Biometrics 60, 16–24
March 2004

Multivariate Multilevel Nonlinear Mixed Effects Models for Timber Yield Predictions

Daniel B. Hall

Department of Statistics, University of Georgia, Athens, Georgia 30602-1952, U.S.A.
email: dhall@stat.uga.edu

and

Michael Clutter

Daniel B. Warnell School of Forest Resources, University of Georgia,
Athens, Georgia 30602, U.S.A.

Summary. Nonlinear mixed effects models have become important tools for growth and yield modeling in forestry. To date, applications have concentrated on modeling single growth variables such as tree height or bole volume. Here, we propose multivariate multilevel nonlinear mixed effects models for describing several plot-level timber quantity characteristics simultaneously. We describe how such models can be used to produce future predictions of timber volume (yield). The class of models and methods of estimation and prediction are developed and then illustrated on data from a University of Georgia study of the effects of various site preparation methods on the growth of slash pine (*Pinus elliottii* Engelm.).

Key words: Clustered data; Growth; Prediction; Random effects; Repeated measures; Volume.

1. Introduction

Recently, nonlinear mixed effects models (NLMMs) have become important tools for growth and yield modeling in forestry. Nonlinear mixed effects models allow the analyst to account for covariate or treatment effects with fixed effects (parameters) as in traditional nonlinear regression models, but they also allow one to account for multiple sources of heterogeneity and/or correlation in data through the inclusion of random effects in the model. Data on timber growth variables such as dominant height and basal area per hectare are typically collected repeatedly through time, often on permanent plots within one or more forest stands. Such a data structure lends itself naturally to analysis via mixed effects models. To date, however, forestry applications of NLMMs have been mainly limited to the analysis of univariate responses, e.g., to model height growth. A more important potential application is in the modeling, and especially prediction, of merchantable volume (yield). In this article, we describe how multilevel (ML) NLMMs can be used to fit multivariate models in which several forest growth variables are modeled simultaneously. Such models can then be used to predict volume at a future age of interest.

This article is organized as follows. In Section 2, we describe the class of multivariate ML-NLMMs that we propose for growth and yield modeling, and discuss the important topic of prediction based on this class of models. In Section 3, we focus on the special case of modeling a trivariate response consisting of dominant height, basal area, and trees per hectare, and describe how this particular model can be used to make volume predictions. Section 4 consists of an extensive example, in which data from a study of site preparation methods for slash pine (*Pinus elliottii* Engelm.) are used to illustrate the methodology. Some final comments are contained in Section 5.

2. Model Formulation and Assumptions

Recently, Pinheiro and Bates (2000, Chapter 7) have presented theory, computational methods, and software for ML-NLMMs. Their methodology is based on that of Lindstrom and Bates (1990). Their formulation is presented and illustrated exclusively as a univariate model. However, a multivariate ML-NLMM can be formulated as a special case of their model and, therefore, can be handled in the framework of their methodology. For consistency, we use the notation of these authors.

Suppose that the data of interest correspond to repeated measures on r response variables that are grouped hierarchically within the levels of multiple nested grouping factors. For example, we might be interested in jointly modeling dominant height, basal area, and trees per hectare, where these variables are measured repeatedly through time (at multiple ages) on multiple plots within several different stands or locations. In this example, the data would consist of a trivariate response in a three-level grouping structure. We follow Pinheiro and Bates (2000) in defining level 1 as the coarsest level of grouping, level 2 as the next coarsest level of grouping, etc. By convention, level 0 corresponds to the population level. In terms of the example, level 0 is the population represented by the stands sampled in the data set, level 1 is the stand level, level 2 is the plot level, level 3 is the individual age-specific measurement, and we have three-level data.

For simplicity, we present the model in the three-level case. Extension to >3 levels is conceptually straightforward, but becomes tedious notationally. Let $y_{ijk\ell}$ denote the ℓth response variable at the kth measurement occasion, on the jth second-level group, and ith first-level group. We assume that there are r response variables, M first-level groups, M_i second-level groups within the ith first-level group, and n_{ij} repeated measures on the jth second-level group within the ith first-level group. The model we consider assumes $y_{ijk\ell} = f_\ell(\boldsymbol{\phi}_{ijk\ell}, \mathbf{v}_{ijk\ell}) + \varepsilon_{ijk\ell}$, for $i = 1, \ldots, M$, $j = 1, \ldots, M_i$, $k = 1, \ldots, n_{ij}$, and $\ell = 1, \ldots, r$. Here f_ℓ is a real-valued, differentiable function of a vector-valued mixed effects "parameter" $\boldsymbol{\phi}_{ijk\ell}$ and a vector of covariates $\mathbf{v}_{ijk\ell}$, and $\varepsilon_{ijk\ell}$ is a mean zero normally distributed error term. Stacking the model equations for the r response variables $y_{ijk1}, \ldots, y_{ijkr}$ we obtain

$$\mathbf{y}_{ijk} = \mathbf{f}(\boldsymbol{\phi}_{ijk}, \mathbf{v}_{ijk}) + \boldsymbol{\varepsilon}_{ijk}, \quad (1)$$

where $\mathbf{y}_{ijk} = (y_{ijk1}, \ldots, y_{ijkr})^{\mathrm{T}}$ and $\mathbf{f}$, $\boldsymbol{\phi}_{ijk}$, $\mathbf{v}_{ijk}$, and $\boldsymbol{\varepsilon}_{ijk}$ are defined similarly. To account for correlation among the r response variables measured on the same occasion and unit of measurement, we assume the elements of $\boldsymbol{\varepsilon}_{ijk}$ are correlated, with variance–covariance matrix $\sigma^2\boldsymbol{\Lambda}$, so that $\boldsymbol{\varepsilon}_{ijk} \sim N(\mathbf{0}, \sigma^2\boldsymbol{\Lambda})$. The mixed effects parameter $\boldsymbol{\phi}_{ijk\ell}$ takes the form

$$\boldsymbol{\phi}_{ijk\ell} = \mathbf{A}_{ijk\ell}\boldsymbol{\beta}_\ell + \mathbf{B}^{(1)}_{ijk\ell}\mathbf{b}^{(1)}_{i,\ell} + \mathbf{B}^{(2)}_{ijk\ell}\mathbf{b}^{(2)}_{ij,\ell},$$

where $\mathbf{b}^{(1)}_{i,\ell}$, $\mathbf{b}^{(2)}_{ij,\ell}$ are first and second level, respectively, random effects vectors specific to the ℓth response variable, and $\mathbf{B}^{(1)}_{ijk\ell}$ and $\mathbf{B}^{(2)}_{ijk\ell}$ are the associated random effects design matrices. The fixed effects design matrix and parameter vector specific to the ℓth response are $\mathbf{A}_{ijk\ell}$ and $\boldsymbol{\beta}_\ell$, respectively. Stacking the fixed and random effects vectors for the r response variables, we define $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^{\mathrm{T}}, \ldots, \boldsymbol{\beta}_r^{\mathrm{T}})^{\mathrm{T}}$, $\mathbf{b}^{(1)}_i = (\mathbf{b}^{(1)\mathrm{T}}_{i,1}, \ldots, \mathbf{b}^{(1)\mathrm{T}}_{i,r})^{\mathrm{T}}$, and $\mathbf{b}^{(2)}_{ij} = (\mathbf{b}^{(2)\mathrm{T}}_{ij,1}, \ldots, \mathbf{b}^{(2)\mathrm{T}}_{ij,r})^{\mathrm{T}}$. In addition, let $\mathbf{A}_{ijk} = \mathrm{diag}(\mathbf{A}_{ijk1}, \ldots, \mathbf{A}_{ijkr})$, $\mathbf{B}^{(1)}_{ijk} = \mathrm{diag}(\mathbf{B}^{(1)}_{ijk1}, \ldots, \mathbf{B}^{(1)}_{ijkr})$, and $\mathbf{B}^{(2)}_{ijk} = \mathrm{diag}(\mathbf{B}^{(2)}_{ijk1}, \ldots, \mathbf{B}^{(2)}_{ijkr})$. Then we can write

$$\boldsymbol{\phi}_{ijk} = \mathbf{A}_{ijk}\boldsymbol{\beta} + \mathbf{B}^{(1)}_{ijk}\mathbf{b}^{(1)}_i + \mathbf{B}^{(2)}_{ijk}\mathbf{b}^{(2)}_{ij}, \quad (2)$$

where we assume

$$\mathbf{b}^{(1)}_1, \ldots, \mathbf{b}^{(1)}_M \overset{\text{i.i.d.}}{\sim} N(\mathbf{0}, \boldsymbol{\Psi}_1), \quad \mathbf{b}^{(2)}_{11}, \ldots, \mathbf{b}^{(2)}_{MM_M} \overset{\text{i.i.d.}}{\sim} N(\mathbf{0}, \boldsymbol{\Psi}_2) \quad (3)$$

and the first- and second-level random effects, $\{\mathbf{b}^{(1)}_i\}$ and $\{\mathbf{b}^{(2)}_{ij}\}$, are assumed independent of each other (cf. Pinheiro and Bates, 2000, equation 7.6, p. 309).

