

---

# Modeling and Prediction of Forest Growth Variables Based on Multilevel Nonlinear Mixed Models

Daniel B. Hall and Robert L. Bailey

---

**ABSTRACT.** In this article, we describe estimation and prediction methods for nonlinear modeling of forest growth variables that are subject to nested sources of variability. The multilevel nonlinear mixed-effects models that we consider are useful for a variety of forestry applications, but we concentrate on the problem of estimating, and making projections from, growth curves for tree height based on longitudinal data grouped by location. Wolfinger and Lin consider estimating equation approaches to fitting more general nonlinear mixed-effects models, and we adapt their zero-expansion estimating equations to the multilevel case. We develop methods of prediction based on these models that allow predictions of future height both for individual trees and for plot averages. We illustrate these methods by fitting and making predictions from a Chapman-Richards type growth model for tree height data from a loblolly pine spacing study in Putnam County, Georgia. The mean and variance of prediction errors based on our methods are examined by means of cross-validation. We provide a more complete and unified presentation of linearization-based estimation and prediction based on multilevel nonlinear mixed-effects models than has previously appeared in the forestry literature, and we argue that these models lead to substantial advantages in growth and yield prediction over traditional forestry methods. *FOR. SCI.* 47(3):311–321.

**Key Words:** Estimating equations, random effects, repeated measures, site index, variance components.

---

**F**OR MANY PURPOSES IN FORESTRY, it is useful to be able to make accurate future predictions of the mean values of growth variables based on repeated measurements through time made on units that are grouped hierarchically. For example, many forest management decisions are based on yield projections that crucially depend on projections of plot level averages of tree height, basal area, and other morphometric variables. In the forestry literature, Lappi and Bailey (1988) have described the use of one example of a nonlinear mixed effects growth curve model to predict dominant tree height both at the plot level and at the individual tree level. In their model, random effects for plots, and for trees nested within plots enter into a Chapman-Richards-type growth model in a linear manner. Although not originally described in these terms, their model is an example of a *multi-level* nonlinear mixed effects model (ML-NLMM), because random effects are included in the model

corresponding to units of heterogeneity (plots and trees) that have a hierarchical structure with multiple levels.

Here we consider a more general version of Lappi and Bailey's model in which multilevel random effects are allowed to enter into the model nonlinearly. We describe methods of fitting such models based on first-order Taylor series linearization techniques that have been developed recently in the rapidly expanding statistical literature on nonlinear mixed-effects models (NLMMs), adapting these methods to the special challenges of the multilevel case. We also develop prediction methods based on these models that allow predictions to be made locally at each level in the model (e.g., to particular trees or plots of interest), as well as globally across the entire population represented in the data. When measurements are available on the units of interest (perhaps from ages prior to the target age), these predictors make use of this relevant information and thus can be ex-

---

Daniel B. Hall, Assistant Professor, University of Georgia Department of Statistics, Athens, Georgia 30602-1952—Phone: (706) 542-3302; Fax: (706) 542-3391; E-mail: dhall@stat.uga.edu. Robert L. Bailey, Professor, University of Georgia Daniel B. Warnell School of Forest Resources, Athens, Georgia 30602—Phone: (706) 542-1187; E-mail: bailey@smokey.forestry.uga.edu.

The authors sincerely thank Dr. Mike Clutter, Hargreaves Professor of Forest Finance at the Daniel B. Warnell School of Forest Resources, University of Georgia, for his encouragement, technical review, and strong support of this work.

pected to outperform some other methods of prediction that have appeared in the literature on NLMMs applied to forestry problems (e.g., Gregoire and Schabenberger 1996a, 1996b).

The methods we describe apply quite generally to nonlinear modeling of responses that are subject to nested sources of variability. However, we will concentrate on the problem of estimating, and making projections from, growth curves for tree height based on longitudinal data (repeated measures through time) that are grouped by location (typically by plots, but possibly by plots within stands, etc.). In the Example section below, we illustrate these methods by fitting Chapman-Richards growth curves to tree height data from an old-field loblolly pine (*Pinus taeda* L.) spacing study conducted in Putnam County, Georgia, by the D.B. Warnell School of Forest Resources, University of Georgia (Pienaar and Shiver 1993). Our goals are twofold: (1) to deliver the state-of-the-art in nonlinear mixed-effects modeling methodology to a forest biometrics audience in the context of an important forestry application of these models; and (2) to illustrate the advantages of this approach over traditional methods for growth and yield modeling based on site-index curves.

## Background

Traditionally, site index curves have been used to predict average height at an age of interest (age at harvest, for example). In this approach, one attempts to select an appropriate growth curve for the site of interest from a family of curves showing height as a function of age. This is done by matching the growth pattern of the site to a particular curve in the family by finding the curve that, at an inventory age, passes through the average (usually dominant) height observed on the site at that age. Thus, the curve from which predictions are made is fit to the data at hand based on the mean response at one point in time. As Garcia put it, "The site-index concept... depends on the assumption that, for a given species and region, variations in the height-growth pattern can be described by a one-parameter family of curves" (Garcia 1983, p. 1061).

The site-index approach is quite old, and through the years many methods of constructing site-index curves have been proposed. These methods are reviewed in Clutter et al. (1983, §2.4); see also Monserud (1984) and references given therein.

In the site-index approach to height prediction, the average height over age curve is localized to the site of interest by forcing it through the average height of the site at the index age. Beginning with Bailey and Clutter (1974), another approach emerged, in which localization of the height over age curve was accomplished by including both local (site-specific) and global parameters. This approach has evolved into mixed-effects growth models through the recognition that site-specific "parameters" are more appropriately thought of as random variables because of their correspondence with sites that can be thought of as randomly selected from, or otherwise representative of, a population of growing locations. Following Bailey and Clutter (1974), the mixed-effects modeling approach to

growth and yield has been developed by a number of authors including Garcia (1983); Biging (1985); Lappi and Bailey (1988); Tait, et al. (1988); Lappi (1991, 1997); Walters, et al. (1991); Bégin and Schütz (1994); Gregoire, et al. (1995); Candy (1997); Duplat and Tran-Ha (1997); and Lappi and Malinen (1997).

## Multilevel Linear Mixed Models

Linear mixed models have been reviewed in the forestry literature by Gregoire et al. (1995). The multilevel case has been covered by several authors in a general context (e.g., Longford 1993, Chap. 6, Goldstein 1995), but we review these models here for completeness and to lay the conceptual and notational groundwork for analogous nonlinear models. Here we will consider the three-level case. Extension to more levels is straightforward, although tedious notationally. Suppose we have multiple observations on trees within plots. We can think of the repeated observations as occurring on level one. These observations are nested within the level two units, the trees, which are nested within the level three units, the plots. Suppose that the  $k$ th measurement on the  $j$ th tree within the  $i$ th plot conforms to the linear mixed model

$$y_{ijk} = \mathbf{x}_{ijk}^T \boldsymbol{\beta} + \mathbf{z}_{2,ijk}^T \mathbf{b}_{2,ij} + \mathbf{z}_{3,ijk}^T \mathbf{b}_{3,i} + \varepsilon_{ijk}, \quad (1)$$

where  $\mathbf{x}_{ijk}$  is a  $p \times 1$  vector of covariates associated with the fixed effects  $\boldsymbol{\beta}$ ,  $\mathbf{z}_{2,ijk}$  is an  $r \times 1$  vector of covariates associated with random effects  $\mathbf{b}_{2,ij}$  operating at the tree level (level 2),  $\mathbf{z}_{3,ijk}$  is an  $s \times 1$  vector of covariates associated with random effects  $\mathbf{b}_{3,i}$  operating at the plot level (level 3), and  $\varepsilon_{ijk}$  is a random error term assumed to have mean 0 and variance  $\phi$ . The random effects,  $\mathbf{b}_{2,ij}$  and  $\mathbf{b}_{3,i}$ , are assumed to be independent of the error term and one another, each with mean 0 and variances  $\phi \Sigma_2$ ,  $\phi \Sigma_3$  at the tree and plot levels, respectively. Covariates  $\mathbf{z}_{m,ijk}$  with random effects at level  $m$  are usually taken to be a subset of the covariates in  $\mathbf{x}_{ijk}$ . These will often be indicator variables for the level  $m$  units. No such restrictions are necessary, though, in what follows. We will assume that there are  $n_3$  plots,  $n_{2,i}$  trees within the  $i$ th plot,  $n_{2,\cdot}$  trees overall,  $n_{1,ij}$  measurements on the  $j$ th tree within the  $i$ th plot,  $n_{1,i\cdot}$  measurements on all trees within the  $i$ th plot, and  $n \equiv n_{1\cdot\cdot}$  measurements overall.

