$$\left. \begin{aligned} \mathbf{z}_{1,ijk} &= (1 \quad 1 \quad 1 \quad 1 \quad 1), \quad \mathbf{b}_{1,ijk} = (b_{ijk,00}), \quad b_{1,ijk} \sim N(0, D_1), \quad D_1 = \varphi_{1,1}^2 \\ \mathbf{z}_{2,ijk} &= \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1/5 & 1/8 & 1/11 & 1/14 & 1/17 \end{pmatrix}, \quad \mathbf{b}_{2,jk} = \begin{pmatrix} b_{jk,00} \\ b_{jk,10} \end{pmatrix} \\ \mathbf{b}_{2,jk} &\sim N(\mathbf{0}, \mathbf{D}_2), \quad \text{and} \quad \mathbf{D}_2 = \begin{pmatrix} \varphi_{2,1}^2 & \varphi_{2,12} \\ \varphi_{2,12} & \varphi_{2,2}^2 \end{pmatrix} \\ \mathbf{z}_{3,ijk} &= \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1/5 & 1/8 & 1/11 & 1/14 & 1/17 \\ \ln(HD_{ijk,1}) & \ln(HD_{ijk,2}) & \ln(HD_{ijk,3}) & \ln(HD_{ijk,4}) & \ln(HD_{ijk,5}) \end{pmatrix} \\ \mathbf{b}_{3,k} &= \begin{pmatrix} b_{k,00} \\ b_{k,10} \\ b_{k,20} \end{pmatrix}, \quad \mathbf{b}_{3,k} \sim N(\mathbf{0}, \mathbf{D}_3), \quad \mathbf{D}_3 = \begin{pmatrix} \varphi_{3,1}^2 & \varphi_{3,12} & \varphi_{3,13} \\ \varphi_{3,12} & \varphi_{3,2}^2 & \varphi_{3,23} \\ \varphi_{3,13} & \varphi_{3,23} & \varphi_{3,3}^2 \end{pmatrix} \end{aligned} \right\} \quad (3.15)$$

Model 7 Mixed effects model with random effects of 2-level nesting and independent homogeneous error terms.

Once there is no deficiency from Model 5 to Model 6 as is the case with our data (Table 14), one may further reduce Model 6 to 2-level nesting in random effects by withdrawing the plot factor. Because most of these silvicultural treatments (10 out of 11) only have one individual plot sampled, treatment within installation coincides with individual plot for most plots.

In this case, every term is the same as in Model 6, except that only 2-level nesting (treatment within installation) is considered in the random effects. Corresponding changes in notation are straightforward.

Model 8 Mixed effects model with random effects of 3-level nesting and heterogeneous and correlated error terms.

This is an extension of Model 6, the most efficient model thus far. Both heteroscedasticity and correlation in the error term (repeated factors) are allowed in this case. First, I determine the variance function and then based on the chosen variance function determine an appropriate within-plot auto-correlation structure for the error term. The most commonly used variance functions are the power and exponential functions (see Chapter 2). Based on a 3-level mixed effects approach such as Model 6 above (we call it Model 8a), the improvement in model fitting using either a power variance function or an exponential variance function is highly significant (Table 11). The model with exponential function (Model 8c) is even better than that with a power function (Model 8b). Based on Model 8c, first order autoregressive function [$AR(1)$],

moving average correlation model with order 2 [$MA(2)$], and mixed autoregressive-moving average model $ARMA(1,1)$ were tried separately to model the within-plot auto-correlation function. These we call Models 8d, 8e and 8f, respectively. Compared with the assumption of independent error, the likelihood ratios for these three correlation functions are respectively, 20.821, 21.298 and 20.817 (Table 12). With 1, 2, and 2 degree of freedom respectively, all three correlation structures are significant. The likelihood ratio values are very close to each other, but $AR(1)$ has only one parameter and the others have two parameters, so we prefer the $AR(1)$ to model the within-plot correlation structure. This choice is also supported by the AIC criteria since the model with $AR(1)$ has the smallest AIC value (Table 12). So the model chosen in this step is Model 8d.

Table 11 Comparison of mixed-effects model performance for slash pine basal area growth data with different variance functions

Model	Variance Function	Number of parameters	<u>AIC</u>	<u>log-likelihood</u>	<u>LRT</u>	<u>p-value</u>
8a	1*	23	-2498.932	1272.466		
8b	Power	24	-2585.397	1316.698	<u>8b vs 8a:</u> 88.465	<.0001
8c	Exponential	24	-2599.804	1323.902	<u>8c vs 8a:</u> 102.873	<.0001

*Variance function 1 means that the variances are homogeneous

Table 12 Comparison of mixed-effects model performance for slash pine basal area growth data with different within-plot correlation structures

Model	Correlation [*] Structure	Number of parameters	AIC	log- likelihood	LRT	p- value
8c	independent	24	-2599.804	1323.902		
8d	AR(1)	25	-2618.625	1334.313	<u>8d vs 8c:</u> 20.821	<.0001
8e	MA(2)	26	-2617.102	1334.551	<u>8e vs 8c:</u> 21.298	<.0001
8f	ARMA(1,1)	26	-2616.621	1334.311	<u>8f vs 8c:</u> 20.817	<.0001

^{*}refer to Pinheiro and Bates 1999;

Model 9 Final model

In Model (8d), the preferred model thus far, the interaction term of $zbed \times \ln(TPH)$ is not significant (the p-value is 0.2046). So the model was refitted without this interaction term. This results in an even larger loglikelihood value (1339.196 vs. 1334.313) and smaller AIC values (-2630.392 vs. -2618.625, see Table 13). So the term of $zbed \times \ln(TPH)$ should not be included in the final model. The final model, Model 9 is summarized as follows:

$$\begin{aligned}
 \ln(BA_{ijk}) &= \beta_{ijk,00} + \beta_{01}zf + \beta_{02}zburn \times zsoil + \beta_{03}zchop \times A \\
 &\quad + (\beta_{jk,10} + \beta_{11}zh + \beta_{12}zburn + \beta_{13}zchop) / A \\
 &\quad + (\beta_{k,20} + \beta_{21}zchop) \ln(HD) + \beta_{30} \ln(TPH) + \varepsilon_{ijk} \\
 \varepsilon_{ijk} &\sim N(0, \sigma^2 R_i) \\
 R_i &= diag(e^{0.5 \times \theta \ln(BA_{ijk})}) \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \rho^4 \\ \rho & 1 & \rho & \rho^2 & \rho^3 \\ \rho^2 & \rho & 1 & \rho & \rho^2 \\ \rho^3 & \rho^2 & \rho & 1 & \rho \\ \rho^4 & \rho^3 & \rho^2 & \rho & 1 \end{pmatrix} diag(e^{0.5 \times \theta \ln(BA_{ijk})}) \\
 \beta_{ijk,00} &= \beta_{00} + z_{1,ijk}^T b_{1,ijk} + z_{2,ijk}^T b_{2,ijk} + z_{3,k}^T b_{3,k} \\
 \beta_{jk,10} &= \beta_{10} + z_{2,ijk}^T b_{2,ijk} + z_{3,k}^T b_{3,k} \\
 \beta_{k,20} &= \beta_{20} + z_{3,k}^T b_{3,k}
 \end{aligned} \tag{3.16}$$

where, the random effects and corresponding design matrix are defined in (3.15).

The comparisons of Model 1 to Model 9 are presented in Table 13. With *REML* estimation, the value of *AIC* and log likelihood (thus likelihood ratio) can only be compared between models with exactly the same fixed effects structure. When models are fitted by maximum likelihood, these criteria can be compared between any models fitted to the same data (Bates and Pinheiro, 1998). Model 2 through Model 8 have the same fixed-effects terms [see (3.14)] and they all are fitted by *REML* (the comparison between *ML* and *REML* see e.g. Verbeke 1997). To compare with the basic basal area growth model (Model 1, without considering covariates of silvicultural treatments and soil type), Model 1 is fitted by *ML* and Model 2, in addition to *REML* fitting, is also fitted by *ML* (designated Model 2'). Covariates for silvicultural treatments and soil type definitely improve the performance of model fitting (Table 14). For example, comparing Model 2 vs. Model 1, the likelihood ratio (by *ML* estimation) is 588.793 with 8 degrees of freedom, which is highly significant. It is interesting to note that Model 3 with unstructured within-plot covariance for the error term but with no random effects is better than Model 4 with *n iid* error term and one-level random effects (individual plot) for three terms (intercept, $1/A$ and $\ln(HD)$]. This is not an unexpected result. As mentioned by several researchers (e.g. Davidian and Giltinan 1995, Bates and Pinheiro 1999), the variance-covariance matrix of the response has two components [see (3.2e)] that can be used to model heteroscedasticity and correlation. These are the random effects and the within group error structure. These two components may “compete” with each other in model specification, in that similar variance-covariance matrices and similar fitting performances may result from a more complex within-group error component

(unstructured, for example) being added to a simpler random effects component, or a simpler within group error component (*iid*, for example) being added to a more complex random effects component. The fact that Model 4 is not as good as Model 3 indicates that usual one-level mixed effects model can not sufficiently explain the heteroscedasticity and correlation in the data. This is supported by the performance of the multilevel mixed effects models (Models 5 through 9). The 4-level mixed-effects model (Model 5) has a much larger likelihood value (and smaller *AIC*) than the usual one-level mixed-effect model (Model 4) ($LRT = 127.391$, p -value < 0.0001). Withdrawing the factor of soil type, the 4-level mixed-effects model is reduced to a 3-level nesting in random effects. The high p -value (0.9967) suggests such a reduction, i.e. Model 6 is better than Model 5. Further reduction in levels of nesting in random effects (from Model 6 to Model 7) is not desirable because there is significance for the nested individual plot factor (with p -value = 0.0007 by LRT).

Table 13 Comparison of mixed-effects model performance for slash pine basal area growth data with different fixed and random effects terms

Model	Method	Number of parameters	<i>AIC</i>	<i>log-likelihood</i>	<i>LRT</i>	<i>p-value</i>
1	ML	5	-1496.394	753.197		
2*	ML	13	-2069.186	1047.593	<u>2* vs 1:</u> 588.793	<.0001
2	REML	13	-1982.681	1004.341		
3	REML	23	-2461.686	1253.843	<u>3 vs 2:</u> 499.005	<.0001
4	REML	19	-2379.541	1208.770	<u>4 vs 3:</u> 90.145	<.0001
5	REML	24	-2496.932	1272.466	<u>5 vs 4:</u> 127.391	<.0001
6	REML	23	-2498.932	1272.466	<u>6 vs 5:</u> 0.0002	0.9967
7	REML	22	-2489.371	1266.686	<u>7 vs 6:</u> 11.561	0.0007
8	REML	25	-2618.625	1334.313	<u>8 vs 7:</u> 123.694	<.0001
9	REML	24	-2630.392	1339.196		

The estimated parameters and the 95% confidence intervals of the estimates for Model 9 are given in Table 14. All terms remaining in the model are significant. The best linear unbiased predictors (*BLUP*'s) for the random effects are also available at three nesting levels. But, only the installation level was given here (Table 15). There are $16 \times 11 = 176$ pairs of *BLUP*'s at treatment level and 191 at individual plot level but both sets are too long to give in detail here. The magnitude of the *BLUP*'s for the random intercept at plot level is very small. This can be explained by the structure of our original data. Because there was only one plot established for most silvicultural treatments within an installation, for most plots (176 out of 191) the level of treatment within installation coincides with the plot level. So after estimating (predicting) the random effects at treatment level, there is little residual variation on the plot level. But the level of plot is still significant by the *LRT* (Table 13), so we keep this level of nesting anyway.

Instead of $\ln(BA)$ the responses for all the above model fitting are $\ln(BA+1)$. We added the constant 1 to the original observations of basal area before log transformation because in metric units for a lot of plots the basal area per hectare (m^2/ha) at age 5 was smaller than 1 and the logarithm function is very steep in $(0,1]$ (Figure 21). Thus, for some plots there were unstable errors in model fitting because a very small change in a fitted value can result a relative large residual. This shortcoming was overcome by adding the positive constant (1.0) to all basal area values before applying the logarithm transformation.

Table 14 Approximate 95% confidence intervals for the estimated parameters of Model 9 (with silvicultural treatment and soil type covariates, 3-level nesting of random effects, heteroscedasticity (exponential function) and correlated [AR(1)] within group errors.)

Fixed effects:

Parameters	Lower bound	Estimates	Upper bound
β_{00}	-4.83715093	-4.42494111	-4.012731298
β_{01}	0.03544299	0.05376615	0.072089299
β_{02}	-0.00248073	0.02936658	0.061213883
β_{03}	-0.01430656	-0.01161126	-0.008915959
β_{10}	-3.14496129	-2.33859482	-1.532228343
β_{11}	1.66069915	1.82791131	1.995123460
β_{12}	0.13182687	0.34608053	0.560334191
β_{13}	-0.81582575	-0.47532666	-0.134827569
β_{20}	1.00742912	1.07945054	1.151471957
β_{21}	0.07907912	0.10326761	0.127456113
β_{30}	0.62679540	0.67282902	0.718862639

Random Effects:

Level 3: installation

Parameters	Lower bound	Estimates	Upper bound
$\varphi_{3,1}$	0.27382111	0.3838393	0.5380615
$\varphi_{3,2}$	0.79716000	1.0210522	1.3078272
$\varphi_{3,3}$	0.06957048	0.1059668	0.1614042
$\varphi_{3,12}$	-0.680065	-0.378761	-0.21095
$\varphi_{3,13}$	-0.0852112	-0.0399088	-0.0186914
$\varphi_{3,23}$	0.0516347	0.10073705	1.965335

Level 2: silvicultural treatments

$\varphi_{2,1}$	0.05066287	0.07013162	0.09708183
$\varphi_{2,2}$	0.12490630	0.31058677	0.77229201
$\varphi_{2,12}$	-0.06750306	-0.0196110	-0.00569742

Level 1: individual plot

$\varphi_{1,1}$	0.007475583	0.02326395	0.0723972
-----------------	-------------	------------	-----------

Correlation structure: [AR(1): $\text{corr}(y_{ij}, y_{ik}) = \rho^{|k-j|}$]

ρ	0.1203269	0.3556445	0.5531224
--------	-----------	-----------	-----------

Variance function: [Exponential: $\exp(\theta E[y_i])$]

θ	-0.5880978	-0.4611205	-0.3341433
----------	------------	------------	------------

Within-group standard error:

σ	0.1120602	0.1583093	0.2236462
----------	-----------	-----------	-----------

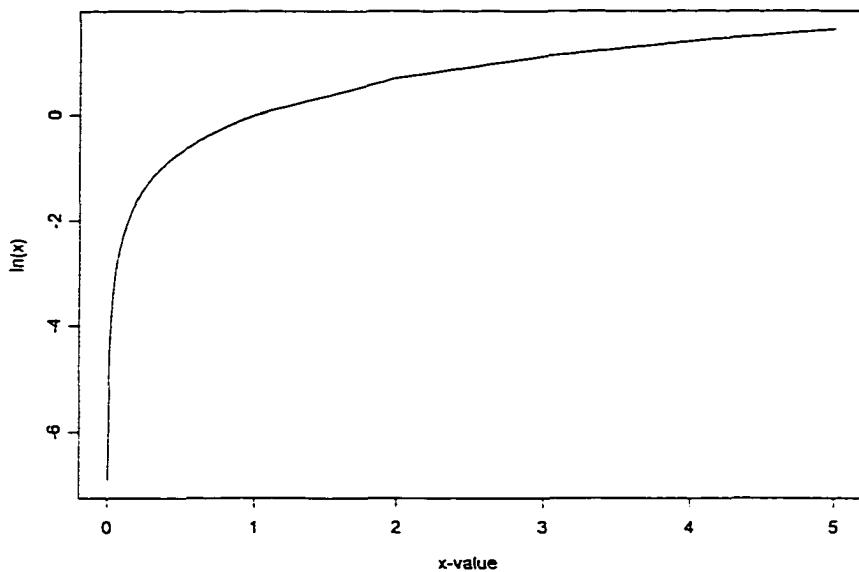


Figure 21: The logarithm curve.

Table 15 Best linear unbiased predictors (BLUP's) for random effects at installation level

<u>Installations</u>	<u>intercept</u>	<u>$1/A$</u>	<u>$\ln(HD)$</u>
1	-0.42659439	0.89609746	0.13990719
4	0.54586289	-1.78161562	-0.10194712
5	0.49164284	-0.87176473	-0.12615299
6	-0.21448373	0.40854570	0.06464945
7	-0.30101010	0.92419057	0.07539121
8	-0.51113516	1.40057492	0.14373405
9	0.20129974	-0.19593652	-0.05181340
10	0.02425167	-0.35289962	0.01261805
11	0.08548702	0.07278229	-0.02823776
12	0.11628054	-0.21005375	-0.05793848
13	0.23094471	-0.57675182	-0.07660199
14	0.09188446	-0.28826864	-0.04071379
15	0.23549982	-0.54044157	-0.07795928
16	-0.47388713	1.19688074	0.12275889
19	0.29699364	-0.90218484	-0.08852709
20	-0.39303680	0.82084543	0.09083304

Plots of the residuals, fitted values, and estimated random effects are often used as diagnostic tools for model checking. The residual plot (residuals or standardized residuals against the fitted values) can be used to check the randomness of the residuals. By the model assumption, the residual plot should not reveal any observable pattern on the whole or by sub-groups (e.g. treatment groups, soil types). There is no obvious violation for Model 9 by residual plots (Figures 22-24). The Q-Q plot is used to check the normality assumptions for the residuals and random effects. For a normal distribution, the Q-Q plot is a straight line. There is no strong violation of the straight line for the Q-Q plot the residuals (Figure 25). For random effects, there are three levels of nesting with no indication of violation of the assumptions here either (Figure 26-28). Another intuitive diagnosis plot is the observed values against the fitted values. If the model fits the data generally well, this plot should form an approximately 45 degree straight line. This is the case for Model 9 (Figure 29).

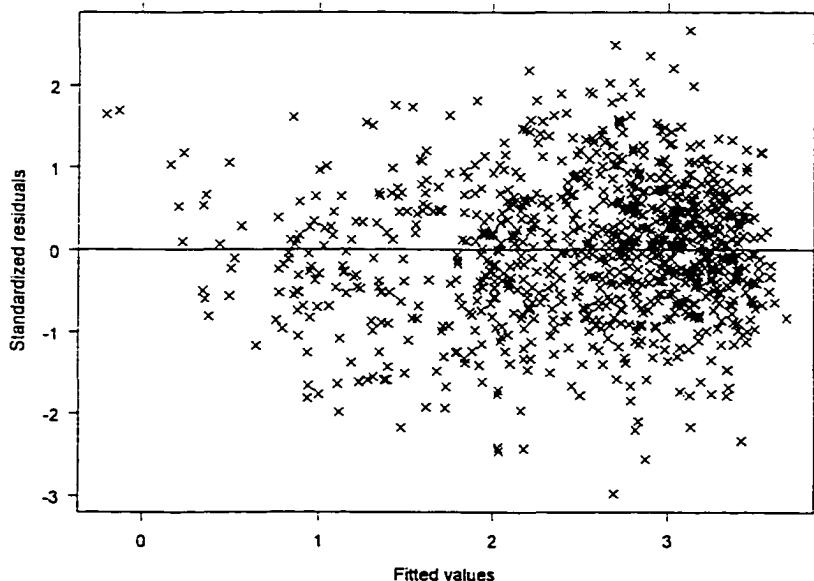


Figure 22: The standardized residual plot of the multilevel linear mixed-effects basal area model [Model 9] for the slash pine data.

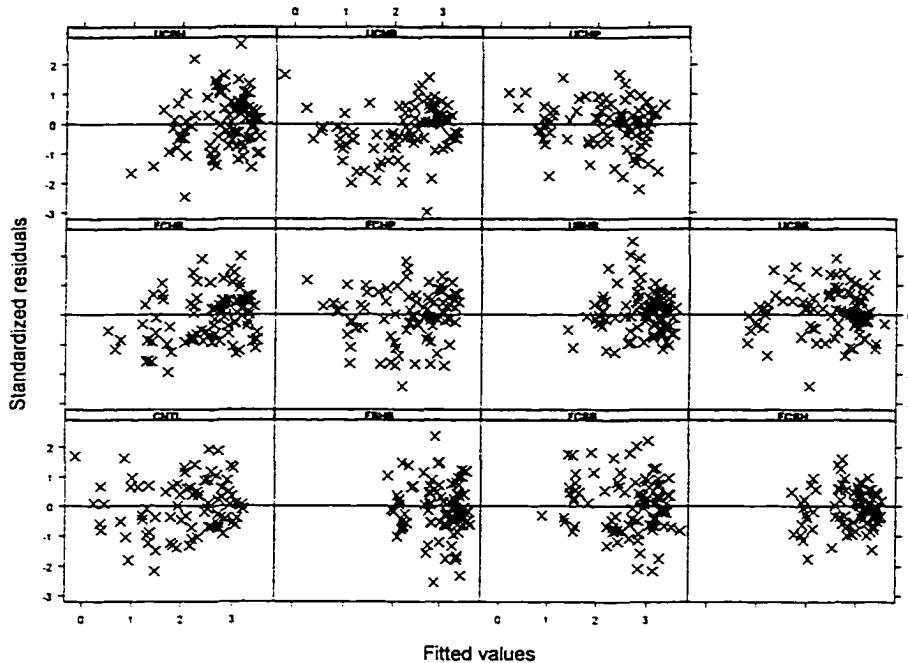


Figure 23: The standardized residual plot conditioning silvicultural treatments with the multilevel linear mixed-effects basal area model.

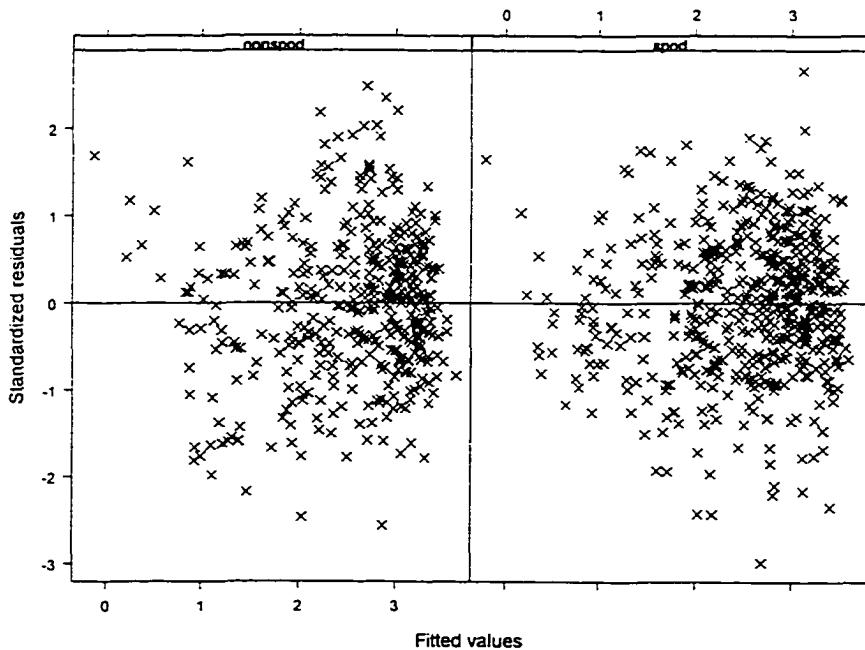


Figure 24: The standardized residual plot by soil type groups (spodosol or nonspdosol) with the multilevel linear mixed-effects basal area model.

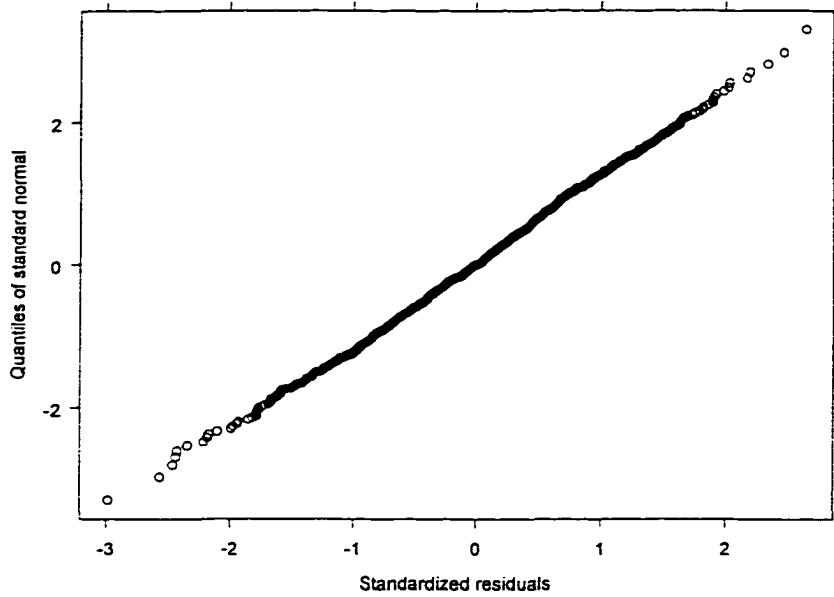


Figure 25: Normal plot of residuals for the multilevel linear mixed-effects basal area growth model with $AR(1)$ autocorrelation and exponential variance function [Model 9].

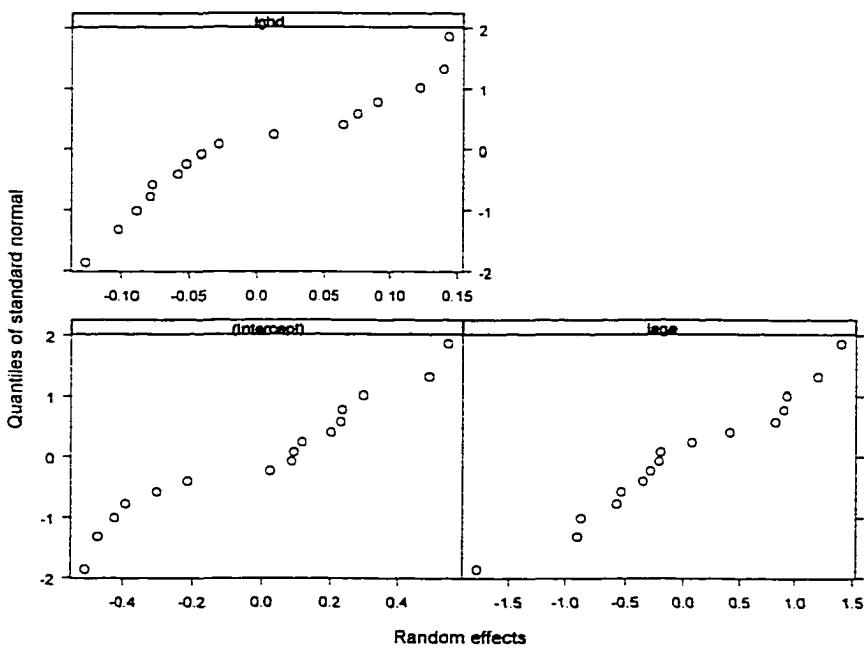


Figure 26: Normal plot of random effects at installation level for the multilevel linear mixed-effects basal area growth model based on the slash pine data.

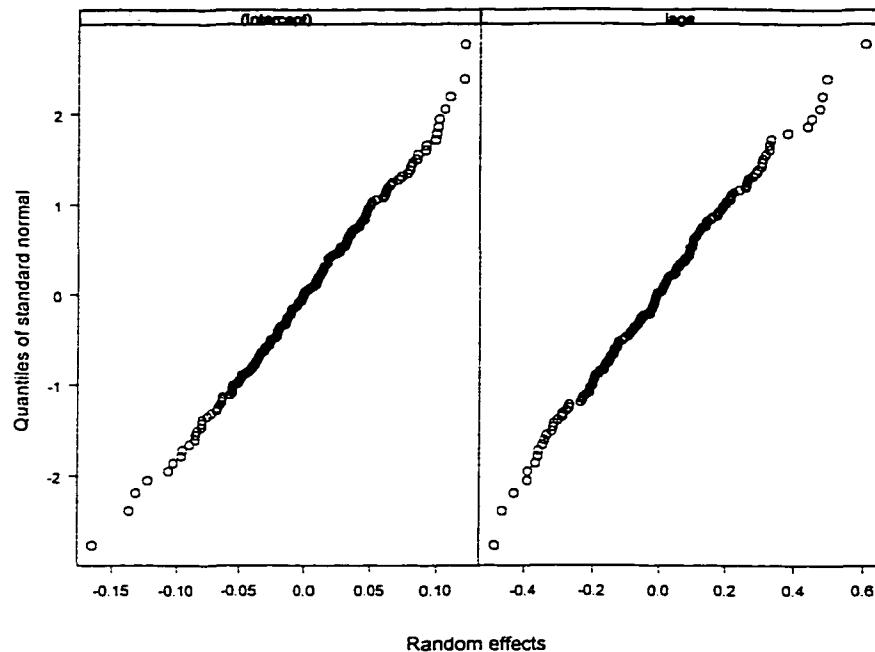


Figure 27: Normal plot of random effects at treatment within installation level for the multilevel linear mixed-effects basal area growth model based on the slash pine data.

