A Mixed-Effects Model for the dbh–Height Relationship of Shortleaf Pine (Pinus echinata Mill.)

Chakra B. Budhathoki, Thomas B. Lynch, and James M. Guldin

Individual tree measurements were available from over 200 permanent plots established during 1985–1987 and later remeasured in naturally regenerated even-aged stands of shortleaf pine (Pinus echinata Mill.) in western Arkansas and eastern Oklahoma. The objective of this study was to model shortleaf pine growth in natural stands for the region. As a major component of the shortleaf pine modeling effort, individual tree-level dbh–total height model was developed in which plot-specific random parameters were fitted using maximum-likelihood methods. The model predicts tree height on the basis of dbh and dominant stand height (which could be obtained from a site-index model). The mixed-effects model approach was found to predict the total height better than the similar models developed previously for this species using ordinary least-squares methods. Moreover, such a model has the appeal of generalization of the results over a region from which the plots were sampled; and also of calibration of parameters for newly sampled stands with minimal measurements.

Keywords: mixed-effects, dbh, total height, dominant height

Shortleaf pine (Pinus echinata Mill.) forests contain standing cubic volume that is second only to loblolly pine (Pinus taeda L.) among the southern pines in the United States. Shortleaf pine grows in 22 states in area more than 1.139,600 km², ranging from southeastern New York to eastern Texas (Willett 1986). Past shortleaf pine growth studies include Murphy (1982, 1986), Lynch et al. (1991), Murphy et al. (1992), Lynch and Murphy (1995), and Lynch et al. (1999). However, there is still relatively little published information concerning the growth of shortleaf pine compared with the quantity of information available for loblolly and other southern pine species. A model for the relationship between dbh and total height is needed for a quantitative description of shortleaf pine forests.

An early model that described the relationship between stand age and the average height of dominants and codominants (Avery and Burkhart 2002) in shortleaf pine forests was the system of site-index curves developed for the US Forest Service (US Forest Service 1929), using graphical techniques. Graney and Burkhart (1973) developed an equation that can be used to predict site index given total height of dominant and codominant trees and age for shortleaf pine. The polymorphic system of site-index curves of Graney and Burkhart (1973) was fitted to shortleaf pine data using nonlinear ordinary least-squares methods. These site-index curves provided information concerning the development of dominant stand height for shortleaf pine forests but could not be used to predict heights of intermediate or suppressed trees or heights of individual dominant and codominant trees of various sizes.

Although the problem of correlated measurements in forestry data has long been recognized (e.g., Ferguson and Leech 1978, West et al. 1984), least-squares techniques assuming a completely random sample have dominated the forest growth and yield modeling literature until quite recently. Lappi and Bailey (1988) and Gregoire et al. (1995) are among those who have proposed mixed modeling as an alternative to ordinary least-squares methods for complex data structures that do not conform well to the assumptions of ordinary least squares. Lynch and Murphy (1995) used seemingly unrelated regression to fit a diameter–height model for natural even-aged shortleaf pine. However, their approach did not account for possible correlations among sample trees located on the same plot. Lynch and Murphy (1995) also provide a comprehensive review of work prior to 1995 on modeling of tree height with dbh and age (or time). Subsequently, Lynch et al. (1999) developed a system to model growth of even-aged shortleaf pine forests on the basis of a distance-independent individual tree basal area growth equation, the dbh–height model of Lynch and Murphy (1995), and a distance-independent individual tree probability of survival equation for shortleaf pine. To date, most models developed to quantitatively describe shortleaf pine forests have been fitted using graphical methods (US Forest Service 1929), ordinary/weighted least-squares, or seemingly unrelated regression methods.

Mixed-Effects Modeling

Shortleaf pine individual-tree models fitted by least-squares techniques have not accounted for correlation among measurements due to plot-level grouping of individual observations. Mixed-effects models can use random plot effects to account for this type of correlation in the data. Trincado and Burkhart (2006) found that the correlated error assumption could be relaxed when tree-level random effects were included in a loblolly pine individual tree taper model fitted using mixed-model techniques. This approach also

Received June 12, 2007; accepted September 26, 2007.

