MODELING *Eucalyptus* STAND GROWTH BASED ON LINEAR AND NONLINEAR MIXED-EFFECTS MODELS

by

NATALINO CALEGARIO

(Under the direction of RICHARD F. DANIELS)

ABSTRACT

Single and multilevel linear and nonlinear mixed-effects model approaches were applied to model *Eucalyptus* stand growth. Most of the data set had a longitudinal, unequally spaced and unbalanced structure. By modeling heteroscedasticity, autocorrelation and correlation among random effects, the precision of these models in explaining data variation was increased. This approach has flexibility in selecting fixed effects which will be significantly associated with random effects, for both linear and nonlinear cases. In the linear case, the logarithm of basal area was linearly associated with some stand covariates. In the nonlinear case the three parameter logistic equation was used to explain dominant height variation over age.

A tree profile equation was developed based on a four parameter logistic equation, including covariates associated with random and fixed effects. The solid of revolution technique was applied to obtain individual tree volume having the individual tree profile. The robustness property of the mixed-effects model was used to estimate tree profiles with information on dbh and total height only. Using this technology, individual tree volume can be estimated with high precision with minimal information.

INDEX WORDS: *Eucalyptus*, Linear mixed-effects models, Nonlinear mixed-effects models, Multilevel models, Longitudinal data, Heteroscedasticity, Individual tree volume estimation, Taper function

MODELING *Eucalyptus* STAND GROWTH BASED ON LINEAR AND NONLINEAR MIXED-EFFECTS MODELS

by

NATALINO CALEGARIO

B.S., Federal University of Viçosa, Brazil, 1985

M.S., Federal University of Viçosa, Brazil, 1992

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

2002

MODELING *Eucalyptus* STAND GROWTH BASED ON LINEAR AND NONLINEAR MIXED-EFFECTS MODELS

by

NATALINO CALEGARIO

Approved:

Major Professor: Richard F. Daniels

Committee: Robert L. Bailey
Barry D. Shiver
Michael L. Clutter
Jaxk H. Reeves

Electronic Version Approved:

Gordhan L. Patel
Dean of the Graduate School
The University of Georgia
May 2002

ACKNOWLEDGMENTS

My sincere gratitude goes to Dr. Richard Daniels, my major professor, who dedicated his time to make this work a reality.

Special thanks to Dr. Robert Bailey, who was my major professor for three years and was the first person to receive me in United States. Dr. Jaxk Reeves, my major professor in my master degree in statistics. Dr. Barry Shiver and Dr. Michael Clutter provided criticism and comments which helped shape this work. Thanks to friends at Warnell School of Forest Resources, my sponsor, CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), and the Forest Science Department, at Federal University of Lavras, Brazil.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The reliability of a forest growth and yield prediction and projection system, with response variables such as dominant height, basal area, trees per acre and volume, has been based on estimator characteristics such as consistency, efficiency and sufficiency. Forest biometricians have been developing and adapting statistical techniques to improve those characteristics and to provide such systems to meet particular objectives for forest management planning and decision making.

The multilevel model approach is a statistical technique that has been used in many fields of study, generating improvements in parameter estimation. Also referred to as multistage models, repeated measurements models, longitudinal data analysis, and mixed models, this type of approach has been developed intensively during the last 20 years and has been applied to forestry, agriculture, ecology, biomedicine, sociology, economics, and other areas. Authors such as Corbeil and Searle (1976), Dempster et al. (1984), Stiratelli (1984), Ware (1985), Goldstein (1986 and 1991), Zeger (1986), Carter and Yang(1986), Chi and Reinsel (1989), Gumpertz (1989), Crowder (1990), Breslow and Clayton (1993), Diggle et al. (1994), Davidian and Giltinan(1995), Burnett et al. (1995), Vonesh and Chinchilli (1997), Wolfinger (1993), Littell et al. (1996), and others have been developing basic and applied studies using multilevel models.

In forestry, studies using multilevel model approaches are relatively recent. As a pioneer study, Biging (1985) improved the estimates of site index curves using a varying-parameter model. In another innovative approach, Lappi and Bailey (1988) described the use of nonlinear mixed-effects growth curve, based on the Richards model, which was fitted to predict dominant and codominant tree height, both at the plot level and at the individual tree level. More recently, other studies based on random effect models have been published in forestry. Studies such as the Kalman filter approach to localizing height-age equations (Walters et al., 1991); linear mixed-effects modeling of the covariance among repeated measurements with random plot effects (Gregoire et al., 1993); bole-volume equations to spatially correlated within-tree data (Gregoire and Schabenberger, 1995); estimating forest yield using functions with random effects (Candy, 1997); a simultaneous system of linear and nonlinear mixed models to predict forest growth and yield (Fang, 1999), and modeling forest growth and yield based on multilevel nonlinear mixed models (Hall and Bailey, 2001) can be cited as recent publications of multilevel models in forestry.

Eucalyptus can be considered, in a worldwide panorama, one of the most important cultivated forestry genra due to the large number of species, adaptation to different edaphic-biologic-climatic situations, and fast growth. Additionally, with the development of silvicultural techniques, intensive management strategies, and genetic improvement, the productivity of these plantations has been improving significantly, leading to different types of products. For modeling purposes, a wide range of studies have been developed to generate prediction and projection systems (Paula Neto, 1991; Campos, 1980 and 1983; Trevizol, 1985; Amaro, 1997; Diaz and Couto, 1999), but there

is no information on applying multilevel theory, including fixed and random effects in the previous studies. Most eucalypt plantations are from uniform genetic stock, having been propagated by asexual reproduction, so that the variation is generated primarily by environmental factors. In this situation, we have strong reasons to believe that by modeling longitudinal and spatial correlation structures, the parameter will have consistent estimates. Therefore, based on the idea of modeling time-within-individual correlation, spatial inter-individual correlation, and adding the fixed and random effects in the models we expect improvements in the quality of the estimates from these linear and nonlinear models.

The data set for this study is from clonal eucalyptus plantations located in the southeast Brazilian Atlantic Coastal Region with a high degree of variation in soil, precipitation, topographic characteristics, silvicultural treatment, and genetic material. The data set is from permanent plots with repeated measurement information from 2 years to 10 years of age. The primary purpose of this work is to evaluate the performance of multilevel linear and nonlinear models relative to previous modeling strategies. These models will include both fixed and random variance effects and include methodology accounting for serial and spatial correlation.

CHAPTER 2

THE DEVELOPMENT OF A LINEAR MIXED-EFFECTS MODEL APPLIED TO *Eucalyptus* PLANTATIONS

Introduction

Linear mixed-effects models have been used in different situations in recent years to model longitudinal, spatial, and spatio-temporal processes in several scientific fields, such as medicine (Verbeke and Lesaffre, 1977), biology (Christman and Jernigan, 1997), engineering (Pinheiro and Bates, 2000), agriculture (Littell et al., 1996), and others.

In forestry, Gregoire (1995) applied the linear mixed-effects model to model eastern white pine (*Pinus strobus* L.) and douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco) basal area growth patterns from permanent plots of irregularly spaced trees from an unbalanced and longitudinal data set. The model fitted shows marked improvement compared with models that do not account for the error structure. In a more recent study, Fang and Bailey (2000) applied this approach to model Slash Pine (*Pinus elliottii* Engelm.) basal area following intensive silvicultural treatments and tested 9 models with different variance and/or correlation structures.

In eucalypt plantations there is no record of using linear mixed-effects to model growth and yield patterns. Due to the genetic and environmental variation in this

cultivation, improvement would be expected in the fitting process. The main purpose of this study is to model the linear tendency between the natural logarithm of basal area as a function of the inverse of age, by plot and clone. Also, distributional assumptions will be examined and, if necessary, the within-subject variance and/or correlation pattern will be included in the estimation process.

Generalized Linear Mixed-Effects Model

The general parametric form presented here is based upon that in Laird and Ware (1982), cited by Davidian and Giltinan (1995), with some adaptations to forest growth and yield studies.

Suppose *m* plots are sampled from a forest population and these plots are measured repeatedly in time, for example, *t* times. If *t* is the same for each plot, we have a balanced data set, generating simple computational features and analysis, with *t* x *m* available values. But unbalanced data sets are more common in forest growth and yield studies. Thus, the number of repeated measurements over time will vary and t_i will represent the number of measurements for the *i*th plot. For example, if plot *i* is measured annually during *j* years, which is a classical situation in a eucalypts plantation, the value of $t_i=j$. In the case of a balanced data set, $t_1=t_2=\ldots=t_m=j$. Let $\mathbf{y}_i$ represents a response vector for the *i*th plot. So, $\mathbf{y}_i$ has dimension of (*j* x 1) and this situation can be modeled using a linear mixed-effects model (2.1).

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\varepsilon}_i \qquad (2.1)$$

The response variable $\mathbf{y}_i$ will be a vector with dimension (t_i x 1), the $\mathbf{X}_i$ will be a matrix (t_i x p): t_i rows with repeated measurements and p columns of covariates, including a column for the intercept. The $\boldsymbol{\beta}$ vector will have (p x 1) dimension representing the parameters of the p fixed effects. For random effects, Z_i is an (t_i x k) design matrix linking $\mathbf{y}_i$ to the random effects $\mathbf{b}_i$, which is up to (p x 1) vector representing the random parameter estimated for plot i. The vector $\mathbf{b}_i$ will have dimension up to (p x 1), representing the random effects of the parameter intercept and the **X** variables. It is common in the mixed-effects literature to represent the fixed effects with a Greek letter and the random effects with a Latin letter. So, in this case, the plot number i will have p fixed effects ($\boldsymbol{\beta}$) including the intercept and up to p mixed effects (**b**). If the total sample includes m plots, every plot will have the same values for fixed effects and possibly different values for the random effects.

In forest growth and yield, one classical linear model relates the logarithm of basal area (ln(BA)), as a response variable, to stand-level variables such as the inverse of age(1/A), logarithm of dominant height (ln(HD)), logarithm of number of trees (ln(N)) and the interactions among these covariates (equation (2.4)). If the plot i, for example, is measured annually for 5 years, the dependent variable $\mathbf{y_i}$=ln(BA) will be a vector of dimension (5 x 1), the $\mathbf{X_i}$ will be a matrix of (5 x p), and the $\boldsymbol{\beta}$ vector will have (p x 1) dimension. The value of p depends on the number of the significant covariates of the right-hand side of the equation plus an intercept.

In standard regression assumptions, $\varepsilon_i \sim N(0,\boldsymbol{\Sigma}_i)$, where $\boldsymbol{\Sigma}_i$ is the within-plot covariance matrix. If the observations are independent, $\boldsymbol{\Sigma}_i = \sigma^2 \mathbf{I}_{ti}$, where I_{ti} is the (t_i x t_i) identity matrix. In our example, if the 5 observations of the same plot through of time are

independent and with same variance, I_{ti} will be (5 x 5) matrix and $\mathbf{\Sigma}_i$ will be a diagonal with σ^2 . In practice, $\mathbf{\Sigma}_i$ has many variations. Conditional on $\mathbf{b}_i$, (2.1) implies

$$E(y_i|b_i) = \mathbf{X_i\beta} + \mathbf{Z_i b_i} \quad (2.2a)$$

$$Cov(y_i|b_i) = \mathbf{\Sigma}_i \quad (2.2b)$$

If the random effects vector $\mathbf{b}_i$ comes from a normal distribution with mean zero, dispersion matrix $\mathbf{\Psi}$ (k x k) and independent of each other and of $\boldsymbol{\varepsilon}_i$, the marginal mean and covariance of $\mathbf{y}_i$ is:

$$E(\mathbf{y}_i) = E\{E(\mathbf{y}_i|\mathbf{b}_i)\} = \mathbf{X_i\beta} \quad (2.2a)$$

$$Cov(\mathbf{y}_i) = E\{Cov(\mathbf{y}_i|\mathbf{b}_i)\} + Cov\{E(\mathbf{y}_i|\mathbf{b}_i)\} = \mathbf{\Sigma}_i + \mathbf{Z}_i\mathbf{\Psi}\mathbf{Z}^T_i = \mathbf{V}_i \quad (2.3b)$$

Expanding our example for two groups (clones=c), we will have 10 plots (i) in each group with 5 observation (j) over time for each plot and assuming the following model,

$$\ln(BA)_{cij} = (\beta_{co} + b_{coi}) + (\beta_{c1} + b_{c1i}) \times \frac{1}{Age_{cij}} + (\beta_{c2} + b_{c2i}) \times \ln(H_{cij}) + (\beta_{c3} + b_{c3i}) \times \frac{1}{Age_{cij}} \times \ln(H_{cij}) + \varepsilon_{cij} \quad (2.4)$$

with c=1…2; i=1…10; j=1…5, we can write the model in matrix form as follows:

$$\mathbf{Y}^T = (\mathbf{Y}^T_{1,1}, \ldots \mathbf{Y}^T_{1,10}, \mathbf{Y}^T_{2,1}, \mathbf{Y}^T_{2,10})$$

$$\mathbf{Y}^T_{c,i} = (Y_{c,i,1}, Y_{c,i,2}, Y_{c,i,3}, Y_{c,i,4}, Y_{c,i,5}).$$

So, $\mathbf{Y}^T$ will have dimension of (1 x 100).

$$\mathbf{X}^T = (\mathbf{X}^T_{1,1}| \ldots|\mathbf{X}^T_{1,10}| \mathbf{X}^T_{2,1}| \ldots|\mathbf{X}^T_{2,10}),$$

Where,

$$\mathbf{X}_{1,i} = \begin{bmatrix} 1 & 0 & 1/Age_{1,i,1} & 0 & \ln(H_{1,i,1}) & 0 & (1/Age_{1,i,1})\ln(H_{1,i,1}) & 0 \\ 1 & 0 & 1/Age_{1,i,2} & 0 & \ln(H_{1,i,2}) & 0 & (1/Age_{1,i,2})\ln(H_{1,i,1}) & 0 \\ 1 & 0 & 1/Age_{1,i,3} & 0 & \ln(H_{1,i,3}) & 0 & (1/Age_{1,i,3})\ln(H_{1,i,1}) & 0 \\ 1 & 0 & 1/Age_{1,i,4} & 0 & \ln(H_{1,i,4}) & 0 & (1/Age_{1,i,4})\ln(H_{1,i,1}) & 0 \\ 1 & 0 & 1/Age_{1,i,5} & 0 & \ln(H_{1,i,5}) & 0 & (1/Age_{1,i,5})\ln(H_{1,i,1}) & 0 \end{bmatrix}$$

$$\mathbf{X}_{2,i} = \begin{bmatrix} 0 & 1 & 0 & 1/Age_{2,i,1} & 0 & \ln(H_{2,i,1}) & 0 & (1/Age_{2,i,1})\ln(H_{2,i,1}) \\ 0 & 1 & 0 & 1/Age_{2,i,2} & 0 & \ln(H_{2,i,2}) & 0 & (1/Age_{2,i,2})\ln(H_{2,i,2}) \\ 0 & 1 & 0 & 1/Age_{2,i,3} & 0 & \ln(H_{2,i,3}) & 0 & (1/Age_{2,i,3})\ln(H_{2,i,3}) \\ 0 & 1 & 0 & 1/Age_{2,i,4} & 0 & \ln(H_{2,i,4}) & 0 & (1/Age_{2,i,4})\ln(H_{2,i,4}) \\ 0 & 1 & 0 & 1/Age_{2,i,5} & 0 & \ln(H_{2,i,5}) & 0 & (1/Age_{2,i,5})\ln(H_{2,i,5}) \end{bmatrix}$$

$$\boldsymbol{\beta} = (\beta_{10}, \beta_{20}, \beta_{11}, \beta_{21}, \beta_{12}, \beta_{22}, \beta_{13}, \beta_{23});$$

$$\mathbf{Z} = I_{20} \otimes \mathbf{1}_{5,}$$

Where I_n is the identity matrix of order n, $\otimes$ is the Kronecker direct product and $\mathbf{1}_n$ denotes (*n* x 1) vector with all entries equal to one.

Further,

$$\mathbf{b} = (b_1, b_2, \ldots, b_{10}, b_{11}, \ldots, b_{20})^T$$

and

$$\varepsilon^T = (\varepsilon_{1,1,1}, \varepsilon_{1,1,2}, \varepsilon_{1,1,3}, \ldots, \varepsilon_{2,10,5})$$

The errors are independent with variance σ^2 and variance-covariance matrix **Σ.** The random effects are independent with variance σ^2_b and **Ψ** representing its the variance-covariance matrix and **V** being the variance-covariance matrix for response variable **Y.** The following representations are useful for computational purposes:

$$\mathbf{\Psi} = \sigma^2_b I_{20},$$

$$\mathbf{\Sigma} = \sigma^2 I_{100},$$

$$\mathbf{V} = \mathbf{Z}\,\mathbf{\Psi}\,\mathbf{Z}^T + \mathbf{\Sigma} = \sigma^2_b(I_{20} \otimes J_5) + \sigma^2 (I_{20} \otimes I_5) = I_{20} \otimes (\sigma^2_b J_5 + \sigma^2 I_5)$$

Estimation of Fixed and Random Effects

In the last section, we found that the marginal values for $\mathbf{y}_i$ are normally distributed with mean $\mathbf{X_i\beta}$ and variance-covariance matrix $\mathbf{V_i} = \mathbf{\Sigma}_i + \mathbf{Z}_i\mathbf{\Psi}\mathbf{Z}^T_i$. Following Verbeck and Molenberghs (1997), let **α** denote the vector of all variance and covariance components present in $\mathbf{V_i}$, *i.e.*, **α** will have all different elements of **Ψ** and all parameters of $\mathbf{\Sigma}_i$. In our example, for the *i*th plot, **Ψ** is represented by 10 different parameters (4*(4+1)/2) and $\mathbf{\Sigma}_i$ by 5, taking into account a diagonal matrix. Letting $\mathbf{\theta} = (\beta^T, \alpha^T)^T$ be the vector of all parameters in the marginal model, the classical approach is to maximize the marginal likelihood function with respect to **θ**.

$$\mathbf{L_{ML}}(\theta) = \prod_{\mathbf{i}=1}^{\mathbf{N}} \left\{ (2\pi)^{-\mathbf{ni}/2} \mid \mathbf{V_i}(\alpha) \mid^{-0.5} \times \exp\left(-\frac{1}{2} (\mathbf{Y_i} - \mathbf{X_i}\beta)^{\mathbf{T}} \mathbf{V_i}^{-1}(\alpha)(\mathbf{Y_i} - \mathbf{X_i}\beta) \right) \right\} \qquad (2.5)$$

If **α** is known, the maximum likelihood estimator of **β**, obtained from maximizing (2.5) is given by

$$\hat{\boldsymbol{\beta}} = \left(\sum_{\mathbf{i}=1}^{\mathbf{N}} \mathbf{X_i^T V_i^{-1}}(\boldsymbol{\alpha})\mathbf{X_i}\right)^{-1} \sum_{\mathbf{i}=1}^{\mathbf{N}} \mathbf{X_i^T V_i^{-1}}(\boldsymbol{\alpha})\mathbf{y_i} \tag{2.6}$$

and its variance-covariance matrix is

$$\operatorname{var}(\hat{\boldsymbol{\beta}}) = \left(\sum_{\mathbf{i}=1}^{\mathbf{N}} \mathbf{X_i^T V_i^{-1}}(\boldsymbol{\alpha})\mathbf{X_i}\right)^{-1} \tag{2.7}$$

When **α** is unknown, but an estimate of **α** is available, $V_i^{-1}(\hat{\alpha})$ can be substituted for $V_i^{-1}(\alpha)$. For estimating **α**, maximum likelihood (ML) and restricted maximum likelihood (REML) methods are used. Details about these methods can be find in Searle, Casella, and McCulloch(1992), Davidian and Giltinan (1995) and Vonesh and Chinchilli(1997).

Using regression analysis when p=rank(**X**) ≤ 4, the maximum likelihood method generates smaller mean squared error for σ^2 than restricted maximum likelihood. The opposite is true if p>4 and n-p is large (Verbeck and Molenberghs, 1997). Also, restricted maximum likelihood estimation adjusts for loss of degrees of freedom due to estimating fixed effects. REML estimation can be viewed as estimating variance components based on residuals calculated after fitting fixed effects only (Davidian and Giltinan, 1995).

Since random effects are assumed to be random variables, it is common to estimate them by Bayesian techniques. The marginal distribution of $\mathbf{b_i}$ is multivariate normal with mean zero and covariance matrix **Ψ** and this distribution is referred to as the prior distribution of $\mathbf{b_i}$. After observed values for $\mathbf{y_i}$ have been collected, the posterior distribution of $\mathbf{b_i}$ can be calculated as:

$$\mathbf{f}(\mathbf{b}_i \mid \mathbf{y}_i) \equiv \mathbf{f}(\mathbf{b}_i \mid \mathbf{Y}_i = \mathbf{y}_i) = \frac{\mathbf{f}(\mathbf{y}_i \mid \mathbf{b}_i)\mathbf{f}(\mathbf{b}_i)}{\int \mathbf{f}(\mathbf{y}_i \mid \mathbf{b}_i)\mathbf{f}(\mathbf{b}_i)d\mathbf{b}_i} \tag{2.8}$$

The expression (2.8) is the density function of the multivariate normal distribution (Smith, 1973) and $\mathbf{b_i}$ is estimated by the posterior mean of $\mathbf{b_i}$.

$$\hat{\mathbf{b}}_i(\theta) = \mathbf{E}[\mathbf{b}_i \mid \mathbf{Y}_i = \mathbf{y}_i] = \int \mathbf{b}_i \mathbf{f}(\mathbf{b}_i \mathbf{y}_i)d\mathbf{b}_i = \Psi \mathbf{Z}_i^{-1}\mathbf{V}_i^{-1}(\alpha)(\mathbf{y}_i - \mathbf{X}_i\hat{\beta}) \tag{2.9}$$

This estimate is Best Linear Unbiased Predictor (BLUP) for $\mathbf{b_i}$ (Searle et al., 1992). The covariance matrix for $\mathbf{b_i}$ is:

$$\text{cov}(\hat{\mathbf{b}}_i) = \Psi \mathbf{Z}_i^{\mathrm{T}} \left\{ \mathbf{W}_i - \mathbf{W}_i \mathbf{X}_i (\sum_{i=1}^{N} \mathbf{X}_i \mathbf{W}_i \mathbf{X}_i)^{-1} \mathbf{X}_i^{\mathrm{T}} \mathbf{W}_i \right\} \mathbf{Z}_i \Psi , \tag{2.10}$$

and to assess the variation of the difference between the random effects estimate and observed (Laird and Ware,1982), the following expression is used.

$$\text{cov}(\hat{\mathbf{b}}_i - \mathbf{b}_i) = \Psi - \text{var}(\hat{\mathbf{b}}_i)$$

Hypothesis Tests and Confidence Intervals

Tests are useful to assess the precision of the estimates and the significance of the terms in the model. The first test discussed here is the likelihood ratio test (LRT). Although called a likelihood test, this test can also be used to compare nested models fitted by restricted maximum likelihood, but the models have to have the same fixed

effects (Pinheiro and Bates, 2000). The nested model occurs when one model represents a special case of another. So, if L_2 is the greater likelihood of the more general model and L_1 is the smaller likelihood of the restricted model, the LRT will be:

$$LRT = 2\log(L_2 / L_1) = 2\,[\log(L_2) - \log(L_1)] \tag{2.11}$$

Since $L_2 > L_1$, the LRT will be positive and if k_i is the number of parameters in model i, the distribution of the LRT is a χ^2 distribution with k_2-k_1 degrees of freedom. The LRT value is compared with a $\chi^2(k_2\text{-}k_1,\alpha)$ critical value and if LRT> $\chi^2(k_2\text{-}k_1,\alpha)$, generating a significant p-value (<0.05), the more general model is preferred over of the restricted model.

The model precision also can be assessed by the information statistic. This statistic is represented by two methods: Akaike Information Criterion (AIC) (Sakamoto et al., 1986) and Bayesian Information Criterion (BIC) (Schwarz, 1978). These criteria are evaluated as

$$AIC = -2\log(L) + 2n_{par} \tag{2.12a}$$

$$BIC = -2\log(L) + n_{par}\log(N) \tag{2.12b}$$

for each model, where L is the likelihood value and n_{par} is the number of parameters in the model. Smaller values for both AIC and BIC are better. Since these tests are conservative (Stram and Lee, 1994), generating p-values greater than they should be, it is appropriate to use an α-value of 10% to select the best model.

Variance Functions and Correlation Structures

Based on the assumptions of the mixed-effects model, the within-subject errors are independent and normally distributed with variance σ^2. The random effects are normally distributed with mean zero and covariance matrix $\mathbf{\Psi}$ and are independent for different groups. When these assumptions are violated, we need to use techniques the model the actual data structure.

The first technique to solve these problems is to model the variance structure of the within-group errors using covariates. Davidian and Giltinan(1995) gave the following expression to define the general variance function for the within-group errors:

$$\text{var}(e_{ij} \mid b_i) = \sigma^2 g^2(\mu_{ij}, \nu_{ij}, \delta), \quad \text{i=1,…,M, j=1,…,n}_\text{i} \tag{2.13}$$

where M is a number of groups, n_i is the number of observations in the ith group, μ_{ij} $=E(y_{ij}|b_i)$, v_{ij} is a vector of variance covariates, δ is a vector of variance parameters and g(.) is the variance function. In forestry it is quite common for the within-group variability to increase with some power of the absolute value of a covariate. For example, the variability of the volume increases with diameter breast height. In this study, we will use the varFunc classes in S-Plus (Pinheiro and Bates, 2000) to specify within-group variance models. Among several classes of variance functions available, the most useful in this specific study case was the varIdent class.

$$\text{var}(\mathbf{e_{ij}} \mid \mathbf{b_i}) = \boldsymbol{\sigma}^2 \boldsymbol{\delta}^2_{\mathbf{sij}} \tag{2.14}$$

The power of this class is based on its capability to model different variances for each level of a stratification variable s. As we will see further, the analysis is based on

different genetic material (clones) with inconsistent variance patterns within-group. This kind of variance class was very useful for modeling such situations.

In correlation analysis, among several families of correlation structures, the autoregressive-moving average structure - ARMA (Box et al., 1994) is the most used and well- known. The general structure is given by:

$$\varepsilon_t = \sum_{i=1}^{p} \phi_i \varepsilon_{t-i} + \sum_{j=1}^{q} \theta_j a_{t-j} + a_t \qquad (2.15)$$

Where ε_t refers to an observation taken at time t and a_t is the noise (error) term. The first part of the expression refers to an autoregressive model (AR(p)) and second part represents the moving average (MA(q)). If p=0 we have the MA(q) situation and, conversely, if q=0 we have the AR(p). In the AR(p) part, ϕ represents the correlation parameters with order p and t-i is the lag, or distance, between two observations. The tendency is for the ϕ values to decrease over time, indicating that observations close in time are more correlated than observations far apart, which is a common in longitudinal studies. In the moving-average part (MA(q)), the model assumes that the current observation is a linear function of the independent and identically distributed noise terms (a_t). The value of q is the number of noise terms included in the model. So, there are q correlation parameters $\boldsymbol{\theta}$ in the model.

Data

The data sets are from commercial hybrid plantations of *Eucalyptus* genus from the Brazilian coastal region, Bahia and Espírito Santo States, located between 17° 48’ S and 40 ° 17’ W. In Figure 2.1, each clone is represented in one square and plots are represented by each line. For example, the clone number 6039 is represented by 4 plots

in this analysis. Each plot was permanently sampled from 3 to 10 times, with age varying from 2 to 10 years, between 1992 and 2001, and its area varied from 131 to 200 m^2. Based on these variations, the data base was longitudinal, irregularly spaced and unbalanced.

Figure 2.1 shows the relationship between 1/age and ln(BA) for each plot within each clone. It shows a fairly consistent linear decrease in ln(BA) with the 1/age, but with variations in intercept and/or slope by plot/clone combinations. Based upon this data set, one could group the clones based on their curve trend. For example, the clones 6039, 6054, 3903, 2747 and 1030 could be grouped based on their growth trend. This was not done because each clone has different management and technological properties and it is important to analyze each one separately. We will see later that these variations are more due to site than to clonal characteristics, but some variation is credited to the clones' growth potential.

Fitting Linear Mixed-Effects Models

To fit the linear mixed-effects model we will follow the methods suggested by Pinheiro and Bates (2000). In this specific case, the response variable is ln(BA) and the fixed effect is represented by 1/age. The random effect will be the plots or sample units, which are random units chosen from a population. So, in this case we will fit a single-level model.

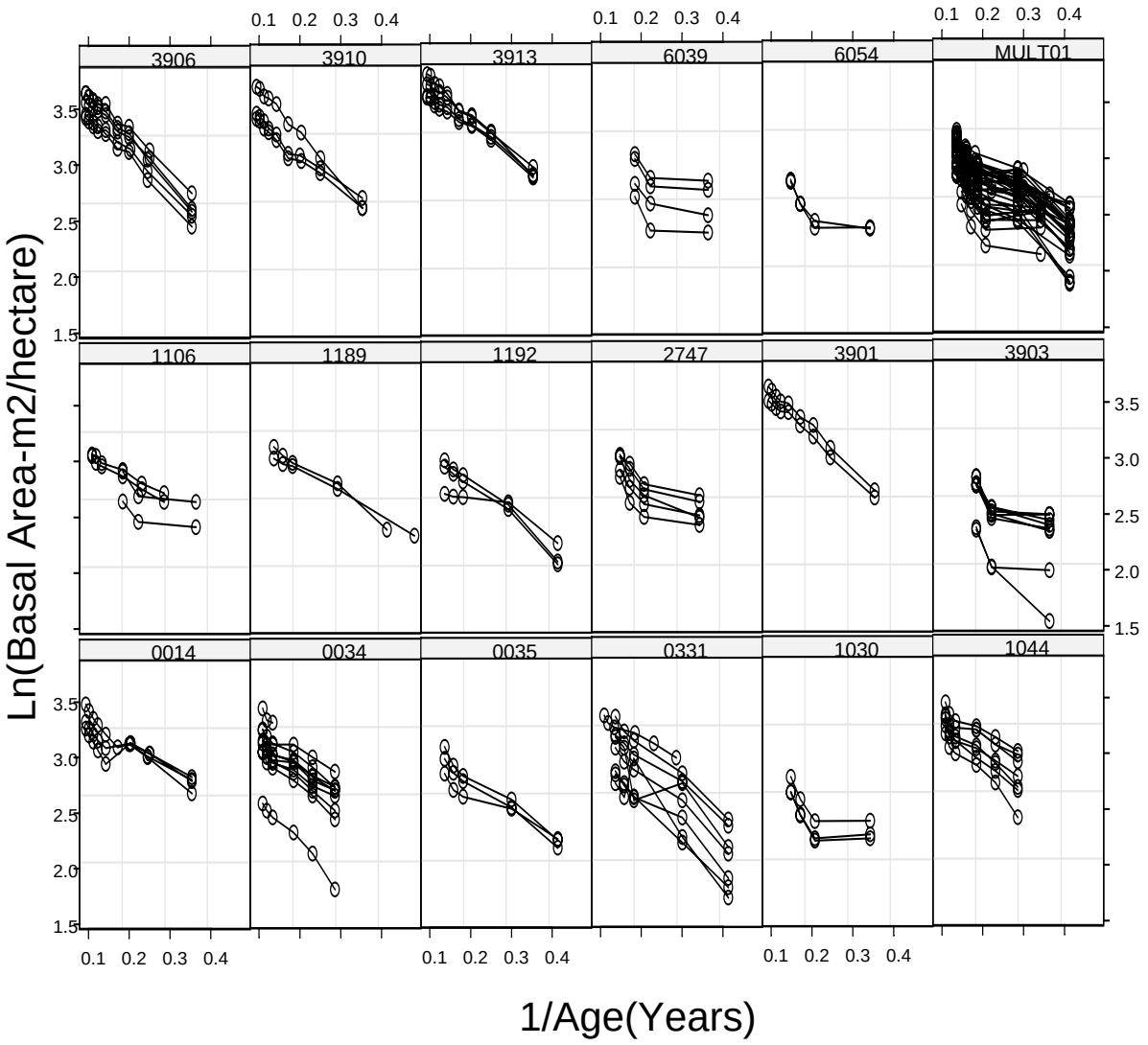

FIGURE 2.1. Linear relationship of the natural logarithm of Basal Area (m2/hectare) measured over time (1/age), for 18 different clones.

