

---

# Nonlinear Mixed Effects Modeling for Slash Pine Dominant Height Growth Following Intensive Silvicultural Treatments

Zixing Fang and Robert L. Bailey

---

**ABSTRACT.** A modified Richard's growth model with nonlinear mixed effects is developed for modeling slash pine (*Pinus elliottii* Engelm.) dominant height growth in conjunction with different silvicultural treatments. All three parameters in the model turn out to have both fixed and random individual plot or silvicultural treatments effects. Moving average correlation with 2° was identified as within-plot error structure. The advantages of the mixed effects model in prediction for new responses are demonstrated in detail by formulations and examples. The modified Richards model has a form that combines dominant height growth and site index into one model form, so the incompatibility between height growth and site index model can be avoided. The general methodologies of nonlinear mixed effects model building, such as which parameters in the model should be considered to be random and which should be purely fixed, how to determine appropriate within-plot variance covariance structure, and how to specify between-plot variation via appropriate covariate modeling, are addressed in detail. Likelihood ratio test and Akaike information criterion (AIC) are used in model performance evaluation. Some useful graphical model diagnosis tools are also presented. *For. Sci.* 47(3):287–300.

**Key Words:** Random effects, site index model, base age invariant, prediction variance, repeated measurements.

---

**P**ERMANENT SAMPLE PLOTS, measured repeatedly over a fixed length of time, are often used for evaluating forest growth and yield and are especially effective as a sampling method to evaluate changes in forest conditions (Avery and Burkhart 1994, p. 208). In the practice of intensive forest management, some new attributes are added into the usual permanent plot sampling method. For example, to accelerate stand growth and development and to increase financial returns, silvicultural treatments such as mechanical and chemical site preparation and herbicide or fertilization applications are very common. To monitor and predict the changes of forest stands with different silvicultural treatments or factors, the permanent plots are usually established in a more careful way: they are actually built up from some standard experimental design. For example, a split-plot design is often used in the southeastern United States. In one such designed study, soil type serves as the whole-plot factor

and silvicultural treatment as the split-plot factor (Shiver et al. 1994). A split-plot design with repeated measurements naturally forms a split-split plot design with the time factor as the within-plot effect (Gumpertz and Brownie 1993).

In the forestry literature, a common approach to modeling such split-plot repeated measurement data from permanent plots is to only “adjust” the fixed part of the model with some additional fixed terms that partially explain the “gains” from the silvicultural treatments (Pienaar and Rheney 1995, Martin et al. 1999, Castleberry 1998). The advantage of this approach is its simplicity. It allows any treatment effects on forest growth and yield to be explicitly expressed in the model. However, the special properties of split-plot repeated-measurement data as discussed above are partially or totally ignored by this approach. For example, the variability of individual plots or within-plot correlation is rarely taken into account in such an approach. Consequently, different plots,

---

Zixing Fang, Biometrician, Division of Biostatistics, Department of Community Medicine and Health Care, University of Connecticut School of Medicine, 270 Farmington Ave., Exchange Suite 260, MC 6325, Farmington, CT 06030—Phone: (860) 679-5477; Fax: (860) 679-5464; E-mail: zfang@neuron.uhc.edu. Robert L. Bailey, Professor, D. B. Warnell School of Forest Resources, The University of Georgia, Athens, Georgia 30605—bailey@smokey.forestry.uga.edu.

Manuscript received February 11, 2000. Accepted January 24, 2001.

Copyright © 2001 by the Society of American Foresters

no matter how different their plot attributes, will obtain exactly the same “gains” as long as they are treated by the same silvicultural treatments. This is usually not realistic, however. Also, any information contained in past observations for a plot or stand is only partially applicable in this approach when predictions are desired for expected future values for a specified plot.

Gregoire et al (1995) recently presented a linear mixed effects model based on data from permanent plots and justified the necessity of an appropriate variance-covariance for the modeling of such data. While linear models can be made robust to assumptions about the correlation, especially when the number of observations per plot are small relative to the number of plots (Diggle et al. 1994, p. 79), this is not always true for nonlinear models. The expectation of an individual mean response usually does not coincide with the marginal mean of the population for a nonlinear model (Zeger et al. 1988, Vonesh and Chinchilli 1997, p. 295). Thus it is obviously inappropriate to use a population mean response as the prediction for an individual whose former information is available.

The idea of a random effects model is not new in forestry. Conceptually it goes back as early as Dr. J.L. Clutter’s Duke University Ph.D. dissertation, which is one of the earliest recognitions of the uniqueness of repeated measurements in a forestry context (Clutter 1961). Clutter noted that the independent randomness assumption in regression analysis is violated by the repeated sampling nature of data from permanent plots, and a corresponding adjustment in regression analysis may be necessary. His original assumption that observations from a particular plot may reflect an underlying pattern which is particular to that plot and a function of time as well is broadly implemented today by a random effects model both in statistics and in widespread applied fields, including forestry. Bailey and Clutter’s (1974) article, in which the concept of varying parameters uniquely identified with sites (plot-specific parameters) was introduced, can be considered the pioneering work in forestry. Thereafter, other models with varying or random coefficients were introduced in forestry (Garcia 1983, Borders and Bailey and Clutter 1984, Biging 1985, Lappi and Bailey, 1988, McDill and Amateis 1992, Lappi and Malinen 1994).

The methodology of linear and nonlinear mixed effects models has been quite well developed, and corresponding software has become publicly available recently (e.g., Vonesh and Chinchilli 1997). However, because of the complexity of model specification and the procedure of parameter estimation for random effects models, mixed effects models, especially nonlinear mixed effects models, have not become as popular with foresters as they should be.

Here we use repeated measures of slash pine (*Pinus elliottii* Engelm.) dominant height from a permanent plot study with a split-plot design to study the effects of intensive silvicultural treatments by using a nonlinear mixed effects model. We also present the techniques of model building and diagnosis for nonlinear mixed effects models with consideration for the special nature of the data.

## Data

The data used in this study come from a designed study for slash pine (*Pinus elliottii* Engelm.) installed by the Plantation Management Research Cooperative (PMRC) of the University of Georgia. The purpose of this study is to evaluate growth, yield, and stand structure of slash pine plantations with different site preparation treatments alone and in combination with fertilization and vegetation control (Pienaar et al. 1998). The original design used in the study is a split-plot with installation locations as whole plots and the permanent plots as the subplots. The whole-plot factor is the soil type, and the subplot factor is the silvicultural treatment. Eleven treatment plots, initially assigned randomly to 20 installation locations, were established on 4 types of soil across the flatwoods of Georgia to north Florida.<sup>1</sup> Some common stand characteristics such as dominant height and basal area have been measured up to 5 times on each treatment plot from age 5 to age 17 with 3 yr intervals since establishment in 1979. Seventeen years after installation, 16 of the original installations remain for the analysis. These comprise 7 nonspodosol and 9 spodosol soil groups. The total number of treatment plots is 191. Details of the data description were given by Pienaar et al. (1998). In the current analysis, we only focus on the response of dominant/codominant height.