In general, we put no constraints on $\boldsymbol{\Psi}_1$ and $\boldsymbol{\Psi}_2$ other than assuming that they are variance–covariance matrices. However, in specific model building, much of the work lies in specifying structures for these matrices. For reasons of computational stability and speed, as well as parsimony, it is advantageous to reduce the number of parameters estimated in these matrices by setting certain covariances to zero. However, given the structure of the model, there is often no a priori reason to impose independence conditions on the random effects operating on a given level. Indeed for the timber growth and yield applications we are mainly concerned with, we expect that plot-level random effects, for example, that pertain to distinct response variables measured on the same plot will typically be correlated. It is a strength of the above model that it is general enough to account for such correlations as opposed to simpler models that assume, for example, only correlated errors through a nondiagonal $\boldsymbol{\Lambda}$ matrix.

As a special case of the model of Pinheiro and Bates (2000, Chapter 7), the model given above by (1)–(3) can be fit using these authors' nlme software for S-PLUS. The methodology implemented in this software is essentially that of Lindstrom and Bates (1990) with extensions due to Pinheiro and Bates (2000, Chapter 7) to handle heteroskedasticity, correlated errors, and nested (multilevel) random effects. Because this methodology is thoroughly described elsewhere (Davidian and Giltinan, 1995, Chapter 6; Vonesh and Chinchilli, 1997, Chapter 7; Wolfinger and Lin, 1997; Pinheiro and Bates, 2000, Chapter 7), we do not repeat that description here in any detail.

Briefly, the Lindstrom and Bates algorithm consists of iteratively fitting a sequence of linear mixed effects models derived as linear approximations to the NLMM given by (1). The linear approximation to (1) is accomplished by taking a first-order Taylor expansion of $\mathbf{f}(\boldsymbol{\phi}_{ijk}, \mathbf{v}_{ijk})$ around a predictor of the combined random effects vector formed based on parameter estimates from the current iteration. It has been shown that this procedure is equivalent to solving a set of estimating equations where the estimating functions are approximate first derivatives of a Laplace approximation to the log likelihood of the original model (Wolfinger, 1993; Wolfinger and Lin, 1997). The approach can be motivated and derived in several ways, but it is essentially an estimating equation methodology which can be thought of either as an extension of generalized least squares (see Vonesh and Chinchilli, 1997, Chapter 9, for details) or as approximate maximum likelihood estimation. The latter interpretation, as approximate maximum likelihood estimation, is more often emphasized in the literature, and likelihood-based methods of inference are advocated by Lindstrom and Bates (1990) and Pinheiro and Bates (2000). In this approach the first-order approximate log likelihood is treated as the true log likelihood, and standard errors for parameter estimates, likelihood ratio tests for nested models, and model selection criteria such as AIC and BIC are formed in the usual way. Although the formal justification of this "approximately asymptotic" approach to inference is an open problem, it is commonly used in practice, and we adopt it for our purposes in this article.

2.1 *Prediction*

Because we are interested in making forecasts of future growth and yield, we will devote more attention to the topic of prediction. By prediction here, we mean prediction of the value of a response vector for a unit and/or measurement occasion not contained in the original sample, based upon the sample data and fitted model. We assume that there is some dependence between the sample data and the unobserved data for which predictions are desired. For example, we may want to predict the value for a unit on which past measurements are available, or for which measurements are available on different same-level units, within a common higher level unit. More specifically, we may be interested in predicting size characteristics of a plot at age 25 in the presence of data on that plot at previous ages, or from age 25 data for another plot within the same stand. Another example is the situation in which we wish to predict one response variable (dominant height, say) from data on the other response variables (e.g., basal area and trees per hectare) on the same unit at the same measurement occasion.

Pinheiro and Bates (2000, Chapter 7) discuss simple plug-in type predictors and implement these predictors in their nlme software. Let $\mathbf{v}_h$ be a vector of values for the model covariates at which a prediction is desired. Then the plug-in predictor at the mth level of grouping is given by $\mathbf{f}(\hat{\boldsymbol{\phi}}_h^{(m)}, \mathbf{v}_h)$, where $\hat{\boldsymbol{\phi}}_h^{(m)} = \mathbf{A}_h^{\mathrm{T}}\hat{\boldsymbol{\beta}} + 1_{\{m>0\}}\sum_{k=1}^{m}\mathbf{B}_h^{(k)}\hat{\mathbf{b}}_h^{(k)}, m = 0, 1, \ldots$. Here, $\mathbf{A}_h$ is the matrix of fixed effects covariates at which a prediction is desired, and $\mathbf{B}_h^{(k)}$ is the design matrix for kth-level random effects $\mathbf{b}_h^{(k)}$ with final model estimate (prediction, really) $\hat{\mathbf{b}}_h^{(k)}$. In addition, $1_{\{C\}}$ denotes the indicator variable taking on the value 1 when condition C is true, 0 otherwise. That is, the plug-in predictor at the mth level simply plugs in new covariate values of interest into the fitted model with random effects at the $(m+1)$th level and higher set equal to their mean value, zero.

Based on a linear mixed model (LMM) approximation of the NLMM, we suggest an alternative predictor analogous to the empirical best linear unbiased predictor (BLUP) of the LMM (e.g., Christensen, 1996, Chapter 12). To describe our predictor, it is convenient to write the NLMM generically as $\mathbf{y} = \mathbf{f}(\boldsymbol{\beta}, \mathbf{b}, \mathbf{A}, \mathbf{B}) + \boldsymbol{\varepsilon}$. Taking a first-order Taylor linearization of $\mathbf{f}(\boldsymbol{\beta}, \mathbf{b}, \mathbf{A}, \mathbf{B})$ around the estimated values $(\boldsymbol{\beta}, \mathbf{b}) = (\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}})$, yields

$$\mathbf{y} \approx \mathbf{f}(\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}}, \mathbf{A}, \mathbf{B}) + \tilde{\mathbf{A}}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + \tilde{\mathbf{B}}(\mathbf{b} - \hat{\mathbf{b}}) + \boldsymbol{\varepsilon}, \qquad (4)$$

where

$$\tilde{\mathbf{A}} = \left.\frac{\partial \mathbf{f}(\boldsymbol{\beta}, \mathbf{b}, \mathbf{A}, \mathbf{B})}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}, \mathbf{b}=\hat{\mathbf{b}}}, \quad \tilde{\mathbf{B}} = \left.\frac{\partial \mathbf{f}(\boldsymbol{\beta}, \mathbf{b}, \mathbf{A}, \mathbf{B})}{\partial \mathbf{b}}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}, \mathbf{b}=\hat{\mathbf{b}}}.$$

Equivalently,

$$\mathbf{z} = \tilde{\mathbf{A}}\boldsymbol{\beta} + \tilde{\mathbf{B}}\mathbf{b} + \boldsymbol{\varepsilon}, \qquad (5)$$

where $\mathbf{z} = \mathbf{y} - \mathbf{f}(\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}}, \mathbf{A}, \mathbf{B}) + \tilde{\mathbf{A}}\hat{\boldsymbol{\beta}} + \tilde{\mathbf{B}}\hat{\mathbf{b}}$. Note that (5) takes the form of an LMM for "pseudo-data" $\mathbf{z}$. The Lindstrom and Bates method of fitting the NLMM consists of iteratively fitting this LMM until convergence, with $\mathbf{z}$ updated at each step based on the current estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{b}}$.