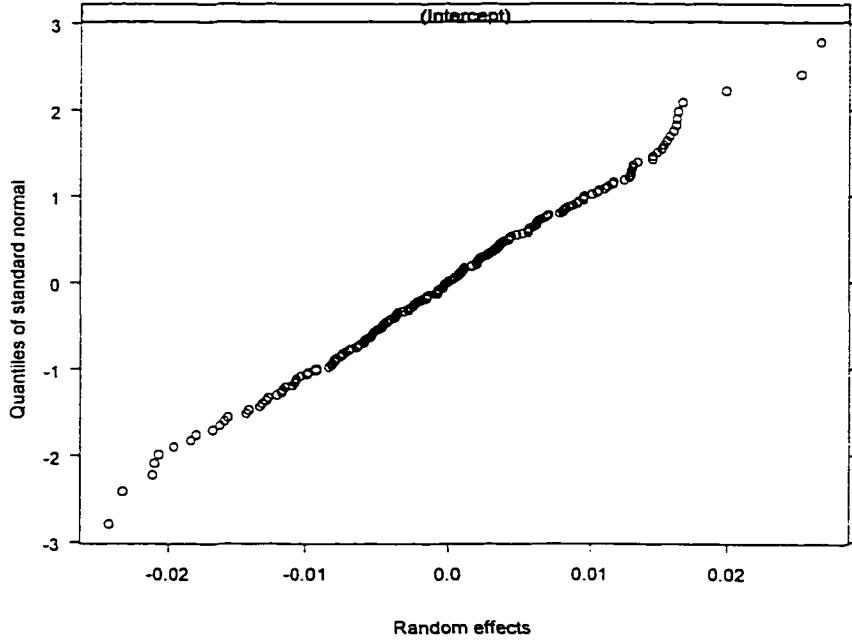


Figure 28: Normal plot of random effects at individual plot level for the multilevel linear mixed-effects basal area growth model based on the slash pine data.

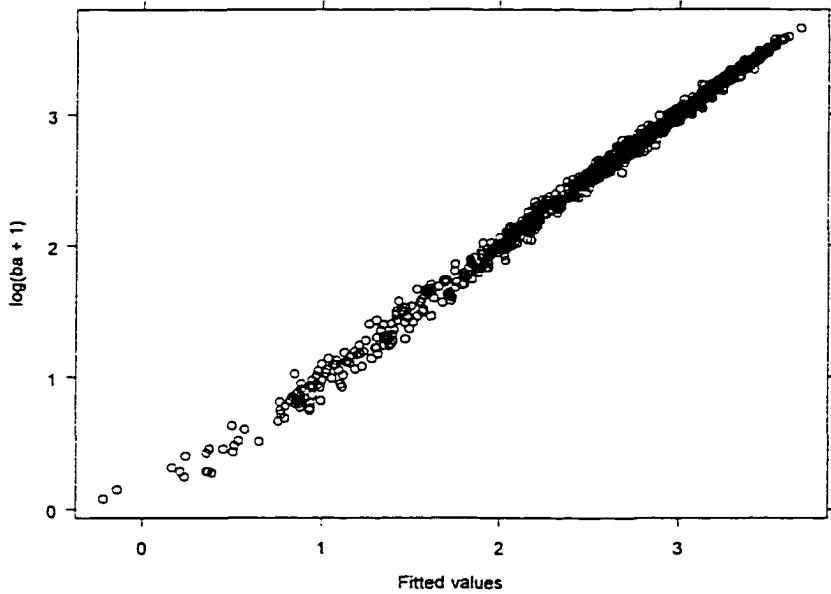


Figure 29: The observed response [$\ln(BA+1)$] versus predicted with the multilevel linear mixed effect basal area growth model.

Predictions

Predictions based on fitted models have been of great interest to the forester.

Predictions with the above model actually present two distinct representations. One refers to the fitted values, which are the predicted values for the responses under the fitted model. As shown above, these quantities are very useful for model checking. The other meaning of prediction refers to a prediction for a new observation based on the fitted model.

In multilevel mixed-effects models such as Model 9, fitted values and prediction for new observations may be obtained at different levels of nesting, or at the population level. For Model 9, there are three levels of nesting in random effects (plot-treatment-installation), so fitted values and prediction can be obtained at 4 distinct levels. Let y_{ijk}

and \mathbf{x}_{ijk} be vectors (with dimension 5×1) of response [$\ln(BA+1)$] and fixed effects covariates of plot i , treatment j and installation k . Variable $\mathbf{z}_{h,ijk}$ is the vector of covariates corresponding to random effects associated with y_{ijk} at h^{th} level of nesting ($h=1,2,3$). Then the marginal expected value and conditional expectations given the random effects are respectively:

Population mean:

$$E[\mathbf{y}_{ijk}] = \mathbf{x}_{ijk}^T \boldsymbol{\beta} \quad (3.17)$$

The installation-level prediction:

$$E[\mathbf{y}_{ijk} | \mathbf{b}_{3,k}] = \mathbf{x}_{ijk}^T \boldsymbol{\beta} + \mathbf{z}_{3,ijk}^T \mathbf{b}_{3,k} \quad (3.18)$$

The treatment-level prediction:

$$E[\mathbf{y}_{ijk} | \mathbf{b}_{3,k}, \mathbf{b}_{2,jk}] = \mathbf{x}_{ijk}^T \boldsymbol{\beta} + \mathbf{z}_{2,ijk}^T \mathbf{b}_{2,jk} + \mathbf{z}_{3,ijk}^T \mathbf{b}_{3,k} \quad (3.19)$$

The plot-level prediction

$$E[\mathbf{y}_{ijk} | \mathbf{b}_{3,k}, \mathbf{b}_{2,jk}, \mathbf{b}_{1,ijk}] = \mathbf{x}_{ijk}^T \boldsymbol{\beta} + \mathbf{z}_{1,ijk}^T \mathbf{b}_{1,ijk} + \mathbf{z}_{2,ijk}^T \mathbf{b}_{2,jk} + \mathbf{z}_{3,ijk}^T \mathbf{b}_{3,k} \quad (3.20)$$

The best linear unbiased predictors (*BLUP*'s) at the above four levels can be obtained by replacing $\boldsymbol{\beta}$ with $\hat{\boldsymbol{\beta}}$, and the random effects with their *BLUP*'s. It will be no surprise that the most precise predictions are obtained at the plot level and the least precise predictions are at the general population mean level. The more specific the information available, the more precise a prediction can be obtained. This can easily be shown as follows:

For Model 9,

$$\mathbf{y}_{ijk} = \mathbf{x}_{ijk}^T \boldsymbol{\beta} + \mathbf{z}_{1,ijk}^T \mathbf{b}_{1,ijk} + \mathbf{z}_{2,ijk}^T \mathbf{b}_{2,jk} + \mathbf{z}_{3,ijk}^T \mathbf{b}_{3,k} + \varepsilon_{ijk} \quad (3.21)$$

The prediction at a population mean level is:

$$\hat{\mathbf{y}}_{ijk} = \mathbf{x}_{ijk}^T \hat{\beta} \quad (3.22a)$$

$$\begin{aligned} \text{var}(\hat{\mathbf{y}}_{ijk} - \mathbf{y}_{ijk}) &= \text{var}[\mathbf{x}_{ijk}^T (\beta - \hat{\beta}) - \mathbf{z}_{1,ijk}^T \hat{\mathbf{b}}_{1,ijk} - \mathbf{z}_{2,ijk}^T \hat{\mathbf{b}}_{2,ijk} - \mathbf{z}_{3,ijk}^T \hat{\mathbf{b}}_{3,k} - \varepsilon_{ijk}] \\ &= \mathbf{x}_{ijk}^T \text{var}(\beta - \hat{\beta}) \mathbf{x}_{ijk} + \mathbf{z}_{1,ijk}^T \hat{\mathbf{D}}_1 \mathbf{z}_{1,ijk} + \mathbf{z}_{2,ijk}^T \hat{\mathbf{D}}_2 \mathbf{z}_{2,ijk} + \mathbf{z}_{3,ijk}^T \hat{\mathbf{D}}_3 \mathbf{z}_{3,ijk} + \hat{\mathbf{R}} \end{aligned}$$

let $\hat{\Omega}(\hat{\beta}) = \text{var}(\hat{\beta} - \beta)$, then

$$\text{var}(\hat{\mathbf{y}}_{ijk} - \mathbf{y}_{ijk}) = \mathbf{x}_{ijk}^T \hat{\Omega}(\hat{\beta}) \mathbf{x}_{ijk} + \mathbf{z}_{1,ijk}^T \mathbf{D}_1 \mathbf{z}_{1,ijk} + \mathbf{z}_{2,ijk}^T \mathbf{D}_2 \mathbf{z}_{2,ijk} + \mathbf{z}_{3,ijk}^T \mathbf{D}_3 \mathbf{z}_{3,ijk} + \mathbf{R} \quad (3.22b)$$

The prediction at an installation level is:

$$E(\hat{\mathbf{y}}_{ijk} | \hat{\mathbf{b}}_{3,k}) = \mathbf{x}_{ijk}^T \hat{\beta} + \mathbf{z}_{3,ijk}^T \hat{\mathbf{b}}_{3,k} \quad (3.23a)$$

Similarly, in this case,

$$\begin{aligned} \text{var}(\hat{\mathbf{y}}_{ijk} - \mathbf{y}_{ijk}) &= \mathbf{x}_{ijk}^T \hat{\Omega}(\hat{\beta}) \mathbf{x}_{ijk} + \mathbf{z}_{1,ijk}^T \mathbf{D}_1 \mathbf{z}_{1,ijk} + \mathbf{z}_{2,ijk}^T \mathbf{D}_2 \mathbf{z}_{2,ijk} + \mathbf{z}_{3,ijk}^T \mathbf{V}_{3,k} \mathbf{z}_{3,ijk} + \\ &\quad \mathbf{x}_{ijk}^T \mathbf{C}_{3,k} \mathbf{z}_{3,ijk} + \mathbf{z}_{3,ijk}^T \mathbf{C}_{3,k}^T \mathbf{x}_{ijk} + \mathbf{R} \end{aligned} \quad (3.23b)$$

where

$$\begin{aligned} \mathbf{V}_{3,k} &= \text{var}(\hat{\mathbf{b}}_{3,k} - \mathbf{b}_{3,k}) \\ &= \mathbf{D}_3 - \mathbf{D}_3 \mathbf{z}_{3,ijk}^T \Sigma_3^{-1} \mathbf{z}_{3,ijk} \mathbf{D}_3 + \mathbf{D}_3 \mathbf{z}_{3,ijk}^T \Sigma_3^{-1} \mathbf{x}_{ijk} \Omega(\hat{\beta}) \mathbf{x}_{ijk}^T \Sigma_3^{-1} \mathbf{z}_{3,ijk} \mathbf{D}_3 \end{aligned} \quad (3.23c)$$

$$\mathbf{C}_{3,k} = \text{Cov}[(\hat{\beta} - \beta), (\hat{\mathbf{b}}_{3,k} - \mathbf{b}_{3,k})^T] = -\Omega(\hat{\beta}) \mathbf{x}_{ijk}^T \Sigma_3^{-1} \mathbf{z}_{3,ijk} \mathbf{D}_3 \quad (3.23d)$$

$$\Sigma_3 = \mathbf{z}_{1,ijk}^T \mathbf{D}_1 \mathbf{z}_{1,ijk} + \mathbf{z}_{2,ijk}^T \mathbf{D}_2 \mathbf{z}_{2,ijk} + \mathbf{z}_{3,ijk}^T \mathbf{D}_3 \mathbf{z}_{3,ijk} + \mathbf{R} \quad (3.23e)$$

The prediction at the treatment level is:

$$E(\hat{\mathbf{y}}_{ijk} | \hat{\mathbf{b}}_{3,k}, \hat{\mathbf{b}}_{2,jk}) = \mathbf{x}_{ijk}^T \hat{\beta} + \mathbf{z}_{2,ijk}^T \hat{\mathbf{b}}_{2,jk} + \mathbf{z}_{3,ijk}^T \hat{\mathbf{b}}_{3,k} \quad (3.24a)$$

with variance

$$\begin{aligned} \text{var}(\hat{\mathbf{y}}_{ijk} - \mathbf{y}_{ijk}) &= \mathbf{x}_{ijk}^T \hat{\Omega}(\hat{\beta}) \mathbf{x}_{ijk} + \mathbf{z}_{1,ijk}^T \mathbf{D}_1 \mathbf{z}_{1,ijk} + \mathbf{z}_{2,ijk}^T \mathbf{V}_{2,jk} \mathbf{z}_{2,ijk} + \mathbf{x}_{ijk}^T \mathbf{C}_{2,jk} \mathbf{z}_{2,ijk} + \\ &\quad \mathbf{z}_{2,ijk}^T \mathbf{C}_{2,jk}^T \mathbf{x}_{ijk} + \mathbf{z}_{3,ijk}^T \mathbf{V}_{3,k} \mathbf{z}_{3,ijk} + \mathbf{x}_{ijk}^T \mathbf{C}_{3,k} \mathbf{z}_{3,ijk} + \mathbf{z}_{3,ijk}^T \mathbf{C}_{3,k}^T \mathbf{x}_{ijk} + \mathbf{R} \end{aligned} \quad (3.24b)$$

Where

$$\begin{aligned}\mathbf{V}_{2,jk} &= \text{var}(\hat{\mathbf{b}}_{2,k} - \mathbf{b}_{2,k}) \\ &= \mathbf{D}_2 - \mathbf{D}_2 \mathbf{z}_{2,jk}^T \Sigma_2^{-1} \mathbf{z}_{2,jk} \mathbf{D}_2 + \mathbf{D}_2 \mathbf{z}_{2,jk}^T \Sigma_2^{-1} \mathbf{x}_{ijk} \hat{\Omega}(\hat{\beta}) \mathbf{x}_{ijk}^T \Sigma_2^{-1} \mathbf{z}_{2,jk} \mathbf{D}_2\end{aligned}\quad (3.24c)$$

$$\mathbf{C}_{2,jk} = \text{Cov}[(\hat{\beta} - \beta), (\hat{\mathbf{b}}_{2,jk} - \mathbf{b}_{2,jk})^T] = -\hat{\Omega}(\hat{\beta}) \mathbf{x}_{ijk}^T \Sigma_2^{-1} \mathbf{z}_{2,jk} \mathbf{D}_2 \quad (3.24d)$$

$$\Sigma_2 = \mathbf{z}_{1,jk}^T \mathbf{D}_1 \mathbf{z}_{1,jk} + \mathbf{z}_{2,jk}^T \mathbf{D}_2 \mathbf{z}_{2,jk} + \mathbf{z}_{3,jk}^T \mathbf{V}_{3,k} \mathbf{z}_{3,jk} + \mathbf{R} \quad (3.24e)$$

The prediction at the plot level is:

$$E(\hat{\mathbf{y}}_{ijk} | \hat{\mathbf{b}}_{3,k}, \hat{\mathbf{b}}_{2,jk}, \hat{b}_{1,jk}) = \mathbf{x}_{ijk}^T \hat{\beta} + z_{1,jk} \hat{b}_{1,jk} + \mathbf{z}_{2,jk}^T \hat{\mathbf{b}}_{2,jk} + \mathbf{z}_{3,jk}^T \hat{\mathbf{b}}_{3,k} \quad (3.25a)$$

with variance

$$\begin{aligned}\text{var}(\hat{\mathbf{y}}_{ijk} - \mathbf{y}_{ijk}) &= \mathbf{x}_{ijk}^T \hat{\Omega}(\hat{\beta}) \mathbf{x}_{ijk} + \mathbf{z}_{1,jk}^T \mathbf{V}_{1,jk} \mathbf{z}_{1,jk} + \mathbf{x}_{ijk}^T \mathbf{C}_{1,jk} \mathbf{z}_{1,jk} + \mathbf{z}_{1,jk}^T \mathbf{C}_{1,jk}^T \mathbf{x}_{ijk} + \\ &\quad \mathbf{z}_{2,jk}^T \mathbf{V}_{2,jk} \mathbf{z}_{2,jk} + \mathbf{x}_{ijk}^T \mathbf{C}_{2,jk} \mathbf{z}_{2,jk} + \mathbf{z}_{2,jk}^T \mathbf{C}_{2,jk}^T \mathbf{x}_{ijk} + \\ &\quad \mathbf{z}_{3,jk}^T \mathbf{V}_{3,k} \mathbf{z}_{3,jk} + \mathbf{x}_{ijk}^T \mathbf{C}_{3,k} \mathbf{z}_{3,jk} + \mathbf{z}_{3,jk}^T \mathbf{C}_{3,k}^T \mathbf{x}_{ijk} + \mathbf{R}\end{aligned}\quad (3.25b)$$

where

$$\begin{aligned}\mathbf{V}_{1,jk} &= \text{var}(\hat{\mathbf{b}}_{1,jk} - \mathbf{b}_{1,jk}) \\ &= \mathbf{D}_1 - \mathbf{D}_1 \mathbf{z}_{1,jk}^T \Sigma_1^{-1} \mathbf{z}_{1,jk} \mathbf{D}_1 + \mathbf{D}_1 \mathbf{z}_{1,jk}^T \Sigma_1^{-1} \mathbf{x}_{ijk} \hat{\Omega}(\hat{\beta}) \mathbf{x}_{ijk}^T \Sigma_1^{-1} \mathbf{z}_{1,jk} \mathbf{D}_1\end{aligned}\quad (3.25c)$$

$$\mathbf{C}_{1,jk} = \text{Cov}[(\hat{\beta} - \beta), (\hat{\mathbf{b}}_{1,jk} - \mathbf{b}_{1,jk})^T] = -\hat{\Omega}(\hat{\beta}) \mathbf{x}_{ijk}^T \Sigma_1^{-1} \mathbf{z}_{1,jk} \mathbf{D}_1 \quad (3.25d)$$

$$\Sigma_1 = \mathbf{z}_{1,jk}^T \mathbf{D}_1 \mathbf{z}_{1,jk} + \mathbf{z}_{2,jk}^T \mathbf{V}_{2,jk} \mathbf{z}_{2,jk} + \mathbf{z}_{3,jk}^T \mathbf{V}_{3,k} \mathbf{z}_{3,jk} + \mathbf{R} \quad (3.25e)$$

The prediction we discussed thus far is “fitting type prediction”. That is to say the prediction is on the same experimental unit as the original data. However, predictions on new units (but in the same population) are often of great interest. In the next section we will focus on this objective.

Individual prediction for new observations

Suppose we are interested in the prediction of a single observation for a new plot.

This plot may or may not contain prior observations, and it may or may not be from one of the original installations and silvicultural treatments. In the following we show how multilevel mixed effects can be applied in these different situations.

Case 1 *Prediction is required for a new plot with no previous observations but the plot is identified with the original installations and treatments.*

Since the new plot matches one of the installations and one of the silvicultural treatments in the original data, the prediction for the conditional expectation for the plot can be obtained either by (3.23a)-(3.23e) at the installation level or (3.24a) – (3.24e) at the silvicultural treatment level, depending on the availability of information about level of nesting.

For example, suppose we are interested in predicting the basal area at age 15 for a plot with silvicultural treatment *FCBB* ($zchop=1$, $zburn=1$, $zbed=1$, $zf=1$ and $zh=0$), nonspodosol soil type ($zsoil=1$). Suppose the plot is in the same region as Installation 4. We also know that the dominant height at age 15 is 16.8 m and the stand density is 680 trees/ha . For this case,

$$\mathbf{x}_h^T = (1 \ 1 \ 1 \ 15 \ 1/15 \ 0 \ 1/15 \ 1/15 \ \ln(16.8) \ \ln(16.8) \ \ln(680))$$

$$z_{1,h} = 1, \quad \mathbf{z}_{2,h}^T = (1 \ 1/15), \quad \mathbf{z}_{3,h}^T = (1 \ 1/15 \ \ln(16.8))$$

$$\hat{\mathbf{b}}_{2,h}^T = (0.098915 \ -0.36387) \text{ (form the BLUP, details not given here)}$$

$$\hat{\mathbf{b}}_{3,h}^T = (0.54586 \ -1.78162 \ -0.10195) \text{ (form the BLUP, see Table 15)}$$

$$\mathbf{D}_1 = 0.000541211,$$

$$\mathbf{D}_2 = \begin{pmatrix} 0.004918444 & -0.01960376 \\ -0.01960376 & 0.09646414 \end{pmatrix}$$

$$\mathbf{D}_3 = \begin{pmatrix} 0.14733261 & -0.3785947 & -0.0399014 \\ -0.3785947 & 1.0425476 & 0.1007320 \\ -0.0399014 & 0.1007320 & 0.01122896 \end{pmatrix}$$

and $\mathbf{R} = \hat{\sigma}^2 \exp(\hat{\theta}\hat{\mu}_h)$, $\hat{\sigma} = 0.1583093$, $\hat{\theta} = -0.4611205$, $\hat{\mu}_h$ is the predicted value(s).

The predictions can be obtained at the following three levels:

1. *At the population level* [using (3.22a)-(3.22b)]:

$$\ln(\hat{BA} + 1) = 3.044656, \text{ thus } \hat{BA} = 20.003 \text{ m}^2/\text{ha}.$$

With variance in $\ln(\hat{BA} + 1)$ [using (26b)] of 0.01355012. The 95% confidence interval for the predicted $\ln(\hat{BA} + 1)$ is approximately:

$3.04829 \pm 1.96 \times \sqrt{0.01355012}$, i.e., [2.8165, 3.27281] Transforming back to the original scale for basal area, the approximate 95% CI for the prediction at the population level is: $[15.718 \text{ m}^2/\text{ha.}, 25.385 \text{ m}^2/\text{ha.}]$

2. *At the installation level* [using (3.23a)-(3.23b)]:

$$\ln(\hat{BA} + 1) = 3.184102, \text{ thus } \hat{BA} = 23.146 \text{ m}^2/\text{ha.}$$

With variance in $\ln(\hat{BA} + 1)$ [using (3.23b)] of 0.01189616, which is 87.8% of the variance at population level. The approximate 95% confidence interval for $\ln(\hat{BA} + 1)$ at the installation level is [2.970325, 3.397878] Transforming back to the original scale, the 95% CI for basal area prediction at installation level is $[18.498 \text{ m}^2/\text{ha.}, 28.90 \text{ m}^2/\text{ha.}]$.

3. *At the treatment within installation level* [using (3.24a)-(3.24b)]:

$$\ln(\hat{BA} + 1) = 3.258759, \text{ thus } \hat{BA} = 25.017 \text{ m}^2/\text{ha.}$$

with variance in $\ln(B\hat{A} + 1)$ [using (26b)] of 0.01095334, which is 80.8% of the variance at population level. The approximate 95% confidence interval for $\ln(B\hat{A} + 1)$ at the installation level is [3.053628, 3.463889]. Transforming back to the original scale, the CI for basal prediction is [$20.192m^2/ha.$, $30.940m^2/ha.$], with mean $25.017 m^2/ha.$

Case 2 *Prediction is required for a new plot with no prior observations available and the plot is not identified with the original installations and treatments.*

In this case, since the new plot does not match any of the original installation and silvicultural treatment combinations, the *BLUP* for the random effects from the original data can not be used . Thus, prediction is only available at the population level. Assuming the same FCBB treatment as in the above example, suppose we do not know which installation the plot belongs to. In this case the prediction is the same as that obtained at population level in Case 1.

Case 3 *Prediction is required for a new plot with prior observations available and the plot is identified with the original installations and treatments.*

As in Case 1, because this plot comes from one of the original installations and treatment nesting combinations, the *BLUP*'s for random effects at installation and treatment level can be applied . However, since there are some former observations available for this plot, we may estimate (predict) the random effects on a plot level with the best linear unbiased predictor (*BLUP*) (see, e.g. Harville 1976) :

$$\hat{\mathbf{b}}_{1,h} = \mathbf{D}_1 \mathbf{z}_{1,h0}^T \Sigma^{-1} (\mathbf{y}_{h0} - \mathbf{x}_{h0}^T \hat{\boldsymbol{\beta}}) \quad (3.26)$$

$$\text{where } \Sigma = \mathbf{z}_{1,h0}^T \mathbf{D}_1 \mathbf{z}_{1,h0} + \mathbf{z}_{2,h0}^T \mathbf{D}_2 \mathbf{z}_{2,h0} + \mathbf{z}_{3,h0}^T \mathbf{D}_3 \mathbf{z}_{3,h0} + \mathbf{R} \quad (3.27)$$

Continuing with the example, suppose that in addition to the information available for Case 1, we also know that the basal area, dominant height and stand density for the plot are respectively 18 m^2 , 12.5 m and 840 stems/ha at age 12, and 13 m^2 , 8.5 m and 960 stems/ha at age 9. We can predict the basal area at age 15 for this plot. For this example,

$$\mathbf{x}_{h0}^T = \begin{pmatrix} 1 & 1 & 1 & 9 & 1/9 & 0 & 1/9 & 1/9 & \ln(8.5) & \ln(8.5) & \ln(960) \\ 1 & 1 & 1 & 12 & 1/12 & 0 & 1/12 & 1/12 & \ln(12.5) & \ln(12.5) & \ln(840) \end{pmatrix}$$

$$\mathbf{z}_{1,h0}^T = (1 \quad 1), \quad \mathbf{z}_{2,h0}^T = \begin{pmatrix} 1 & 1/9 \\ 1 & 1/12 \end{pmatrix}, \quad \mathbf{z}_{3,h0}^T = \begin{pmatrix} 1 & 1/9 & \ln(8.5) \\ 1 & 1/12 & \ln(12.5) \end{pmatrix}$$

From (25a), the prediction values at age 9 and 12 are:

$$\hat{\mu}_{1,h0} = \begin{pmatrix} 2.430857 \\ 2.830862 \end{pmatrix}, \quad \text{So } \hat{e}_{1,h0} = y_{h0} - \hat{\mu}_{1,h0} = \begin{pmatrix} \ln(13+1) - 2.430857 \\ \ln(18+1) - 2.830862 \end{pmatrix} = \begin{pmatrix} 0.2082001 \\ 0.1135771 \end{pmatrix}$$

$$\mathbf{R} = \hat{\sigma}^2 \text{diag}[\exp(\hat{\theta}\hat{\mu}_{h0})] \begin{pmatrix} 1 & \hat{\rho} \\ \hat{\rho} & 1 \end{pmatrix}, \quad \hat{\sigma} = 0.1583093, \quad \hat{\theta} = -0.4611205, \quad \hat{\rho} = 0.3556445 \text{ (see}$$

$$\text{Table 14), i.e. } \mathbf{R} = \begin{pmatrix} 0.008169666 & 0.002649522 \\ 0.002649522 & 0.006793578 \end{pmatrix}$$

According to (3.25e) :

$$\Sigma = \begin{pmatrix} 0.015084416 & 0.008776294 \\ 0.008776294 & 0.013605554 \end{pmatrix}$$

Therefore [using (3.26)], $\hat{\mathbf{b}}_{1,h} = 0.007268766$.

So the prediction for basal area at age 15 can be obtained at the plot level [using (3.25a)]: $\ln(B\hat{A} + 1) = 3.26776$, thus $B\hat{A} = 25.252 \text{ m}^2/\text{ha}$.

The empirical residuals at age 9 and 12 (see \hat{e}_{1,h_0}) indicate that the basal areas at ages 9 and 12 are underpredicted with the population-level model. The plot level random effects, $\hat{b}_{1,h}$, adjusted the underestimation up a little bit, but the magnitude is small compared to the residuals. This is because the variation at the plot nesting level is relatively small compared with that at the installation and treatment levels (comparing D_1 with D_2 and D_3). The weight of the adjustment given to the plot nesting level is also relatively small [see equation (3.26)]. In the original data, the main variations are expected on installation and treatment within installation levels, so it is no surprise that *BLUP* at a plot nesting level has only a very little effect on the prediction. Now that the individual adjustment of response mainly relies on the *BLUP*'s for the random effects at installation and treatment within installation levels, it is critical that the *BLUP*'s with the original data at installation and treatment levels should reflect a current individual's variability. Otherwise, the prediction based on the *BLUP*'s with the original data may have a large error. For instance, the residuals of prediction at the treatment level are, respectively, 0.01999264 and -0.09490905 at ages 9 and 12 for the above example. These are acceptable with any usual criteria. However, suppose that the dominant heights at ages 9 and 12 are 10.5 m and 14.5 m (instead of 8.5m and 12.5 m) and all the other information is the same as above. The residuals increase to -0.2084335 and -0.2555343, which are obviously overestimated. Therefore, the *BLUP*'s from the original data should be applied with caution to new individual plots. When repeated measurements are available, as will be illustrated in the next case, the random effects at the installation and treatment level, as well as individual plot level should be estimated directly from these repeated measurements on the specific plot.