## Method

### The General Dominant Height Growth Model

The well-known three-parameter Richards model (1959) serves as the basic dominant height growth model. The statistical expectation of the Richards model may be expressed as:

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

where  $t$  is stand age,  $f(t)$  is the mean response function of dominant height,  $A$  is the asymptote parameter which denotes the asymptotic value of the dominant height, and  $\beta_1$  is the rate parameter. Since  $A$  is the most unstable parameter to estimate with our data, it was replaced by an expected-value parameter,  $\beta_0$ , corresponding to the expected height when  $t = t_0$ . This allows model (1) to be represented as:

$$f(t, \beta) = \beta_0 \left( \frac{1 - e^{-\beta_1 t}}{1 - e^{-\beta_1 t_0}} \right)^{\beta_2} \quad (2)$$

Model (2) has exactly the same form as an algebraic difference project model (Borders and Bailey 1984), but with

<sup>1</sup> **Treatments:** CNTL: control (harvest and plant, no site preparation); UCHP: chop (single pass with a rolling drum chopper); FCHP: chop, fertilize; UCHB: chop, burn (chop followed by a broadcast burn); FCHB: chop, burn, fertilize; UCBB: chop, burn, bed; FCBB: chop, burn, bed, fertilize; UCBH: chop, burn, herbicide; FCBH: chop, burn, herbicide, fertilize; UBHB: chop, burn, bed, herbicide; FBHB: Chop, burn, bed, herbicide, fertilize. **Soil type**—A: poorly drained, nonspodosol; B: somewhat poorly to moderately well drained, nonspodosol; C: poorly to moderately well drained spodosol with an underlying argillic horizon; D: poorly to moderately well drained spodosol without an underlying argillic horizon.

one big difference in interpretation. In (2),  $\beta_0$  is treated as an unknown parameter (rather than a measured height) and  $t_0$  is treated as a given age, which may not necessarily be included in the data. There is no limit on  $t_0$  as long as it is positive, although a reasonable  $t_0$  should be chosen within the range of the data. The properties of model (2) can be summarized as follows:

1.  $\beta_0$  is the expected dominant/co-dominant height at given age  $t_0$ , thus, if  $t_0$  is taken as the site index reference age (25 yr for slash pine) then  $\beta_0$  corresponds to the site index; if  $t_0$  is taken as a reasonably large value, then  $\beta_0$  corresponds to the asymptote and (2) has the same form as (1).
2. For a given data set, the estimates of  $\beta_1$  and  $\beta_2$  will be unchanged for different choices of  $t_0$ , i.e. changing  $t_0$  only affects the estimate of  $\beta_0$ ;
3. The asymptotic height is  $\beta_0/[1 - \exp(-\beta_1 t_0)]^{\beta_2}$ .

The advantages of models with expected-value parameter(s) were discussed by Schnute (1981), Cieszewski and Bella (1989), and Cieszewski and Bailey (2000).

### Nonlinear Mixed Effects Model

To apply Equation (2) to the real data, various statistical models will result in different statistical assumptions. Let  $y_{ij}$  denote a dominant height measurement at occasion  $j$  for the  $i$ th plot ( $i = 1, \dots, n; j = 1, \dots, m_i$ ).  $m_i$  is the total number of measurements on plot  $i$ ;  $e_{ij}$  denotes the corresponding residual for  $y_{ij}$  with the model; and the stand age for plot  $i$  at occasion  $j$  is  $t_{ij}$ . The function  $f$  (Equation 2) is assumed to be common to all plots, but the parameter estimates may vary across plots. In vector form, let  $\mathbf{y}_i = (y_{i1}, \dots, y_{im_i})^T$ ,  $\mathbf{e}_i = (e_{i1}, \dots, e_{im_i})^T$ ,  $\mathbf{t}_i = (t_{i1}, \dots, t_{im_i})^T$  and  $\boldsymbol{\beta}_i = (\beta_{i0}, \beta_{i1}, \beta_{i2})^T$ . There are some stand characteristics corresponding to an individual plot, such as installation number of the plot, the soil group, and the silvicultural treatment, which have not been explicitly expressed thus far, but they will be addressed in the section on model specification. We adopt the two-stage model formulation (Lindstrom and Bates 1990, Davidian and Giltinan 1995, p. 99) in order to write the general nonlinear model as:

#### Stage 1: Within-Plot Variation

$$\left. \begin{aligned} y_i &= f(t_i, \boldsymbol{\beta}_i) + e_i \\ e_i | \boldsymbol{\beta}_i &\stackrel{iid}{\sim} N[0, R_i(\boldsymbol{\beta}_i, \boldsymbol{\alpha}, \boldsymbol{\theta})] \\ \mu_i &= E[y_i | \boldsymbol{\beta}_i] = f(t_i, \boldsymbol{\beta}_i) \\ f(t_i, \boldsymbol{\beta}_i) &= \beta_{0i} \left( \frac{1 - e^{-\beta_{1i} t_i}}{1 - e^{-\beta_{1i} t_0}} \right)^{\beta_{2i}} \end{aligned} \right\} \quad (3)$$

In (3),  $R_i(\boldsymbol{\beta}_i, \boldsymbol{\alpha}, \boldsymbol{\theta})$  is an  $m_i \times m_i$  within-plot variance-covariance matrix, which may depend on parameter  $\boldsymbol{\beta}_i$  (via the mean function) and some other covariance parameters  $\boldsymbol{\theta}$ . From Davidian and Giltinan (1995, p. 26), one may formalize the within-individual variance-covariance as:

$$\begin{aligned} Cov(e_i | \boldsymbol{\beta}_i) &= R_i(\boldsymbol{\beta}_i, \boldsymbol{\alpha}, \boldsymbol{\theta}) \\ &= \sigma^2 G_i^{1/2}(\boldsymbol{\beta}_i, \boldsymbol{\alpha}) \Gamma_i(\boldsymbol{\theta}) G_i^{1/2}(\boldsymbol{\beta}_i, \boldsymbol{\alpha}) \end{aligned} \quad (4)$$

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

#### Stage 2: Between-Plot Variation

$$\left. \begin{aligned} \boldsymbol{\beta}_i &= A_i \boldsymbol{\beta} + B_i b_i \\ b_i &\sim N(0, D) \end{aligned} \right\} \quad (5)$$

In (5),  $A_i$  is a  $3 \times p$  design matrix for  $p \times 1$  fixed effects  $\boldsymbol{\beta}$ ,  $b_i$  is the  $q \times 1$  vector of random effects associated with the  $i$ th plot, and  $B_i$  is the corresponding design matrix.  $D$  is a  $(q \times q)$  covariance matrix for the random effects. The design matrix  $A_i$  and matrix  $B_i$  are determined by the stand characteristics, such as silvicultural treatments and soil types, which are usually, but not necessarily, matrices containing only zeroes and ones as elements.

Equations (3) to (5) represent the general formulation of a nonlinear mixed effects model based on the deterministic dominant height growth equation, Equation (2). Different models can be obtained with distinct variance-covariance specifications.

#### Model Specification

To apply the model represented by (3)–(5) with a real set of data, one needs to:

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

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

#### Step 1: Determine Parameter Effects

In modeling, which effects should be considered mixed (both fixed and random) and which should be considered purely fixed are generally data dependent. For example, Lappi and Bailey (1988) chose the rate parameters  $\beta_1$  as fixed while the asymptote parameter and  $\beta_2$  were considered random when they used Richards model for dominant height growth. They did so simply because it is impossible to obtain reasonable estimates with all parameters random due to collinearity problems in their data. Intuitively, one simple

approach to this question is to obtain separate fits for each plot and assess the variability of estimated parameters across plots by considering the individual confidence intervals for the parameters. The parameters with high variability and less overlap in confidence intervals across plots should be considered as mixed effects. This approach requires sufficient observations on each plot to give meaningful parameter estimates by separate fitting. However, in forestry, repeated measurements from permanent plots usually do not cover a long enough time span or contain enough degrees of freedom to produce stable individual parameter estimates. In our case there are only five observations on each plot and three parameters to be estimated. In this situation, the separate-fittings approach is not likely to be helpful in judging the nature of the parameters.

If no prior information about the random effects variance-covariance structure is available, and convergence is possible, Pinheiro and Bates (1998) suggested that all parameters in the model should first be considered mixed (both random and fixed). After the initial estimates are available, the eigenvalues of the estimated covariance matrix of the effects ( $D$  matrix) may be studied to see if one or more are close to zero. The associated eigenvector(s) for such eigenvalues will then give an indication of the linear combinations of parameters that could be taken as fixed. Since small eigenvalues may arise when the relative magnitudes of the scales of the parameters in the model are quite different, the “normalized” variance-covariance matrix [i.e., the coefficient of variation (CV) matrix  $D_{cv}$ ] was recommended instead of the variance-covariance matrix [Pinheiro and Bates (1998)].

$$[D_{cv}]_{lm} = \frac{[D]_{lm}}{|\beta_{k(l)}\beta_{k(m)}|} \quad (6)$$

In (6),  $\beta_k$  is the  $k$ th fixed effect and  $k(l)$ ,  $k(m)$  are the fixed effects associated with the  $l$ th and the  $m$ th random effects. This approach is still an ad hoc method, for there is no decision rule to indicate how close an eigenvalue may be to zero and not indicate rank deficiency.