The first step is to fit a single linear regression model of ln(BA) on 1/age to the data from all the plots, ignoring the grouping structure. After the fit, there is considerable variability, as shown in the residual plots on Figure 2.2. Another important point is the unusual influence on the fit by some observations, mainly observation number 229, 233 and 502. In addition, the normal probability plot indicates that the error distribution has

heavier tails. The Cook's Distance shows two more important observations which are not outliers but have significant influence in the fit results. These features indicate that the simple linear regression model does not represent the structure of the data.

To check if there are differences among clones, we fitted a model which has specific intercept and slope for each clone. The results of the interaction effect are shown in Table 2.1. All Clone:(1/age) interaction had a significant p-value, suggesting that growth patterns are different for different clones. Because the data are from repeated measures on each plot, the basic assumption of independence of linear models may be violated. Three clones are represented in Figure 2.3 for easy visualization of the subject effect. When we plot the residual for each plot, the signal tended to be the same. This characteristic is the motivation for using the linear mixed-effects models.

The next step is to perform a preliminary analysis to decide which random effects to include in the model and what covariance structure will be most appropriate. To eliminate the correlation between slopes and intercepts for each plot, the data were centered at 4 years (1/age=0.25). Figure 2.4 shows no correlations among the pairs of slope-intercept for centered data. Both intercept and slope appear to vary per plot. The estimated value and its interval were plotted for twelve plots representing three clones to visualize how the parameter estimates vary among individuals (Figure 2.5). The confidence intervals for each plot give a clear indication that it is necessary to estimate separate intercepts and, in some cases, separate slopes to account for plot-to-plot variability. Because only 12 plots were represented, the confidence intervals for slope had a high overlap. If we had used all 115 plots, the overlap would be less frequent, indicating that both parameters could be considered as random effects.

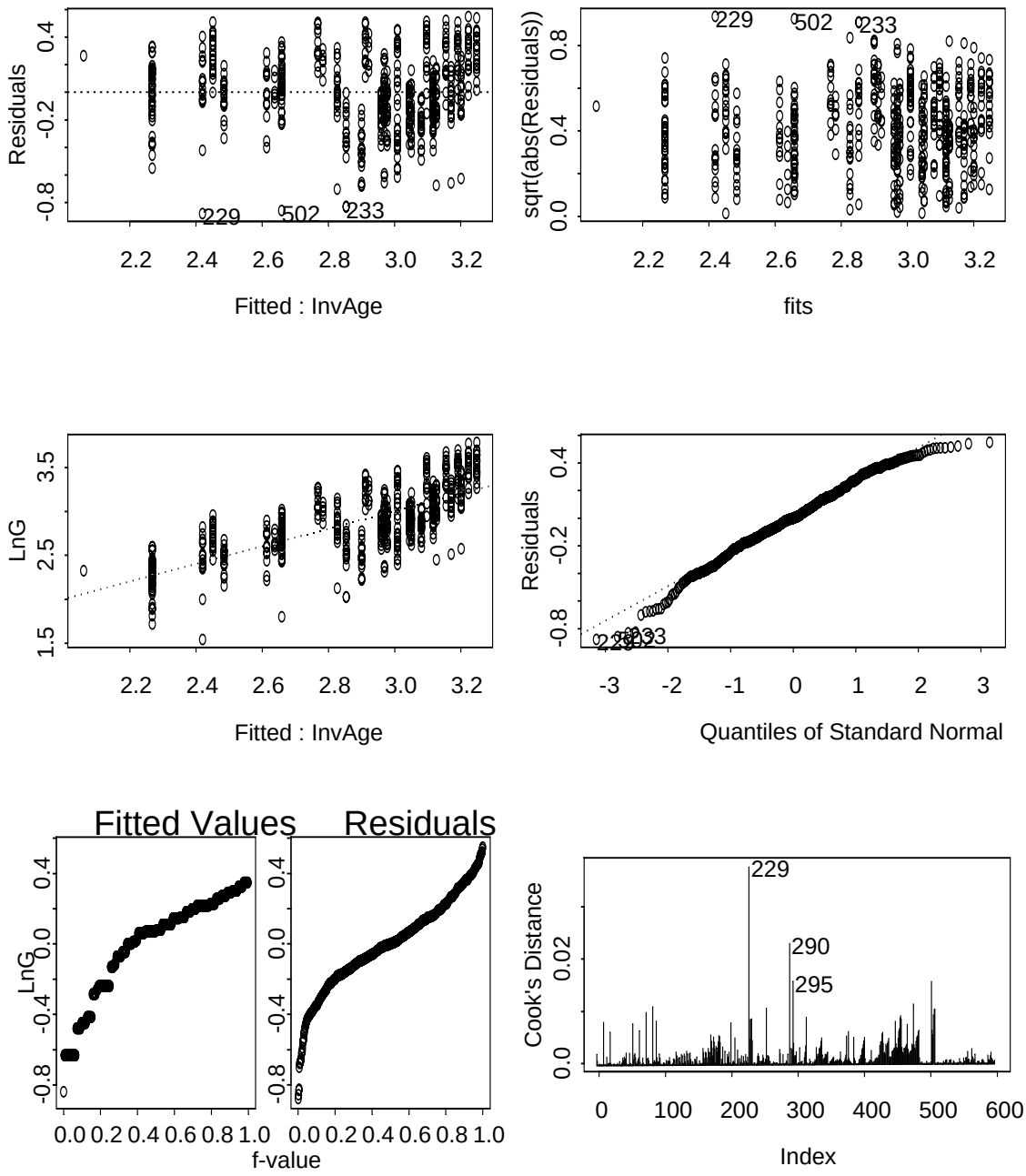

FIGURE 2.2. Diagnostic plots for the simple linear regression model fitted on ln(BA) versus 1/age, for 595 plots representing 18 clones.

TABLE 2.1. Significance of the Clones and interaction between Clone and 1/age representing the data set used in the analysis.

Coefficients:	Value	Std. Error	t value	Pr(>|t|)
(Intercept)	2.7628	0.0205	134.8305	< 0.0001
Clone0014	-0.1425	0.0504	-2.8278	0.0057
Clone0034	-0.0715	0.0365	-1.9599	0.0529
Clone0331	-0.0706	0.0210	-3.3541	0.0011
Clone1030	0.0361	0.0124	2.9166	0.0044
Clone2747	0.0312	0.0111	2.8066	0.0061
Clone3901	-0.0235	0.0055	-4.2767	< 0.0001
Clone3903	0.0212	0.0061	3.4765	0.0008
Clone3906	0.0162	0.0072	2.2644	0.0258
Clone3910	0.0306	0.0058	5.2548	< 0.0001
Clone6039	-0.0165	0.0072	-2.2912	0.0241
Clone6054	-0.0035	0.0019	-1.7990	0.0751
InvAgeC	-2.4075	0.0764	-31.5007	< 0.0001
Clone0014:InvAge	-1.8937	0.2103	-9.0046	< 0.0001
Clone0034:InvAge	-3.2333	0.1793	-18.0317	< 0.0001
Clone0035:InvAge	-3.2125	0.2125	-15.1206	< 0.0001
Clone0331:InvAge	-3.1377	0.1470	-21.3491	< 0.0001
Clone1030:InvAge	-4.3190	0.2659	-16.2444	< 0.0001
Clone1044:InvAge	-2.1929	0.1994	-10.9973	< 0.0001
Clone1106:InvAge	-3.0534	0.2345	-13.0193	< 0.0001
Clone1189:InvAge	-2.5108	0.2452	-10.2378	< 0.0001
Clone1192:InvAge	-3.3173	0.2140	-15.5048	< 0.0001
Clone2747:InvAge	-3.1551	0.2133	-14.7900	< 0.0001
Clone3901:InvAge	-1.2704	0.2689	-4.7238	< 0.0001
Clone3903:InvAge	-3.7837	0.1703	-22.2177	< 0.0001
Clone3906:InvAge	-1.5131	0.1877	-8.0601	< 0.0001
Clone3910:InvAge	-1.7063	0.2267	-7.5266	< 0.0001
Clone3913:InvAge	-0.6142	0.2027	-3.0296	0.0026
Clone6039:InvAge	-2.7458	0.2284	-12.0233	< 0.0001
Clone6054:InvAge	-3.8854	0.3196	-12.1560	< 0.0001
CloneMUL1:InvAge	-2.9523	0.1002	-29.4767	< 0.0001

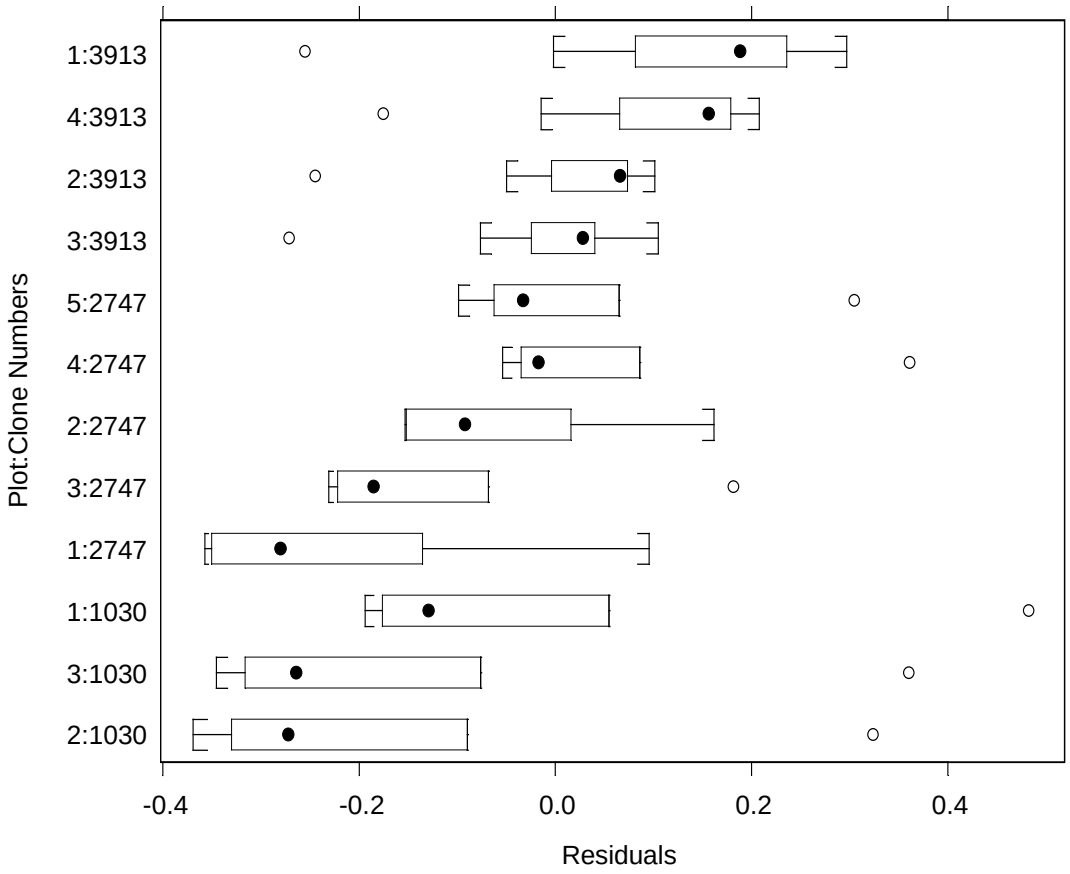

FIGURE 2.3. Residual plots of the linear model by subject.

Following from these analyses, we have arguments to fit the model as a mixed-effects model, considering plots as a random unit.

The mixed-effects model was first fitted considering just 1/age as a fixed covariate. The AIC, BIC and REML were -747.3824, -721.0611, and 379.6912, respectively. To make sure if the basal area growth patterns are different among clones, a new model was fitted including the categorical variable Clone and the interaction

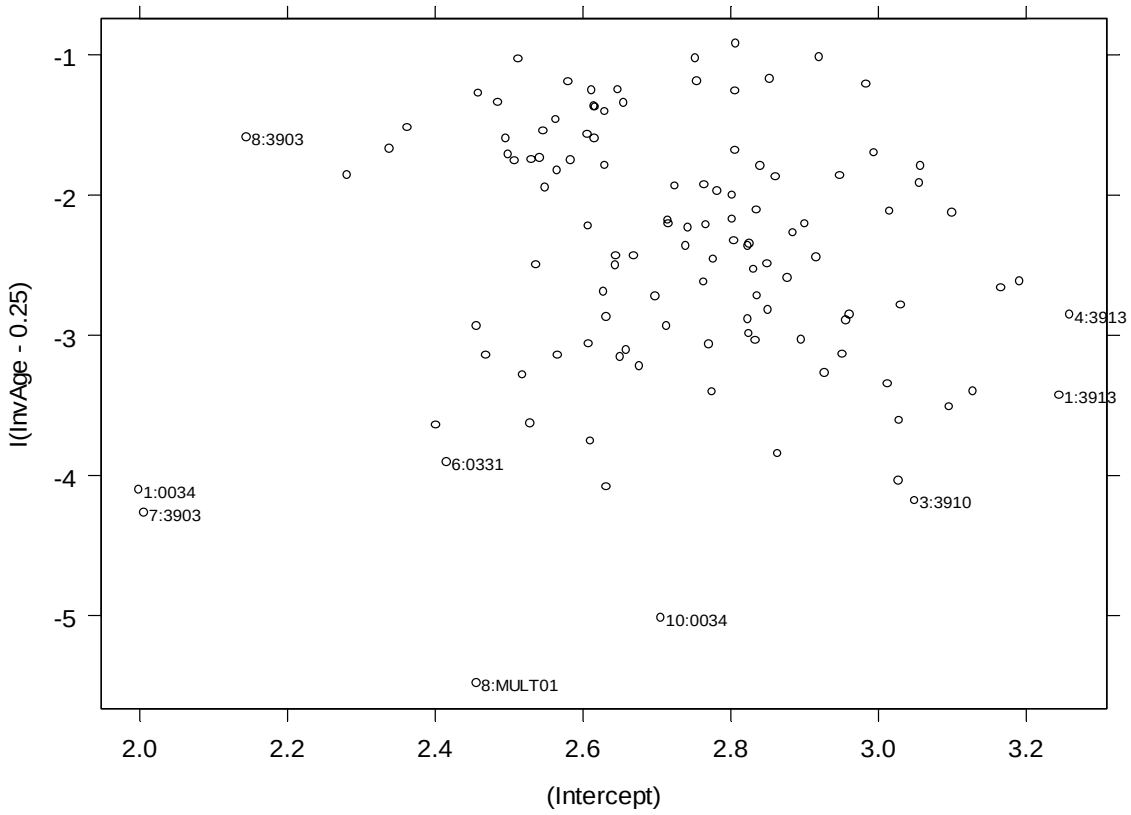

FIGURE 2.4. Pairs plot relationship between intercept and slope fitted by plot.

Clone*(1/age). The new values for AIC, BIC and REML were -644.8446, -471.7272, and 362.4223. Based on these estimated parameters, the first model had better performance. But, when we check the p-values for each clone and for the interaction we found that 10 clones of 18 had significant p-value < 0.05, indicating different growth patterns among clones. So, we decided to maintain the variables Clone and Clone*(1/age) in the model. Further, our final purpose is to develop a model that accounts for the variability among clones.

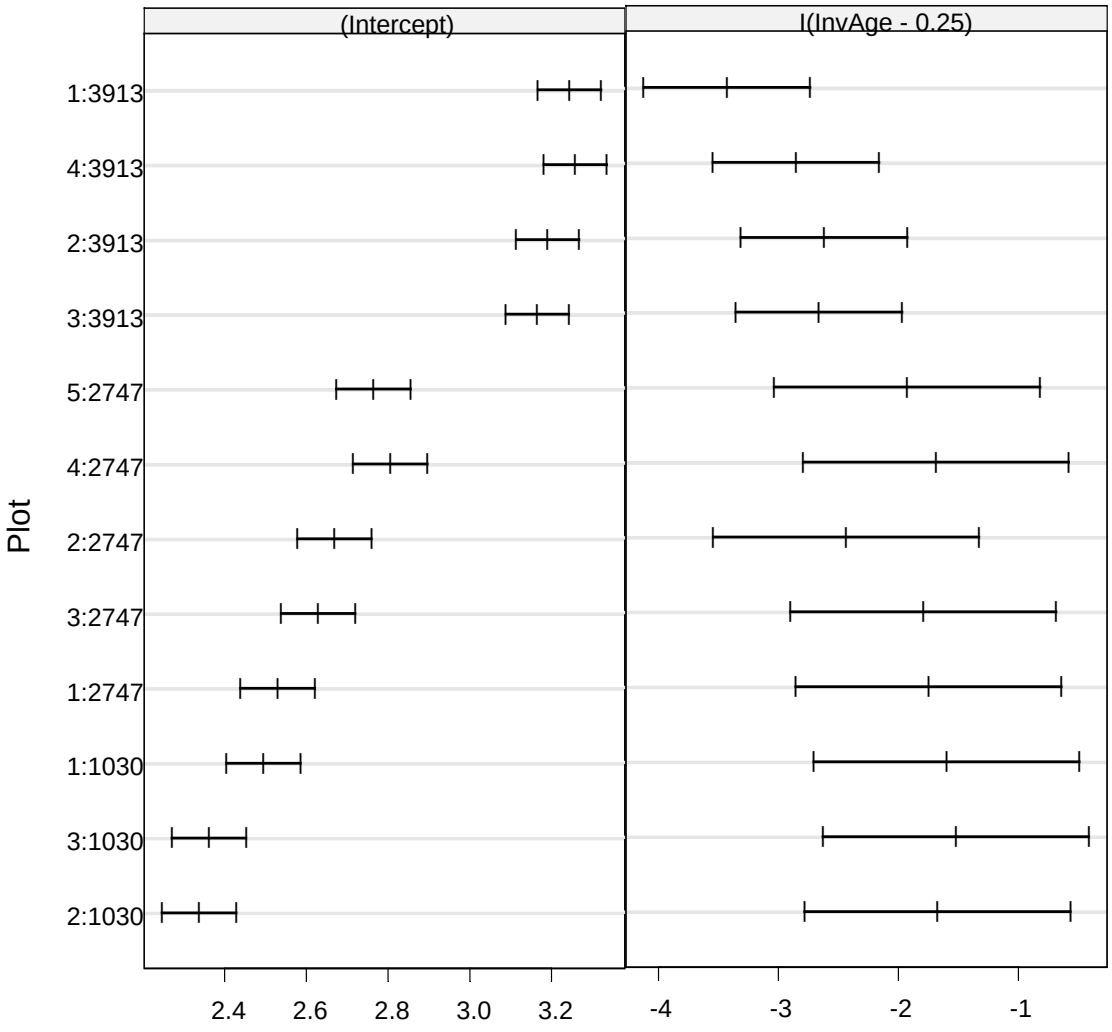

FIGURE 2.5. Ninety-five percent confidence intervals on intercept and slope for each plot.

To verify the effect of using mixed-effect techniques in fitting the model, the values of the parameters estimated using just fixed effects were plotted against parameters estimated by the mixed-effect model (Figure 2.6). It can be seen that the individual estimates from the linear mixed-effect model, represented by “+”, tend to be “pulled toward” the fixed-effects estimates, represented by “o”, mainly in larger

residuals. This occurs for the reason that mixed-effects estimations represent the effect of the individual fits and the fixed-effects estimates, associated with the population averages. These are often referred to as “shrinkage estimates”. In plots with an outlying basal area growth pattern, the shrinkage is more evident. This characteristic gives a certain robustness to outlying plot behavior. When we compare the lines fitted by the fixed-effect and mixed-effect models (Figure 2.7; representing 36 plots of a total of 115), this attribute is better visualized. In some plots, like 7:3903 and 1:0331, we can verify the greater sensibility of the fixed-effect in fitting individual plots. In every plot, the distances from observed and estimated values are smaller when the plot effect is included in the model, generating smaller residual mean squares.

Checking Distributional Assumptions

Two basic distributional assumptions will be checked: within groups and random effects. Within-group errors are considered independent and identically normally distributed, with mean zero and variance σ^2 and independent of the random effects. The random effects will be verified if they are normally distributed with mean zero and variance-covariance matrix $\boldsymbol{\psi}$, which do not depend on the group and are independent for different groups. As pointed out by Pinheiro and Bates (2000), the most useful of the methods for assessing the validity of these assumptions are based on plots of the residuals, the fitted values, and the estimated random effects. Also, the tests could be performed by using hypothesis tests, but the conclusions rarely contradict the information displayed in the plots.

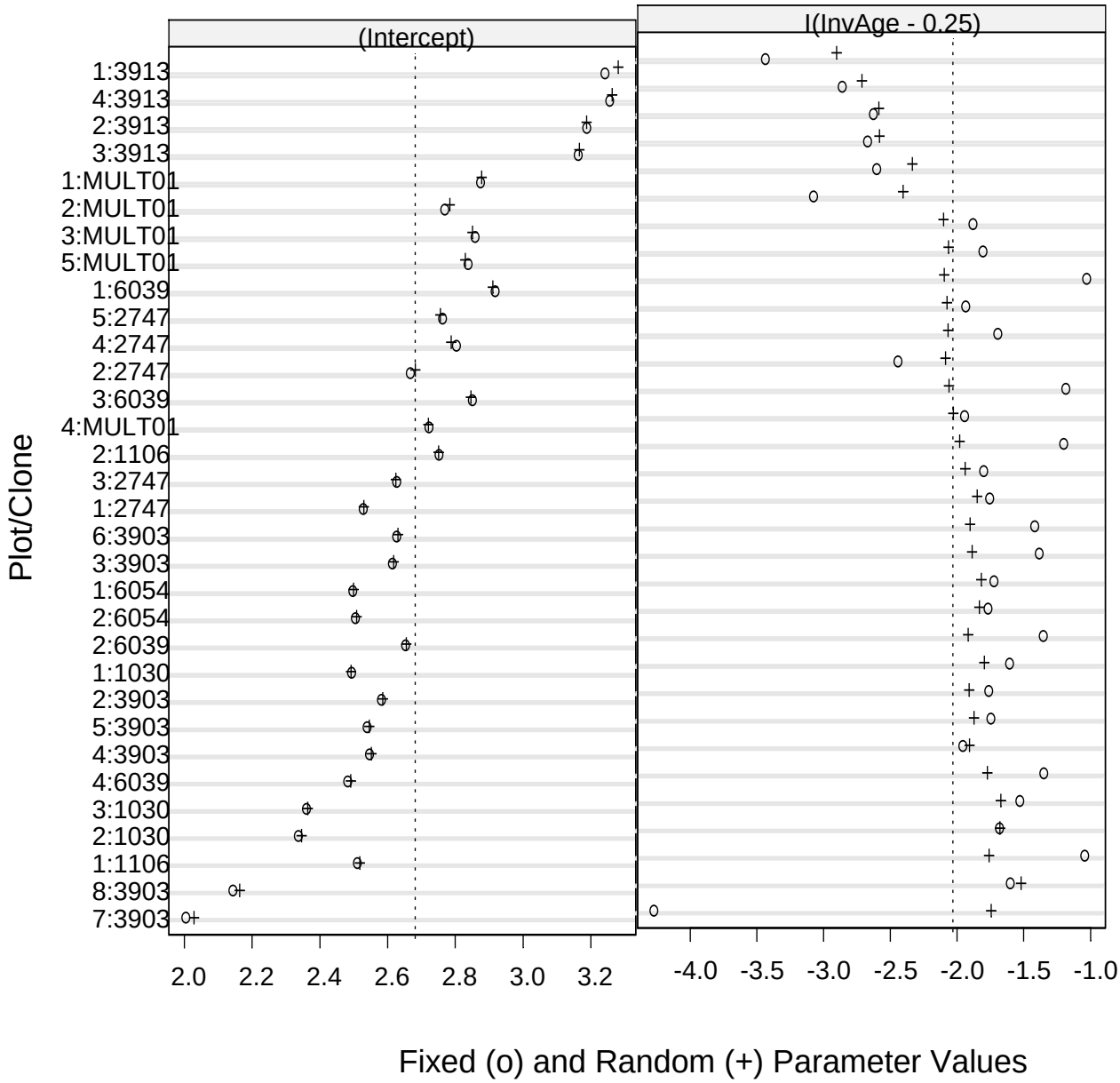

FIGURE 2.6. Comparing individual estimates from linear fixed-effects model (o) and linear mixed-effects model (+).

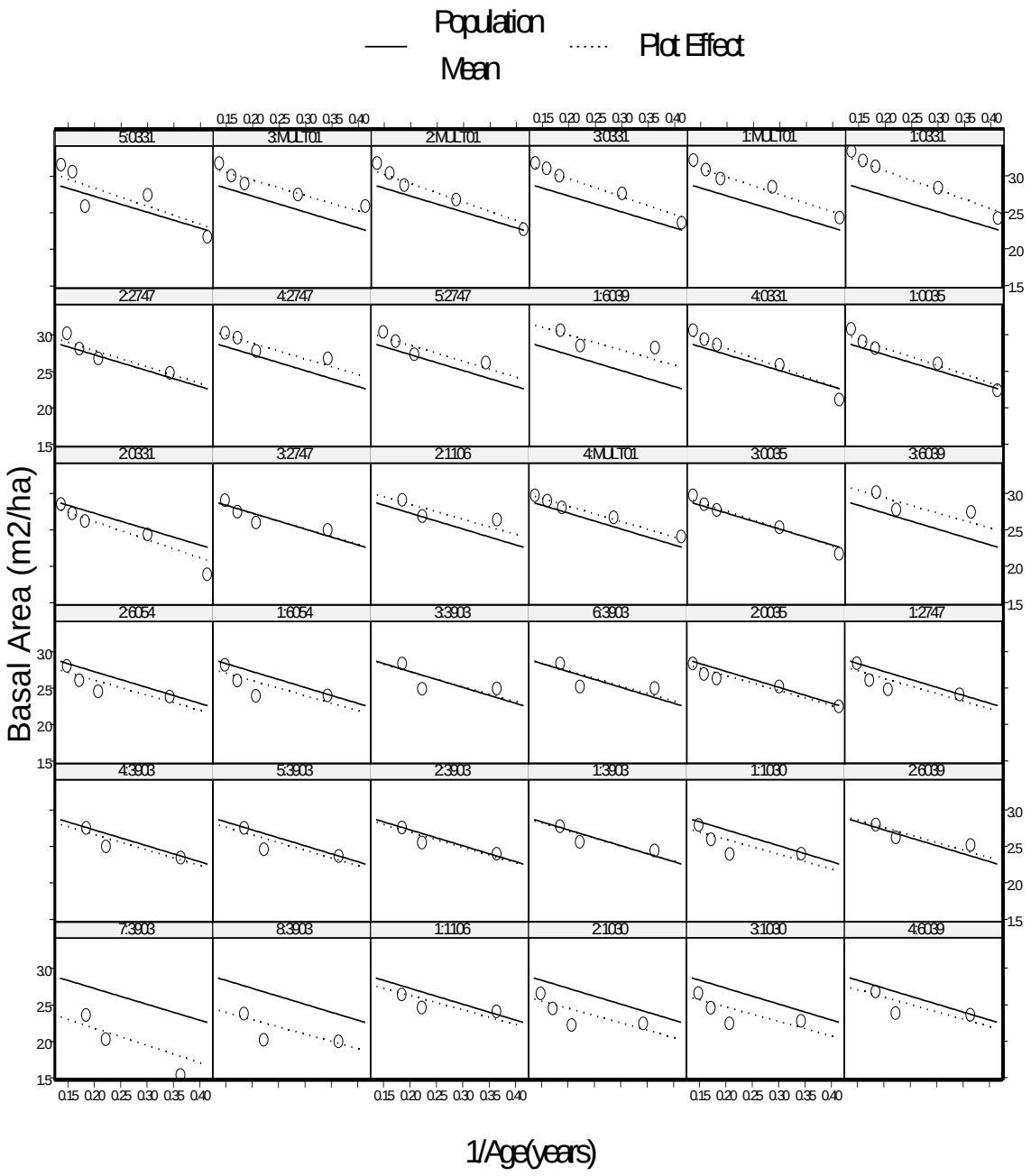

FIGURE 2.7. Individual predicted values from fixed (solid), mixed-effects model (dashed) and observed data (o).

As we can see on Figure 2.8, in 12 plots representing 3 of the 18 clones in the analysis, the residuals are distributed around zero, confirming the assumption that $E[\boldsymbol{\varepsilon}]=\mathbf{0}$. In residual analysis, it is apparent that the variability is among plots. So, the within-plot constant variance assumption was violated and it will be necessary to model this variability to improve the model. If we use all 17 clones in the analysis, the residual distribution becomes more clustered. Also, Figure 2.8 shows some outlier observations in plot clone 1:3913 and 2:2747 and larger residuals for clone 1030. The box plots indicate that the variability is greater among plots representing the clone 1030. If the standard residuals versus fitted values for all 18 clones were plotted (Figure 2.9), it is clear that the variability among plots from some clones are greater than plot from others.

Based on observations of Figures 2.8 and 2.9, the first idea was to model the variance by clone for the within-group error. It is possible to do this by using the S-plus varIdent function, from varFunc Classes in the nlme library. This procedure allows a variance model with different variances for each level or clone (Pinheiro and Bates, 2000). Table 2.2 represents the different values estimated for homoscedastic and heteroscedastic models. The smaller values of AIC and BIC, the greater value of log-likelihood and the very small p-value of the likelihood ratio statistic confirm that the heteroscedastic model improved on adjusting and explaining the data over the homoscedastic model. Graphics based on residuals and quantiles of standard normal are presented to confirm the improvement (Figures 2.10 and 2.11).