To fix notation and to understand the structure of the data, consider the simple case where we have  $n_3 = 3$  plots,  $n_{2,i} = 2$  trees within each plot, and  $n_{1,ij} = 4$  measurements per tree. A simple model of the form given by (1) includes  $n_{2,\cdot} = 6$  tree-level, and  $n_3 = 3$  plot-level random effects as follows:

$$\begin{aligned} y_{ijk} = & \mathbf{x}_{ijk}^T \boldsymbol{\beta} + I\{i = 1, j = 1\}b_{2,11} + I\{i = 1, j = 2\}b_{2,12} + \dots \\ & + I\{i = 3, j = 2\}b_{2,32} + I\{i = 1\}b_{3,1} + I\{i = 2\}b_{3,2} \\ & + I\{i = 3\}b_{3,3} + \varepsilon_{ijk}, \end{aligned}$$

where  $\text{var}(b_{2,ij}) = \phi \Sigma_2$  for all  $i, j$ ,  $\text{var}(b_{3,ij}) = \phi \Sigma_3$  for all  $i$ , and  $I\{C\}$  is the indicator variable taking the value 1 when condition  $C$  is true, and 0 otherwise.

Under the assumption of independence across the highest level units (plots), it is useful to rewrite this model in matrix notation in terms of the vector-valued response

$$\mathbf{y}_i = (y_{i11}, y_{i12}, \dots, y_{in_2, n_{1,ij}})^T$$

for the  $i$ th plot. In general, this model representation is

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_{2,i} \mathbf{b}_{2,i} + \mathbf{Z}_{3,i} \mathbf{b}_{3,i} + \boldsymbol{\varepsilon}_i, \quad (2)$$

where  $\mathbf{X}_i$  is  $n_{1,i} \times p$  with rows formed from the  $x_{ijk}^T$ ;  $\mathbf{Z}_{2,i}$  is a  $n_{1,i} \times m_{2,i}$  matrix with columns formed from the coefficients on the tree-level random effects (the  $b_{2,ij}$ 's); and  $\mathbf{Z}_{3,i}$  is a  $n_{1,i} \times s$  matrix with columns formed from the coefficients on the plot-level random effects (the  $b_{3,ij}$ 's). In the example given above, we have only random intercepts at the tree and plot levels, so  $r = s = 1$  and we have  $\mathbf{b}_{2,i} = (b_{2,i1}, b_{2,i2})^T$ ;  $\mathbf{b}_{3,i} = b_{3,i}$ , a scalar;  $\mathbf{Z}_{2,i}$  is  $8 \times 2$  with columns formed from the indicator variables  $I\{i = i, j = 1\}$ ,  $I\{i = i, j = 2\}$  above; and  $\mathbf{Z}_{3,i}$  is an  $8 \times 1$  vector of ones corresponding to  $I\{i = i\}$ , the coefficient on  $b_{3,i}$ . Here,  $\text{var}(\mathbf{b}_{2,i}) = \phi \boldsymbol{\Sigma}_2 \mathbf{I}_2$  and  $\text{var}(b_{3,i}) = \phi \boldsymbol{\Sigma}_3$ , where  $\mathbf{I}_a$  represents the  $a \times a$  identity matrix. A more succinct representation for (2) is

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i, \quad (3)$$

where

$$\mathbf{Z}_i = (\mathbf{Z}_{2,i}, \mathbf{Z}_{3,i}) \text{ and } \mathbf{b}_i = (\mathbf{b}_{2,i}^T, \mathbf{b}_{3,i}^T)^T.$$

Based on (2) we see that

$$\text{var}(\mathbf{y}_i) = \phi (\boldsymbol{\Sigma}_2 \mathbf{Z}_{2,i} \mathbf{Z}_{2,i}^T + \boldsymbol{\Sigma}_3 \mathbf{Z}_{3,i} \mathbf{Z}_{3,i}^T + \mathbf{I}_{n_{1,i}})$$

for this case in which  $b_{2,ij}, b_{3,i}$  are scalar, for all  $i, j$ . In general,

$$\text{var}(\mathbf{y}_i) = \phi \mathbf{Z}_i \boldsymbol{\Sigma}_i \mathbf{Z}_i^T + \phi \mathbf{I}_{n_{1,i}},$$

based on (3), where  $\phi \boldsymbol{\Sigma}_i = \text{var}(\mathbf{b}_i)$  is the block-diagonal matrix with diagonal elements  $\boldsymbol{\Sigma}_2$  repeated  $n_{2,i}$  times down the diagonal,  $\boldsymbol{\Sigma}_3$  forming the bottom-right block, and zeros elsewhere.

The random effects' variance-covariance matrices are assumed to be known up to a  $q \times 1$  parameter  $\boldsymbol{\theta}$ , so we will often write  $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_i(\boldsymbol{\theta})$ , to emphasize this parameterization. Often, both  $\mathbf{b}_{2,ij}$  and  $\mathbf{b}_{3,i}$  are assumed to have uncorrelated components with distinct variances, so that  $q = r + s$  and  $\boldsymbol{\theta}$  contains the diagonal elements of  $\boldsymbol{\Sigma}_2$  and  $\boldsymbol{\Sigma}_3$ . This case is commonly referred to as a *variance components model* (e.g., Searle et al. 1992, §4.6). Another important case is when the elements of  $\mathbf{b}_{2,ij}$  and  $\mathbf{b}_{3,i}$  are allowed to covary in a general way so that  $\boldsymbol{\Sigma}_2$  and  $\boldsymbol{\Sigma}_3$  are completely unstructured. This case can be parameterized in a number of ways (Pinheiro and Bates 1996), but a simple choice that we adopt here is for  $\boldsymbol{\theta}$  to consist of the  $q = r(r+1)/2 + s(s+1)/2$  elements in the upper triangles of  $\boldsymbol{\Sigma}_2$  and  $\boldsymbol{\Sigma}_3$ .

## Multilevel Nonlinear Mixed Models

Growth curves for trees are typically sigmoidal, so that linear models such as (3) are seldom adequate. Instead,

nonlinear growth models are commonly used in forestry. The ML-NLMM considered here has the general form

$$\mathbf{y}_i = \mathbf{f}(\boldsymbol{\beta}, \mathbf{b}_i, \mathbf{X}_i; \mathbf{Z}_i) + \boldsymbol{\varepsilon}_i, \quad i = 1, \dots, n_3, \quad (4)$$

where  $\mathbf{f}$  is a nonlinear function with arguments  $\boldsymbol{\beta}, \mathbf{b}_i, \mathbf{X}_i$ , and  $\mathbf{Z}_i$  defined as in the previous section, and  $\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_{n_3}$  are independent error terms each distributed as  $N(\mathbf{0}, \phi \mathbf{I})$ .

When computationally feasible, an attractive approach to fitting model (4) is maximum likelihood. According to model (4), conditional on the random effects vector  $\mathbf{b}_i$ ,  $\mathbf{y}_i | \mathbf{b}_i \sim N\{\mathbf{f}(\boldsymbol{\beta}, \mathbf{b}_i, \mathbf{X}_i, \mathbf{Z}_i), \phi \mathbf{I}\}$ , so that the (marginal) likelihood for  $\mathbf{y}_i$  can be obtained by integrating a normal density with respect to the distribution of the random effects. Maximization of the resulting likelihood function is complicated by the presence of a multidimensional integral in this function. For simple random effects structures, the necessary integration can be done by Gaussian quadrature, or some other numerical technique without too much difficulty. In particular, for the two-level case (e.g., data from several trees within a single plot) SAS's PROC NL MIXED (Wolfinger 1999) can perform maximum likelihood estimation for models with a small number of correlated random effects that are specific to the level two units (e.g., tree-specific random effects). However, for more than two or three random effects, or when random effects are crossed or nested (as in the three-level and higher ML-NLMM), the computational burden of evaluating the integral numerically becomes too great.

Various authors have proposed simulation-based methods to approximate the MLEs in generalized linear mixed-effects models (GLMMs) with complex random effects structures. Examples of such methods include various Monte Carlo versions of the EM algorithm (e.g., McCulloch 1992, 1997, Booth and Hobert 1999), Monte Carlo Newton-Raphson (McCulloch 1997), simulated maximum likelihood (e.g., Geyer and Thompson 1992, Gelfand and Karlin 1993), and various importance sampling algorithms (e.g., Zhang 1996, Kuk 1999). In principle, these methods can be adapted to the ML-NLMM context, but this remains to be done. Furthermore, these methods are typically difficult to implement and too expensive computationally to be practical tools for routine use.