Now suppose we decompose the response vector $\mathbf{y} = (\mathbf{y}_s^{\mathrm{T}}, \mathbf{y}_h^{\mathrm{T}})^{\mathrm{T}}$ into observed and unobserved components (s for sample, denoting the observed component, h denoting the unobserved component) and all other model quantities accordingly:

$$\mathbf{z} = \begin{pmatrix}\mathbf{z}_s \\ \mathbf{z}_h\end{pmatrix}, \quad \tilde{\mathbf{A}} = \begin{pmatrix}\tilde{\mathbf{A}}_s \\ \tilde{\mathbf{A}}_h\end{pmatrix}, \quad \tilde{\mathbf{B}} = \begin{pmatrix}\tilde{\mathbf{B}}_s \\ \tilde{\mathbf{B}}_h\end{pmatrix},$$

$$\mathbf{f}(\boldsymbol{\beta}, \mathbf{b}, \tilde{\mathbf{A}}, \tilde{\mathbf{B}}) = \begin{pmatrix}\mathbf{f}_s(\boldsymbol{\beta}, \mathbf{b}, \tilde{\mathbf{A}}_s, \tilde{\mathbf{B}}_s) \\ \mathbf{f}_h(\boldsymbol{\beta}, \mathbf{b}, \tilde{\mathbf{A}}_h, \tilde{\mathbf{B}}_h)\end{pmatrix}, \quad \text{etc.}$$

Then under the LMM (5), the empirical BLUP of $\mathbf{z}_h$ based on $\mathbf{z}_s$ is given by

$$\hat{\mathbf{z}}_h = \tilde{\mathbf{A}}_h\hat{\boldsymbol{\beta}} + \tilde{\mathbf{V}}_{hs}\tilde{\mathbf{V}}_{ss}^{-1}(\mathbf{z}_s - \tilde{\mathbf{A}}_s\hat{\boldsymbol{\beta}}), \qquad (6)$$

where $\tilde{\mathbf{V}} = \tilde{\mathbf{B}}\widehat{\mathrm{var}}(\mathbf{b})\tilde{\mathbf{B}}^{\mathrm{T}} + \widehat{\mathrm{var}}(\boldsymbol{\varepsilon})$, the estimated variance–covariance matrix of $\mathbf{z}$ according to the LMM approximation (5), decomposes as

$$\tilde{\mathbf{V}} = \begin{pmatrix}\tilde{\mathbf{V}}_{ss} & \tilde{\mathbf{V}}_{sh} \\ \tilde{\mathbf{V}}_{hs} & \tilde{\mathbf{V}}_{hh}\end{pmatrix}.$$

Using the relationship between pseudo-data $\mathbf{z}$ and real data $\mathbf{y}$, we rearrange (6) to yield our predictor for $\mathbf{y}_h$:

$$\begin{aligned}\hat{\mathbf{y}}_h &= \mathbf{f}_h(\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}}, \mathbf{A}_h, \mathbf{B}_h) - \tilde{\mathbf{B}}_h\hat{\mathbf{b}} \\ &\quad + \tilde{\mathbf{V}}_{hs}\tilde{\mathbf{V}}_{ss}^{-1}\{\mathbf{y}_s - \mathbf{f}_s(\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}}, \mathbf{A}_s, \mathbf{B}_s) + \tilde{\mathbf{B}}_s\hat{\mathbf{b}}\}. \qquad (7)\end{aligned}$$

An approximation for the prediction variance $\mathrm{var}(\hat{\mathbf{y}}_h - \mathbf{y}_h)$ can be obtained by again utilizing the LMM approximation (4). Applying (4) to $\mathbf{y}_h$, we have (after some rearrangement) $\mathbf{y}_h - \hat{\mathbf{y}}_h \approx (-\tilde{\mathbf{V}}_{hs}\tilde{\mathbf{V}}_{ss}^{-1} \vdots \mathbf{I}_h)\{\tilde{\mathbf{A}}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + \tilde{\mathbf{B}}\mathbf{b} + \boldsymbol{\varepsilon}\}$. Therefore,

$$\begin{aligned}\mathrm{var}(\mathbf{y}_h - \hat{\mathbf{y}}_h) &\approx \left(-\tilde{\mathbf{V}}_{hs}\tilde{\mathbf{V}}_{ss}^{-1} \vdots \mathbf{I}_h\right)\mathrm{var}\{\tilde{\mathbf{A}}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + \tilde{\mathbf{B}}\mathbf{b} + \boldsymbol{\varepsilon}\} \\ &\quad \times \left(-\tilde{\mathbf{V}}_{hs}\tilde{\mathbf{V}}_{ss}^{-1} \vdots \mathbf{I}_h\right)^{\mathrm{T}}.\end{aligned}$$

Using the approximations

$$\mathrm{var}\begin{pmatrix}\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \\ \hat{\mathbf{b}} - \mathbf{b}\end{pmatrix} \approx \begin{pmatrix}\tilde{\mathbf{A}}_s^{\mathrm{T}}\mathbf{R}_s^{-1}\tilde{\mathbf{A}}_s & \tilde{\mathbf{A}}_s^{\mathrm{T}}\mathbf{R}_s^{-1}\tilde{\mathbf{B}}_s \\ \tilde{\mathbf{B}}_s^{\mathrm{T}}\mathbf{R}_s^{-1}\tilde{\mathbf{A}}_s & \mathbf{D}^{-1} + \tilde{\mathbf{B}}_s^{\mathrm{T}}\mathbf{R}_s^{-1}\tilde{\mathbf{B}}_s\end{pmatrix}^{-}$$

$$\equiv \tilde{\mathbf{G}} = \begin{pmatrix}\tilde{\mathbf{G}}_{11} & \tilde{\mathbf{G}}_{12} \\ \tilde{\mathbf{G}}_{12}^{\mathrm{T}} & \tilde{\mathbf{G}}_{22}\end{pmatrix},$$

and $\mathrm{cov}(\hat{\boldsymbol{\beta}}, \boldsymbol{\varepsilon}_s) \approx \tilde{\mathbf{G}}_{11}\tilde{\mathbf{A}}_s + \tilde{\mathbf{G}}_{12}\tilde{\mathbf{B}}_s^{\mathrm{T}}$, we obtain

$$\begin{aligned}&\mathrm{var}(\mathbf{y}_h - \hat{\mathbf{y}}_h) \\ &\approx \left(-\tilde{\mathbf{V}}_{hs}\tilde{\mathbf{V}}_{ss}^{-1} \vdots \mathbf{I}_h\right) \\ &\quad \times \left(\tilde{\mathbf{F}}\tilde{\mathbf{G}}\tilde{\mathbf{F}}^{\mathrm{T}} - \tilde{\mathbf{B}}\tilde{\mathbf{G}}_{22}\tilde{\mathbf{B}}^{\mathrm{T}} + \tilde{\mathbf{B}}\mathbf{D}\tilde{\mathbf{B}}^{\mathrm{T}} + \mathbf{R}_s - \mathbf{M} - \mathbf{M}^{\mathrm{T}}\right) \\ &\quad \times \left(-\tilde{\mathbf{V}}_{hs}\tilde{\mathbf{V}}_{ss}^{-1} \vdots \mathbf{I}_h\right)^{\mathrm{T}}. \qquad (8)\end{aligned}$$

Here, $\mathbf{R}_s = \mathrm{var}(\boldsymbol{\varepsilon}_s), \tilde{\mathbf{F}} = (\tilde{\mathbf{A}} \vdots \tilde{\mathbf{B}})$, and $\mathbf{M} = \{\tilde{\mathbf{A}}(\tilde{\mathbf{G}}_{11}\tilde{\mathbf{A}}_s^{\mathrm{T}} + \tilde{\mathbf{G}}_{12}\tilde{\mathbf{B}}_s^{\mathrm{T}}) \vdots \mathbf{0}_{\dim(\mathbf{y}),\dim(\mathbf{y}_h)}\}$.

The predictor given by (7) can be thought of as an approximation to $\mathrm{E}(\mathbf{y}_h \mid \mathbf{y}_s, \mathbf{A}, \mathbf{B})$ based on a first-order linearization of $\mathbf{f}(\boldsymbol{\beta}, \mathbf{b}, \mathbf{A}, \mathbf{B})$. In the case of the LMM, $\mathbf{f}(\boldsymbol{\beta}, \mathbf{b}, \mathbf{A}, \mathbf{B}) = \mathbf{A}\boldsymbol{\beta} + \mathbf{B}\mathbf{b}$, and we recover the empirical BLUP $\mathbf{A}_h\hat{\boldsymbol{\beta}} + \mathbf{V}_{hs}\mathbf{V}_{ss}^{-1}(\mathbf{y}_s - \mathbf{A}_s\hat{\boldsymbol{\beta}})$ from expression (7). When $\mathrm{cov}(\boldsymbol{\varepsilon}_h, \boldsymbol{\varepsilon}_s) = \mathbf{0}$, $\mathbf{y}_h$ and $\mathbf{y}_s$ are correlated only through shared random effects in $\mathbf{b}$. Therefore, in this situation $\mathrm{E}(\mathbf{y}_h \mid \mathbf{y}_s, \mathbf{A}, \mathbf{B}) = \mathrm{E}(\mathbf{y}_h \mid \mathbf{b}, \mathbf{A}_h, \mathbf{B}_h) = \mathbf{f}(\boldsymbol{\beta}, \mathbf{b}, \mathbf{A}_h, \mathbf{B}_h)$, which is approximated by the plug-in predictor $\mathbf{f}(\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}}, \mathbf{A}_h, \mathbf{B}_h)$. In such a situation, we expect the plug-in predictor and (7) to differ very little. However, when $\mathrm{cov}(\boldsymbol{\varepsilon}_h, \boldsymbol{\varepsilon}_s) \neq \mathbf{0}$ this dependence is taken into account in (7) through $\tilde{\mathbf{v}}_{hs}$, but ignored in the plug-in predictor. So, if it is of interest to predict the value of one of the response variables from data that include the other response variables measured at the same time on the same unit, or if we assume some autocorrelation model in $\text{var}(\boldsymbol{\varepsilon})$, then we should expect (7) to improve upon $\mathbf{f}(\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}}, \mathbf{A}, \mathbf{B})$. We compare these predictors in Section 4 on our example data set. A more complete evaluation of the relative performance of these predictors is worthwhile, but will be pursued elsewhere.

