Comparing Traditional Longitudinal Modeling Strategies of Forestry with Mixed-Effects Models: Restrictions in Model Formulation

Comparing Traditional Longitudinal Modeling Strategies of Forestry with Mixed-Effects Models: Restrictions in Model Formulation

Brock Stewart, Lewis Jordan, and Rongxia Li

Abstract: Generalized algebraic difference approach and site index models, which are commonly fitted to repeated-measures data in forestry, are formulated by specifying that one or more regression parameters depend on a single, common subject-specific quantity. Their formulation leads to shortcomings not present with "effects" types of formulations, which involve specifying each subject-specific regression parameter as depending on at least one local quantity that no other parameters depend on. We give conceptual illustrations as well as empirical examples using published models. FOR. SCI. 57(5):408–415.

Keywords: longitudinal models, height–age, basal area, wood properties, tree taper

LONGITUDINAL DATA REPRESENTING repeated measurements on multiple individuals (trees, plots, and stands) are now often used to calibrate forestry models. Examples include basal area over time (Barrio-Anta et al. 2006), tree taper (Strub et al. 2005), and wood specific gravity (Phillips 2002). To simplify the presentation, we focus on height–age models, although our main points apply more broadly.

Models for longitudinal data can be represented by

$$y_{ij} = f(x_{ij}, \boldsymbol{\beta}_i) + \varepsilon_{ij}, \quad (1)$$

where y_{ij} is the jth measurement of the response variable (height) on the ith subject, x_{ij} is the corresponding independent variable (age), $\boldsymbol{\beta}_i$ is a vector of p parameters, k of which are subject-specific, f is a linear or nonlinear function, and ε_{ij} is mean-zero random error. In general terms, $\boldsymbol{\beta}$ is made to take subject-specific values by being specified as a function of one or more subject-specific quantities. For example, in mixed-effects models a specification such as (Lindstrom and Bates 1990)

$$\boldsymbol{\beta}_i = \mathbf{A}_i\boldsymbol{\gamma} + \mathbf{B}_i\boldsymbol{\lambda}_i \quad (2)$$

is often made, where $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$ are vectors of global parameters and q random effects, respectively, and $\mathbf{A}_i$ and $\mathbf{B}_i$ are design matrices. A more general notation (Davidian and Giltinan 1995) is

$$\boldsymbol{\beta}_i = \boldsymbol{d}_i(\boldsymbol{\gamma}_i\ \boldsymbol{\lambda}_i), \quad (3)$$

where $\boldsymbol{d}_i$ is a p-dimensional function. However, mixed-effects models are not the only types of longitudinal models: $\boldsymbol{\lambda}$ could be a q vector of fixed effects or other subject-specific quantities.

This general notation subsumes the traditional height–age-site modeling method. In site index (SI) models $k \geq 1$, $q = 1$, and λ_i is height of the ith subject at a certain age. For example, Payandeh (1974) paired the Chapman-Richards function

$$f(x_{ij}, \boldsymbol{\beta}_i) = \beta_1(1 - e^{-\beta_2 x_{ij}})^{\beta_3} \quad (4)$$

with the formulation

$$\boldsymbol{\beta}_i = (\gamma_1\lambda_i^{\gamma_2},\ \gamma_3,\ \gamma_4\lambda_i^{\gamma_5})^{\mathrm{T}}, \quad (5)$$

where λ_i is height at age 50. Likewise, Cieszewski and Bailey (2000) suggested a class of models (CB00) having $k \geq 1$, $q = 1$, but λ_i is a fixed effect. For example, Dieguez-Aranda et al. (2006) developed a CB00 model from the Chapman-Richards function after specifying that

$$\boldsymbol{\beta}_i = (e^{\lambda_i},\ \gamma_1,\ \gamma_2 + \gamma_3/\lambda_i)^{\mathrm{T}}. \quad (6)$$

Bailey and Clutter (1974) suggested a class of models (BC74) with $q = 1$, λ_i is a fixed effect, but $k = 1$.

A common aspect of SI, BC74, and CB00 models is that they have a single local quantity regardless of the number of varying base parameters. An alternative formulation commonly used with mixed models is to assign one local quantity to each varying base parameter. Continuing with the Chapman-Richards examples, we might specify

$$\boldsymbol{\beta}_i = (\gamma_1 + \lambda_{1i},\ \gamma_2,\ \gamma_3 + \lambda_{2i})^{\mathrm{T}} \quad (7)$$

although the linear parameterization is not required. In general, we call these *effects* formulations.

In this article, we illustrate how SI, BC74, and CB00 formulations lead to inflexibility and bias not present with effects (random or fixed) types of formulations. Although estimation techniques have been debated (Wang et al. 2004, 2007, Cieszewski and Strub 2007, Ni and Zhang 2007), model formulation is more fundamental: parameter estimation associates parameter values to each data trend; a model's formulation can limit the available parameter values. After giving a brief background on traditional methods, we discuss useful and

Brock Stewart, University of Georgia, Warnell School of Forestry and Natural Resources, 180 East Green Street, Athens, GA 30605—bs@warnell.uga.edu. Lewis Jordan, Weyerhaeuser Company—lewis.jordan@weyerhaeuser.com. Rongxia Li, University of Maine—rongxia_li@umit.maine.edu.

Acknowledgments: We are grateful to Harold Burkhart for providing the data set examined in this work. We also thank Robert L. Bailey, Bruce Borders, Daniel B. Hall, Mike Strub, and the editors and reviewers for useful comments on earlier versions of the manuscript, which was in need of much improvement.

Manuscript received August 3, 2009, accepted March 17, 2011

a priori types of parameter value restrictions. We then provide an illustration using model-fitting results on empirical data.