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

$$LRT = 2\log(L_0 / L_1) = 2 [\log(L_0) - \log(L_1)] \quad (7)$$

Under the null hypothesis that the restricted model is adequate, the asymptotic (large sample) distribution of  $LRT$  is a  $\chi^2$  with  $k_0 - k_1$  degree of freedom. If  $LRT > \chi^2(k_0 - k_1, \alpha)$  then one rejects the null hypothesis with  $(1 - \alpha)\%$  confidence that the full model should not be reduced to the restricted model. Here  $\alpha$  is some specified percentile of the Chi-squared distribution.

The Akaike Information Criterion (AIC) is defined as (Akaike 1972):

$$AIC = -2 \log(L) + 2k \quad (8)$$

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

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

$$\beta_i = \beta + b_i \quad (9)$$

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

It is worth noting that the interplot variation Equation (9) is just a special case of the general formulation of interindividual variation [see (5)], in which no covariate was put into the random effect. In other words, it is simply a random coefficient model and both design matrices  $A_i$  and  $B_i$  in (5) are  $3 \times 3$  identity matrixes in this case.

When all three parameters are mixed, the estimated coefficient of variation matrix  $D_{cv}$  is<sup>2</sup>:

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

The eigenvalues of  $\hat{D}_{cv}$  are 0.02181, 0.018286, and 5.2014  $\times 10^{-6}$ . There is some evidence of rank deficiency, but the evidence is not so strong. The eigenvector corresponding to the smallest eigenvalue, converted back to the original scale of the random effects and normalized, is  $(-0.002193, -0.99948, -0.03201)$ . Even though both the first and the third component of this eigenvector are very small, it is hard to conclude that they are close to each other because one is more than ten times the other in magnitude. So a formal likelihood ratio test is necessary in this case to determine the effects of the parameters. We first took all three parameters as mixed effects without considering any covariates (this is referred as the full model), and then picked each of the three parameters as purely fixed in turn. This resulted in three reduced models. Since the reduced models are nested within the full model, likelihood ratio tests can be carried out to check if the

<sup>2</sup> Maximum likelihood method was used in the parameter estimation (ref. Pinheiro and Bates 2000).  $t_0$  in (3) was taken as  $t_0 = 25$  yr all through this study.

**Table 1. Likelihood ratio tests (LRT) for nonlinear mixed-effects models of slash pine dominant height with different fixed and random effects components.**

Model		No.* of parameters	AIC	Log-likelihood †	LRT	P-value
Mixed effects	Fixed only					
$\beta_0, \beta_1, \beta_2$	None	10	1,731.185	-855.593		
$\beta_0, \beta_1$	$\beta_2$	7	1,785.984	-885.992	60.799	<0.0001
$\beta_1, \beta_2$	$\beta_0$	7	2,160.942	-1,056.455	401.72	<0.0001
$\beta_0, \beta_2$	$\beta_1$	7	1,739.007	-862.504	13.822	<0.0032

\* The number of parameters here includes both the parameters in the mean function (three in this case) and variance-covariance (six for the full model and three for the reduced model) and add another deviance parameter  $\sigma$ .

† Maximum likelihood method was applied in parameter estimation.

reduction in parameters caused any significant changes in model performance. Replacing any mixed effects parameter with a purely fixed effect significantly lowered the likelihood values (Table 1). Therefore, preference must be given to all three parameters being considered as mixed effects. Of course, results may be different with different covariates for the fixed effects and variance-covariance structure for the random effects, but the above procedure should be the same.

**Step 2: Determine Within-Plot Variance-Covariance Structure**

To specify the within-plot variance-covariance structure, as implied in (4), two components should be addressed: one is the heteroscedasticity, and the other is autocorrelation structure. Forest growth and yield data from permanent plots usually exhibit both autocorrelation and heteroscedasticity (Gregoire 1987).

**1. Variance Function**

The variances of errors around growth and yield models are often found to be dependent on the means; larger means usually have larger variance. Some frequently used variance functions for growth modeling include (Davidian and Giltinan 1995, p. 23) the power function model,

$$g(\mu_{ij}, \alpha) = g(\mu_{ij}, \alpha) = \mu_{ij}^\alpha \tag{10}$$

and the exponential function model,

$$g(\mu_{ij}, \alpha) = \exp(\alpha\mu_{ij}), \tag{11}$$

where  $\mu_{ij}$  is the mean function, which is defined in (3).

Note that both of the above functions imply that the variance of responses depends on the regression parameters through the means. Random effects in the mean function may remove some heterogeneity in variance. This is not hard to understand. Since the variances depend on the mean response through the parameters, random effects in parameters will definitely affect the underlying distribution of error of the

model and thus the variance. LRT (if the alternative models are nested) or AIC (if not nested) may be used to determine an appropriate variance function for the model.

The fitting comparison of a full mixed-effects model with both exponential and power variance functions indicates that the LRT tests are significant for both of these variance functions (Table 2). So even with random effects in the parameters, heteroscedasticity still exists in the mixed-effects dominant height model. With a power function to model the variance, a smaller AIC value results than with an exponential function. Since both have the same number of parameters in the model, the power function is judged superior in this case.

**2. Serial Correlation Structure**

Forest growth and yield data from permanent plots may be correlated over time, and thus some adjustment in regression techniques might be necessary (Leak 1966, Gregoire 1987, Lappi and Malinen 1994). The empirical autocorrelation function (Box et al. 1994) can be used to investigate serial correlation. Let  $r_{ij}$  denote the standardized residuals from a fitted mixed model for individual  $i(i = 1, \dots, n)$  at occasion  $j(j = 1, \dots, n_i)$ . The empirical autocorrelation at lag  $l$  is defined as (Pinheiro and Bates 2000):

$$\hat{\rho}(l) = \frac{\frac{1}{N(l)} \sum_{i=1}^n \sum_j^{n_i-l} r_{ij} r_{i(j+l)}}{\frac{1}{N} \sum_{i=1}^n \sum_j^{n_i} r_{ij}^2}, \tag{12}$$

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

$$N = \sum_{i=1}^n n_i$$

is the total number of the observations.

An approximate two-sided critical value for autocorrelation  $\hat{\rho}(l)$  at significance level  $\alpha$  is given by  $z_{(1-\alpha/2)} / \sqrt{N(l)}$ , where

**Table 2. Comparisons of mixed-effects model performance for slash pine dominant height growth data with different variance functions.**

Model	Variance function	No. of parameters	AIC	Log-likelihood	LRT†	P-value
1	1*	10	1,731.185	-855.593		
2	Exponential	11	1,703.053	-840.527	30.132	<0.0001
3	Power	11	1,697.299	-837.650	35.886	<0.0001

\* Variance function 1 means that the variances are homogeneous.

† Likelihood ratio is calculated with respect to Model 1.

$z_{(1-\alpha/2)}$  is the standard normal quantile at percentile  $1 - \alpha/2$  and  $N(l)$  is defined in (12).

The estimated empirical correlation for Equation (3) in Table 2 is

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

The autocorrelation looks “strange” in that the first three lags are all negatively correlated, and the fourth lag is significantly positively correlated. This is not consistent with our intuition, because one usually expects that observations spaced closely in time should be positively associated. However, such a counter-intuitive autocorrelation is not uncommon in the real world (e.g., Davidian and Giltinan 1995, p. 133). This is an excellent example to show the complexity of within-plot autocorrelation. Knowing that the empirical autocorrelation is significant, it is necessary to modify Model 3 (in Table 2) to reflect this within-plot autocorrelation.

One of the most commonly used within-individual autocorrelation structures with repeated measurement data is  $AR(1)$ , i.e., autoregressive model with order 1 (Gregoire et al. 1995). However, there was no significant improvement when we included an  $AR(1)$  autocorrelation structure with mixed effects Model 3 (see Table 3). Therefore, an inappropriate correlation structure can result in no improvement in fitting even though there is some evidence of correlation in the error. This may partially explain why several authors in the forestry literature concluded that modeling the within-plot autocorrelation produced no improvements in forest growth and yield prediction (see, for example, Sullivan and Clutter 1972). After trying several models, we found that a moving average correlation model was the best for improvement in the mixed effects version of Model 3 (see Table 3).

