2011

A mixed-effects height-diameter model for individual loblolly and slash pine trees in east Texas

Dean W. Coble
Arthur Temple College of Forestry and Agriculture, Stephen F. Austin State University, dcoble@sfasu.edu

Young Jin Lee

A Mixed-Effects Height–Diameter Model for Individual Loblolly and Slash Pine Trees in East Texas

Dean W. Coble and Young-Jin Lee

ABSTRACT

A new mixed-effects model was developed that predicts individual-tree total height for loblolly (*Pinus taeda*) and slash pine (*Pinus elliottii*) as a function of individual-tree diameter (in.), dominant height (ft), quadratic mean diameter (in.), and maximum stand diameter (in.). Data from 119,983 loblolly pine and 42,697 slash pine height–diameter observations collected on 185 loblolly pine and 84 slash pine permanent plots located in plantations throughout East Texas were used for model fitting. This new model is an improvement over earlier models fit with ordinary least squares, in that it can be calibrated to a new stand with observed height–diameter pairs, thus improving height prediction. An example is provided that describes how to calibrate the model to a new stand with observed data.

Keywords: *Pinus taeda*, *Pinus elliottii*, height prediction, missing heights, random effects

Total tree height represents an important independent variable in volume or biomass prediction models. However, height is often subsampled in forest inventories because it can be time-consuming to estimate. In these situations, height–diameter models are constructed to predict height from diameter. However, height–diameter relationships are often quite variable between forest stands, so this makes the use of local or single-entry volume tables or equations restricted to the forest stand for which they were developed. The use of mixed-effects modeling techniques has made it possible to build height–diameter models that can be calibrated to local stands with a subsample of heights and diameters. This type of model has the potential to be useful for forest inventory or growth and yield models that require predicted heights by diameter class (e.g., diameter distribution models).

Mixed-effects models have previously been applied to height–diameter estimation. Lappi (1997) used a mixed-effects model to predict height from diameter for jack pine (*Pinus banksiana*). Sharma and Parton (2007) developed mixed-effects height–diameter models for boreal tree species in Ontario, Canada. Mehtatalo (2004) used the Korf equation (Zeide 1989, 1993) to model height from diameter for Norway spruce (*Picea abies*). Calama and Montero (2004) developed a mixed height–diameter model for stone pine (*Pinus pinea*) in Spain. Lynch et al. (2005) developed a mixed-effects model also based on the Korf equation for height prediction of cherrybark oak (*Quercus pagoda*). Trincado et al. (2007) developed a regional mixed-effects height–diameter model for loblolly pine in the southeastern United States. Budhathoki et al. (2008) predicted height from diameter for shortleaf pine (*Pinus echinata*) with a mixed-effects model using stand-level independent variables. Budhathoki et al. (2008) found that including dominant height and basal area per hectare along with individual tree diameter improved the predictions of individual tree height. Trincado et al. (2007), however, did not include any stand-level variables, arguing that the calibration of the regional model to the local stand accounts for differences in stand density and site quality. The objective of this study was to develop a mixed-effects height–diameter model for loblolly and slash pine plantations in East Texas. No such model has been developed for this region. We examined models that included stand- and tree-level independent variables as well as tree-level variables only to determine the best predictive model that can be calibrated to local stands.

Methods

In this study, 119,983 height–diameter observations were measured on loblolly pine trees repeatedly sampled on 185 permanent plots (Table 1) as well as 42,697 height–diameter observations measured on slash pine trees repeatedly sampled on 84 permanent plots (Table 2) located in East Texas pine plantations. On each plot, the same trees have been remeasured since 1982 as part of the East Texas Pine Plantation Research Project (Lenhart et al. 1985). Plots ranged in age from 2 to 40 years old and represented a wide range of site quality and density. Because the height–diameter pairs were repeated observations, the assumption of independent residuals was violated. Thus, parameter estimates from ordinary least squares

Received January 20, 2009; accepted April 29, 2010.

Dean Coble (dcoble@sfasu.edu), Stephen F. Austin State University, Arthur Temple College of Forestry and Agriculture, Box 6109, SFA Station, Nacogdoches, TX 75962. Young-Jin W Lee (leeyj@kongju.ac.kr), Department of Forest Resources, Kongju National University, Yesan, Chungnam 340-802, South Korea. *The authors thank Dr. Jung-Kee Choi for allowing them to present preliminary results of this research at the 2008 International Union of Forestry Research Organizations (IUFRO) International Conference meeting. They also thank three anonymous reviews for their very helpful comments in improving the article; Lewis Jordan of Weyerhaeuser, Curtis VanderSchaaf of the Texas Forest Service, and Shongming Huang of the Alberta Department of Sustainable Resource Development for statistical advice that improved the article; and the East Texas Pine Plantation Research Project for providing the data.*

Table 1. Observed individual tree and stand characteristics for East Texas loblolly pine plantation dataset (n = 119, 983 observations from 185 remeasured permanent plots across eight measurement cycles).