There are several closely related alternatives to full ML estimation in the NLMM that are all based on first order Taylor linearization of the model. Reviews of the extensive literature on linearization methods are available in several sources (Davidian and Giltinan 1995, Chap. 6, Schabenberger and Gregoire 1996, Vonesh and Chinchilli 1997, Chap. 7–9, Wolfinger and Lin 1997), so we do not repeat that material here except as necessary to place our approach in proper context. The main distinction is in the point of expansion for the Taylor series linearization that is an element of all of these methods. Expansion is done either around the expected value of the vector of random effects (which is 0), or around some estimate of the random effects vector, usually the so-called *empirical best linear unbiased predictor* (EBLUP, see Robinson 1991).

The approach we adopt here is to adapt the zero-expansion Taylor series method of Wolfinger and Lin (1997) to produce estimating equations appropriate to the ML-NLMM. We then use those estimating equations to fit, and produce predictions from, a specific example of the ML-NLMM based on a Chapman-Richards growth model. The zero-expansion method that we use is termed *quasi-extended least squares* by Vonesh and Chinchilli (1997, §9.2.5), and is equivalent to the extended generalized estimating equation (GEE) methodology of Hall and Severini (1998) and Hall (2001) with first-order approximations used in place of the true marginal mean and variance-covariance matrix of  $y_i$ . The GEE connection was first explored by Zeger, et al. (1988) in a GLMM context, but has since been discussed by several authors in both the GLMM and NLMM settings (e.g., Schabenberger 1994, Schabenberger and Gregoire 1996). Further connections with the existing literature are made in Davidian and Giltinan (1995, §6.2), Wolfinger and Lin (1997), and Vonesh and Chinchilli (1997, §7.4.2.i and §8.2, and Chap. 9).

### Estimation

Based on a first-order Taylor expansion around the random effects' mean,  $\mathbf{b}_i = \mathbf{0}$ , model (4) can be linearized through the approximation

$$\begin{aligned} \mathbf{y}_i &\doteq \mathbf{f}(\boldsymbol{\beta}, \mathbf{0}, \mathbf{X}_i, \mathbf{Z}_i) + \tilde{\mathbf{Z}}_{2,i} \mathbf{b}_{2,i} + \tilde{\mathbf{Z}}_{3,i} \mathbf{b}_{3,i} + \boldsymbol{\varepsilon}_i \\ &= \mathbf{f}(\boldsymbol{\beta}, \mathbf{0}, \mathbf{X}_i, \mathbf{Z}_i) + \tilde{\mathbf{Z}}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i \end{aligned} \quad (5)$$

where  $\tilde{\mathbf{Z}}_i = (\tilde{\mathbf{Z}}_{2,i}, \tilde{\mathbf{Z}}_{3,i})$  and

$$\tilde{\mathbf{Z}}_{m,i} = \left. \frac{\partial \mathbf{f}(\boldsymbol{\beta}, \mathbf{b}_i, \mathbf{X}_i, \mathbf{Z}_i)}{\partial \mathbf{b}_{m,i}^T} \right|_{\mathbf{b}_i = \mathbf{0}}, \quad m = 2, 3$$

Based on (5) we have

$$\begin{aligned} E(\mathbf{y}_i) &\doteq \mathbf{f}(\boldsymbol{\beta}, \mathbf{0}, \mathbf{X}_i, \mathbf{Z}_i) \\ \text{var}(\mathbf{y}_i) &\doteq \phi \left\{ \mathbf{I}_{n_i} + \tilde{\mathbf{Z}}_i \Sigma_i(\boldsymbol{\theta}) \tilde{\mathbf{Z}}_i^T \right\} \equiv \mathbf{V}_i(\boldsymbol{\delta}) \end{aligned} \quad (6)$$

where  $\boldsymbol{\delta} = (\boldsymbol{\theta}^T, \phi)^T$ . Utilizing these approximations in place of the true marginal moments, the extended GEEs are

$$\sum_{i=1}^{n_3} \tilde{\mathbf{X}}_i^T \mathbf{V}_i^{-1}(\boldsymbol{\delta}) (\mathbf{y}_i - \mathbf{f}(\boldsymbol{\beta}, \mathbf{0}, \mathbf{X}_i, \mathbf{Z}_i)) = \mathbf{0}, \quad (7)$$

where

$$\tilde{\mathbf{X}}_i = \left. \frac{\partial \mathbf{f}(\boldsymbol{\beta}, \mathbf{b}_i, \mathbf{X}_i, \mathbf{Z}_i)}{\partial \boldsymbol{\beta}^T} \right|_{\mathbf{b}_i = \mathbf{0}}$$

and

$$\begin{aligned} &\frac{1}{2} \sum_{i=1}^{n_3} \left[ \{ \mathbf{y}_i - \mathbf{f}(\hat{\boldsymbol{\beta}}, \mathbf{0}, \mathbf{X}_i, \mathbf{Z}_i) \}^T \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \boldsymbol{\delta}_j} \mathbf{V}_i^{-1} \{ \mathbf{y}_i - \mathbf{f}(\hat{\boldsymbol{\beta}}, \mathbf{0}, \mathbf{X}_i, \mathbf{Z}_i) \} \right. \\ &\left. - \text{tr} \left( \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \boldsymbol{\delta}_j} \right) \right] = 0, \quad j = 1, \dots, q+1, \end{aligned} \quad (8)$$

for  $\boldsymbol{\beta}$  and  $\boldsymbol{\delta}$ , respectively (cf. Wolfinger and Lin 1997, equations 8 and 15).

In linear mixed-effects models, an often-preferred alternative to maximum likelihood estimation is the method of restricted maximum likelihood (REML) estimation (see Harville 1977, and references therein). The main advantage to REML estimation is that, unlike ML, it produces the standard analysis of variance-based quadratic unbiased estimators of variance components in the balanced data cases of the linear mixed-effects model for which such estimators are available. That is, the REML approach produces a bias correction or degree of freedom adjustment that takes into account information lost in having to estimate the elements of  $\boldsymbol{\beta}$ , the fixed effects parameters in the model. For example, the REML estimator of the error variance in the simple balanced one-way ANOVA model is the usual one,  $SS_E / (n - p)$  where  $SS_E$  is the within-group, or error, sum of squares, and  $n$  and  $p$  are the total sample size and dimension of  $\boldsymbol{\beta}$ , respectively. In contrast, the ML estimator is  $SS_E / n$ . In the nonlinear context of this paper, a REML-type alternative to (8) is available by replacing  $\mathbf{V}_i^{-1}$  in the trace term with

$$\mathbf{P}_i = \mathbf{V}_i^{-1} - \mathbf{V}_i^{-1} \tilde{\mathbf{X}}_i \left( \sum_{i=1}^{n_3} \tilde{\mathbf{X}}_i^T \mathbf{V}_i^{-1} \tilde{\mathbf{X}}_i \right)^{-1} \tilde{\mathbf{X}}_i^T \mathbf{V}_i^{-1}.$$

Estimation proceeds by iteratively solving (7) and (8), with Equation (8) modified as described above for the REML-type version of the procedure. Consistent estimators of the asymptotic variance-covariance matrix of the resulting parameter estimators are given in Hall (2001). These expressions are based on the sandwich variance estimator of Royall (1986) and have the advantage of being robust to misspecification of the second and higher moment structure in the estimating equations. Because the marginal variance is approximated by (6) and to guard against possible misspecification of the random effects structure in the model, such robust variance estimators are preferable to the model-based estimators that are output by commonly used NLMM software such as nlme (Pinheiro and Bates 2000) and the SAS macro NLINMIX (Littell, et al. 1996).

At convergence, a predictor for the random effects  $\mathbf{b}_i$ ,  $i = 1, \dots, n_3$ , can be obtained from the EBLUP of the linearized model. This EBLUP predictor is given by

$$\hat{\mathbf{b}}_i = \hat{\Sigma}_i \hat{\mathbf{Z}}_i^T \hat{\mathbf{V}}_i^{-1} (\mathbf{y}_i - \mathbf{f}(\hat{\boldsymbol{\beta}}, \mathbf{0}, \mathbf{X}_i, \mathbf{Z}_i)), \quad (9)$$

where the hats indicate that the corresponding quantities are evaluated at the final parameter estimates. The predictor (9) has links to several other areas of statistical methodology, including empirical Bayes methods, Kalman filtering, and Kriging (see Robinson 1991 for details).