3. A Multivariate Three-Level NLMM for Timber Growth and Yield

Timber growth and yield models are often based on repeated measures of readily measured plot-level variables related to tree size and abundance. Although volume at harvest is often of primary interest, it is infeasible to directly measure the volume of standing trees; instead average height of all or a subclass of all trees (e.g., just dominant and codominant trees) is measured, as well as total basal area or average basal area, and number of trees per unit area. Predictions of such variables as well as predictions of volume at potential harvest ages are desired. Traditionally, nonlinear sigmoidal growth curves such as the Chapman–Richards model, logistic model, and others (e.g., Ratkowsky, 1990, Chapter 4) have been used to model individual growth variables such as average dominant height (hd) and basal area per hectare (ba). Both linear and nonlinear models are commonly used for modeling trees per hectare. We propose to utilize a trivariate NLMM for modeling these three variables simultaneously. We assume that the available data have a multilevel structure. For example, plot-level measurements of hd, ba, and trees per hectare (tph) obtained repeatedly through time on each of several plots within each of several stands or locations. We use this example as the motivating problem, although a variety of other multilevel multivariate problems can be handled similarly. In the example, we therefore have a trivariate three-level NLMM, formed by combining commonly used NLMMs for hd, ba, and tph.

Specifically, we propose to use models of a modified Chapman–Richards form for both hd and ba, and to use a modified negative exponential model for tph. In the notation of Section 2, $r = 3$ and $\ell = 1, 2, 3$ correspond to responses hd, ba, and tph, respectively. We take

$$f_\ell(\boldsymbol{\phi}_{ijk\ell}, \mathbf{v}_{ijk\ell}) = \theta^{(\text{EV})}_{ijk\ell} \left\{ \frac{1 - \exp\left(-\theta^{(\text{Rate})}_{ijk\ell} \text{Age}_{ijk}\right)}{1 - \exp\left(-\theta^{(\text{Rate})}_{ijk\ell} \text{Age}_R\right)} \right\}^{\theta^{(\text{Shape})}_{ijk\ell}}, \quad \ell = 1, 2, \qquad (9)$$

and

$$f_\ell(\boldsymbol{\phi}_{ijk\ell}, \mathbf{v}_{ijk\ell}) = \text{Itph}_{ij} \left(\frac{\text{Age}_{ijk}}{\text{Age}_0} \right)^{\theta^{(1)}_{ijk\ell}} \times \exp\left\{\theta^{(2)}_{ijk\ell}(\text{Age}_{ijk} - \text{Age}_0)\right\}, \quad \ell = 3. \qquad (10)$$

Here, Age_{ijk} is the age of the trees at measurement occasion k, Age_R is a fixed reference age, Itph_{ij} is the "initial" tph measured at a young age after the first thinning (in our example Itph is measured at age two years), and Age_0 is the age at which Itph is measured. Equation (9) corresponds to a Chapman–Richards model with an expected value parameterization for the asymptote (cf. Ratkowsky, 1990, equation 4.3.6, p. 110). For a fixed effects model, parameter $\theta^{(\text{EV})}_{ijk\ell}$ would have an interpretation as the expected value of response ℓ at the reference age Age_R. In addition, $\theta^{(\text{Rate})}_{ijk\ell}$ and $\theta^{(\text{Shape})}_{ijk\ell}$ are rate of growth and curve shape parameters, respectively. Equation (10) can be derived from the assumption that the proportional mortality rate over time is inversely related to age (cf. Clutter et al., 1983, equation 4.68, p. 133). In equation (9), $\boldsymbol{\phi}_{ijk\ell} = (\theta^{(\text{EV})}_{ijk\ell}, \theta^{(\text{Rate})}_{ijk\ell}, \theta^{(\text{Shape})}_{ijk\ell})^{\text{T}}$, and $\mathbf{v}_{ijk\ell} = \text{Age}_{ijk}$, $\ell = 1, 2$. In (10), we have $\boldsymbol{\phi}_{ijk3} = (\theta^{(1)}_{ijk3}, \theta^{(2)}_{ijk3})^{\text{T}}$ and $\mathbf{v}_{ijk3} = (\text{Itph}_{ij}, \text{Age}_{ijk})^{\text{T}}$. We will consider specific forms for the mixed effects parameters $\boldsymbol{\phi}_{ijk\ell}$, $\ell = 1, 2, 3$, and the random effects variance–covariance matrices in the example of Section 4.

3.1 *Volume Predictions*

Using the trivariate three-level NLMM described above combined with the methods of Section 2.1, we can obtain predictions of the plot-level variables hd, ba, and tph at a future age of interest (e.g., a possible age of harvest). We propose to use these predictions as inputs into either a known or a fitted model for volume as a function of these variables. The idea here is to think of the standing timber on a plot as a three-dimensional object whose volume is computable as a function of its dimensions. This object is an aggregation of the individual trees on the plot, each of which has a volume which is a function of its basal area and its height. Therefore, the "dimensions" of the standing timber on a given plot are basal area, average tree height, and trees per acre. By this reasoning, we propose to use a plot-level (or whole stand) volume equation in the same way as an individual-tree volume equation (volume as a function of height and basal diameter) is used. That is, we take plot volume as a function of plot hd, ba, and tph as a known geometric relationship. Predictions of future volume are obtained simply by inputting predicted hd, ba, and tph values into this function.

Conceivably, standard plot-level volume equations could be developed for each species of interest and taken as known functions. For a known plot-level volume function, volume predictions would be obtained by evaluating the function at the predicted values of the arguments (hd, ba, tph). A prediction error variance could then be obtained from the variances of the predictions of hd, ba, and tph using the delta method. In the absence of a known volume equation, a model for volume as a function of hd, ba, and tph must be developed. Ideally, such a model would be fit to directly measured volume data. Because such data are rarely available, it will typically be necessary to use derived plot-level volumes. These derived plot-level volumes can be computed in several ways (Clutter et al., 1983, Section 1.3). For example, individual tree-level diameters and heights may be measured on all trees and then plugged into an individual-tree volume equation (taken as known). The resulting tree-level volumes are then summed to produce a plot-level volume. Derived plot-level volumes can then be used to fit a model for volume as a function of hd, ba, and tph.

In our example of the following section, such a volume model is fit to the same data to which our trivariate NLMM for hd, ba, and tph is fit. Alternatively, the volume equation could be developed from another data source. Either way, we propose to treat the fitted volume equation as fixed, and we ignore the variance of the volume equation parameter estimates and their covariance with predicted hd, ba, and tph in computing prediction error variances for predicted volumes. Of course, we are ignoring a nonnegligible source of error with this approach and we expect that observed prediction errors will vary more than our estimated prediction error variance. However, what we will have ignored has nothing to do with uncertainty about growth, which is our primary interest. Rather, the ignored variability pertains only to uncertainty concerning the geometric relationship between the volume of a timber plot object and that object's dimensions (hd, ba, and tph).

4. Example—A Slash Pine Site Preparation Study

In 1979, the Plantation Management Research Cooperative of the University of Georgia's Daniel B. Warnell School of Forest Resources initiated a study of various site preparation and intensive management regimes on the growth of slash pine in the lower coastal plain region of Georgia and Florida. Data consisting of repeated plot-level measurements of hd (m), ba (m^2/ha), tph (100s of trees/ha), derived volume (total volume outside bark in m^3/ha), and other variables at ages 2, 5, 8, 11, 14, 17, and 20 years were available for analysis. These repeated measures were taken on a total of 191 0.2-ha plots nested within 16 sites in the study region. At each site, plots were randomized to 11 treatments consisting of a subset of the $2^5 = 32$ combinations of five two-level (absent/present) treatment factors. These factors are A = Chop, site preparation with a single pass of a rolling drum chopper; B = Fert, fertilizer applications following the first, 12th, and 17th growing seasons; C = Burn, a broadcast burn of the site prior to planting; D = Bed, a double pass bedding of the site; and E = Herb, vegetation control with a chemical herbicide (Shiver and Harrison, 2000).