SI, BC74, and CB00 Models

The idea for SI models was conceived in the 1800s and early 1900s as an alternative to using observed volume as a localizing factor in yield tables before the advent of computers and introduction of statistical methods into the forestry sciences (see Roth 1916, Frothingham 1918). The idea was carried forward to the computer age (Lundgren and Dolid 1970, Beck 1971, Payandeh 1974) when hand-drawn curves were being replaced with nonlinear functions. One difficulty encountered with SI models is the question of how to obtain site index when height at the model's base age has not been measured.

BC74 and CB00 models are typically fitted under an expected value parameterization (EVP) (Ratkowsky 1990, p. 31) wherein λ_i is replaced with its algebraic solution given a hypothetical data point (y_{i0}, x_{i0}) and the global parameters; y_{i0} is then treated as the fixed effect. This method was originally proposed by Bailey and Clutter (1974) for $k = 1$ as an alternative to SI models that did not require height measurements at a common age; it was extended to $k \geq 1$ by Cieszewski and Bailey (2000). The algebraic parameterization of BC74 and CB00 models has led them to be called algebraic difference approach (ADA) and generalized ADA (GADA) models, a terminology we avoid for reasons listed later in the Discussion section.

Informed versus A Priori Parameter Space Restrictions

A case can be made for all base parameters being local in longitudinal models; still, one or more could vary across subjects to such a small degree that making them global leads to little loss in information relative to practical gains (Davidian and Giltinan 1995, p. 149). Zhao et al. (2005) suggested data analytic techniques for deciding which, if any, base parameters should be global in mixed-effects models. In addition, we may restrict global and/or local parameter values to only those meaningful for the given application. For example, we may restrict the asymptote and/or rate parameters of a height–age model to be positive.

Because SI and CB00 formulations specify that multiple βs are functions of the same λ, their local values are functionally related. For example, with formulation 6 we must have that $\beta_{3i} = \gamma_2 + \gamma_3/\ln(\beta_{1i})$ for every ith subject. This is easily verified by algebraically solving for λ_i in $\beta_{1i} = \exp(\lambda_i)$ and inserting the result into $\beta_{3i} = \gamma_2 + \gamma_3/\lambda_i$. In contrast, effects formulation 7 allows any value of (β_1, β_3) in R^2 (two-dimensional set of real numbers).

Figure 1 illustrates possible local parameter spaces (P) of the Chapman-Richards function. An effects formulation with $k = 3$ has $P = R^3$. Effects formulation 7 gives a plane as P, for example,

$$\beta_{1i} = \gamma_1 + \lambda_{1i} \quad \text{and} \quad \beta_2 = 0.02912 \quad \text{and} \quad \beta_{3i} = \gamma_3 + \lambda_{2i}. \tag{8}$$

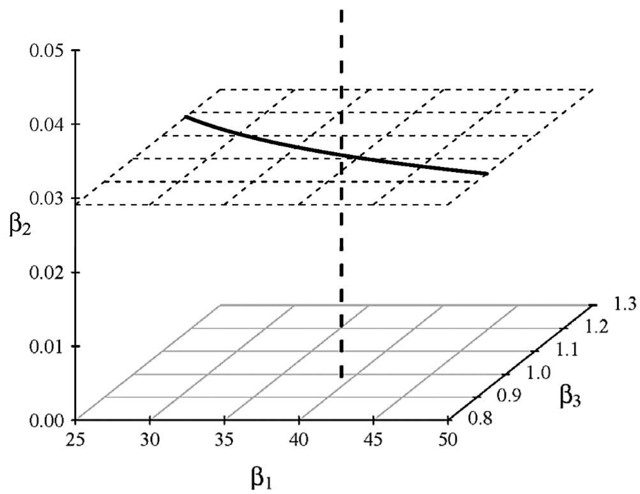

Figure 1. Local parameter spaces of different formulations of a three-parameter base function.

P is reduced further by the CB00 formulation used by Dieguez-Aranda et al. (2006), which gives

$$\beta_{1i} = \exp(\lambda_i) \quad \text{and} \quad \beta_2 = 0.02912 \quad \text{and} \quad \beta_{3i} = -0.2142 + 4.494/\lambda_i \tag{9}$$

and is represented by the curve in Figure 1; the numerical global parameter values are those reported by Dieguez-Aranda et al. The BC74 model of Dieguez-Aranda et al. has P equal to a straight line in R^3, defined by

$$\beta_1 = 39.13 \quad \text{and} \quad \beta_{2i} = \lambda_i \quad \text{and} \quad \beta_3 = 0.9916. \tag{10}$$

In Table 1 we describe local parameter spaces of several formulations of a three-parameter function.

To translate how restricted parameter spaces affect the set of available height–age curves, consider Figure 2. If the Schumacher function

$$f(x_{ij}, \boldsymbol{\beta}_i) = \beta_1 e^{-\beta_2 x_{ij}^{-1}} \tag{11}$$

is given an effects formulation, then the local parameter space is R^2, which could be reduced to a two-dimensional subset of R^2 by, e.g., placing bounds on the local values of β_1 and β_2. A SI or CB00 formulation (depending on whether λ_i is treated as site index or a fixed effect) could be

$$\boldsymbol{\beta}_i = (\lambda_i, \gamma/\lambda_i)^{\mathrm{T}} \tag{12}$$

Clearly, the only values of β_1 and β_2 available for any ith subject are those such that $\beta_{2i} = \gamma/\beta_{1i}$. For example, using a round number, if the estimated value of γ turned out to be 100, then every ith subject's fitted value of $\beta_i = (\beta_{1i}, \beta_{2i})^{\mathrm{T}}$ must correspond to the coordinates of a point on the bold line in Figure 2A. Suppose that three hypothetical subjects' growth patterns are best described by the Schumacher function with parameter values corresponding to the coordinates of points I, II, and III in Figure 2A, respectively; the best curves for these subjects are then curves I, II, and III in Figure 2B. As an illustration, we used 26 points from $x = 5$ to $x = 30$ on line II in Figure 2B as pseudodata to find the solution to $\min_\lambda \|\boldsymbol{y} - \lambda \exp(-100(\lambda \boldsymbol{x})^{-1})\|^2$, where $\boldsymbol{x} = (5, 6, \ldots, 29, 30)^{\mathrm{T}}$ and $\boldsymbol{y} = (25 \exp(-7/5), \ldots, 25 \exp(-7/30))^{\mathrm{T}}$