Chakra B. Budhathoki (chakra.budhathoki@uchsc.edu) is statistician, University of Colorado Health Sciences Center, 4200 East 9th Avenue B-139, Denver, CO 80262. Thomas B. Lynch is professor, Department of Natural Resource Ecology and Management, Oklahoma State University, Stillwater, OK 74078. James M. Guldin is supervisory ecologist and project leader, Southern Research Station, US Forest Service, P.O. Box 1270, Hot Springs, AR 71902. This article is approved for publication by the Director of the Oklahoma Agricultural Experiment Station and supported by Project MS-1887. The assistance received from Department of Forestry (now renamed to Natural Resource Ecology and Management) at Oklahoma State University is highly appreciated. The cooperation of the US Forest Service Southern Research Station and the Ozark and Ouachita National Forests in data collection and financial support for the study is much appreciated. We are also thankful for the comments of the reviewers to improve this article.

Copyright © 2008 by the Society of American Foresters.

SOUTH. J. APPL. FOR. 32(1) 2008 5
facilitated calibration of taper curves to specific locations with new data. Lappi and Bailey (1988) presented mixed modeling as a promising alternative to the methods that were then conventional for development of site-index curves. Gregoire et al. (1995) used mixed-effects modeling to account for correlation due to grouping in data structures that commonly occur in forestry applications. Gregoire et al. (1995) cited lack of easily available and user-friendly software as an important reason why there was still not much application of mixed-effects models in forestry at that time. Thus, it is expected that mixed effects modeling may provide better results than ordinary least-squares when used to develop a dbh–height model for shortleaf pine.

Lappi (1997) used a mixed-model approach to analyze dbh–height relationships for two jack pine data sets, one from plantations and the other from naturally regenerated stands. The diameter–height curve parameters were partitioned into an age-dependent trend (population mean), a random stand effect, and a random time effect. A practical advantage of this type of model would be calibration in which the random stand and time effects could be predicted with some additional measurements from new forest stands of interest without requiring the detailed observations normally needed for new dbh–height equations in new stands. Fang and Bailey (2001) modeled dominant height growth of slash pine (Pinus elliottii Engelm.), using data obtained from a study with several silvicultural treatments installed in Georgia and north Florida. They parameterized the three-parameter Richards equation (Richards 1959) and used a nonlinear mixed-effects model approach to predict dominant height growth in presence of silvicultural treatments, such as chopping, fertilization, and burning. Hall and Bailey (2001) used multilevel nonlinear mixed models to describe forest growth relationships.

Mehtätalo (2004) used a mixed model with longitudinal height and diameter data for Norway spruce (Picea abies [L.] Karst.). The Korf growth curve was used as a basic growth function for the height–diameter relationship. Calama and Montero (2004) used a mixed-model approach to model the individual-tree diameter–height relationship for stone pine (Pinus pinea L.) in Spain. Lynch et al. (2005) used a random-parameter approach to analyze dbh–height data for cherrybark oak (Quercus pagoda Raf.) from East Texas, fitting a model similar to that reported by Lappi (1991). Uzoh and Oliver (2006) used a composite approach (as described by Wykoff 1990) for height increment modeling of managed even-aged stands of ponderosa pine (Pinus ponderosa Dougl.). Random effects for locations, plots, and trees were used in the model, and an autoregressive covariance structure was used to model the repeated measurements. Transformation of the periodic annual height increment to a logarithmic scale enabled Uzoh and Oliver (2006) to construct a linear mixed model. Site index was found to have more effect on height growth than other variables used in the study. A nonlinear mixed model was developed for height growth for Eucalyptus plantations in Brazil by Calegario et al. (2005). They modeled dominant height as a logistic function of age with plot random effects.

This brief review indicates increasing use of mixed-model techniques for forest growth and yield modeling in recent years. However, prior to the preliminary work of Budhathoki et al. (2006) involving linear mixed models for basal area growth, no published work has applied mixed-model techniques to shortleaf pine growth modeling.

### Methods

#### Data

Individual tree measurements (e.g., total height, dbh, crown height) on shortleaf pine were available from 208 permanent plots established in naturally regenerated even-aged stands in western Arkansas and eastern Oklahoma. The study plots were established from 1985 to 1987 by Oklahoma State University in collaboration with the US Forest Service Southern Research Station and the Ouachita and Ozark National Forests. Ranges for study design variables (stand basal area, site index, and stand age) that were used in establishing permanent plots are given in Table 1. These data were used in the development of a diameter—height relationship for even-aged, naturally occurring shortleaf pine. Additional data were also available from a thinning study, which was modified to comply with the same design criteria described above (Lynch et al. 1999). Three measurements at an interval of 4 to 5 years were available for over 8,000 trees. However, total height and crown data were available from only subsample of trees selected in each plot to span the range of tree diameters on that plot (Table 2). Ring count was used to determine individual tree age, and stand age was assumed to be the average age of the representative dominant and codominant (Avery and Burkhardt 2002) trees in the plot, assuming a plot would represent the entire stand. Site index (average total height of dominant and codominant trees at base age 50 years) was calculated using the equation of Graney and Burkhardt (1973). The dominant height calculation is based on a site-index value obtained by averaging over all the three measurements.