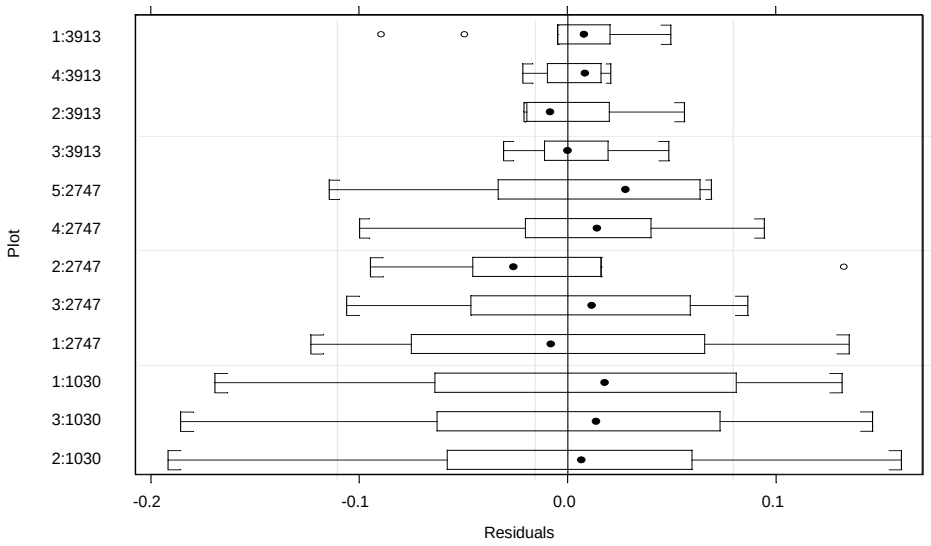

FIGURE 2.8. Boxplots of the residuals for linear mixed-effects model.

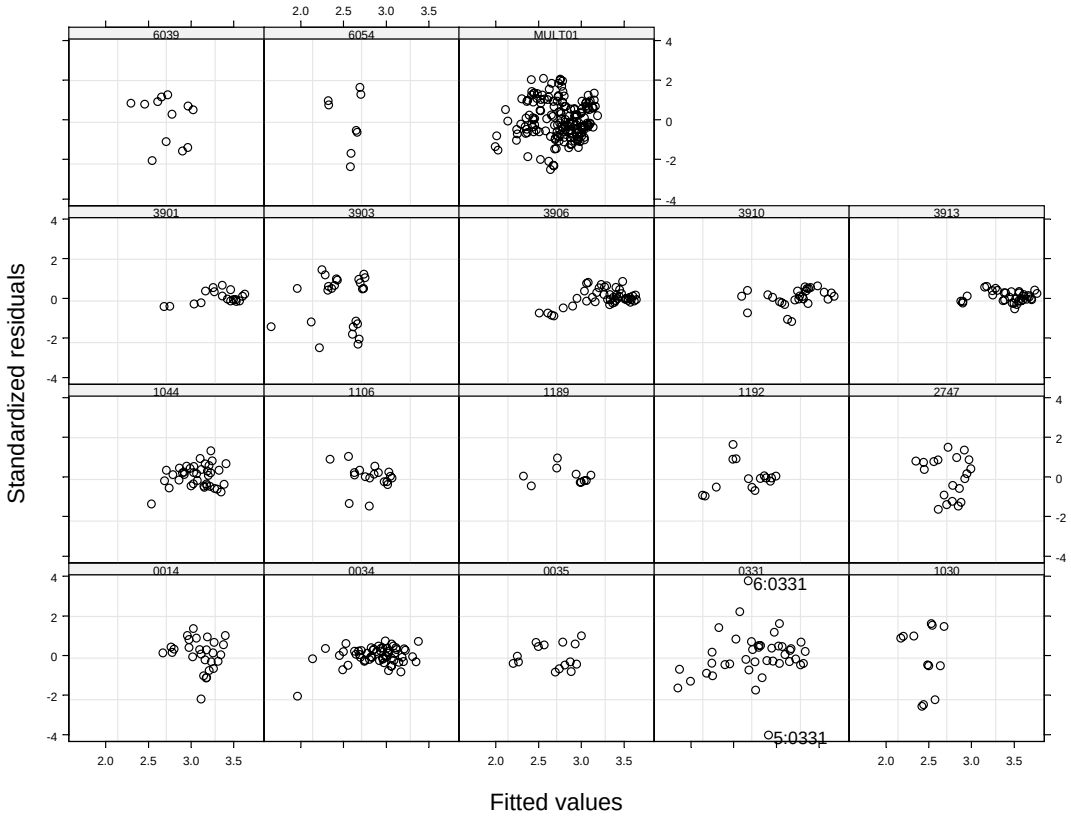

FIGURE 2.9. Residuals versus fitted values for linear mixed-effects model by clone.

TABLE 2.2. Comparing Akaike Information Criterion (AIC), Bayesian Information Criterion(BIC) and LogLikelihood (logLik) for homoscedastic and heteroscedastic models.

Model	DF	AIC	BIC	LogLik	Test LogLik Ratio		P-Value
1 - Homoscedastic	6	-747.38	-721.06	379.69	-	-	
2 - Heteroscedastic	70	-901.24	-798.28	520.62	1 vs 2	281.86	<0.0001

The next step is to asses the assumptions on the random effects. Figure 2.10 represents the normal plot of estimated random effects for the heteroscedastic model. The assumption of normality appears reasonable for random effects, despite the fact that there are two outliers: one for intercept (1:0034) and one for slope(8:0331). Considering the variability of the data set, representing different clones in distinct regions, the presence of just two outliers is not a concern. The second assumption related to random effects can be checked in Figure 2.11. It can be seen that the pairs slope-intercept for all plot:clone combinations have mean close to zero and constant variance.

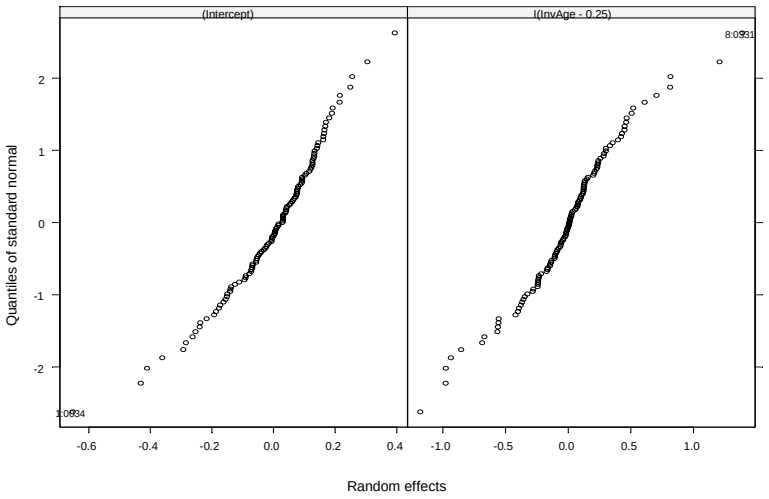

FIGURE 2.10. Normal plot of estimated random effect for heteroscedastic fitted model.

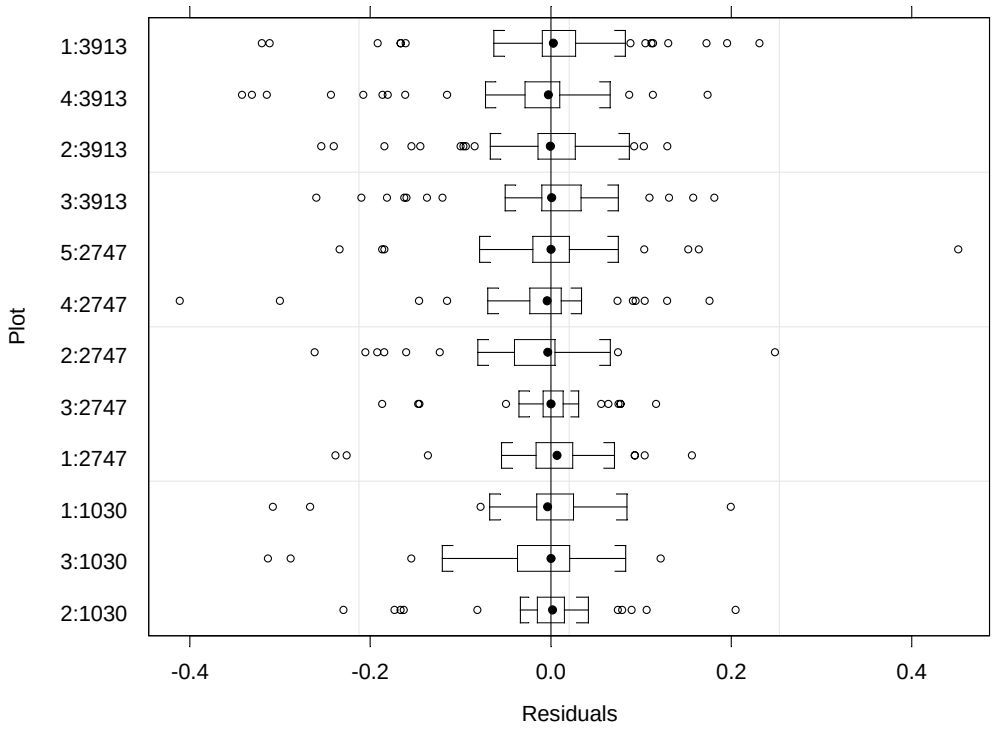

FIGURE 2.11. Boxplots of the residuals for heteroscedastic linear mixed-effects model.

Since the data sets are from longitudinal information, with repeated measurements by plots and/or spatial data, with observations indexed by spatial location, the next step is to verify the correlation structures for modeling possible correlation dependencies. Because these data are not equally spaced in time, we used the spatial correlation to fit continuous-time correlation models. Figure 2.12 is a graphical representation of sample semivariogram. The semivariogram values appear to increase up to 0.10 and then decrease. We tried to model this pattern using five spatial correlation models: Exponential, Gaussian, Linear, Rational quadratic, and Spherical. The Spherical spatial model had better fit in this specific situation. The resulting plot, shown in Figure 2.13,

appears to vary randomly around y=0.5, with no observed patterns. This suggests that the spherical model is adequate. The AIC and BIC values are smaller indicating that heteroscedastic-autocorrelation model had an improvement in representing these data. Also, the large value of the likelihood ratio test indicate the evidence of dependence, generating a p-value=0.0004.

TABLE 2.3. Comparing Akaike Information Criterion (AIC), Bayesian Information Criterion(BIC) and LogLikelihood (logLik) for homoscedastic, heteroscedastic and heteroscedastic-autocorrelation models.

Model	DF	AIC	BIC	LogLik	Test LogLik Ratio		P-Value
1- Homoscedastic	6	-747.38	-721.06	379.69	-	-	-
2- Heteroscedastic	70	-901.24	-798.28	520.62	1 vs 2	281.86	< 0.0001
3- Heteroscedastic/ Autocorrelation	37	-1034.72	-872.41	554.36	2 vs 3	67.49	0.0004

The assumption of normalized residuals for each clone can be confirmed by Figure 2.14. With some exceptions, the residuals have fairly normal distribution, confirming the normality assumption after the modeling heteroscedastic and autocorrelation.

The final parameters estimated for fixed and random effects can be seen on Table 2.4. Considering all 115 plots, both random intercept and slope had positive or negative values. The parameter estimates for each plot are generated by adding the fixed and random effects. So, considering the parameter values, the curves for each plot had different intercepts and slopes.

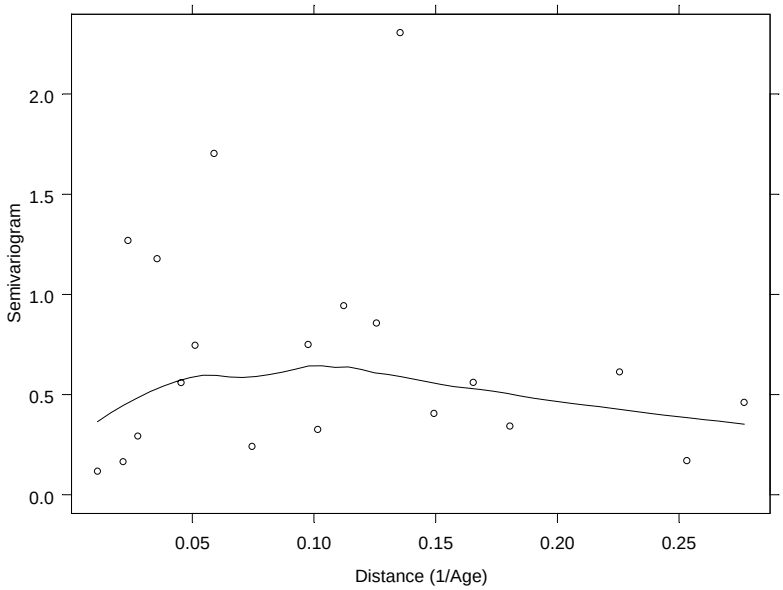

FIGURE 2.12. Sample semivariogram estimates for linear mixed-effects model before modeling the spatial autocorrelation.

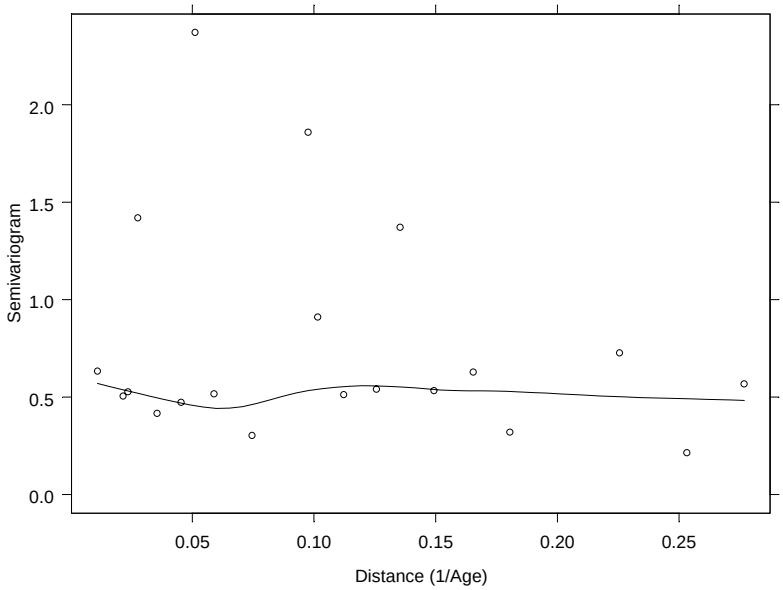

FIGURE 2.13. Sample semivariogram estimates for linear mixed-effects model after modeling the spatial autocorrelation.

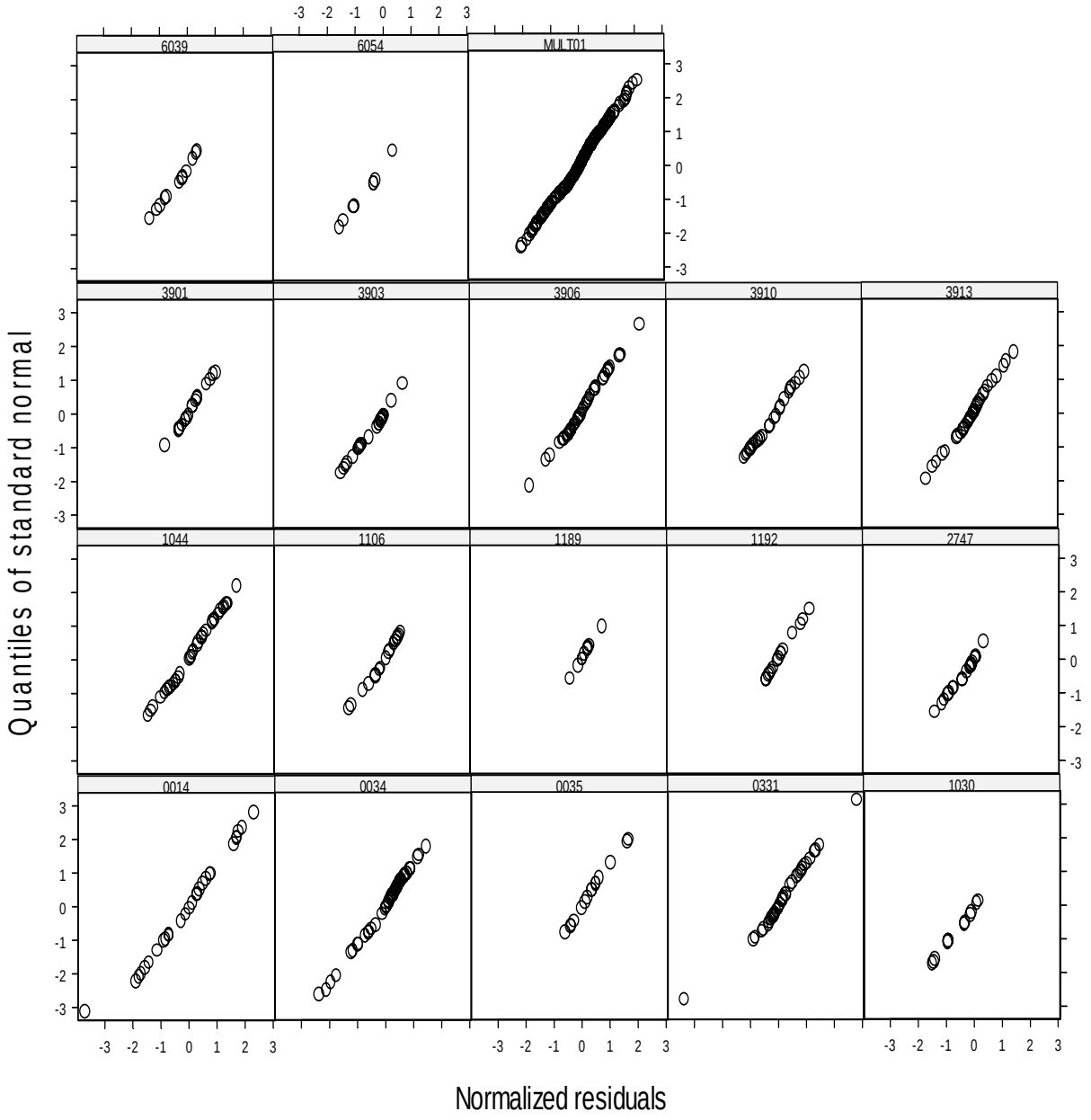

FIGURE 2.14. Normal plots of residuals for each clone.

TABLE 2.4. Fixed and mixed parameters estimates for the final model, with heteroscedastic and autocorrelation modeled, representing 44 plots in a total of the 115.

Plot:Clone	Fixed Intercept - β_o	Random Intercept- b_{oi}	Fixed Slope - β_1	Random Slope- b_{1i}
1:0331	2.7505	0.173135	-2.494754	-0.602890
2:0331	2.7505	-0.326031	-2.494754	-0.820808
3:0331	2.7505	0.086599	-2.494754	-0.437708
4:0331	2.7505	-0.112138	-2.494754	-0.759980
5:0331	2.7505	0.003331	-2.494754	-1.161210
6:0331	2.7505	-0.468503	-2.494754	-0.934774
7:0331	2.7505	-0.339641	-2.494754	-1.133078
8:0331	2.7505	0.303319	-2.494754	0.701643
1:1030	2.7505	-0.153020	-2.494754	0.456104
1:1044	2.7505	-0.062380	-2.494754	-0.320059
2:1030	2.7505	-0.300282	-2.494754	0.375090
2:1044	2.7505	0.037841	-2.494754	-0.644155
1:1106	2.7505	-0.172280	-2.494754	0.907539
3:1030	2.7505	-0.279085	-2.494754	0.517948
3:1044	2.7505	0.113330	-2.494754	0.077612
2:1106	2.7505	0.075706	-2.494754	0.766970
1:1189	2.7505	0.031074	-2.494754	0.462230
1:1192	2.7505	-0.135062	-2.494754	-0.746007
4:1044	2.7505	0.273362	-2.494754	0.690275
3:1106	2.7505	0.001544	-2.494754	0.186253
2:1189	2.7505	0.054275	-2.494754	-0.143285
2:1192	2.7505	-0.166382	-2.494754	-0.713032
5:1044	2.7505	0.343911	-2.494754	0.488235
4:1106	2.7505	-0.043321	-2.494754	0.330997
3:1192	2.7505	-0.227671	-2.494754	0.774958
6:1044	2.7505	0.137643	-2.494754	0.491384
7:1044	2.7505	0.273095	-2.494754	-0.186133
1:2747	2.7505	-0.125751	-2.494754	0.242588
2:2747	2.7505	-0.007582	-2.494754	-0.292900
3:2747	2.7505	-0.054132	-2.494754	0.383382
4:2747	2.7505	0.095186	-2.494754	0.686109
5:2747	2.7505	0.073741	-2.494754	0.340994
1:3901	2.7505	0.330668	-2.494754	-1.054732
1:3903	2.7505	-0.079383	-2.494754	0.504073
1:3906	2.7505	0.098306	-2.494754	-1.391277
1:3910	2.7505	0.168650	-2.494754	-0.603410
1:3913	2.7505	0.496449	-2.494754	-1.043936
2:3901	2.7505	0.244551	-2.494754	-0.847973
2:3903	2.7505	-0.108126	-2.494754	0.385165
2:3906	2.7505	0.168061	-2.494754	-0.724138
2:3910	2.7505	0.234550	-2.494754	-0.388806
2:3913	2.7505	0.419233	-2.494754	-0.204785
3:3903	2.7505	-0.015414	-2.494754	0.491923

Relating the Random Parameters with Plot Characteristics.

The focal point here is to relate the random intercept and slope estimates to plot characteristics. This method is similar to parameter prediction by Clutter(1983), that related fixed parameters of the fitted curves to site index and other variables through linear or nonlinear regression procedures. The crucial difference is that the parameters used are random, representing each plot sampled. The application of this method is to estimate the random-effect parameters for other plots, which represent a specific stand, based upon plot properties, such as site index and trees per hectare. In Figures 2.15 and 2.16 it can be seen that the random intercept is linear-positively related to site index (base-age = 7 years) and the random slope is linear-negatively related to site index. This suggests a linear model to represent this relationship. The variable N (number tree per hectare) does not have a significant effect in explaining the variation.

The results of the simple linear regression between site index and random parameters are presented on Figures 2.17 and 2.18. We had a better fit for the intercept variable with good residual distribution and relatively high R-squared (0.63). Even though the R-square for slope was not strong (0.12), the important result was the residual distribution, which had uniform distribution around zero.

In order to check the precision of relating random parameters with plot characteristics, we compare estimated and observed basal-area for some plots which were not included in the analysis (Table 2.4). The percentage difference between observed and estimated basal-area varies from 0% to 31.6% and most of them were less than 10 percent. In view of the fact that the clones in this analysis are different from those used to estimate the mixed-effect model for basal area, this result can be considered

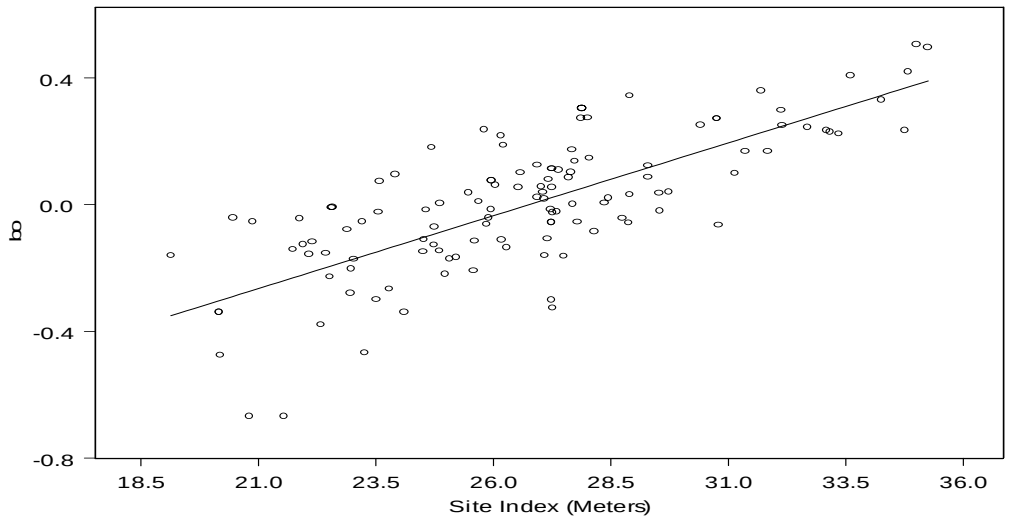

FIGURE 2.15. Relationship between Site Index and Random Effect b_o (intercept) estimated based on mixed-effects model.

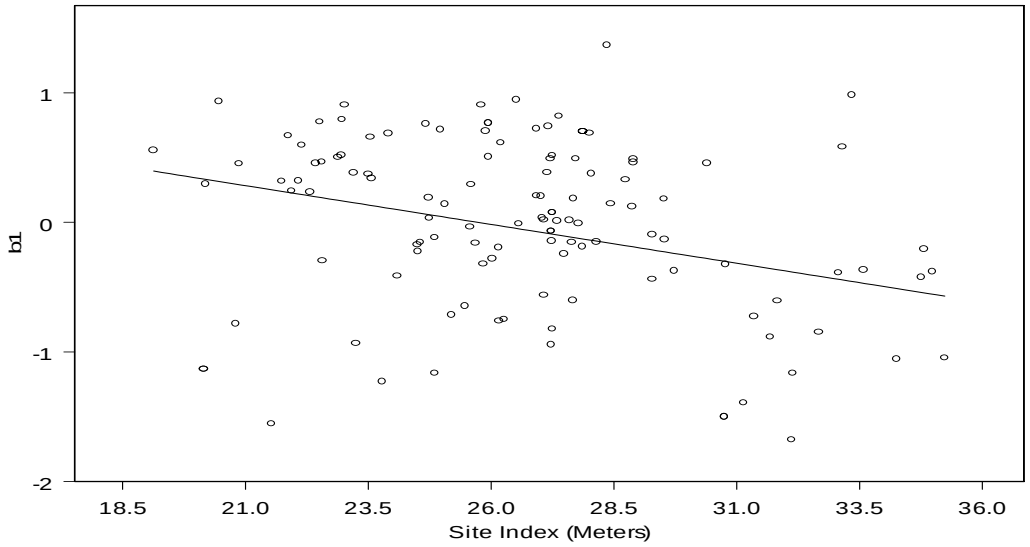

FIGURE 2.16. Relationship between Site Index and the random effect b_1(slope) estimated based on mixed-effect model.

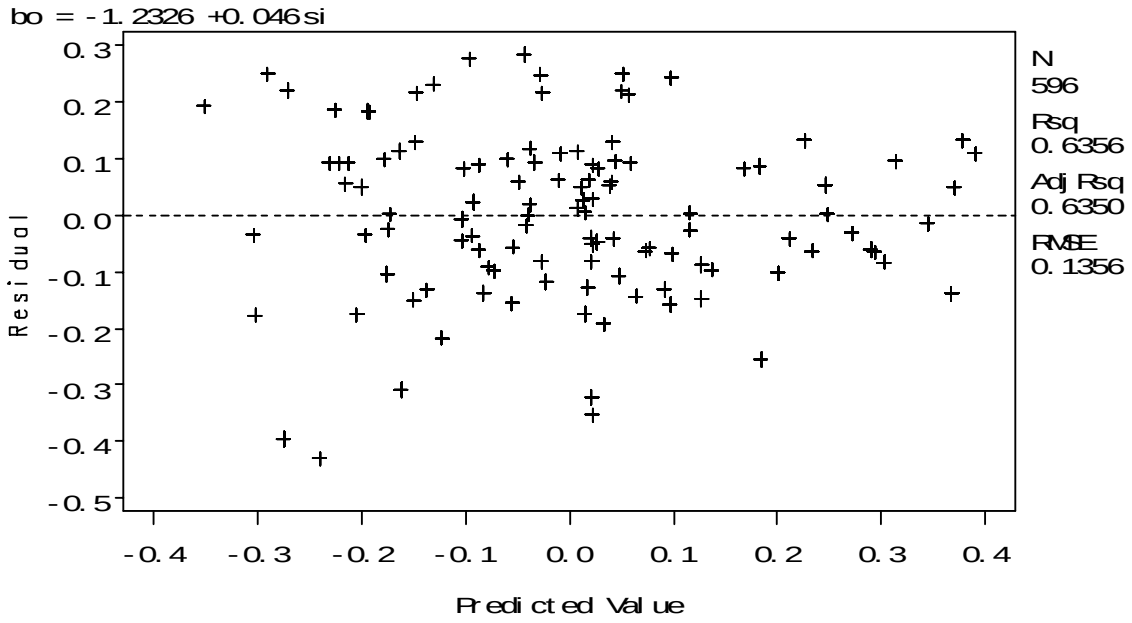

FIGURE 2.17. Parameters estimated, statistics and residual distribution of linear relationship between Site Index and the random effect b_o.

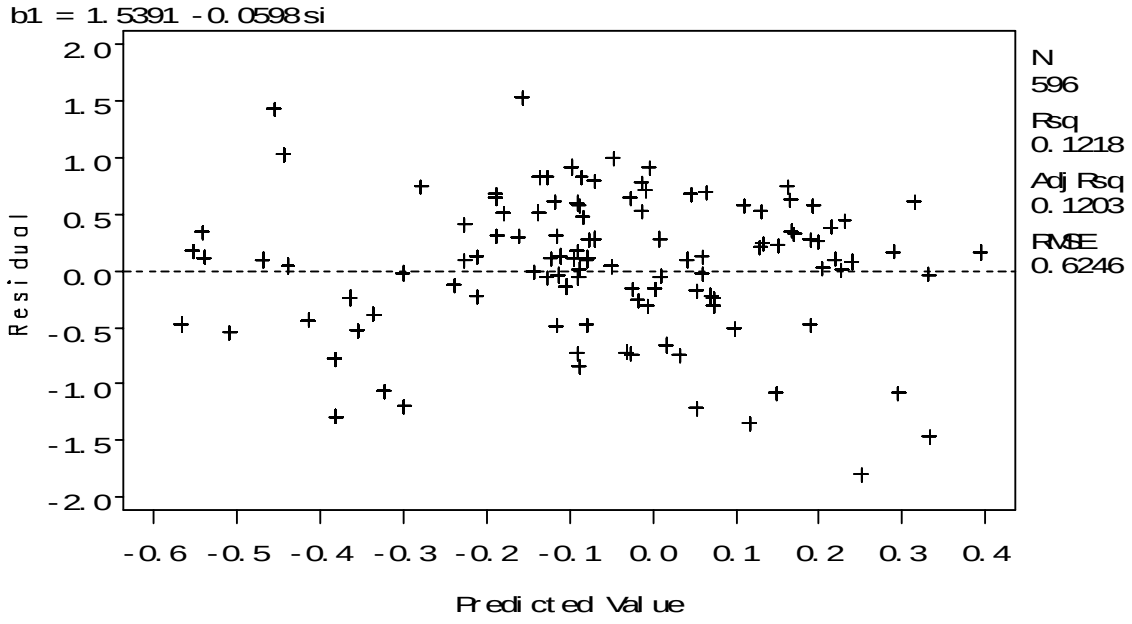

FIGURE 2.18. Parameters estimated, statistics and residual distribution of linear relationship between Site Index and the random effect b_1.

precise in terms of projection. The signals of the difference had greater variation in positive and negative values, demonstrating a compensation in overestimation and underestimation of basal area, decreasing the final error of estimate for the population.