Variables	Mean	Standard deviation	Minimum	Maximum
Age (yr)	14.9	7.2	2.0	40.0
Dominant height (ft)	47.1	19.8	5.0	99.0
Site index (25-yr, ft)	69.2	11.2	23.0	108.0
Trees per acre	486.7	147.1	78.0	1002.0
BA (ft^2) per acre	98.0	53.0	0.5	222.1
Quadratic mean diameter (in.)	6.0	2.4	0.1	13.4
dbh (in.)	5.8	2.8	0.1	20.0
Total tree height (ft)	40.2	18.8	5.0	102.0

Table 2. Observed individual tree and stand characteristics for East Texas slash pine plantation dataset (n = 42,697 observations from 84 remeasured permanent plots across eight measurement cycles).

Variables	Mean	Standard deviation	Minimum	Maximum
Age (yr)	13.6	6.7	2.0	37.0
Dominant height (ft)	43.5	19.9	3.0	92.0
Site Index (25-yr, ft)	85.9	25.9	15.0	110.0
Trees per acre	436.0	183.5	22.0	1002.0
BA (ft^2) per acre	73.5	45.9	0.5	184.6
Quadratic mean diameter (in.)	5.6	2.4	0.1	12.6
dbh (in.)	5.4	2.7	0.1	18.3
Total tree height (ft)	37.2	18.8	5.0	97.0

Table 3. Parameter estimates and fit statistics of the mixed-effects height-diameter Equation 2 for East Texas loblolly and slash pine plantations.

Species	Parameter	Parameter estimate	Standard error	Pr (parameter = 0)
Loblolly	β_5	0.026490	0.002516	<0.0001
	β_6	0.449000	0.013880	<0.0001
	β_7	−0.019530	0.008019	0.0159
	$s^2(u_5)$	0.001091	0.000123	<0.0001
	$s^2(u_6)$	0.031320	0.004094	<0.0001
	$s^2(u_7)$	0.010560	0.001288	<0.0001
	$s^2(e)$	0.009614	0.000039	<0.0001
Slash	β_5	0.051750	0.006225	<0.0001
	β_6	0.498000	0.02148	<0.0001
	β_7	0.005485	0.01242	0.6599
	$s^2(u_5)$	0.003130	0.000505	<0.0001
	$s^2(u_6)$	0.032490	0.007194	<0.0001
	$s^2(u_7)$	0.011040	0.002149	<0.0001
	$s^2(e)$	0.010790	0.000074	<0.0001

would be biased. The mixed-model approach uses maximum likelihood (ML) estimation, which is appropriate for data with this serially correlated structure.

This study examined a model developed by Budhathoki et al. (2008) that includes dominant height and basal area (BA) per acre as well as individual tree diameter to predict individual tree total height. A random effect is included for tree diameter:

$$H_{ij} = \beta_0(Hd_i)^{\beta_1} e^{(-(\beta_2+u_{2i})\, D_{ij}^{\beta_3} + \beta_4 BA_i)} + \varepsilon_{ij}, \qquad (1)$$

where H_{ij} is the total tree height (ft) of tree j in plot i, Hd_i is the dominant height (ft) in plot i, D_{ij} is the dbh (in.) of tree j in plot i, BA_i is the basal area (ft^2) per acre of plot i, β_k is the fixed-effects regression parameters, u_i is the plot-specific random effects, and ε_{ij} is the random error of tree j in plot i, $\varepsilon \sim N(0, \sigma^2)$.

We also examined the logarithmic height–diameter model of Lenhart (1968) that uses both individual-tree and stand-level variables to predict tree height. Lee and Coble (2006) and Coble and Lee (2008) also used this model to estimate height from diameter for loblolly and slash pine, respectively. In this study, one to three random effects were included to account for plot-level variation. However, initial screening suggested that the model with three random effects worked best:

$$\ln(H_{ij}) = \ln(Hd_i) + (\beta_5+u_{5i}) + (\ln(D_{ij}) - \ln(D_{\max_i}))$$
$$* ((\beta_6 + u_{6i}) + (\beta_7 + u_{7i}) * \ln(Dq_i)) + \varepsilon_{ij},$$

or after rearranging,

$$\ln\left(\frac{H_{ij}}{Hd_i}\right) = (\beta_5+u_{5i}) + (\beta_6+u_{6i})$$
$$* \ln\left(\frac{D_{ij}}{D_{\max_i}}\right) + (\beta_7 + u_{7i})$$
$$* \ln\left(\frac{D_{ij}}{D_{\max_i}}\right) * \ln(Dq_i) + \varepsilon_{ij}, \qquad (2)$$

where $D_{\max_i}$ is the maximum observed diameter (in.) in plot i, Dq_i is the quadratic mean diameter (in.) of plot i, and ln is the natural logarithm. All other variables are defined as before.