Case 4 *Prediction is required for a new plot with prior observations available and the plot is not identified with the original installations and treatments.*

As with Case 2, there is no matching installation and treatment nesting in from the original design associated with the new individual (plot). Thus, the *BLUP*'s with the original data are not applicable. However, because repeated measurement information is available, we can estimate (predict) the random effects at three distinct nesting levels:

At the installation level,

$$\hat{\mathbf{b}}_{3,h} = \mathbf{D}_3 \mathbf{z}_{3,h0}^T \Sigma^{-1} (\mathbf{y}_{h0} - \mathbf{x}_{h0}^T \hat{\boldsymbol{\beta}}) . \quad (3.28)$$

At the treatment within installation level,

$$\hat{\mathbf{b}}_{2,h} = \mathbf{D}_2 \mathbf{z}_{2,h0}^T \Sigma^{-1} (\mathbf{y}_{h0} - \mathbf{x}_{h0}^T \hat{\boldsymbol{\beta}}) . \quad (3.29)$$

The formula for plot within treatment level was given as (3.26).

Continuing with the example information as presented in Case 3, instead of using *BLUP*'s from the original data we obtain *BLUP*'s with (3.28) and (3.29) for installation level and treatment within installation level.

$$\hat{\mathbf{b}}_{3,h} = \begin{pmatrix} 0.27126175 \\ -0.64248409 \\ -0.06419382 \end{pmatrix}, \quad \hat{\mathbf{b}}_{2,h} = \begin{pmatrix} 0.03631892 \\ -0.11695520 \end{pmatrix}, \text{ and } \hat{\mathbf{b}}_{1,h} = (0.007268766)$$

The residuals of the prediction at treatment within installation level are 0.12238061 and 0.031419180 respectively for ages 9 and 12. Compared with those using *BLUP*'s from the original data (0.01999264 and -0.09490905, see Case 3), this plot has a larger residual at age 9 but a smaller one at age 12. However, as shown above, if the dominant heights at ages 9 and 12 are 10.5 m and 14.5 m (instead of 8.5 m and 12.5

m), the residuals with *BLUP*'s from the original data are -0.2083837 and -0.2553168.

With individual *BLUP*'s based on repeated measurement data [using (3.28) and (3.29)],

the residuals are -0.01521545 and -0.03316591, which are much smaller.

Now that there are previous observations available, *BLUP*'s for random effects can be obtained, in this case and predictions can be obtained at the following four levels:

1. *At the population level* [using (3.22a)-(3.22d)]:

The predicted value, prediction variance and 95% confidence interval for the prediction is the same as those in Case 1.

2. *At the installation level* [using (3.23a)-(3.23d)]:

$$\ln(\hat{BA} + 1) = 3.09197, \text{ thus } \hat{BA} = 21.020 \text{ m}^2/\text{ha}.$$

with variance in $\ln(\hat{BA} + 1)$ [using (3.23b)] of 0.01181591, which is 87.2% of the variance at population level. The approximate 95% confidence interval for the prediction of $\ln(\hat{BA} + 1)$ at the installation level is [2.878916, 3.305024]. When transformed back to the original scale, it is: [16.795m²/ha., 26.249m²/ha.].

3. *At the treatment level* [using (3.24a)-(3.24d)]:

$$\ln(\hat{BA} + 1) = 3.120492, \text{ thus } \hat{BA} = 21.658 \text{ m}^2/\text{ha}.$$

with variance in $\ln(\hat{BA} + 1)$ [using (3.24b)] of 0.01087634, which is 80.3% of the variance at population level. The approximate 95% confidence interval for the prediction of $\ln(\hat{BA} + 1)$ at the installation level is [2.916084, 3.3249].

Transformed back to the original scale, it is: [17.469m²/ha., 26.796 m²/ha.].

4. *At the plot level* [using (3.25a)-(3.25d)]:

$$\ln(B\hat{A} + 1) = 3.127761, \text{ thus } B\hat{A} = 21.823 \text{ m}^2/\text{ha}.$$

The variance is almost the same as that obtained at the treatment within installation level (3. above).

A graphical comparison for above four situations in this sub-case is presented in Figure 30.

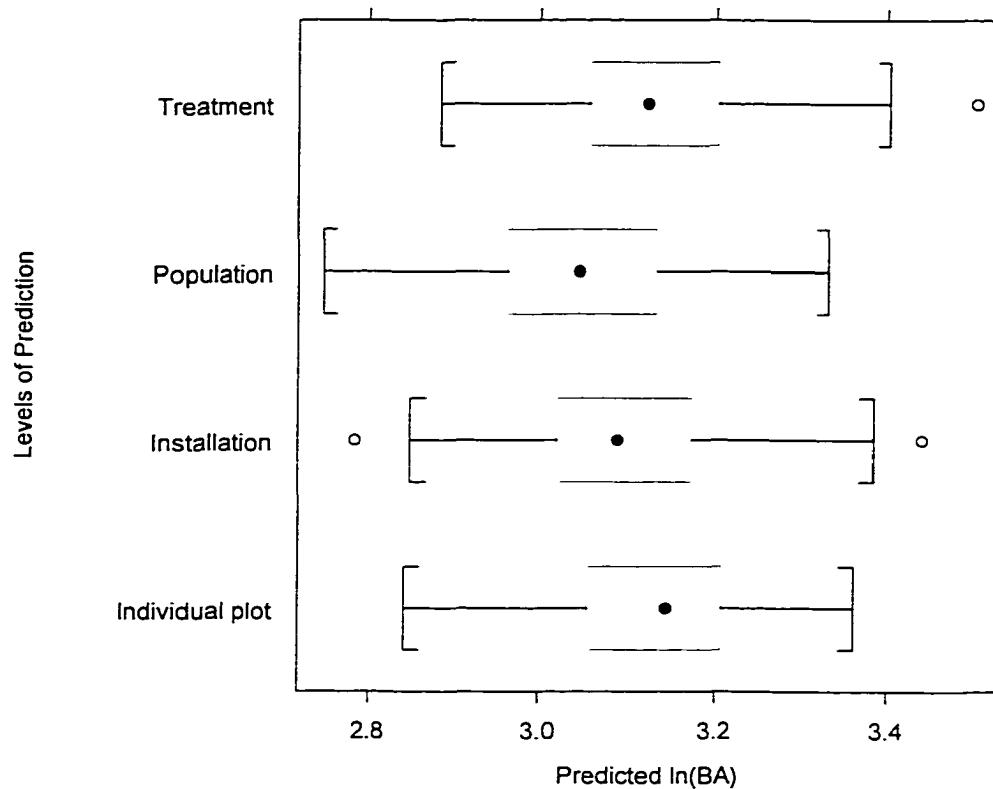


Figure 30. A comparison of the predictions and corresponding confidence intervals for a new observation at four nesting levels with the linear multilevel mixed-effects basal area growth model.

Discussion

Data with one or more classification factors often have some common properties. For example, heteroscedasticity between groups and correlation within groups are often observed on these kinds of data. Since there is usually more than one source of error in the model for such data, a multi-level mixed effects model is intuitively appealing. Based on the comparisons of model performance, a linear mixed effects model with three-levels of nesting (in random effects) was developed and applied to the slash pine basal area growth data from a designed study installed by The University of Georgia *PMRC*. Specifically, random intercepts are allowed at installation, silvicultural treatment within installation and individual plot level. The effects of covariate age^{-1} can be random at both treatment and installation levels and the effects of $\ln(HD)$ are allowed to be random at installation level. Conditioning on the random effects, a power variance function and $AR(1)$ correlation structure are specified for within-plot error. This is a complex model in the sense that both (multi-level) random effects and serial correlations among the measurements are included in the model. Since the random effects are often able to represent the serial correlations (Chi and Reinsel 1989), the current model can be overparameterized, especially with small samples. The methods for determination of the best combination of serial correlation components and random effects are an important topic that deserves further study (see e.g. Diem and Liukkonen 1988, Rochon 1992). A common decision method is to rely upon some statistical criterion, such as the *AIC* and the likelihood ratio test.

Since the error for the model has multi-level sources, it is logical that any prior information at each level will be helpful in model predictions. The current model

contains such advantages in prediction applications for a new observation since: ① The *BLUP*'s for random effects at different levels may be directly applicable in prediction, ② Repeated observations on a new plot can be used to estimate/predict the random effects associated with that plot so that a more precise prediction can be obtained, and ③ Previous observations on a new plot are also informative in predictions for the serial correlation existing among the measurements. Thus, in the current framework, predictions for new observations can be obtained at different levels of precision in different cases. Generally speaking, the more specific the information available for a new case (plot), the more precise the prediction obtained.

The prediction and the standard error of the prediction, thus the confidence intervals of the prediction, are available in the logarithmic transform of basal area per hectare. By reversing the transformation, one may obtain corresponding quantities of basal area per hectare in untransformed units (m^2/ha). However, as discussed in similar work by Gregoire et al (1995) such reversing of the transformation results in a systematic underestimate of mean basal area, even though the resulting median of basal area per hectare is still consistent (Miller 1984). Flewelling and Pienaar (1981) addressed this transformation bias and recommended some general guidance for the correction to this problem in a forest context. An alternative is to forgo the logarithmic transformation and fit a nonlinear mixed effects model to the untransformed response.

For dimension reduction and analysis/application conveniences, 11 silvicultural treatment combinations in the original design are expressed additively using 5 distinct silvicultural activities (chop, burn, bed, fertilize and herbicide). While this is most often the approach taken in forest growth and yield modeling in conjunction with silvicultural

treatments, interpretation of such a model should be done with caution. The effects of one silvicultural activity, such as “chop”, should not be interpreted as the real effects of this activity when applied alone. For example, one may wonder why there are so many significant terms involved with the “chop” activity in the final model (Model 9). Of course, this does not mean that “chop” is the most significant activity influencing slash pine basal area growth. With close look into the original design of this silvicultural study the answer becomes obvious. Because “chop” is the most extensively applied activity in the study (except for the control, all other silvicultural treatments/combinations contain chop), the effects of “chop” actually form a valid contrast between the silvicultural treatments and the control (no treatment). The effects of other silvicultural activities should also be interpreted in a similar way: a contrast of the *combination* of all the treatments with an activity as a component to those without this activity as a component. Because these are not orthogonal contrasts, a detailed interpretation is difficult. This stimulates another approach to silvicultural treatment modeling as presented in the next chapter.

CHAPTER 4

A NONLINEAR MIXED MODEL FOR SLASH PINE DOMINANT HEIGHT GROWTH: A GIBBS SAMPLER APPROACH

Introduction

In previous chapters silvicultural treatments have been included as terms in a forest growth model using a dummy variable approach. For dimension reduction, silvicultural treatments were re-coded using 5 binary variables ($zchop$, $zburn$, $zbed$, zf and zh) which respectively have values of 1 if the silvicultural activities chop, burn, bed, fertilizer, or herbicide were applied, and 0 otherwise. For example, silvicultural treatment $FCBH$, a combination of treatments chop, burn and fertilize, has been coded as $zchop=1$, $zburn=1$, $zbed=0$, $zf=1$ and $zh=0$. The significant dummy variables remained in the final model. This is a very common and convenient approach. However, this approach has its serious limitation. It is not realistic that the effects of these 5 silvicultural activities will all be additive. The cumulative effects attributed to the dummy variables are not necessarily the same as the effects of silvicultural treatments where the corresponding activities are combined. For the above example the summation of the effects of $zchop$, $zburn$ and zf will not strictly account for the effects contributed by silvicultural treatment $FCBH$, because all the interaction terms and the potential error terms for such a treatments combination have not been tracked. This problem doesn't

exist if the original treatment combinations (including control, there are 11 treatment combinations) are directly coded in the model. But as mentioned in previous chapters, this creates a dimensionality problem for model identifiability in the usual approach to modeling data from such studies.

As an alternative approach, in this chapter I revisit the nonlinear mixed effects dominant height growth model using a Bayesian approach. The 11 silvicultural treatments are directly represented in the formulation and a Gibbs sampler is used for model identification.

Bayesian model formulation for a nonlinear mixed effects model

In a Bayesian model framework, in addition to the two-stage formulation in Chapter 2 another stage that specifies the prior distributions of the parameters is also required for the nonlinear mixed effects model formulation. Let y_i be a $n_i \times 1$ vector of repeated measurements of dominant height available on plot i . The three-stage hierarchical model formulation for a nonlinear mixed effects model can be summarized as follows:

Stage 1: *within-plot variation*

$$y_i = f(t_i, \beta_i) + e_i, \quad \text{var}(e_i | \beta_i) = R_i(\beta_i, \theta)$$

(4.1a)

Stage 2: *between-plot variation*

$$\left. \begin{array}{l} \beta_i = X_i \beta + Z_i b_i \\ b_i \sim N(0, D) \end{array} \right\} \quad (4.1b)$$

Stage 3: *hyperprior distribution*

$$(\beta, \theta, D) \sim p(\beta, \theta, D) \quad (4.1c)$$

The function $p(\beta, \theta, D)$ is the probability density function of the parameters and generally reflects weak knowledge about the parameters at this stage of hierarchy. Stage 1 specifies the general mean function of the response and the within-plot error structure due to repeated measurements. Stage 2 of the hierarchy gives an explicit relationship between random parameters β_i , and the random effects b_i , in which the variability in response among the plots is modeled by allowing the parameters in the model to vary systematically or randomly from plot to plot. Similarly as in Chapter 2, in Stage 1, the modified Richards equation is the mean response function for dominant height growth, i.e.

$$E[HD_i(t_{ij})] = f(t_{ij}, \beta_i) = \beta_{i,0} \left(\frac{1 - e^{-\beta_{i,1} t_{ij}}}{1 - e^{-\beta_{i,1} t_0}} \right)^{\beta_{i,2}}, \quad (4.2a)$$

where $HD_i(t_{ij})$ is the dominant height of plot i at occasion j .

For simplicity, within-plot errors are assumed to be in the simplest form given the random effects, i.e.,

$$e_i \sim N(0, \sigma^2). \quad (4.2b)$$

In Stage 2, we assume that the three parameters in the model are common for the same silvicultural treatment, but vary randomly from plot to plot. That is,

$$\beta_i = X_i \beta + Z_i b_i, \quad (4.2c)$$

where

$$\beta_i = \begin{pmatrix} \beta_{i,0} \\ \beta_{i,1} \\ \beta_{i,2} \end{pmatrix}, \quad X_i = \begin{pmatrix} x_{i,1} & x_{i,2} & \dots & x_{i,11} & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & x_{i,1} & x_{i,2} & \dots & x_{i,11} & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & x_{i,1} & x_{i,2} & \dots & x_{i,11} \end{pmatrix}_{3 \times 11}$$

$$\beta^T = (\beta_{1,0} \ \dots \ \beta_{11,0} \ \ \beta_{1,1} \ \dots \ \beta_{11,1} \ \ \beta_{1,2} \ \dots \ \beta_{11,2})_{1 \times 11}$$

$$Z_i = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad b_i = \begin{pmatrix} b_{i,0} \\ b_{i,1} \\ b_{i,2} \end{pmatrix} \text{ and } \begin{pmatrix} b_{i,0} \\ b_{i,1} \\ b_{i,2} \end{pmatrix} \sim MVN(0, D)$$

Parameters $\beta_{0,1}, \dots, \beta_{0,11}$ are the fixed effects contained in parameter β_0 in the model reflecting the effects of the 11 different silvicultural treatments on this parameter. Similarly, $\beta_{1,1}, \dots, \beta_{1,11}$ and $\beta_{2,1}, \dots, \beta_{2,11}$ are fixed effects on parameters β_1 and β_2 , respectively, in the original mean response function (the modified Richards equation) for different treatments. Variables $x_{i,1}, \dots, x_{i,11}$ are dummy variables that indicate which silvicultural treatment has been applied on plot i . Note that $\sum_{j=1}^{11} x_{i,j} = 1$. i.e. only one silvicultural treatment (i.e. combination of the 5 silvicultural activities) applies on a plot. Parameters $b_{0,i}$, $b_{1,i}$ and $b_{2,i}$ are random effects due to plot i . They are assumed to follow a multivariate normal distribution with mean 0 and covariance D .

In Stage 3, the prior distributions of the parameters are assumed as follows:

$$\beta_{j,0} \sim N(\beta_{j,0}^*, \sigma_0^2), \quad j=1, \dots, 11. \quad (4.2d)$$

$$\beta_{j,1} \sim N(\beta_{j,1}^*, \sigma_1^2), \quad j=1, \dots, 11. \quad (4.2e)$$

$$\beta_{j,2} \sim N(\beta_{j,2}^*, \sigma_2^2), \quad j=1, \dots, 11. \quad (4.2f)$$

$$D^{-1} \sim Wishart(m_0, \Psi(\theta)), \quad (4.2g)$$

$$\sigma_j^{-2} \sim Gamma(\nu_j, \tau_j), \quad j=0,1,2. \quad , \quad (4.2h)$$

where the parameters $\beta_{j,..}^*$, ν_j , τ_j , m_0 , $\Psi(\theta)$ that characterize the hyperprior distribution are known. The matrix D^{-1} follows a Wishart distribution with degrees of freedom m_0

($m_0 \geq 3$) and 3×3 precision matrix Ψ with parameters θ . Without former information, vague priors are often used. This issue will be discussed latter in a subsequent section. Under the model assumptions (4.2a-4.2g), the joint posterior distribution of all the parameters is (Rao 1987, Davidian and Giltinan 1995, page 220):

$$p_{post}(\beta, \theta, \sigma^2 | y) = \frac{p_{prior}(\beta, \theta, \sigma^2) L(\beta, \theta, \sigma^2 | y)}{\int \int \int p_{prior}(\beta, \theta, \sigma^2) L(\beta, \theta, \sigma^2 | y) d\beta d\theta d\sigma^2} \quad (4.3)$$

where $L(\beta, \theta, \sigma^2 | y)$ is the likelihood of parameters given all the observed data, i.e.

$$L(\beta, \theta, \sigma^2 | y) = \prod_{i=1}^n \frac{1}{\sqrt{(2\pi\sigma^2)^{n_i} |D|}} \int \frac{\exp\left[\frac{-1}{2\sigma^2} \left(\|y_i - f(\beta, b_i, t_i)\|^2 + b_i^T D^{-1} b_i \right)\right]}{(2\pi)^{q/2}} db_i , \quad (4.4)$$

and $n = 191$ (the number of plots), $n_i = 5$ (the number of repeated measurements), $q=3$ (the dimension of random effects). All other terms are defined above.

Since the random effects are nonlinear in the mean response function, explicit expressions for the likelihood function and the marginal posterior distribution of β_i and β are not possible. As shown in (4.3) and (4.4), this requires a multi-level integration that is intractable because some complicated nonlinear parameters get involved. However, if conditional distributions of the parameters can be obtained explicitly, this difficulty may be overcome by the Gibbs sampler.

The Gibbs sampler is a technique for generating random variables (vectors) from a (marginal) distribution indirectly, without having to calculate the density (Casella and George 1992). This is obviously a big advantage, especially when the number of variables (parameters) is large, because tedious or impossible integration is avoided by Gibbs sampling, which is exactly the happy event needed in our case.

Gibbs sampling

Gibbs sampling (Geman and Geman 1984) is a Markov Chain Monte Carlo (*MCMC*) method (Tanner 1996, Ch 6.) that is extensively applied for estimating complex Bayesian statistical models (Gelfand and Smith 1990). Hastings (1970) first proposed the method as a generalization of the Metropolis algorithm (Metropolis et al 1953). In a repeated measurement context, Zeger and Karim (1991) applied Gibbs sampling in the generalized linear model with random effects. Wakefield et al (1994) develop linear and nonlinear mixed models in a Bayesian framework using Gibbs sampling. Carter and Kohn (1994) applied Gibbs sampling to state space models and comparisons are given between Gibbs sampler and Kalman filter. Gilks et al. (1993) proposed linear multilevel-random-effects models for longitudinal data modeling with Gibbs sampling as a parameter estimation method. Green and Valentine (1998) present a linear model with heterogeneous variance for individual tree foliar dry matter in a Bayesian framework with Gibbs sampling.

To review Gibbs sampling, suppose conditional distributions of the parameters are as follows:

$$\beta | b, \theta, \sigma^2 \sim D_\beta \quad (4.5a)$$

$$b | \beta, \theta, \sigma^2 \sim D_b \quad (4.5b)$$

$$\sigma^2 | \beta, b, \theta \sim D_\sigma \quad (4.5c)$$

$$\theta | \beta, b, \sigma^2 \sim D_\theta \quad (4.5d)$$

The Gibbs sampler involves drawing random samples from all full conditional densities. The Gibbs sampler involves drawing random samples from all full conditional densities from (4.5a) to (4.5d). Given initial value(s) of each parameter(s), say $\sigma^{(0)}, \theta^{(0)}$, $\beta^{(0)}, b^{(0)}$, generate $\beta^{(1)}$ from $\beta^{(1)} | b^{(0)}, \sigma^{(0)}, \theta^{(0)}$, then generate $b^{(1)}$ from $b^{(1)} | \beta^{(1)}, \sigma^{(0)}, \theta^{(0)}$,

then generate $\sigma^{(1)}$ from $\sigma^{(1)} | \beta^{(1)}, b^{(1)}, \theta^{(0)}$, and then generate $\theta^{(1)}$ from $\theta^{(1)} | \beta^{(1)}, b^{(1)}, \sigma^{(1)}$. This completes the first cycle of the sampler. The values from the first cycle are used as the initial values for the second cycle. This procedure keeps going for a large number of times until the chain converges. It can be shown that the generated values for $(\beta, b, \sigma^2, \theta)$ are, asymptotically, realizations from the joint posterior distribution of $p(\beta, b, \sigma^2, \theta)$ (Geman and Geman 1984). Thus, traditional sample moments, such as the mean and variance of the respective generated values, can be applied to obtain consistent estimators.

Detailed conditional distributions for the general hierarchical nonlinear mixed effects are discussed by Davidian and Giltinan (1995, Chapter 8). Spiegelhalter *et al* (1996) note that appropriate parameterization is very important in *MCMC* work in order to improve convergence and stability of the samples. This has also been the case in my analysis. The process of sampling the original parameters in the Model (4.2a-4.2g) broke down after some 800 runs due to unreasonable numbers being generated. From Chapter 2, we know that the three parameters in the modified Richard equations are quite different in magnitude for dominant height growth (β_0 is around 20, β_1 is around 0.07 and β_2 is around 1.7). Thus we make the following reparameterisation:

$$\left. \begin{array}{l} \beta_{i,0} = e^{\phi_{i,0}} \\ \beta_{i,1} = e^{\phi_{i,1}} - 1 \\ \beta_{i,2} = e^{\phi_{i,2}} \end{array} \right\} \quad (4.6)$$

and then

$$\left. \begin{array}{l} \phi_{i,0} = a_{j,0} + b_{i,0} \\ \phi_{i,1} = a_{j,1} + b_{i,1} \\ \phi_{i,2} = a_{j,2} + b_{i,2} \end{array} \right\} \quad (4.7)$$

where, i indexes the plot and j indexes treatments. The b_i denotes random effects of plot i ($i = 1, 2, \dots, 191$) and a_j denotes the fixed effects due to silvicultural treatment combination j ($j = 1, 2, \dots, 11$).

The reparameterisation indicated by (4.6) implies that

$$\left. \begin{array}{l} \phi_{i,0} = \ln \beta_{i,0} \\ \phi_{i,1} = \ln(\beta_{i,1} + 1) \\ \phi_{i,2} = \ln \beta_{i,2} \end{array} \right\} \quad (4.8)$$

Thus, the magnitudes of the components in the chain are closer to each other after reparameterisation. This can improve the stability of the samples (Spiegelhalter et al 1996).

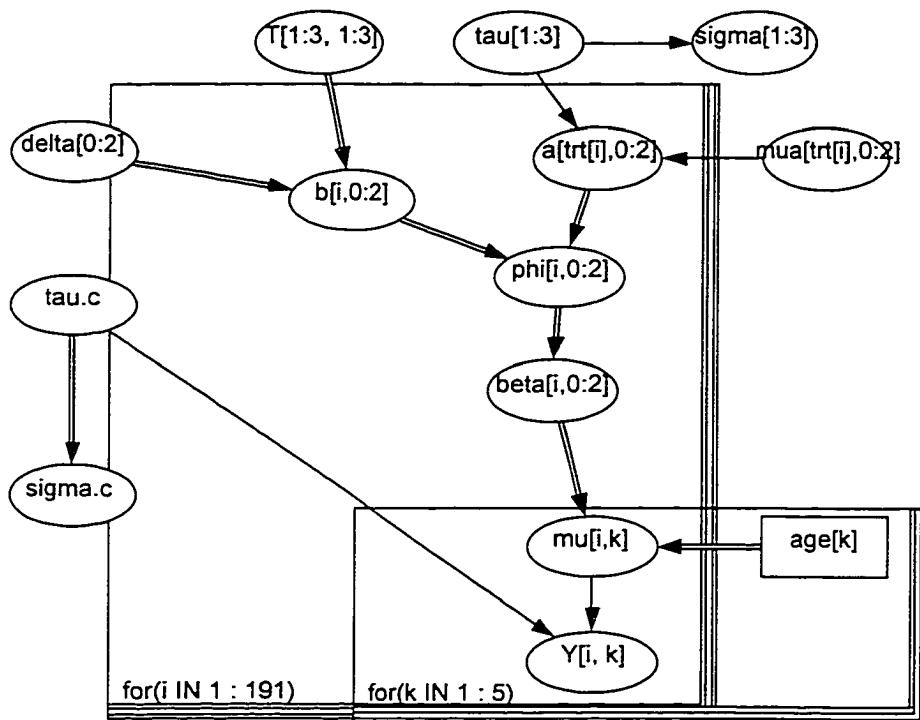


Figure 31: A graphical representing of the nonlinear mixed effects dominant height growth model with a Bayesian reformulation.

A graphical version of the above nonlinear mixed effects model for dominant height growth is shown in Figure 31. Rectangles in the graph denote constant nodes (the age of the stand) and ellipses may denote stochastic nodes or a logical function. The stochastic nodes are variables that have been given a distribution and a logical function is a function of nodes (can be stochastic or constant). The child-parent nodes may have two types of links: a solid arrow indicates a stochastic dependence and a hollow arrow indicates a logical function. So, the parallel model to the graphical model in Figure 1 is as follows:

$$y_{ik} \sim N(\mu_{ik}, \tau_c) \quad (4.9a)$$

$$\mu_{ik} = E[y_{ik}] = f(t_{ik}, \beta_i) = \beta_{i,0} \left(\frac{1 - e^{-\beta_{i,1} t_{ik}}}{1 - e^{-\beta_{i,1} t_0}} \right)^{\beta_{i,2}} \quad (4.9b)$$

$$\left. \begin{array}{l} \beta_{i,0} = e^{\phi_{i,0}} \\ \beta_{i,1} = e^{\phi_{i,1}} - 1 \\ \beta_{i,2} = e^{\phi_{i,2}} \end{array} \right\} \quad (4.9c)$$

$$\left. \begin{array}{l} \phi_{i,0} = \alpha_{j,0} + b_{i,0} \\ \phi_{i,1} = \alpha_{j,1} + b_{i,1} \\ \phi_{i,2} = \alpha_{j,2} + b_{i,2} \end{array} \right\} \quad (4.9d)$$

$$\left. \begin{array}{l} \alpha_{j,0} \sim N(\mu_{a_0}[j], \tau_0) \\ \alpha_{j,1} \sim N(\mu_{a_1}[j], \tau_1) \\ \alpha_{j,2} \sim N(\mu_{a_2}[j], \tau_2) \end{array} \right\} \quad j = 1, \dots, 11 \quad (4.9f)$$

and $\begin{pmatrix} b_{i,0} \\ b_{i,1} \\ b_{i,2} \end{pmatrix} \sim MVN(\Delta_{3 \times 1}, T_{3 \times 3})$ (4.9e)

where τ_c , τ_0 , τ_1 , τ_2 are the precision measures of y , a_0 , a_1 and a_2 , respectively.

Here the fixed effects by different silvicultural treatments are assumed to have different means but the same variance (1/precise). Matrix T is a 3×3 precise matrix for the random effects (i.e. $T = D^{-1}$).

For implementation, the Gibbs sampling for the above nonlinear mixed effects model for dominant height growth mainly relies on the software called *BUGS*¹ (Bayesian inference using Gibbs sampling, Spiegelhalter 1995). Based on conditional independence, in *BUGS* the full joint distribution, $p(V)$, of all the quantities V is the product of all conditional distributions, $p(v|parents[v])$, of the node given their parents, $p(V) = \prod_{v \in V} p(v | parents[v])$. Thus one only needs to provide the parent-child distribution in order to fully specify the model via *BUGS*.