The special case of the multilevel version of the NLMM results in a clustered data structure in which the vectors of observations at the highest level (plot) each have correlated elements, but are independent from one plot to another. This allows us to write the estimating functions as sums of independent terms in contrast to the general formulation of

Wolfinger and Lin (1997). An additional special feature of the multilevel case is that, although solution of the estimating equations requires us to invert  $\mathbf{V}_i$  for each  $i$ , this operation is computationally less demanding than inverting a  $n_{1,i} \times n_{1,i}$  variance-covariance matrix of general form. The nested structure of the random effects gives  $\mathbf{V}_i$  a particular structure that can be taken advantage of for computational savings in its inversion. This and other opportunities for computational shortcuts carry over unchanged from the multilevel linear mixed model case. For extensive details see Longford (1993, Chap. 6).

### Prediction

For purposes of forest management, yield predictions are important. Prediction of the average height of a plot, the height of a particular tree, etc., can be based on nonlinear prediction theory (Valliant 1985). In this section we present the prediction technique in its general form. Prediction of a plot average height at a given age, or an individual tree height, occurs as a special case. Throughout, we assume that predictions are desired for future observations that are not necessarily independent of those already observed. That is, our predictor applies for units (e.g., trees or plots) on which past measurements are available or for which measurements are available on different same-level units within a common higher-level unit (e.g., we want to predict the height of tree at age 20 in the presence of height data on other trees at age 20 from the same plot). In such cases, the correlation between the observation to be predicted and the observations available provides an opportunity to improve upon the predictor advocated by Gregoire and Schabenberger (1996a, 1996b) in an NLMM context. For an approach to prediction similar to ours, see Vonesh and Chinchilli (1997, §7.4.5).

Let

$$\mathbf{y} = (\mathbf{y}_s^T, \mathbf{y}_r^T)^T$$

be the combined vector of observed sample units ( $s$  for sample) and unobserved nonsample units (indicated by the  $r$  subscript). Let

$$\mathbf{X} = (\mathbf{X}_s^T, \mathbf{X}_r^T)^T, \quad \mathbf{Z} = (\mathbf{Z}_s^T, \mathbf{Z}_r^T)^T$$

be matrices of fixed effects covariates and random effects covariates, respectively, partitioned as is  $\mathbf{y}$  into portions corresponding to sample and nonsample measurements. Let  $\mathbf{W}$  denote the covariance matrix for the combined sample and nonsample response vector with the appropriate partitioning:

$$\mathbf{W} = \begin{pmatrix} \mathbf{W}_{ss} & \mathbf{W}_{sr} \\ \mathbf{W}_{rs} & \mathbf{W}_{rr} \end{pmatrix}$$

For the results of this section,  $\mathbf{W}$  is initially assumed to be known.

Let

$$\alpha_0 = (\beta_0^T, \mathbf{b}_0^T)^T$$

denote the combined vector of the true fixed effects  $\beta_0$  and true random effects  $\mathbf{b}_0$ . By “true” random effects we mean the

realized values of the random vector  $\mathbf{b}$  for the sampled trees and plots. According to a first order linearization about  $\mathbf{b}_0$  similar to (5),

$$E(\mathbf{y}) \doteq \mathbf{f}(\alpha_0, \mathbf{X}, \mathbf{Z}) = \{\mathbf{f}_s(\alpha_0, \mathbf{X}_s, \mathbf{Z}_s)^T, \mathbf{f}_r(\alpha_0, \mathbf{X}_r, \mathbf{Z}_r)^T\}^T.$$

The best (minimum variance) linear unbiased predictor (BLUP) of  $\mathbf{y}_r$  based on the linearized model is

$$\hat{\mathbf{y}}_{r0} = \mathbf{f}_r(\alpha_0, \mathbf{X}_r, \mathbf{Z}_r) + \mathbf{W}_{rs} \mathbf{W}_{ss}^{-1} \{\mathbf{y}_s - \mathbf{f}_s(\alpha_0, \mathbf{X}_s, \mathbf{Z}_s)\}. \quad (10)$$

Because  $\alpha_0$  is unknown, this predictor is unavailable. Instead, following Valliant (1985), we will consider the “empirical BLUP”  $\hat{\mathbf{y}}_r$  obtained by substituting  $\hat{\alpha}$  for  $\alpha_0$  in (10), leading to

$$\hat{\mathbf{y}}_r = \mathbf{f}_r(\hat{\alpha}, \mathbf{X}_r, \mathbf{Z}_r) + \mathbf{W}_{rs} \mathbf{W}_{ss}^{-1} \{\mathbf{y}_s - \mathbf{f}_s(\hat{\alpha}, \mathbf{X}_s, \mathbf{Z}_s)\}, \quad (11)$$

where  $\hat{\alpha} = (\hat{\beta}^T, \hat{\mathbf{b}}^T)^T$  is the zero-expansion solution described above.

As described in Valliant (1985),  $\hat{\mathbf{y}}_r$  is asymptotically normally distributed, so approximate inference can be based on the normal distribution. Standard errors for prediction can be obtained from the sample estimator of the approximate prediction variance of  $\hat{\mathbf{y}}_r$ . The prediction variance can be approximated by making use of several first-order approximations. The derivation of, and expressions for, this variance are mathematically tedious, so they are omitted here. These and other details of the methodology of this article can be found in Hall and Bailey (2000). Although the predictor (11), and its approximate variance, are derived assuming a known covariance matrix  $\mathbf{W}$ , in practice such knowledge is not available. Instead, the sample version  $\hat{W}(\hat{\alpha})$  must be used in place of  $\mathbf{W}(\alpha)$  and estimators  $\hat{\phi}$  and  $\hat{\Sigma}$  used to obtain predictions and standard errors for prediction intervals.

### Example

In the southeastern United States, loblolly pine forms the largest component of the timber resource, much of it consisting of plantations grown under intensive management practices. To date, there has been limited study of height growth in this population. One relevant study is the D.B. Warnell School of Forest Resources’ Old-field Loblolly Pine Spacing Study currently ongoing in the B.F. Grant Memorial Forest in Putnam County, Georgia, in the Piedmont region of the state. In this study, herbaceous competition was controlled during the first three growing seasons using mowing and herbicide application. The 24 0.081 ha plots on the study site were randomized to six stocking densities of 247, 494, 988, 1483, 1977, and 2471 trees/ha in a balanced (four plots per density), completely randomized design. Currently, repeated measurements of the heights of all trees in 0.040 ha, interior plots within each of the 24 original plots are available for ages 3, 5, 8, 10, 12, and 14 yr. Further details of the study and results following 8 yr of growth are reported by Pienaar and Shiver (1993).

As pointed out by these authors, “height growth curves that would be implied by existing site index equations for old-field loblolly plantations in the Piedmont are not likely to be

appropriate for plantations such as these with accelerated early height growth” (Pienaar and Shiver 1993, p. 193–194). Instead there is a need for growth curves developed from data like those from the B.F. Grant Study, which come from the population of interest. In addition, it is highly desirable that the fitted height growth over age relationship generates predictions of future height that are as precise as possible and whose precision can be quantified. These considerations motivate the development of a ML-NLMM based on the B.F. Grant Study data.

Another reason that the B.F. Grant Study data were chosen to illustrate the methods of this article is that data on all trees in each measurement plot are available. In typical applications, only a subset of trees in each plot will be remeasured; so for realism we develop our model below based on a subset of the data consisting of a random sample of 30% of the trees in each plot (365 sampled trees, in total). However, access to the complete data on all trees will allow us to compare plot-average predictions generated from the subset with corresponding true values.