Discussion

The linear mixed-effects model generated precise estimates for both fixed (1/Age), and random effects (plot). More variability was found among different clones than among plots of the same clone, meaning it was necessary to include the clone effect in the analysis. The individual estimates from the linear mixed-effect model tended to be "pulled toward" ("shrinkage estimates") the fixed-effects estimates because mixed-effects estimates represent the effect of the individual fits and the fixed-effects estimates, associated with the population averages, giving a certain robustness to outlying plot behavior.

Although it was not verified that there were problems with residual distribution within plots, the heteroscedasticity among plots was modeled and the information criteria statistics and likelihood values had a significant improvement. Moreover, because the plots are from different locations with environmental variations, the spatial correlation pattern was modeled. Once more the information criteria statistic and likelihood values showed significant improvement. Semivariogram and normal plots by sample unit showed the superiority of the heteroscedastic-autocorrelation model.

The random parameters estimated in the mixed-effects model were related to site index of each plot, and two linear regression equations were generated to estimate these

TABLE 2.5. Basal-area estimated for new plots using mixed-effects model.

Clone	Age(years)	Dominant Height (meters)	Basal Area (m2/hectare)	Dominant Height in Age 7 (meters)	Predicted Basal Area (m2/hectare)	% Difference
HUGA45	2.50	17.50	11.0900	29.24	11.6738	-5.3
HUGA45	2.50	18.50	13.7600	30.24	12.1142	12.0
HUGA45	2.50	15.50	12.2000	27.24	10.8405	11.1
EGRBS1	2.50	16.75	9.1500	28.49	11.3541	-24.1
EGRBS1	2.50	18.75	12.3200	30.49	12.2269	0.8
EGRBS1	2.50	18.75	12.4900	30.49	12.2269	2.1
EGRBS1	2.50	21.25	12.8200	32.99	13.4128	-4.6
EGRBS1	2.50	18.00	14.0500	29.74	11.8920	15.4
EGRBS1	2.50	15.50	11.6000	27.24	10.8405	6.5
EGRBS1	2.50	19.25	12.9900	30.99	12.4554	4.1
1172	2.17	13.00	8.4900	26.37	8.9981	-6.0
1172	2.17	15.25	13.0100	28.62	9.7002	25.4
1172	2.17	14.75	10.1400	28.12	9.5396	5.9
1172	2.17	14.25	8.9800	27.62	9.3817	-4.5
1172	2.17	12.75	8.0600	26.12	8.9233	-10.7
6024	2.17	13.50	8.0100	26.87	9.1496	-14.2
6024	2.17	13.00	6.8400	26.37	8.9981	-31.6
6024	2.17	14.75	8.7500	28.12	9.5396	-9.0
6024	2.17	14.00	8.3200	27.37	9.3037	-11.8
3902	2.67	14.75	9.1300	25.76	10.9359	-19.8
1052	2.67	18.25	14.4600	29.26	12.5158	13.4
1052	2.67	18.25	16.6000	29.26	12.5158	24.6
1052	2.67	18.00	13.2600	29.01	12.3957	6.5
1052	2.67	18.00	14.7500	29.01	12.3957	16.0
1052	2.67	18.50	15.1200	29.51	12.6370	16.4
1052	2.67	17.00	13.4000	28.01	11.9269	11.0
1052	2.67	16.25	10.9800	27.26	11.5870	-5.5
3902	2.67	17.25	13.3200	28.26	12.0424	9.6
3902	2.67	15.00	10.2500	26.01	11.0419	-7.7
3902	2.67	14.25	8.8800	25.26	10.7272	-20.8
0010	2.67	16.50	12.7100	27.51	11.6992	8.0
3902	2.67	14.50	8.3600	25.51	10.8310	-29.6
3902	2.67	18.00	15.5600	29.01	12.3957	20.3
0010	2.67	17.00	11.7600	28.01	11.9269	-1.4
0010	2.67	18.00	13.1100	29.01	12.3957	5.4
0010	2.67	16.50	12.6600	27.51	11.6992	7.6
0010	2.67	16.25	14.6400	27.26	11.5870	20.9
MISTSE	2.00	14.00	8.7400	28.29	8.6489	1.0
MISTSE	2.00	13.25	8.7200	27.54	8.4498	3.1
MISTSE	2.00	12.75	9.7900	27.04	8.3196	15.0

TABLE 2.5. Continued.

MISTSE	2.00	13.00	8.0200	27.29	8.3845	-4.5
EURCHM	2.00	12.75	8.4000	27.04	8.3196	1.0
EURCHM	2.00	13.50	8.9600	27.79	8.5156	5.0
EGRANH	2.00	14.00	9.2100	28.29	8.6489	6.1
EGRBS1	2.00	14.00	8.9400	28.29	8.6489	3.3
EGRBS1	2.00	14.50	9.3800	28.79	8.7842	6.4
EGRBS1	2.00	13.50	7.2800	27.79	8.5156	-17.0
EGRBS1	2.00	13.75	7.2200	28.04	8.5820	-18.9
3908	2.83	20.00	16.6600	30.33	13.8306	17.0
3908	4.08	27.75	20.5300	33.91	22.0304	-7.3
3908	4.92	28.75	23.5800	32.77	23.6094	-0.1
3908	5.83	31.75	24.8300	33.84	27.3406	-10.1
3908	7.00	33.50	28.1800	33.50	29.2526	-3.8
3908	8.08	34.75	28.4800	33.11	30.3133	-6.4
3908	8.92	36.25	29.2800	33.49	32.0217	-9.4
3908	9.83	36.75	30.9300	32.87	31.9157	-3.2
3908	10.67	37.25	31.3700	32.45	31.9205	-1.8
1255	3.50	16.50	11.9900	24.40	12.8597	-7.3
1255	4.33	22.25	15.4500	27.73	17.1727	-11.2
1255	5.42	24.50	19.1000	27.42	19.0916	0.0
1255	7.50	26.50	18.7500	25.71	19.9148	-6.2
1255	8.42	27.50	19.7600	25.40	20.3095	-2.8
1255	9.17	29.00	21.2300	25.92	21.4008	-0.8
1165	3.50	24.00	17.1000	31.90	17.8692	-4.5
1165	4.33	24.75	19.0000	30.23	19.3206	-1.7
1165	5.42	25.50	21.5100	28.42	20.0688	6.7
1165	7.50	27.00	22.4400	26.21	20.4493	8.9
1165	8.42	28.00	23.2900	25.90	20.8638	10.4
1165	9.17	30.00	23.9600	26.92	22.5979	5.7
209	3.50	20.50	13.1800	28.40	15.3261	-16.3
209	4.33	22.25	16.0300	27.73	17.1727	-7.1
209	5.42	23.00	18.0500	25.92	17.7144	1.9
209	7.50	24.50	19.5800	23.71	17.9127	8.5
209	8.42	25.25	20.1700	23.15	17.9921	10.8
209	9.17	27.00	21.6800	23.92	19.1935	11.5

parameters based upon site information. The fitted equations were tested in various new plots representing different clones. The method generated precise results with prediction error in general of less than 10 percent.

CHAPTER 3

NONLINEAR MIXED-EFFECTS MODEL APPLIED TO CLONAL *Eucalyptus* PLANTATIONS

Introduction

Similar to the linear case, nonlinear mixed-effects models also have been applied in modeling an assortment of situations in recent years. Due to the nonlinear characteristics of growth curves and the variation among random subjects, the nonlinear mixed-effects model approach is a rich tool for modeling such curves. In addition, the biological interpretability of both the fixed and random parameters make this approach directly connected with population feature.

In forestry, the application of the nonlinear mixed-effect theory has been increasing in recent years. In one of the pioneer studies, Lappi and Bailey (1988) used the nonlinear Richards equation to model tree height with random stand and tree parameters. Also, Walters et al. (1991), Gregoire and Schabenberger (1995), Fang (1999) and Hall and Bailey (2001) have published studies based on this approach.

The main purpose of this study is to fit the logistic equation to model height growth of eucalypt plantations using random and fixed parameters, accounting for variation within plots and by clone and modeling the autocorrelation within subject.

Generalized Nonlinear Mixed-Effects Models

As in linear case, we will follow the methodology presented by Davidian and Giltinan (1995) and Lindstrom and Bates (1990) to present the basic framework features about nonlinear mixed-effect models. Again, we will consider the response variable y_{ij} representing the group i (or plot i in our case) measured over time j. So, i=1,…,m, j=1,…,n_i, m is the total number of plots and n_i is the number of repeated measurents of the response variable. In a balanced data set, $n_1=n_2=\ldots=n_m$. We can use the nonlinear function $y_{ij} = f(\phi_{ij}, \upsilon_{ij}) + \varepsilon_{ij}$ to represent the relationship between the response variable and the covariates within the ith group, where f is a general, real-valued, differentiable function of a group specific parameter vector ϕ_{ij} and a covariate vector υ_{ij}, and ε_{ij} is a normally distributed within-group error term. The function f has to be nonlinear in at least one component of the group-specific parameter vector ϕ_{ij}, which has the form

$$\phi_{ij} = A_{ij}\beta + B_{ij}b_i, \qquad b_i \approx N(o, \Psi), \tag{3.1}$$

where β is a (p x 1) vector of fixed effects, b_i is a (q x 1) vector of random effects associated with the ith group, and A_{ij} and B_{ij} are incidence matrices . The assumptions are the same as those in the linear case. The within-group errors are independently distributed with mean zero and variance σ^2 and independent of the random effects.

Further we will use the logistic equation with three parameters to model the dominant height growth pattern as a function of age. If we assume that the final model will have three fixed and three random effects, the model and the matrices of the expression $\phi_{ij} = A_{ij}\beta + B_{ij}b_i$ will have the following format:

$$\mathbf{HD_{ij}} = \frac{\boldsymbol{\phi}_{1i}}{1+\exp[-(\mathbf{a_{ij}} - \boldsymbol{\phi}_{2i})/\boldsymbol{\phi}_{3i}]} + \boldsymbol{\varepsilon}_{ij} \qquad (3.2)$$

$$\begin{bmatrix} \boldsymbol{\phi}_{1i} \\ \boldsymbol{\phi}_{2i} \\ \boldsymbol{\phi}_{3i} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \boldsymbol{\beta}_3 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{b}_{1i} \\ \mathbf{b}_{2i} \\ \mathbf{b}_{3i} \end{bmatrix}$$

$$\mathbf{b}_i \sim N\left(0, \begin{bmatrix} \boldsymbol{\psi}_{11} & \boldsymbol{\psi}_{12} & \boldsymbol{\psi}_{13} \\ \boldsymbol{\psi}_{12} & \boldsymbol{\psi}_{22} & \boldsymbol{\psi}_{23} \\ \boldsymbol{\psi}_{31} & \boldsymbol{\psi}_{32} & \boldsymbol{\psi}_{33} \end{bmatrix}\right), \varepsilon_{ij} \sim N\ (0, \sigma^2).$$

Where HD represents Dominant Height, which is the mean of the 100 trees of largest diameter per hectare, a_{ij} is the age of plot **i** in time **j** and $\boldsymbol{\phi}_{kj}$ are parameters to be estimated.

The approach in modeling variance and correlation structures is basically the same as for the linear mixed-effects model. Details can be find in Pinheiro and Bates(2000), Davidian and Giltinan(1995) and Littell et al. (1999).

Data

The data set, the same used for the basal-area study, is from permanent plots of *Eucalyptus* hybrid, propagated as clones, cultivated in the Brazilian Coastal Region, in Bahia and Espírito Santo states. Each plot was measured from 3 to 10 times, between 1992 and 2001. So, the data set is represented by longitudinal, irregularly spaced and unbalanced plots.

The response variable for each plot will be Dominant Height (HD) and the covariate associated with the response variable will be the Age(A). In Figure 3.1 we can see a distinct nonlinear increase of HD with Age. Also, this trend has different patterns

for different Plot x Clone combinations, including the situation in which the number of measurements was not sufficient for HD to approach an asymptote. We will further examine the robustness of the mixed-effects model for modeling this situation by using auxiliary information from other plots to model insufficient repeated measurements.

Some variation between Plot x Clone combinations are due to site quality and/or genetic differences between clones, leading to different curve shapes (polymorphism). In Figure 3.1 it is clear that some plots have lower asymptotes and/or early inflection points, representing a polymorphic system of site equations. Therefore, the focus is to model the fixed and random parameters to generate a family of polymorphic site equations, that will represent the site and clone variations.

Fitting Nonlinear Mixed-Effects Models

As seen in Figure 3.1, nonlinear trends and the variation among plots is the primary motivation to consider a nonlinear model. Similar to the linear mixed-effects model, we will use a single level mixed-effects model where the response variable is dominant height (HD).

We have noted the variations among plots to justify the use of the mixed-effects model. After trying some models to fit the data set, such as Gompertz, Richards and Weibull-type models, we chose the logistic model, which proved to be precise and flexible in this case:

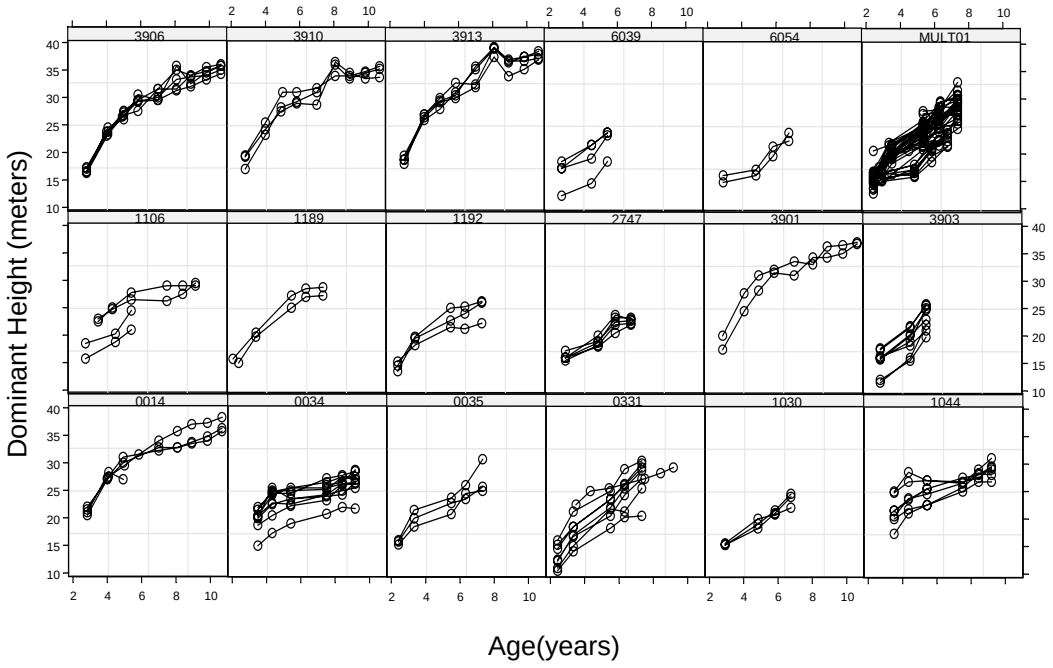

FIGURE 3.1. Dominant height (HD) growth for different plots and clones

$$\mathbf{HD_{ij}} = \frac{\boldsymbol{\phi}_{1\mathbf{i}}}{1+\exp[-(\mathbf{a_{ij}} - \boldsymbol{\phi}_{2\mathbf{i}})/\boldsymbol{\phi}_{3\mathbf{i}}]} + \boldsymbol{\varepsilon}_{\mathbf{ij}} \qquad (3.3)$$

Where:

HD_{ij} = Dominant Height (meters) for i-th plot on time j;

a_{ij} = Age(years) for i-th plot on time j;

ε_{ij} = Random error;

$$\boldsymbol{\Phi}\mathrm{i} = \begin{bmatrix} \boldsymbol{\phi}_{1\mathbf{i}} \\ \boldsymbol{\phi}_{2\mathbf{i}} \\ \boldsymbol{\phi}_{3\mathbf{i}} \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} \mathbf{b}_{1\mathbf{i}} \\ \mathbf{b}_{2\mathbf{i}} \\ \mathbf{b}_{3\mathbf{i}} \end{bmatrix} = \boldsymbol{\beta} + \boldsymbol{b_i}$$

$\boldsymbol{b_i} \sim N(0,\Psi)$ and $\varepsilon_{ij} \sim N(0,\sigma^2)$

Here, $\boldsymbol{\beta}$ is a vector of fixed effects and $\boldsymbol{b_i}$ represents the vector of random effects. $\boldsymbol{\varepsilon_{ij}}$ and $\boldsymbol{b_i}$ are independents. ϕ_{1i} is the horizontal asymptote as age goes to infinity. ϕ_{2i} is the age value at which the response is $\phi_{1i}/2$. It is the inflection point of the curve. ϕ_{3i} is the scale parameter, which represents the distance on the x-axis between the inflection point and the point where the response is $\phi_{1i}/(1+e^{-1}) = 0.73\,\phi_{1i}$. In the basic assumptions, the random effects are normally distributed and independent for different groups and the within-group errors are independent and identically normally distributed and independent of the random effects.

Ignoring the grouping of the dominant height measurements and the random effect, the model (3.1) was fitted using the entire data set, using standard nonlinear least squares methodology. The boxplots in Figure 3.2 show that the residuals tend to be mostly negative for some plots, positive for others, and the plots have different variations. If the model (3.1) is adjusted for each plot separately (Figure 3.3), the pattern for each plot is evident. Some gaps in Figure 3.3 are because the plot does not have enough observations to fit the model. When the model was fitted as a fixed effect model for all plots together, the residual standard error was 4.38646. In contrast, the residual standard error had value 1.565119 when the model was adjusted for each plot separately, indicating that adjustment for each plot accounted for the Plot x Clone effect.

The drawback in fitting each plot separately is that we had an over-parameterized model. In our case, we have 596 plots with 3 parameters per plot and the model does not take into account similarities among plots and variability among and within individuals.

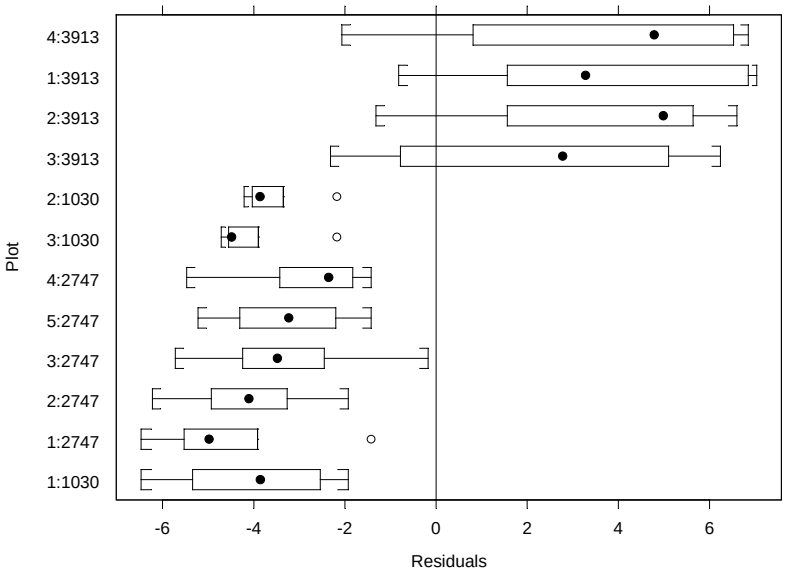

FIGURE 3.2. Residual boxplots representing plot:clone, fitted with part of the data set

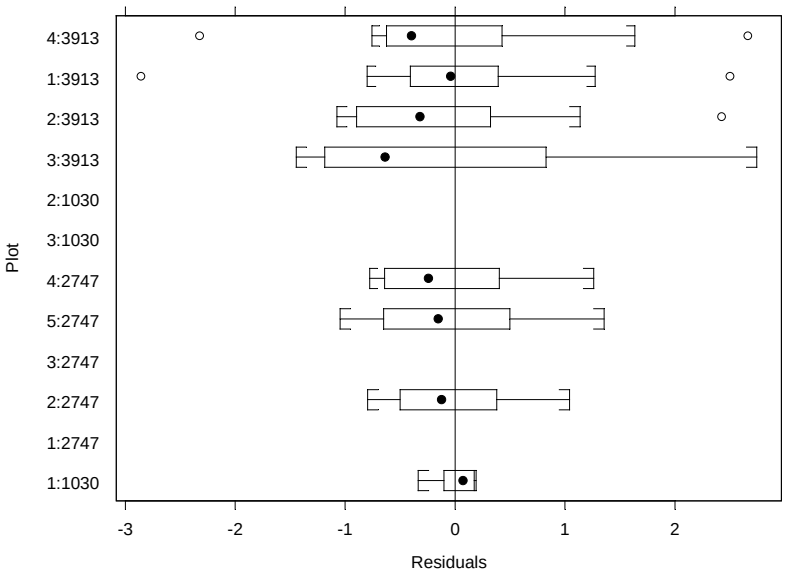

FIGURE 3.3. Boxplots of the residuals fitted by subject

The mixed-effects models were designed to consider these and other characteristics of the group analysis.

The mixed-effects model (3.1) was fitted using our data set (Table 3.1). The large estimate for the standard deviation for the three random effects suggests that they are needed in the model. The AIC, BIC and LogLikelihood values were 2467.838, 2511.740 and -1223.919, respectively. Based on the relatively high correlation between the ϕ_{1i} and ϕ_{2i} parameter estimates, we tried to eliminate the ϕ_{2i} random parameter, keeping it just as a fixed effect to avoid ill-conditioning problems associated with the variance-covariance matrix and over-parameterized random-effects. The AIC, BIC and LogLikelihood values were 2470.304, 2518.597 and -1224.152, indicating that the random effect is needed in the model.

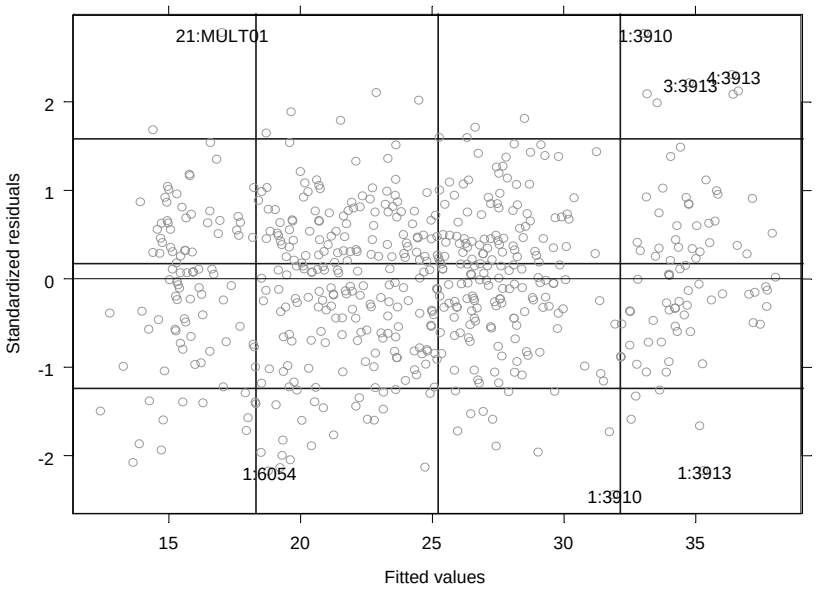

FIGURE 3.4. Scatter plots of standardized residuals versus fitted values

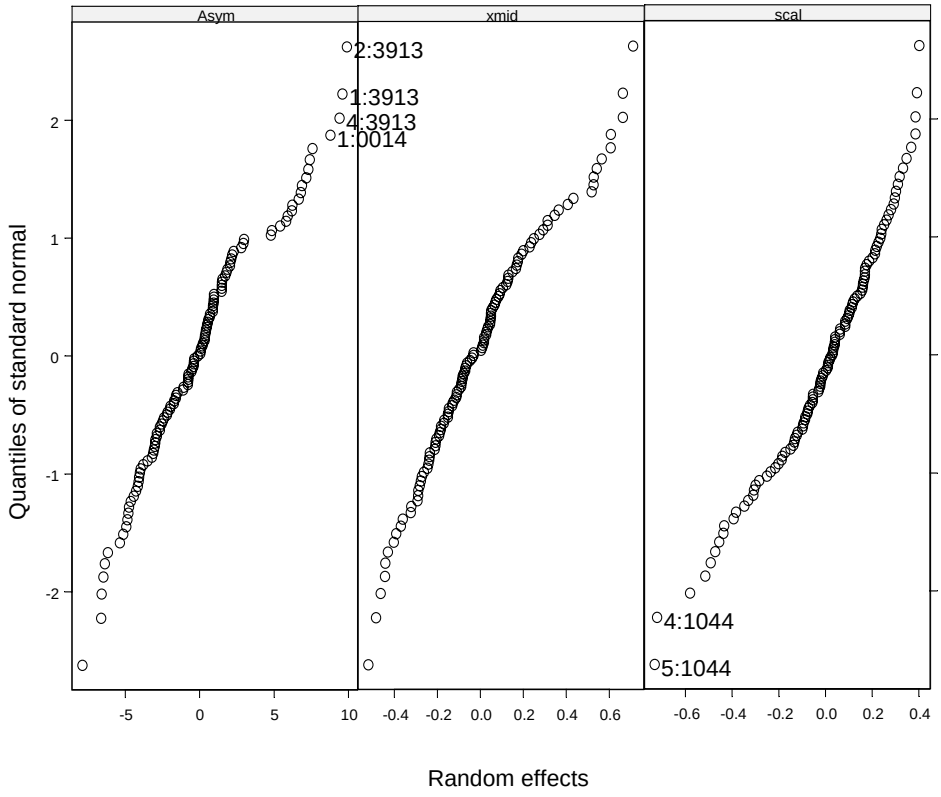

FIGURE 3.5. Normal plot of estimated random effects for the fitted model

TABLE 3.1. Information of fixed and random effects and correlation

	Fixed Effect		Random Effect	Correlation	
Parameter	Value	Stand. Error	Stand. Error	ϕ_2	ϕ_3
ϕ_1	28.62547	0.4511514	4.157212	0.807	0.017
ϕ_2	2.12669	0.0521297	0.362975	x	0.604
ϕ_3	1.82543	0.0714603	0.337231	x	X

Checking Distributional Assumptions

Similar to the linear mixed-effects model, the nonlinear distributional assumptions require that the random effects are independent and normally distributed with mean zero and variance-covariance matrix $\mathbf{\Psi}$. The within-group errors ε_{ij} are independent normally distributed with mean zero and variance σ^2 and independent of the random effects.

The plot of the standardized residuals versus the fitted values, presented in Figure 3.4, shows that the residuals are distributed symmetrically around zero, with approximately constant variance. This does not indicate any departure from the assumptions for within-group error. With the exception of some possible outliers, the homoscedastic model provides a good representation of the data. Some heteroscedastic models were fit with variable variance-covariance structure but the values of AIC, BIC and LogLikelihood did not show significant improvements.

The distribution of the random effects can be examined in Figure 3.5. The result does not indicate any serious violation of the assumption of normality for random effects, with some outliers present on the ϕ_{1i} and ϕ_{3i} parameters.

The adequacy of the fitted model can be visualized in Figure 3.6, representing 36 plots from a total of 115. Both the population predictions, with random effects set to zero, and the within-group predictions can be compared. The plot-specific estimates are close to the observed values, indicating that the logistic mixed-effects model adequately represents the dominant height growth data.

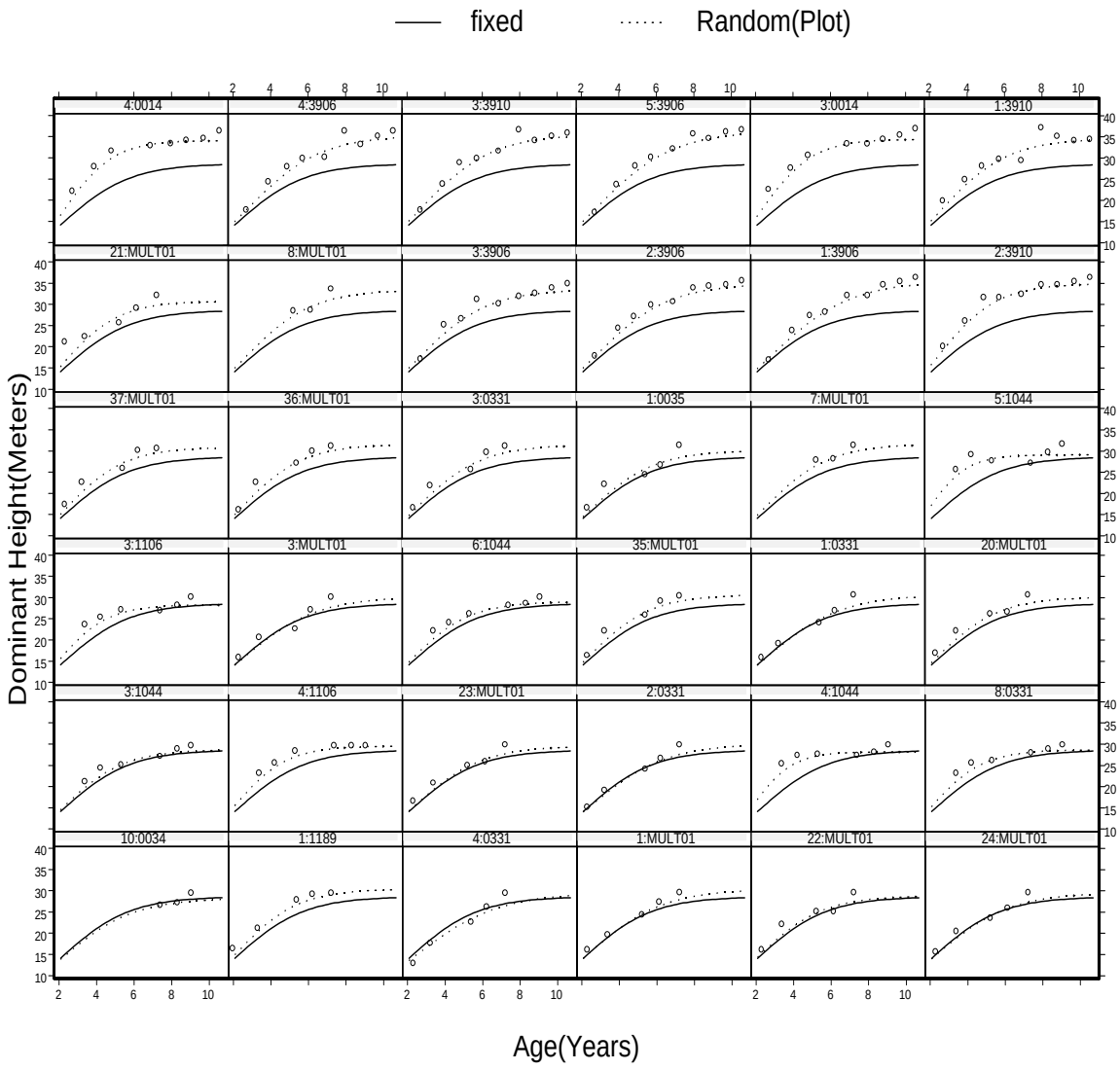

FIGURE 3.6. Observed and fitted dominant height (HD) as a function of age for fixed and random (plot) effects

Plots like 4:1044 and 5:1044, for example, reach the asymptote early. Such trends were described only by the mixed-effects model.