The summary statistics for variables used in modeling growth and development of even-aged shortleaf pine forests are presented in Table 2. These data were used in the development of a diameter–height relationship for even-aged, naturally occurring shortleaf pine.

#### Statistical Analysis

Total height for individual trees can be modeled as an explicit function of tree age (Curtis 1967, Lappi and Bailey 1988, Meng et al. 1997), and it can also be modeled using dominant height and dbh as predictors where dominant height is a function of tree age and site index (Lynch and Murphy 1995, Lynch et al. 1999). The analysis for this article included data from third measurements that were not available previously for development of the diameter–height relationship model of Lynch et al. (1999). The
Table 2. Summary of stand-level and tree variables recorded/observed in this study.

<table>
<thead>
<tr>
<th>Variable</th>
<th>No. of observations</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basal area (m²/ha at establishment)</td>
<td>208</td>
<td>21.33</td>
<td>6.68</td>
<td>6.27</td>
<td>29.62</td>
</tr>
<tr>
<td>Stand age (year at establishment)</td>
<td>208</td>
<td>41.8</td>
<td>19.7</td>
<td>18.0</td>
<td>93.0</td>
</tr>
<tr>
<td>Site index (m at age 50 years)</td>
<td>208</td>
<td>17.5</td>
<td>2.9</td>
<td>12.2</td>
<td>26.6</td>
</tr>
<tr>
<td>Dominant height (m)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First measurement</td>
<td>208</td>
<td>19.9</td>
<td>5.8</td>
<td>7.3</td>
<td>31.3</td>
</tr>
<tr>
<td>Second measurement</td>
<td>208</td>
<td>20.8</td>
<td>5.6</td>
<td>8.6</td>
<td>31.8</td>
</tr>
<tr>
<td>Third measurement</td>
<td>208</td>
<td>21.6</td>
<td>5.4</td>
<td>9.8</td>
<td>32.6</td>
</tr>
<tr>
<td>Total height (m)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First measurement</td>
<td>2,688</td>
<td>17.4</td>
<td>6.7</td>
<td>3.0</td>
<td>34.1</td>
</tr>
<tr>
<td>Second measurement</td>
<td>3,049</td>
<td>18.6</td>
<td>6.4</td>
<td>3.0</td>
<td>34.4</td>
</tr>
<tr>
<td>Third measurement</td>
<td>3,235</td>
<td>19.8</td>
<td>6.1</td>
<td>3.9</td>
<td>36.3</td>
</tr>
<tr>
<td>dbh (cm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First measurement</td>
<td>8,284</td>
<td>18.8</td>
<td>9.9</td>
<td>2.8</td>
<td>61.9</td>
</tr>
<tr>
<td>Second measurement</td>
<td>8,092</td>
<td>20.8</td>
<td>9.9</td>
<td>3.0</td>
<td>64.5</td>
</tr>
<tr>
<td>Third measurement</td>
<td>7,591</td>
<td>23.1</td>
<td>10.2</td>
<td>3.8</td>
<td>67.6</td>
</tr>
</tbody>
</table>

Table 3. Parameter estimates from SAS PROC NLMIXED.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard error</th>
<th>t value</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$</td>
<td>2.9111</td>
<td>0.0853</td>
<td>34.1</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.7853</td>
<td>0.00432</td>
<td>181.8</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>7.0281</td>
<td>0.3280</td>
<td>21.4</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>0.0031</td>
<td>0.00241</td>
<td>41.6</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>$\beta_4$</td>
<td>0.00166</td>
<td>0.000107</td>
<td>15.5</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

Total observations = 8,964; residual variance = 1.9522.

Akaike information criterion (AIC) = 31,435; Bayes information criterion (BIC) = 31,471; $-2 \times \log$-likelihood = 31,425.