We fit a trivariate three-level NLMM as described in Section 3 using age 5–20 data from this data set. Age two was regarded as an initial condition from which Itph and Age_0 were obtained for model (10). Because of its common use as a reference age in site index models, we used $\text{Age}_R = 25$ years in (9). To develop the trivariate model, first models (9) and (10) were fit to hd, ba, and tph separately as univariate three-level NLMMs. In the separate fittings, specifications for the mixed effects parameters $\boldsymbol{\phi}_{ijkl}$, $\ell = 1, 2, 3$ were chosen using the AIC model selection criterion. All eight mixed effects parameters, $\theta_{ijk1}^{(\text{EV})}$, $\theta_{ijk1}^{(\text{Rate})}$, $\theta_{ijk1}^{(\text{Shape})}$, $\theta_{ijk2}^{(\text{EV})}$, $\theta_{ijk2}^{(\text{Rate})}$, $\theta_{ijk2}^{(\text{Shape})}$, $\theta_{ijk3}^{(1)}$, and $\theta_{ijk3}^{(2)}$ were taken as mixed, with random plot-level intercepts and random site-level intercepts in all of these parameters. In addition, the fixed effects specifications obtained for these eight parameters is given in Table 1. In Table 1, we use the Wilkinson and Rogers notation to specify the fixed effects portion of the linear predictors (see McCullagh and Nelder, 1989, Section 3.4 for details on this notation). In Table 1, Soil represents a soil-type factor with two levels: poorly drained nonspodosol soil, versus other soil (moderately well-drained nonspodosol and spodosol soil types); and SI represents a plot-specific site index computed based on the previous rotation.

In the univariate models, we allowed for heteroskedasticity by utilizing a conditional error variance specification as in Davidian and Giltinan (1995, Chapter 4) and Pinheiro and Bates (2000, Section 5.2). That is, we assume $\text{var}(\varepsilon_{ijk\ell} \mid \mathbf{b}_i, \mathbf{b}_{ij}) = \sigma^2 g^2(\mu_{ijk\ell}, \mathbf{w}_{ijk\ell}, \boldsymbol{\delta})$, where $\mu_{ijk\ell} = \text{E}(y_{ijk\ell} \mid \mathbf{b}_{i,2}^{(1)}, \mathbf{b}_{ij,2}^{(2)})$, $\mathbf{w}_{ijk\ell}$ is a vector of variance covariates, and $\boldsymbol{\delta}$ is a possibly vector-valued variance parameter. The nlme software allows the specification of a wide variety of variance functions. In the univariate models for hd, ba, and tph, we found evidence of heteroskedasticity only for ba. Therefore, we used a power of the fitted value variance function for ba: $\text{var}(\varepsilon_{ijk2} \mid \mathbf{b}_{i,2}^{(1)}, \mathbf{b}_{ij,2}^{(2)}) = \sigma^2 |\hat{\mu}_{ijk2}|^{2\delta}$. In the univariate NLMMs, we used model selection criteria (AIC and BIC) to choose variance–covariance structures for $\text{var}(\mathbf{b}_{i,2}^{(1)})$ and $\text{var}(\mathbf{b}_{ij,2}^{(2)})$. In all three cases, general (unstructured) positive definite forms for these variance–covariance matrices were selected. Log-likelihood values for the three separately fit univariate NLMMs, referred to as models hd1, ba1, and tph1, are given in Table 2.

Table 1
Specifications for fixed effects portions of the mixed effects parameters in separate univariate three-level NLMM fits

Mixed effect parameter	Fixed effects model formula
$\theta_{ijk1}^{(\text{EV})}$	1 + Burn + Fert × Herb
$\theta_{ijk1}^{(\text{Rate})}$	1 + Bed + Chop + Fert + Soil
$\theta_{ijk1}^{(\text{Shape})}$	1 + Chop + Soil + Fert × Herb + Itph + Itph^2
$\theta_{ijk2}^{(\text{EV})}$	1 + Bed + Burn + Fert + Herb + Itph
$\theta_{ijk2}^{(\text{Rate})}$	1 + Chop + Fert + Soil + Itph + Itph^2
$\theta_{ijk2}^{(\text{Shape})}$	1 + Bed + Fert + Herb + Soil + Itph + Itph^2
$\theta_{ijk3}^{(1)}$	1
$\theta_{ijk3}^{(2)}$	1 + SI

In addition, in Table 2 we also display log likelihoods for the model hbt1 = hd1 + ba1 + tph1, a simultaneous fit of hd1, ba1, and tph1, obtained in a single call to nlme. The separate univariate NLMMs can be equivalently fit simultaneously by stacking the response variables hd, ba, and tph into a single response vector as described in Section 2. Then indicator variables for the three responses can be used to construct the model function by combining $f_1(\cdot)$, $f_2(\cdot)$, and $f_3(\cdot)$. We obtain distinct error variances for the three responses and heteroskedasticity for ba, by utilizing the varComb, varIdent, and varPower classes in nlme to produce the variance function

$$\begin{aligned} \text{var}\left(\varepsilon_{ijk\ell} \mid \mathbf{b}_{i,\ell}^{(1)}, \mathbf{b}_{ij,\ell}^{(2)}\right) &= \sigma^2 \left(\delta_1^2\right)^{1_{\{\ell=2\}}} \left(\delta_2^2\right)^{1_{\{\ell=3\}}} |\hat{\mu}_{ijk2}|^{2\delta_3 1_{\{\ell=2\}}} \\ &= \begin{cases} \sigma^2 & \text{if } \ell = 1, \\ \sigma^2 \delta_1^2 |\hat{\mu}_{ijk2}|^{2\delta_3} & \text{if } \ell = 2, \\ \sigma^2 \delta_2^2 & \text{if } \ell = 3. \end{cases} \end{aligned}$$

Table 2
Log likelihoods for separately fit and simultaneously fit univariate models

Method of fitting	Model	Log likelihood
Separate	hd1	−933.28
Separate	ba1	−1636.84
Separate	tph1	−947.11
Separate	hd1 + ba1 + tph1	−3517.23
Simultaneous	hbt1 = hd1 + ba1 + tph1	−3521.02

Table 3
Log likelihoods and model selection criteria for trivariate models

Model	Log likelihood	AIC	BIC
hbt1	−3521.02	7190.04	7644.14
hbt2	−3360.25	6874.49	7347.01
hbt3	−3160.58	6511.16	7094.13
hbt4	−3054.22	6336.44	7036.00

In addition, we separate the random effects structures by using a block-diagonal form for $\mathbf{\Psi}_1$ and $\mathbf{\Psi}_2$; e.g., we set

$$\mathbf{\Psi}_1 = \begin{pmatrix} \mathrm{var}\big(\mathbf{b}_{i,1}^{(1)}\big) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathrm{var}\big(\mathbf{b}_{i,2}^{(1)}\big) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathrm{var}\big(\mathbf{b}_{i,3}^{(1)}\big) \end{pmatrix}.$$

From Table 2, we see that there is only a slight difference between the log likelihood for the three models fit simultaneously as model hbt1 ($\ell = -3521.02$) and the sum of the log likelihoods for the three separately fit models ($\ell = -3517.23$). This discrepancy is due solely to parameterization effects, as the likelihood for the combined model factors perfectly into components corresponding to the likelihoods of the three univariate models.

Of course, the advantage of fitting the model simultaneously is that we can now specify the variance–covariance matrices $\mathbf{\Lambda}$, $\mathbf{\Psi}_1$, and $\mathbf{\Psi}_2$ to build correlation among the errors and among the random effects associated with distinct responses into the model. We begin by adding contemporaneous correlation among the responses by changing $\mathbf{\Lambda}$ from the identity matrix to an arbitrary symmetric positive-definite matrix. As can be seen in Table 3, the resulting model, hbt2, vastly improves the fit. Next, we allow for correlation among the random effects associated with hd and ba. That is, we set

$$\mathbf{\Psi}_1 = \begin{pmatrix} \mathrm{var}\left\{\begin{pmatrix}\mathbf{b}_{i,1}^{(1)} \\ \mathbf{b}_{i,2}^{(1)}\end{pmatrix}\right\} & \mathbf{0} \\ \mathbf{0} & \mathrm{var}\big(\mathbf{b}_{i,3}^{(1)}\big) \end{pmatrix},$$

$$\mathbf{\Psi}_2 = \begin{pmatrix} \mathrm{var}\left\{\begin{pmatrix}\mathbf{b}_{ij,1}^{(2)} \\ \mathbf{b}_{ij,2}^{(2)}\end{pmatrix}\right\} & \mathbf{0} \\ \mathbf{0} & \mathrm{var}\big(\mathbf{b}_{ij,3}^{(2)}\big) \end{pmatrix},$$

with unstructured upper-right and lower-left blocks. The resulting model, hbt3, again drastically improves the fit.