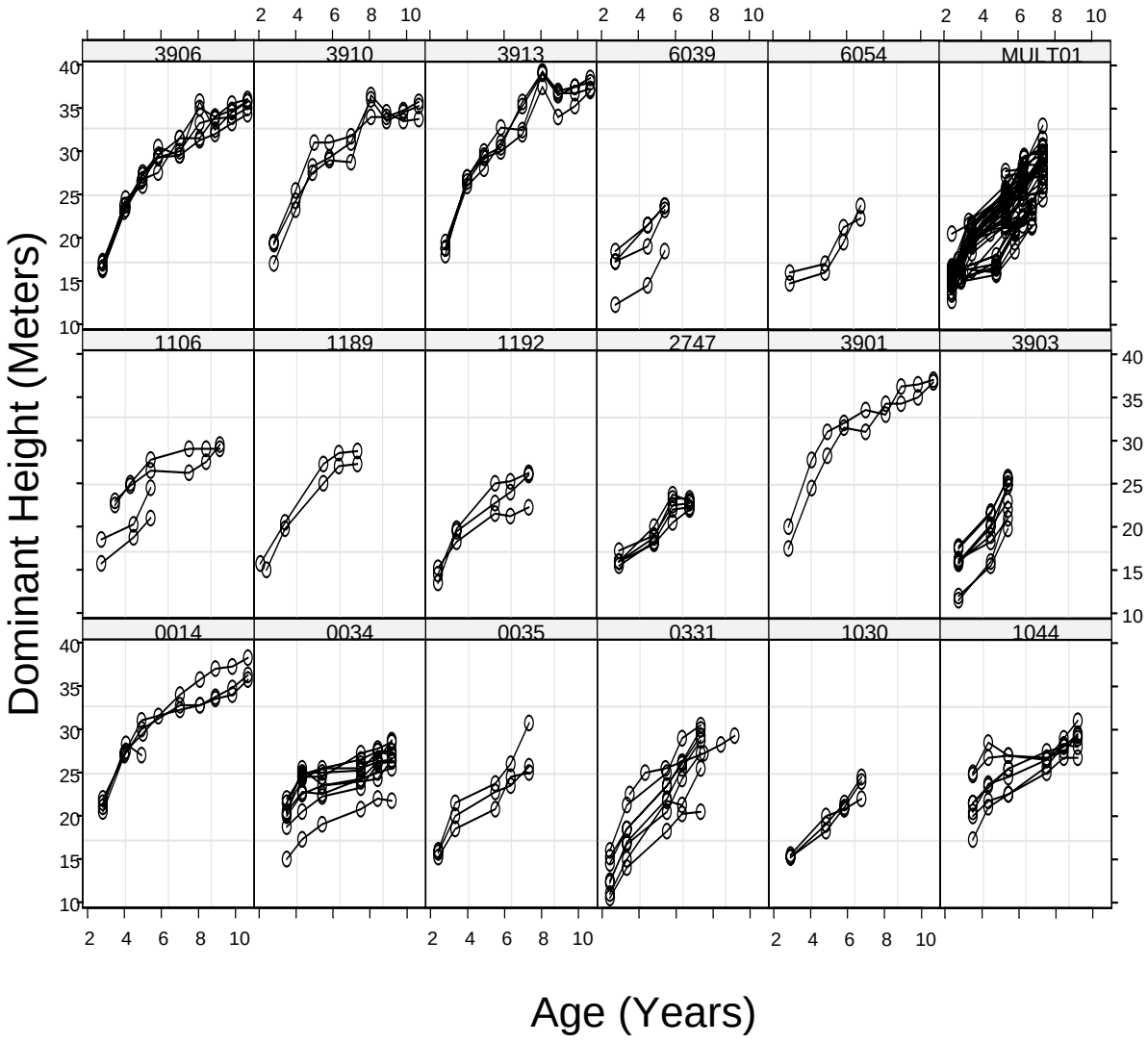

FIGURE 3.7. Observed values of dominant height by age and plot

Figures 3.7 and 3.8 show the observed and estimated curves for different combinations of clone/plot. Figure 3.8 represents an assortment of shapes for dominant height growth, including anamorphic and polymorphic curves. Thus, we had strong variations in both fixed and random effects for the three parameters of the logistic equation. This suggests a biological interpretation for the parameters.

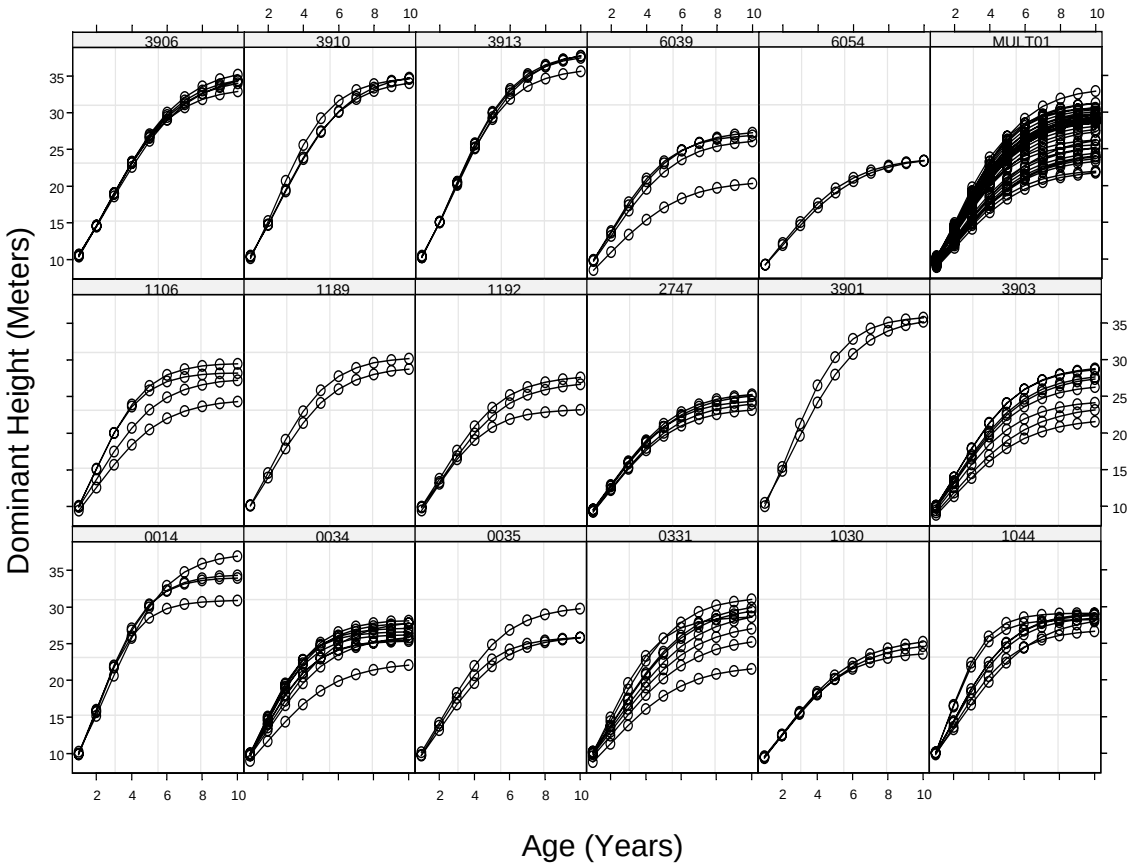

FIGURE 3.8. Fitted profile for dominant height as a function of age

Biological Interpretation to Fixed and Random Effects

Biological interpretations can be given to the parameters of the logistic equation (3.3), where ϕ_1 is the asymptote parameter (maximum height), ϕ_2 is the inflection point (age at which ½ of maximum height occurs) and ϕ_3 is the scale parameter (distance in years from the inflection point to the point where the height is 73% of the maximum

height). The intention here is to provide a biological interpretation for fixed and random effects which generate these parameters. Figure 3.9 represents the dominant height growth curves for the clones labeled 0014, 1044, and 3903. The classification of the curves are either anamorphic or polymorphic and, sometimes, “quasi-anamorphic”. In clone 3903, the different curves have nearly the same pattern and the curves with a greater asymptote parameter have smaller middle response and scale parameters. There is a high correlation between the asymptote and the other two parameters. In this case, for a new observation representing this clone, we could estimate the asymptote parameter based on an early inventory which will give us information about ϕ_2 and ϕ_3. The same situation is not clear when we analyze the height growth profiles for clones 0014 and 1044. Due to nondisjoint polymorphism of these curves, sometimes a curve with smaller ϕ_2 and ϕ_3 parameters, which indicate good site, does not have a larger value for the ϕ_1 parameter. Thus, if a new observation, with height/age information for an early age, is projected in this type of curve, the curve generated could have two or more different shapes, generating a projection problem. Here, the problem is more evident because the analyses were carried out based on plots with subjects (trees) with the same genotype and the variations are due to environment only.

Clutter (1983) commented that the solution to this problem is to include other variable(s) than height and age in the site index system. Studies such as Zahner (1962) and Newberry and Pienaar (1978) include soil categories as a discrete variable to explain the polymorphism of the curves. However, as pointed out by Clutter (1983), when the additional variables are quantitative and continuous, there is no alternative to viewing the curve system as being polymorphic-nondisjoint. Thus, one could include quantitative

soil chemical and physical information as covariates and fit a system more precisely. Other information, such as precipitation, site preparation, fertilization, and other silvicultural treatments would be useful as covariates to explain the anamorphism/polymorphism behavior of the dominant height growth curves.

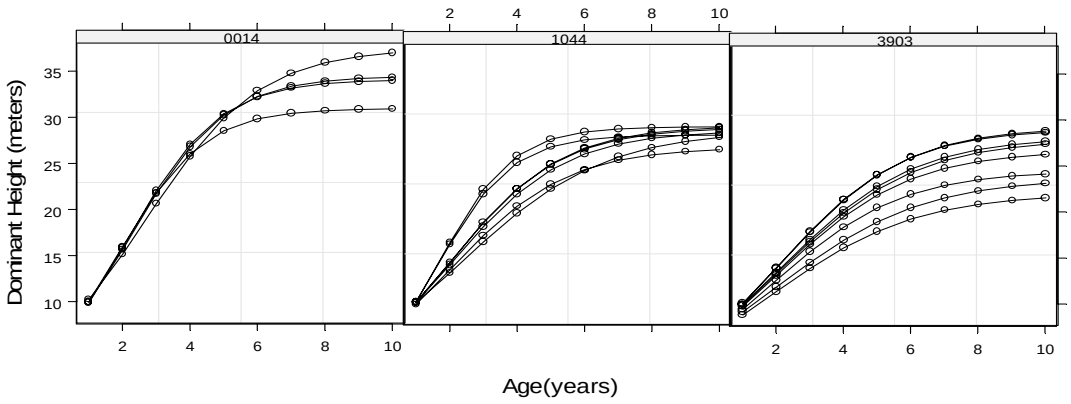

FIGURE 3.9. Dominant height growth for three different clones

Discussion

The nonlinear mixed-effects model, represented here by the logistic equation with three parameters, has considerable flexibility in represent the dominant height growth pattern for eucalypt clones, generating either an anamorphic (or “quasi-anamorphic”) or polymorphic sets of site index curves, depending on the combination of clone and environment.

Using information statistics (AIC and BIC), logLikelihood and correlation analysis, we conclude that the logistic model with three fixed and three random effects

does not generate an ill-conditioned variance-covariance matrix and, consequently, that the model is not over-parameterized.

The analysis of the scatter plot of the standard within-group residuals shows a uniform distribution around zero with approximately constant variance, indicating that the homoscedastic model provides a good representation of the data. Also, the assumptions related to random effects was examined and it was concluded that the three parameters are approximately normally distributed, with no serious violation of this assumption.

The fitted values representing 115 plots confirmed the adequacy of the model to represent the data set, with strong variation among the three random parameters, indicating that the plots have different asymptotes, inflection points and scale.

Due to the differences among the random parameters of the logistic model, within plots and among plots, the system of site curves generated varied substantially across location and clones, with anamorphism or polymorphism, depending on the site/clone combination.

CHAPTER 4

LINEAR AND NONLINEAR MULTILEVEL MIXED-EFFECTS MODEL APPLIED TO *Eucalyptus* PLANTATIONS

Introduction

Multilevel mixed-effects models are based on nested classification factors. In Chapters 2 and 3, the mixed effects model was developed with just one level, which was plot. Here, supported by the availability of the individual tree data set, the focus will be on modeling two levels: plot (level 1) and trees within plot (level 2). Based on definitions by Daniels and Burkhart (1988), this situation would have an intermediate level of resolution between distance-independent and distance-dependent individual-tree models. Although the data sets have a distance-independent level of resolution, the modelling approach including random effects and spatio-temporal process could be considered a superior level of resolution compared to distance-independent level without these approaches.

In forestry, multilevel mixed-effects model, taking in to account the linear and the nonlinear multilevel approach, was fitted by Fang(1999) to model the growth of slash pine with different cultural treatments.

Again, we do not have information about applying multilevel mixed-effects approach to model *Eucalyptus* stand growth. The general purpose of this study is to assess linear and nonlinear two-level (plot and tree within plot) mixed-effects approach to model basal area and dominant height growth.

Generalized Multilevel Linear Mixed-Effects Models

The formulation presented in chapter 2 for single-level linear mixed-effects models can be extended to the two nested levels case. The response variable y_{ijk} will be measured for the first level i, second level j and occasion (time) k. The general model has the follow formulation:

$$\mathbf{y}_{ij} = \mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{Z}_{i,j}\mathbf{b}_i + \mathbf{Z}_{ij}\mathbf{b}_{ij} + \boldsymbol{\varepsilon}_{ij} \qquad (4.1)$$

Where $\mathbf{X}_{ij}$ are matrices of fixed effects with dimension (n_{ij} x p); $\boldsymbol{\beta}$ is a vector of the fixed parameters with dimension (p x 1); $\mathbf{Z}_{i,j}$ are matrices with dimension (n_i x q_1) associated with the first level random effects $\mathbf{b}_i$; $\mathbf{Z}_{ij}$ are matrices of dimension (n_i x q_2) associated with the second level random effects $\mathbf{b}_{ij;}$ and $\boldsymbol{\varepsilon}_{ij}$ are the error terms. The assumptions are the same as those for single level model.

Generalized Multilevel Nonlinear Mixed-Effects Models

In the single level nonlinear approach we had two random components in the model which were represented by the random error within group (e_{ij}) and the subject effect (b_i). In the multilevel nonlinear approach the model is extended including more

than one level of random subject effects and these effects are nested. For example, plots are random and trees are random and nested within plots, providing a case of two-level random effects model. The model will be represented by y_{ijk} which is the response variable of the kth observation in time on the jth second-level group (tree) and ith first-level group(plot). The expression representing this situation is

$$\mathbf{y_{ijk} = f(\phi_{ij}, \upsilon_{ijk}) + \varepsilon_{ijk}} \ . \qquad (4.2)$$

Where i=1,…,m, j=1,…,n_i, and k=1,…,n_{ij}. Again, f is a general, real-valued, differentiable function of a group specific parameter vector ϕ_{ij} and a covariate vector υ_{ijk}, and ε_{ijk} is a normally distributed within-group error term. The function f must be nonlinear in at least one component of the group-specific parameter vector ϕ_{ij}, which has the form

$$\boldsymbol{\phi}_{ij} = \mathbf{A}_{ijk}\boldsymbol{\beta} + \mathbf{B}_{i,jk}\mathbf{b}_i + \mathbf{B}_{ijk}\mathbf{b}_{ij}, \qquad \mathbf{b}_i \approx \mathbf{N}(0, \Psi_1) \ \textbf{and} \ \mathbf{b}_{ij} \approx \mathbf{N}(0, \Psi_2) \qquad (4.3)$$

where $\boldsymbol{\beta}$ is a (p x 1) vector of fixed effects and $\mathbf{b}_i$ is a (q_1 x 1) vector of random effects associated with the ith group, $\mathbf{b}_{ij}$ is a (q_2 x 1) vector associated with the second level random effects and assumed to be independent of the first-level random effects, and A_{ijk} and B_{ijk} are incidence matrices. The assumptions are the same as those in the linear case. The within-group errors are independently distributed with mean zero and variance σ^2 and are independent of the random effects.

The logistic equation with three fixed parameters will be used to model dominant height growth pattern as a function of age. Suppose that the final model will have three fixed and three random effects, the model and the matrices of the expression $\boldsymbol{\phi}_{ij} = \mathbf{A}_{ijk}\boldsymbol{\beta} + \mathbf{B}_{i,jk}\mathbf{b}_i + \mathbf{B}_{ijk}\mathbf{b}_{ij}$ will have the following format:

$$\mathbf{HD}_{ijk} = \frac{\boldsymbol{\phi}_{1ij}}{1+\exp[-(\mathbf{a}_{ijk} - \boldsymbol{\phi}_{2ij})/\boldsymbol{\phi}_{3ij}]} + \boldsymbol{\varepsilon}_{ijk} \tag{4.4}$$

$$\begin{bmatrix} \boldsymbol{\phi}_{1ij} \\ \boldsymbol{\phi}_{2ij} \\ \boldsymbol{\phi}_{3ij} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \boldsymbol{\beta}_3 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{b}_{1i} \\ \mathbf{b}_{2i} \\ \mathbf{b}_{3i} \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{b}_{1ij} \\ \mathbf{b}_{2ij} \\ \mathbf{b}_{3ij} \end{bmatrix}$$

$$\mathbf{b}_i \sim N\left(0, \begin{bmatrix} \psi_{111} & \psi_{112} & \psi_{113} \\ \psi_{112} & \psi_{122} & \psi_{123} \\ \psi_{131} & \psi_{132} & \psi_{133} \end{bmatrix}\right),$$

$$\mathbf{b}_{ij} \sim N\left(0, \begin{bmatrix} \psi_{211} & \psi_{212} & \psi_{213} \\ \psi_{212} & \psi_{222} & \psi_{223} \\ \psi_{231} & \psi_{232} & \psi_{233} \end{bmatrix}\right),$$

$$\varepsilon_{ij} \sim N\ (0, \sigma^2\ ).$$

Extensions to more than two levels are straightforward. The approach to modeling variance and correlation structures is basically the same given in the linear mixed-effects model.

Data

The data sets are from the same regions as those used in modeling basal area and dominant height. Here, we have data on individual trees, measured in plots, representing seven different genetic materials (clones). The total was 4289 observations, representing

254 trees and 8 plots. The ages ranged from 1.8 to 7.5 years with variations by plot (Figure 4.1). For example, the trees on plot 18764, representing the clone AR4, were measured monthly from 3.2 to 4.8 years, with a total of 18 repeated measurements. This research data set is atypical, compared with those generated by ordinary forest inventory in eucalypts plantations.

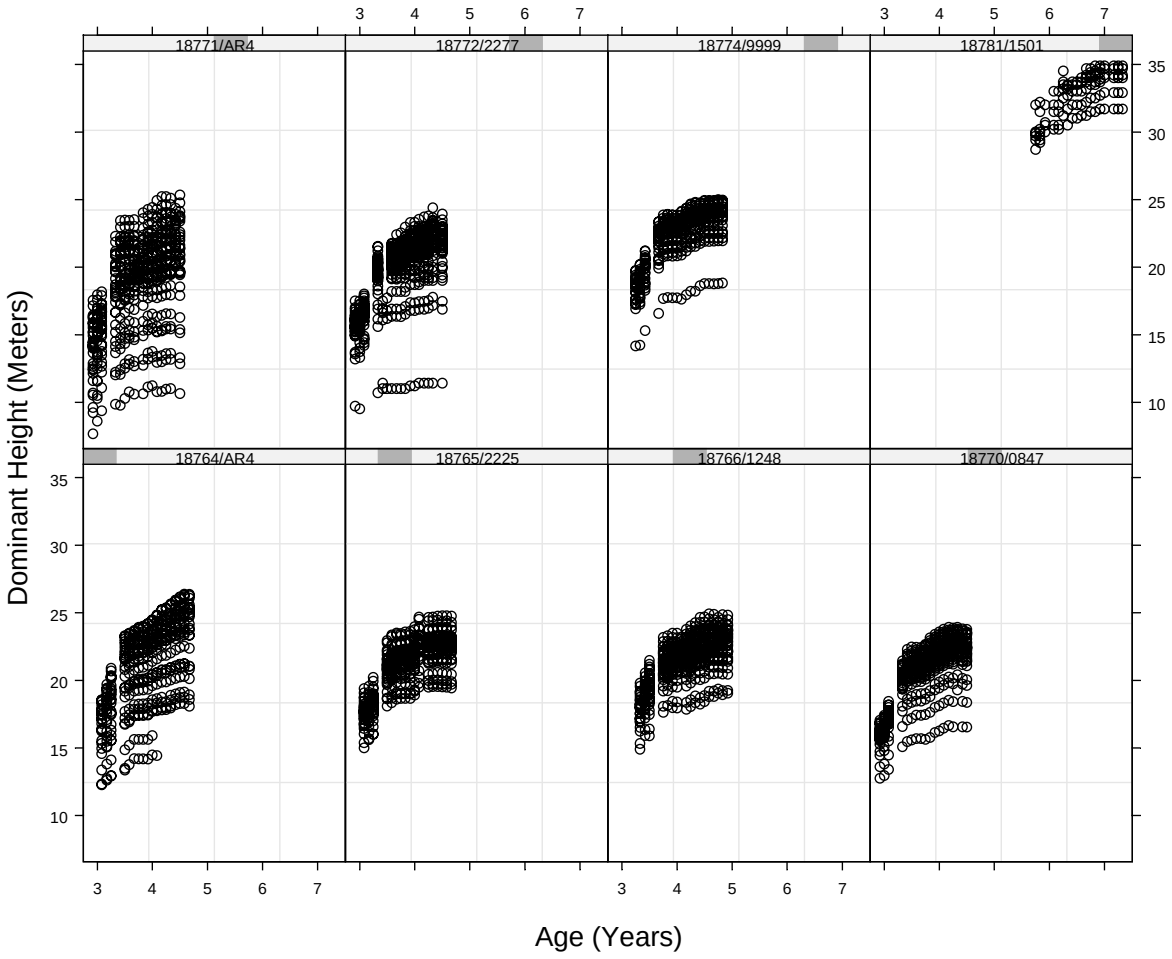

FIGURE 4.1. Nonlinear relationship of dominant height measured over time for individual trees, representing different combinations plot:clone.

Fitting A Linear Two Levels Mixed-Effect Model.

Based on the data set presented in first section and previous studies of modeling basal area yield, we used the ln(BA) as response variable and 1/age, ln(H) and their interaction as covariates. Where BA represents the basal area in square meters, age is in years and H is total height in meters. First, we considered a full mixed-effects model, with all terms having random effects at plot and tree within plot levels. We had 12362 observations, j=8 plots and number of trees i varying by plot. The two level model is presented in (4.1).

$$\mathbf{ln(BA)_{ijk} = (\beta_o + b_{oi} + b_{oij}) + (\beta_1 + b_{1i} + b_{1ij})\frac{1}{A_{ijk}} + (\beta_2 + b_{2i} + b_{2ij})ln(H)_{ijk} + (\beta_3 + b_{3i} + b_{3ij})\frac{1}{A_{ijk}}ln(H)_{ijk} + \varepsilon_{ijk}}$$

$$\mathbf{b}_i = \begin{bmatrix} b_{oi} \\ b_{1i} \\ b_{2i} \end{bmatrix} \sim N(0, \Psi_1),\ \mathbf{b}_{ij} = \begin{bmatrix} b_{oij} \\ b_{1ij} \\ b_{2ij} \end{bmatrix} \sim N(0, \Psi_2) \text{ and } \varepsilon_{ijk} \sim N(0, \sigma^2). \quad (4.5)$$

The parameters β_o, β_1, and β_2 are fixed effects and $\mathbf{b}_i$ is the plot-level random-effects vector, $\mathbf{b}_{ij}$ is the tree within plot-level random-effects vector, and ε_{ijk} is the within-group error. The $\mathbf{b}_i$ are assumed to be independent for different plots, the $\mathbf{b}_{ij}$ are assumed to be independent for different trees, different plots and independent of the $\mathbf{b}_{i.}$ $\boldsymbol{\varepsilon}_{ijk}$ are assumed to be independent for different plots, trees and observations, and independent of the random effects.

Due to the great number of observations, trees within plot and 20 variance-covariance components, the first mixed-effects model (4.1) was fitted assuming that $\boldsymbol{\Psi}_1$

and $\mathbf{\Psi}_2$ are diagonal matrices, which makes the optimization of the profiled log-restricted-likelihood more stable. The estimated parameters and standard deviations are presented in Table 4.1. The fixed effect had significant p-value (<0.05) with the exception of $\beta_{2,}$ which had a p-value=0.0603. The variable ln(H) was kept in the model because the interaction had significant p-value. The covariance components among random effects are assumed to be zero in a diagonal structure. The estimate for standard deviations for both levels were not small, meaning that these random effects should not be dropped from the model.

Table 4.1. Parameter estimates for fixed effects, random effects and standard deviations for multilevel linear mixed-effects model

	Fixed Effects					Random Effect Std. Deviations	
Parameter Estimated	Value	Std. Error	DF	t-value	p-value	Plot	Tree within Plot
β_o	-4.1151	0.084581	4032	-48.65	<0.0001	0.16484	0.45943
β_1	-19.3416	3.713849	4032	-5.20	<0.0001	10.40285	1.12316
β_2	-0.0442	0.023544	4032	-1.87	0.0603	0.03953	0.129426
β_3	5.5481	1.260866	4032	4.40	<0.0001	3.53865	0.318676
Residual	-	-	-	-	-	-	0.013350

Checking Distributional Assumptions

Again, the distributional assumptions are based on within-group errors (ε_{ijk}) and random effects ($\mathbf{b}_i$ and $\mathbf{b}_{ij}$). The within-group errors are independent and normally distributed with mean zero and variance σ^2 and the random effects are normally distributed with mean zero and covariance $\mathbf{\Psi}$ and are independent for different groups.

We can check these assumptions graphically with residual analysis. Figure 4.2 shows the within-group residual distribution, which indicates that the residuals are

approximately symmetrically distributed around zero, but may not have constant variance. Some outlying and/or influential observations are present. The same pattern can be seen for the standardized residual distribution by plot (Figure 4.3). Even though the total residuals are symmetrically distributed, the residuals by plot have different patterns, which violates the assumption of homoscedasticity. So, it was necessary to model this pattern.

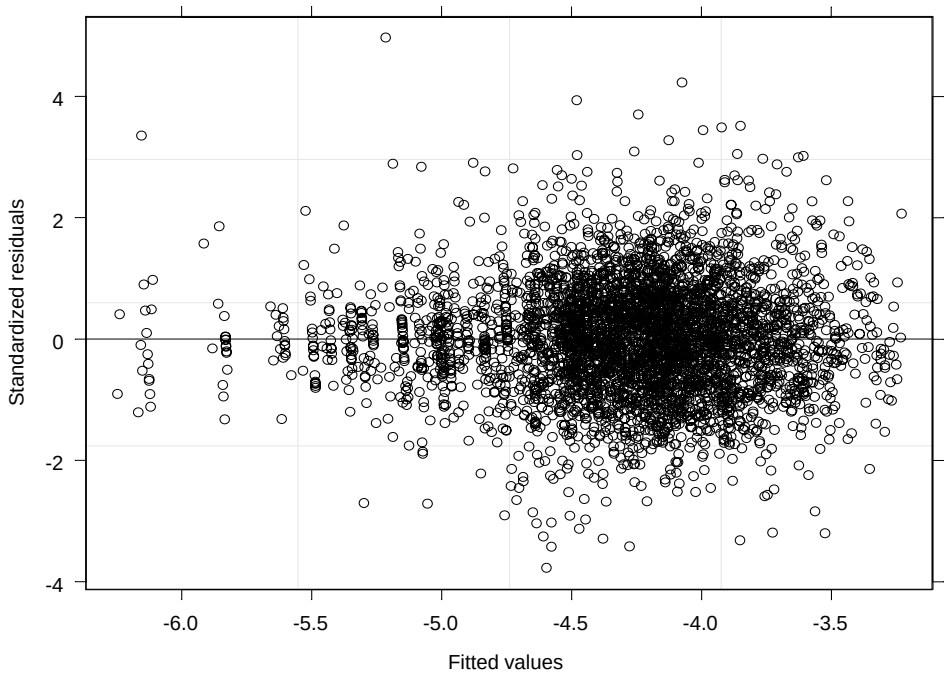

FIGURE 4.2. Standardized residuals versus fitted values for the linear multilevel model

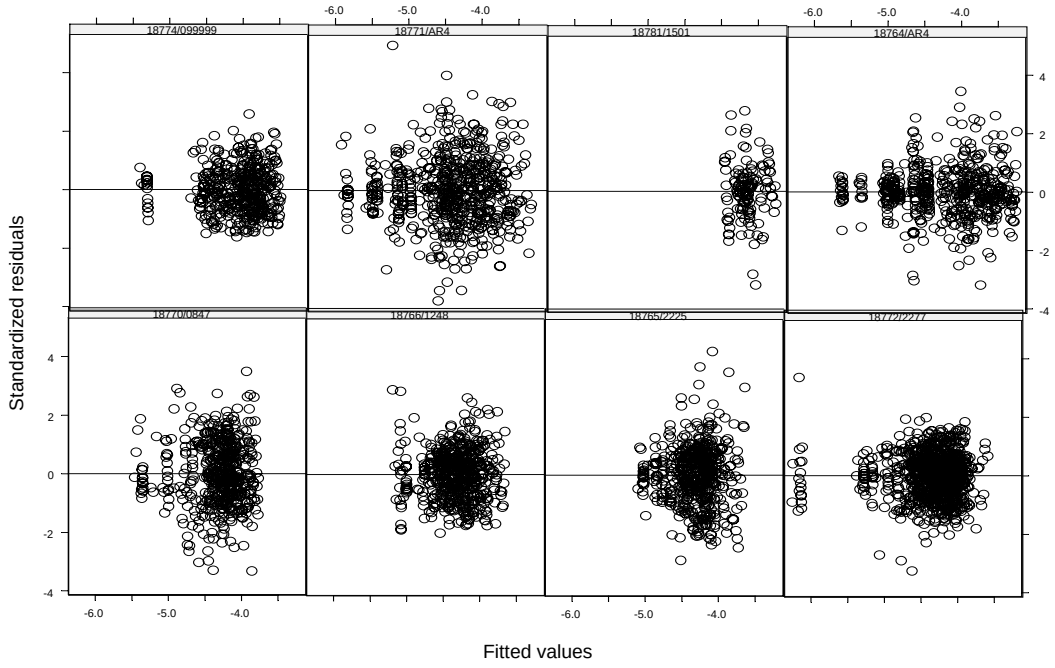

FIGURE 4.3. Standard residual by plot for linear multilevel model

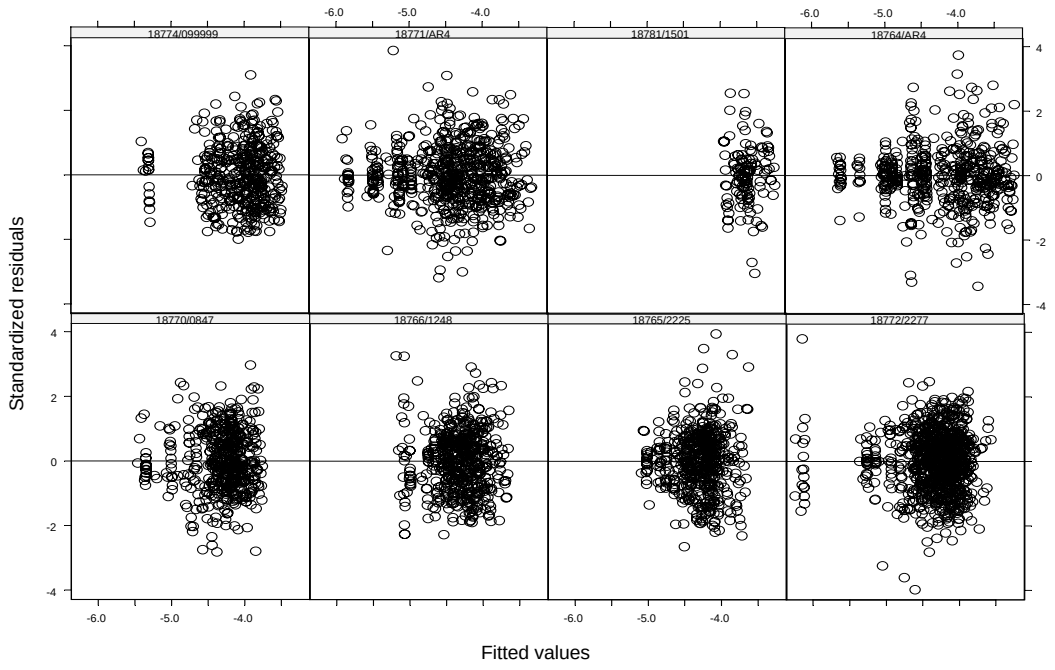

FIGURE 4.4. Standard residuals by plot for linear multilevel model after modeling variance among plots (heteroscedastic model)

The residual distributions of the heteroscedastic model are showed in Figure 4.4. The distribution among plots are more similar and have about the same variability. The logLikelihood value increased from 10,261 to 10,362, generating a Ratio-Likelihood value of 203.15 with p-value < 0.0001, indicating that the heteroscedastic model explains the data significantly better than the homoscedastic model.