Equations (4.9a) – (4.9e) show the parent-child relationships of the of interest. To start the Gibbs sampling, one only need specify the prior distributions for the parameters of the founder nodes (those with no parents) in the graph (i.e. the prior distributions of τ_c , τ_0 , τ_1 , τ_2 , $T_{3 \times 3}$, μ_{a_0} , μ_{a_1} , μ_{a_2} and Δ) and then give the initial values of those nodes.

It is usually helpful to distinguish parameters of primary interest from those used to specify secondary structure of the model when one tries to figure out the prior for some parameters. Parameters of primary interest will generally be location parameters, such as regression parameters. In many cases a normal prior with an extremely large variance, thus a small precision, is adequate for those parameters (Spiegelhalter et al

¹ The software and related documents are available free of charge on internet at: <http://www.mrc-bsu.cam.ac.uk/bugs>

1996). Thus we choose $\mu_{a_0}, \mu_{a_1}, \mu_{a_2} \sim N(0, 0.0001)$. This indicates that priors for $\mu_{a_0}, \mu_{a_1}, \mu_{a_2}$ all have variance 10000. Alternatively, one may choose $\mu_{a_0}, \mu_{a_1}, \mu_{a_2} \sim Uniform(-200, 200)$ as the non-informative priors for those parameters.

For the secondary aspects of the model, such as the precision parameters, instead of $p(\tau) \sim 1/\tau$ for a small τ so that the scale parameter has a uniform distribution on $(-\infty, \infty)$, Spiegelhalter et al (1996) have shown that $\tau \sim Gamma(\varepsilon, \varepsilon)$ with mean 1 and variance $1/\varepsilon$ is preferable. Thus, we choose $\tau_c, \tau_0, \tau_1, \tau_2, \sim Gamma(0.001, 0.001)$ as the non-informative prior for scale parameters.

The random plot effects, b_i 's, are assumed to arise from a multivariate normal distribution having an unknown mean Δ with precision matrix $T = D^{-1}$ assumed to follow a Wishart distribution. To represent vague prior knowledge, we choose the degrees of freedom for the distribution to be as small as possible (i.e. 3, the rank of D). The scale matrix R was specified as:

$$R = \begin{pmatrix} 0.02 & -0.0007 & 0.0003 \\ -0.0007 & 0.001 & 0.001 \\ 0.0003 & 0.001 & 0.01 \end{pmatrix},$$

which represents the prior guess at the order of magnitude of D .

Convergence of Gibbs sampler

The Gibbs sampler generates a Markov chain of random variables (the stochastic nodes in the graph) that converge to the marginal distributions. So once the chain is running, an immediate question is how to determine the convergence of the chain, i.e.

how long or how many chains are needed to justify that the samples do represent the marginal posterior distributions. There is no general answer to this question. A common approach is to run multiple parallel chains from over-dispersed starting values. An empirical judgment is then to check scalar quantities such as first and second moments of the samples from different chains. The quantities of the same parameter from the chains with different starting values should be close to each other when a condition of convergence is reached. Gelman and Rubin (1992) suggest checking both within- and between-output series variations for such multiple chains. The Gelman-Rubin convergence statistic has been modified recently by Brook and Gelman(1998). The width of the central 80% interval of pooled runs and that of within the individual runs are compared. Their ratio should converge to 1 and the interval width (both pooled and within) should be stable in a condition of convergence.

With two sets of distinct starting values, most parameters converge after 6,000 iterations. For example, the three parameters of fixed effects for silvicultural treatment *UCHP* converge after 6,100 iterations, and $\beta_{1,0}$, $\beta_{1,1}$ and $\beta_{1,2}$ (i.e. the parameters of dominant height growth in plot 1) converged after about 5,200 iterations. (see Figure 32).

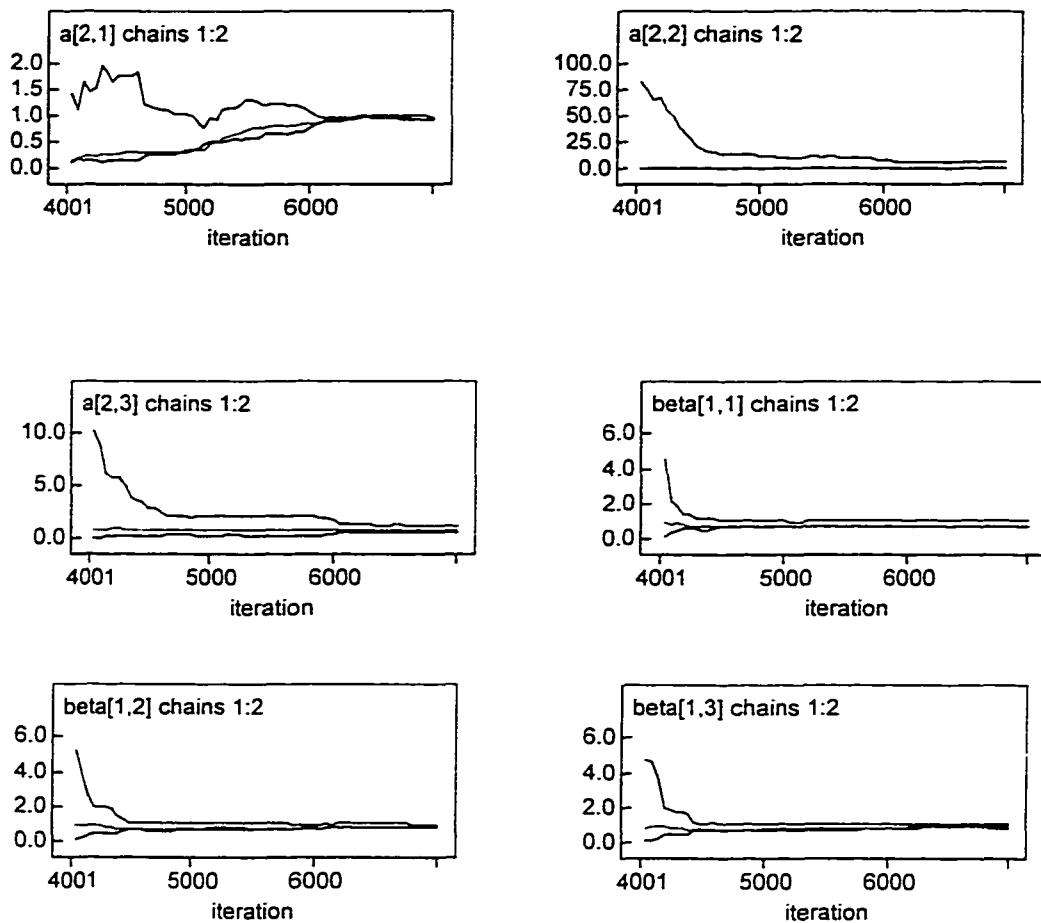


Figure 32: Convergence diagnostics of the Gibbs sampler by modified Gelman-Rubin statistics for the fixed effects of *UCHP* and the parameters in dominant height growth equation on Plot 1. Values $a[2,1]$, $a[2,2]$ and $a[2,3]$ denote the fixed effects of silvicultural treatment *UCHP*. Values $\beta[1,1]$, $\beta[1,2]$ and $\beta[1,3]$ denote the three parameters in the dominant height growth equation for Plot 1. The lower line is the average width of the 80% intervals within the individual runs (the intervals are normalized to have maximum value 1, same as the follows), the middle line is the width of the central 80% interval of the pooled runs and the upper line is the ratio (pooled/within).

Posterior distributions

A 1,000 update burn-in followed by a further 10,000 updates, gave the parameter estimates (empirical marginal posterior quantities) as shown in Table 16. Quantities such as mean, standard error of samples (std.), median, 2.5% and 97.5% for each parameter are summarized in Table 16. Due to space limitations, the mixed effects ($\beta_{i,0}$, $\beta_{i,1}$ and $\beta_{i,2}$, i

$=1, \dots, 191$, indexing individual plot) and fixed effects ($a_{j,0}$, $a_{j,1}$ and $a_{j,2}, j = 1, \dots, 11$, indexing 11 original silvicultural treatment combinations) are only partially given in Table 16. The Metropolis MCMC algorithm (see, e.g. Smith and Robert 1993) was applied in the sampling, in which the first 4,000 iterations as the adapting phase have not been included in the calculations of the summary statistics. The parameters for fixed and random effects are in transformed scale (logarithm). The long tails in the posterior distributions of the variance parameters (σ_0^2 , σ_1^2 and σ_2^2 , see Figure 33) may indicate that the fixed or mixed effects parameters (the α 's and thus the β_i 's) estimated by the mean of the samples and those by median of the samples do not necessarily coincide with each other (see Table 16).

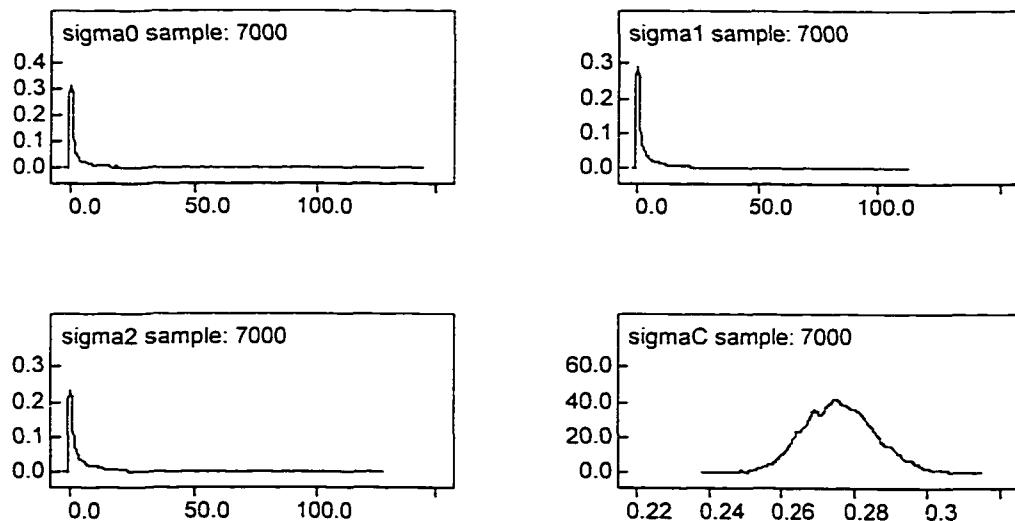


Figure 33: Posterior probability distributions of the variance (precision) parameters: Sigma0, sigma1, sigma2 and sigmaC are the standard errors respectively for fixed effects of silvicultural treatment ($a_{j,0}$, $a_{j,1}$ and $a_{j,2}, j=1, \dots, 11$) and dominant height model.

Table 16 The estimated parameters and quantiles for the slash pine nonlinear mixed effects-dominant height growth model fitted with Gibbs sampler approach.

Parameters	mean	std.	MC error*	2.5%	median	97.5%
$\beta_{1,0}$	19.85	1.251	0.1049	17.61	19.76	22.38
$\beta_{1,1}$	0.03944	0.01787	0.001699	0.003137	0.0411	0.07091
$\beta_{1,2}$	1.932	0.2183	0.02011	1.52	1.937	2.387
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$\beta_{191,0}$	17.56	0.9098	0.07687	16.01	17.45	19.68
$\beta_{191,1}$	0.08621	0.0183	0.001722	0.04294	0.0876	0.1202
$\beta_{191,2}$	2.092	0.2402	0.02185	1.608	2.09	2.579
$a_{1,0}$	3.035	0.03175	0.003052	2.968	3.035	3.093
$a_{1,1}$	0.2339	0.01404	0.001402	0.2147	0.2286	0.2613
$a_{1,2}$	0.6064	0.0593	0.005775	0.4888	0.6014	0.728
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$a_{11,0}$	3.191	0.03479	0.003435	3.13	3.188	3.259
$a_{11,1}$	0.2433	0.008843	8.788E-4	0.2249	0.2444	0.2578
$a_{11,2}$	0.3092	0.04688	0.004553	0.2276	0.3024	0.4072
σ_0^2	5.024	10.76	0.729	0.03625	0.7539	36.33
σ_1^2	5.424	10.62	0.7155	0.0396	1.198	36.75
σ_2^2	6.234	11.61	0.7734	0.03395	1.543	42.45
σ_c^2	0.2744	0.008387	5.66E-4	0.2583	0.2742	0.2914
Δ_0	-0.09192	0.01514	0.001189	-0.1214	-0.09185	-0.06223
Δ_1	-0.1816	0.003944	3.214E-4	-0.1893	-0.1816	-0.174
Δ_2	-0.00858	0.02771	0.002645	-0.0595	-0.007034	0.04431
$T_{1,1}$	70.73	10.3	0.6927	53.96	69.46	94.75
$T_{1,2}$	64.44	24.21	0.7901	17.18	64.04	112.2
$T_{1,3}$	-11.74	17.49	1.57	-54.65	-8.803	15.14
$T_{2,1}$	64.44	24.21	0.7901	17.18	64.04	112.2
$T_{2,2}$	1331.0	172.9	10.53	975.8	1334.0	1663.0
$T_{2,3}$	-146.7	36.41	1.742	-217.5	-146.8	-72.62
$T_{3,1}$	-11.74	17.49	1.57	-54.65	-8.803	15.14
$T_{3,2}$	-146.7	36.41	1.742	-217.5	-146.8	-72.62
$T_{3,3}$	117.5	37.77	3.5	67.57	108.2	199.2

* MC error refers to the Monte Carlo standard error of the mean, = $\sigma/N^{1/2}$. N is the number of samples in calculation and σ is estimated by the bath means method outlined by Roberts (1996).

Enriched graphical outputs are available from the software used. As examples, Figure 34 shows the posterior distributions for the three mixed effects parameters on plots 4 and 5. The 95% empirical confidence interval for fixed silvicultural treatment effects for treatment *UCHB* ($\alpha[4,1]$ to $\alpha[4,3]$ in Figure 35) and *FCBH* ($\alpha[5,1]$ to $\alpha[5,3]$ in Figure 35) are displayed in Figure 35. Figure 36 traces the auto-correlation in samples for parameters σ_v^2 , σ_i^2 , σ_e^2 and σ_c^2 . Usually, higher correlation in the samples requires more iterations to reach convergence of the chain.

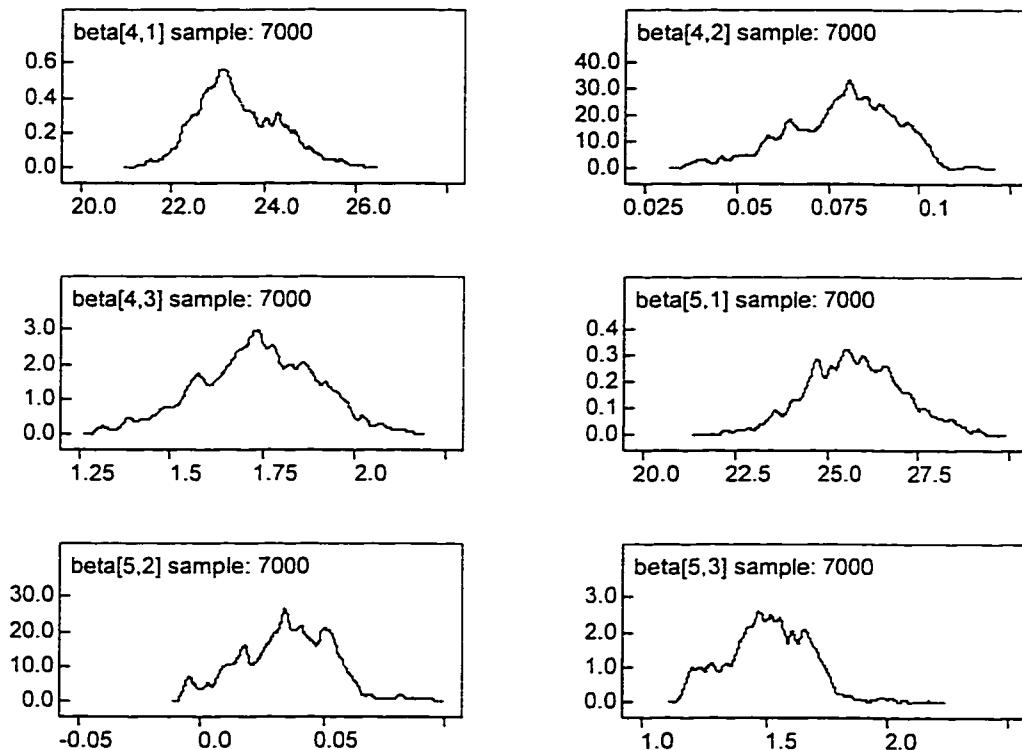


Figure 34: Posterior probability distributions of for the three mixed effects parameters (beta's in the graph) on treatment Plots 4 and 5.

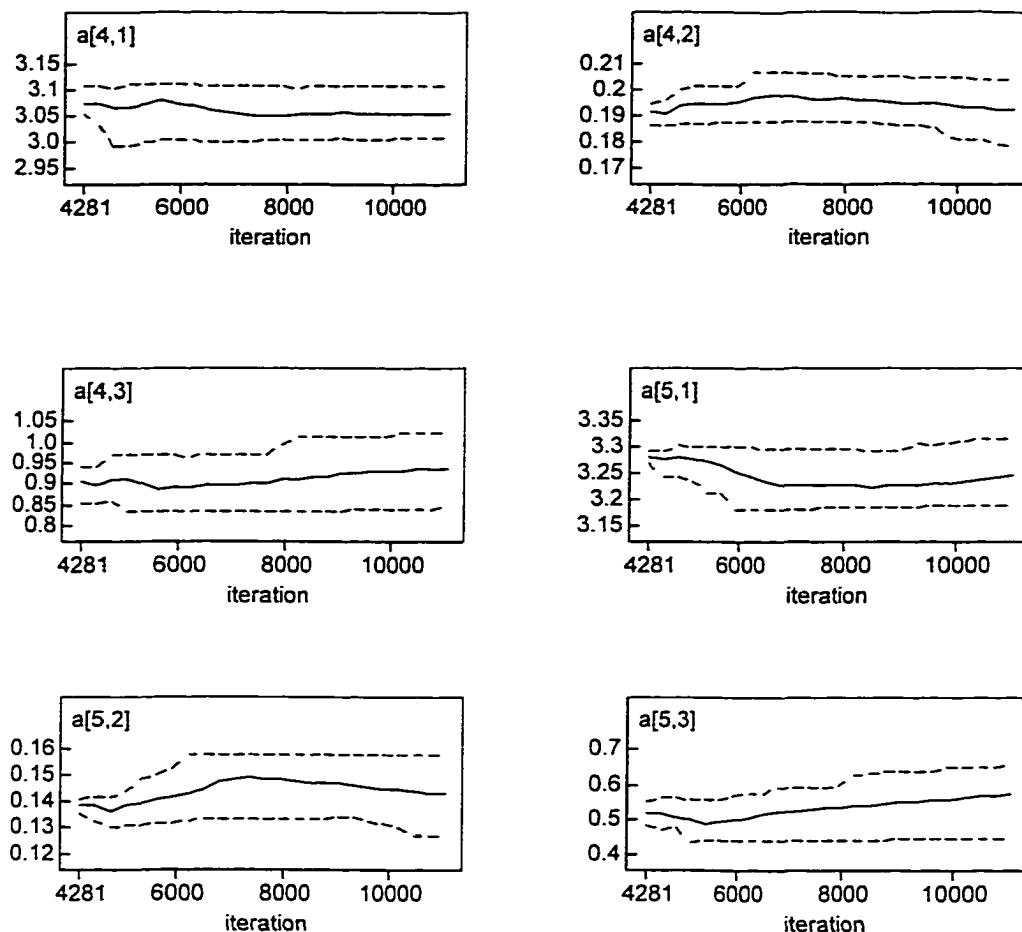


Figure 35: The 95% empirical confidence intervals for fixed silvicultural treatment effects by *UCHB* ($a[4,1], a[4,2]$ and $a[4,3]$) and *FCBH* ($a[5,1], a[5,2]$, $a[5,3]$).

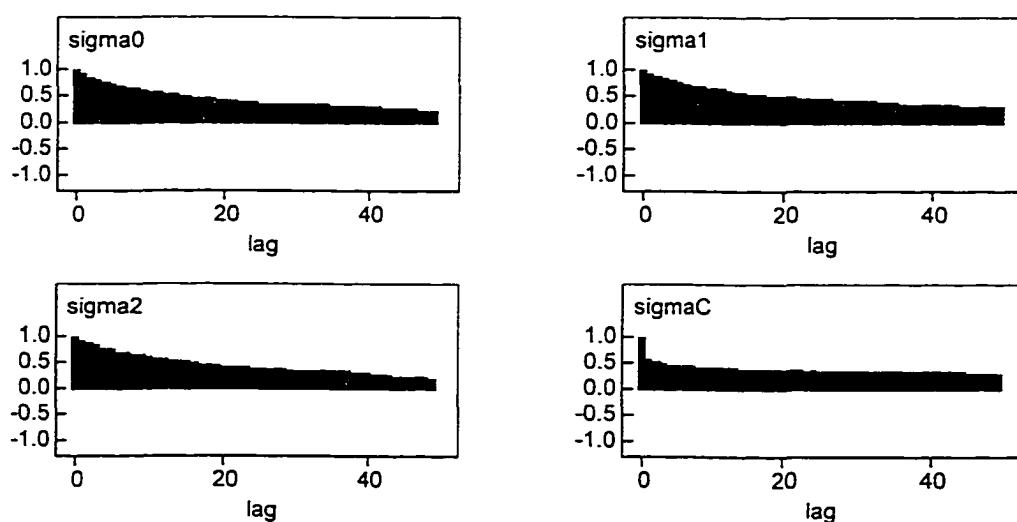


Figure 36: Autocorrelation plots for the samples of the variance (precision) parameters.

Model diagnostics and comparison

As usual, we mainly rely on the residuals to check the general performance for the model with the estimated parameters. Let y_{ij} denote the dominant height observation on plot i at occasion j , \hat{y}_{ij} is the corresponding predicted value by using the estimated parameters. Then the residual is $e_{i,j} = y_{ij} - \hat{y}_{ij}$. As shown above (Table 16), the estimated parameters by mean and median of the samples are not coincident. So residuals are calculated separately based on the parameters from these two distinct quantities. The residual plots (residual vs fitted values) are displayed in Figure 37. These plots show that the residuals based on either quantity (mean or median of the samples) are generally homogeneous. The magnitude of the residuals is small, so the plot of the fitted values against the original observations forms an almost 45° straight line (see Figure 38), which implies a generally nice fitting. To compare the fitted model using Gibbs sampling with the nonlinear mixed effects model approach (Model 10 in Chapter 2), summaries for the residuals(e), absolute value of the residuals (ASE) and squares of the residuals (SQE) were calculated (Table 17). For Table 17,

$$ASE_{ij} = |e_{ij}| = |y_{ij} - \hat{y}_{ij}| \quad (4.10)$$

$$SQE_{ij} = e_{ij}^2 = (y_{ij} - \hat{y}_{ij})^2 \quad (4.11)$$

These measures are often used as criteria for assessing forest growth and yield model performance. For instance, the mean of residual is a measure of bias of the model, the mean of SQE is a estimated stand error (SEE) type measure. The mean of ASE , known as mean absolute bias (MAB), is also favored in model comparison (see, e.g. Tasissa and Burkhart 1998, Zhang and Borders 1999).

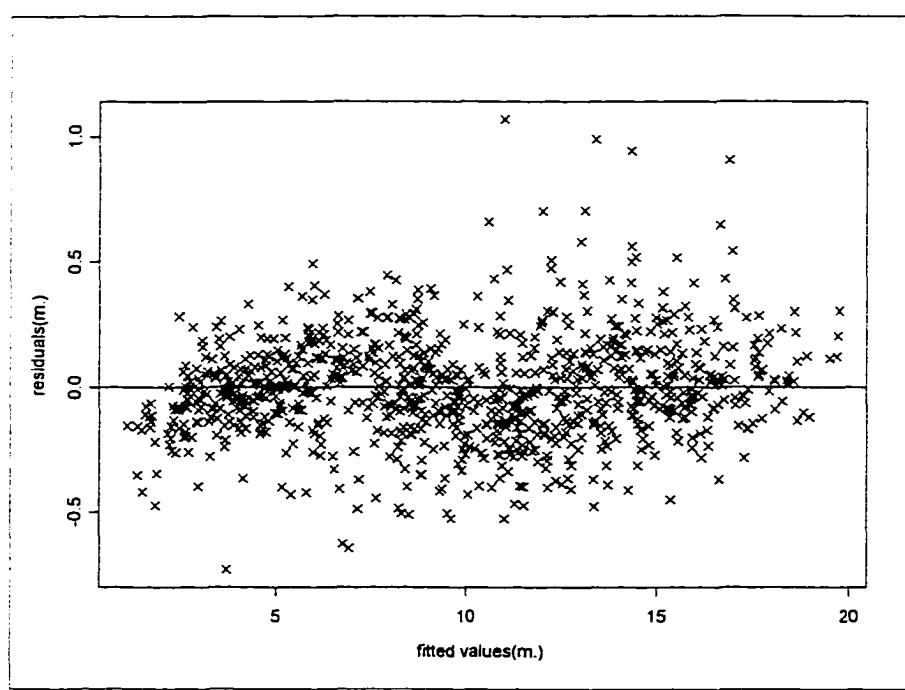
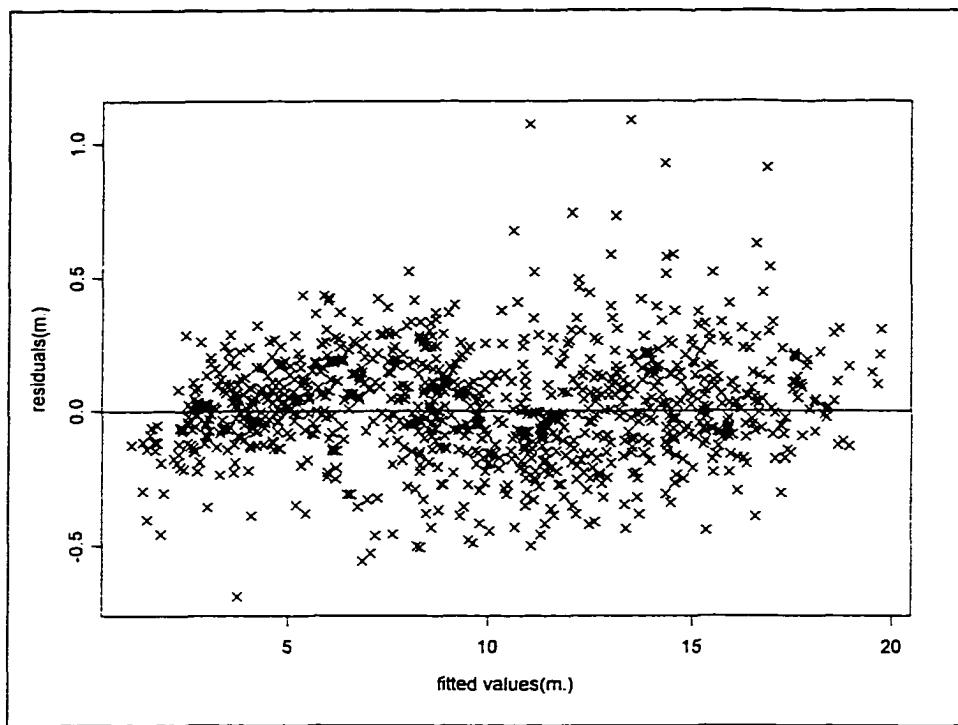


Figure 37: Residual plots for the modified Richards dominant height growth equation with mixed-effects fitted with the Gibbs sampler approach with the mean of samples as estimation (upper) and the median of the samples as estimation (lower).