The final and best fitting model hbt4 assumes that both $\mathbf{\Psi}_1$ and $\mathbf{\Psi}_2$ are completely unstructured in form. That is, in hbt4 we allow all stand-level random effects to be correlated with distinct correlations for all unique pairs, and similarly for all plot-level random effects. As displayed in Table 3, this model fits significantly better than any other trivariate NLMM considered. Therefore, we adopt hbt4 as the model from which we will make future predictions of hd, ba, tph, and volume.

4.1 *Volume Prediction*

The volume model that we propose is based on the assumption that volume is related multiplicatively to unknown powers of hd, ba, and tph. That is, the basic relationship used to model volume (V) is

$$\mathrm{V} = e^{\delta_1}\mathrm{hd}^{\delta_2}\mathrm{ba}^{\delta_3}\mathrm{tph}^{\delta_4},$$

where we assume that the parameters δ_i, $i = 1, 2, 3, 4$, each depend on Age; specifically, we assume the δ_i's are each polynomials in the reciprocal of Age. This volume equation can be linearized by taking logarithms, and we choose to do so to form a linear mixed effects model for volume as follows:

$$\begin{aligned}\log(\mathrm{V}_{ijk}) = {} & \beta_1 + b_{i1}^{(1)} + b_{ij1}^{(2)} + \big(\beta_2 + b_{i2}^{(1)} + b_{ij2}^{(2)}\big)\mathrm{Age}_{ijk}^{-1} \\ & + \big(\beta_3 + b_{i3}^{(1)} + b_{ij3}^{(2)}\big)\mathrm{Age}_{ijk}^{-2} + \beta_4 \log(\mathrm{hd}_{ijk}) \\ & + \beta_5\{\log(\mathrm{hd}_{ijk})/\mathrm{Age}_{ijk}\} + \beta_6\big\{\log(\mathrm{hd}_{ijk})\big/\mathrm{Age}_{ijk}^2\big\} \\ & + \beta_7 \log(\mathrm{ba}_{ijk}) + \big(\beta_8 + b_{i4}^{(1)} + b_{ij4}^{(2)}\big) \\ & \times \{\log(\mathrm{ba}_{ijk})/\mathrm{Age}_{ijk}\} + \beta_9\big\{\log(\mathrm{ba}_{ijk})\big/\mathrm{Age}_{ijk}^2\big\} \\ & + \beta_{10}\log(\mathrm{tph}_{ijk}) + \beta_{11}\mathrm{Chop}_{ij} + \beta_{12}\mathrm{Fert}_{ij} \\ & + \beta_{13}\mathrm{Bed}_{ij} + \varepsilon_{ijk}. \end{aligned} \tag{11}$$

Model (11) was selected based on a model-building process in which we considered various fixed effects and random effects specifications. Multiplicative terms involving log(tph) and linear and quadratic terms in the reciprocal of age were found nonsignificant and dropped from the model. In addition, main effects of all of the treatments were considered, and nonsignificant effects corresponding to Herb and Burn were dropped from the model. Model (11) includes random intercepts at both the plot and site level, and random slopes at both levels for Age^{-1}, Age^{-2}, and log(ba)/Age. The 4×4 variance–covariance matrices $\mathbf{\Psi}_1 = \mathrm{var}(\mathbf{b}_i^{(1)})$ and $\mathbf{\Psi}_2 = \mathrm{var}(\mathbf{b}_{ij}^{(2)})$ were both taken as general (unstructured) positive-definite matrices.

If we represent model (11) simply as $\log(V_{ijk}) = f(\mathrm{Age}_k, \mathbf{y}_{ijk}, \boldsymbol{\beta}, \mathbf{b}_i^{(1)}, \mathbf{b}_{ij}^{(2)}) + \varepsilon_{ijk}$, where $\mathbf{y}_{ijk} = (\mathrm{hd}_{ijk}, \mathrm{ba}_{ijk}, \mathrm{tph}_{ijk})^{\mathrm{T}}$, then we predict log(V) at age Age_h at the plot level as $\widehat{\log(V_{ijh})} = f(\mathrm{Age}_h, \hat{\mathbf{y}}_{ijh}, \hat{\boldsymbol{\beta}}, \hat{\mathbf{b}}_i^{(1)}, \hat{\mathbf{b}}_{ij}^{(2)})$. Here $\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}}_i^{(1)}, \hat{\mathbf{b}}_{ij}^{(2)}$ are estimated and predicted values from model (11), and $\hat{\mathbf{y}}_{ijh}$ is the vector of plot-level predicted values at age Age_h based on the trivariate three-level NLMM hbt4. Stand-level predictions at Age_h are made similarly using $\widehat{\log(V_{ih})} = f(\mathrm{Age}_h, \hat{\mathbf{y}}_{ih}, \hat{\boldsymbol{\beta}}, \hat{\mathbf{b}}_i^{(1)}, \mathbf{0})$, where now $\hat{\mathbf{y}}_{ih}$ is the vector of stand-level predicted values from model hbt4.

Using the delta method, we can obtain an approximate prediction error variance for $\widehat{\log(\mathrm{V}_{ijh})}$ or $\widehat{\log(\mathrm{V}_{ih})}$ that ignores the error introduced by the estimation of model (11). Assuming model (11) is known without error, we have $\log(\mathrm{V}_{ijh}) = \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{w}_{ijh}$, where $\boldsymbol{\alpha} = \{\boldsymbol{\beta}^{\mathrm{T}}, (\mathbf{b}_i^{(1)})^{\mathrm{T}}, (\mathbf{b}_{ij}^{(2)})^{\mathrm{T}}\}^{\mathrm{T}}$ is the combined vector of (assumed known) fixed and random effects, and $\mathbf{w}_{ijh} = (1, \ldots, \mathrm{Bed}_{ij}, 1, \ldots, \log(\mathrm{ba}_{ijh})/\mathrm{Age}_{ijh}, 1, \ldots, \log(\mathrm{ba}_{ijh})/\mathrm{Age}_{ijh})^{\mathrm{T}}$ is the combined vector of fixed effects covariates, stand-level random effects covariates, and plot-level random effects covariates. The covariate vector $\mathbf{w}_{ijh}$ is a function of $\mathbf{y}_{ijh}$, so we write $\mathbf{w}(\mathbf{y}_{ijh}) \equiv \mathbf{w}_{ijh}$. The prediction error is $\log(\mathrm{V}_{ijh}) - \widehat{\log(\mathrm{V}_{ijh})} = \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{w}(\mathbf{y}_{ijh}) - \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{w}(\hat{\mathbf{y}}_{ijh}) = \hat{\boldsymbol{\alpha}}^{\mathrm{T}}\{\mathbf{w}(\mathbf{y}_{ijh}) - \mathbf{w}(\hat{\mathbf{y}}_{ijh})\}$. Taking a first-order Taylor expansion of $\mathbf{w}(\hat{\mathbf{y}}_{ijh})$, the prediction error can be approximated as $\log(\mathrm{V}_{ijh}) - \widehat{\log(\mathrm{V}_{ijh})} \approx \boldsymbol{\alpha}^{\mathrm{T}}\dot{\mathbf{w}}(\hat{\mathbf{y}}_{ijh})(\hat{\mathbf{y}}_{ijh} - \mathbf{y}_{ijh})$, where $\dot{\mathbf{w}}(\hat{\mathbf{y}}_{ijh}) \equiv \frac{\partial \mathbf{w}_{ijh}(\mathbf{y}_{ijh})}{\partial \mathbf{y}_{ijh}^{\mathrm{T}}}|_{\mathbf{y}_{ijh}=\hat{\mathbf{y}}_{ijh}}$, for $\mathbf{y}_{ijh}$ close to $\hat{\mathbf{y}}_{ijh}$. Therefore, an approximate prediction variance for $\widehat{\log(\mathrm{V}_{ijh})}$ can be obtained from the prediction error variance of $\hat{\mathbf{y}}_{ijh}$ via

$$\begin{aligned}&\widehat{\mathrm{var}}\{\log(\mathrm{V}_{ijh}) - \widehat{\log(\mathrm{V}_{ijh})}\}\\ &\quad = \boldsymbol{\alpha}^{\mathrm{T}}\dot{\mathbf{w}}(\hat{\mathbf{y}}_{ijh})\widehat{\mathrm{var}}(\mathbf{y}_{ijh} - \hat{\mathbf{y}}_{ijh})\dot{\mathbf{w}}(\hat{\mathbf{y}}_{ijh})^{\mathrm{T}}\boldsymbol{\alpha}|_{\boldsymbol{\alpha}=\hat{\boldsymbol{\alpha}}}. \qquad (12)\end{aligned}$$