For  $y_{ijk}$ , the height of the  $j$ th tree in the  $i$ th plot at the  $k$ th measurement occasion, we fit a model of the form

$$y_{ijk} = \gamma_{1ij} \{1 - \exp(-\gamma_{2i} \text{age}_{ijk})\}^{\gamma_{3ij}} + \varepsilon_{ijk} \quad (12)$$

where

$$\begin{aligned} \gamma_{1ij} &= \beta_1 + \beta_2 \text{tph}_i + b_{2,ij}^{(1)} + b_{3,i}^{(1)} \\ \gamma_{2,i} &= \beta_3 + \beta_4 \text{tph}_i \end{aligned}$$

and

$$\gamma_{3ij} = \beta_5 + \beta_6 \text{tph}_i + b_{2,ij}^{(2)} + b_{3,i}^{(2)}.$$

In this model, all three parameters of a Chapman-Richards type growth model are assumed to depend linearly on the covariate,  $\text{tph}_i$  = trees per hectare/100 in plot  $i$ . In addition, both the asymptote parameter,  $\gamma_1$ , and the shape parameter,  $\gamma_3$ , are allowed to vary from plot to plot and from tree to tree through dependencies on random tree effects,  $b_{2,ij}^{(1)}$  and  $b_{2,ij}^{(2)}$ , and random plot effects,  $b_{3,i}^{(1)}$  and  $b_{3,i}^{(2)}$ . In addition to normal errors, we assume bivariate normal distributions for both the tree-specific and plot-specific random effects as follows:

$$\begin{pmatrix} b_{2,ij}^{(1)} \\ b_{2,ij}^{(2)} \end{pmatrix} \stackrel{iid}{\sim} N_2(\mathbf{0}, \phi \Sigma_2), \quad \begin{pmatrix} b_{3,i}^{(1)} \\ b_{3,i}^{(2)} \end{pmatrix} \stackrel{iid}{\sim} N_2(\mathbf{0}, \phi \Sigma_3), \quad \varepsilon_{ijk} \stackrel{iid}{\sim} N(0, \phi).$$

The specification of the Chapman-Richards parameters  $\gamma_1$ ,  $\gamma_2$ ,  $\gamma_3$  in model (12) deserves some discussion. Conceivably, each parameter could be allowed to depend on available covariates in a complex manner, and all parameters could be allowed to vary across both plots and trees through dependencies on both plot-specific and tree-specific random effects. However, there are both statistical and practical reasons to limit the complexity of the model. From a statistical standpoint, the inclusion of complex parametric functions of covariates and nested random effects in each of  $\gamma_1$ ,  $\gamma_2$ , and  $\gamma_3$  will typically lead to nonidentifiability and ill-conditioning.

All of the complications associated with these problems extend from linear and nonlinear fixed effects models (see Seber and Wild 1989, §3.4) to NLMMs, but generally speaking, they become more severe. Fortunately, nonidentifiability and ill-conditioning can usually be detected by stubborn nonconvergence of the estimation procedure or convergence to a solution that has an estimated variance-covariance matrix that is nearly singular and/or includes extremely large or small variances or near perfect correlation. A simple, practical approach is to avoid models that yield such results by simplifying dependence structures for covariates and/or removing random effects from one or more of the model parameters. From a practical standpoint, this is often desirable so that the model is not too complex to be understood.

In the current investigation, we chose linear functions of  $\text{tph}$  for each  $\gamma$  parameter based on an examination of the fitted coefficients from a fixed effects, no covariates version of (12) fit separately to the data from each  $\text{tph}$  value. Plots of nonlinear least squares (NLLS) estimates of  $\hat{\gamma}_1$ ,  $\hat{\gamma}_2$  and  $\hat{\gamma}_3$  against  $\text{tph}$  suggested decreasing linear, increasing linear, and increasing linear trends, respectively. Simple linear regression fits to these plots were also helpful in suggesting starting values for the parameters  $\beta_1, \dots, \beta_6$  in model (12). Initially, we attempted to fit a version of (12) in which random plot-effects and random tree-effects entered into each  $\gamma$  parameter. Because we could not obtain convergence with such a model and because simpler models with a single random effect were consistently most difficult to fit when the random effect appeared in  $\gamma_2$ , we simplified the initial model by eliminating random effects in  $\gamma_2$ . Starting values for the variance component parameters of this model,  $\theta$  and  $\phi$ , were obtained in several steps. First, the basic fixed effects model given in expression (12) with  $\gamma_1, \gamma_2, \gamma_3$  all specified as fixed-effects parameters not depending on covariates, was fit separately for each plot in the data set using NLLS. The plot-specific parameter estimates were then regressed on the covariate  $\text{tph}$ , and the residual mean squares from these regressions were taken as starting values for plot-specific variance components (for the diagonal elements of  $\phi \Sigma_3$ ). These starting values were further refined by utilizing SAS' PROC NL MIXED to fit a simplification of model (12) that included  $b_{3,i}^{(1)}$  and  $b_{3,i}^{(2)}$  but not  $b_{2,ij}^{(1)}$  and  $b_{2,ij}^{(2)}$ . The same process was repeated at the tree level. The basic fixed-effects, no covariate version of (12) was fitted separately by tree, parameter estimates were regressed on  $\text{tph}$  to obtain starting values for the diagonal elements of  $\phi \Sigma_2$ , and finally these starting values were refined by fitting simplified two-level NLMMs with PROC NL MIXED.

After selecting a model and obtaining starting values as described above, we utilized the REML version of the zero-expansion estimation method to obtain the following fitted model:

$$\hat{y}_{ijk} = \hat{\gamma}_{1ij} \{1 - \exp(-\hat{\gamma}_{2i} \text{age}_{ijk})\}^{\hat{\gamma}_{3ij}},$$

where

$$\begin{aligned} \hat{\gamma}_{1ij} &= 42.1 - 0.757 \text{tph}_i + \hat{b}_{2,ij}^{(1)} + \hat{b}_{3,i}^{(1)} \\ \hat{\gamma}_{2i} &= 0.0385 + 0.00219 \text{tph}_i \end{aligned}$$

**Table 1. Parameter estimates and robust asymptotic standard errors (ASEs) for model (12).**

| Parameter | Estimate | ASE      | Parameter  | Estimate | ASE     |
|-----------|----------|----------|------------|----------|---------|
| $\beta_1$ | 42.1     | 1.70     | $\theta_1$ | 36.0     | 6.25    |
| $\beta_2$ | -0.757   | 0.0996   | $\theta_2$ | 0.579    | 0.126   |
| $\beta_3$ | 0.0385   | 0.00432  | $\theta_3$ | 0.0290   | 0.00536 |
| $\beta_4$ | 0.00219  | 0.000399 | $\theta_4$ | 5.16     | 1.51    |
| $\beta_5$ | 1.22     | 0.0475   | $\theta_5$ | -0.0494  | 0.0527  |
| $\beta_6$ | 0.00696  | 0.00350  | $\theta_6$ | 0.00811  | 0.00215 |
|           |          |          | $\phi$     | 0.229    | 0.0124  |

and

$$\hat{\gamma}_{3ij} = 1.22 + 0.00696\beta_6 h_i + \hat{b}_{2,ij}^{(2)} + \hat{b}_{3,i}^{(2)}.$$

Estimates of variance components corresponding to trees, plots and residual error are

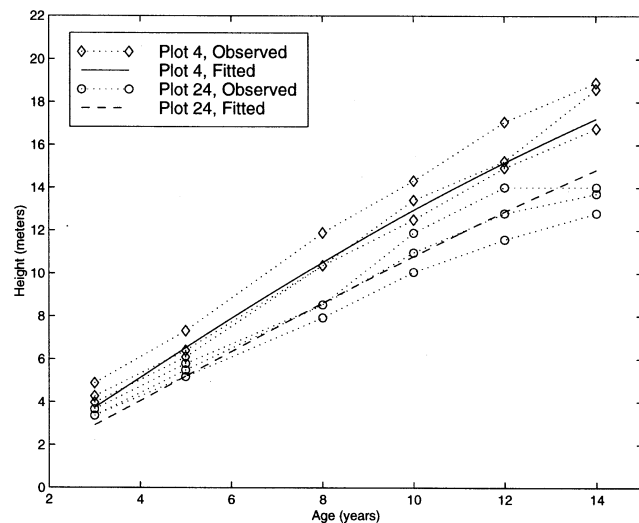
$$\hat{\Sigma}_2 = \begin{pmatrix} 36.0 & 0.579 \\ 0.579 & 0.00290 \end{pmatrix},$$

$$\hat{\Sigma}_3 = \begin{pmatrix} 5.16 & -0.0494 \\ -0.0494 & 0.00811 \end{pmatrix},$$

$$\hat{\phi} = 0.229.$$

These parameter estimates and their robust asymptotic standard errors are summarized in Table 1.