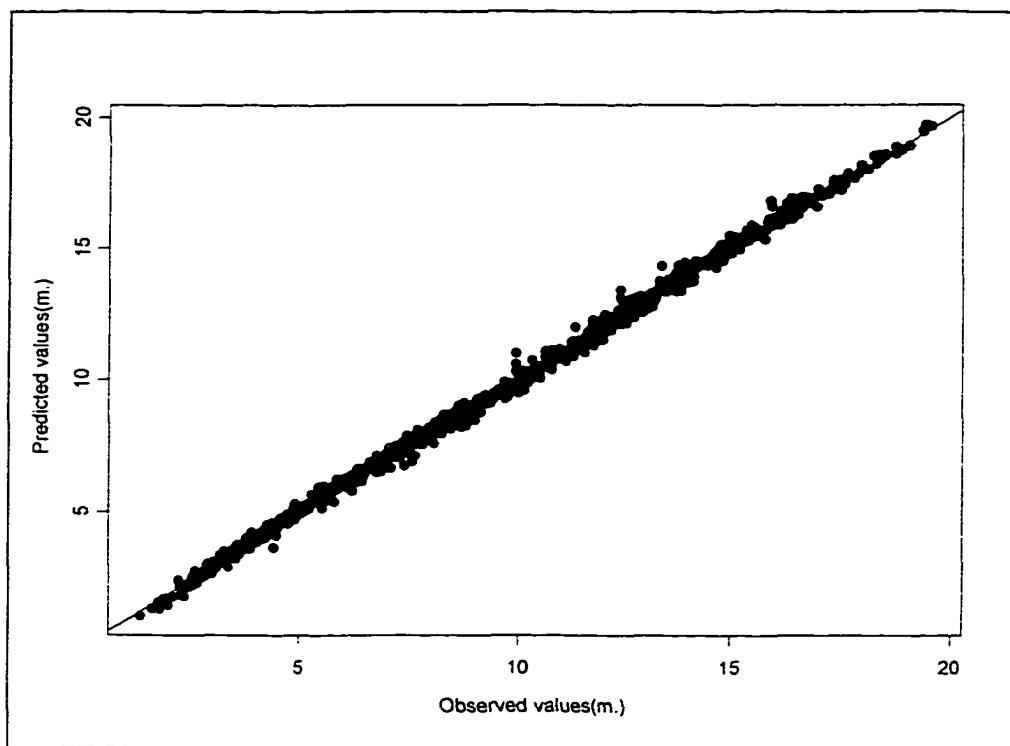


Figure 38: The observed values versus the fitted using the estimated parameters from the median of the Gibbs samples.

Table 17 Comparisons on the basic statistics of residuals, absolute values of the residuals and squares of residual respectively from the Gibbs sampler with mean (e_1) and median (e_2) of the samples and the conventional nonlinear mixed effects approach (e_3).

Residuals	N	Mean	Std Dev	Minimum	Maximum
e_1	955	0.0168931	0.2042622	-0.6868949	1.0886485
e_2	955	-0.0029520	0.2060407	-0.7295301	1.0691387
e_3	955	-0.0057521	0.2156698	-0.7393094	1.1450379
$ e_1 $	955	0.1541542	0.1349826	0.000017115	1.0886485
$ e_2 $	955	0.1545346	0.1362186	0.000390384	1.0691387
$ e_3 $	955	0.1605506	0.1440246	0.000128472	1.1450379
$(e_1)^2$	955	0.0419647	0.0886816	2.929117E-10	1.1851556
$(e_2)^2$	955	0.0424170	0.0874771	1.5239951E-7	1.1430575
$(e_3)^2$	955	0.0464979	0.0995786	1.650513E-8	1.3111117

In Table 17, e_1 and e_2 are the residuals from the Gibbs sampler approach with parameters estimated respectively by the mean and the median of the samples. The e_3 's are the residuals by the nonlinear mixed effects approach. Values $|e_1|$, $|e_2|$, $|e_3|$ are the corresponding absolute residuals and $(e_1)^2$, $(e_2)^2$, $(e_3)^2$ are the corresponding residuals squared.

Generally the residuals with the nonlinear mixed effects models are very small (Table 17). For example, the standard errors of the residuals (a *MSE* type measure) are all around 0.2 m and the largest error among all 955 observations is only in 1.1 m . The summarized statistics of the residuals are close to each other for the three methods. The Gibbs sampler results show an even better fitting than that achieved with the conventional nonlinear mixed effects approach. For example, the *MAB* and standard error of the residuals from the Gibbs sampler are smaller than those from the conventional approach. The mean of the residuals based on the median of the samples (which is preferable) is also smaller than that with the conventional approach. The numbers of parameters in the respective models are close to each other with these two approaches. The model based on the Gibbs sampler has 11 fixed effects parameters with a simple within-plot variance-covariance structure and the model using a conventional mixed-effects approach has 10 fixed effects parameters and 3 other within-plot covariance parameters. The random effects are the same in both approaches. This may indicate that the assumption of additive effects of silvicultural treatment combination in conventional approach (see Chapter 2) may not precisely describe the actual treatment effects on dominant height growth. The mean effects of the three parameters in the dominant height

growth model based on the Gibbs sampler approach (the median of the samples) are displayed in Table 18.

Table 18 The estimated parameters with the Gibbs sampler for dominant height growth with modified Richard's equation by silvicultural treatments groups

Silvicultural Treatments	Estimated parameters		
	β_0	β_1	β_2
CNTL	18.8637	0.060755	1.82181
UCHP	18.0118	0.091815	2.11824
FCHP	21.3872	0.057409	1.68733
UCHB	18.9122	0.083619	2.09256
FCHB	22.8265	0.041451	1.53206
UCBB	19.2868	0.089085	1.92121
FCBB	21.5353	0.066518	1.59676
UBHB	20.2756	0.094824	1.64039
FCBH	21.6494	0.078169	1.60525
UCBH	20.0353	0.093971	1.75159
FBHB	22.2450	0.064349	1.35778

Within the Gibbs sampler approach, the model with the median of the samples as the estimated parameter has a smaller mean bias but a little larger standard error of the residuals than a model with the mean of the samples as the estimated parameter. However, the differences are very small in magnitude.

Discussion

One of the great advantages of the Bayesian approach is flexibility given the modeler. Here we have presented a new version of a nonlinear mixed effects model for

dominant height growth based on the modified Richards equation (see Chapter 2). In this approach, the simplest within-plot error structure and multivariate structure between plot random effects are specified respectively in the first- and second model stages. In a hierarchical model structure, different mean effects (on parameters) are allowed for distinct silvicultural treatments (or combinations). This will give a more precise evaluation of the effects of silvicultural treatments on dominant height growth, which is one of the main goals of the original design in the slash pine plantation growth and yield research.

The Gibbs sampler has been extensively applied in complex Bayesian model computing due to its relatively easy implementation and the enriched outputs that can be used in further reference or prediction. For example, the graphical output from Gibbs sampling provides both informative diagnostic aids and easily understood inferential summaries. The key points of Gibbs sampling are that the full conditional distributions for unknown quantities (parameters or missing values) are available and conditional independence is pre-assumed. The joint distribution over all unobserved (parameters and missing values) and observed quantities (the data) is evaluated indirectly by sampling from the conditional distributions. Thus, prior conditional distributions and initial values for unknown quantities must be given in advance so that sampling can start. Without former information, vague priors are often used. The parameters estimated from some fundamental approaches, such as the OLS, may serve as initial values. See Gilks et al(1996) for a detailed discussion of these topics.

In addition to simplicity in implementation, the Gibbs sampler approach has at least two other advantages. First, it enables any posterior joint or marginal distribution of

interest to be constructed, in principle, to any degree of accuracy. The Gibbs sampler doesn't suffer from approximations or deficiencies inherent in other approaches, such as *EM* algorithm. It is always preferable to have available full posterior distributions instead of normal approximations to them, especially for highly skewed posteriors where maximum posterior estimates are misleading (Gilk et 1993). This property is even valuable for the fact that once the marginal posterior for parameters has been obtained, the marginal posterior distribution of any function of those parameters is also available through a direct transformation of the values of the posterior samples (Gelfand and Smith 1990, Green et al 1994). The second additional advantage is flexible in the sense that particular features of a given set of data can be accommodated with only minor changes to the set of full conditional distributions. For example, missing values could be handled simply by constructing their full conditional distributions with no changes to the other full conditional distributions being necessary. As another example, prediction for a new observation from the same population can be readily obtained from the sample-based approach. Suppose y^* is a vector of new observations to be predicted, y is the available observed data, $\theta^{(i)}$ is the sample of parameter vector θ at the i^{th} sample ($i = 1, N$). It is straightforward to obtain the predictive distribution as: (see Smith and Robert 1993):

$$p(y^* | y) = \int p(y^* | \theta) \pi(\theta | y) d\theta \approx \frac{1}{N} \sum_i^N p(y^* | \theta^{(i)}) \quad (4.12)$$

Therefore, instead of a prediction procedure based on the estimated parameters as discussed in previous chapters, one may use (4.12) directly in the Gibbs sampler so that parameter estimation and prediction of values for new observations can be carried out simultaneously

CHAPTER 5

A MULTIVARIATE SIMULTANEOUS PREDICTION SYSTEM FOR STAND GROWTH AND YIELD WITH FIXED AND RANDOM EFFECTS

Introduction

Simultaneous model systems are familiar to foresters. Examples can be found in forest growth and yield modeling either at the stand level (e.g. Borders and Bailey 1986) or at the individual tree level (Hasenauer et al 1998). There are perhaps two main reasons that simultaneous equations are popular in forest biometrics. One is the compatibility requirement on forest growth and yield modeling introduced first by Clutter (1963). This concept is that the output by integration of a growth model should be consistent with that from the yield model. This property can easily and naturally result in a system with simultaneous growth and yield models. The other reason that simultaneous model systems are very common in forest growth and yield modeling is an awareness of the advantages obtained when parameter estimation proceeds with a consideration for contemporaneous correlation among components of the system. To date, heavy reliance has been placed on the estimation techniques first developed in econometrics.

Several people have discussed simultaneous growth and yield systems in the forestry literature, but primary interests have been given to parameter estimation and

regression type prediction (Clutter 1963, Furnival and Wilson 1971, Borders and Bailey 1986, Borders 1989, LeMay 1990). Predicting growth and yield for subsequent periods based on all past and current measurements, however, is one of the most important topics in forestry, especially in an inventory updating context. This type of prediction was addressed in conjunction with dominant height growth by Lappi and Bailey (1988), and further defined and explored with linear and nonlinear mixed models in conjunction with forestry engineering by Liski and Nummi (1995,1996). But, rare work, if any, has ever been focused on this type of prediction in conjunction with a simultaneous stand growth and yield system (Tait et al 1988).

In previous chapters I have shown the advantages in prediction of new responses by using a mixed-effects model approach. In this chapter, I will extend the univariate mixed effects model to a simultaneous multivariate system and show how contemporaneous correlations in the simultaneous system can be combined with the random effects to improve the precision of the prediction for new observations.

Generalized simultaneous stand growth and yield model system

Suppose p responses (components of stand growth and yield) have been measured for each of the n individuals (stands) on m occasions. Let $y_{ij}(t_k)$ denote the response j of individual i at occasion k and $x_{ij}(t_k)$ denote the $(q_j \times 1)$ predetermined vector corresponding to response $y_{ij}(t_k)$. The simultaneous stand growth and yield model system may read as:

$$y_{ij}(t_k) = f(y_{ij}(t_k), \alpha_j) + f(x_{ij}(t_k), \beta_j) + \varepsilon_{ij}(t_k) \quad , \quad (5.1a)$$

where, $i=1, \dots, n; j=1, \dots, p; y_{ij}(t_k) = (y_{il})'$ is a $(p_j \times 1)$ vector that consists of p_j ($p_j \leq p-1$) possible endogenous variables other than response j , i.e. $l=1, \dots, p$ and $l \neq j$; $k=1, \dots, m$;

t_k is the stand age of response $y_{ij}(t_k)$ at occasion k ; $f(y_{ij}(t_k), \alpha)$ and $f(x_{ij}(t_k), \beta)$ are appropriate functions of endogenous variables and exogenous variables for response $y_{ij}(t_k)$; α_j and β_j are appropriate dimension parameter vectors to response j , respectively corresponding to the endogenous variable function and exogenous variable function; and $\varepsilon_{ij}(t_k)$ is an error term whose distribution must be specified. The p equations may be written simultaneously as:

$$y_i(t_k) = \begin{pmatrix} y_{i1}(t_k) \\ y_{i2}(t_k) \\ \vdots \\ y_{ip}(t_k) \end{pmatrix} = \begin{pmatrix} f(y_{i1}(t_k), \alpha_1) \\ f(y_{i2}(t_k), \alpha_2) \\ \vdots \\ f(y_{ip}(t_k), \alpha_p) \end{pmatrix} + \begin{pmatrix} f(x_{i1}(t_k), \beta_1) \\ f(x_{i2}(t_k), \beta_2) \\ \vdots \\ f(x_{ip}(t_k), \beta_p) \end{pmatrix} + \begin{pmatrix} \varepsilon_{i1}(t_k) \\ \varepsilon_{i2}(t_k) \\ \vdots \\ \varepsilon_{ip}(t_k) \end{pmatrix} \quad (5.1b)$$

The simultaneous stand growth and yield Model (5.1a-5.1b) is in a formal and general form. In empirical applications, specification is required. If the functions on the right hand side of (5.1a-5.1b) are linear, then they reduce to:

$$y_{ij}(t_k) = y_{ij}^T(t_k)\alpha_j + x_{ij}^T(t_k)\beta_j + \varepsilon_{ij}(t_k) \quad (5.2a)$$

$$\begin{pmatrix} y_{i1}(t_k) \\ y_{i2}(t_k) \\ \vdots \\ y_{ip}(t_k) \end{pmatrix} = \begin{pmatrix} y_{i1}^T(t_k) & 0 & \cdots & 0 \\ 0 & y_{i2}^T(t_k) & \cdots & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & 0 & \cdots & y_{ip}^T(t_k) \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{pmatrix} + \begin{pmatrix} x_{i1}^T(t_k) & 0 & \cdots & 0 \\ 0 & x_{i2}^T(t_k) & \cdots & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & 0 & \cdots & x_{ip}^T(t_k) \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_{i1}(t_k) \\ \varepsilon_{i2}(t_k) \\ \vdots \\ \varepsilon_{ip}(t_k) \end{pmatrix} \quad (5.2b)$$

Such a linear-system growth and yield model has been extensively applied and presented in the forestry literature with various components and functional forms (Clutter 1963, Furnival and Wilson 1971, Sullivan and Clutter 1972, Murphy and Sternizke 1979, Borders and Bailey 1986, Borders 1989, LeMay 1990). Most of the published models

are easy to write down in this formal and general linear system by specifying appropriate terms in (5.2b).

Another appealing special situation arises with the general simultaneous model system (5.1) when response j ($j = 1, \dots, p$) is linear in endogenous variables, but nonlinear with exogenous variables. Namely,

$$y_{ij}(t_k) = y_{ij-}^T(t_k)\alpha_j + f(x_{ij}(t_k), \beta_j) + \varepsilon_{ij}(t_k) \quad (5.3a)$$

i.e.

$$\begin{pmatrix} y_{i1}(t_k) \\ y_{i2}(t_k) \\ \vdots \\ y_{ip}(t_k) \end{pmatrix} = \begin{pmatrix} y_{i1-}^T(t_k) & 0 & \cdots & 0 \\ 0 & y_{i2-}^T(t_k) & \cdots & 0 \\ \vdots & 0 & \vdots & \vdots \\ 0 & 0 & \cdots & y_{ip-}^T(t_k) \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{pmatrix} + \begin{pmatrix} f(x_{i1}(t_k), \beta_1) \\ f(x_{i2}(t_k), \beta_2) \\ \vdots \\ f(x_{ip}(t_k), \beta_p) \end{pmatrix} + \begin{pmatrix} \varepsilon_{i1}(t_k) \\ \varepsilon_{i2}(t_k) \\ \vdots \\ \varepsilon_{ip}(t_k) \end{pmatrix} \quad (5.3b)$$

Equation (5.3a-5.3b) may be more appropriate than (5.2a-5.2b) in some situations. For instance, dominant height is usually modeled as a function of age in the system. Nonlinear growth models for dominant height should be more flexible (e.g. see, Cieszewski and Bailey 1999); (5.3) is more tractable than (5.1) and more flexible than (5.2). Thus (5.3a-5.3b) was chosen as the general formula for a simultaneous stand growth and yield model system to be examined here.

After exploring the data set, a simultaneous model system will be developed with form (5.3b) including components of volume, basal area and dominant height.

The data

The same data from The University of Georgia PMRC research as discussed in previous chapters are used for estimating parameters in the system. The measurements of dominant height (m), basal area ($m^2/ha.$), total-stem outside-bark volume ($m^3/ha.$) are

the stand characteristics of interest in the simultaneous predicted system. Other covariates used in the analysis include stand age, stand density (trees/ha), stand records such as installations (which region the treatment plot belongs to), silvicultural treatment, and soil type.

The simultaneous growth and yield model

Three response components are considered in this simultaneous growth and yield model system. They are dominant height (m), basal area ($m^2/ha.$), and total volume ($m^3/ha.$). The basic models (fixed effects terms) for dominant height and basal area are borrowed directly from Chapters 1 and 2. The basic total volume model is based on the after-transformation linear model presented by Borders and Bailey(1986):

$$\ln(V) = \beta_{3,0} + \beta_{3,1} \ln(HD) + \beta_{3,2} \ln(BA) + \epsilon_v \quad (5.4)$$

Where V = volume per hectare (m^3);

BA = basal area per hectare (m^2);

HD = dominant height (m);

Following a stepwise regression procedure with consideration for the silvicultural treatments, soil types, and corresponding interaction terms (e.g. soil types and silvicultural treatments, treatments and stand age, and so forth, the following terms were significant at $\alpha = 0.05$ level (see Table 19): $\ln(BA)$, $\ln(HD)$, $zchop$, $zchop \times age$, $zh \times zsoil$. (variables $zchop$, zh , and $zsoil$ are defined in Chapter 2). Thus, the basic total volume model is:

$$\begin{aligned} \ln(V) = & \beta_{3,0} + \beta_{3,1} \ln(HD) + \beta_{3,2} \ln(BA) \\ & + \beta_{3,3} zchop + \beta_{3,4} zchop \times age + \beta_{3,5} zh \times zsoil + \epsilon_v \end{aligned} \quad (5.5)$$

Table 19 The significant terms (at $\alpha = 0.05$) resulting from a stepwise regression procedure for choosing the fixed effects in the total volume component of the simultaneous system .

Step	Variable Entered	Number Removed	Model					
			In	Partial R ²	R ²	C(p)	F	Prob>F
1	ln(BA)	1	0.9910	0.9910	13378.288	105288.292	0.0001	
2	ln(HD)	2	0.0083	0.9993	130.0711	11689.8750	0.0001	
3	zchop	3	0.0000	0.9993	100.3630	28.7907	0.0001	
4	zchop×age	4	0.0001	0.9994	20.9123	80.1090	0.0001	
5	zh×zsoil	5	0.0000	0.9994	14.1675	8.6702	0.0033	

Therefore, the basic three-components simultaneous forest growth and yield system in conjunction with silvicultural treatments is as follows:

$$\begin{aligned}
 HD &= (\beta_{1,0} + \beta_{1,1}zburn + \beta_{1,2}zf) \left(\frac{1 - e^{-(\beta_{1,3} + \beta_{1,4}zf + \beta_{1,5}zh)t}}{1 - e^{-(\beta_{1,3} + \beta_{1,4}zf + \beta_{1,5}zh)t_0}} \right)^{(\beta_{1,6} + \beta_{1,7}zf + \beta_{1,8}zbed + \beta_{1,9}zh)} + \varepsilon_H \\
 \ln(BA) &= \beta_{2,0} + \beta_{2,1}zf + \beta_{2,2}zburn \times zsoil + \beta_{2,3}zchop \times t \\
 &\quad + (\beta_{2,4} + \beta_{2,5}zh + \beta_{2,6}zburn + \beta_{2,7}zchop) / t \\
 &\quad + (\beta_{2,8} + \beta_{2,9}zchop) \ln(HD) \\
 &\quad + (\beta_{2,10} + \beta_{2,11}zbed) \ln(TPH) + \varepsilon_B \\
 \ln(V) &= \beta_{3,0} + \beta_{3,1} \ln(HD) + \beta_{3,2} \ln(BA) + \beta_{3,3}zchop \\
 &\quad + \beta_{3,4}zchop \times t + \beta_{3,5}zh \times zsoil + \varepsilon_V
 \end{aligned} \tag{5.6}$$

where $\beta_{1,0}$ through $\beta_{3,5}$ are the parameters to be estimated and all other terms are as defined above in Chapters 2 and 3.

In terms of the general simultaneous model system (5.3a-5.3b), let $y_{ij}(t_k)$ denote the response j of plot i at occasion k and $x_{ij}(t_k)$ the $(q_j \times 1)$ predetermined vector corresponding to response $y_{ij}(t_k)$. Further, $y_{ij}(t_k) = (y_{il})'$ is a $(p_j \times 1)$ vector of endogenous variables that have been used as predictors to response j , t_k is the stand age of response $y_{ij}(t_k)$ at occasion k ; then $i=1, \dots, 191; j=1, 2, 3; k=1, \dots, 5$, indexing 5 repeated

measurements; $\{t_k, k=1, \dots, 5\} = \{5, 8, 11, 14, 17\}$. Writing the three-component simultaneous model system in general form it becomes [see (5.3b)]:

$$y_i(t_k) = \begin{pmatrix} y_{i1}(t_k) \\ y_{i2}(t_k) \\ y_{i3}(t_k) \end{pmatrix} = \begin{pmatrix} y_{i1-}^T(t_k) & 0 & 0 \\ 0 & y_{i2-}^T(t_k) & 0 \\ 0 & 0 & y_{ip-}^T(t_k) \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} + \begin{pmatrix} f_1(x_{i1}(t_k), \beta_1) \\ f_2(x_{i2}(t_k), \beta_2) \\ f_3(x_{i3}(t_k), \beta_3) \end{pmatrix} + \begin{pmatrix} \varepsilon_{i1}(t_k) \\ \varepsilon_{i2}(t_k) \\ \varepsilon_{i3}(t_k) \end{pmatrix} \quad (5.7)$$

Where, $y_i(t_k) = \begin{pmatrix} y_{i1}(t_k) \\ y_{i2}(t_k) \\ y_{i3}(t_k) \end{pmatrix} = \begin{pmatrix} HD_i(t_k) \\ \ln[BA_i(t_k)] \\ \ln[V_i(t_k)] \end{pmatrix}$,

$$\begin{pmatrix} y_{i1-}^T(t_k) & 0 & 0 \\ 0 & y_{i2-}^T(t_k) & 0 \\ 0 & 0 & y_{ip-}^T(t_k) \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & \ln[HD_i(t_k)] & zchop_i \times \ln[HD_i(t_k)] & 0 & 0 \\ 0 & 0 & 0 & \ln[HD_i(t_k)] & \ln[BA_i(t_k)] \end{pmatrix}$$

$$\alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} = \begin{pmatrix} 0 \\ \beta_{2.8} \\ \beta_{2.9} \\ \beta_{3.1} \\ \beta_{3.2} \end{pmatrix}, \text{ i.e., } \alpha_1 \text{ is empty, I put 0 here for convenience in expression, both } \alpha_2 \text{ and } \alpha_3 \text{ have dimension 2. (i.e. } p_1 = 0, p_2 = p_3 = 2\text{)}$$

$$f_1(x_{i1}, \beta_1) = (\beta_{1,0} + \beta_{1,1}zburn + \beta_{1,2}zf) \left(\frac{1 - e^{-(\beta_{1,3} + \beta_{1,4}zf + \beta_{1,5}zh)t}}{1 - e^{-(\beta_{1,3} + \beta_{1,4}zf + \beta_{1,5}zh)t_0}} \right)^{(\beta_{1,6} + \beta_{1,7}zf + \beta_{1,8}zbed + \beta_{1,9}zh)} \quad (5.8a)$$

$$f_2(x_{i2}, \beta_2) = x_{i2}^T \beta_2 \quad (5.8b)$$

$$f_3(x_{i3}, \beta_3) = x_{i3}^T \beta_3 \quad (5.8c)$$

where,

$$x_{i1}(t_k) = (zburn_i, zbed_i, zf_i, zh_i, t_k)^T, \quad \beta_1 = (\beta_{1,0} \ \beta_{1,1} \ \beta_{1,3} \ \beta_{1,4} \ \beta_{1,5} \ \beta_{1,6} \ \beta_{1,7} \ \beta_{1,8} \ \beta_{1,9})^T$$

$$x_{i2}(t_k) =$$

$$(1 \ zf_i \ zburn \times zsoil_i \ zchop_i \times t_k \ 1/t_k \ zh_i / t_k \ zburn_i / t_k \ zchop_i / t_k \ \ln[TPH_i(t_k)] \ zbed_i \times \ln[TPH_i(t_k)])^T$$

$$\beta_2 = (\beta_{2,0} \ \beta_{2,1} \ \beta_{2,2} \ \beta_{2,3} \ \beta_{2,4} \ \beta_{2,5} \ \beta_{2,6} \ \beta_{2,7} \ \beta_{2,10} \ \beta_{2,11})^T,$$

$$x_{i3} = (1 \ zchop_i \ zchop \times t_k \ zh_i \times zsoil_i)^T, \quad \beta_3 = (\beta_{3,0} \ \beta_{3,1} \ \beta_{3,3} \ \beta_{3,4} \ \beta_{3,5})^T$$

Model (5.6) [equivalently Model (5.7)] is a modified version of a common simultaneous growth and yield model system in forestry in that three components (dominant height, basal area and volume) are the focus and silvicultural treatment and soil group differences are included. Endogenous variables *HD* and *BA* are also used as predictors in the system and they are usually correlated with the response on the left hand side of the equation. For example, dominant height is correlated with basal area and both dominant height and basal area are correlated with volume. To avoid the “simultaneous bias” in estimation, various parameter estimation methods (such as three-stage least squares, full information likelihood, the moment method) have been recommended, a fact that is well documented both in forestry and economics (see e.g, Borders 1989, LeMay 1990, Amemiya 1985).

As was presented in Chapters 2 and 3, I believe that a mixed effects model is more appropriate for explaining the variations in forest growth and thus more efficient in prediction. Based on the results from earlier chapters, and a comparison of model performance (likelihood ratio test, *AIC* criterion), for the first approach the intercepts in the three components are taken as mixed effects. Namely, the parameters of $\beta_{1,0}$, $\beta_{1,3}$, and $\beta_{1,6}$ in the dominant height equation, $\beta_{2,0}$ in the basal area equation and $\beta_{3,0}$ in the total volume equation are assumed to vary randomly from plot to plot. That is, in Model (5.6) and (5.7),

$$\beta_{1,0,i} = \beta_{1,0} + b_{1,0,i} \tag{5.9a}$$

$$\beta_{1,3,i} = \beta_{1,3} + b_{1,3,i} \tag{5.9b}$$

$$\beta_{1,6,i} = \beta_{1,6} + b_{1,6,i} \quad (5.9c)$$

$$\beta_{2,0,i} = \beta_{2,0} + b_{2,0,i} \quad (5.9d)$$

$$\beta_{3,0,i} = \beta_{3,0} + b_{3,0,i} \quad (5.9e)$$

The variance covariance structure for the random effects of the domain height model were assumed to be positive symmetric. The random intercepts of the basal area model and the volume model are correlated, but both are independent of the random effects of dominant height .

That is,

$$\begin{pmatrix} b_{1,0,i} \\ b_{1,3,i} \\ b_{1,6,i} \end{pmatrix} \sim N(\mathbf{0}, \mathbf{D}_H), \text{ where } \mathbf{D}_H = \begin{pmatrix} \varphi_{1,11} & \varphi_{1,12} & \varphi_{1,13} \\ \varphi_{1,12} & \varphi_{1,22} & \varphi_{1,23} \\ \varphi_{1,13} & \varphi_{1,23} & \varphi_{1,33} \end{pmatrix}$$

$$\begin{pmatrix} b_{2,0,i} \\ b_{3,0,i} \end{pmatrix} \sim N(\mathbf{0}, \mathbf{D}_{B'}), \text{ where } \mathbf{D}_{B'} = \begin{pmatrix} \varphi_{2,11} & \varphi_{2,12} \\ \varphi_{2,12} & \varphi_{2,22} \end{pmatrix}$$

The endogenous variables that appear on the RHS of an equation were assumed to be independent of the error term of the LHS responses for given values of the random effects. For example, the term of $\ln(HD)$ is no longer to be correlated with the error term of the basal area equation given the random effects. Thus the “simultaneous error” problem is overcome by including random effects in the model and the three-component simultaneous equation system is specified to a seemingly unrelated equation system (Srivastava and Giles 1987) in this case. Given the random effects, however, the error terms of ε_H , ε_B and ε_V are assumed to be correlated to each other, which are the so called contemporaneous correlations in a simultaneous or seemingly unrelated equation system (e.g. Borders and Bailey 1986, LeMay 1990). But unlike what was done in prior

chapters, for simplicity, errors within plot for the same response are assumed to be independent. The variances of the models are assumed to be proportional to a power of the mean response and the proportionality parameters of the power are distinct for the different responses. Let ε_{H_i} , ε_{B_i} , and ε_{V_i} denote the error of model dominant height, basal area and volume for plot i at occasion j . Then,

$$\begin{pmatrix} \varepsilon_{H_i} \\ \varepsilon_{B_i} \\ \varepsilon_{V_i} \end{pmatrix} | \mathbf{b}_i \sim N(\mathbf{0}, \mathbf{R}_i),$$

$$\text{where } \mathbf{R}_i = \sigma^2 G_i^{1/2}(\beta_i, \theta) \Gamma_i(\rho) G_i^{1/2}(\beta_i, \theta) \quad (5.10a)$$

$$G_i(\beta_i, \theta) = \text{diag}(f_{HD_i}^{\theta_1}, f_{BA_i}^{\theta_2}, f_{V_i}^{\theta_3})$$

$$G_i(\beta_i, \theta) = \text{diag}(f_{HD_i}^{\theta_1}, f_{BA_i}^{\theta_2}, f_{V_i}^{\theta_3}), \quad (5.10b)$$

which is the variance function (see definitions in Chapter 2) for the three models;

$$\Gamma_i(\rho) = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{pmatrix} \quad (5.10c)$$

which is the contemporaneous correlation matrix.