Note that expression (12) is evaluated at $\hat{\boldsymbol{\alpha}}$, the estimated vector of fixed and random effects from model (11), assuming that these are the true values. Similarly, our prediction error variance for the stand-level predictor $\widehat{\log(\mathrm{V}_{ih})}$ is given by

$$\begin{aligned}&\widehat{\mathrm{var}}\{\log(\mathrm{V}_{ih}) - \widehat{\log(\mathrm{V}_{ih})}\}\\ &\quad = \boldsymbol{\alpha}^{\mathrm{T}}\dot{\mathbf{w}}(\hat{\mathbf{y}}_{ih})\widehat{\mathrm{var}}(\mathbf{y}_{ih} - \hat{\mathbf{y}}_{ih})\dot{\mathbf{w}}(\hat{\mathbf{y}}_{ih})^{\mathrm{T}}\boldsymbol{\alpha}\Big|_{\boldsymbol{\alpha}=\left(\hat{\boldsymbol{\beta}}^{\mathrm{T}},\hat{\mathbf{b}}_i^{(1)},\mathbf{0}^{\mathrm{T}}\right)^{\mathrm{T}}}.\end{aligned}$$

Approximate $100(1-\alpha)\%$ prediction limits for V_{ijh} can be obtained by exponentiating the endpoints of the corresponding normal-theory interval for $\log(V_{ijh})$: $\widehat{\log(V_{ijh})} \pm z_{1-\alpha/2}[\widehat{\mathrm{var}}\{\log(\mathrm{V}_{ijh}) - \widehat{\log(\mathrm{V}_{ijh})}\}]^{1/2}$; and similarly for V_{ih}.

4.2 *Illustration*

To illustrate our approach to prediction of future timber volume, we chose to generate predictions at the plot level at age 20, for one of the sites included in the data set. Although age 20 is a relatively young age for volume prediction, we wanted to be able to compare our predictions with observed values, so we chose the oldest age within the age range of our data. Perhaps a more realistic application would be to predict stand-level rather than plot-level volume. However, for the data set at hand, covariates in the model, such as the treatment indicators, vary from plot to plot within each stand. Since any prediction of stand-level volume must be made at a given set of covariate values, and those covariate values will not be constant within the stand, there is no observed (or even potentially observable) value against which a stand-level prediction could be compared. This is a peculiarity of the data set at hand, stemming from the fact that it corresponds to a designed experiment in which plots were assigned different treatments. For data from operational timber stands, it will be more typical that each stand will be grown under a constant set of conditions. In such cases, stand-level predictions will be both more meaningful and more appropriate.

Since a realistic application would involve predicting a value not included in the data set from which the model was fit, we refit our model hbt4 to a new data set consisting of all of the data from the first 15 of the 16 sites in the original data set, plus data from ages 5 and 14 from site 16. From the resulting fitted model, we predicted hd, ba, and tph at age 20 for site 16. We also refit model (11) for volume to the reduced data set, and then generated volume predictions and prediction intervals for volume at age 20 in stand 16 based on the predicted values of hd, ba, and tph and their prediction error variance–covariance. The resulting predictions and 95% prediction intervals are displayed in Table 4. Also given are the observed values for all four variables, and a root mean square prediction error (RMSPE) for each variable.

Table 4

Observed and predicted values for hd, ba, tph, and volume at age 20 *in stand* 16, *with* 95% *prediction intervals in parentheses. Note that tph is in hundreds of trees per hectare and volume is total volume outside of bark. RMSPE is root mean squared prediction error.*

Plot	hd (m)	ba (m^2/ha)	tph	Volume (m^3/ha)	hd	ba	tph	Volume
1	10.62	10.37	10.30	56.76	9.44	10.03	10.70	50.16
	(9.39,11.86)	(8.24,12.51)	(9.04,11.57)	(49.36,65.26)				
2	15.60	25.40	12.39	185.67	14.94	24.28	12.13	170.51
	(14.35,16.85)	(22.53,28.27)	(11.10,13.69)	(158.15,217.97)				
3	14.16	16.80	10.12	113.64	13.45	15.99	10.28	102.35
	(12.93,15.40)	(14.34,19.26)	(8.89,11.35)	(95.01,135.93)				
4	17.47	29.83	11.97	240.98	17.25	28.80	11.39	229.58
	(16.23,18.71)	(26.93,32.74)	(10.69,13.24)	(209.62,277.04)				
5	12.32	13.06	10.25	78.00	11.77	11.15	9.81	62.51
	(11.09,13.56)	(10.73,15.40)	(9.00,11.50)	(67.01,90.79)				
6	16.33	24.80	11.59	190.25	15.61	24.42	11.32	181.47
	(15.06,17.59)	(21.97,27.63)	(10.32,12.86)	(164.58,219.93)				
7	13.97	20.99	12.96	140.99	12.43	19.98	13.12	121.06
	(12.75,15.19)	(18.51,23.47)	(11.61,14.30)	(122.36,162.45)				
8	17.85	33.07	12.18	269.49	17.58	34.48	12.28	275.81
	(16.64,19.05)	(30.24,35.90)	(10.89,13.58)	(240.21,302.33)				
9	16.24	25.83	10.71	195.43	15.65	24.21	10.72	176.30
	(15.00,17.47)	(23.14,28.52)	(9.45,11.98)	(171.50,222.70)				
10	17.95	34.49	12.72	282.33	17.10	34.80	12.45	278.64
	(16.76,18.42)	(31.75,37.23)	(11.39,14.05)	(253.68,314.23)				
11	17.24	30.50	12.96	240.79	16.58	30.39	12.58	231.43
	(16.05,18.42)	(27.96,33.05)	(11.63,14.29)	(214.90,269.81)				
12	17.54	31.12	12.25	246.95	17.52	31.49	11.98	249.36
	(16.35,18.73)	(28.44,33.81)	(10.94,13.56)	(221.73,275.04)				
RMSPE	0.7721	1.032	0.3130	12.10				

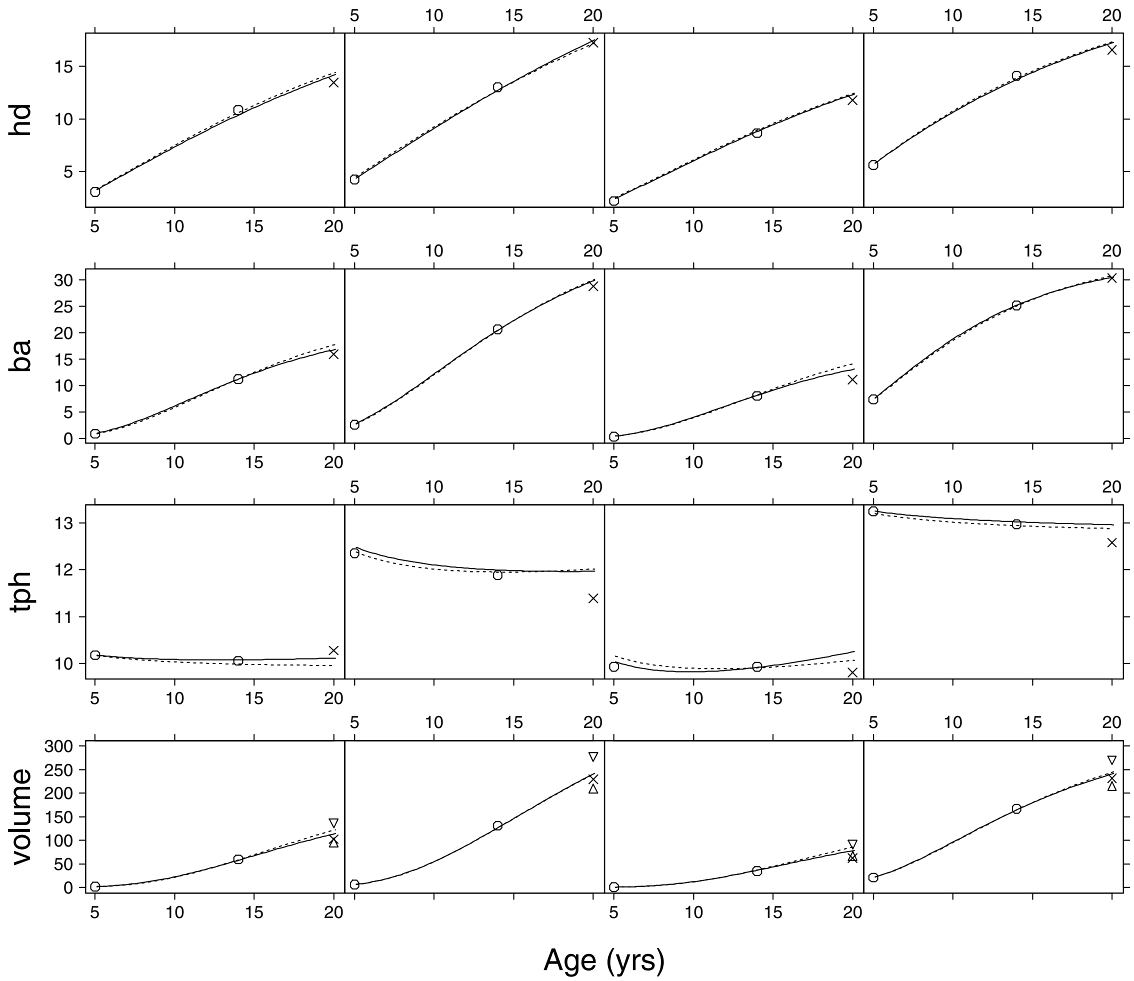

Figure 1. Observed data (circles) and hold-out data ("x"s) and predicted growth curves for plots 3, 4, 5, and 11 in stand 16. Solid lines show predictions based on the multivariate model hbt4, and dashed lines show predictions based on univariate modeling of hd, ba, and tph. For volume, 95% prediction limits at age 20 are given (triangles).