Table 1. Flexibility as measured by the dimensionality of local parameter spaces: example using any three-parameter regression function.

Formulation of local base parameters and Example formulation*	Dimensionality of local parameter space	No. local base parameters	Formulation type	Description of local parameter space
Three functionally independent				
$\beta_{1i} = d_1(\lambda_{1i}, \boldsymbol{\gamma})$	3	3	Effects	all of R^3
$\beta_{2i} = d_2(\lambda_{2i}, \boldsymbol{\gamma})$				
$\beta_{3i} = d_3(\lambda_{3i}, \boldsymbol{\gamma})$				
Two functionally independent				
$\beta_{1i} = d_1(\lambda_{1i}, \boldsymbol{\gamma})$	2	2	Effects	Plane in R^3 perpendicular to β_3 axis
$\beta_{2i} = d_2(\lambda_{2i}, \boldsymbol{\gamma})$				
$\beta_3 = \gamma_3$				
Two functionally dependent, one varying functionally independently				
$\beta_{1i} = d_1(\lambda_{1i}, \boldsymbol{\gamma})$	2	3	NA	Sheet formed by expanding a curve or line on the $\beta_1 \times \beta_2$ plane perpendicularly to the β_3 axis
$\beta_{2i} = d_2(\lambda_{1i}, \boldsymbol{\gamma})$				
$\beta_{3i} = d_3(\lambda_{2i}, \boldsymbol{\gamma})$				
One local base parameter				
$\beta_{1i} = d_1(\lambda_i, \boldsymbol{\gamma})$	1	1	Effects, BC74, SI	Straight line in R^3 that is perpendicular to both the β_2 and β_3 axes and passes through the $\beta_2 \times \beta_3$ plane at $(\gamma_2, \gamma_3)^T$
$\beta_2 = \gamma_2$				
$\beta_3 = \gamma_3$				
Two functionally dependent				
$\beta_{1i} = d_1(\lambda_i, \boldsymbol{\gamma})$	1	2	CB00, SI	Curve or line on a plane in R^3; the plane is perpendicular to β_3 axis
$\beta_{2i} = d_2(\lambda_i, \boldsymbol{\gamma})$				
$\beta_3 = \gamma_3$				
Three functionally dependent				
$\beta_{1i} = d_1(\lambda_i, \boldsymbol{\gamma})$	1	3	CB00, SI	Curve or line in R^3
$\beta_{2i} = d_2(\lambda_i, \boldsymbol{\gamma})$				
$\beta_{3i} = d_3(\lambda_i, \boldsymbol{\gamma})$				

* General notation: $f(x, \boldsymbol{\beta})$, regression function; $\boldsymbol{\beta}$, p-dimensional vector of base parameters; f is formulated via $\boldsymbol{\beta}_i = \boldsymbol{d}(\lambda_i, \gamma) = (d_1(\boldsymbol{\lambda}_i, \boldsymbol{\gamma}), \ldots, d_p(\boldsymbol{\lambda}_i, \boldsymbol{\gamma}))^T$, where each sth base parameter is replaced with a function d_s of a vector of local quantities $\boldsymbol{\lambda}_i$ and a vector of global parameters $\boldsymbol{\gamma}$.

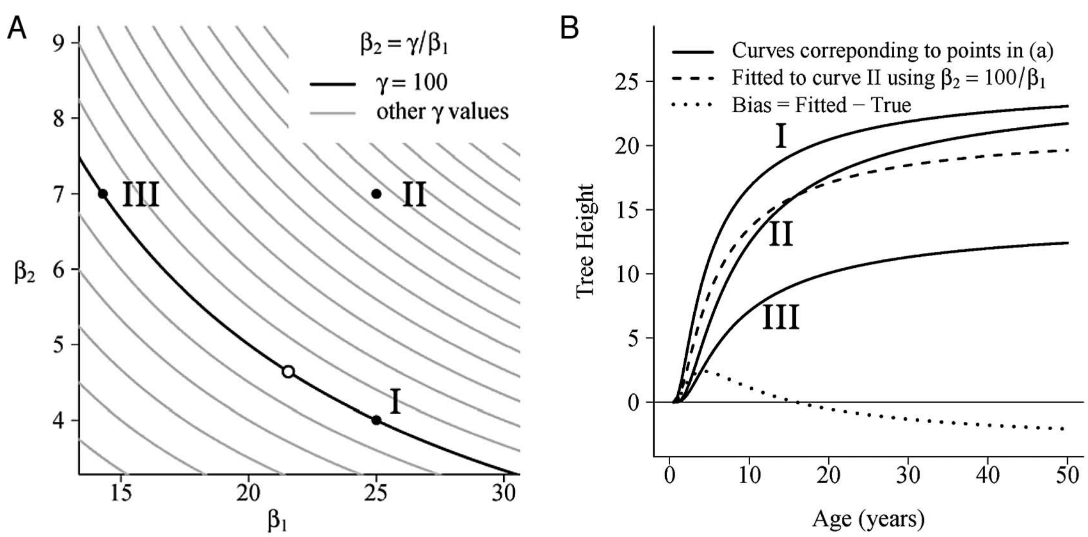

Figure 2. A. Curves: one-dimensional parameter spaces defined by $\beta_2 = \gamma/\beta_1$ for different values of γ; coordinates of points I, II, and III: "true" or "best" values for three hypothetical subjects; coordinates of ○: obtained by fitting Schumacher's function with $\beta_2 = 100/\beta_1$ to points on subject II's true curve. B. The true curves of the three hypothetical subjects, the fitted curve obtained as just described, and bias of the fitted curve.