Next we assess the assumptions on the random effects. Model (4.1) was initially fitted using the diagonal structure for variance-covariance matrix, assuming that the random effects are independent. We can see in Figure 4.5 that the random effects for level 1 (plot) are not independent, with strong correlation between b_{1i} (1/Age) and b_{3i}

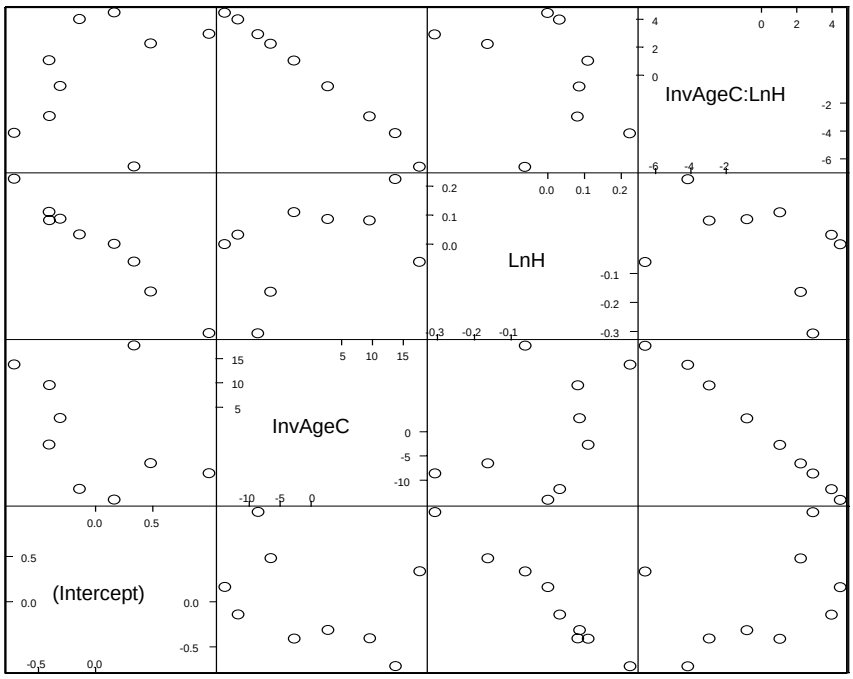

FIGURE 4.5. Pairs plot with correlation for the random-effects of first level estimated for linear multilevel effect model.

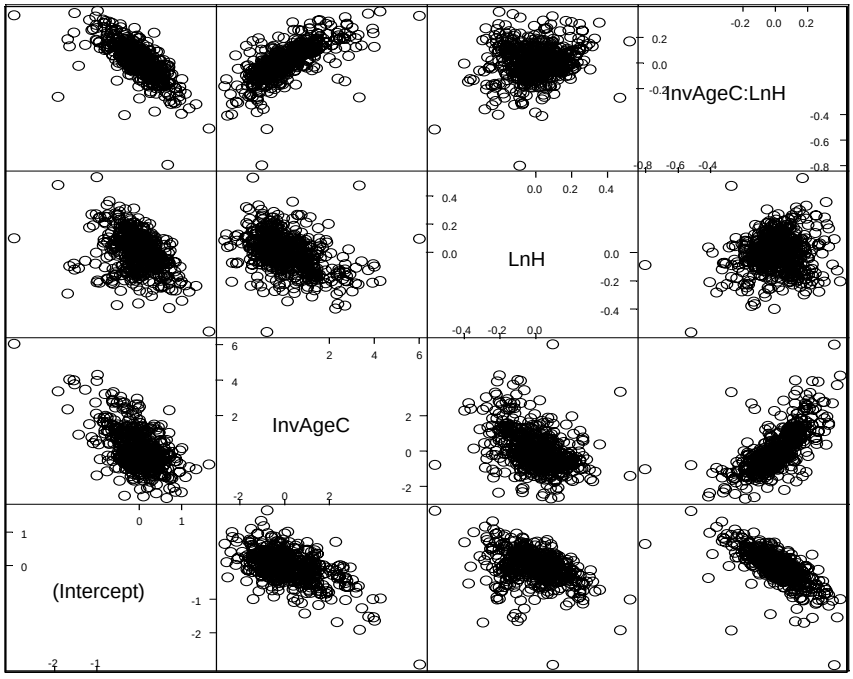

FIGURE 4.6. Pairs plot with correlation for the random-effects of second level estimated for linear multilevel effect model.

TABLE 4.2. Information statistics and loglikelihood for 6 different linear multilevel models.

Model	*df*	*AIC*	*BIC*	*logLik*	*Test*	*L.Ratio*	*p-value*
Heteroscedastic/Positive Definite Level1/Block Level 2	28	-20125	-19947	10090	-	-	-
Homocedastic/Diagonal Levels1,2	13	-20496	-20413	10261	1vs2	340.56	<0.0001
Heteroscedastic/Positive Definite Levels 1 and 2	32	-20596	-20392	10330	2vs3	137.93	<0.0001
Heteroscedastic./Diagonal Levels 1,2	20	-20685	-20557	10362	3vs4	64.99	<0.0001
Heteroscedastic./Positive Definite Level 1	26	-20707	-20541	10379	4vs5	30.08	<0.0001
Heteroscedastic./Positive Definite Leve1/AR(1)	27	-21851	-21680	10952	5vs6	1146.70	<0.0001

(Interaction). The model with a general positive-definite structure for level 1 was fitted to check the independence assumptions. The results are showed in Table 4.2.

The Table 4.2 present the logLikelihood values in ascending order. Based on information about correlation within-group errors and random effect variance-covariance, we tried to model these characteristics taking into account the levels 1 (plot) and 2 (tree within plot). The best model was generated by modeling the heteroscedastic pattern among plots and accounting for the correlation among random effects for level 1, which had a logLikelihood value of 10379 and a significant p-value (<0.0001) when compared with the heteroscedastic model, which has diagonal structure for both levels 1 and 2.

The empirical correlation structure was modeled based on the visualization showed in Figure 4.7. The values are positive in the first two lags, suggesting that an AR(1) model may be suitable for modeling the within-group correlation. An initial value of 0.26 as used for the AR(1) parameter, which is the value of the empirical autocorrelation at lag-1. The result is represented in Table 4.2. The logLikelihood values increased from 10359 to 10952, when compared heteroscedastic with positive definite structure without and with modeling correlation structure, respectively. The likelihood ratio test was 1146.70, with a significant p-value (<0.0001), indicating that the AR(1) represents within-subject dependence.

The final curve estimates for both level 1 (plot) and level 2 (tree) are presented in Figures 4.8 and 4.9. For level 1, even though the fitted curve passes through the center of

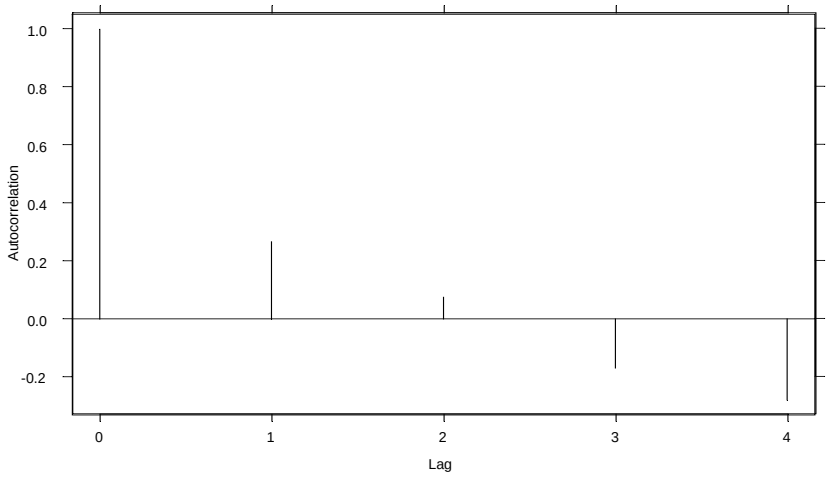

FIGURE 4.7. Empirical autocorrelation corresponding to the normalized residuals with 4 lags

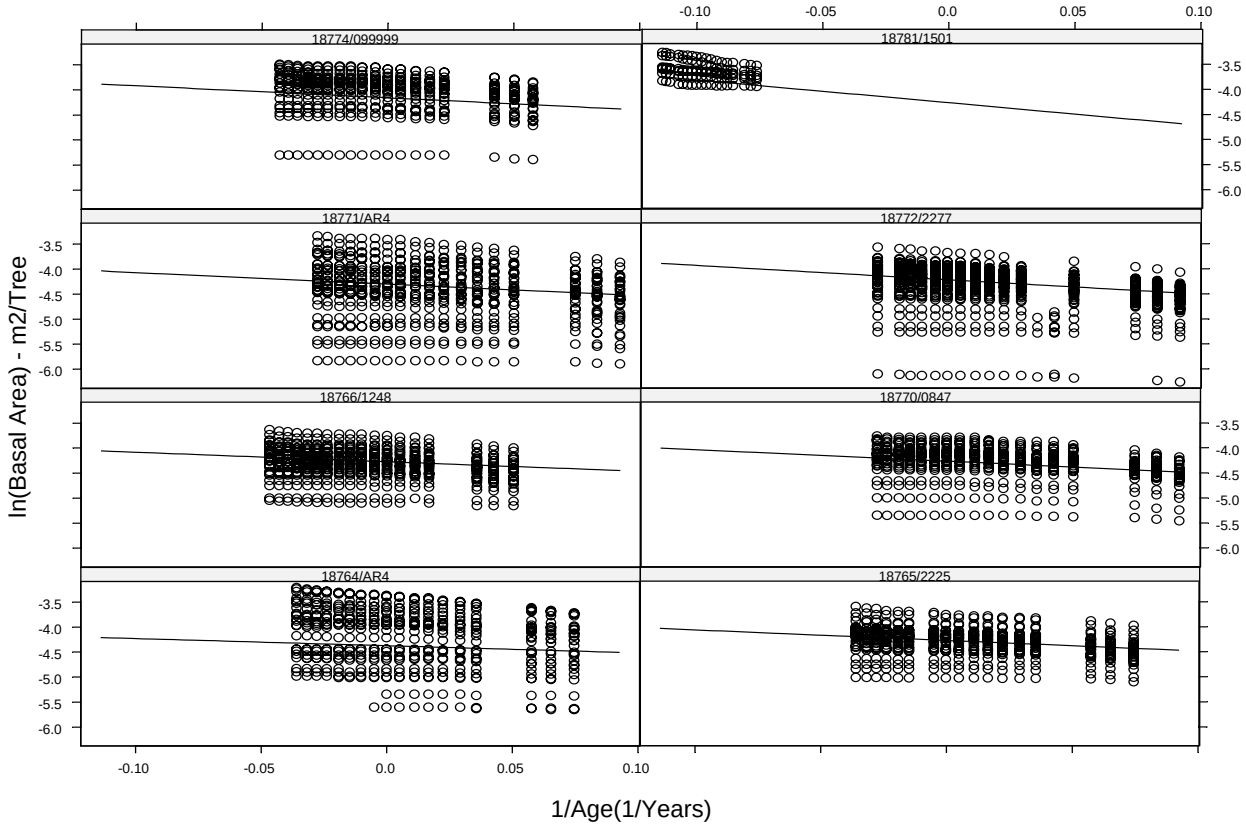

FIGURE 4.8. Estimated values of ln(BA) for the first level (plot)

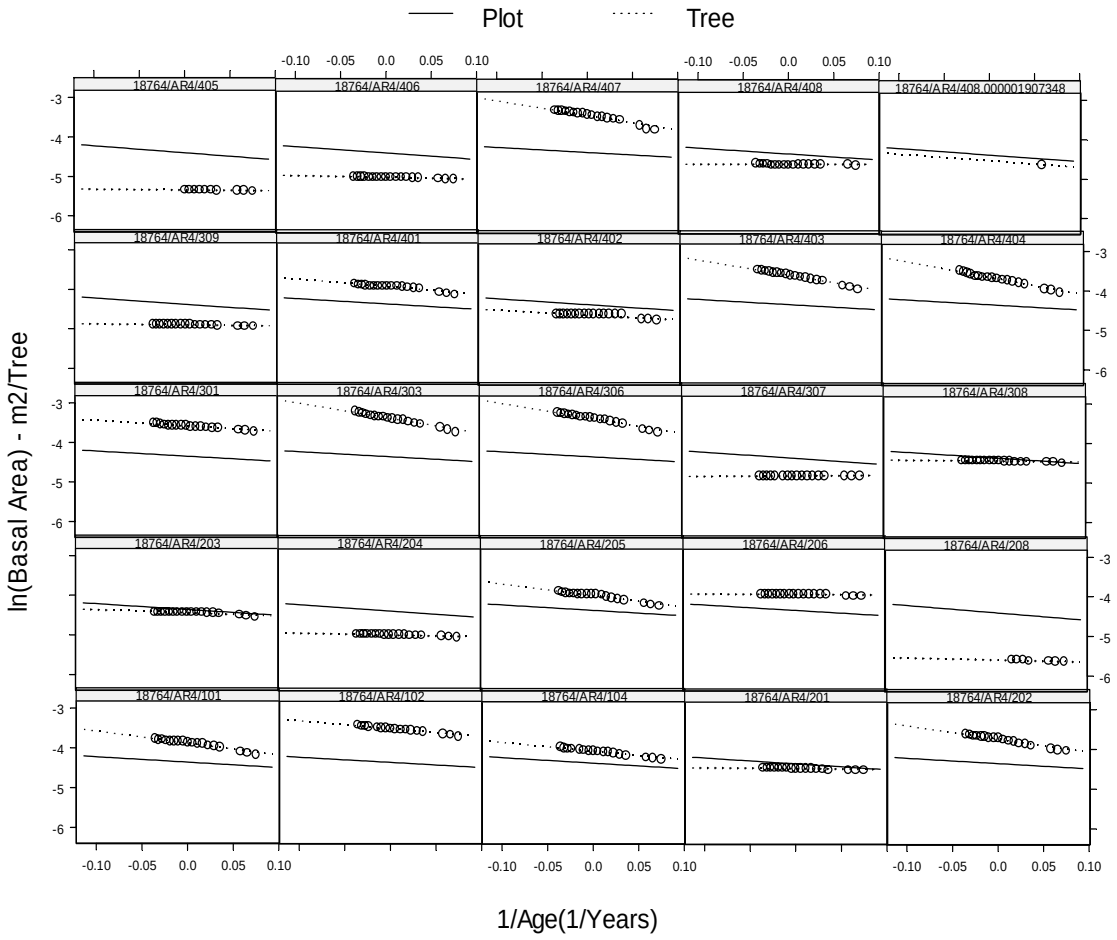

FIGURE 4.9. Estimated values of ln(BA) for the first and second levels

the observed data for each plot, the variability around the line is larger and different among plots. In the level 2, represented by Figure 4.9 for the first 25 trees, we can verify that the two level model generated predictions that follow the observed values closely, indicating that the model explains the basal area growth very accurately. Moreover, when we compare the slope between level 1 and level 2, the variation in the pattern is strong, having parallel or non-parallel lines with different distances between them. The greater the distance and/or slope between these lines, the worse the predictions will be.

Fitting A Nonlinear Two-Levels Mixed-Effects Model.

The data set used here was the same presented to fit two levels linear mixed-effects models, with some random reduction in number of trees by plot to allowed faster convergence. The response variable will be Height (H) in meters and the covariate is represented by Age(years). The Figure 4.10 shows the relationship between these two variables for Plot/Tree levels. As with single level (plot) model presented previously, we can notice a clear nonlinear relation with possible different parameters among both tree and plot. The logistic model with three parameters was used here due to the easy interpretation of its parameters and capability to generate good adjustments in this situation (4.6).

$$\mathbf{H_{ijk}} = \frac{\boldsymbol{\phi}_{\mathbf{1ij}}}{1+\exp[-(\mathbf{a_{ijk}} - \boldsymbol{\phi}_{\mathbf{2ij}})/\boldsymbol{\phi}_{\mathbf{3ij}}]} + \boldsymbol{\varepsilon}_{\mathbf{ijk}} \qquad (4.6)$$

Where:

H_{ijk} = Height (meters) for i-th plot, j-th tree and time k;

a_{ijk} = Age(years) for i-th plot, j-the tree and time k;

ε_{ijk} = Random error;

$$\boldsymbol{\Phi}\text{ij} = \begin{bmatrix} \boldsymbol{\phi}_{1ij} \\ \boldsymbol{\phi}_{2ij} \\ \boldsymbol{\phi}_{3ij} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \boldsymbol{\beta}_3 \end{bmatrix} + \begin{bmatrix} \mathbf{b}_{1i} \\ \mathbf{b}_{2i} \\ \mathbf{b}_{3i} \end{bmatrix} + \begin{bmatrix} \mathbf{b}_{1ij} \\ \mathbf{b}_{2ij} \\ \mathbf{b}_{3ij} \end{bmatrix} = \boldsymbol{\beta} + \boldsymbol{b_i} + \boldsymbol{b_{ij}}$$

$\boldsymbol{b_i} \sim N(0,\boldsymbol{\Psi_1})$, $\boldsymbol{b_{ij}} \sim N(0,\boldsymbol{\Psi_2})$ and $\varepsilon_{ijk} \sim N(0,\sigma^2)$

Here, $\boldsymbol{ß}$ is a vector of fixed effects and $\boldsymbol{b_i}$ represents the vector of random effects for level 1(plot) and $\boldsymbol{b_{ij}}$ is the vector of random effects of level 2 (tree). The interpretation of the parameter are the same as that for one level: Φ_{1ij} (asymptote) is the horizontal

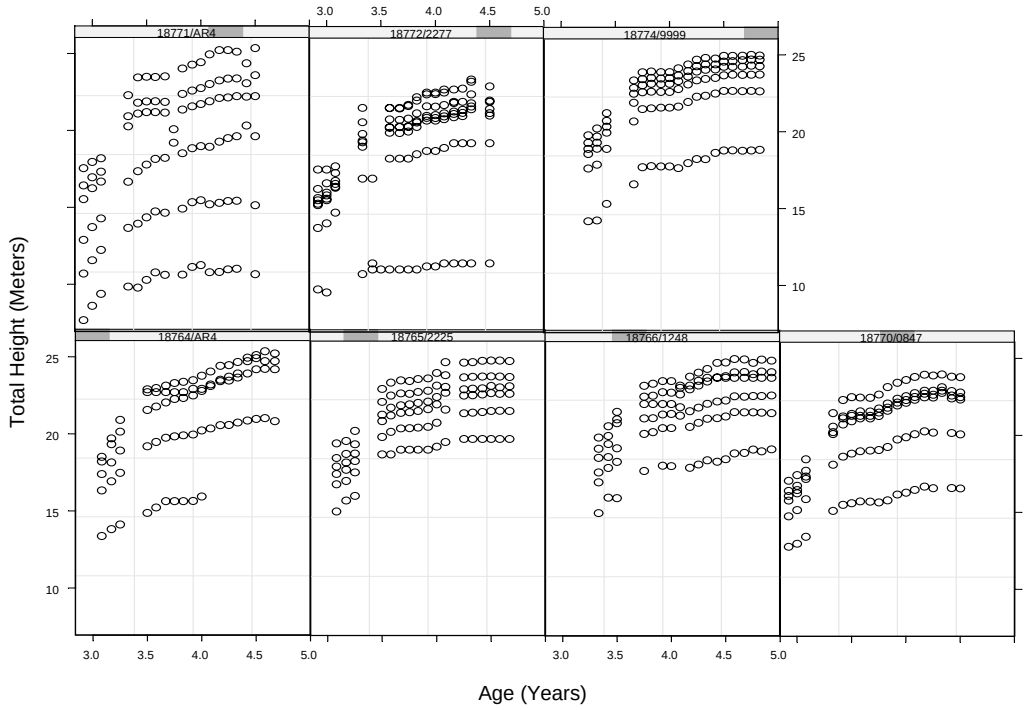

FIGURE 4.10. Observed total height for individual trees within plots

asymptote as age goes to infinity. Φ_{2ij} (middle response) is the age value which the response is Φ_{1ij} /2, representing the inflection point of the curve. Φ_{3ij} (scale) is the scale parameter, which represents the distance on the x-axis between the inflection point and the point where the response is asymptote/$(1+e^{-1})$ =0.73 of the asymptote. Also, the basic assumptions are: the random effects are normally distributed and independent for different groups and the within-group errors are independent and identically, normally

distributed and independent of random effects. We will check the assumptions that Ψ_1 and Ψ_2 are normally distributed and are diagonal matrices, i.e., the random effects are independent.

The nonlinear mixed-effects model, with two levels, was fitted considering the variance-covariance matrix for both levels as diagonal, meaning that the random effects for both levels are independent. The results are presented in Table 4.3. The three fixed parameters had significant p-values (<0.0001), indicating their importance in the model.

TABLE 4.3. Fixed parameters estimated and random standard deviations for model (4.6)

	Fixed Effect					Random Effect Std. Deviations	
Parameter Estimated	Value	Std. Error	DF	t-value	p-value	Plot	Tree within Plot
Φ_1	21.69225	0.509483	684	42.57	<0.0001	0.22063	3.28324
Φ_2	2.67183	0.033037	684	80.87	<0.0001	0.08285	0.04256
Φ_3	0.37902	0.012367	684	30.64	<0.0001	0.02788	6.82e-7
Residual	-	-	-	-	-	-	0.29723

An important observation is the very small value of the standard error for the scale parameter on level 2 (Tree within Plot). This value probably indicates that this parameter might be dropped from this level. This will be checked later, when we will try to model the level 2 with two random parameters.

Initially, the correlation among the random effects was checked. The Figure 4.11 shows a positive correlation between middle response and scale parameters for level 2. To account for this correlation, variance-covariance matrix for level 2 was modeled as a block diagonal matrix, having Asym as the first block and xMid and Scal as the second block. The results for diagonal and block diagonal models are presented in Table 4.3.

TABLE 4.4. Information statistics and likelihood for 6 different models.

Model	*df*	*AIC*	*BIC*	*logLik*	*Test*	*L.Ratio*	*p-value*
Homoscedastic/Diagonal Levels1,2	10	746	791	-363	-	-	-
Homosc./Diag. Lev.1 and Block 2	11	745	795	-361	1vs2	2.602	<0.1067
Heteroscedastic./Diag. Levels 1 and 2	16	721	794	-344	2vs3	34.086	<0.0001
Heter./Diag. Lev. 1,2/AR(1)	17	562	640	-264	3vs4	161.178	<0.0001
Heter./Diag. Lev. 1,2/ARMA(1,1)	18	550	632	-257	4vs5	13.967	<0.0002
Heter./Diag. Lev 1,2/ARMA(2,1)	19	547	634	-254	5vs6	4.8190	<0.0281

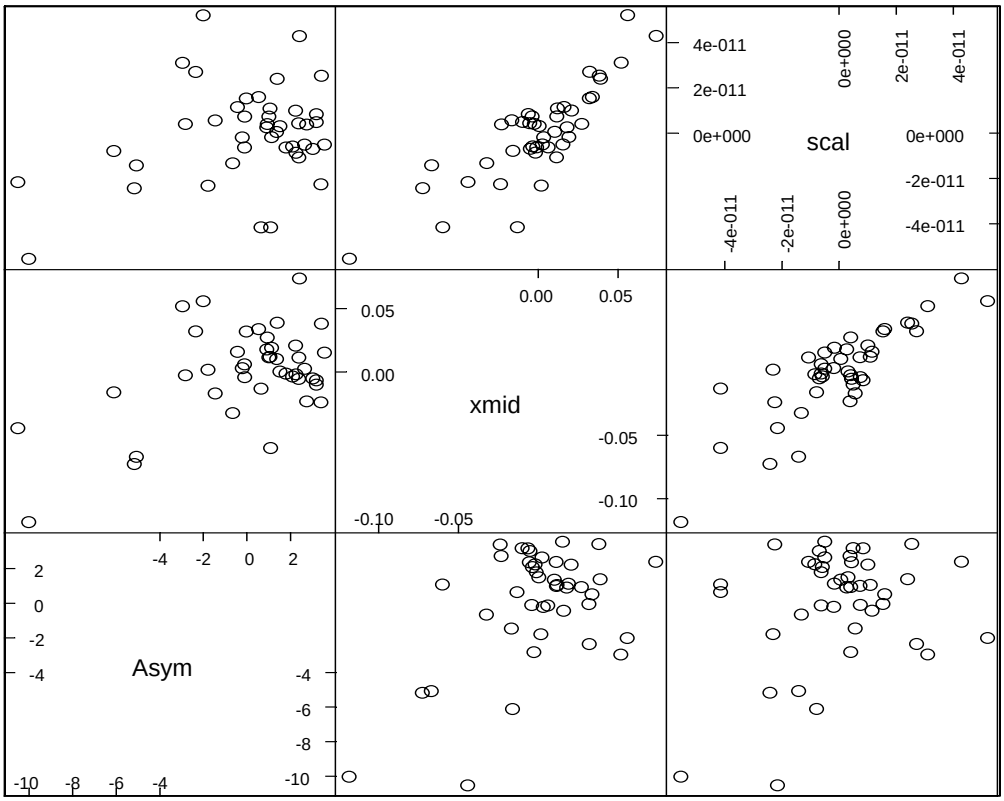

FIGURE 4.11. Pairs plot showing correlation for the random effects for level 2

The gain in modeling the variance-covariance as block diagonal was not significant (p-value=0.1067), indicating that the diagonal structure is the better model.

Next, the heteroscedasticity was evaluated. The first important characteristic of the residuals shown in Figure 4.12 is that, for each plot, the distribution around zero is uniform, indicating no violation of the homoscedasticity assumption by plot. However, comparing the distribution among plots, the pattern is different. Some plots have greater variability than others. Based on this pattern, the strategy was to model the within-plot variance (heteroscedastic model).

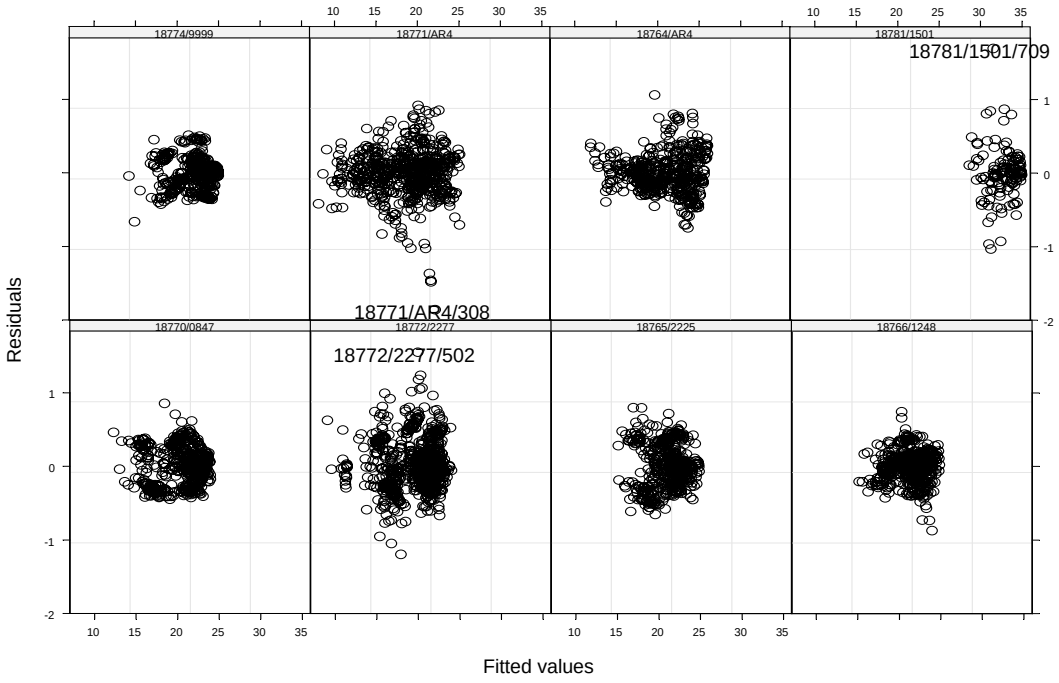

FIGURE 4.12. Scatter plot of standardized residuals distribution by plot

Table 4.3 shows the differences between the two models. The smaller value for AIC and greater LogLikelihood indicates that the heteroscedastic model explains the data

better than the homoscedastic model. This is confirmed by the very small p-value (<0.0001).