Some of these results are also portrayed graphically in Figure 1. Figure 1 shows the observed data (circles) at ages 5 and 14 and the held-out responses at age 20 (plotted with "x"s) for hd, ba, tph, and volume for four of the plots in site 16 (plots 3, 4, 5, and 11, from left to right). In the panels for hd, ba and tph, solid lines plot the predicted responses based on the multivariate model hbt4 and dotted lines plot the predicted curves based on corresponding univariate models fit one at a time. Solid lines in the bottom panel are predicted volumes based on substituting the predicted values from hbt4 into (11). Dotted lines in this panel are predicted volumes based on substituting the predicted volumes from the three univariate models. Also included in the bottom panel are the 95% prediction limits for volume at age 20 (plotted with triangles) based on model (11) and the age 20 predicted values from hbt4.

Figure 1 demonstrates the flexibility of the NLMMs used here. The inclusion of plot-specific random effects allows the model to fit a wide variety of growth curve forms among the plots. In addition, Figure 1 and Table 4 demonstrate that the future predictions are quite accurate. From Figure 1, predictions based upon a univariate approach appear to be quite similar to those from the multivariate approach developed here. The RMSPEs based on the univariate approach are 0.7189, 1.399, 0.3218, and 14.83 for hd, ba, tph, and volume, respectively. Surprisingly, the univariate model produces slightly more accurate predictions of hd than does the multivariate model. However, the multivariate approach leads to an 18.4% reduction in RMSPE for the variable of primary interest, volume. More importantly perhaps, the univariate approach does not yield prediction limits for volume. With separate models for predicting hd, ba, and tph at a future age of interest, the covariances between these predictions cannot be estimated. That is, the univariate approach does not lead to an estimator of $\text{var}(\mathbf{y}_{ijh} - \hat{\mathbf{y}}_{ijh})$ in formula (12).

The predictions in Table 4 and Figure 1 are based upon (7). For comparison, we also generated predictions from the multivariate model hbt4 based on the simple plug-in predictor. In this prediction scenario, we are predicting values of the response variables at a future age based on observations of those variables from the same plot and stand. In this case, the model implies that $\mathbf{y}_h$ and $\mathbf{y}_s$ are correlated only through shared plot and stand random effects. That is, $\text{cov}(\boldsymbol{\varepsilon}_h, \boldsymbol{\varepsilon}_s) = \mathbf{0}$ here. Therefore, we expect little difference between predictor (7) and the plug-in predictor. Indeed, the RMSPEs for the plug-in predictor are nearly identical to those given in Table 4 (0.7726, 1.032, and 0.3130 for hd, ba, and tph, respectively).

However, to illustrate that equation (7) can improve upon the plug-in predictor when $\text{cov}(\boldsymbol{\varepsilon}_h, \boldsymbol{\varepsilon}_s) \neq \mathbf{0}$, we refit the model to data that included ages 5, 14, and 20 for hd and tph in stand 16 but only ages 5 and 14 for ba in that stand. The RMSPE for predicting ba at age 20 for the 12 plots in this stand was 0.8296 using the plug-in predictor and was 0.8009 using (7). As another example, we fit the univariate three-level NLMM corresponding to the hd part of hbt4, but we added a serial correlation structure to model $\text{corr}(\varepsilon_{ijk}, \varepsilon_{ijk'}), k \neq k'$. Specifically, we assumed the banded correlation structure of order 3 (corBand(ord=3) in the nlme software), which allows distinct correlations for all pairs $(\varepsilon_{ijk}, \varepsilon_{ijk'})$ where $|k - k'| < 3$. Fitting this model to dominant height data from ages 5 and 14 only in stand 16, the RMSPEs were 0.6060 and 0.5682 for the plug-in predictor and predictor (7), respectively, for predicting hd at age 20 for this stand. Thus, when $\text{cov}(\boldsymbol{\varepsilon}_h, \boldsymbol{\varepsilon}_s) \neq \mathbf{0}$ it does appear that (7) improves on the plug-in predictor. A more systematic and comprehensive comparison of these predictors based on simulation results will be pursued elsewhere.

5. Summary

We have proposed to extend the methodology of ML-NLMMs from the univariate case in order to simultaneously model a set of plot-level timber growth variables from which merchantable yields can be determined. We have presented methodology for fitting multivariate ML-NLMMs and for producing predictions and prediction intervals from those models. We have described how these predictions and their error variances can then be used, in conjunction with a whole stand yield equation, to produce predictions and prediction intervals for timber yields at future ages of interest. This methodology can be viewed as an alternative to the use of simultaneous systems of nonlinear equations for forest growth and yield modeling (Borders, 1989). The primary advantages of our methodology are (1) the ability to model and predict timber growth and yield at the plot, stand, and population level, and (2) the availability of a prediction variance estimator, which allows for the quantification of the uncertainty in yield predictions.

Acknowledgements

The authors wish to extend their sincere thanks to José Pinheiro for his help with questions concerning his nlme software, and the Plantation Management Research Cooperative of the Daniel B. Warnell School of Forest Resources at the University of Georgia for providing data on the slash pine site preparation study discussed in Section 4. In addition, the authors thank The Timber Company for its support of this research.

Résumé

Les modèles mixtes non linéaires sont devenus des outils importants pour modéliser des croissances et des rendements en sylviculture. Les applications sont restreintes actuellement à la modélisation de variables simples de croissance telles que la hauteurs des arbres et le volume de bois. Nous proposons ici des modèles multivariables multi-niveaux non linéaires mixtes pour décrire simultanément plusieurs caractéristiques de relevés de bois. Nous décrivons comment de tels modèles peuvent être utilisés pour fournir des prédictions de rendements futurs de bois. La classe des modèles et les méthodes d'estimation et de prédiction sont développés et illustrées à partir de données d'une étude de l'université de Georgia portant sur les effets de différentes méthodes de préparation de sites sur la croissance de pins (*pinus elliottii Engelm.*)

References

Borders, B. E. (1989). Systems of equations in forest stand modeling. *Forest Science* **35,** 548–556.

Christensen, R. (1996). *Plane Answers to Complex Questions: The Theory of Linear Models,* 2nd edition. New York: Springer-Verlag.

Clutter, J. L., Fortson, F. C., Pienaar, L. V., Brister, G. H., and Bailey, R. L. (1983). *Timber Management: A Quantitative Approach.* New York: Wiley.

Davidian, M. and Giltinan, D. M. (1995). *Nonlinear Models for Repeated Measurement Data.* London: Chapman & Hall.

Lindstrom, M. J. and Bates, D. M. (1990). Nonlinear mixed-effects models for repeated measures data. *Biometrics* **46,** 673–687.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models,* 2nd edition. London: Chapman & Hall.

Pinheiro, J. C. and Bates, D. M. (2000). *Mixed-Effects Models in S and S-PLUS.* New York: Springer.

Ratkowsky, D. A. (1990). *Handbook of Nonlinear Regression Models.* New York: Marcel Dekker.

Shiver, B. D. and Harrison, W. M. (2000). *Slash pine site preparation study: Age 20 results.* Plantation Management Research Cooperative Technical Report 2000–4, Daniel B. Warnell School of Forest Resources, University of Georgia.

Vonesh, E. F. and Chinchilli, V. M. (1997). *Linear and Nonlinear Models for the Analysis of Repeated Measures.* New York: Marcel Dekker.

Wolfinger, R. D. (1993). Laplace's approximation for nonlinear mixed models. *Biometrika* **80,** 791–795.

Wolfinger, R. D. and Lin, X. (1997). Two Taylor-series approximation methods for nonlinear mixed models. *Computational Statistics and Data Analysis* **25,** 465–490.

Received January 2003. *Revised May* 2003.
Accepted July 2003.