To implement the above assumptions in the model formulation, I define the response vector as the general form [see (5.7)], in which three vector components [dominant height, $\ln(BA)$ and $\ln(V)$] are combined into one long vector and the components are identified by the function factor: Level 1 of the function factor indicates dominant height, Level 2 indicates $\ln(BA)$, and Level 3 indicates $\ln(V)$. Different levels have different mean functions and error variances. Computing mainly relied on *NLME* (a s-plus library) by Bates and Pinheiro (1999).

The REML estimates (Bates and Pinheiro 1999) for the above covariance parameters for the random effects are as follows.

$$\hat{\mathbf{D}}_H = \text{cov} \left(\begin{pmatrix} b_{1,0,i} \\ b_{1,3,i} \\ b_{1,6,i} \end{pmatrix} \left(b_{1,0,i} \ b_{1,3,i} \ b_{1,6,i} \right) \right) = \begin{pmatrix} 4.923500635 & -0.00196176707 & 0.1577507530 \\ -0.001961767 & 0.00006827368 & -0.0004576125 \\ 0.157750753 & -0.00045761250 & 0.0075353092 \end{pmatrix}$$

$$\hat{\mathbf{D}}_{BV} = \text{cov} \left(\begin{pmatrix} b_{2,0,i} \\ b_{3,0,i} \end{pmatrix} \left(b_{2,0,i} \ b_{3,0,i} \right) \right) = \begin{pmatrix} 0.00016101623 & 0.00001092752 \\ 0.00001092752 & 0.00001245645 \end{pmatrix}$$

The relatively small variance-covariance of the random effects in basal area and volume indicates that the random intercepts have made a very small adjustment in the models and their contributions may be negligible. So, the random effects in basal area and volume models may be ignored thus reducing the system to one where only the three parameters in dominant height are taken as random. A likelihood ratio test supported such a model reduction. The log likelihood values of the two model systems are almost the same (3032.141 vs. 3032.143), but the former system has three more parameters. So, preference should be given to the reduced model. It is interesting that random effects in basal area and volume models are not important any more as long as the parameters in dominant height are considered to be random and contemporaneous correlation among the three components is also considered. Therefore, in the final model, only the three parameters in the dominant height model are taken to be mixed (both random and fixed), and the errors of the components in the system are correlated to each other, with variance proportional to the power function of its mean. Since the mean function and the power parameters are different for different components, the variances for the three components are different from each other. After considering the random

effects and contemporaneous correlation, the term $zbed \times \ln(TPH)$ was no longer significant (with p-value = 0.3599), so it was dropped from the system. This produced a log-likelihood value of 3031.799. Again, a likelihood ratio test supported the change.

Therefore, the final model system is as follows:

$$\left. \begin{aligned} y_{1,i}(t_k) &= HD_i(t_k) = (\beta_{1,0} + b_{1,i} + \beta_{1,1}zburn + \beta_{1,2}zf) \\ &\quad \times \left(\frac{1 - e^{-(\beta_{1,3} + b_{2,i} + \beta_{1,4}zf + \beta_{1,5}zh)t_k}}{1 - e^{-(\beta_{1,3} + b_{2,i} + \beta_{1,4}zf + \beta_{1,5}zh)t_0}} \right)^{(\beta_{1,6} + b_{3,i} + \beta_{1,7}zf + \beta_{1,8}zbed + \beta_{1,9}zh)} + \varepsilon_{HD} \\ y_{2,i}(t_k) &= \ln[BA_i(t_k)] = \beta_{2,0} + \beta_{2,1}zf + \beta_{2,2}zburn \times zsoil + \beta_{2,3}zchop \times t_k \\ &\quad + (\beta_{2,4} + \beta_{2,5}zh + \beta_{2,6}zburn + \beta_{2,7}zchop) / t_k \\ &\quad + (\beta_{2,8} + \beta_{2,9}zchop) \ln(HD) + \beta_{2,10} \ln(TPH) + \varepsilon_{BA} \\ y_{3,i}(t_k) &= \ln[V_i(t_k)] = \beta_{3,0} + \beta_{3,1} \ln(HD) + \beta_{3,2} \ln(BA) + \beta_{3,3}zchop \\ &\quad + \beta_{3,4}zchop \times t_k + \beta_{3,5}zh \times zsoil + \varepsilon_V \end{aligned} \right\} \quad (5.11a)$$

$$\begin{pmatrix} b_{1,i} \\ b_{2,i} \\ b_{3,i} \end{pmatrix} \sim N(\mathbf{0}, \mathbf{D}), \text{ where } \mathbf{D} = \begin{pmatrix} \varphi_{1,11} & \varphi_{1,12} & \varphi_{1,13} \\ \varphi_{1,12} & \varphi_{1,22} & \varphi_{1,23} \\ \varphi_{1,13} & \varphi_{1,23} & \varphi_{1,33} \end{pmatrix} \quad (5.11b)$$

$$(\varepsilon_{HD}, \varepsilon_{BA}, \varepsilon_V)^T | \mathbf{b}_i \sim N(\mathbf{0}, \mathbf{R}_i), \quad (5.11c)$$

$$\text{where } \mathbf{R}_i = \sigma^2 \mathbf{G}_i^{-1/2} (\beta_i, \theta) \Gamma_i(\rho) \mathbf{G}_i^{-1/2} (\beta_i, \theta) \quad (5.11d)$$

$$\mathbf{G}_i(\beta_i, \theta) = \text{diag}(f_{HD_i}^{\theta_1}, f_{BA_i}^{\theta_2}, f_{V_i}^{\theta_3}) \quad (5.11e)$$

$$\Gamma_i(\rho) = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{pmatrix} \quad (5.11f)$$

$i = 1, \dots, 191$, indexing individual plot; $t_k = (5, 8, 11, 14, 17)^T$ is the age of the stand.

Matrix R is the variance covariance of the three response components in the system given the random effects and G is the power variance function that may explain the within-plot heteroscedasticity. Matrix Γ is the contemporaneous correlation matrix.

The estimated parameters (by REML) are as follows:

$$\hat{\mathbf{D}} = \text{cov} \left(\begin{pmatrix} b_{1,i} \\ b_{2,i} \\ b_{3,i} \end{pmatrix} \begin{pmatrix} b_{1,i} & b_{2,i} & b_{3,i} \end{pmatrix} \right) = \begin{pmatrix} 4.921527284 & -0.00192995263 & 0.1583067656 \\ -0.001929953 & 0.00006735684 & -0.0004605135 \\ 0.158306766 & -0.00046051346 & 0.0075915639 \end{pmatrix}$$

$$\hat{\Gamma}_i = \begin{pmatrix} 1 & 0.441 & 0.330 \\ 0.441 & 1 & 0.568 \\ 0.330 & 0.568 & 1 \end{pmatrix}$$

$$\hat{\theta}_1 = 0.4695512,$$

$$\hat{\theta}_2 = -0.2272813$$

$$\hat{\theta}_3 = -1.228734$$

$$\hat{\sigma} = 0.096753489$$

The estimated parameters for the fixed effects are all significant (Table 20).

Similarly as was the case in Chapter 3, to avoid unstable residuals (a small change in the original value results in a large residual) due to log transformations on small basal area and total volume values for younger stands (e.g age 5), I replace $\ln(BA)$ and $\ln(V)$ with $\ln(BA+1)$ and $\ln(V+1)$ respectively in model fitting.

Table 20 The estimated parameters for the fixed effects in the dominant height, basal area and total volume simultaneous model system

Parameters	Value	Std.Error	t-value	p-value
$\beta_{1,0}$	18.36704	0.3627850	50.6279	<.0001
$\beta_{1,1}$	1.19048	0.3713349	3.2060	0.0014
$\beta_{1,2}$	2.38687	0.3908644	6.1066	<.0001
$\beta_{1,3}$	0.08115	0.0030647	26.4791	<.0001
$\beta_{1,4}$	-0.02192	0.0041760	-5.2491	<.0001
$\beta_{1,5}$	0.01540	0.0038456	4.0040	0.0001
$\beta_{1,6}$	1.98367	0.0393179	50.4522	<.0001
$\beta_{1,7}$	-0.26349	0.0452981	-5.8168	<.0001
$\beta_{1,8}$	-0.16069	0.0246817	-6.5103	<.0001
$\beta_{1,9}$	-0.16502	0.0446935	-3.6922	0.0002
$\beta_{2,0}$	-4.61006	0.1547915	-29.7824	<.0001
$\beta_{2,1}$	0.03259	0.0101457	3.2117	0.0013
$\beta_{2,2}$	0.03880	0.0111913	3.4666	0.0005
$\beta_{2,3}$	-0.01806	0.0014064	-12.8408	<.0001
$\beta_{2,4}$	-0.77866	0.2881011	-2.7027	0.0069
$\beta_{2,5}$	1.51645	0.0827950	18.3158	<.0001
$\beta_{2,6}$	0.27581	0.1046136	2.6365	0.0084
$\beta_{2,7}$	-0.58888	0.1700667	-3.4626	0.0005
$\beta_{2,8}$	1.21727	0.0222883	54.6145	<.0001
$\beta_{2,9}$	0.14209	0.0120492	11.7922	<.0001
$\beta_{2,10}$	0.63554	0.0177388	35.8276	<.0001
$\beta_{3,0}$	-0.47398	0.0084069	-56.3803	<.0001
$\beta_{3,1}$	0.79417	0.0067315	117.9793	<.0001
$\beta_{3,2}$	1.06569	0.0044231	240.9372	<.0001
$\beta_{3,3}$	-0.01620	0.0054913	-2.9492	0.0032
$\beta_{3,4}$	0.00097	0.0003228	3.0030	0.0027
$\beta_{3,5}$	0.00543	0.0023153	2.3432	0.0192

Prediction

The motivation for my work is prediction. As discussed in former chapters, at least two situations should be considered:

Situation 1 -- prediction is for an individual unit (plot) on which no previous observations have ever been made

Situation 2 -- prediction is for an individual unit on which observations at prior times are available.

For each of these situations there are two sub-cases when prediction is made in conjunction with a simultaneous growth and yield model system:

Sub-case 1 -- all the endogenous and exogenous variables on the RHS of the system are observed on the occasion for which prediction is wanted.

Sub-case 2 -- only partial variables on the RHS are observed on the occasion for which prediction is wanted.

For Situation 2 there may also be two other special cases. One occurs when the previous observations are complete on each occasion when the observations were made. That is to say, all the responses, as well as the exogenous variables, were measured on those past occasions for which observations are available. Another special case is that for some past occasions completely measured data are available, but on others only partial data are recorded. The latter is perhaps the most common situation in forestry practice. For example, it is not uncommon in forestry that projections are required for future volume or basal area on a stand for which inventory records contain volume and basal area as well as dominant height, mean diameter and age for some past occasions. For other past occasions only dominant height, mean diameter and age, or maybe only age, will be known. The results of the individual predictions in these different cases generally will be different. Generally speaking, the observed components of endogenous variables in the simultaneous system can be used to improve the precision of prediction of the unobserved components in the system because of the contemporaneous correlation in the simultaneous system. For a unit on which previous observations are known, individual

specified parameters can be estimated, which can be applied to improve the precision of the prediction.

Suppose we are interested in predicting the components in the simultaneous system for a new plot at age t_h . Development of the prediction case by case follows:

Case -1a. No previous observation and all the endogenous variables and exogenous variables on RHS are observed.

In this case, dominant height and basal area at age t_k are available, so prediction is focused on total volume only. One may just put the observed values into the total volume equation and obtain the prediction. However, the contemporaneous correlation estimated by the population data will be ignored in such an approach. Actually, a more precise prediction can be obtained by applying the contemporaneous correlation information. Let $y_1(t_k)$, $y_2(t_k)$ and $y_3(t_k)$ denote random variables of dominant height, $\ln(BA+1)$ and $\ln(V+1)$ at stand age t_k , $f_{H(t_k)}$,

$f_{BA(t_k)}$, and $f_{V(t_k)}$, are the mean function of the three components in the system that have been expressed by the dominant height, basal area and volume models respectively. We can obtain a Bayesian best linear predictor for total volume as follows:

$$\hat{y}_3(t_k) = \hat{f}_{V(t_k)} + \Sigma_{21}\Sigma_{11}^{-1}[y_1(t_k) - \hat{f}_1(t_k)] \quad (5.12a)$$

with approximate prediction variance:

$$\text{var}[y_3(t_k) - \hat{y}_3(t_k)] = \hat{\sigma}^2(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}) + \mathbf{A}_3^T\Omega(\hat{\beta}_{3...})\mathbf{A}_3 \quad (5.12b)$$

where

$$\mathbf{y}_1(t_k) = \begin{pmatrix} HD(t_k) \\ \ln[BA(t_k) + 1] \end{pmatrix}, \quad \mathbf{f}_1(t_k) = \begin{pmatrix} \hat{f}_{HD}(t_k) \\ \hat{f}_{BA}(t_k) \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} \Sigma_{11} & | & \Sigma_{12} \\ \hline \Sigma_{21} & | & \Sigma_{22} \end{pmatrix} = \hat{\sigma}^2 \begin{pmatrix} \hat{v}_{11} & \hat{g}_{11}^{1/2} \hat{\rho}_{12} \hat{g}_{22}^{1/2} & | & \hat{g}_{11}^{1/2} \hat{\rho}_{13} \hat{g}_{33}^{1/2} \\ \hat{g}_{22}^{1/2} \hat{\rho}_{21} \hat{g}_{11}^{1/2} & \hat{g}_{22} & | & \hat{g}_{22}^{1/2} \hat{\rho}_{23} \hat{g}_{33}^{1/2} \\ \hline \hat{g}_{33}^{1/2} \hat{\rho}_{31} \hat{g}_{11}^{1/2} & \hat{g}_{33}^{1/2} \hat{\rho}_{32} \hat{g}_{22}^{1/2} & | & \hat{g}_{33} \end{pmatrix}$$

$$\hat{v}_{11} = Z^T(\hat{\beta}_{1,j}) \hat{D} Z(\hat{\beta}_{1,j}) + \hat{\sigma}^2 \hat{g}_{11} \quad (5.12c)$$

$$Z^T(\hat{\beta}_{1,j}) = \frac{\partial f_{HD}(t_k, \beta_{1,j})}{\partial \beta_k^T} \Big|_{\beta_{1,j} = \hat{\beta}_{1,j}}, \quad (5.12d)$$

$$\beta_{1,j} = (\beta_{1,0}, \dots, \beta_{1,9})^T \text{ and } \beta_k = (\beta_{1,0} \ \beta_{1,3} \ \beta_{1,6})^T$$

Variables \hat{g}_{11} , \hat{g}_{22} and \hat{g}_{33} are, respectively, the estimated variance functions for dominant height, basal area and volume which are defined in (5.11e).

Estimates $\hat{\rho}_{12}$, $\hat{\rho}_{13}$ and $\hat{\rho}_{23}$ are the estimated contemporaneous correlation parameters, which are defined in (5.10c). Note that $\hat{\rho}_{21} = \hat{\rho}_{12}$, $\hat{\rho}_{31} = \hat{\rho}_{13}$, and $\hat{\rho}_{32} = \hat{\rho}_{23}$.

Vector A_3 is the 6×1 covariates vector (including the intercept) for the component of volume. Matrix $\Omega(\hat{\beta}_{3...})$ is a 6×6 covariance matrix for parameters (fixed effects) of total volume model.

Case -1b. No previous observation and only part of the endogenous variables and exogenous variables on the RHS are observed.

In this case, one usually assumes that all the exogenous variables are observed and some endogenous variables on RHS need to be estimated by the exogenous variables before the prediction for some other endogenous variables on the LHS can be made. Based on this assumption, the prediction for unobserved endogenous variables that are

only functions of exogenous variables can be obtained first by the corresponding mean function and then the predicted values may be used to take the place of the unobserved endogenous variables on the RHS. To do so, it is required that the simultaneous system of equations be sequentially related. Fortunately, this is usually the case in forestry (Borders, 1989). The variance of the prediction will increase due to the predicted errors of the endogenous variables if a response is a function of some unobserved endogenous variables. On the other hand, if contemporaneous observations are available for some RHS endogenous variables, then similar to *Case-1a*, one may use the information of population contemporaneous correlation to improve the prediction(s).

For the above three-component simultaneous system, I discuss the following three situations for this sub-case:

1. Only exogenous variables are observed at time point t_k , at which prediction is required. So all the endogenous variables (dominant height, basal area and total volume) need to be predicted. In this sub-case, we first predict dominant height from the information in the exogenous variables. Since there is no repeated observations in this case, this prediction is only based on the marginal mean function of dominant height. Basal area prediction is based on the predicted dominant height and other exogenous variables. Volume prediction is unavailable until the predictions of dominant height and basal area have been achieved. Of course, the prediction error of basal area will increase because the predicted dominant height was used in the prediction of basal area. Using the predicted values for dominant height and basal area, instead of real observations, in volume prediction will also unavoidably increase the prediction error of volume.

Dominant height prediction

The predicted value is given by

$$\begin{aligned}\hat{y}_1(t_k) &= H\hat{D}(t_k) = \hat{f}_{HD}(t_k) \\ &= (\hat{\beta}_{1,0} + \hat{\beta}_{1,1}zburn + \hat{\beta}_{1,2}zf) \left(\frac{1 - e^{-(\hat{\beta}_{1,3} + \hat{\beta}_{1,4}zf + \hat{\beta}_{1,5}zh)t_k}}{1 - e^{-(\hat{\beta}_{1,3} + \hat{\beta}_{1,4}zf + \hat{\beta}_{1,5}zh)t_0}} \right)^{(\hat{\beta}_{1,6} + \hat{\beta}_{1,7}zf + \hat{\beta}_{1,8}zbed + \hat{\beta}_{1,9}zh)}\end{aligned}\quad (5.13a)$$

with prediction variance

$$\text{var}[y_1(t_k) - \hat{y}_1(t_k)] = Z^T(\hat{\beta}_{1,j}) \hat{D} Z(\hat{\beta}_{1,j}) + \hat{\sigma}^2 \hat{g}_{11} + \mathbf{A}_1^T \Omega(\hat{\beta}_{1...}) \mathbf{A}_1 \quad (5.13b)$$

where

\mathbf{A}_1 is a 10×1 vector with components $a_{1,0}$ to $a_{1,9}$ (given below), and $\Omega(\hat{\beta}_{1...})$ is a 10×10 covariance matrix for parameters (fixed effects) of the dominant height model.

$$a_{1,j} = \left. \frac{\partial f_{HD}(t_k, \beta_{1,j})}{\partial \beta_{1,j}^T} \right|_{\beta_{1,j} = \hat{\beta}_{1,j}}, \quad j = 0, \dots, 9 \quad (5.13c)$$

and all other terms defined above.

Basal area prediction

The prediction is

$$\begin{aligned}\hat{y}_2(t_k) &= \ln(B\hat{A} + 1) = \hat{\beta}_{2,0} + \hat{\beta}_{2,1}zf + \hat{\beta}_{2,2}zburn \times zsoil + \hat{\beta}_{2,3}zchop \times t_k \\ &\quad + (\hat{\beta}_{2,4} + \hat{\beta}_{2,5}zh + \hat{\beta}_{2,6}zburn + \hat{\beta}_{2,7}zchop) / t_k \\ &\quad + (\hat{\beta}_{2,8} + \hat{\beta}_{2,9}zchop) \ln(H\hat{D}) + \hat{\beta}_{2,10} \ln(TPH)\end{aligned}\quad (5.14a)$$

with prediction variance

$$\text{var}[y_2(t_k) - \hat{y}_2(t_k)] = \hat{\sigma}^2 \hat{g}_{22} + \mathbf{A}_2^T \Omega(\hat{\beta}_{2...}) \mathbf{A}_2 + V_{BA}(H\hat{D}) \quad (5.14b)$$

where

\mathbf{A}_2 is a 11×1 covariates vector (including the intercept) for components of basal area.

$\Omega(\hat{\beta}_{2...})$ is a 11×11 covariance matrix for parameters (fixed effects) of basal area model.

$V_{BA}(H\hat{D})$ is the variance due to the fact that dominant height in basal area model was replaced by the fitted value in this case, which unavoidably introduces error into the model.

We know the marginal mean function of dominant height has error with variance of \hat{v}_{11} , however, the exact variance due to the dominant height model is not so obvious.

The dominant height in the basal area model was input in logarithm form, i.e., $\ln(HD)$ while \hat{v}_{11} is the variance of the dominant height model error in HD scale. Note

$\hat{e}_{HD} = HD - H\hat{D}$, and $\text{var}[\hat{e}_{HD}] = \hat{v}_{11}$ at the marginal mean level. Let

$\hat{e}_T = \ln(HD) - \ln(H\hat{D})$, now I derive an approximation to $V_{BA}(H\hat{D})$ as the follows:

$$\begin{aligned}\hat{e}_T &= \ln(HD) - \ln(H\hat{D}) \\ &= \ln(H\hat{D} + \hat{e}_{HD}) - \ln(H\hat{D}) \\ &= \ln\left(1 + \frac{\hat{e}_{HD}}{H\hat{D}}\right) \\ &= \frac{\hat{e}_{HD}}{H\hat{D}} - \frac{1}{2}\left(\frac{\hat{e}_{HD}}{H\hat{D}}\right)^2 + \frac{1}{3}\left(\frac{\hat{e}_{HD}}{H\hat{D}}\right)^3 - \frac{1}{4}\left(\frac{\hat{e}_{HD}}{H\hat{D}}\right)^4 + \dots + (-1)^n \frac{1}{n+1}\left(\frac{\hat{e}_{HD}}{H\hat{D}}\right)^{n+1} + \dots\end{aligned}\tag{5.14c}$$

If we ignore all the terms after the first one in (5.14c), then it is simplified as follows:

$$\hat{e}_T \approx \frac{\hat{e}_{HD}}{H\hat{D}}, \text{ thus } \text{var}(\hat{e}_T) \approx \frac{1}{H\hat{D}^2} \text{var}(\hat{e}_{HD}) = \frac{\hat{v}_{11}}{H\hat{D}^2}\tag{5.14d}$$

$$\text{Thus, } V_{BA}(H\hat{D}) \approx (\hat{\beta}_{2.8} + \hat{\beta}_{2.9} zchop)^2 \frac{\hat{v}_{11}}{H\hat{D}^2}\tag{5.14e}$$

Total volume prediction

The predicted volume is

$$\hat{y}_3(t_k) = \ln(\hat{V} + 1) = \hat{\beta}_{3,0} + \hat{\beta}_{3,1} \ln(H\hat{D}) + \hat{\beta}_{3,2} \ln(B\hat{A}) + \hat{\beta}_{3,3} zchop + \hat{\beta}_{3,4} zchop \times t_k + \hat{\beta}_{3,5} zh \times zsoil \quad (5.15a)$$

with prediction variance

$$\text{var}[y_3(t_k) - \hat{y}_3(t_k)] = \hat{\sigma}^2 \hat{g}_{33} + \mathbf{A}_3^T \Omega(\hat{\beta}_{3,\dots}) \mathbf{A}_3 + V_V(H\hat{D}) + V_V(B\hat{A}) \quad (5.15b)$$

Expressions $V_V(H\hat{D})$ and $V_V(B\hat{A})$ are the variances due to the replacement of dominant height and basal area with corresponding predicted values in the volume model.

$$V_V(H\hat{D}) \approx \beta_{3,1}^2 \frac{\hat{v}_{11}}{H\hat{D}^2} \quad (5.15c)$$

$$V_V(B\hat{A}) \approx \beta_{3,2}^2 \left(\hat{\sigma}^2 \hat{g}_{22} + (\hat{\beta}_{2,8} + \hat{\beta}_{2,9} zchop)^2 \frac{\hat{v}_{11}}{H\hat{D}^2} \right) \quad (5.15d)$$

2. In this second special situation, suppose that dominant height at t_k was observed and predictions are needed for basal area and volume .

For this situation we may combine the approaches in *Case 1a* and *Case 1b-1*. The contemporaneous correlation of dominant height and basal area can be applied to improve the basal area prediction. Then predicted basal area can be directly applied in volume prediction. Contemporaneous correlation of dominant height and volume also apply in this sub-case. i.e.

Basal area prediction

The predicted value is

$$\hat{y}_2(t_k) = \hat{f}_{BA}(t_k) + \frac{\hat{\rho}_{12}^2 \hat{\sigma}^4 \hat{g}_{11} \hat{g}_{22}}{\hat{v}_{11}} [y_1(t_k) - \hat{f}_{HD}(t_k)] \quad (5.16a)$$

with prediction variance:

$$\text{var}[y_2(t_k) - \hat{y}_2(t_k)] = \hat{\sigma}^2 \left(1 - \frac{\hat{\rho}_{12}^2 \hat{\sigma}^2 \hat{g}_{11}}{\hat{v}_{11}}\right) \hat{g}_{22} + \mathbf{A}_2^T \Omega(\hat{\beta}_{2...}) \mathbf{A}_2 \quad (5.16b)$$

Total volume prediction

The predicted value is

$$\hat{y}_3(t_k) = \hat{f}_v^*(t_k) + \frac{\hat{\rho}_{13}^2 \hat{\sigma}^4 \hat{g}_{11} \hat{g}_{33}}{\hat{v}_{11}} [y_1(t_k) - \hat{f}_{HD}(t_k)] \quad (5.17a)$$

with prediction variance

$$\begin{aligned} \text{var}[y_3(t_k) - \hat{y}_3(t_k)] &= \\ \hat{\sigma}^2 \left[\left(1 - \frac{\hat{\rho}_{13}^2 \hat{\sigma}^2 \hat{g}_{11}}{\hat{v}_{11}}\right) \hat{g}_{33} + \hat{\beta}_{3,2}^2 \left(1 - \frac{\hat{\rho}_{12}^2 \hat{\sigma}^2 \hat{g}_{11}}{\hat{v}_{11}}\right) \hat{g}_{22} \right] &+ \mathbf{A}_3^T \Omega(\hat{\beta}_{3...}) \mathbf{A}_3 \end{aligned} \quad (5.17b)$$

where,

$$\begin{aligned} \hat{f}_v^*(t_k) &= \hat{\beta}_{3,0} + \hat{\beta}_{3,1} \ln(HD) + \hat{\beta}_{3,2} \ln(BA) + \hat{\beta}_{3,3} zchop \\ &+ \hat{\beta}_{3,4} zchop \times t_k + \hat{\beta}_{3,5} zh \times zsoil \end{aligned} \quad (5.17c)$$

3. In special situation 3, suppose basal area at t_k was observed and predictions are needed for dominant height and volume.

This sub-case is parallel to *Case 1b-2*.