are 26 error-free height values for hypothetical subject II. The resulting fitted curve and bias are drawn in Figure 2B.

The three take-home points here are the following: SI, BC74, and CB00 models have a priori one-dimensional parameter spaces; this disallows a great number of height–age curves (inflexibility); and the remaining curves that are available must systematically deviate from true trends unless every ith subject's best parameter values lie on the one-dimensional parameter space (bias).

Empirical Model-Fitting Example

Data

We use the data examined by Dieguez-Aranda et al. (2006). The response variable is the arithmetic average of the heights of dominant and codominant trees that were always dominant or codominant over the life of the stand. There are 1,063 observations, which were taken every 3 years at 186 locations as part of a long-term thinning experiment. More information on these data and the experiment they come from can be found in Dieguez-Aranda et al. (2006) and Burkhart et al. (1985).

Methods

We fitted the four models considered by Dieguez-Aranda et al. (2006): one BC74 and one CB00 formulation each of the Chapman-Richards (4) and log-logistic

$$f(x_{ij}, \boldsymbol{\beta}_i) = \frac{\beta_1}{1 + \beta_2 x_{if}^{-\beta_3}} \tag{13}$$

functions (Table 2). In addition, we fitted three CB00 formulations of the Chapman-Richards function presented by Cieszewski and Strub (2008), as well as another BC74 formulation and an effects-formulated version of both base functions (Table 2). We used no heterogeneity or correlation parameters in these models: $\mathrm{Var}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}$, where $\boldsymbol{\varepsilon}$ is a vector of all error terms, $\sigma^2 > 0$ is error variance, and $\mathbf{I}$ is an identity matrix. We fitted the effects models by treating the λs as normally distributed random effects with a diagonal covariance structure using the maximum likelihood method of Pinheiro and Bates (2000). We used least squares to fit each BC74 and CB00 model both in non-EVP form, treating each λ_i as a fixed effect, and in EVP form, treating each y_{i0} as a fixed effect. We calculated the sum of squared residual (SSR) and mean absolute residual (MAR) of all fitted models. All analysis was conducted using the R statistical package (R Development Core Team 2009).

Results and Discussion

Although we could reproduce the descriptive statistics listed in Table 1 of Dieguez-Aranda et al. (2006), we somehow obtained slightly lower SSR when fitting their four models to their data; this is probably mostly due to the autocorrelation structure they used and possibly to differences in software.

There was no difference, to all available decimal places, in fitted parameter values, SSR, or MAR between corresponding EVP and non-EVP versions of the BC74 and CB00 models. Among the models formulated from the Chapman-Richards function, those having a one-dimensional local parameter space, CB00 models 15–18 and BC74 models 19 and 20, resulted in twice the SSR and 38% larger MAR than random effects model 14 (Table 2). Likewise, for log-logistic formulations, CB00 model 22 and BC74 models 23 and 24 resulted in 1.5 times the SSR and 22% larger MAR than random effects model 21.

The fitted local parameter values of models 14–19 are plotted in Figure 3: lines correspond to local parameter spaces of BC74 model 19 and CB00 models 15–18, whereas symbols represent parameter estimates. This figure illustrates how effects formulations allow local parameter values to vary in multiple dimensions, whereas estimates fall on a straight line with BC74 models and a straight or curvaceous line with CB00 models. In addition, all four CB00 models have local parameter spaces that almost overlap within the range of parameter estimates, leading to almost identical fits.

We attribute the reduced fits of the BC74 and CB00 models to their restricted local parameter spaces or, equivalently, to their restricted families of curves, as illustrated in Figure 4. There we plotted the data from two subjects (remeasurement field plots) and superimposed fitted subject-specific curves. The fitted BC74 and CB00 curves appear to bisect each observed trend rather than mimic it. Systematic deviations, such as bisections, induce patterns in residuals that could lead a modeler to incorporate unnecessary autocorrelation error structures that might indicate better fits than are actually occurring; many CB00 model users have decided to use autocorrelation error structures, including Dieguez-Aranda et al. (2006).

Figure 4 also illustrates the similarity in the flexibility of the BC74 and BC00 models. The dotted curves were obtained by replacing λ with ±1, 2, and 3 times (λ_{136}–λ_{186}) the difference between the fitted λ-values for these two subjects. Because none of these curves follow the observed trends, the parameter estimation routine has essentially picked the curves "balancing" the bias. In general, these types of bias could be difficult to detect using numerical residual analysis techniques when data series are short or

Table 2. Formulations of the Chapman-Richards and log-logistic functions considered here, with fit statistics.