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

$$e_t = \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t, \quad (13)$$

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

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

$$\Gamma(k, \theta) = \begin{cases} \frac{\theta_k + \theta_1 \theta_{k-1} + \dots + \theta_{k-q} \theta_q}{1 + \theta_1^2 + \dots + \theta_q^2}, & k = 1, \dots, q \\ 0 & k = q + 1, q + 2, \dots \end{cases} \quad (14)$$

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

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

### Step 3: Specify Between-Plot Variation: Covariate Modeling

After determining the nature of the parameters in the model (random or fixed), an immediate question is how to track the random effects parameters. In this step, we address the question of which variables in the survey or experiment are potentially useful in explaining random effects variation and how the random effects are explained by those variables [i.e., determine the design matrix in Equation (5)]. This is a procedure similar to variable selection in ordinary regression analysis and should be primarily determined by the design of the experiment and the variables in the survey.

It will be informative as a first look to check the correlation matrix of the random effects from the final model in Step 2. High correlation among the random effects usually indicates that some similar patterns exist among the design components which may be explained by some other covariates.

All correlations among the three random effects are moderately high giving us a strong message that a covariate model for the mixed effects would be useful in explaining variation (Figure 1).

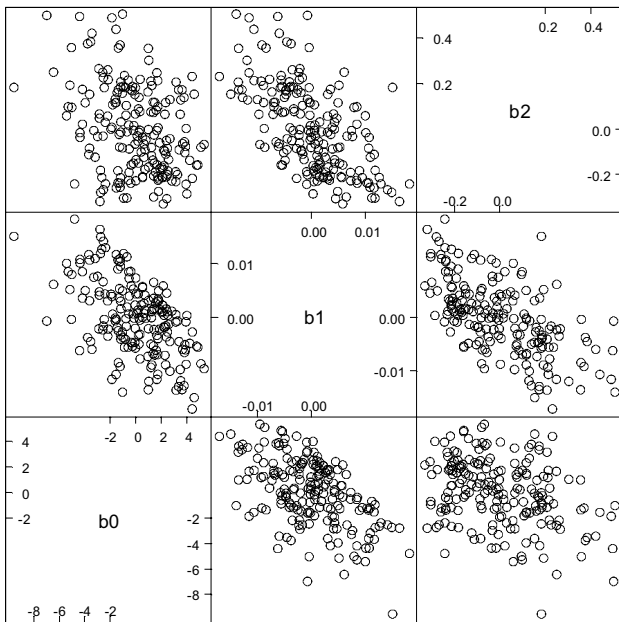
Another useful graphical device for investigating possible relationships between mixed effects and individual plot attributes (such as silvicultural treatments, soil type, etc.) is to plot estimates of the random effects against potential covariates. The empirical Bayes estimates of random effects in Equation (4) from Step 2 are plotted conditionally on silvicultural treatments (see Figure 2) and soil type (see Figure 3). Visually, silvicultural treatments are more likely to be the significant factor than the soil types. A formal split-plot analysis of variance (e.g., Lentner and Bishop, 1986, Chap. 11) to the three random effects shows that silvicultural treatment is the only significant term to the random effects. Soil type and the

**Table 3. Comparisons of mixed-effects model performance for slash pine dominant height growth data with different within-plot correlation structures and no interplot covariates.**

Model	Correlation* structure	No. of parameters	AIC	Log-likelihood	LRT†	P-value
3	Independent	11	1,697.299	-837.650		
	$AR(1)$	12	1,697.092	-836.544	2.2078	0.1373
	$ARMA(1,1)$	13	1,695.872	-834.936	5.4278	0.0663
4	$MA(2)$	13	1,694.808	-834.404	6.4919	0.0389

\* Refer to Bates and Pinheiro 2000.

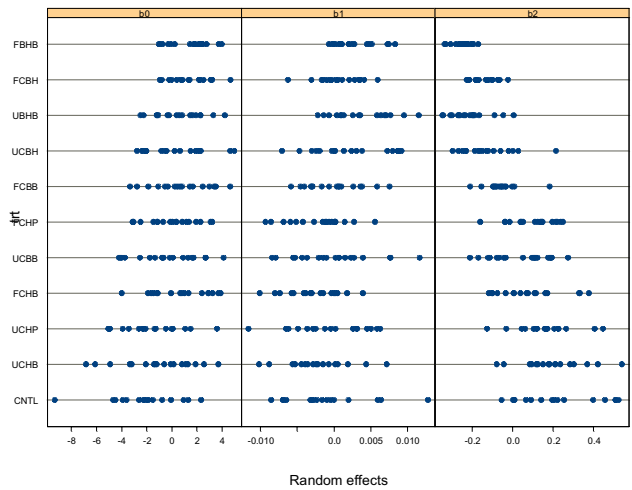
† Likelihood ratio is calculated with respect to Model 3.



**Figure 1.** Scatter plot of estimated random effects from Model 4 for the modified Richards dominant height growth model with random effects and independent normal within-plot errors.

interactions are not significant. This conclusion is also consistent with reports by other researchers using different approaches with these same data (e.g., Castleberry 1998).

Even though the general silvicultural treatments are significant for all the mixed effects, it may well be an overparameterization to consider every silvicultural treatment as a fixed effect. There are 11 levels of treatments (including the control), and the differences between response curves for some of them are very small. One way around this is to take the significant treatments as fixed effects and the others as random. But some problems still remain because terms significant in the analysis of variance (ANOVA) are not necessarily significant when they are put into the model as fixed effects. One of the attributes of the original study is that the silvicultural treatments are arranged in an additive way. Five distinct silvicultural activities (chop, burn, bed, fertilizer, and herbicide) are applied to individual treatment plots individually or in combination. For example, treatment FCHB is the combination of chopping, burning, and fertilization. By assuming that the effects of the silvicultural treatments on the response (dominant height growth) are additive (e.g., Pienaar et al. 1998) some dimension reduction can be obtained. Only 5 fixed effects will be involved instead of 11. However, it is still unrealistic to take all 5 silvicultural treatments as fixed in the 3 mixed effects ( $\beta_{i,0}$ ,  $\beta_{i,1}$  and  $\beta_{i,2}$ ). After initial efforts, effects of burning and fertilizer were taken as fixed to  $\beta_{i,0}$ , effects of fertilizer and herbicide fixed to  $\beta_{i,1}$ , and effects of fertilizer, bed, and herbicide fixed to  $\beta_{i,2}$ . For the random effects terms, the intercepts of the first two parameters ( $\beta_0$  and  $\beta_1$ ) are chosen as random, which may explain the combined random effects such as soil types, silvicultural treatment combinations, installations, or individual plot. Chopping has a random effect on parameter  $\beta_2$ . Let  $zchop_i$ ,  $zburn_i$ ,  $zf_i$ ,  $zbed_i$  and  $zh_i$  be respective dummy

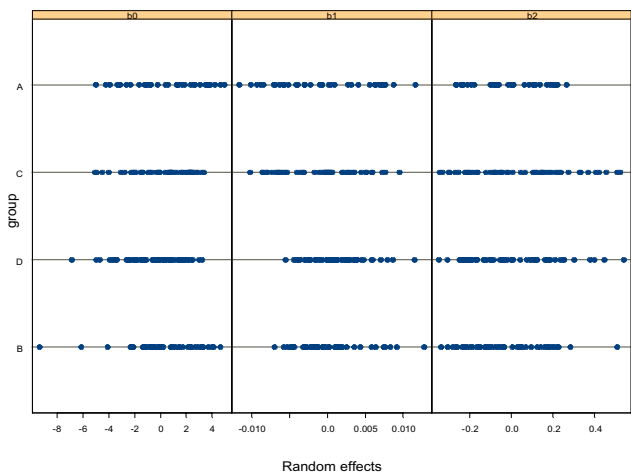


**Figure 2.** Comparisons of estimated random effects by silvicultural treatments for the modified Richards dominant height growth model with random effects and independent normal within-plot errors.