Dominant height prediction

The prediction is

$$\hat{y}_1(t_k) = \hat{f}_{HD}(t_k) + \frac{\hat{\rho}_{12}^2 \hat{\sigma}^4 \hat{g}_{22} \hat{g}_{11}}{\hat{\sigma}^2 \hat{g}_{22} + V_{BA}(H\hat{D})} [y_2(t_k) - \hat{f}_{BA}^*(t_k)] \quad (5.18a)$$

with prediction variance:

$$\text{var}[y_1(t_k) - \hat{y}_1(t_k)] = \hat{v}_{11} - \frac{\hat{\rho}_{12}^2 \hat{\sigma}^4 \hat{g}_{11} \hat{g}_{22}}{\hat{\sigma}^2 \hat{g}_{22} + V_{BA}(H\hat{D})} + \mathbf{A}_1^T \Omega(\hat{\beta}_{1...}) \mathbf{A}_1 \quad (5.18b)$$

where,

$$\begin{aligned}\hat{f}_{BA}^*(t_k) = & \hat{\beta}_{2,0} + \hat{\beta}_{2,1}zf + \hat{\beta}_{2,2}zburn \times zsoil + \hat{\beta}_{2,3}zchop \times t_k \\ & + (\hat{\beta}_{2,4} + \hat{\beta}_{2,5}zh + \hat{\beta}_{2,6}zburn + \hat{\beta}_{2,7}zchop) / t_k \\ & + (\hat{\beta}_{2,8} + \hat{\beta}_{2,9}zchop) \ln(H\hat{D}) + \hat{\beta}_{2,10} \ln(TPH)\end{aligned}\quad (5.18c)$$

$V_{BA}(H\hat{D})$ is defined in (5.14e);

Total volume prediction

The predicted value is

$$\hat{y}_3(t_k) = \hat{f}_v^*(t_k) + \frac{\hat{\rho}_{23}^2 \hat{\sigma}^4 \hat{g}_{22} \hat{g}_{33}}{\hat{\sigma}^2 \hat{g}_{22} + V_{BA}(H\hat{D})} [y_2(t_k) - \hat{f}_{BA}^*(t_k)] \quad (5.19a)$$

with prediction variance

$$\text{var}[y_3(t_k) - \hat{y}_3(t_k)] = \hat{\beta}_{3,1}^2 V_v(H\hat{D}) + \hat{\sigma}^2 \hat{g}_{33} \left(1 - \frac{\hat{\rho}_{23}^2 \hat{\sigma}^2 \hat{g}_{22}}{\hat{\sigma}^2 \hat{g}_{22} + V_{BA}(H\hat{D})} \right) + \mathbf{A}_3^T \Omega(\hat{\beta}_{3,\dots}) \mathbf{A}_3 \quad (5.19b)$$

where,

$$\begin{aligned}\hat{f}_v^*(t_k) = & \hat{\beta}_{3,0} + \hat{\beta}_{3,1} \ln(H\hat{D}) + \hat{\beta}_{3,2} \ln(BA) + \hat{\beta}_{3,3} zchop \\ & + \hat{\beta}_{3,4} zchop \times t_k + \hat{\beta}_{3,5} zh \times zsoil\end{aligned}\quad (5.19c)$$

and \hat{f}_{BA}^* was defined in (5.18c).

Case -2a. Repeated measurements are available and all the endogenous and exogenous variables on the RHS are observed at time t_k .

This is the same as Case 1a except that some previous observations are available, so random effects can be estimated (predicted) by BLUP and prediction can be obtained at the individual plot level. Dominant height is the only component in the simultaneous system that contains random effects, so one just needs to rebuild $\hat{f}_{HD}(t_k)$ and \hat{v}_{11} in (13a-13c) so that the BLUP of random effects is applied in prediction.

Suppose there are m dominant height measurements up to stand age t_k available on the plot on which total volume prediction at t_k is required. As in *Case 1a*, dominant height and basal area are observed in this case also. Then the *BLUP* for the random effects is (see Chapter 2)

$$\hat{b}_m(t_k) = \hat{D} \hat{Z}_m^T (\hat{Z}_m \hat{D} \hat{Z}_m^T + \hat{\sigma}^2 \hat{g}_{m,11})^{-1} \hat{e}_m \quad (5.20a)$$

where \hat{Z}_m is a $m \times 3$ design matrix for random effects in the dominant height component.

$$\hat{Z}_m^T = \left. \frac{\partial f_{HD}(t_i, \beta_{1,j})}{\partial \beta_k^T} \right|_{\beta_j = \hat{\beta}_j}, \quad i = 1, \dots, m; j = 1, \dots, 9; \beta_k = (\beta_{1,0} \ \beta_{1,3} \ \beta_{1,6})^T \quad (5.20b)$$

\hat{e}_m is a $m \times 1$ vector of residuals for the mean function of dominant height and m repeated measurements, i.e.

$$\hat{e}_m = HD_{t_i} - f_{HD}(t_i) \quad (5.20c)$$

$\hat{g}_{m,11}$ is a $m \times 1$ vector of variance function for the dominant height model. i.e.,

$$\hat{g}_{m,11} = (f_{HD}^{\hat{\theta}_1}(t_1), \dots, f_{HD}^{\hat{\theta}_1}(t_m))^T$$

The variance of the *BLUP* is

$$\hat{U}_m = \text{var}[\hat{b}_m(t_k)] = \hat{D} - \hat{D} \hat{Z}_m^T \hat{\Sigma}_m^{-1} \hat{Z}_m \hat{D} \quad (5.20d)$$

where $\hat{\Sigma}_m = \hat{Z}_m \hat{D} \hat{Z}_m^T + \hat{\sigma}^2 \hat{g}_{m,11}$.

Now equations (13a)- (13c) can be applied to this case by replacing $\hat{f}_{HD}(t_k)$ with $\hat{f}_{HD}^*(t_k)$ and \hat{v}_{11} with \hat{v}_{11}^* as follows:

$$\begin{aligned} \hat{f}_{HD}^*(t_k) &= (\hat{\beta}_{1,0} + \hat{b}_1 + \hat{\beta}_{1,1} zburn + \hat{\beta}_{1,2} zf) \\ &\times \left(\frac{1 - e^{-(\hat{\beta}_{1,3} + \hat{b}_2 + \hat{\beta}_{1,4} zf + \hat{\beta}_{1,5} zh)t_k}}{1 - e^{-(\hat{\beta}_{1,3} + \hat{b}_2 + \hat{\beta}_{1,4} zf + \hat{\beta}_{1,5} zh)t_0}} \right)^{(\hat{\beta}_{1,6} + \hat{b}_3 + \hat{\beta}_{1,7} zf + \hat{\beta}_{1,8} zbed + \hat{\beta}_{1,9} zh)} \end{aligned} \quad (5.20e)$$

$$\hat{v}_{11}^* = z_k^T(\hat{\beta}_j) \hat{U}_m z_k(\hat{\beta}_j) + \hat{\sigma}^2 \hat{g}_{11} \quad (5.20f)$$

Case -2b. Repeated measurements are available and only part of the endogenous variables and exogenous variables on RHS are observed at time t_k

This case is parallel to Case 1b but the prior observations of dominant height can be used to estimate the plot-specific random effects [using (5.20a)], and thus the predictions may be improved . The formulas for different situations discussed in Case 1b are still valid as long as the following adjustments are also carried out:

① Similarly as in Case 2a, replace $\hat{f}_{HD}(t_k)$ with $\hat{f}_{HD}^*(t_k)$ and \hat{v}_{11} with \hat{v}_{11}^* in equations (5.13a) –(5.19c), if applicable. Expressions $\hat{f}_{HD}^*(t_k)$ and \hat{v}_{11}^* are given in (5.20e) and (5.20f). For example, replace (5.13a) with (5.20e), change \hat{v}_{11} to \hat{v}_{11}^* in (5.16b), etc.

The prediction variance of dominant height [(5.13b) for Case 1b-1 and (5.18b) for Case 1b-3] should provide some adjustments to predictions to account for the given random effects. Namely, (5.13b) changes to:

$$\text{var}[\hat{y}_1(t_k) - \hat{y}_1(t_k)] = \hat{A}_1^T \hat{\Omega}(\hat{\beta}_{1...}) \hat{A}_1 + \hat{Z}_k^T \hat{V}_m \hat{Z}_k + \hat{A}_1^T \hat{C}_m \hat{Z}_k + \hat{Z}_k^T \hat{C}_m^T \hat{A}_1 + \hat{\sigma}^2 \hat{g}_{11} \quad (5.21a)$$

and (5.18b) changes to

$$\begin{aligned} \text{var}[\hat{y}_1(t_k) - \hat{y}_1(t_k)] &= \hat{A}_1^T \hat{\Omega}(\hat{\beta}_{1...}) \hat{A}_1 + \hat{Z}_k^T \hat{V}_m \hat{Z}_k + \hat{A}_1^T \hat{C}_m \hat{Z}_k + \hat{Z}_k^T \hat{C}_m^T \hat{A}_1 + \hat{\sigma}^2 \hat{g}_{11} \\ &\quad - \frac{\hat{\rho}_{12}^2 \hat{\sigma}^4 \hat{g}_{11} \hat{g}_{22}}{\hat{\sigma}^2 \hat{g}_{22} + V_{BA}(H\hat{D})} \end{aligned} \quad (5.21b)$$

Where

$$V_m = \text{var}(\hat{b}_m - b_m) = \hat{D} - \hat{D} \hat{Z}_m^T \hat{\Sigma}_m^{-1} \hat{Z}_m \hat{D} + \hat{D} \hat{Z}_m^T \hat{A}_{1m} \Omega(\hat{\beta}_{1...}) \hat{A}_{1m}^T \hat{Z}_m \hat{D} \quad (5.21c)$$

$$C_m = Cov[(\hat{\beta}_{1...} - \beta_{1...}), (\hat{b}_m - b_m)^T] \equiv -\Omega(\hat{\beta}_{1...}) \hat{A}_{1m}^T \hat{\Sigma}_m^{-1} \hat{Z}_m \hat{D} \quad (5.21d)$$

$$\hat{\Sigma}_m = \hat{Z}_m \hat{D} \hat{Z}_m^T + \hat{\sigma}^2 \hat{g}_{m,11} \quad (5.21e)$$

$$V_{BA}(H\hat{D}) = (\hat{\beta}_{2,8} + \hat{\beta}_{2,9} zchop)^2 \frac{\hat{v}_{11}}{H\hat{D}^2} \quad (5.21f)$$

Examples

Example 1. Suppose we want to predict the slash pine dominant height, basal area and total volume per hectare at age 20 on a new plot with silvicultural treatments *FCBH* (chop, burn, herbicide, fertilize). We also know the soil type is nonspodosol and the stand density is 720 trees/ha at age 20 (measured or projected by some other models). There is no other information available on this plot. So this is Case 1b-1 as discussed above.

The predicted dominant height is 19.18 m [using (5.13a)]:

$$Z_k^T(\hat{\beta}_j) = (0.8741891 \ 36.43489 \ -1.658602), \quad \hat{g}_{11} = f_{HD}^{\hat{\beta}_1} = 19.18355^{0.4695512} = 4.00314$$

$$\hat{A}_1^T = (0.8742 \ 0.8742 \ 0.8742 \ 36.435 \ 36.435 \ 36.435 \ -1.6586 \ -1.6586 \ -1.6586 \ -1.6586)$$

$$\text{From (5.12c), } \hat{v}_{11} = 3.382487$$

The variance of the dominant height prediction is 3.435234 [using (5.13b)], in which random effects (variation between plots) are the main component (97.37%). The standard error of the prediction is 1.85 m.

The predicted value for $\ln(BA + 1)$ is 3.318365, so predicted basal area is 26.615 (m^2/ha).

The estimated variance of the error due dominant height in the basal area model [using (5.14e)] is $V_{BA}(\hat{HD}) = 0.01698428$.

$$\hat{g}_{22} = f_{BA}^{\hat{\theta}_2} = 0.761384, \mathbf{A}_2^T \Omega(\beta_{2...}) \mathbf{A}_2 = 0.0001484329$$

The variance of the predicted value for $\ln(\hat{BA} + 1)$ is 0.02426021 and the error due to dominant height prediction is the main component (70.01%).

The predicted value for $\ln(\hat{V} + 1)$ is 5.417018, so $\hat{V} = 224.2067 (m^3/ha)$

The estimated variance of the error due dominant height in the volume prediction model [using (5.15c)] is

$$V_V(\hat{HD}) = 0.005797407, V_V(\hat{BA}) = 0.02738224 \quad \hat{g}_{33} = f_V^{\hat{\theta}_3} = 0.125431,$$

The value of $\mathbf{A}_3^T \Omega(\beta_{3...}) \mathbf{A}_3$ is 6.103989e-006, so the variance of the prediction for $\ln(\hat{V} + 1)$ is 0.03436095. Errors due to dominant height prediction and basal area prediction contribute 16.87% and 79.69%, respectively, in this prediction variance.

Example 2. Suppose we also know that the dominant height at age 20 of the stand discussed in Example 1 is actually 19.6 m and all other information is the same. Now predictions are required for the basal area and volume of the stand at age 20. This is Case 1b-2 discussed above. In this case, the observed dominant height can be directly applied in basal area prediction and then predicted basal error can be applied in volume prediction. The information on population contemporaneous correlation between dominant height and basal area as well as between volume and dominant height may be applied for the prediction in this case.

From (5.16a), $\ln(\hat{BA} + 1) = 3.347559 + 0.00001532666(19.6 - 19.18) = 3.347566$, so $\hat{BA} = 27.433 (m^2/ha)$, with prediction variance [using (5.16b)] on $\ln(\hat{BA} + 1)$ of 0.007243122. The prediction variance is much smaller than that in Example 1 due to the addition of known information about current height.

Similarly, by (5.17a), $\ln(\hat{V} + 1) = 5.465193$, so $\hat{V} = 235.3215 (m^3/ha)$, with prediction variance on $\ln(\hat{V} + 1)$ [using (5.17b)] of 0.00922726. This prediction variance is also much smaller than the corresponding value in Example 1.

Example 3. Suppose dominant height is unknown but we know that the basal area at age 20 is $25.0 (m^2/ha)$. We want to predict dominant height and volume at age 20 for the stand. This is Case 1b-3 discussed above.

From (5.18a), $\hat{HD} = 19.18 \text{ m}$. This is almost the same predicted height as in Example 1, 19.18342 m vs. 19.1835 m. The prediction variance [using (5.18b)] is 3.43308, a slightly smaller value than in Example 1 due to the available information on basal area.

Similarly, using (5.19a), $\ln(\hat{V} + 1) = 5.352775$ and $\hat{V} = 210.1935 (m^3/ha)$. The prediction variance [using (5.19b)] in $\ln(\hat{V} + 1)$ is 0.006881516, which is smaller than that in Example 2.

Example 4. Suppose we know that the dominant height is 19.6 m and basal area is $25.0 (m^2/ha)$ at age 20, and all other information is the same as in Example 1. Now prediction is required for the volume at age 20. This is Case 1a. So, with (5.12a),

$$\ln(\hat{V} + 1) = 5.3698 + (0.2351 \ 0.1765) \begin{pmatrix} 3.3825 & 0.76991 \\ 0.7699 & 0.7614 \end{pmatrix}^{-1} \begin{pmatrix} 19.6 - 19.1835 \\ \ln(25 + 1) - 3.318 \end{pmatrix} = 5.366263$$

with prediction variance [using (5.12b)] in $\ln(\hat{V} + 1)$ of 0.001496376, which is much smaller than the corresponding values from any of the above examples due to the availability of both basal area and height information.

Suppose for the examples shown above, we also know that the dominant heights were 12.5 m, 14.8 m and 17.6 m at ages 11, 14 and 17 respectively for the slash pine stand on which predictions of stand characteristics at age 20 are required. Then we may estimate/predict the random effects with the information from these repeated measurements and apply these to adjust the above predictions.

Example 5. The information available in Example 1 is augmented by the repeated measurements on dominant height.

This is the Case 2b-1. In this case, one has no information about dominant height, basal area, or volume at age 20. Only the values of the exogenous variables are known at this age as in Example 1. However, the dominant height observations for prior ages can be used to estimate the random effects associated with this individual stand and the precision of the predictions in the simultaneous system will therefore be improved. The calculation details follow:

The residuals of dominant height prediction from the marginal mean model (e_m) are:

$$\hat{e}_m = \begin{pmatrix} \hat{e}(t_1 = 11) \\ \hat{e}(t_1 = 14) \\ \hat{e}(t_1 = 17) \end{pmatrix} = \begin{pmatrix} 12.5 - 11.568 \\ 14.8 - 14.525 \\ 17.6 - 17.057 \end{pmatrix} = \begin{pmatrix} 0.932 \\ 0.275 \\ 0.543 \end{pmatrix} \quad [\text{by (5.20c)}]$$

$$\hat{Z}_m = \begin{pmatrix} 0.5271558 & 73.14009 & -4.762590 \\ 0.6619085 & 68.18880 & -3.853941 \\ 0.7772879 & 54.96920 & -2.763346 \end{pmatrix} \quad [\text{by (5.20b)}], \quad \hat{\sigma}^2 \hat{g}_{m,11} = \begin{pmatrix} 0.05555101 \\ 0.06181747 \\ 0.06666201 \end{pmatrix}$$

$$\hat{\Sigma}_m = \hat{Z}_m \hat{D} \hat{Z}_m^T + \hat{\sigma}^2 \hat{g}_{m,11} = \begin{pmatrix} 1.332833 & 1.488424 & 1.618667 \\ 1.488424 & 1.904156 & 2.113452 \\ 1.618667 & 2.113452 & 2.596551 \end{pmatrix}$$

$$\hat{\Sigma}_m^{-1} = \begin{pmatrix} 6.0629122 & -5.634250 & 0.8064075 \\ -5.6342496 & 10.673013 & -5.1749119 \\ 0.8064075 & -5.174912 & 4.0945165 \end{pmatrix}$$

Thus the *BLUP* based on the dominant height measurements at ages 11,14 and 17 is:

$$\hat{b}_m = \begin{pmatrix} -0.090494047 \\ 0.006997612 \\ -0.045273954 \end{pmatrix} \quad [\text{by (5.20a)}], \quad H\hat{D}(t_m = (11,14,17)^T | \hat{b}_m) = \begin{pmatrix} 12.23693 \\ 15.09876 \\ 17.47668 \end{pmatrix}. \quad \text{Without}$$

surprise, they are closer to the observed values than those predicted with the marginal mean function.

$$\hat{U}_m = \text{var}[\hat{b}_m] = \begin{pmatrix} 0.394224726 & -0.00216267630 & 0.0247216597 \\ -0.002162676 & 0.00001449415 & -0.0001464945 \\ 0.024721660 & -0.00014649445 & 0.0017016494 \end{pmatrix} \quad [\text{by (5.20d)}]$$

$$Z_k(t_k = 20) = \begin{pmatrix} 0.8885566 \\ 32.96657 \\ -1.5196 \end{pmatrix} \quad [\text{by (5.20b)}], \quad \text{Thus } \hat{v}_{11}^* = 0.1898397 \quad [\text{by (5.20f)}]$$

The predicted dominant height at age 20 given the random effects [using (5.20e)] is:

$$H\hat{D}(t_k = 20 | \hat{b}_m) = 19.42 \text{ m}$$

To obtain the prediction variance, one must calculate V_m and C_m using the m ($m=3$ in this case) repeated dominant height observations. This can be done with formulas (5.21b) and (5.21c):

$$V_m = \begin{pmatrix} 0.458974305 & -0.00214581887 & 0.0265500167 \\ -0.002145819 & 0.00001588834 & -0.0001544563 \\ 0.026550017 & -0.00015445626 & 0.0018045053 \end{pmatrix} \quad [\text{by (5.21c)}]$$

$$C_m = \begin{pmatrix} 0.01399797609 & -1.941947e-005 & 5.352615e-004 \\ -0.02072013121 & 4.440343e-005 & -8.873856e-004 \\ -0.06525178909 & 3.899050e-005 & -2.182265e-003 \\ 0.00006295375 & 2.731233e-007 & 2.200854e-007 \\ 0.00021966928 & -1.303234e-006 & 1.445937e-005 \\ -0.00019405972 & -1.643734e-006 & 4.189981e-006 \\ 0.00086686295 & -6.430498e-006 & 6.490735e-005 \\ 0.00041514561 & -8.874901e-007 & 1.773317e-005 \\ -0.00025592281 & 1.172260e-005 & -7.880934e-005 \\ -0.00219117301 & -3.935788e-006 & -4.147894e-005 \end{pmatrix} \quad [\text{by (5.21d)}]$$

Therefore, the prediction variance $\text{Var}(\hat{H}\hat{D}[l_k = 20 | \hat{b}_m - HD]) = 0.1924406$

[using (5.21a)], and the standard error of the prediction is 0.439 m, which is only 23% of standard error (1.85 m) for the comparable situation (Case 1b-1) without the availability of previous observations on height.

Basal area and total volume predictions can be obtained using (5.14a-5.14b) and (5.15a-5.15b) by replacing the marginal mean of dominant height with the conditional mean and \hat{v}_{11} with \hat{v}_{11}^* .

$$\ln(B\hat{A} + 1) = 3.334907, \text{ so } B\hat{A} = 27.076 \text{ (m}^2/\text{ha)}$$

The variance of the prediction for $\ln(B\hat{A} + 1)$ is 0.008196244. The error due to dominant height prediction contributes only 11.35% and, unlike the other examples, is not the main contributor in this instance.

$$\text{The predicted total volume } \ln(\hat{V} + 1) = 5.444312, \text{ so } \hat{V} = 230.438 \text{ (m}^3/\text{ha})$$

With prediction variance for $\ln(\hat{V} + 1)$ is 0.01063261 . Errors due to dominant height prediction and basal area prediction contribute 2.99% and 85.98%, respectively.

Example 6. The same information as in Example 2 is available with the addition of the prior measurements on dominant height.

In this case, dominant height at age 20 was observed, 19.6 m, say. Obviously, it can be combined with previous observations to estimate the random effects. So the number of repeated measurements is $m = 4$ in this case. Using (5.20a), the BLUP's for the random effects are:

$$\hat{b}_m = \begin{pmatrix} 0.092020247 \\ 0.006088786 \\ -0.034353997 \end{pmatrix} \text{ [by(5.20a)]}, \quad HD[t_m | \hat{b}_m] = \begin{pmatrix} 12.22237 \\ 15.12482 \\ 17.54725 \\ 19.53333 \end{pmatrix} \text{ [by(5.20e)]}$$

$$\hat{U}_m \text{ is updated to: } \hat{U}_m = \begin{pmatrix} 0.1463035425 & -9.281586e-004 & 0.0098883631 \\ -0.0009281586 & 8.346899e-006 & -0.0000726324 \\ 0.0098883631 & -7.263240e-005 & 0.0008141629 \end{pmatrix}$$

$\hat{v}_{11}^* = 0.08910379$, which is smaller than that in example 5.

The basal area and volume predictions based on the updated dominant height estimated and \hat{v}_{11}^* are as the follows:

$$\ln(B\hat{A} + 1) = 3.347833, \text{ so } B\hat{A} = 27.441 \text{ (m}^2/\text{ha)}$$

$$\text{Variance}[\ln(B\hat{A} + 1)] = 0.006671673$$

$$\ln(\hat{V} + 1) = 5.465995, \text{ so } \hat{V} = 235.511 \text{ (m}^3/\text{ha)}$$

$$\text{Variance}[\ln(\hat{V} + 1)] = 0.008525947$$

Example 7. The same information as in Example 3 is available with the addition of the prior measurements on dominant height.

In this case, basal area at age 20 was observed, say $25 (m^2/ha)$. So, the random effects and \hat{v}_{11}^* estimated in Example 5 are applicable again in the calculations. We must predict dominant height and volume at age 20 with (5.18a)-(5.19c).

$$H\hat{D}[t_k = 20 | \hat{b}_m, BA(t_k = 20)] = 19.418 \text{ m with prediction variance } 0.1859669,$$

i.e. the standard error of the prediction is 0.43 m .

The predicted total volume $\ln(\hat{V} + 1) = 5.362409$, so $\hat{V} = 212.238 (m^3/ha)$ with prediction variance of 0.001173914 for $\ln(\hat{V} + 1)$.

Example 8. The same information as in Example 4 is available with the addition of the prior measurements on dominant height.

In addition to the three repeated prior measurements of dominant height, both dominant height and basal area at age 20 are observed, say 19.6 m and $25 (m^2/ha)$ respectively. So we can use the height observations up to age 20 to estimate the random effects. This procedure is the same as that in Example 6. Then the updated \hat{v}_{11}^* , observed dominant height, and basal area can be applied in total volume prediction using [(5.12a)-(5.12d)].

The predicted total volume $\ln(\hat{V} + 1) = 5.303519$, so $\hat{V} = 200.043 (m^3/ha)$. With prediction variance is 0.0008542908 for $\ln(\hat{V} + 1)$.

In summarization of the examples it is interesting to compare “bound on error” defined as half the approximate 95% confidence interval divided by the predicted value for volume and expressed as a percent (see Table 21 and Figure 39). For the case with only the exogenous variables known at age 20 (Example 1), the bound is 37.3%. By including the information on prior heights at ages 11, 14, and 17 (Example 5), this bound

is reduced to 20.4%. In addition, if observed values for basal area and dominant height are also available at age 20 (Example 8), the bound is reduced to 5.8%.

The precision of dominant height prediction mainly depends on the repeated measurement factor (Figure 40) and dominant height itself is the main factor for precise basal area predicting (Figure 41). Basal area and the repeated measurement factor in dominant height both heavily affect total stem volume prediction (Figure 42).

Table 21 Prediction values and confidence intervals with the simultaneous system based on 8 examples.

Available Information	Dominant Height		Basal Area		Total Volume	
	Prediction (m)	95%CI.*	Prediction ($m^2/ha.$)	95% CI.	Prediction ($m^3/ha.$)	95% CI.
Exp1 exogenous variables only	19.18 (1.85)	[15.55, 22.81]	3.3184 (0.1558)	26.62 [20.35, 36.48]	5.4170 (0.1854)	224.2 [155.6, 322.8]
Exp2 exp1+HD	-	-	3.3476 (0.0851)	27.43 [24.06, 33.60]	5.4652 (0.09606)	235.3 [194.8, 284.3]
Exp3 exp1+BA	19.18 (1.85)	[15.55, 22.81]	-	-	-	5.3528 (0.08295)
Exp4 exp1 +HD+BA	-	-	-	-	-	5.3663 (0.03868)
Exp5 exp1 +repeated	19.42 (0.439)	[18.56, 20.28]	3.3349 (0.09053)	27.08 [22.51, 32.53]	5.4443 (0.10311)	230.4 [188.1, 282.3]
Exp6 exp2 +repeated	-	-	3.3478 (0.08168)	27.44 [23.23, 32.38]	5.4660 (0.09234)	235.5 [196.4, 282.4]
Exp7 exp3 +repeated	19.42 (0.431)	[18.57, 20.26]	-	-	-	5.3624 (0.03426)
Exp8 exp4 - +repeated	-	-	-	-	-	5.3035 (0.02923)

*95%CI. is the approximated confidence intervals at $\alpha=95\%$ level, CI. = Mean \pm $1.96 \times \text{std}$. For basal area and total volume, the CI's are obtained by transforming back from the CI. in logarithm.

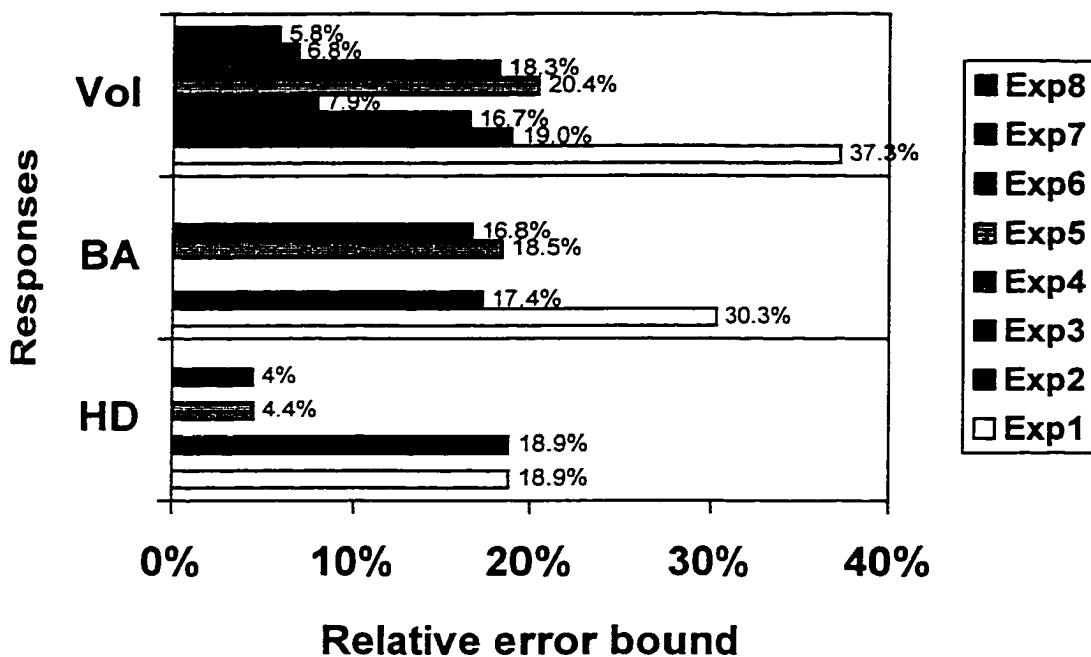


Figure 39. A comparison of "Bound on Error" for different prediction situations (8 examples) in the simultaneous growth and yield model system. "Bound on Error" is defined as half the approximate 95% confidence interval divided by the predicted value.