To illustrate the flexibility of this model, in Figures 1 and 2 we have plotted observed and fitted growth curves for several trees from two of the plots in the data set. In Figure 1 we plot observed growth curves (dotted lines) for three trees in plot 4 (988 trees/ha), and three trees in plot 24 (247 trees/ha). In addition, we plot the fitted plot average growth curve (solid lines) for the two plots based on the predicted plot effects for plots 4 and 24, and tree effects set equal to their average, zero. In Figure 2, tree-specific fitted growth curves are displayed for the same trees as in Figure 1. These tree-

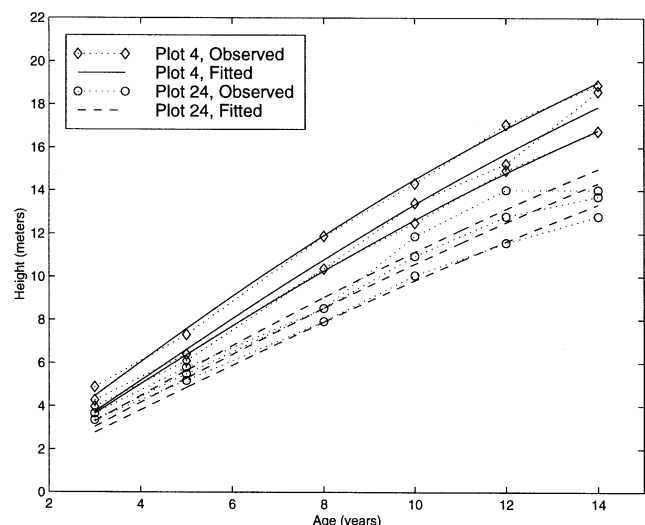


**Figure 1. Observed growth curves from 3 trees from plot 4 (988 trees/ha) and 3 trees from plot 24 (247 trees/ha), plus fitted plot average curves.**

specific curves are generated based on the predicted plot-level and tree-level random effects corresponding to the six selected trees from plots 4 and 24. Figures 1 and 2 are meant to emphasize the flexibility of the nonlinear multilevel mixed model; so to convey this point and to keep the plots uncluttered, we have plotted observed and fitted curves only for selected trees and plots with well-differentiated growth curves. While capturing the overall population growth trend, the model also allows the estimated growth pattern to be individualized to each level of heterogeneity contained in the data structure, that is, to each plot (Figure 1) and each tree (Figure 2).

To investigate the properties of the predictor proposed above, a small cross-validation study was conducted. Based on the Chapman-Richards model described above, we predicted both individual tree height at age 14 and plot average tree height at age 14. The accuracy of these predictions was assessed via cross-validation. That is, for tree-specific height predictions, we fit model (12) 365 times, once for each tree in the data set. For each fit of the model, one observation, the height at age 14 for one tree, was omitted from the data, and a prediction made for that observation. This was done for each tree so that a set of 365 predictions were obtained, one for each tree, where none of the predictions were made based on data that included the observation to be predicted. Similarly, for plot average height predictions, for each plot in turn, we omitted all observations at age 14 and predicted that plot's average height. Twenty-four predicted plot average heights were obtained in this way and compared with the true plot-average height computed from the complete data set.

Cross-validation results appear in Table 2. For comparison purposes, we report also the results based on methods of Lappi and Bailey (1988). Both Lappi and Bailey's and our approaches are first-order approximation methods (Vonesh and Chinchilli 1997, §7.4.2). Lappi and Bailey also fit a model that is an approximation to the ML-NLMM based on a Taylor series expansion about  $E(\mathbf{b}) = \mathbf{0}$ . Lappi and Bailey fit this model by combining ordinary least squares for the fixed



**Figure 2. Tree-specific observed and fitted growth curves for trees plotted in Figure 1.**

**Table 2. Cross-validation results. Prediction errors (true-predicted) of height at age 14 for individual trees and for plot averages based on Lappi and Bailey's (1988) method ("LB"), and the methods of this article ("HB"). All units are meters.**

| Predictions    | No. of predictions | Method | Mean error | SD    | Mean theoretical SD |
|----------------|--------------------|--------|------------|-------|---------------------|
| Stand averages | 24                 | HB     | -0.0305    | 0.643 | 0.628               |
|                |                    | LB     | -0.0710    | 0.625 | 0.311               |
| Single trees   | 365                | HB     | -0.261     | 0.765 | 0.698               |
|                |                    | LB     | -0.297     | 0.762 | 0.643               |

effects parameters with ad hoc methods of estimating the variance components associated with random effects. Essentially, what they do is use Liang and Zeger's (1986) original GEE approach with an independence working covariance structure, supplemented by ad hoc estimation of covariance parameters based on the residuals of the fitted fixed effects portion of the model. In contrast, we use an extended GEE approach (Hall and Severini, 1998), simultaneously solving estimating equations for fixed effects parameters and variance-covariance parameters associated with the random effects. In the latter approach, within-plot correlation is accounted for in the fixed effects parameter estimators, yielding greater efficiency.

With respect to prediction, Lappi and Bailey utilize the first-order linearized model (5) to generate their predictor and its prediction error variance. These authors ignore the estimation error that arises from using  $\mathbf{f}(\hat{\beta}, \mathbf{0}, \mathbf{X}_i, \mathbf{Z}_i)$  rather than  $\mathbf{f}(\beta, \mathbf{0}, \mathbf{X}_i, \mathbf{Z}_i)$  to construct their predictor. As a consequence, we see in Table 2 that the standard deviation of the prediction errors is under-estimated by Lappi and Bailey's theoretical standard deviation of prediction error, particularly in the prediction of plot averages. Otherwise, the current method provides a modest improvement in mean prediction error and similar prediction error variances to Lappi and Bailey's technique. Note that, for both methods, the prediction of plot average height at a given age is prediction of the average height of all trees in that plot at that age, rather than prediction of the average height of sampled trees at the age of interest.

Given the similarity in the basic approaches to modeling presented here and in Lappi and Bailey (1988), the small size of the improvement associated with our techniques seen in Table 2 is not surprising. In comparison to Lappi and Bailey's work, our approach offers improvements in efficiency with respect to estimation of the regression parameter  $\beta$ , and the covariance parameter  $\delta$ ; estimators of  $\text{var}(\beta)$  and  $\text{var}(\delta)$  that are more robust with respect to model misspecification; an improved estimator of the prediction error variance; but, most importantly, a unified methodology for estimation and prediction of growth curves based on ML-NLMMs as opposed to a collection of ad hoc techniques. That is, relative to Lappi and Bailey's approach, the improvements associated with our methodology do not come at a cost of increased complexity. Instead, we would argue that the reverse is true.

Perhaps a more relevant comparison is between our methodology and a traditional site index approach. Such approaches are both simpler and much more common in applied forestry than ML-NLMMs such as ours and Lappi and Bailey's. This type of comparison is complicated, however, by having to choose among the many published variants on the basic site index methodology. Lappi and Bailey (1988)

presented a comparison of their approach for predicting individual tree height and plot-average height with that of Borders et al. (1984), which was found by Borders et al. (1984) to be comparable with other site index methods. Lappi and Bailey's results demonstrate a clear improvement in both mean prediction error and prediction variance at both the tree and plot levels. Given the results of Table 2, the advantages of our approach over that of Borders, et al. can only be greater. Perhaps foremost among those advantages is the availability of a prediction variance estimate with our method. In the Borders, et al. technique, as in most site index approaches, no quantification of the uncertainty associated with their predictor is available.

## Discussion

In our approach to estimation we have utilized a first-order linearization based on a Taylor expansion about the random effects' mean,  $\mathbf{0}$ . Another possibility is to take the expansion about  $\bar{\mathbf{b}}$ , an estimate of  $\mathbf{b}$ , typically taken to be an EBLUP-type predictor similar to (9) based on the current parameter estimates (Vonesh and Chinchilli 1997, eq. 7.4.51). This EBLUP-expansion approach is described by Wolfinger and Lin (1997), Vonesh and Chinchilli (1997, §7.4.2.ii) and Lindstrom and Bates (1990). The essential difference from our zero-expansion approach is that the EBLUP-expansion method solves a set of extended GEEs based on different marginal moment approximations. Instead of the expressions given in (6), the marginal mean is approximated by

$$\mathbf{f}(\beta, \bar{\mathbf{b}}, \mathbf{X}_i, \mathbf{Z}_i) + \bar{\mathbf{Z}}_{2,i} \bar{\mathbf{b}}_{2,i} + \bar{\mathbf{Z}}_{3,i} \bar{\mathbf{b}}_{3,i}$$

where

$$\bar{\mathbf{Z}}_{m,i} = \left. \frac{\partial \mathbf{f}(\beta, \mathbf{b}_i, \mathbf{X}_i, \mathbf{Z}_i)}{\partial \mathbf{b}_{m,i}^T} \right|_{\mathbf{b}_i = \bar{\mathbf{b}}}, \quad m = 2, 3,$$

and the marginal variance is approximated as in (6) but with  $\bar{\mathbf{Z}}_{m,i}$  replaced by  $\bar{\mathbf{Z}}_{m,i}$ ,  $m = 1, 2$ . These approximations for the marginal first and second moments of model (4) are the mean and variance of the linearization of the model obtained by expanding around  $\bar{\mathbf{b}}$ , evaluated conditionally upon  $\bar{\mathbf{b}}$ .