Formulation type	Parameter of base function			Fit statistics		Equation
	β_1	β_2	β_3	SSR	MAR	
Formulations of the Chapman-Richards function						
Effects*	$\gamma_1 + \lambda_{1i}$	γ_2	$\gamma_3 + \lambda_{2i}$	155.9	0.2970	14
CB00	$\exp(\lambda_i)$	γ_1	$\gamma_2 + \gamma_3 \lambda_i$	297.7	0.4098	15
			$\gamma_2 + \gamma_3/\lambda_i$	297.9	0.4100	16†
			$\gamma_2+\gamma_3\lambda_i+\gamma_4\lambda_i^2$	297.2	0.4093	17
			$\gamma_2 (\gamma_3 + \lambda_i)^{-1}$	298.0	0.4100	18
BC74	λ_i	γ_1	γ_2	306.4	0.4125	19
	γ_1	λ_i	γ_2	296.1	0.4097	20†
Formulations of the log-logistic function						
Effects*	$\gamma_1 + \lambda_{1i}$	$\gamma_2 + \lambda_{2i}$	$\gamma_3 + \lambda_{3i}$	201.0	0.3352	21
CB00	$\gamma_1 + \lambda_i$	γ_2/λ_i	γ_3	293.0	0.4073	22†‡
BC74	λ_i	γ_1	γ_2	303.7	0.4108	23
	γ_1	λ_i	γ_2	294.8	0.4089	24†

* Fitted as random effects models; see text for more details.
† Models considered by Dieguez-Aranda et al. (2006) for these data.
‡ Final model chosen by Dieguez-Aranda et al. (2006) for these data.

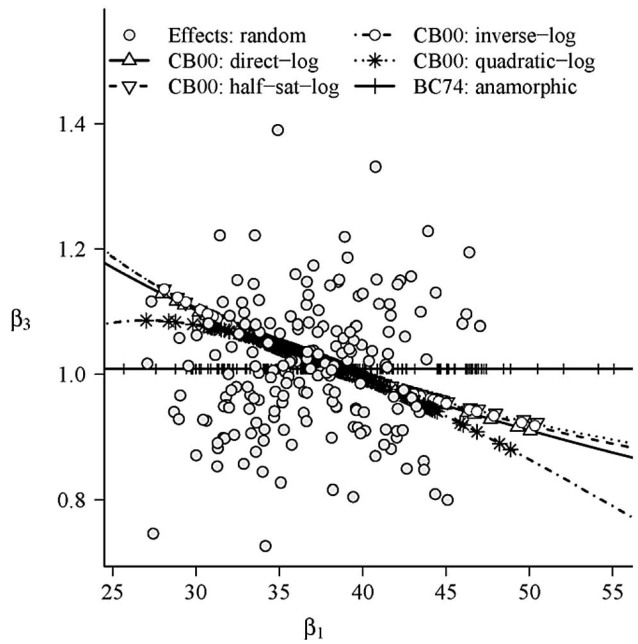

Figure 3. Local parameter estimates (symbols) resulting from fitting different formulations of the Chapman-Richards function to the data of Dieguez et al. (2006); lines correspond to local parameter spaces.

when masked by autocorrelation structures accounting for systematic trends in residuals induced by a systematically deviating curve. In addition, with moderate- to large-sized longitudinal data sets, when all subjects' residuals are viewed graphically together (Dieguez-Aranda et al. 2006, p. 270) it may be difficult to detect subject-level bias.

Such systematically biased fitted trends generally hold important implications for conclusions made from these types of models, including but not limited to management prescriptions, silvicultural investigations (Weiskittel et al. 2009), climatic considerations (Bravo-Oviedo et al. 2008, p. 2356), and carbon sequestration (James et al. 2007). In the current example, the heights on plot 186 are lower than those for plot 136 at earlier ages and appear to converge and or surpass those of plot 136 by about age 30 years. The effects-formulated model mimics these patterns closely. Conversely, the fitted trends from the one-dimensional models are closer together than the data for early ages and do not converge.

Discussion

Model formulation is an important aspect of longitudinal regression modeling. It is easy to take formulations for granted without considering the implied consequences. This stage of model development should be considered at least as important as parameter estimation because bias and inflexibility in a model's family of curves cannot be mitigated when parameters are estimated.

Some authors have considered the major distinction between effects and CB00 models to be treatment of λs as fixed or random effects (Wang et al. 2008, p. 2666–2667). However, the dimensionality of their local parameter spaces is more fundamental and has heretofore been largely overlooked. Furthermore, we are only aware of functional parameter relationships being given a cursory mention (Krumland and Eng 2005, p. 28) or even lauded (Cieszewski 2002, 2003) in the literature. Although CB00 and BC74 models are "fixed-effects models" in one sense, because each λ_i is a fixed effect, it is important to consider their one-dimensional parameter spaces, and we thus distinguish between them and effects-formulated fixed-effects models.

The act of replacing multiple regression parameters with functions of a single, common fixed effect is the essence of the CB00 method (Cieszewski 2003, p. 543); the same could be said of SI models, the only difference being that the local quantity is site index. CB00 models with multiple varying base parameters have repeatedly been called more flexible in the literature than those with a single varying base parameter (Cieszewski 2002, p. 8, Cieszewski and Bailey 2000, p. 125, and re-cited in several articles referencing that work). However, it is clear from Figures 1–4 and Table 1 that this is not true: all models with one local quantity have a one-dimensional parameter space, which is directly related to a model's flexibility (range of available curves).

We propose that a more rigorous definition of flexibility be used: size of local parameter space. For example, consider three local parameter spaces: $P_1 = R^k$; P_2, a k-dimensional subset of R^k; and $P_3 \subset R^m$, where $m < k$. It seems reasonable to call a model based on P_1 more flexible than one based on P_2 and to call a model based on P_3 the least flexible. Because parameter spaces like P_2 would be chosen for reasons such as restricting parameter estimates to meaningful values, in general, we can say a model's *degree of flexibility* equals the dimensionality of its parameter space. This statement implies that all SI, BC74, and CB00 models are equally flexible, a sentiment not currently found in the literature.