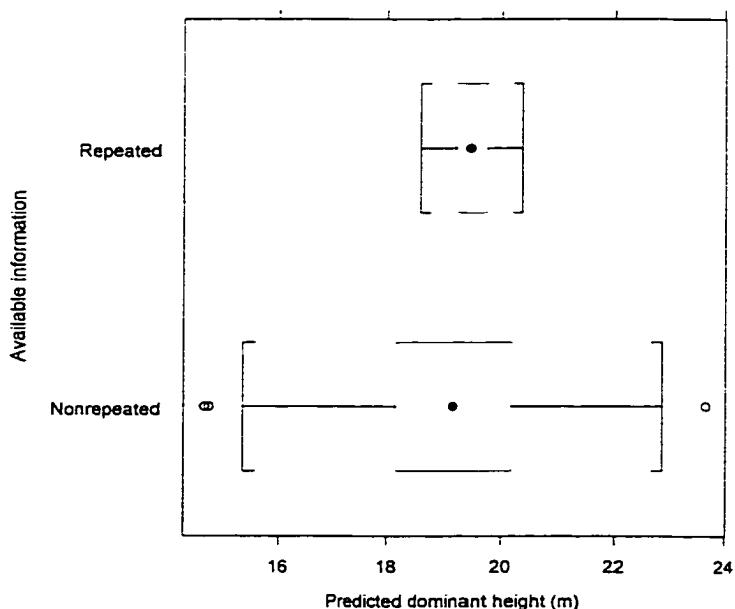


Figure 40. A comparison of the predictions and corresponding confidence intervals for slash pine dominant height in the simultaneous growth and yield model system with or without the information on repeated measurements.

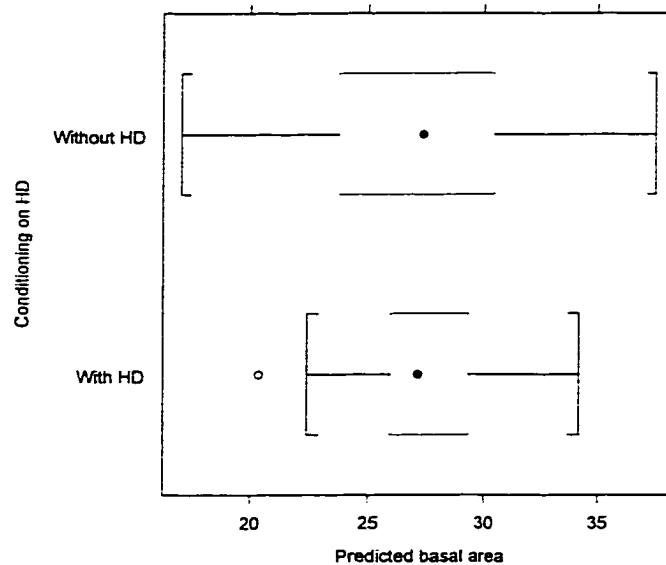


Figure 41. A comparison of the predictions and corresponding confidence intervals for slash pine basal area (in logarithm) in the simultaneous growth and yield model system with or without the contemporary dominant height measured.

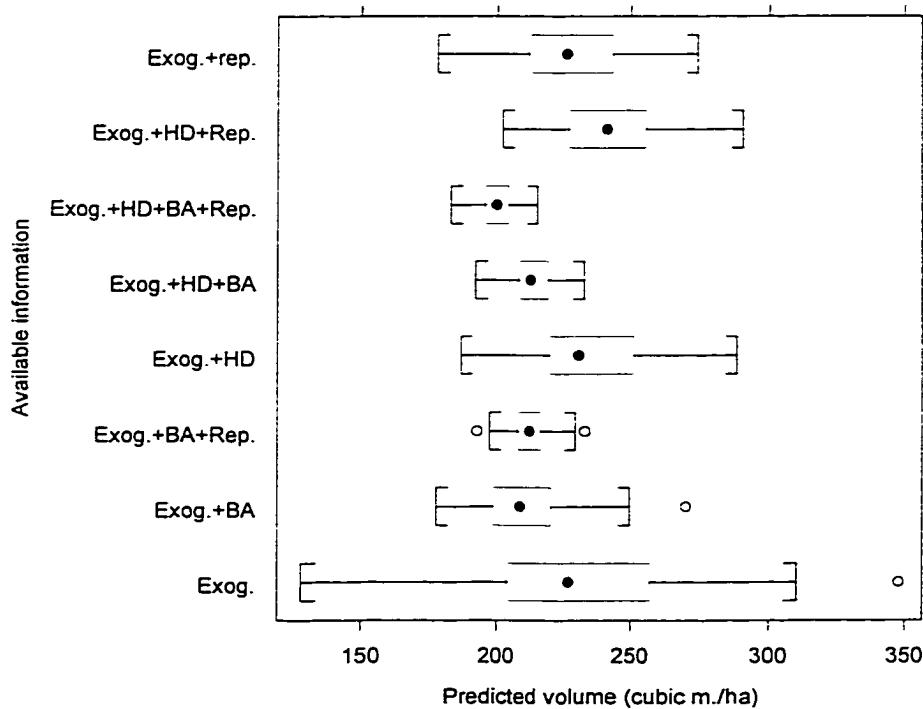


Figure 42. A comparison of the predictions and corresponding confidence intervals for slash pine total stem volume (in logarithm) in the simultaneous growth and yield model system based on 8 distinct situations with different information available.

Discussion

Two basic sources of variation in stand growth and yield are very common for survey data, such as those from permanent plots in a forest inventory. One is the within-plot error and the other is the variation from plot to plot. Within-plot variation is usually modeled by a reasonable variance function that accounts for within-plot heteroscedasticity and correlation. The between plot variation can be modeled by random effects that allow the parameters in the model to be varied from plot to plot. The more information on these sources of variation available, the more precise the predictions of the stand characteristics for a new observation. This is the intuitive logic behind the improvements in prediction of stand growth and yield for a new plot based on the simultaneous system with random effects. The interdependency among the components in the system (dominant height, basal area and volume) is another key to the prediction in that the observed components in the system can be used to improve the prediction of the unobserved components by application of the contemporaneous correlation among the components.

Dominant height is the fundamental component in the simultaneous system. This being that dominant height is a predictor in both the basal area model and the volume model. In addition, the prediction error for dominant height is one of the main sources of error in both basal area and total volume predictions when observed dominant height is unavailable. Thus, precise prediction of dominant height is critically important in the simultaneous system. For example, comparing *Example 1* and *Example 2*, with dominant height observations (so error of dominant height is 0), the variances of prediction error for basal area and volume decrease from 0.02426 to 0.007243, and 0.03436 to 0.009227,

respectively. When current dominant height observation is unavailable but previous observations for dominant height are available, the variance of prediction error for basal area and total volume still decrease from 0.02426 to 0.008196, and 0.03436 to 0.01063, respectively (see *Example 5*). This is because the estimated random effects from previously observed dominant heights dramatically decreased the prediction error for dominant height (from 1.85 m to 0.44 m). When observed basal area is unavailable, the prediction error of basal area is the main source of the error in volume prediction (see *Examples 2 and 5*). So, the three components in the simultaneous system form a pyramid with dominant height at the base and volume at the top. Once the random effects in dominant height and the contemporaneous correlations among the components are reasonable modeled, accounting for random effects in basal area and total volume have been proved to be unnecessary. Formal statistical criteria, such as *AIC* and the likelihood ratio test, have been used in the justification of these results. Intuitively, this makes logical sense in that dominant height is a predictor in both the basal area model and the volume model. Random effects are reasonably included in the dominant height model. Therefore, it is no surprise that random effects in the basal area and volume models are not significant any more.

From the error structure of dominant height, random effects (between-plot variation) is the main source of error. For example, in the marginal mean response random effects account for 97.4% of the variation in the total error (see *Example 1*). For simplicity, within-plot correlation for the same component of the response in the system has not be included in the model. This does not mean that autocorrelation is generally not important. Actually, the random effects are often able to represent the serial correlation

(Jones 1990). Random effects and autocorrelation can compete with each other in covariance modeling (see previous chapters). In order to avoid this complexity, modeling serial correlation has given way to approaches using random effects.

CHAPTER 6

SUMMARY AND CONCLUSIONS

For longitudinal data, two basic sources of variation are often presented in growth and yield models. One is within observational units and the other is between observational units. With these data it is clear that between-plot variability dominates the uncertainty in predictions of slash pine growth and yield for intensively site-prepared plantations. Therefore, the mixed-effects model is intuitively appealing in that the parameters in the model are allowed to be varied from plot to plot by mixed effects model approach. Mixed effects model presents a lot of flexibility in tracking the variation among grouping factors, such as silvicultural treatment groups, in the original experimental design. This results in a more precise model than can be developed by the contemporary approaches based on traditional regression.

A nonlinear mixed-effects model and a linear multi-level mixed-effects model were developed for slash pine dominant height and basal area growth in conjunction with different silvicultural treatments. The mixed-effects model is extended to a simultaneous growth model system including the components of dominant height, basal area and total volume. Predictions for new cases (i.e. plots) is one of the main focuses of the study. The advantages of a mixed-effects model and the simultaneous system in prediction are illustrated with detailed formulas and specific examples. As a comparison, a Bayesian

model formulation with a Gibbs sampling estimation process are presented for the nonlinear mixed-effects dominant height growth model.

The general techniques illustrated and applied include:

- a. Model building for a nonlinear mixed-effects model and for a linear multi-level mixed-effects model,
- b. Model building in a Bayesian framework with a Gibbs sampler for estimation,
- c. Predictions of variable values and determination of variance estimates in different situations,
- d. Development of a multi-variate simultaneous model system with fixed and random effects.

These are all relevant to situations beyond the primary focus of the work, that being forest growth and yield modeling. Potential applications to longitudinal data or other grouped data in other fields, such as medical trials, are highly likely.

Specific results, findings and conclusions obtained in the current work for the slash pine growth data include:

- ① Between-plot variation is the main source of error for dominant height growth prediction.
- ② Within-plot heterogeneity and correlation still exist even after introducing random components into all three parameters in a modified Richard's dominant height growth equation. A power variance function and moving average correlation

with order 2 [$MV(2)$] are justified for the within-plot error structure in a conventional nonlinear mixed-effects model approach.

③ Predictions for new observations can be obtained at different precision levels in distinct situations. Past observations can be used to estimate the individual-specific parameters and thus improve the precision of the prediction for the specific individual experimental unit (i.e. plot).

④ The multi-level mixed effects model is more appropriate for a design with nested factors, such as a split-plot. In the basal area growth example, the installation area and silvicultural treatment within installation are the most important sources of variation in the model. An exponential variance function and autoregression correlation with order 1 [$AR(1)$] is justified for the with-in plot covariance structure.

⑤ Predictions for a new observation in the multi-level mixed effects model can be obtained at different nesting levels. Generally speaking, the more specific information available, the more precise the prediction obtained. When observations recorded at prior times on the same unit are available, prediction for an individual unit with a nesting factor known should be based on the *BLUP*'s from the estimates based on these prior observations for this specific individual instead of *BLUP*'s of random effects based on the original data used to fit the model.

⑥ The silvicultural treatments tested are significant to slash pine plantation growth. For the nonlinear mixed-effects dominant height growth model, more precise analysis of the effects due to silvicultural treatment were obtained with the Bayesian formulation and a Gibbs sampling estimation method.

⑦ Accounting for both contemporaneous correlation and random effects improves the precision of the prediction of new observations in a simultaneous growth and yield model system.

⑧ In a simultaneous system containing models for dominant height, basal area and total volume the error in dominant height prediction is the main contributor to error for the basal area prediction. Moreover, the error in basal area prediction is the main error source for the total volume prediction.

⑨ Dominant height is the fundamental component in the three-component system. With contemporaneous correlation and random effects for dominant height considered, including random effects in basal area and total volume model proved to be unnecessary.

⑩ Prediction for a new observation is illustrated with several different examples. Prediction variance and thus confidence intervals of the prediction are also available for each case illustrated.

LITERATURE CITED

- Amemiya, T. 1985. Advanced econometrics. Cambridge, MA: Harvard University Press.
- Avery, T. E., and H.E. Burkhart. 1994. Forest measurements. McGraw-Hill, New York.
- Bailey, R.L., and J.L. Clutter. 1974. Base-age invariant polymorphic site curves. For. Sci. 20:155-159.
- Bailey, R.L., T.M. Burgan, and E. J. Jokela. 1989. Fertilized midrotation-aged slash pine plantations-stand structure and yield prediction models. South. J. Appl. For. 13(2):76-80.
- Bates, D. M. and J. C. Pinheiro 1999. Mixed effects models in S. Springer, New York.
- Bondeson, J. 1990. Prediction in random coefficient regression models. Biom. J. 32: 387-405.
- Bondeson, J. and J. Lanke. 1990. A medical application of the general random coefficient regression model. Biom. J. 32: 407-426.
- Borders, B. E., R.L. Bailey, and J.D. Ware. 1984. Slash pine site-index from a polymorphic model by joining (splining) nonpolynomial segments with an algebraic difference method. For. Sci. 30:411-423.
- Borders E. B. 1989. System of equations in forest stand modelling. For. Sci. 35:548-556.
- Borders E. B., R. L. Bailey, 1986. A compatible system of growth and yield equations for Slash pine fitted with restricted three-stage least squares. For. Sci. 32:185-201.
- Brooks, S. P., and A. Gelman 1998. Alternative methods for monitoring convergence of iterative simulations. J. of Comp. and Graph. Stat. 7:434-455.
- Bryk, A. , and S. Raudenbush 1992. Hierarchical linear models for social and behavioral research. Sage, Newbury Park.

- Burkhart, H. E. and P. E. Sprinz 1984. Compatible cubic volume and basal area projection equations for thinned old-field loblolly pine plantations. *For. Sci.* 30:86-93.
- Box, G.E.P., G.M. Jenkins, and G.C. Reinsel. 1994. Time series analysis: forecasting and control, 3rd edn. Holden-Day, San Francisco.
- Cao, Q.V. 1993. Estimating coefficients of Base-age-invariant site index equations. *Can. J. For. Res.* 23:2343-2347.
- Carter, C. K. and R. Kohn 1994. On Gibbs sampling for state space models. *Biometrika* 81(3): 541-553.
- Casella, G. and E. I. George. 1992. Explaining the Gibbs Sampler. *The Amer. Statistician* 46(3):167-173.
- Castleberry, J. P. 1998. Evaluation of models developed on a designed study data set applied to operational stands. The university of Georgia, M.S. thesis.
- Chi, E.M., and G.C. Reinsel. 1989. Models for longitudinal data with random effects and AR(1) errors. *J. of the Amer. Stat. Assoc.* 84:452-459.
- Cieszewski, C.J., and I.E. Bella. 1989. Polymorphic height and site index curves for lodgepole pine in Alberta. *Can. J. For. Res.* 19:1151-1160.
- Cieszewski, C. J. and R. L. Bailey, 1999, Generalized algebraic difference approach: derivation of base-age invariant polymorphic site equations with variable asymptotes. Submitted to Forest Science
- Clutter, J. L. 1961, The development of compatible analytic models for growth and yield of Loblolly pine. Duke university, PhD dissertation
- Clutter J. L. 1963. Compatible growth and yield models for loblolly pine. *For. Sci.* 9(3):354-370.
- Clutter J. L., J.C. Fortson, L.V. Pienaar, G. H. Brister, and R. L. Bailey 1992. Timber Management: a Quantitative approach. Krieger, Malabar Florida. 333p.
- Crowder, M.J. and D. J. Hand 1990. Analysis of repeated measures. Chapman & Hall, New York.
- Curtis, R.O., D.J. Demars, and F. R. Herman. 1974. Which dependent varaiable in site index-height-age regressions? *For. Sci.* 20:74-87.
- Davidian, M. and D. M. Giltinan 1995. Nonlinear models for repeated measurement data. London: Chapman and Hall.

- Diem, J.E., and J.R. Liukkonen. 1988. A comparative study of three methods for analysis longitudinal pulmonary function data. *Statistics in Medicine* 7:19-28
- Diggle, P.J., K. Y. Liang and S. L. Zeger 1994. Analysis of Longitudinal data. Oxford University Press, Oxford.
- Diggle, P. J. and A. P. Verbyla 1998. Nonparametric estimation of covariance structure in longitudinal data. *Biometrics* 54(2): 401-415.
- Fang, Z. and R. L. Bailey 1999. Compatible volume and taper models with coefficients for tropical species on Hainan island in Southern China. *For. Sci.* 45(1) 85-100.
- Fang, Z., B. E. Borders and R. L. Bailey 1999, Compatible volume-taper models for loblolly pine and slash pine based on a system with segmented-stem factors. *For. Sci. in press.*
- Flewelling, J. W., and R.D. Jong, 1994. Considerations in simultaneous curve fitting for repeated height-diameter measurements. *Can. J. For. Res.* 24: 1408-1414.
- Flewelling, J.W., and L.V. Pienaar. 1981. Multiplicative regression with lognormal error. *For. Sci.* 27:281-289.
- Furnival, G. M. and R. W. Wilson, JR. 1971. System of equations for predicting forest growth and yield. *Statist. Ecol.* 3: 43-55.
- Geisser, S. 1981. Sample reuse procedures for prediction of unobserved portion of a partially observed vector. *Biometrika* 68:243-250.
- Geisser, S. 1993. Predictive inference: an introduction. New York: Chapman & Hall. 264p.
- Geman, A. and D. Geman 1984. Stochastic relaxation, Gibbs distributions and Bayesian restoration of images. *IEEE Trans. Pattn. Anal. Mach. Intell.* 6: 721-741.
- Gelfand, A. E. and A. F. M., Smith 1990. Sampling based approaches to calculating marginal densities. *J. of Amer. Statist. Assoc.* 85: 398-409.
- Gelman, A., and D. Rubin. 1992. Inference from iterative simulation using multiple sequences. *Stat. Sci.* 7: 457-511.
- Gilks, W. R., C. C. Wang, B. Yvonnet and P. Coursaget 1993. Random-effects models for longitudinal data using Gibbs sampling. *Biometrics* 49: 441-453.
- Gilks, W.R., S. Richardson, and D.J. Spiegelhalter. 1996. Markov chain Monte Carlo in practice. Chapman & Hall.

- Goldstein, H. 1986. Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika* 73(1): 43-56.
- Goldstein, H. 1995. Multilevel statistical models. Halstead Press, New York.
- Godambe, H. 1991. Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika* 73(1): 43-56.
- Green, J. E. and H. T. Valentine 1998. Bayesian analysis of the linear model with heterogeneous variance. *For. Sci.* 44(1) 134-138.
- Green, J. E., F.A. Roesch, A.F.M. Smith, and W.E. Strawderman. 1994. Bayesian estimation for the three-parameter Weibull distribution with tree diameter data. *Biometrics* 50:254-269
- Gregoire, T. G., O. Schabenberger 1989. Model fitting under patterned heterogeneity of variance. *For. Sci.* 35: 105-125.
- Gregoire, T. G. 1987. Generalized error structure for forestry yield models. *For. Sci.* 33: 423-444.
- Gregoire, T. G., O. Schabenberger and J. P. Barrett. 1995. Linear modelling of irregularly spaced, unbalanced, longitudinal data from permanent-plot measurements. *Can. J. For. Res.* 25: 137-156.
- Gumpertz, M.L., and C. Brownie. 1993. Repeated measures in randomized block and split-plot experiments. *Can. J. For. Res.* 23: 625-639.
- Hall, B. D. and R. L. Bailey 1997. on the use of nonlinear multilevel mixed models to predict site productivity in forest stands. Unpublished report. 12p.
- Harville, D. 1976. Extension of the Gauss-Markov theorem to include the estimation of random effects. *The Annals of Statistics* 2: 384-395.
- Hocking, R. R. 1996. Models and Applications of linear models: regression and analysis of variances. J. Wiley, New York. 731p.
- Hasenauer, H., R.A. Monserud, and T.G. Gregoire. 1997. Using simultaneous regression techniques with individual-tree growth models. *For. Sci.* 44(1): 87-95.
- Hastings, W.K. 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57:97-109.
- Jones, R.H. 1990. Serial correlation or random subject effects? *Commu. in Stat. A* 19: 1105-1123.

- Jones, R. H. 1993. Longitudinal data with serial correlation: A state space approach. London: Chapman Hall.
- Karlsson, M. O. and L. B. Sheiner 1994. The importance of modeling inter-occasion variation in population pharmacokinetics. *J. of Pharmacokinetics and Biopharmaceutics*. 22: 735-750.
- Laird, N.M. and J. H. Ware 1982, Random Effects Models for Longitudinal Data. *Biometrics* 38:963-974.
- 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.* 38(2): 409-429.
- Lappi, J. and J. Malinen 1994. Random-parameter height/age models when stand parameters and stand age are correlated. *For. Sci.* 40: 715-731.
- Leak, W. 1966. Analysis of multiple systematic remeasurements. *For. Sci.* 12:69-73.
- Lee, C. J. 1988. Prediction and estimation of growth curves with special covariance structures. *J. of Amer. Statist. Assoc.* 83:432-440.
- Lee, J. C. and S. Geisser 1975. Applications of growth curves prediction. *Sankhya, Series A* 37: 239-256.
- LeMay, M. V. 1990. MSLS: a linear least squares technique for fitting a simultaneous system of equations with a generalized error structure. *Can. J. For. Res.* 20: 1830-1839.
- Lentner, M., and T. Bishop 1986. Experimental design and analysis. Valley Book Co. Blacksburg, VA.
- Lindsey, J. K. 1993. Models for repeated measurements. Clarendon press, Oxford. 413p.
- Lindsey, J.K. 1997. Applying generalized linear models. New York: Springer. 256p.
- Lindstrom, J. M. and D. M. Bates. 1990. Nonlinear mixed effects models for repeated measures data. *Biometrics* 46: 673-687.
- Liski, E. P. and T. Nummi 1990. Prediction in growth curve models using the EM algorithm. *Computational Statistics & Data Analysis* 10: 99-108.
- Liski, E. P. and T. Nummi 1995. Prediction of tree stems to improve efficiency in automatized harvesting of Forests. *Scandinavian Journal of Statistics* 22: 255-269.

- Liski, E. P. and T. Nummi 1996. Prediction in repeated-measures models with engineering applications. *Technometrics* 38(1):25-36.
- Longford, N. T. 1993. Random coefficient models. Oxford: Oxford University Press.
- Maitre, P.o., M. Buhrer, D. Thomson, and D.R. Stanski. 1991. A three-step approach combining Bayesian regression and NONMEM population analysis: application to midazolam. *J. of Pharmacokinetics and Biopharmaceutics* 19: 377-384.
- Martin, S. W., R. L. Bailey, and E. J. Jokela. 1998. Models for unfertilized and fertilized slash pine plantations: CRIFF B400-500 series. *South J. Appl. For.*
- Metropolis, N, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller 1953. Equations of state calculations by fast computing machine. *J. Chem. Phys.* 21: 1087-1091.
- Miller,D.M. 1984. Reducing transformation bias in curve fitting. *Am. stat.* 38:124-126
- Murphy, P. A. and H. S. Sternizke 1979. Growth and yield estimation for loblolly pine in the West Gulf. *USDA For. Serv. Res. Pap. So-154.* 8p.
- Northway, S.M. 1985. Fitting site index equations and other self-inferencing functions. *For. Sci.* 31:233-235.
- Omule, S.A.Y., and R.N. MacDonald, 1991. Simultaneous curve fitting for repeated height-diameter measurements. *Can. J. For. Res.* 21: 1418-1422.
- Pienaar, L.V., and J. W. Rheney. 1995. Modeling stand level growth and yield response to silvicultural treatments. *For. Sci.* 41(3) 629-638.
- Pienaar, L.V., B.D. Shiver, and W. M. Harrison. 1998. Slash pine site preparation study: age 17 results. *Sch. Forest Resour., Univ. Ga., Athens, PMRC Tech. Rep.* 1998-4.
- Pinheiro, J.C. and D.M. Bates 1995. Approximation to the log-likelihood function in the nonlinear mixed effects model. *Journal of Computational and Graphical Statistics* 4: 12-35.
- Pinheiro, J.C. and D.M. Bates 1998. Model building for nonlinear mixed effects models. Technical report, Department of statistics, University of Wisconsin.
- Rao, C. R. 1976. Prediction of future observations with special reference to linear models. In *Multivariate Analysis IV* (P. R. Krishnaiah, ed.) 193-208. North-Holland, Amsterdam.

- Rao, C. R. 1987. Prediction of future observations in growth curve models. *Statist. Sci.* 2(4): 434-471.
- Rao, C. R. and R. Boudreau 1985. Prediction of future observations in factor analytic type growth model. In *Multivariate Analysis VI* (P. R. Krishnaiah, ed.) 449-466. North-Holland, Amsterdam.
- Ratkowsky, D. A. 1990. *Handbook of nonlinear regression models*. Marcel Dekker, New York.
- Reed, D. D. 1982. Simultaneous estimation of tree taper and merchantable volume in loblolly pine. *Va. J. Sci.* 33:85 (abst.)
- Richards, F.J. 1959. A flexible growth function for empirical use. *J.Exp.Bot.* 10:290-300.
- Roberts, G.O. 1996. Markov chain concepts related to sampling algorithms. In W.R. Gilks, S. Richardson and D. J. Spiegelhalter (Eds) *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London.
- Robinson, G. K. 1991. That BLUP is a good thing: the estimation of random effects. *Statist. Sci.* 6(1): 15-51. (with discussions)
- Rochon, J. 1992. ARMA covariance structures with time heteroscedasticity for repeated measures experiments. *J. of the Amer. Stat. Assoc.* 87:777-784.
- Schnute, J. 1981. A versatile growth model with statistically stable parameters. *Can. J. Fish. Aquat. Sci.* 38(9):1128-1140.
- Shiver, B.D., L.V. Pienarr, K.L. Hitch and J.W. Rheney. 1994. Slash pine site preparation study: age 14 results. Sch. Forest Resour., Univ. Ga., Athens, PMRC Tech. Rep. 1994-2.
- Smith, A.F.M., and G.O. Roberts 1993. Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *J. R. Statist. Soc. B* 55(1):3-23.
- Spiegelhalter, D.J., A. Thomas, N. Best, and W. Gilks. 1996. BUGS 0.5: Bayesian Inference Using Gibbs Sampling Manual (version *ii*). MRC Biostatistics Unit, Cambridge.
- Spiegelhalter, D.J., A. Thomas, N.G. Best. 1995. Computing on Bayesian graphical models. In *Bayesian Statistics 5*. Clarendon Press, Oxford, UK.
- Srivastava, V. K. and D. E. A. Giles. 1987. *Seemingly unrelated regression equations models*. New York: Dekker.

- Sullivan, A. D. and J. L. Clutter 1972. A simultaneous Growth and yield model for loblolly pine. *For. Sci.* 18(1):76-86.
- Tait D.E., C.J. Cieszewski, I. E. Bella. 1988. The stand dynamics of lodgepole pine. *Can. J. For. Res.* 18(10) 1255-1260.
- Tanner, A. M. 1996, Tools for statistical inference: method for the exploration of posterior distributions and likelihood functions. Springer-Verlag New York.
- Tasissa, G. and H.E. Burkhart. 1998. Thinning effects on ring specific gravity of loblolly pine (*Pinus taeda* L.). *For. Sci.* 44(2):212-223.
- Valliant, R. 1985. Nonlinear prediction theory and the estimation of proportions in a finite population. *J. of the Amer. Statist. Assoc.* 80(391): 631-641.
- Verbeke, G., and G. Molenberghs. 1997. Linear mixed effects model in practice: a SAS-oriented approach. Springer-verlag, New York.
- Vonesh, E.F. and R. L. Carter 1992. Mixed-effects nonlinear regression for unbalanced repeated measures. *Biometrics* 48(1):1-17.
- Vonesh, E. F. 1996. A note on the use of Laplace's approximation for nonlinear mixed-effects models. *Biometrika* 83: 447-452.
- Vonesh, E.F., and V.M. Chinchilli, 1997. Linear and nonlinear models for the analysis of repeated measurements. Marcel Dekker, New York.
- Wakefield J. 1996. The Bayesian analysis of population pharmacokinetic models. *J. of the Amer. Statist. Assoc.* 91(433): 62-75.
- Wakefield, J. C., A. F. M. Smith, A. Racine-Poon, and A. E. Gelfand. 1994. Bayesian analysis of linear and non-linear population models by using Gibbs sampler. *Appl. Statist.* 43(1): 201-221.
- Wolfinger, R. 1993. Laplace's approximation for nonlinear mixed models. *Biometrika* 80: 791-795.
- Zeger, S.L., K. Liang, and P. Albert. 1988. Models for longitudinal data: a generalized estimating equation approach. *Biometrics*. 44: 1049-1060.
- Zeger, L. S. and M. R. Karim 1991. Generalized linear models with random effects: A Gibbs sampling approach. *J. of the Amer. Statist Assoc.* 86(413): 79-86.
- Zhang, Y., and B. E. Borders. 1999. Evaluating basal area growth response to midrotation fertilization. Submitted to *Forest Ecology and Management*.