Several authors have compared the performance of zero-expansion-type estimators with EBLUP-expansion-type estimators via simulation in both the GLMM and NLMM contexts (Vonesh 1992, Breslow and Clayton 1993, Pinheiro and Bates 1995, Wolfinger and Lin 1997). In the most recent and comprehensive of these studies, Wolfinger and Lin found that, "Overall, the EBLUP-expansion estimates are slightly better than the zero-expansion estimates in terms of bias and [mean squared error]" (Wolfinger and Lin 1997, p. 488). This statement is consistent with the results of other authors.

However, the magnitude of the improvement associated with the EBLUP expansion is highly dependent on the underlying model and data structure, ranging from negligible to moderate improvement. In addition, the potential advantages of an EBLUP expansion come at a significant cost in terms of computational expense and instability (Wolfinger and Lin 1997), and some decreased robustness with respect to the assumed random effects structure (Vonesh 1992). Such considerations suggest that further Monte Carlo-based comparisons of the two linearization methods would be useful in the context of specific forestry applications of the ML-NLMM; e.g., for height projections based on a Chapman-Richards model fit to longitudinal multilevel data of the sort typically available for growth and yield modeling.

In model (4) we have assumed that the errors are independent normal random variables with constant variance. As a consequence, the model implies that the responses are conditionally independent, given the values of the random effects. Marginally, however, the model implies that observations that share a random effect are correlated. For example, in a ML-NLMM with random tree-specific and plot-specific effects, observations from two different trees within the same plot are correlated because they share the same plot effect; observations from the same tree are correlated to an even higher degree because they share a plot effect and a tree effect. In this approach the correlation structure for the data is accounted for entirely through the random effects. Although the presence of the random effects implies a correlation structure, precisely what that marginal correlation structure is within a particular tree or plot is difficult to determine. That is, there is typically no closed-form solution for the within-tree or within-plot correlation matrix when the random effects enter into the model in a nonlinear fashion. For this reason, our approach is an example of indirect modeling of the variance-covariance structure (Schabenberger and Gregoire 1996).

Another approach to accounting for a dependence variance-covariance structure is to specify it directly through the variance-covariance matrix of  $\epsilon_i$ . That is, we could assume  $\text{var}(\epsilon_i) = \phi \mathbf{R}_i(\theta)$  rather than  $\phi \mathbf{I}$ , where  $\mathbf{R}_i$  is a correlation matrix with an assumed structure, parameterized by  $\theta$ , which has now been expanded to characterize both  $\Sigma_i$  and  $\mathbf{R}_i$ . For example,  $\mathbf{R}_i$  could be assumed to have a block-diagonal structure with blocks corresponding to observations from each tree in the plot. Furthermore, each block could be assumed to have the correlation structure of a continuous time first-order autoregressive (AR-1) process to capture the decaying nature of the pairwise correlations within a tree across increasing lags. Such a change poses no problem in the estimation and prediction procedures we have outlined above. The only differences are an increase in the dimension of  $\theta$  and an obvious change to  $\mathbf{V}_i(\delta)$  which becomes

$$\mathbf{V}_i(\delta) = \phi \{ \mathbf{R}_i(\theta) + \tilde{\mathbf{Z}}_i \Sigma_i(\theta) \tilde{\mathbf{Z}}_i^T \}.$$

There is typically a trade-off between the modeling of a variance-covariance structure indirectly through the random effects structure, and directly through the variance-covariance matrix of the errors. Although both types of

variance-covariance modeling are easily accommodated in the ML-NLMM, this is often unnecessary. As has been noted by Jones (1990), Gregoire and Schabenberger (1996b), and others, in the presence of random subject effects, additional within-subject correlation is often negligible. In particular, we expect that this will be true in our application of primary interest, growth modeling based on tree size variables measured repeatedly through time with relatively few measurements per tree and relatively long inter-measurement intervals.

In the context of single level NLMMs for cumulative bole volume of trees, Gregoire and Schabenberger (1996a, 1996b) compare certain zero-expansion and EBLUP-expansion linearization methods of estimation with a GEE-type approach due to Schabenberger (1994). These authors emphasize the connection between the linearization methods that they consider and REML, and they implement their linearization methods via an algorithm that fits a sequence of linear mixed models to pseudodata. However, in fact, the zero-expansion method that they utilize is equivalent to solving extended GEEs of the form given by (7) and (8) (with the REML-type adjustment), and their EBLUP-expansion method is equivalent to solving a similar extended GEE based on the first-order EBLUP approximation to the marginal mean and variance described above. The implication of this is that their zero-expansion estimation method and the GEE procedure that they consider differ only by the estimating equation that is utilized to obtain estimators of the variance-covariance parameters of the model.

Like the zero-expansion approximate REML procedure considered by Gregoire and Schabenberger (1996a), our zero-expansion method of estimation simply replaces the method of moments estimator of Schabenberger (1994) by an estimating equation for variance-covariance parameters. Since any method of moments estimator can be expressed as the root of an estimating equation that has a closed form solution, the methods differ in the substitution of one estimating equation for another. Thus, our method is not a fully second-order estimating equation method like the GEE2 approach of Prentice and Zhao (1991; see also Liang et al. 1992, Vonesh and Chinchilli 1997, §9.2.7, and Hall 2001). This is an important point because of the absence of a consistency result for the first-moment estimators based on GEE2 in the presence of second moment misspecification. Instead, our extended GEE approach yields a consistent estimator of  $\beta$  provided only that the first moment structure is correctly specified. In addition, based on results from the fixed effects generalized linear model (Hall and Severini 1998, Hall 2001), we can expect our extended GEE estimators to have greater finite sample efficiency than those based on a method of moments approach.

Using ML-NLMMs for growth and yield modeling provides several important advantages over traditional approaches, including the recently popular two-stage and three-stage least-squares methods (Borders 1989). Such methods, and others (Clutter et al. 1983), were motivated by the need to devise techniques appropriate for modeling remeasured tree and plot data. Invariably, how to formulate the best

system of related equations for modeling is problematic in such methods. In particular, the choice of which endogenous variables to include as predictors (right-hand-side variables) of the response is rarely easy. For example, the so-called “difference equation” methods (Clutter et al. 1983) require an early measurement on a unit (tree or plot) as a predictor in a model with a later measurement of the same variable as the response. No clear rationale exists for determining which pair of early and later observations (e.g., shortest intervals, longest intervals, all possible intervals) should be used in these models (Borders et al. 1988). In addition, with some exceptions, no provision is made in the fitting methods for the obvious violation of the usual independence assumption on the error structure. Moreover, when fitted models are used for projections of expected future values of the response (i.e., growth projections), no well-justified method exists for quantifying the uncertainty in these predictions.

The ML-NLMM approach we advocate herein obviates formulation of a system of equations for modeling and any decisions concerning which endogenous variables to use as predictors. By including random variables to represent tree and plot effects, dependence among current and past observations is accounted for within a unified regression model. Furthermore, the estimation methods and model structure lead naturally to a method of obtaining projected values of the response and corresponding standard errors.