The local parameter space of all BC74 models is a straight line; the local parameter space for CB00 and SI models is a straight or curvaceous line. Whether one or the other performs better depends on the distribution of the "true" or "best" local parameter values. A straight-line parameter space could dissect a cloud of true parameter values as well as a curved line. In the results above, the BC74 Chapman-Richards model performed virtually identically to all CB00 models; their local parameter spaces pass through the cloud of estimates obtained with the random effects model essentially in a straight-line fashion, but on rotated axes.

As illustrated by Figure 2, SI, BC74, and CB00 models can only produce systematically biased fitted and prediction curves, the only caveat being if the entire population adheres precisely to the specified functional parameter relationship. When inferential focus is retrospective, e.g., determining whether growth curves vary across levels of an experimental factor, false conclusions could result, which we illustrated using Figure 4. In the case of growth and yield, in which inferential focus is on predictions, this type of bias will be carried forward, and, e.g., suboptimal management decisions could be reached. Because effects-formulated models allow all curves possible with a base function, it seems more fruitful to

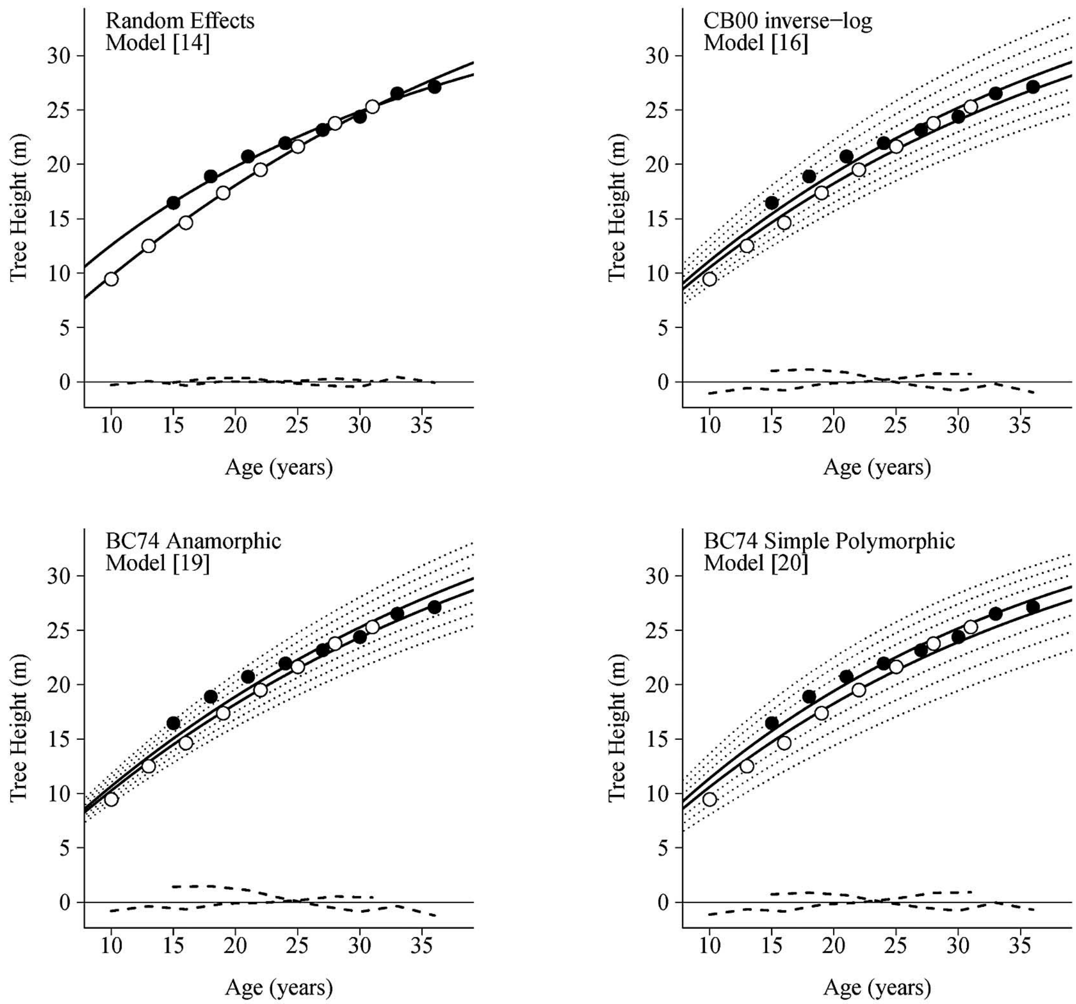

Figure 4. Data and fitted trends using four formulations of the Chapman-Richards function. Symbols: observed values from permanent sample plots 136 (●) and 186 (○); Solid lines: fitted curves; dashed lines: residuals; dotted lines: other possible curves.

research model-fitting and prediction estimation strategies for them.

When models are fitted to longitudinal forestry data, we recommend random-effects models (mixed models). This recommendation is especially true for unbalanced data, where data trends with few observations and outlying trends are handled by the random distribution placed on the λs. Random-effects models are also useful for predictions with few observations. Fixed-effects models are useful in model-fitting situations when each data trend has many observations and for predictions when many previous observations are available. The definition of "few" and "many" here will be situation dependent and could be a productive path of research.