variables indicating if chopping, burning, fertilizer, bed, or herbicide treatment were applied on a specified treatment plot  $i$ . The interplot formulation for the slash pine dominant height growth may be represented as:

$$\beta_i = \begin{pmatrix} \beta_{i,0} \\ \beta_{i,1} \\ \beta_{i,2} \end{pmatrix} = \begin{pmatrix} 1 & zburn_i & zf_i & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & zf_i & zh_i & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & zf_i & zbed_i & zh_i \end{pmatrix} \begin{pmatrix} \beta_{00} \\ \beta_{01} \\ \beta_{02} \\ \beta_{10} \\ \beta_{11} \\ \beta_{12} \\ \beta_{20} \\ \beta_{21} \\ \beta_{22} \\ \beta_{23} \end{pmatrix} + \begin{pmatrix} b_{i,0} \\ b_{i,1} \\ zchop_i \times b_{i,2} \end{pmatrix} \quad (15)$$

where  $\beta_{00}, \dots, \beta_{23}$  denote the fixed effects. Parameters  $\beta_{00}, \beta_{10}, \beta_{20}$  are, respectively, the intercepts for fixed effects. Parameter vectors  $(\beta_{01}, \beta_{02}), (\beta_{11}, \beta_{12}), (\beta_{21}, \beta_{22}, \beta_{23})$  are the fixed “gains” on parameters  $\beta_{i,0}, \beta_{i,1}, \beta_{i,2}$ , respectively, because of burning and fertilization, fertilization and herbicide, fertilizer, bed, and herbicide.



**Figure 3. Comparisons of estimated random effects by soil types for the modified Richards dominant height growth model with random effects and independent normal within-plot errors.**

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

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

Therefore, matrices  $A_i$  and  $B_i$  in (5) are specified as:

$$A_i =$$

$$\begin{pmatrix} 1 & zburn_i & zfi & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & zfi & zhi & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & zfi & zbed_i & zhi \end{pmatrix}$$

$$B_i = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & zchop_i \end{pmatrix}$$

Once appropriate covariate models (design matrices  $A_i$ ,  $B_i$ ) are determined, one still needs to go back to Step 2, to choose an appropriate within-plot variance function and correlation structure. Usually, Step 3 and Step 2 need to be repeated until a satisfactory model is obtained. Just as before, the likelihood ratio test and the AIC values indicate that both heterogeneity and correlation are evident even after including covariates in the model. For example, both the exponen-

tial and the power functions significantly improve model fitting. Moreover, correlation structures  $AR(1)$ ,  $ARMA(1,1)$ , and  $MA(2)$  are all significant compared to the model with an independent structure. A power variance function combined with the moving average correlation with  $q = 2$  denote the best among these models (Table 4).

Thus the final specified model has the following form:

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

$$\mu_i = E[y_i | \beta_i] = f(t_i, \beta_i)$$

$$\mu_i = (\beta_{00} + \beta_{01} zburn_i + \beta_{02} zfi + b_{i,0})$$

$$\left( \frac{1 - e^{-(\beta_{10} + \beta_{11} zfi + \beta_{12} zhi + b_{i,1}) t_i}}{1 - e^{-(\beta_{10} + \beta_{11} zfi + \beta_{12} zhi + b_{i,1}) t_0}} \right)^{(\beta_{20} + \beta_{21} zfi + \beta_{22} zbed_i + \beta_{23} zhi + b_{i,2} zchop_i)}$$

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

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

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

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

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

where  $MA(2)$  denotes a moving average correlation model with two parameters as defined in (13) and (14). There are 20 parameters (including the parameters in the variance-covariance matrices) to estimate in the above mixed effects model. Individual random effects can be predicted by first-order approximation (Lindstrom and Bates 1990)

#### Parameter Estimation and Model Diagnosis

Parameter estimation for nonlinear mixed effects models requires numerical integration of random effects in the model. This procedure can be computationally difficult as the random effects often enter the model nonlinearly (e.g.,  $b_{i,1}$  and  $b_{i,2}$  in the above model). Different approximations have been used to deal with interindividual random effects, among those linear approximation with a first-order Taylor expansion in all the interindividual random effects is the most common method. The Taylor expansion can be either at 0 (the mean of the random effects) or at the empirical best linear unbiased predictor (EBLUP) of the random effects. The first approach is cheaper in term of computing time, but the second approach can be more accurate

**Table 4. Comparisons of mixed effects model performance for slash pine dominant height growth data with between-plot covariates [Equation (15)] and different within-plot variance function and correlation structures.**

Model	Variance function	Correlation structure	No. of parameters	AIC	Log-likelihood	LRT <sup>†</sup>	P-value
5	1*	Independent	17	1,524.069	-745.035		
6	Exp.	Independent	18	1,499.589	-731.795	26.480	<0.0001
7	Power	Independent	18	1,493.780	-728.890	32.289	<0.0001
8	Power	AR(1)	19	1,489.443	-725.722	6.337	0.0118
9	Power	ARMA(1,1)	20	1,489.831	-724.916	7.949	0.0188
10	Power	MA(2)	20	1,487.933	-723.966	9.848	0.0073

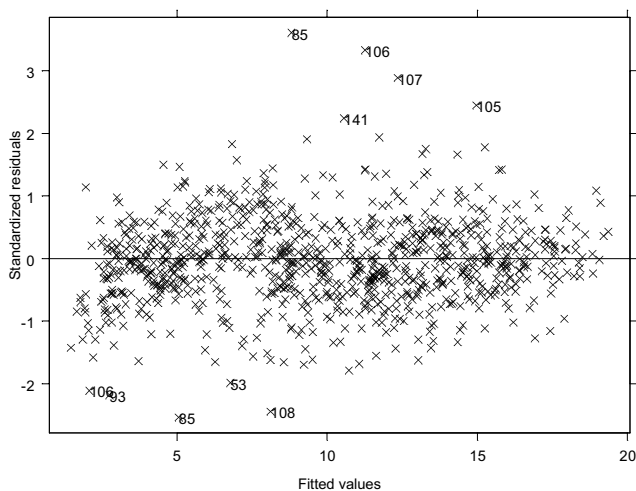
<sup>†</sup> For Models 6 and 7, the likelihood ratios are calculated with respect to Model 5; for Models 8, 9, and 10, the likelihood ratios are calculated with respect to Model 7.

\* Variance function 1 means that the variances are homogeneous.



although possibly more unstable (Wolfinger and Lin 1997). Other alternative first-order Taylor series expansions have been proposed. For example, Vonesh and Carter (1992) proposed the use of estimated generalized least squares and established the asymptotic properties of the resulting estimates. Except for the first-order approximation, an adaptive Gaussian quadrature approximation has been optionally available within a new SAS procedure NL MIXED (available SAS 7.0 or later). Different parameter estimation methods have been implemented into various commercial software package, such as S-plus and SAS. A comprehensive review has been presented by Roe (1997). We use the EBLUP approximation approach, which uses a first-order Taylor series expansion about conditional estimates of the interindividual random effects. Detailed descriptions for this method can be found in Davidian and Giltinan (1995). S-plus software implementations are available from the NLME library by Pinheiro and Bates (2000). Model 10 fits well to the data and all ten fixed-effect parameters were significant with generally low covariances between pairs of parameters.

Comparing Model 10 with Model 4 (see Tables 3 and 4), one can see that the model performance has remarkably improved by including the interplot covariates (silvicultural treatments). For example, the log-likelihood value increased from  $-834.404$  to  $-723.966$ . This gives a LRT of 220.876 with  $20 - 13 = 7$  degrees of freedom, which is significant at any common significance level. A closer diagnosis of the final model can be obtained by checking plots of the residuals, the fitted values, and the estimated random effects. Because the within-plot errors are assumed to be independent and normally distributed, given the appropriate within-plot variance-covariance [ $R$  matrix in (4)] the residual plot (residuals or standardized residuals against the fitted values) should not reveal any observable pattern, and the Q-Q plot should be approximately linear. The residual and Q-Q plots of Model 10 do not indicate serious departures of randomness and normality. The numbers identified on the residual plots are the observations which are potential outliers in that the corresponding standardized residuals are outside of the range 2.5 to 97.5 standard normal quantiles (see Figure 4). Overall, the standardized residuals are small, thus suggesting the



**Figure 4.** The residual plot for the modified Richards dominant height growth model with random between-plot effects and MA(2) within-plot autocorrelation and power variance function [Model (10)].

nonlinear mixed effects Model 10 was successful in explaining the dominant height growth data.