## Literature Cited

- BAILEY, R.L., AND J.L. CLUTTER. 1974. Base-age invariant polymorphic site curves. *For. Sci.* 20:155–159.
- BÉGIN, J., AND J.P. SCHÜTZ. 1994. Estimation of total yield of Douglas-fir by means of incomplete growth series. *Annales des Sciences Forestières.* 51:345–355.
- BIGING, G.S. 1985. Improved estimates of site index curves using a varying-parameter model. *For. Sci.* 31:248–259.
- BOOTH, J.G., AND J.P. HOBERT. 1999. Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. *J.R. Stat. Soc. Ser. B Stat. Methodol.* 61:265–285.
- Borders, B.E. 1989. Systems of equations in forest stand modeling. *For. Sci.* 35:548–556.
- BORDERS, B.E., R.L. BAILEY, AND M.L. CLUTTER. 1988. Forest growth models: Parameter estimation using real growth series. P. 660–667 in *Forest growth modelling and prediction, Vol 2.* USDA For. Serv. Gen. Tech. Rep. NC-120.
- BRESLOW, N.E., AND D.G. CLAYTON. 1993. Approximate inference in generalized linear mixed models. *J. Am. Stat. Assoc.* 88:9–25.
- CANDY, S.G. 1997. Estimation in forest yield models using composite link functions with random effects. *Biometrics.* 53:146–160.
- CURTIS, R.O. 1964. A stem-analysis approach to site-index curves. *For. Sci.* 10:241–256.
- CLUTTER, J.L., J.C. FORTSON, L.V. PIENAAR, G.H. BRISTER, AND R.L. BAILEY. 1983. *Timber management—a quantitative approach.* Wiley, New York. 333 p.
- DAVIDIAN, M., AND D.M. GILTINAN. 1995. *Nonlinear models for repeated measurement data.* Chapman and Hall, London. 359 p.
- DUPLAT, P., AND M. TRAN-HA. 1997. Modelling the dominant height growth of sessile oak (*Quercus petraea* Liebl) in France. Inter-regional variability and effect of the recent period (1959–1993). *Annales des Sciences Forestières* 54:611–634.
- GARCIA, O. 1983. A stochastic differential equation model for the height growth of forest stands. *Biometrics* 39:1059–1072.
- GELFAND, A., AND B.P. CARLIN. 1993. Maximum likelihood for constrained- and missing-data models. *Can. J. Stat.* 21:303–311.
- GEYER, C.J., AND E.A. THOMPSON. 1992. Constrained Monte Carlo maximum likelihood for dependent data. *J.R. Stat. Soc. Ser. B Stat. Methodol.* 54:657–699.
- GOLDSTEIN, H. 1995. *Multilevel statistical models.* Ed. 2. Halstead Press, New York. 178 p.
- GREGOIRE, T.G., AND O. SCHABENBERGER. 1996a. Nonlinear mixed-effects modeling of cumulative bole volume with spatially correlated within-tree data. *J. Agric. Biol. Environ. Stat.* 1:107–119.
- GREGOIRE, T.G., AND O. SCHABENBERGER. 1996b. A non-linear mixed-effects model to predict cumulative bole volume of standing trees. *J. Appl. Statist.* 23:257–271.
- 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.
- HALL, D.B. 2001. On the application of extended quasilielihood to the clustered data case. *Can. J. Stat.* In press.
- HALL, D.B., AND R.L. BAILEY. 2000. Prediction of dominant height growth in forest plantations based on multilevel nonlinear mixed models. *Univ. of Georgia Stat. Dep. Tech. Rep. #2000-5.*
- HALL, D.B., AND T.A. SEVERINI. 1998. Extended generalized estimating equations for clustered data. *J. Am. Stat. Assoc.* 93:1365–1375.
- HARVILLE, D.A. 1977. Maximum likelihood approaches to variance component estimation and to related problems. *J. Am. Stat. Assoc.* 72:320–338.
- JONES, R.H. 1990. Serial correlation or random subject effects? *Comm. Stat. Simul. Comput.* 19:1105–1123.
- KUK, A.Y.C. 1999. The use of approximating models in Monte Carlo maximum likelihood estimation. *Stat. Probab. Lett.* 45:325–333.
- LAPPI, J. 1991. Calibration of height and volume equations with random parameters. *For. Sci.* 37:781–801.
- LAPPI, J. 1997. A longitudinal analysis of height/diameter curves. *For. Sci.* 43:555–570.
- LAPPI, J., AND R.L. BAILEY. 1988. A height prediction model with random stand and tree parameters: An alternative to traditional site index methods. *For. Sci.* 34:907–927.
- LAPPI, J., AND J. MALINEN. 1994. Random-parameter height/age models when stand parameters and stand age are correlated. *For. Sci.* 40:715–731.
- LIANG, K.-Y., AND S.L. ZEGER. 1986. Longitudinal analysis using generalized linear models. *Biometrika* 73:13–22.
- LIANG, K.-Y., S.L. ZEGER, AND B. QAQISH. 1992. Multivariate regression analyses for categorical data (with discussion). *J.R. Stat. Soc. Ser. B Stat. Methodol.* 54:3–40.
- LINDSTROM, M.J., AND D.M. BATES. 1990. Nonlinear mixed effects models for repeated measures data. *Biometrics.* 46:673–687.
- LITTELL, R.C., G.A. MILLIKEN, W.W. STROUP, AND R.D. WOLFINGER. 1996. *SAS<sup>®</sup> system for mixed models.* SAS Institute Inc., Cary, NC. 633 p.
- LONGFORD, N.T. 1993. *Random coefficient models.* Clarendon Press, Oxford. 270 p.
- MCCULLOCH, C.E. 1994. Maximum likelihood estimation of variance components for binary data. *J. Am. Stat. Assoc.* 89:330–335.
- MCCULLOCH, C.E. 1997. Maximum likelihood algorithms for generalized linear mixed models. *J. Am. Stat. Assoc.* 92:162–170.
- MONSERUD, R.A. 1984. Height growth and site index curves for inland Douglas-fir based on stem analysis data and forest habitat type. *For. Sci.* 30:943–965.
- PIENAAR, L.V., AND B.D. SHIVER. 1993. Early results from an old-field loblolly pine spacing study in the Georgia Piedmont with competition control. *South. J. Appl. For.* 17:193–196.

- PINHEIRO, J.C., AND D.M. BATES. 1995. Approximations to the log-likelihood function in the mixed-effects model. *J. Comput. Graph. Stat.* 4:12–35.
- PINHEIRO, J.C., AND D.M. BATES. 1996. Unconstrained parameterizations for variance-covariance matrices. *Stat. Comput.* 6:289–296.
- PINHEIRO, J.C., AND D.M. BATES. 2000. *Mixed-effects models in S and S-PLUS*. Springer, New York. 528 p.
- PRENTICE, R.L., AND L.P. ZHAO. 1991. Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses. *Biometrics*. 47:825–839.
- ROBINSON, G.K. 1991. That BLUP is a good thing: The estimation of random effects. *Stat. Sci.* 6:15–51.
- ROYALL, R.M. 1986. Model robust confidence intervals using maximum likelihood estimators. *Int. Stat. Rev.* 54:221–226.
- SCHABENBERGER, O. 1994. Nonlinear mixed effects growth models for repeated measures in ecology. P. 156–161 *in* Proc. of the 1994 Joint Statistical Meetings, section on statistics and the environment.
- SCHABENBERGER, O., AND T.G. GREGOIRE. 1996. Population-averaged and subject-specific approaches for clustered categorical data. *J. Statist. Comput. Simul.* 54:231–253.
- SEARLE, S.R., G. CASELLA, AND C.E. MCCULLOCH. 1992. *Variance components*. Wiley, New York. 501 p.
- SEBER, G.A.F., AND C.J. WILD. 1989. *Nonlinear regression*. Wiley, New York. 768 p.
- TAIT, D.E., C.J. CIESZEWSKI, AND I.E. BELLA. 1988. The stand dynamics of lodgepole pine. *Can. J. For. Res.* 18:1255–1260.
- VALLIANT, R. 1985. Nonlinear prediction theory and the estimation of proportions in a finite population. *J. Am. Stat. Assoc.* 80:631–641.
- VONESH, E.F. 1992. Non-linear models for the analysis of longitudinal data. *Stat. Med.* 11:1929–1954.
- VONESH, E.F., AND V.M. CHINCHILLI. 1997. *Linear and nonlinear models for the analysis of repeated measurements*. Marcel Dekker, New York. 560 p.
- WALTERS, D.K., H.E. BURKHART, M.R. REYNOLDS, AND T.G. GREGOIRE. 1991. A Kalman filter approach to localizing height-age equations. *For. Sci.* 37:1526–1537.
- WOLFINGER, R. 1999. Fitting nonlinear mixed models with the new NLMIXED procedure. *SAS User's Group Internat. Pap.* #287.
- WOLFINGER, R.D., AND X. LIN. 1997. Two Taylor-series approximation methods for nonlinear mixed models. *Comput. Stat. Data Anal.* 25:465–490.
- ZEGER, S.L., K.-Y. LIANG, AND P.S. ALBERT. 1988. Models for longitudinal data: a generalized estimating equation approach. *Biometrics*. 44:1049–1060.
- ZHANG, P. 1996. Nonparametric importance sampling. *J. Am. Stat. Assoc.* 91:1245–1253.