After defining the variance-covariance as a diagonal structure and assuming a heteroscedastic model, the next step was to model the correlation. The characteristic of the data set as longitudinal repeated measurements is a good reason to expect serial correlation within the subjects (Tree). Thus, several models were tested, with variations of autoregressive (AR(.)), moving average (MA(.)) and the combination of them (ARMA(.,.)). The results are presented in Table 4.3. The greater LogLikelihood value associated with the model ARMA(2,1), generating a likelihood ration test of 4.81 when compared with the model ARMA(1,1) and the significant p-value of 0.0281, indicates that this model is preferred. For further analysis, the ARMA(2,1) model will be used.

A general assessment of the capability of the model (4.1) is provided by plotting both estimated plot and tree levels against the observed values of trees (Figure 4.13). The figure indicates that the heteroscedastic model, with diagonal correlation among random effects, for both levels, and with autoregressive and moving average correlation structure describes the individual tree height growth patterns most precisely. In addition, we observe that the tree-level model has the ability to represent with great flexibility even in radical situations when the tree is far from the plot average. For example, even though the profile18771/AR4/406 and 18774/9999/302 are different when compared with the plot profile, the model captured this variation and generated precise estimates (The values X/Y/Z represent the plot number, the clone number and the tree number, respectively).

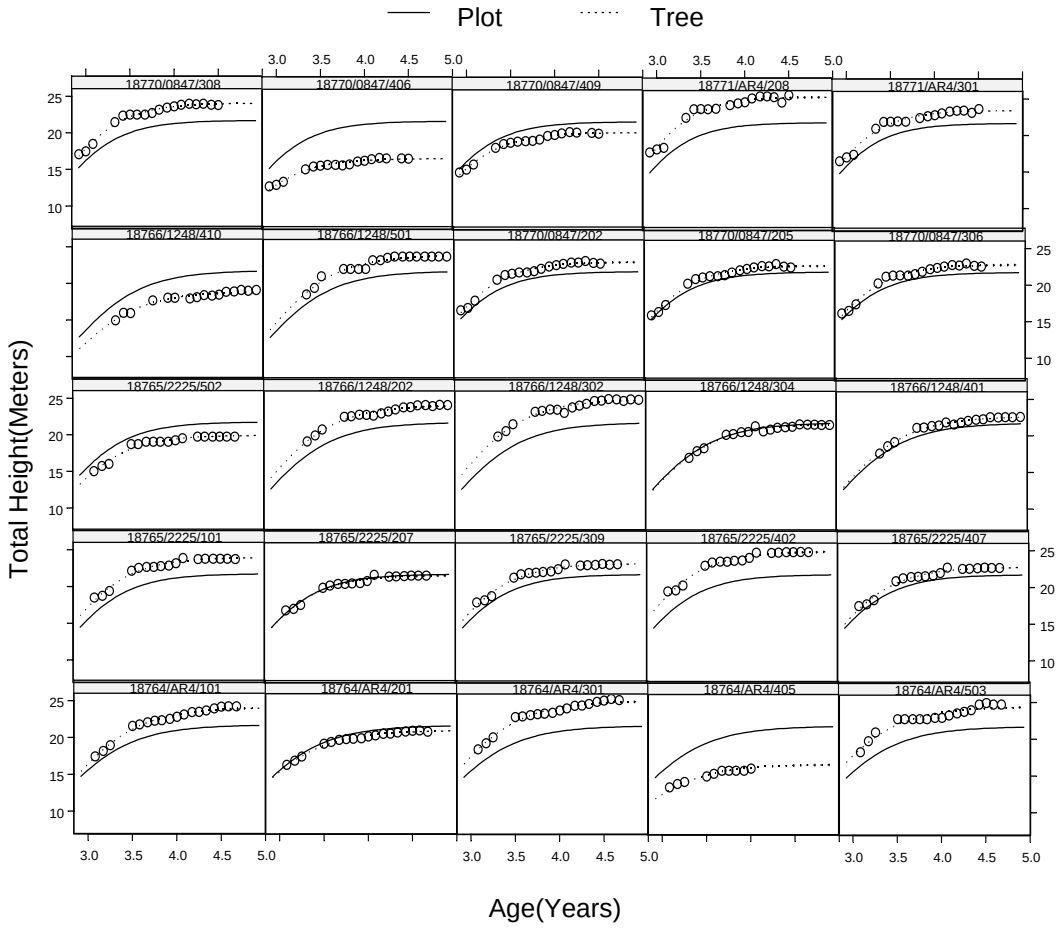

FIGURE 4.13. Fitted values for height (H) by plot (level 1) and by tree (level 2)

Discussion

In linear multilevel mixed-effects approach, we used the logarithm of basal area as a response variable associated with inverse of age, logarithm of height and the interaction between these two covariates. As the residuals by plot had different variances, the heteroscedastic model was used to model this situation and avoid a violation of the assumption that there was the same variance among plots. Also, the normality assumption of the random effects was assessed and a positive-definite structure for level 1 was used.

After modeling heteroscedasticity and autocorrelation using AR(1) structure, the loglikelihood value increased from 10090 to 10952, with p-value <0.0001 for the likelihood ratio test.

In the nonlinear multilevel mixed-effects approach, the three parameter logistic equation was used to model the variation of the tree height as a function of tree age. After modeling heteroscedasticity, random effects correlation and autocorrelation, using ARMA(2,1) structure, the loglikelihood value increased from -363 to -254, with a significant p-value of the 0.0281.

Based on these results, we can conclude that it is necessary to model heteroscedasticity, autocorrelation and correlation and the distribution of the random effects in mixed-effects model approach to obtain models that better explain data variations.

CHAPTER 5

MODELING INDIVIDUAL TREE PROFILE BASED ON A NONLINEAR MIXED EFFECT MODEL: AN APPLICATION IN *Eucalyptus* STANDS

Introduction

Tree profile functions, also cited as taper relationships, have been investigated intensively in past century (Behre, 1923; Bruce, 1968; Kozak et al., 1969; Demaerschalk, 1972; Max and Burkhart, 1976; Clutter, 1980; McTague, 1986; Bailey, 1994 and 1995).

A function that can describe the tree profile precisely can also be used to obtain precise estimates of total or merchantable tree volume by integration. The focus here is not limited to selecting or to developing a taper relationship, but to interpret the estimated parameters of such functions and to use a mixed-models approach to capture and explain the variability among trees included in the analysis.

Based on visual analysis of stem profiles and previous data processing, a nonlinear four-parameter logistic model was selected for use in this analysis. One of the many qualities of this model is the biological interpretability of its parameters. The fact that the parameters can be interpreted to match the tree profile characteristics facilitates the estimation analysis.

The main purpose of this study is to use the nonlinear mixed-effects model approach, based on the four-parameter logistic model, to describe individual tree profile; to relate these profiles with population characteristics; to obtain individual tree volumes; and to predict tree profiles, exploring the robustness proprieties of the mixed-effects model approach.

Data

The data are from the same regions of those used in previous chapters, representing upper-stem diameter measurements of 133 trees, 10 different eucalypt clones and a total of 2494 observations (Table 5.1). The intention was to locate the sample trees strategically to capture the site variation. Also, a 300 square meter sample plot was located close to each sample tree and dbh, height and soil variables were recorded. Figure 5.1 shows the profile of 16 trees randomly sampled from the data set. We can see nonlinearity between height and diameter representing tree profile and some variations among tree profiles. These variations are more evident when the total data base is visualized. In these different profiles, we can recognize that upper-asymptotes for the curves may be different and that inflection points may occur at different heights on the tree stem.

Model

The model used in this analysis will be the nonlinear four-parameter logistic, which is an extension of the three-parameter logistic model used in previous chapters. The inclusion of the fourth parameter was motivated by the fact that the tree profiles, represented in Figure 5.1, have some evidence of upper and lower asymptotes. Other

Table 5.1 – Number of trees by diameter classes and clones used in the analysis.

	Diameter Classes (Cm)						
Clone	5⊣ 8	8⊣ 11	11⊣ 14	14⊣ 17	17⊣ 20	20⊣ 23	Total Trees
0204	2	2	5	2	3	1	15
1205	3	4	3	3	2		15
1248	3	4	3	3	1		14
1501	2	3	3	3	2		13
2225	2	3	3	3	1		12
2277	2	2	2	2	2		10
3918	3	3	3	3	2		14
4619	3	3	3	2	2		13
AR4	2	2	3	3	2	2	14
AR9	2	3	2	2	2	2	13
Total	24	29	30	26	19	5	133

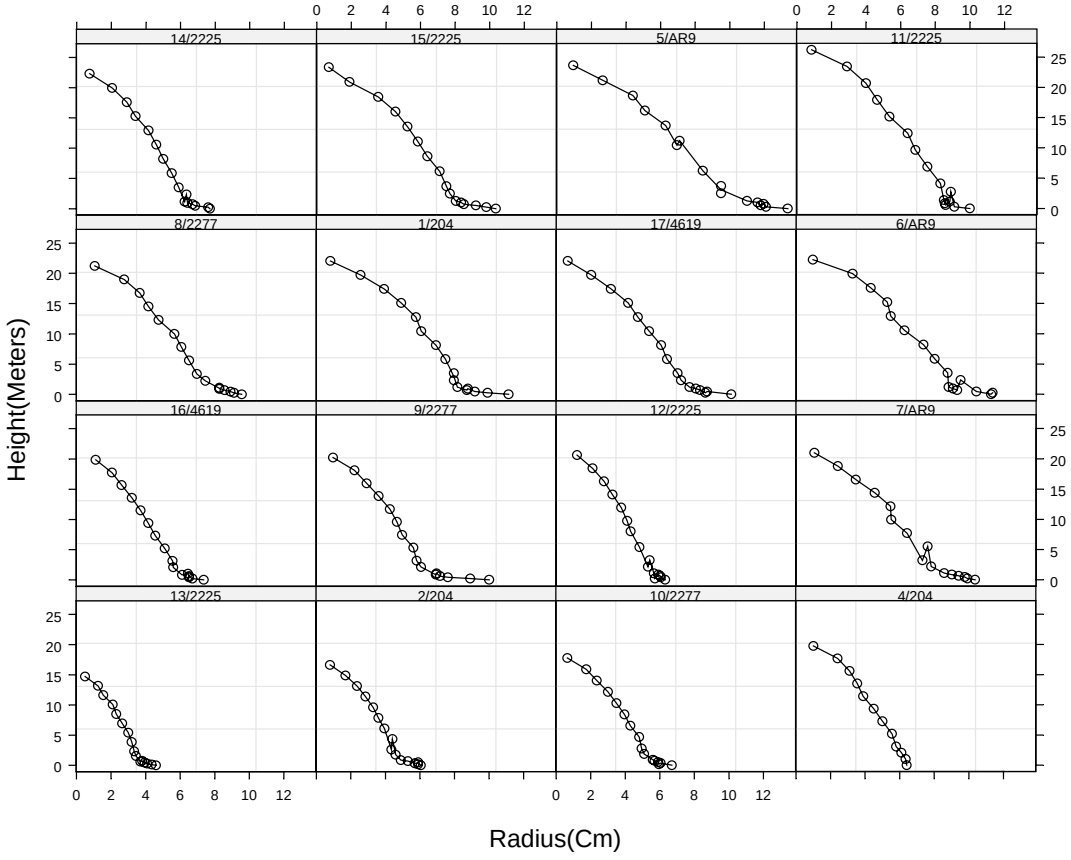

FIGURE 5.1 – Profiles representing upper-stem radius of 16 sample trees

advantages of using this model were pointed out by Ratkowsky and Reedy (1986), who describe that this model scarcely increases the nonlinearity when the fourth parameter is included. The authors compare the logistic model with Richards model (logistic model with an exponent parameter) and conclude that the Richards model is an unfortunate model because not only is its parameter-effects nonlinearity high but also its intrinsic nonlinearity. We tested the close-to-linear property of the model by fitting about 100 random samples taken from the data set (Figure 5.2). Except for the lower asymptote parameter, the parameter distributions had normal shape confirming the close-to-linear property found by Ratkowsky and Reedy (1986). As we will see further, the lower asymptote parameter was kept in the model to facilitate the convergence process. The final nonlinear four-parameter mixed-effects logistic model had the following form:

$$\mathbf{h_{ij}} = \boldsymbol{\phi}_{1\mathbf{i}} + \frac{\boldsymbol{\phi}_{2\mathbf{i}} - \boldsymbol{\phi}_{1\mathbf{i}}}{1 + \exp[(\boldsymbol{\phi}_{3\mathbf{i}} - \mathbf{r_{ij}})/\boldsymbol{\phi}_{4\mathbf{i}}]} + \boldsymbol{\varepsilon}_{\mathbf{ij}}; \qquad (5.1)$$

Where:

h_{ij} = Height of the ith tree in jth stem position;

r_{ij} = Radius of the ith tree in jth stem position;

ε_{ij} = Random error;

$$\boldsymbol{\Phi}_i = \begin{bmatrix} \phi_{1i} \\ \phi_{2i} \\ \phi_{3i} \\ \phi_{4i} \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} + \begin{bmatrix} b_{1i} \\ b_{2i} \\ b_{3i} \\ b_{4i} \end{bmatrix} = \boldsymbol{\beta} + \boldsymbol{b_i}$$

$$\boldsymbol{b_i} \sim N(0, \boldsymbol{\Psi}) \text{ and } \varepsilon_{ij} \sim N(0, \sigma^2)$$

The $\boldsymbol{\beta}$ is a vector of fixed effects and $\boldsymbol{b_i}$ represents the vector of random effects. The parameter ϕ_{1i} is the upper horizontal asymptote as tree radius goes to zero. The

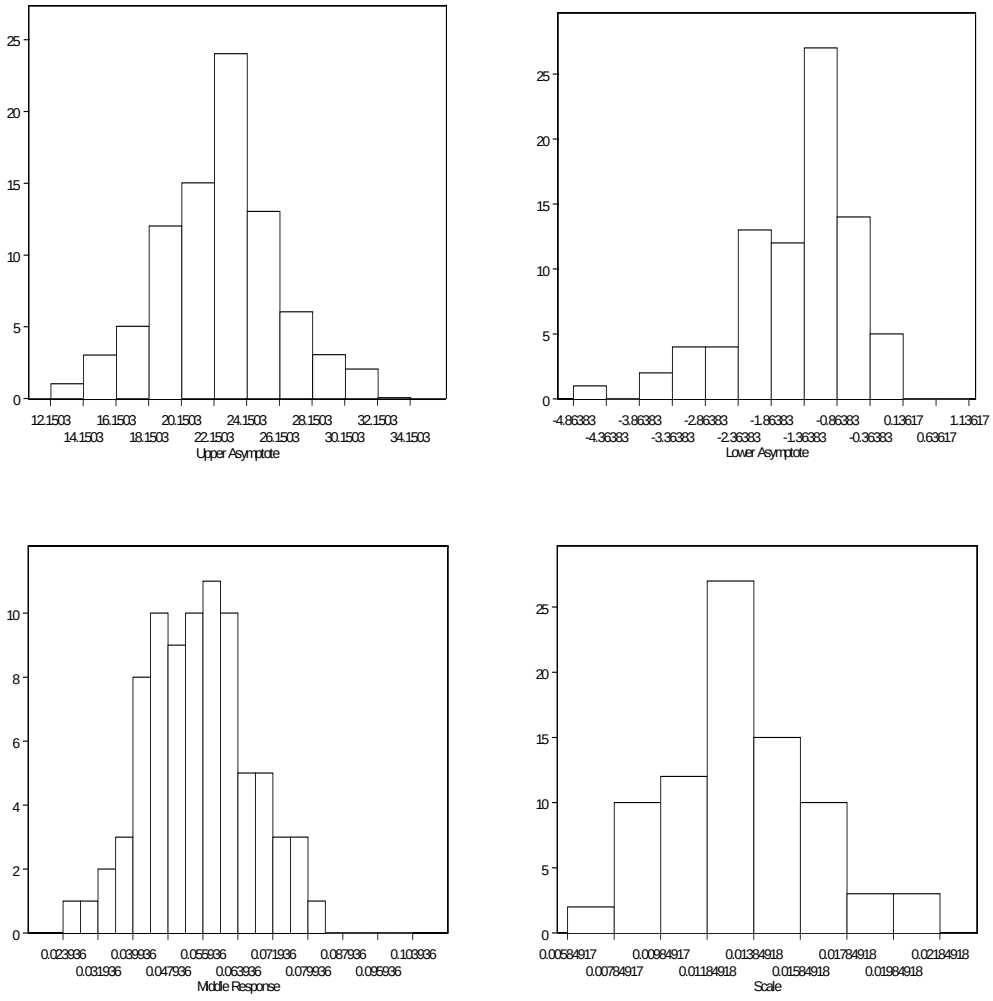

FIGURE 5.2. Distribution of the four-parameter logistic equation

parameter ϕ_{2i} is the lower horizontal asymptote as tree radius goes to positive infinity. The parameter ϕ_{3i} is the middle response parameter, which represents the radius value at the inflection point, when the response is $(\phi_{1i} - \phi_{2i})/2$, *i.e.* halfway between the upper and lower asymptote. The parameter ϕ_{4i} is the scale parameter or, when $r = \phi_{3i} + \phi_{4i}$, the response is $\phi_{1i} + (\phi_{2i} - \phi_{1i})/(1+e^{-1}) \approx 0.75(\phi_{1i} - \phi_{2i})$. Again, as basic assumptions, the random effects are normally distributed and independent for different groups and the within-group errors are independent and identically normally distributed and independent of the random effects.

Fitting the Four-Parameter Logistic Mixed-Effects Model

The four-parameter logistic model was fitted and the results are presented in Table 5.2. The significant p-value for fixed effects indicates that the four-parameter logistic model represents these data well. The small values for the correlations between random effects indicate that the variance-covariance matrix is not ill-conditioned and that

TABLE 5.2 – Parameter estimates, statistics, random effects and correlation

	Fixed Effect					Random Effect Std. Deviations	Correlations		
Parameter Estimated	Value	Std. Error	DF	t-value	p-value	Tree	Φ_2	Φ_3	Φ_4
Φ_1	20.23546	0.357992	2335	56.5249	<0.0001	4.361702	-0.14	-0.014	0.075
Φ_2	-1.13249	0.055895	2335	-20.2610	<0.0001	0.087209	X	-0.019	-0.363
Φ_3	4.80151	0.125702	2335	38.1975	<0.0001	1.561621	-0.019	X	-0.011
Φ_4	1.15481	0.025969	2335	44.4678	<0.0001	0.288901	-0.363	-0.011	X
Residual	-	-	-	-	-	0.711355			

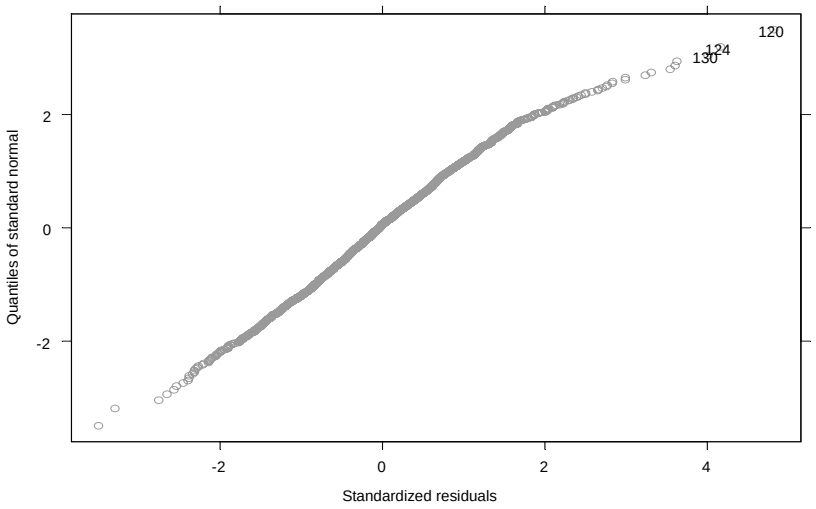

FIGURE 5.3. Normal probability plot of the standardized residuals from the four-parameter logistic model.

the random-effects structure is not over-parameterized. Based on this information, we kept the four parameters as both fixed and random in the model. Figure 5.3, with exception of some outliers, does not indicate violations of the normality assumption for the within-tree errors.

Based on information from plot samples located close to the tree sample, we tried to include some covariates in the model, such as age, dominant height, trees per hectare, basal area, soil characteristics (class, percentage of sand, silt and clay). The new formulation for model parameters is:

$$\phi_{ij} = \beta_i + \gamma_{1i}x_{1i} + \gamma_{2i}x_{2i} + ... + \gamma_{ki}x_{ki} + b_{ij} \qquad (5.2)$$

Where **i** represents the ith parameters, varying from 1 to 4; **j** represents the random tree number; β_i represents the average of the ith parameter; γ_{ki} represents the

effect of the covariates $\mathbf{x}_{ij}$; and $\mathbf{b}_{ij}$ is the ith random effect associated with jth tree. Table 5.3 indicates that the covariates for dominant height and clone had significant influence on the parameter estimates. The covariate dominant height had significant influence on the upper asymptote, middle response and scale parameters, while the covariate clone had significant influence only on the scale parameter.

Figure 5.4 provides the plot of the augmented predictions, which is the final assessment of the quality of the fitted model of the 16 randomly selected trees from a total of 133 trees. The fitted curves representing the four-parameter logistic mixed-effects model generated precise estimates attesting to the adequacy of the model. Also, this adequacy can be seen in Figure 5.5. The variation in the fixed effects among trees is due to including dominant height and clone effect in the model. Also, there are strong variations in the random effect among trees. These variations are because the curves have different values for upper asymptote, middle response and scale parameters.

TABLE 5.3 – ANOVA table representing the significance for intercept and covariates effects in the logistic equation

Parameter	Variable	DF1	DF2	F-value	p-value
β_1	-	1	1976	2803.15	<.0001
γ_{11}	HD	1	1976	35.27	<.0001
Φ_2	-	1	1976	1.93	0.1647
β_3	-	1	1976	1360.84	<.0001
γ_{31}	HD	1	1976	14.58	0.0001
β_4	-	1	1976	2104.38	<.0001
γ_{41}	HD	1	1976	12.15	0.0005
γ_{42}	Clone	9	1976	2.87	0.0023

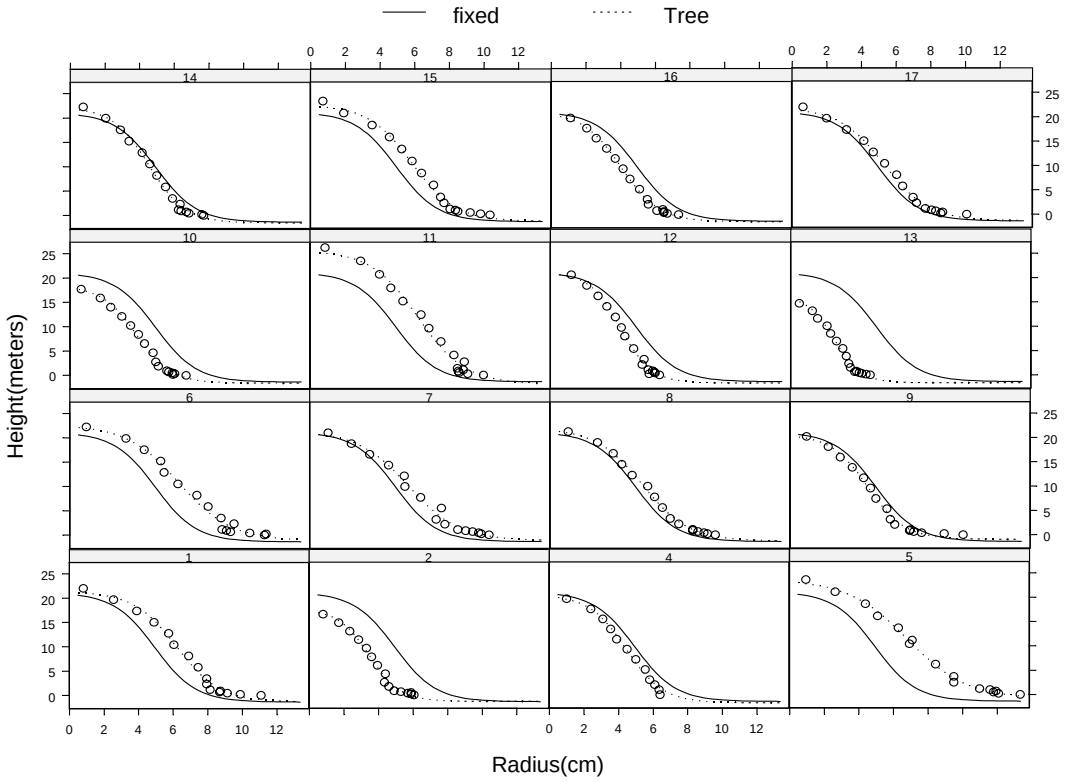

FIGURE 5.4. Individual predicted values for fixed (solid) and mixed-effects model (dashed).

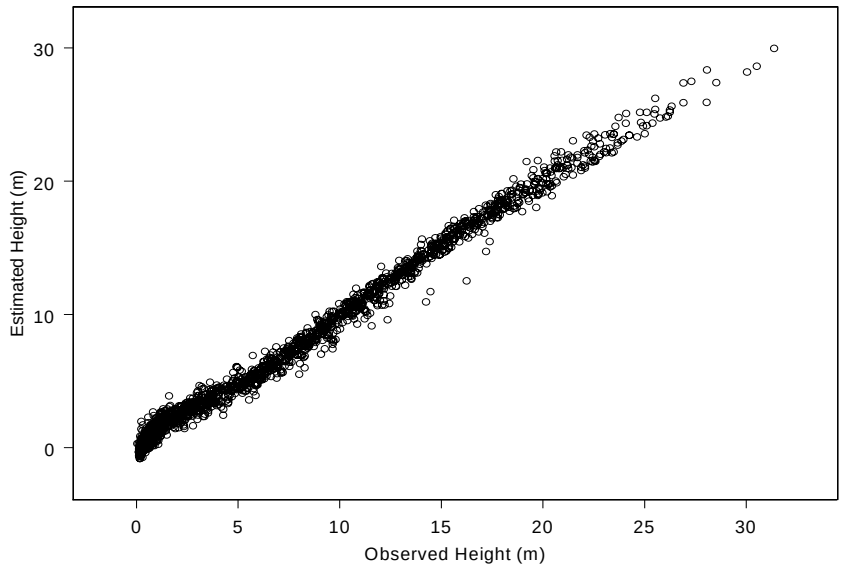

FIGURE 5.5. Comparing observed and estimated height by logistic equation

Estimating Individual Tree Volumes Using Random Parameters and Solid of Revolution Technique

The objective here is to take the tree profiles generated with the four-parameter logistic equation with fixed and random effects and rotate them around the y-axis to derive an estimation equation for volume of individual trees with different profiles. As the logistic equation has the response variable represented by tree height and tree radius representing the covariate, the method of cylindrical shells will be used for computing tree volumes. In general, the interval between radius minimum and maximum is partitioned into *n* subintervals, all of the same length. If R^*_i denotes the midpoint of the ith subinterval, a rectangle in the x-y plane with base $[R_{i-1},R_i]$ and height $f(R^*_i)$ will be formed. The cylindrical shell will be obtained by revolving the region under y=f(x) and over $[R_{i-1},R_i]$ with volume ΔV_i . The expression that gives the volume is

$$\Delta V_i \approx 2\pi R_i^* f(R_i^*)\Delta R \tag{5.2a}$$

$$V = \sum_{i=1}^{n} \Delta V_i \approx \sum_{i=1}^{n} 2\pi R_i^* f(R_i^*)\Delta R\,. \tag{5.2b}$$

This approximation is a Riemann sum that approaches the integral, which gives the volume of the solid of revolution. We can generate the merchantable volume by integrating from radius minimum to radius maximum, or vice-versa, using the following expression:

$$V = \int_{R=\min}^{R=\max} 2\pi R f(R)\,dR\,. \tag{5.2c}$$

In our specific case, the f(x) function is the logistic with four parameters, represented by the fixed and random effects. R_{min} represents the tree radius (inside or outside-bark) at height f(R_{min}) and R_{max} the tree radius at height f(R_{max}). V is the volume (inside or outside-bark, depending on whether R is specified IB or OB) of the section between R_{min} and R_{max}. Substituting the logistic function, we have the following expression:

$$V = \int_{R=\min}^{R=\max} 2\pi R \left\{ \phi_1 + \frac{\phi_2 - \phi_1}{1 + \exp[(\phi_3 - R_i)/\phi_4]} \right\} dR \qquad (5.2d)$$

In reality, the logistic equation in this situation is a taper function, which gives the estimated height of the determined radius in the tree. If the equation is solved for the variable radius, we can estimate the radius as a function of height. The expression is:

$$r_{ij}(h_{ij}) = \phi_{3i} - \ln\left[-\frac{h_{ij} - \phi_{2i}}{h_{ij} - \phi_{1i}} \right] \phi_{4i} \qquad (5.3)$$

Where r_{ij} represents the radius of the ith tree at the jth height and the Φ parameters are as defined earlier.To check the precision of equation (5.3), the dbh values for each tree at height of 1.3 meters were estimated (Figure 5.6). The estimated values are concentrated around the absolute line implying reasonably precise estimates.

By integrating the expression (5.2d), including fixed and random effects in the logistic equation, we can generate stem profiles such as those shown in Figure 5.7. The variations among trees are due to the inclusion of the random parameters, which capture the individual tree profile variations. Also, the integration technique gives us the ability to obtain merchantable volumes to a specific upper-stem radius (or diameter). If we constrain the R_{min} in (5.2d) as a minimum merchantable radius, the volume generated

will be merchantable volume with minimum radius specified. Another flexible feature of the integration process is the ability to generate multiple products from a single tree (Figure 5.8). Figure 5.9 shows the precision of using this integration technique. We can integrate the same tree again at a different merchantability position and obtain volumes for different products.