Table 4. Parameter estimates from PROC NLINX.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard error</th>
<th>t value</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$</td>
<td>1.8296</td>
<td>0.05302</td>
<td>34.5</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.8427</td>
<td>0.008108</td>
<td>103.9</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>17.1483</td>
<td>1.5877</td>
<td>10.8</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>1.3613</td>
<td>0.03822</td>
<td>35.6</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>$\beta_4$</td>
<td>0.001699</td>
<td>0.000186</td>
<td>9.1</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>$\sigma^2_{\epsilon}$</td>
<td>10.6011</td>
<td>3.075</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Number of plots = 208; residual degrees of freedom = 207; residual variance = 1.4408.

Akaike information criterion (AIC) = 29,263; Bayes information criterion (BIC) = 29,286; $-2 \times \log$-likelihood = 29,249.

Basic objective of this work is to provide improved parameter estimates in a diameter—height model similar to that developed by Lynch et al. (1999) using mixed modeling techniques. The results of Lynch et al. (1999) were originally reported in English units. However, the analysis presented here uses metric units, and random effects for plots were added. Furthermore, analysis of the data including the third measurement period indicated that the dbh—height relationship is significantly affected by stand density in terms of basal area per hectare.

A modified total height prediction model having the same form as that given by Lynch et al. (1999) but including stand basal area as an additional independent variable and using metric units is given in

$$(H_j - 1.37) = \beta_0 (H_D - 1.37)^{\beta_b} \exp(-{(\beta_2 D_j)^{\beta_b}}) + (\beta_b) B_3$$

where $H_j$ = total height (m) of tree $j$; $D_j$ = dbh (cm) of tree $j$ (breast height = 1.37 m); $H_D$ = dominant height (m) for a plot as per Graney and Burkhart (1973); $B_3$ = stand basal area (m²/ha); $\beta_0$, $\beta_1$, $\beta_2$, $\beta_3$, $\beta_4$ = model parameters; and $\epsilon_j$ = random error for tree $j$.

It is assumed that $\epsilon_j \sim N(0, \sigma^2_{\epsilon})$. Lynch et al. (1999) fitted Model 1 without stand basal area in the model using first two measurements of the data summarized in Table 2 that were then available. They found that contribution of stand basal area as an explanatory variable to predict total height was limited with the first two measurements. Lynch and Murphy (1995) discussed the limitation of stand density variable in predicting total height in managed natural stands such as this study, as compared with plantations in which stand density is expected to have more significant contribution. However, the addition of a third measurement may have allowed for additional diameter response due to time elapsed since plot establishment, resulting in significance of basal area for this analysis. The modified nonlinear model described above, including stand basal area, was initially fitted using ordinary least-squares (OLS) methods.

Model 1 can be modified to obtain a mixed-effects model that includes random effects for plots. Difference in residual mean squares can be used to compare these two models in addition to comparison of the fit statistics Akaike information criterion (AIC) (Akaike 1974) and Bayes information criterion (BIC) (Schwarz 1978). Moreover, the variance component for random effects can be used to test the statistical significance of the mixed model versus the model without random effects. As indicated above, the mixed model has several attractive properties compared with a model that is developed by the ordinary least-squares method, including a more realistic representation of the data structure (grouping of trees on plots) than would typically be the case with ordinary least squares.

The following mixed model was developed using Model 1 as a basis. This model including plot-specific random effects and one random effects variance component was fitted to all measurements:

$$(H_j - 1.37) = \beta_0 (H_D - 1.37)^{\beta_b} \exp(-{(\beta_2 + b_2) D_j^{\beta_b}} + \beta_4 B_3) + \epsilon_j$$

where $H_j$ = total height (m) of tree $j$ in plot $i$; $D_j$ = dbh (cm) of tree $j$ in plot $i$ (breast height = 1.37 m); $H_D$ = dominant height (m) for plot $i$ as given above in Model 1; $B_3$ = stand basal area (m²/ha) for stand $i$; $\beta_0$, $\beta_1$, $\beta_2$, $\beta_3$, $\beta_4$ = fixed-effects parameters; $b_2$ = random effect associated with $B_2$ (dbh), specific to $i$th plot; and $\epsilon_j$ = within-plot error (random error for tree $j$ in plot $i$).