## Interpretation of Model Parameters

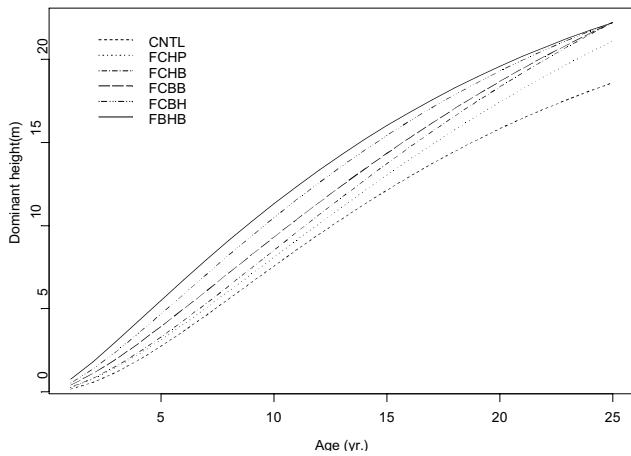
### “Typical” Responses for Different Silvicultural Treatments

Parameter interpretation for a nonlinear mixed effects model will usually be quite different depending on whether one chooses to model the response  $y_i$  at the individual level or to consider population behavior and the marginal moments directly. Zeger et al. (1988) refer to the case of the hierarchical model for an individual as a *subject-specific* model, and that of a model for marginal moments as *population-averaged* model. We have chosen a subject-specific approach thus far. A higher precision for individual prediction usually results with this approach. However, it is still possible in this approach to assess the “typical” response for an individual or some grouping of individuals. For example, we assume the random individual treatment plot effects to be 0 and then this “typical” dominant height growth for a given individual (or group) can be obtained by just substituting corresponding fixed effects into the final model (Model 10). We obtained parameters in the modified Richards model for 11 silvicultural treatment groups with random effects set to zero (Table 5).

One may use the parameters in Table 5 to obtain the “typical” dominant height growth curve for each individual silvicultural treatment. For example, we constructed a comparison of the “typical” dominant height growth for the following six silvicultural treatments: FCHP, FCHB, FCBB, FCBH, FBHB, and CNTL. Treatment FBHB has highest growth response (Figure 5). It is also the most comprehensive treatment combination. CNTL has lowest growth response and has no silvicultural treatment included. While this model is still informative to assess the “typical” or mean responses for individual prediction, information on random effects is always helpful. As an example, we graphed a comparison of predicted curves with and without random effects for an experimental plot in the original data (a treatment plot in Installation 7, soil type C) with silvicultural treatment FCBH. From this it is clear that one obtains much more precise prediction with estimated random effects included in the model (Figure 6).

**Table 5.** The estimated parameters for “typical” dominant height growth (random effects input as 0) with modified Richards model by silvicultural treatment groups.

Silvicultural treatments	Estimated parameters		
	$\beta_0$	$\beta_1$	$\beta_2$
CNTL	18.57601	0.07859702	1.967684
UCHP	18.57601	0.07859702	1.967684
FCHP	21.10557	0.05432134	1.675987
UCHB	19.67853	0.07859702	1.967684
FCHB	22.20809	0.05432134	1.675987
UCBB	19.67853	0.07859702	1.813103
FCBB	22.20809	0.05432134	1.521406
UCBH	19.67853	0.09306127	1.792522
FCBH	22.20809	0.06878559	1.500825
UBHB	19.67853	0.09306127	1.637942
FBHB	22.20809	0.06878559	1.346245



**Figure 5. A comparison of the mean responses (typical responses) of dominant height growth by six different silvicultural treatments (CNTL, FCHP, FCHB, FCBB, FCBH, FBHB) with the nonlinear mixed-effects dominant height growth model.**

### Localization of Site Index Curves

Age  $t_0$  in the modified dominant height model [model (2)] was chosen to be 25, which is often used as the reference age for slash pine. The estimated parameter  $\hat{\beta}_{i,0}$  is naturally interpreted to be the estimated site index for plot  $i$ . So the dominant height growth and the site index prediction models are combined into one model in the current approach. For example, to obtain the site index curve, one may solve for  $\beta_0$  directly from the current dominant height growth model or estimate a series of  $\beta_0$ 's by changing  $t_0$ . With the second approach, one can not only obtain a site index curve for an individual plot, but also estimate upper and lower bounds on this curve. This naturally provides a method to label site index curves for slash pine plantations with silvicultural treatments.

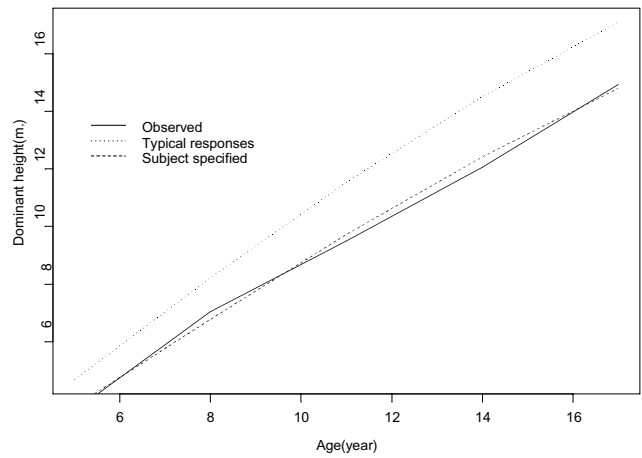
It is easy to show that the nonlinear mixed effects model based on the Richards model (10) is polymorphic in that varying rates of height growth on different sites are reflected. Model 10 is also base age invariant. That is, predictions and curves are invariant with respect to the changes in the base age for site index (Bailey and Clutter 1974). It is also interesting to note that site index is affected by silvicultural treatments. This has been demonstrated not only because some silvicultural treatments (e.g., burning and fertilization) are significant for parameter  $\beta_0$ , but also because the dominant height growth curve is affected by silvicultural treatments (see Figure 5).

### Individual Prediction for New Observations

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

#### Case 1: Prediction Is Required for a New Individual with No Prior Observations

Suppose we are interested in the prediction of the dominant height on plot  $k$  at age  $t_{kh}$ , and there is no prior



**Figure 6. A comparison of individual dominant height growth predictions by typical response and subject-specified response with the nonlinear mixed-effects dominant height growth model.**

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

$$\hat{y}_{kh} = f(\hat{\beta}_k, t_{kh}) = \hat{\beta}_{k0} \left( \frac{1 - e^{-\hat{\beta}_{k1} t_{kh}}}{1 - e^{-\hat{\beta}_{k1} t_0}} \right)^{\hat{\beta}_{k2}} \quad (16)$$

with prediction variance:

$$\text{var}(y_{kh} - \hat{y}_{kh}) = F_k^T(\hat{\beta}_k) \left( A_k \hat{\Omega} A_k^T + B_k \hat{D} B_k^T \right) F_k(\hat{\beta}_k) + \hat{\sigma}^2 R_{kh} \quad (17)$$

where

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

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

$R_{kh} = f^{\hat{\alpha}}(\hat{\beta}_k, t_{kh})$ ,  $t_0 = 25$ ,  $\hat{\Omega}$  is a  $10 \times 10$  covariance matrix corresponding to the estimated parameter vector  $\hat{\beta}_K$ . The  $\hat{\beta}_{\dots}$ 's,  $\hat{D}$ ,  $\hat{\sigma}$ , and  $\hat{\alpha}$  are the estimated parameters. All other terms are defined above.