The random-effects models analyzed above (14 and 21) were fitted using basic mixed-model methods without a formal data-analytic model-building procedure: a standard parameterization of the base function and normally distributed random effects without correlation parameters between the random effects were used. Furthermore, we restricted our attention only to effects models having the same varying base parameters as the CB00 models considered by Dieguez-Aranda et al. (2006); unlike SI, BC74, and CB00 models, it is straightforward to extend effects (either random or fixed) models to allow all parameters to vary. For example, the Chapman-Richards function cannot be formulated as an EVP-parameterized CB00 model wherein β_2 and either of the other parameters vary, because in that case algebraically solving for λ_i is not possible. In a more formal analysis, these options should be considered more closely. Reparameterizations, such as EVPs, could lead to a more suitable model; e.g., replacing an asymptote parameter with a parameter representing the expected value of the heights within the data range. For example, as a subanalysis we fitted the log-logistic function under a two-point EVP,

$$f(x_{ij}, \boldsymbol{\beta}_i) = \beta_1 \frac{g(25)}{g(x_{ij})},$$

$$g(z) = \beta_1(25^{-\beta_3} - z^{-\beta_3}) + \beta_2(z^{-\beta_3} - 10^{-\beta_3}) \qquad (14)$$

(i.e., β_1 and β_2 represent expected height at age 25 and 10, respectively) and obtained much improved fit statistics, depending on whether only β_1 and β_2 varied (SSR = 136.5, MAR = 0.2763) or all three base parameters varied (SSR = 93.2, MAR = 0.2267). In general, the fitted random-effects distributions will change, depending on parameterization; e.g., the distribution of asymptotes is different from, say, the distribution of heights at age 25.

In addition, more flexible distributional assumptions, such as a mixture of normals, can highlight patterns between local parameter values (Davidian and Giltinan 1995, Chapter 7, Verbeke and Lesaffre 1996, Zhang and Davidian 2001, Proust and Jacqmin-Gadda 2005). That is, in situations in which multimodality or nonlinear tendencies between local parameter values persist after reparameterization, heterogeneous random effects distributions could be useful. Whereas Cieszewski (2003, p. 544), Nord-Larsen (2006, p. 179), and Nord-Larsen and Johannsen (2007, p. 368) are of the opinion that normally distributed random effects are insufficient for forestry height–age populations, the results presented here suggest that the normality assumption is far less restrictive than the one-dimensionality of CB00 models; in addition, these authors did not consider the fact that random-effects models could incorporate heterogeneous and other non-normal effects distributions.

The algebraic reparameterization of BC74 and CB00 models provides no benefit for model fittings, could falsely give the appearance of biologically based "initial condition" relationships, and could mask the one-dimensionality of the local parameter space. Moreover, BC74 models are fixed-effect models wherein $k = 1$ under an EVP. We have thus chosen to depart from the convention of calling these "ADA" or "GADA" models; perhaps a more appropriate terminology would be *fixed-effects site index models.*

Conclusions

The traditional height–age modeling methods of forestry (SI, BC74, and CB00) have one-dimensional local parameter spaces, which are based on weak or even baseless theoretical considerations that have little or no empirical support. Their ad hoc model formulations lead to inflexible and biased families of curves. Furthermore, the algebraic reparameterization of BC74 and CB00 models is unnecessary and even hinders the set of available models. These concepts are perhaps known to those familiar with these methods, yet, to our knowledge, are undocumented in the literature. The current state of random-parameter (mixed-effects) modeling theory, combined with present-day computing power and a wide availability of relevant software makes effects-formulated models readily available. We suggest the use of fixed- or random-effects models for longitudinal data modeling in forestry. We have given a conceptual basis for and empirical examples of reduced model fittings due to an a priori one-dimensional parameter space, which is a characteristic of SI, BC74, and CB00 (site index, algebraic difference approach, and generalized algebraic difference approach, respectively) models. Results showing similar reductions for predictions are forthcoming, and preliminary findings have been presented elsewhere [1].

Endnote

[1] Prediction validations comparing GADA and mixed-models were presented by the first and second authors at the 2009 Southern Mensurationists Meeting (SOMENS), San Antonio, TX, Oct. 25–27, 2009, and the 2010 SOMENS, Knoxville, TN, Oct. 5–7, 2010, using height–age and tree-taper data sets (available upon request).

Literature Cited

BAILEY, R.L., AND J.L. CLUTTER. 1974. Base-age invariant polymorphic site curves. *For. Sci.* 20(2):155–159.

BARRIO-ANTA, M., F. CASTEDO-DORADO, U. DIEGUEZ-ARANDA, J.G. ALVAREZ-GONZALEZ, B.R. PARRESOL, AND R. RODRIGUEZ-SOALLEIRO. 2006. Development of a basal area growth system for maritime pine in northwestern Spain using the generalized algebraic difference approach. *Can. J. For. Res.* 36(6):1461–1474.

BECK, D.E. 1971. Height-growth patterns and site index of white pine in the southern in the southern Appalachians. *For. Sci.* 17:252–260.

BRAVO-OVIEDO, A., M. TOME, F. BRAVO, G. MONTERO, AND M. DEL RIO. 2008. Dominant height growth equations including site attributes in the generalized algebraic difference approach. *Can. J. For. Res.* 38(9):2348–2358.

BURKHART, H.E., D.C. CLOEREN, AND R.L. AMATEIS. 1985. Yield relationships in unthinned loblolly pine plantations on cutover, site-prepared lands. *South. J. Appl. For.* 9:84–91.

CIESZEWSKI, C.J. 2002. Comparing fixed- and variable-base-age polymorphic site equations having single versus multiple asymptotes. *For. Sci.* 48(1):7–23.

CIESZEWSKI, C.J. 2003. Developing a well-behaved dynamic site equation using a modified Hossfeld IV function $Y^3 = (ax^m)/(c + x^{m-1})$, a simplified mixed-model and scant subalpine fir data. *For. Sci.* 49(4):539–554.