It is assumed that $b_{ij} \sim N(0, \sigma^2_{b2})$, $\epsilon_j \sim N(0, \sigma^2_{\epsilon})$, and cov$(b_{ij}, \epsilon_j) = 0$. We would usually be interested in an estimate of var$(b_{ij})$, that is, $\sigma^2_{b2}$, a variance component describing the spread of the random
coefficients. The parameters were fitted using maximum-likelihood methods. This model also makes it possible to use calibration techniques to predict a plot- or stand-specific random effect ($b_2$) using dbh–height measurements from a particular plot or stand in a forest of interest. In the notation above, a plot would represent a typical stand. Random effects were associated with the fixed-effect coefficient $b_2$, which was associated with the tree-level variable dbh. Since the dominant height and stand basal area are both plot-level variables, no plot random effects were associated with these variables. Models 1 and 2 were fitted using SAS PROC NLIN and PROC NLMIXED, respectively (Tao 2002, p. 411–462). Because the OLS likelihood at its maximum can be expressed as a function of the
residual sum of squares, the residual sum of squares from PROC NLIN was used to calculate values of $-2 \times \log$-likelihood, AIC, and BIC for Model 1.

### Results and Discussion

The fixed-effect parameter estimates and associated statistics for the fitted models are presented in Tables 3 and 4. Fit statistics and results obtained by fitting Model 1 using PROC NLIN are given in Table 3. Similarly, fit statistics, parameter estimates, and testing information from fitting Model 2 are presented in Table 4.

All the model coefficients are significantly different from zero for both the models (Tables 3 and 4). However, Model 2 also includes variance component quantifying the variability among the plot-specific random effects associated with the variable dbh, i.e., parameter...
This variance component is significant ($P = 0.0007$), with a 95% confidence interval of [4.5387, 16.6635]. The fixed parameter estimates are similar in magnitude and algebraic sign to those of the height–diameter model reported by Lynch et al. (1999) with two measurements, and the updated parameter estimates of Model 1 using all three measurements. The residual variance is reduced in Model 2 compared to Model 1. Model 2 has smaller values of AIC, BIC, and $-2 \times \log$-likelihood. Moreover, Model 2 accounts for the data structure more completely than Model 1 because the plot random effects account for the fact that trees are selected as a cluster within a plot, and not individually at random as would be implicitly assumed by the least-squares estimation used to fit parameters in Model 1. Therefore, Model 2 with plot random effects is preferred for prediction of individual tree heights in even-aged shortleaf pine natural stands.

The standardized residuals from Model 2 are plotted against predicted or fitted total height values (Figure 1) and also against dbh (Figure 2). These two plots do not reveal any systematic patterns. The standardized residuals are also plotted against design variables and are presented in Figures 3 to 5. Model 2 appears to make very good predictions over the range of design variables with a minimum bias, although there appears to be slight overprediction for small trees.

Model 2, a mixed-effects model for total height, is clearly a better alternative to the Model 1 fitted by OLS since the variance component associated with plot random effects for a dbh parameter is statistically significant. This argument is also supported by smaller values of AIC, BIC, and $-2 \times \log$-likelihood compared with Model 1. The fitted mixed model is similar in concept to those reported by Lappi (1991) and Lappi and Bailey (1988). Model 2 can also be used with calibration to predict random effects for stands not used in model fitting with a minimal number of measurements of dbh and height from sample trees located in shortleaf pine forests of interest. For example, Lynch et al. (2005) used a random-parameter model and calibration for cherrybark oak data from Texas. Similarly, Mehtätalo (2004) fitted a mixed model with diameter and height data for Norway spruce and used calibration techniques to predict random parameters, resulting in improved height predictions for a new stand.

Conclusions

The statistical significance of the variance component in the mixed-effects Model 2 indicates that it is a more realistic representation of the shortleaf pine data structure than an OLS fit of Model 1. Model 1 ignores correlation among individuals located on the same plot. However, Model 2 uses plot random effects to account for within-plot correlation. Trincado and Burkhart (2006) suggested that where tree-level random effects were present in a taper model, the correlated error assumption could be relaxed. This can simplify calibration to localize the model with supplemental data from a specific new location. Although they were considering multiple observations on the same tree stem, this should be similar in principle to multiple observations on the same plot in time, where random plot effects are present. The mixed model with random effects for plots has applicability for predictions in the forest population from which plots/stands were selected. The parameter estimates obtained from this new model can be used to help develop information used for practical forest management decision making. For example, the shortleaf pine total height estimates from Model 2 could be incorporated in a revision of the Shortleaf Pine Stand Simulator (Huebschmann et al. 1998) along with other revised components of the system, resulting in improved estimates of future conditions in naturally occurring even-aged shortleaf pine forests.