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

$$\hat{y}_{kh} \pm t(n - q, 1 - \alpha / 2) \sqrt{\text{var}(\hat{y}_{kh})} \quad (18)$$

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

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

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

thus

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

$$= \begin{pmatrix} 22.20809 \\ 0.068786 \\ 1.500825 \end{pmatrix}$$

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

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

$$F_k^T = (0.8686262 \quad 37.81246 \quad -1.810287).$$

Substitute  $\hat{\Omega}$ ,  $\hat{D}$ ,  $\hat{\sigma}$ ,  $\hat{\alpha}$ , and design matrices into (17) to obtain a variance of this prediction:  $\text{var}(\hat{y}_{kh}) = 3.8758$ . The 95% confidence interval for the prediction is [15.43 m, 23.15 m]

**Case 2 Prediction Is Required for a New Individual with Prior Observations**

An alternate, and perhaps more practical, situation in forestry occurs when a prediction is required for a known individual for which a collection of formerly observed data is

available. One may still use the procedure in Case 1 to obtain the prediction and corresponding prediction limits just as with traditional regression. However, this approach would ignore any information contained in the prior observations of the specified individual. Instead, we may use the information contained in former observations to estimate the subject-specific random effects for this individual and incorporate this directly into our prediction.

Suppose the information available for individual plot  $k$  is  $\{y_k, t_k, A_k, B_k\}$ , where  $y_k$  is a vector of former observed dominant heights at corresponding age vector  $t_k$ . We wish to predict a future dominant height  $y_{kh}$  at a given age  $t_{kh}$ . We first estimate the random effect  $b_k$ . We can take a first order expansion of  $f(t_k, A_k\beta_k + B_k b_k)$  about  $b_k = 0$ . From the residual,  $e_k = y_k - f(t_k, A_k\beta_k)$ , one can estimate  $b_k$  by the post mean and variance of  $(b_k | y_k)$ ,  $(\beta_k, \xi_k)$ , where  $\xi_k$  is the vector of variance covariance parameters for individual  $k$  [see the general model form (4)]. For details refer to Vonesh and Chinchilli (1997, Chap. 7).

$$\hat{b}_k = \hat{D}\hat{Z}_k^T (\hat{Z}_k\hat{D}\hat{Z}_k^T + \hat{R}_k)^{-1} \hat{e}_k \tag{19}$$

where

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

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

with variance

$$V_k = \text{var}(b_k - \hat{b}_k) = \hat{D} - \hat{D}\hat{Z}_k^T \Sigma_k^{-1} (\hat{\beta}, \hat{\xi}) \hat{Z}_k \hat{D} + \hat{D}\hat{Z}_k^T \hat{F}_k \Omega (\hat{\beta}) \hat{F}_k^T \hat{Z}_k \hat{D} \tag{20}$$

where

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

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

After estimating  $b_k$ , we may take a first-order expansion of  $f(t_k, A_k\beta_k + B_k b_k)$  about  $\beta = \hat{\beta}$  and  $b_k = \hat{b}_k$ , with a little algebra, the prediction for  $y_{kh}$  can be realized as:

$$\hat{y}_{kh} = f(t_{kh}, A_k\hat{\beta}_k + B_k\hat{b}_k) \tag{21}$$

with prediction variance:

$$\text{var}(y_{kh} - \hat{y}_{kh}) = \hat{F}_{kh}^T \hat{\Omega}(\hat{\beta}) \hat{F}_{kh} + \hat{z}_{kh}^T \hat{V}_k \hat{z}_{kh} + \hat{F}_{kh}^T \hat{C}_k \hat{z}_{kh} + \hat{z}_{kh}^T \hat{C}_k^T \hat{F}_{kh} + \hat{R}_{kh} \tag{22}$$

where

$$C_k = \text{Cov}[(\hat{\beta} - \beta), (\hat{b}_k - b_k)^T] \cong -\Omega(\hat{\beta}) \hat{F}_k^T (\hat{\beta}) \Sigma_k^{-1} (\hat{\beta}, \hat{\xi}) \hat{Z}_k \hat{D}$$

and  $\hat{F}_{kh}$ ,  $\hat{z}_{kh}$  are, respectively, the component of a column vector in matrix  $\hat{F}_k$ ,  $\hat{z}_k$ , corresponding to the  $h$ th occasion (on which prediction is required).

Similarly, the approximate  $100(1 - \alpha)\%$  can be obtained by (18).

For a specific example, suppose we know the dominant heights at age 11,14,17 to be 12 m,16 m, and 18 m for a plot

with the FCBH treatment. We want to predict the dominant height of this plot at age 20. Thus,

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

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

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

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

Hence, by (19),

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

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

The estimated variance-covariance matrix for the random effects is [by (20)]:

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

Therefore, with the random effects considered, the predicted dominant height at age 20 [by (21)] is:  $\hat{y}_{kh} = 20.45$  m, the estimated variance is [by (22)]:  $\text{var}(\hat{y}_{kh}) = 2.097$  m, thus the 95% confidence interval for the prediction is :[17.61 m, 23.68 m].

Comparing with Case 1, we can see that former observations for a specified individual can improve the precision of the

future prediction for that individual with the mixed effects model approach. In this example, the standard error of the prediction decreased by 26.5% from 1.969 m to 1.448 m. Generally, the more special an individual is (i.e., the more different from the “typical” response with  $b_i$ 's of zero), the more “gain” will result by using the mixed effects prediction.

It is worth pointing out that the correct within plot variance-covariance structure also plays an important role in prediction. For example, if we ignore the random effects and within-plot heterogeneity and correlation, and use the ordinary least square approach to the final chosen model (Model 10), the effects of fertilizer and herbicide will not be significant any more (with  $P$  value = 0.80 and 0.15 respectively). The prediction dominant height for the above example will be 18.72 m, with error 2.07 m, which is 5% higher than that obtained by appropriate variance-covariance structure (Case 1).

## Discussion and Conclusion

A mixed effects model is appealing for the analysis of repeated measurement data from a given experimental design because of its flexibility in representing the covariance structures that are not readily identified by the traditional regression approach. A modified Richards model with nonlinear mixed effects is recommended for modeling slash pine dominant height growth in conjunction with different silvicultural treatments. The analysis results show that a random effects model provided better model fitting and more precise predictions compared to a similar model without random effects.

Mixed effects reasonably explain the individual random variation that is quite common for repeated measurement data from permanent plots with an experimental design. If prior observations are available, the random effects corresponding to this specified individual can be estimated, thus more precise prediction for future observations can be obtained. This makes intuitive sense since more information is applied in the prediction by using the random effects approach. Random effects may relieve such problem as nonhomogeneous variance and within-plot autocorrelation error in modeling forest growth and yield repeated measurement data. But heterogeneity and correlation still occur for slash pine dominant height growth even after considering the appropriate random effects.

Heteroscedasticity is usually easier to handle. Once the trend of the variance (variance function) is identified, an appropriate individual weight (inverse of the variance) can be applied to individual observations to justify the process of parameter estimation. Comparatively, within-plot autocorrelation is not so easy to justify. Because many other factors in the model (or data) affect the autocorrelation, a reasonable structure may not be obvious for a given collection of data and model. No benefits will accrue from an inappropriate structure when autocorrelation is evident in the data. So, the model for autocorrelation structure should usually be as simple as possible when an appropriate one is not available (Davidian and Giltinan, p. 133, 1995). Moving average correlation with two parameters was used to model the within-plot correlation for the slash pine data. The estimated  $\theta_1$  is smaller than that of  $\theta_2$  (0.1853 vs. 0.3386). This

seems counterintuitive. One would expect observations spaced closer in time to exhibit higher correlation. Actually, this is also demonstrated on the empirical autocorrelation function in that negative correlation was estimated for the first three lags. Since the plot to plot variation is the main source of error (i.e., the random effects are much higher than the errors of within-plot error components.), any pattern of plot to plot variation may lead to unexpected within-plot patterns. Another explanation is that the effects of silvicultural treatment effects can be time-varied, and the autocorrelation pattern might reflect such variations. In such sense, a time-varied mixed effects model can be more desirable.