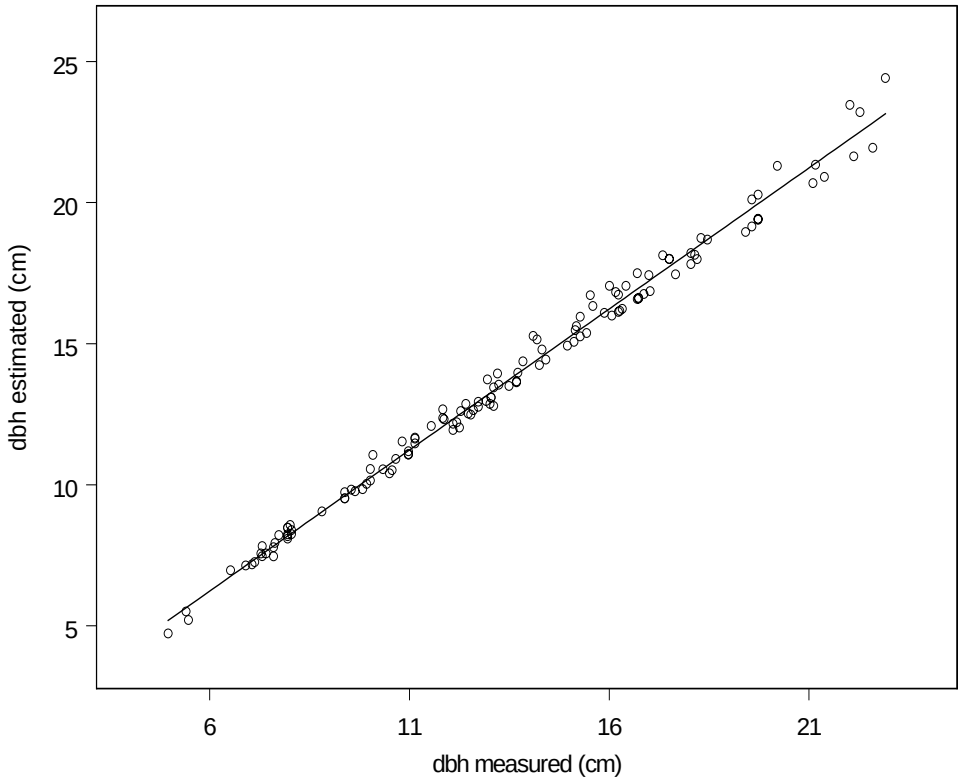

FIGURE 5.6 – Diameter at breast height estimated by mixed-effects model and observed for each tree

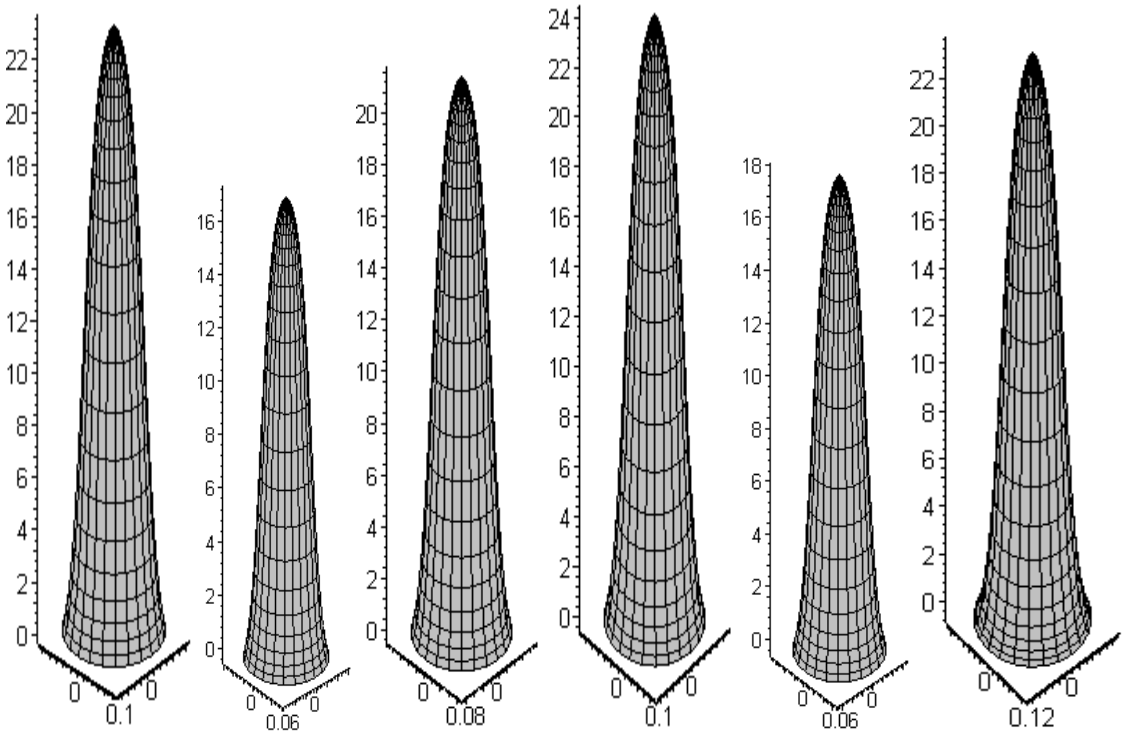

FIGURE 5.7. Variations in stem form generated by integrating the four parameter logistic equation.

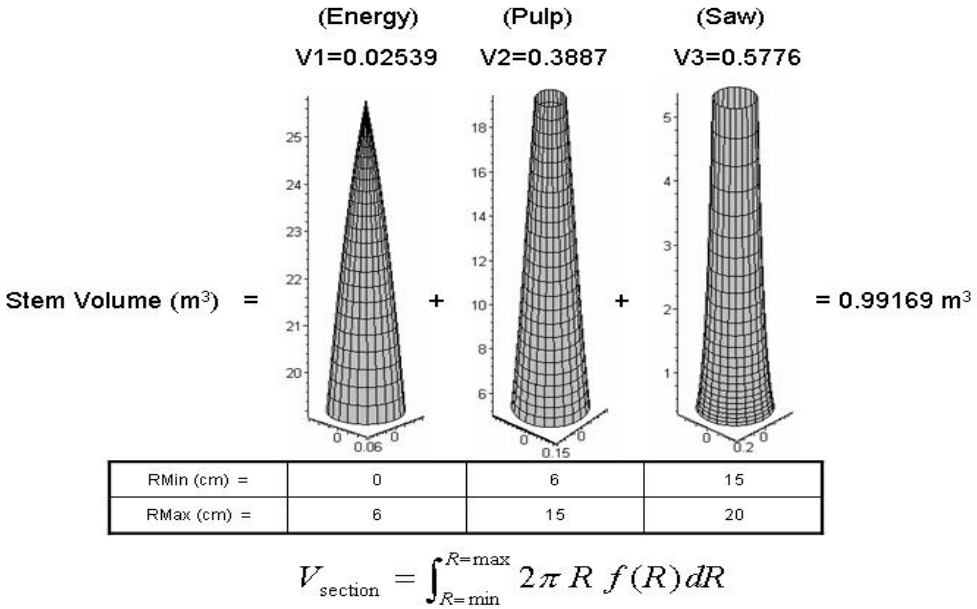

FIGURE 5.8. Partial integration representing two sub-products of a 20 cm-radius tree.

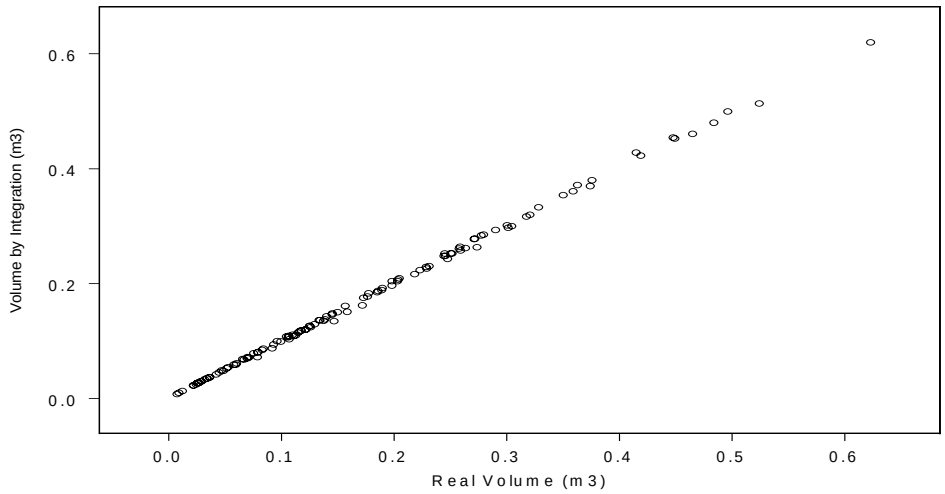

FIGURE 5.9. Comparing real volumes with those obtained by integration technique

Using the Robustness of the Mixed-Effects Model to Predict Height Based on Upper-Stem Diameter of Partially Observed Tree

In practice, the data gathered in typical inventory methods do not have complete upper-stem diameter measurements of the trees. Normally, we have a relatively small data base with complete measurements and a more representative data base with restricted information of the stem, such as dbh and total height. Hence, the focus here is to use the mixed-model theory to estimate the tree profile (upper-stem diameter or radius) of trees with restricted information, based on a data set from trees with complete upper-stem diameter measurements.

In the data set used in this analysis, the average number of observations per tree is approximately 15. Here the observations are represented by height above of the ground

associated with some upper-stem radius, which will represent the covariate. Let $\mathbf{y}_i$ denote the vector of observations for the ith tree. In our prior analysis, the data set was composed of 133 trees with complete measurements of the upper-stem height and diameter. Suppose a new tree $\mathbf{y}_k$ is included in the data base, but we have limited information about this tree, say dbh and total height. In this situation, the observation vector is $y_k = (y_{k(o)}^T, y_{k(u)}^T)^T$. Where T represents the transpose of the associated vector. The first element $(y_{k(o)})$ represents the vector of observed values for the kth tree and the second $(y_{k(u)})$ represents the vector of the values of unobserved data. If we consider 15 measurements for each tree, for example, and we have available 2 values for the kth tree, diameter at breast height and diameter (zero) at total height, the first vector will have two components and the second one 13 components.

In the linear case, the prediction process is performed by using Estimation and Maximization (EM) or Newton-Raphson algorithms, considering the vector of the unobserved data as a incomplete data (Rao ,1973; Dempster et al.,1977; Laird et al., 1987; Lindstrom and Bates, 1988; Liski, 1985, 1990, and 1996) .

In the nonlinear case, Lindstrom and Bates (1990) presented a two-step iterative procedure to estimate the parameters of the general nonlinear model (3.1). The first step, called the pseudo-data step, uses an iterative process to obtain the estimate of fixed effects $\boldsymbol{ß}$, random effects **b**, variance and covariance components $\boldsymbol{\theta}$, covariates **X** and the matrix associated with random effects **Z.** These estimates are based on Cholesky factors. The second step, called the linear mixed effects step, uses the estimate obtained in first step and, also using an iterative process, generates the desired estimate of $\boldsymbol{\theta}$, σ and $\boldsymbol{ß}$.

To perform the prediction analysis, we eliminated the the upper-stem height-diameter measurements of some trees, leaving only measures of the dbh and total height. So, the data set used to perform this analysis is a mixture of trees with complete and restricted information. This procedure allows us to compare the estimated with the real tree profile.

Figure 5.7 shows that the method is adequate to estimate both tree profiles with complete upper-stem diameter information and tree profiles with limited information. Again, the mixed-effects model expressed a strong degree of robustness, using information from trees with complete data and estimating the parameters, fixed and random, for trees with limited information. This feature is especially useful in the forest estimation process, in which a monumented data set with measurements of dbh and total height is available.

The precision of the estimate can be seen in Figure 5.8. The estimated curve fitted with both the complete and restricted data set for each tree had a sigmoid shape with one inflection point and upper and lower asymptotes. For every tree with restricted data (just dbh and total height measurements) the curve crossed the two sample points. This occurs due to the robustness of the mixed-effects approach. In another approach, such as ordinary least squares, this situation would be a problem because this approach fits either a mean curve for the population or one curve independently for each tree. In the first case, the missing observations could strongly affect the estimation of the population mean value. In the second situation, the adjustment for an individual tree with missing observations, in this specific case, would generate a straight line, which does not represent the tree profile.

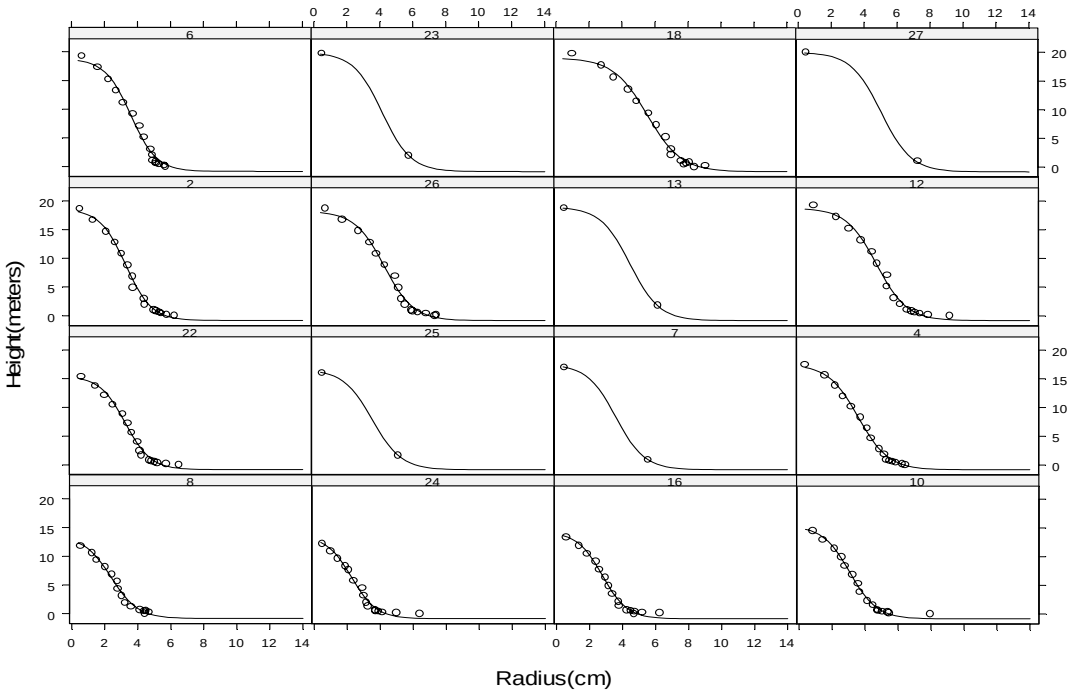

FIGURE 5.10. Random estimate profile for 16 trees, with 5 trees with restricted information.

When the tree profiles estimated by mixed-effects model are compared with the real values (Figure 5.9), we can see that the model generates a sigmoid curve close to the observed trend. This situation has a strong practical application in forest prediction studies. With a reliable data set, which represents some forest population and describes the variability in tree profiles represented in the population, we can generate precise estimates for individual trees, using minimal tree information, such as dbh and total height. The more complete the data base, the more reliable the predictions.

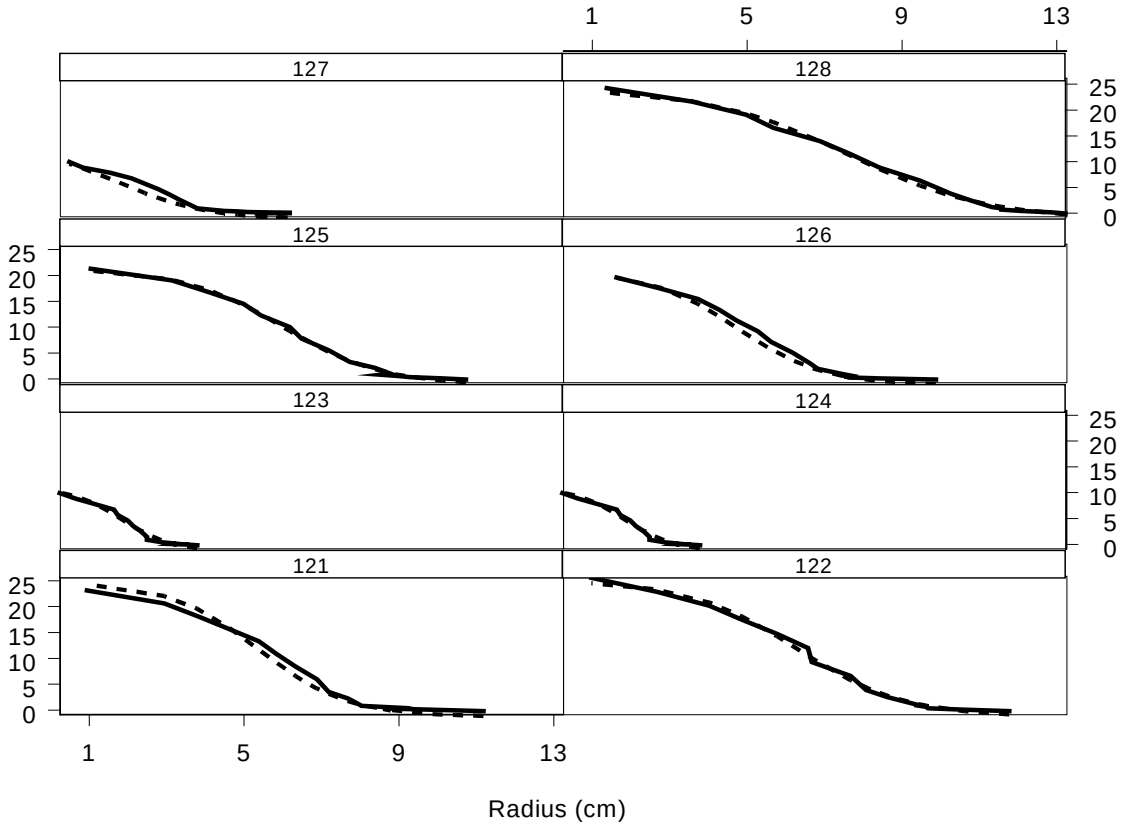

FIGURE 5.11. Observed (solid) and estimated (dashed) height values for eight trees with restricted information

Discussion

The logistic four-parameter equation was fitted based on the nonlinear mixed-effects model approach to represent the individual tree profiles. The four parameters, represented by upper asymptote, lower asymptote, inflection point and middle response, in both fixed and random effects, had a significant contribution in explaining the variation of the tree height as a function of the upper-stem radius. The less significant parameter was the lower asymptote, but it was kept in the model to improve the

convergence process. One could set the parameter value to zero, based on the logic that for large diameters the height values approach to zero. It was verified that strong variability exists in the parameter values among trees.

Variables such as clone type, dominant height (HD), soil characteristics, age, basal area, and others were used as covariates to explain the variation among trees. After selection based on analysis of variance procedures, the variables HD and clone type were included in the model. HD was associated with the parameters upper asymptote, inflection point and middle response, and the clone type was associated with the middle response only.

The solid of revolution technique was used to obtain the individual tree volumes for both total tree and merchantable volumes. Although this technique is flexible and relatively simple, the precision of the result is directly related to the quality of the fitted function to describe the tree profile.

Using an iterative prediction technique, the robustness of the nonlinear mixed-effect was used to estimate the profile for trees with restricted observations, in this case represented by dbh and total height. When we compared the profiles generated with the prediction process to real profiles, the estimates were very close and the profiles estimated had parameters approximating those of the real profiles. This technique, compared with generalized nonlinear least squares estimate, has the advantage in capturing individual tree variations, translating them into parameter estimates and, through integration, generating the individual tree volume estimates. One practical application of this methodology would be to estimate plot volume based upon

information on the dbh and total of height the trees in the plots. In application, it would be necessary to have available a data set with complete information about tree profiles representing the population.

CHAPTER 6

SUMMARY AND CONCLUSIONS

The linear and nonlinear multilevel mixed-effects model approach has been used in many fields of study. In this study, this approach was used to model *Eucalyptus* clonal stand growth, considering plots and/or tree within plots as the random effects, using a data set with longitudinal, irregularly spaced and unbalanced information. The following conclusions were generated from this study:

1. Modeling basal area using single level linear mixed-effects approach:
    - The logarithm of basal area as a response variable was associated with the inverse of age as a covariate. Both slope and intercept random effects had significant differences when plot was included in the model as a random subject. Other fixed covariates were included in the model, such as logarithm of dominant height, logarithm of the number of trees per hectare and some interaction among them. The clone effect as a covariate also had a significant effect in explaining variation in basal area;

- Estimates from the mixed-effects model tended to be pulled toward those of the fixed effects (shrinkage estimates), giving a certain robustness, an important characteristic of the mixed-effects approach;
- In examining distributional assumptions, the within-plot constant variance assumption was violated and it was necessary to model it. After modeling this assumption, the loglikelihood value had a significant increment. The normality of random effects was considered reasonable. Also the correlation structure was modeled and the loglikelihood increased significantly again;
- The random parameters representing the slope and intercept were related to and could be predicted from the site index value for each plot. With this method, it was possible to estimate basal area for different plots, not included in the analysis;

2. Modeling dominant height growth using a single level nonlinear mixed-effects approach:
    - Dominant height growth was modeled as a function of time (years) using a three parameter logistic equation (upper asymptote, scale and middle response). All three parameters had significant fixed and random effects, with no indication of an over-parameterized model;
    - Neither homoscedasticity within-plot nor normality of the random effects were violated and the homoscedastic model provided a good representation of the data;

- As result of modeling dominant height growth, sets of anamorphic, polymorphic and “quasi-anamorphic” height average curves were developed;

3. Modeling basal area and dominant height based on multilevel linear and nonlinear mixed-effects approach:
    - In studying basal area, fixed and random effects related to inverse of age, logarithm of dominant height and the interaction of logarithm of dominant height and inverse of age had significance influences in explaining logarithm of basal area variation;
    - Checking distributional assumptions, it was found that the residuals within plot had different patterns, and it was necessary to model this pattern. After modeling these different patterns, the logLikelihood value had significant increment, indicating that the heteroscedastic model explains the data better than the homoscedastic model. Also, modeled were the variance-covariance structure representing the random effects and the empirical correlation structure of within group. Due to the high correlation between two random effects in level 1, the general positive-definite structure was used. In the empirical correlation modeling, the autoregressive structure had a better improvement in the model. After modeling the variance-covariance of random effects and empirical autocorrelation, the loglikelihood again increased significantly;

- In studying nonlinear two level mixed-effects, where the tree height growth was modeled, the three parameter logistic-equation was used. Again, there were three significant fixed and random parameters for the model. Better improvement was reached when we modeled the within-plot variance and the autocorrelation were modeled. For autocorrelation, the better model was ARMA(2,1);

4. Modeling individual tree profiles based on a nonlinear mixed-effects model:
    - The four parameter logistic equation was used to estimate tree height as function of tree radius. As a first result, the model had significant fixed and random effects to represent the data set. In a simulation study, the four parameter model had a close-to-linear behavior, confirming the adequacy of this equation;
    - Using information from plot samples located close to the tree sample, some new covariates were included in the model to improve the representation of the data. The plot-level dominant height had significant influence in the model when associated with upper asymptote, middle response and scale parameters. The variable representing the tree's clone had a significant effect on the scale parameter;
    - Using the random parameters and solid of revolution technique, volumes were generated for each tree providing values close to actual tree volumes, confirming the precision of the nonlinear mixed-effects model in estimating the tree profile;

- The robustness of the mixed-effects model was used to estimate the profile for trees with only dbh and height measurements. The method generated precise estimation of the tree profile, as an alternative to estimating individual tree and plot volumes;

LITERATURE CITED

Amaro, A., Reed, D., Tomé, M., and Themido, I. 1998. Modeling dominant height growth: *Eucalyptus* plantations in Portugual. *For. Sci.* 44(1):37-46.

Avery, T. E. and H.E.Burkhart. 2002. *Forest measurements.* McGraw-Hill, New York. 408p.

Bailey, R.L. and J.L. Clutter. 1974. Base-age invariant polymorphic site curves. *For. Sci.* 20:155-159.

Biging. G. S. 1985. Improved estimates of site index curves using a varying-parameter model. *For. Sci.* 31:411-423.

Borders B.E. 1989. System of equations in forest stand modeling. *For. Sci*. 35:548-556.

Borders B.E., 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.

Breslow, N.E. and D.G.Clayton. 1993. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association.* 88:9-25.

Burnett, R.T., W,H. Ross and D. Krewksi. 1995. Nonlinear random effects regression models. *Environmetrics.* 6:85-99.

Campos, J.C.C. and J.C.Ribeiro. 1983. Produção dos povoamentos de Eucalyptus spp. Na região do Triângulo Mineiro. *Imprensa Universitaria*. Viçosa. 43p.

Carter, R.L. and M.C.K. Yang. 1986. Large-sample inference in random coefficient regression models. *Communications in Statistics – Theory and Methods*, 8, 2507-2526.

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.

Christman, M.C. and R. J. Jernigan. 1977. Spatial correlation models as applied to evolutionary biology. *Modeling Longitudinal and Spatially Correlated Data*. Springer-Verlag, New York. 410 p.

Clutter J. L., 1963. Compatible growth and yield models for loblolly pine. *For. Sci.* 9:354-371.

Clutter J. L., J.C. Fortson, L.V. Pienaar, G. H. Brister, and R. L. Bailey 1983. *Timber Management: a Quantitative approach*. Krieger, Malabar Florida. 333p.

Corbeil , R.R. and S.R. Searle.1976. Restricted maximum likelihood (REML) estimation of variance components in the mixed model. *Technometrics.* 18:31-38.

Crowder, M.J. and D. J. Hand 1990. *Analysis of repeated measures*. Chapman & Hall, New York.

Daniels, R. F. and H. E. Burkhart. 1988. An integrated system of forest stand models. *For. Ecol. Man.* 23:159-177.

Davidian, M. and D. M. Giltinan 1995. *Nonlinear models for repeated measurement data. London*: Chapman and Hall.

Demaerschalk, J.P. 1972. Converting volume equations to compatible taper equations. For. Sci. 18:180-191.

Dempster, A.P., M.R. Selwyn, C.M. Patel and A.J. Roth. 1984. Statistical and computational aspects of mixed model analysis. *Applied Statistics*. 33:203-214.

Díaz, M. P. and H. T. Z. Couto. 1999. Modelos generalizados para a mortalidade de árvores de Eucalyptus grandis no Estado de São Paulo, Brasil. *Scientia Forestalis*. 56:101-111.

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. 1999. *A simultaneous system of linear and nonlinear mixed effects models for forest growth and yield prediction*. University of Georgia.(PhD dissertation).

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. and R.L.Bailey. 2001. Nonlinear mixed effects modeling for slash pine dominant height growth following intensive silvicultural treatments. *For. Sci.* 47(3):287-300.

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.

Goldstein, H. 1986. Multilevel mixed linear model analysis using iterative generalized least squares. *Biometricka.* 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. *Biometricka.* 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.

Gregoire, T. G. and O. Schabenberger 1989. Model fitting under patterned heterogeneity of variance. *For. Sci*. 35:105-125.

Gregoire, T. G. and O. Schabenberger 1994. Fitting bole-volume equations to spatially correlated within-tree data. *Procedings of the 6th Annual Conference on Applied Statistics in Agriculture*, Manhattan, Kansas.

Gregoire, T. G., O. Schabenberger and J. P. Barrett. 1995. Linear modeling of irregularly spaced, unbalanced, longitudinal data from permanent-plot measurements. *Can. J. For. Res.* 25: 137-156.

Gumpertz, M.L. and S.G. Pantula.1989. A simple approach to inference in random coefficient models. *The American Statistician.* 43: 203-210.

Gumpertz, M.L., and C. Brownie. 1993. Repeated measures in randomized block and split-plot experiments. *Can. J. For. Res.* 23: 625-639.

Hall, D.B. and R. L. Bailey. 2001. Modeling and prediction of forest growth variables based on multilevel nonlinear mixed models. *Forest Science 47*(3).

Jones, R. H. 1993. *Longitudinal data with serial correlation: A state space approach.* London: Chapman Hall.

Kozak, A., D. D. Munro, and J. H. G. Smith. 1969. Taper functions and their applications in forest inventory. *For. Chron.* 45:278-283.

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.

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.

Littell, R.C., G. A. Milliken, W.W. Stroup, and R. D. Wolfinger. 1996. *SAS System for mixed models*. Cary, NC: SAS Institute Inc. 633 p.

Longford, N. T. 1993. *Random coefficient models*. Oxford: Oxford University Press.

McTague, J.P. and R.L. Bailey. 1987. Simultaneous total and merchantable volume equations and a compatible taper function for loblolly pine. *Can. J. For. Res.* 17:87-92.

Newberry, J.D. and L. V. Pienaar. 1978. Dominant height growth models and site index curves for site-prepared slash pine plantations in the lower coastal plain of Georgia and North Florida. *Univ. of Ga. Plantation Mgt. Res. Coop. Res.* Paper No. 4.

Pienaar, L.V., and J. W. Rheney. 1995. Modeling stand level growth and yield response to silvicultural treatments. *For. Sci.* 41(3) 629-638.

Pinheiro, J.C. and D.M. Bates. 2000. *Mixed-effects models in S and S-Plus.* Springer-Verlag. New York. 528 p.

Rao, C. R. 1987. Prediction of future observations in growth curve models. *Statist. Sci.* 2(4): 434-471.

Ratkowsky, D.A. 1983. *Nonlinear regression modeling: A unified practical approach.* Marcel Dekker. New York.

Ratkowsky, D.A. and T.J. Reedy. 1986. Choosing near-linear parameters in the four-parameter logistic model for radioligand and related assays. *Biometrics*, 42, 575-583.

Ratkowsky, D. A. 1990. *Handbook of nonlinear regression models*. Marcel Dekker, New York.

Sakamoto, Y., M. Ishiguro, and G. Kitagawa. 1986. Akaike Information Criterion Statistics, Kluwer Academic Publishers, Dordrecht, Holland.

Shiver, B.D., L.V. Pienaar, 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.

Srivastava, V. K. and D. E. A. Giles. 1987. *Seemingly unrelated regression equations models*. New York: Dekker.

Stiratelli, R., N.M. Laird, and J. Ware. 1984. Random effects models for serial observations with binary response. *Biometrics*, 40, 961-971.

Trevizol Jr., T.L. 1985. *Análise de um modelo compatível de crescimento e produção em plantações de Eucalyptus grandis* (W. Hill ex-Maiden). Viçosa, UFV, 1985. 74p.

Verbeke, G. and E. Lesaffre. 1996. A linear mixed-effects model with heterogeneity in the random-effects population. *Journal of the American Statistical Association*, 91, 217-221.

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. and V.M. Chinchilli, 1997. *Linear and nonlinear models for the analysis of repeated measurements.* Marcel Dekker, New York.

Walters, D.K., H.E. Burkhart, M.R. Reynolds, T.G. Gregoire. 1991. A Kalman filter approach to localizing height-age equations. *For. Sci*. 37:1526-1537.

Ware, J.H. 1985. Linear models for the analysis of longitudinal studies. *The American Statistician,* 39, 95-101.

Zahner, R. 1962. Loblolly pine site curves by soil groups. *For. Sci.* 8:104-110.

Zeger, S.L. and K.Y. Liang. 1986. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, 42, 121-130.

Zeger, S.L., K. Liang, and P. Albert. 1988. Models for longitudinal data: a generalized estimating equation approach. *Biometrics* 44: 1049-1060.