CIESZEWSKI, C.J., AND R.L. BAILEY. 2000. Generalized algebraic difference approach: theory based derivation of dynamic site equations with polymorphism and variable asymptotes. *For. Sci.* 46(1):116–126.

CIESZEWSKI, C.J., AND M. STRUB. 2007. Parameter estimation of base-age invariant site index models: Which data structure to use?—A discussion. *For. Sci.* 53(5):552–555.

CIESZEWSKI, C.J., AND M. STRUB. 2008. Generalized algebraic difference approach derivation of dynamic site equations with polymorphism and variable asymptotes from exponential and logarithmic functions. *For. Sci.* 54(3):303–315.

DAVIDIAN, M., AND D.M. GILTINAN. 1995. *Nonlinear models for repeated measurement data.* Chapman & Hall, London, UK. 359 p.

DIEGUEZ-ARANDA, U., H.E. BURKHART, AND R.L. AMATEIS. 2006. Dynamic site models for loblolly pine (*Pinus taeda* L.) plantations in the United States. *For. Sci.* 52(3):262–272.

FROTHINGHAM, E.H. 1918. Height growth as a key to site. *J. For.* 16(7):754–760.

JAMES, C., B. KRUMLAND, AND P.J. ECKERT. 2007. *Carbon sequestration in Californian forests; Two case studies in managed watersheds.* Sierra Pacific Industries, Redding, CA. Available online at www.spi-ind.com/html/pdf_forests/CarbonSequestration.pdf; last accessed July 19, 2009.

KRUMLAND, B., AND H. ENG. 2005. *Site index systems for major young-growth forest and woodland species in northern California.* California For. Rep. No. 4. California Department of Forestry and Fire Protection, Sacramento, CA. 219 p. Available online at www.demoforests.net/Warehouse/Docs/ForestryReports/ForestryReport4.pdf; last accessed July 19, 2009.

LINDSTROM, J.L., AND M.B. BATES. 1990. Nonlinear mixed effects models for repeated measures data. *Biometrics* 46(3):673–687.

LUNDGREN, A.L., AND W.A. DOLID. 1970. Biological growth functions describe published site index curves for Lake States timber species. US For. Serv. Res. Paper NC-36., North Central For. Exp. Stn., St. Paul, MN. 9 p.

Ni, C., and L. Zhang. 2007. An analysis and comparison of estimation methods for self-referencing equations. *Can. J. For. Res.* 37(8):1472–1484.

Nord-Larsen, T. 2006. Developing dynamic site index curves for European beech (*Fagus sylvatica* L.) in Denmark. *For. Sci.* 52(2):173–181.

Nord-Larsen, T., and V.K. Johannsen. 2007. A state-space approach to stand growth modeling of European beech. *Ann. For. Sci.* 64(4):365–374.

Payandeh, B. 1974. Formulated site index curves for major timber species in Ontario. *For. Sci.* 20(2):143–144.

Phillips, K.M. 2002. *Modeling within-tree changes in wood specific gravity and moisture content for loblolly pine in Georgia.* MSc thesis, University of Georgia, Athens, GA. 83 p.

Pinheiro, J.C., and D.M. Bates. 2000. *Mixed effects models in S and S-plus.* Statistics and Computing Series, Springer, New York. 528 p.

Proust, C., and H. Jacqmin-Gadda. 2005. Estimation of linear mixed models with a mixture of distributions for the random effects. *Comput. Methods Programs Biomed.* 78:165–173.

Ratkowsky, D.A. 1990. *Handbook of nonlinear regression models.* Marcel Dekker, New York. 241 p.

Roth, F. 1916. Concerning site. *J. For.* 14(1):3–12.

R Development Core Team. 2009. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available online at http://www.R-project.org; last accessed July 30, 2009.

Strub, M., C.J. Cieszewski, and D. Hyink. 2005. Self-referencing taper curves for loblolly pine. P. 59–64 In *Proc. of the 4th annual FIA symposium*, McRoberts, R.E., G.A. Reams, P. Van Deusen, H. McWilliams, and C.J. Cieszewski (eds.). US For. Serv. Gen. Tech. Rep. NC-252. North Central Res. Stn., St. Paul, MN.

Verbeke, G., and E. Lesaffre. 1996. A linear mixed-effects model with heterogeneity in the random-effects population. *J. Am. Stat. Assoc.* 91(433):217–221.

Wang, M., B.E. Borders, and D. Zhao. 2007. Parameter estimation of base-age invariant site index models: Which data structure to use? *For. Sci.* 53(5):541–551.

Wang, M., B.E. Borders, and D. Zhao. 2008. An empirical comparison of two subject-specific approaches to dominant heights modeling: The dummy variable method and the mixed model method. *For. Ecol. Manag.* 255(7):2659–2669.

Wang, Y., S. Huang, R.C. Yang, and S. Tang. 2004. Error-in-variable method to estimate parameters for reciprocal base-age invariant site index models. *Can. J. For. Res.* 34(9):1929–1937.

Weiskittel, A.R., D.W. Hann, D.E Hibbs, T.Y. Lam, and A.A. Bluhm. 2009. Modeling top height growth of red alder plantations. *For. Ecol. Manag.* 258(3):323–331.

Zhang, D., and M. Davidian. 2001. Linear mixed models with flexible distributions of random effects for longitudinal data. *Biometrics* 57(3):795–802.

Zhao, D., M. Wilson, and B.E. Borders. 2005. Modeling response curves and testing treatment effects in repeated measures experiments: a multilevel nonlinear mixed-effects model approach. *Can. J. For. Res.* 35(1):122–132.