Incorrect covariance structure not only results in an incorrect model, but also produces a larger prediction error for a new observation prediction. For example, when we used the traditional ordinary least squares approach and ignored the random effects and within-plot heterogeneity, the prediction error increased 5% as compared to population level prediction with the covariance structure appropriately addressed.

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

I should note that the modified Richards model was chosen without comparing it to other growth models. As one anonymous reviewer pointed out, the Richards model does not fit slash or loblolly pine well at early ages because of initial rapid growth of these species. So it is helpful to emphasize that the purpose of this work is to demonstrate the useful method rather than to advocate the final product.

Three parameters in the modified Richards model are all chosen to be mixed (both fixed and random). Silvicultural treatments are significant in covariate modeling, but soil types are not. Likelihood ratio and *AIC* values are used in model performance evaluation. The procedures of model specification with the nonlinear mixed effects model is discussed in detail, and some useful graphical diagnosis methods are also presented. In summary, (1) between-plot variation is the main source of error for dominant height

growth prediction. (2) Within-plot heterogeneity and correlation still exist even after introducing random components into all three parameters in a modified Richards dominant height growth model. A power variance function and moving average correlation with order 2 [ $MV(2)$ ] are justified for the within-plot error structure in a conventional nonlinear mixed-effects model approach. (3) Predictions (including response value and prediction variance and thus confidence interval for the predictions) for new observations can be obtained at different precision levels in distinct situations. Past observations can be used to estimate the individual-specific parameters and thus improve the precision of the prediction for the specific individual experimental unit.

## Literature Cited

- AKAIKE, H. 1972. A new look at statistical model identification. P. 716–722 in *IEEE Trans. on Automatic Control*. AU-19.
- AVERY, T.E., AND H.E. BURKHART. 1994. *Forest measurements*. McGraw-Hill, New York.
- BAILEY, R.L., AND J.L. CLUTTER. 1974. Base-age invariant polymorphic site curves. *For. Sci.* 20:155–159.
- BIGING, G.S. 1985. Improved estimates of site index curves using a varying-parameter model. *For. Sci.* 31:248–259.
- BORDERS, B. E., R.L. BAILEY, AND J.D. WARE. 1984. Slash pine site-index from a polymorphic model by joining (splining) nonpolynomial segments with an algebraic difference method. *For. Sci.* 30:411–423.
- BOX, G.E.P., G.M. JENKINS, AND G.C. REINSEL. 1994. *Time series analysis: Forecasting and control*. Ed. 3. Holden-Day, San Francisco.
- CASTLEBERRY, J.P. 1998. Evaluation of models developed on a designed study data set applied to operational stands. M.S. thesis, The University of Georgia.
- CIESZEWSKI, C.J., AND I.E. BELLA. 1989. Polymorphic height and site index curves for lodgepole pine in Alberta. *Can. J. For. Res.* 19:1151–1160.
- CIESZEWSKI, C.J., AND R.L. BAILEY. 2000. Generalized algebraic difference approach: Theory based derivation of dynamic site equations with polymorphism and variable asymptotes. *For. Sci.* 46(1):116–126.
- CLUTTER, J.L. 1961. The development of compatible analytic models for growth and yield of Loblolly pine. Ph.D. dissertation, Duke University.
- CLUTTER J.L., J.C. FORTSON, L.V. PIENAAR, G.H. BRISTER, AND R.L. BAILEY. 1992. *Timber management: A quantitative approach*. Krieger, Malabar, Florida. 333 p.
- CURTIS, R.O., D.J. DEMARS, AND F.R. HERMAN. 1974. Which dependent variable in site index-height-age regressions? *For. Sci.* 20:74–87.
- DAVIDIAN, M., AND D.M. GILTINAN. 1995. *Nonlinear models for repeated measurement data*. Chapman and Hall, London.
- DIGGLE, P.J., K.Y. LIANG, AND S.L. ZEGER. 1994. *Analysis of longitudinal data*. Oxford University Press, Oxford.
- GARCIA, O. 1983. A stochastic differential equation for height growth of forest stands. *Biometrics* 39: 1059–1072.
- GREGOIRE, T.G. 1987. Generalized error structure for forestry yield models. *For. Sci.* 33:423–444.
- GREGOIRE, T.G., O. SCHABENBERGER, AND J.P. BARRETT. 1995. Linear modeling of irregularly spaced, unbalanced, longitudinal data from permanent-plot measurements. *Can. J. For. Res.* 25:137–156.
- GUMPERTZ, M.L., AND C. BROWNIE. 1993. Repeated measures in randomized block and split-plot experiments. *Can. J. For. Res.* 23:625–639.
- LAPPI, J., AND R.L. BAILEY. 1988. A height prediction model with random stand and tree parameters: An alternative to traditional site index methods. *For. Sci.* 38(2): 409–429.

- LAPPI, J., AND J. MALINEN. 1994. Random-parameter height/age models when stand parameters and stand age are correlated. *For. Sci.* 40: 715–731.
- LEAK, W. 1966. Analysis of multiple systematic remeasurements. *For. Sci.* 12:69–73.
- LENTNER, M., AND T. BISHOP. 1986. *Experimental design and analysis*. Valley Book Co., Blacksburg, VA.
- LINDSTROM, J.M., AND D.M. BATES. 1990. Nonlinear mixed effects models for repeated measures data. *Biometrics* 46:673–687.
- MAITRE, P.O., M. BUHRER, D. THOMSON, AND D.R. STANSKI. 1991. A three-step approach combining Bayesian regression and NONMEM population analysis: Application to midazolam. *J. Pharmacokinet. Biopharmaceut.* 19:377–384.
- MARTIN, S.W., R.L. BAILEY, AND E.J. JOKELA. 1999. Growth and yield predictions for lower coastal plain slash pine plantations fertilized at mid-rotation. *South. J. Appl. For.* 23(1):39–45.
- MCDILL, M.E., AND R.L. AMATEIS. 1992. Measuring forest site quality using the parameters of dimensionally compatible height growth function. *For. Sci.* 34(4):907–927.
- NORTHWAY, S.M. 1985. Fitting site index equations and other self-inferencing functions. *For. Sci.* 31:233–235.
- PIENARR, L.V., AND J.W. RHENEY. 1995. Modeling stand level growth and yield response to silvicultural treatments. *For. Sci.* 41(3):629–638.
- PIENARR, L.V., B.D. SHIVER, AND W.M. HARRISON. 1998. Slash pine site preparation study: Age 17 results. *Sch. For. Resour., PMRC Tech. Rep.* 1998-4. Univ. of Georgia, Athens.
- PINHEIRO, J.C., AND D.M. BATES. 1998. Model building for nonlinear mixed effects models. Tech. rep., Dep. of Stat., Univ. of Wisconsin.
- PINHEIRO, J.C., AND D.M. BATES. 2000. *Mixed effects models in S and S-plus*. Stat. and Comput. Series. Springer, New York.
- RICHARDS, F.J. 1959. A flexible growth function for empirical use. *J. Exp. Bot.* 10:290–300.
- ROE, D.J. 1997. Comparison of population pharmacokinetic modeling methods using simulated data: Results from the population modeling workgroup. *Stat. Med.* 16:1241–1262.
- SCHNUTE, J. 1981. A versatile growth model with statistically stable parameters. *Can. J. Fish. Aquat. Sci.* 38(9):1128–1140.
- SHIVER, B.D., L.V. PIENARR, K.L. HITCH, AND J.W. RHENEY. 1994. Slash pine site preparation study: Age 14 results. *Sch. For. Resour., PMRC Tech. Rep.* 1994-2. Univ. Georgia, Athens.
- SULLIVAN, A.D., AND J.L. CLUTTER. 1972. A simultaneous growth and yield model for loblolly pine. *For. Sci.* 18(1):76–86.
- 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.
- WOLFINGER, R.D., AND X. LIN. 1997. Two Taylor-series approximation method for the nonlinear mixed models. *Computational Stat. Data Analysis* 25:465–490.
- ZEGER, S.L., K. LIANG, AND P. ALBERT. 1988. Models for longitudinal data: a generalized estimating equation approach. *Biometrics* 44:1049–1060.