Avery and Burkhart (2002) and Clutter et al. (1983) presented a modified form of the Korf equation that has been widely used to predict individual tree height from individual tree diameter alone. We added two random effects to their model for this study,

$$\ln(H_{ij}) = (\beta_8 + u_{8i}) + (\beta_9 + u_{9i})D_{ij}^{-1} + \varepsilon_{ij}, \qquad (3)$$

where all variables are defined as before.

The NLMIXED procedure (for Equation 1) and MIXED (for Equations 2 and 3) of the SAS Institute, Inc. (SAS Institute, Inc., 2000–2004) were used to estimate the model parameters. The NLMIXED procedure uses ML while the MIXED procedure uses restricted/residual ML (REML) to estimate the parameters and their variances. The best model was chosen based on Furnival's Index of Fit (FI; Furnival 1961). This index is a modified likelihood criterion that reflects both the size of the residuals and the possible departures from normality and homoscedasticity. FI can be used to compare any number of models where the dependent variable (Y') represents different transformations of the original dependent variable. Lower values of FI indicate a better fit model. FI is defined as

$$FI = [f'(Y)]^{-1}\sqrt{MSE},$$

where $f'(Y)$ is the first derivative of the transformed dependent variable with respect to Y, and $[f'(Y)]^{-1}$ is $[z]$ = geometric mean of z,

$$z = \exp\left[\frac{\sum_{i=1}^{n} \ln z_i}{n}\right].$$

Results and Discussion

Based on FI, the mixed-effects model of Lenhart (1968) (Equation 2) performed best for individual loblolly and slash pine tree total height prediction (Tables 3 and 4). For loblolly pine, the pa-

Table 4. FI values for the three height-diameter models considered in this study: (1) Budhathoki et al. (2008), (2) Lenhart (1968), and (3) Korf lower values of FI indicate a better fit.

Model	Loblolly	Slash
Budhathoki et al. (2008) (Equation 1)	3.6854	3.6921
Lenhart (1968) (Equation 2)	0.0813	0.0863
Korf equation (Equation 3)	12.4055	11.6273

rameter estimates for both individual-tree (i.e., diameter) and stand-level (i.e., dominant height, quadratic mean diameter, and maximum observed diameter) independent variables were significant at the 0.05 level. The results were similar for slash pine, except the coefficient with quadratic mean diameter was not significant (P = 0.6599; Table 3). However, quadratic mean diameter was retained in the model rather than removed to refit the model. Removal of this term would alter the theoretical form of the model, which is not something we desire. Perhaps the nature of the slash pine data caused the nonsignificant result, rather than a limitation of the theoretical model, which we believe should include a measure of average tree size. The residual plots for both the loblolly and the slash pine predictions did not indicate any departure from normality or nonconstant variance (Figures 1 and 2).

The data used in this study were sampled in a two-level hierarchical design. Individual trees were repeatedly measured through time on the same plots. Thus, random effects for plots (spatial autocorrelation) as well as for the repeatedly measured trees (serial autocorrelation) are theoretically necessary to characterize the variance–covariance structure of the loblolly and slash pine data sets. However, we found (results not shown) that predictions from fitting the model in the two-level design were not statistically different from predictions from a one-level design that considered only the random effects for plots (spatial autocorrelation). In fact, in some cases, the predictions were worse. Other researchers have found similar results for mixed-effects prediction models (Trincado and Burkhart 2006, Huang 2009). Trincado and Burkhart (2006) concluded that the inclusion of the plot-level random effects accounted for most of the serial autocorrelation associated with repeatedly measured trees. Because we wanted to develop a prediction model with high precision rather than test statistical hypotheses, we chose the mixed-effects height–diameter model that included random effects for plots only. This worked best for our data set, but other researchers should investigate the variance–covariance structure of their data before arbitrarily making the same conclusion.

An application example serves to show the benefits from using mixed-effects models to calibrate a population-level (fixed-effects) model to local stands. Ten loblolly pine trees with observed dbh and total heights were randomly selected from a data set that was independent of the fitting data set. These 10 trees were measured during a timber cruise of a 23-year-old, twice-thinned loblolly pine plantation, with a site index =70 ft, 78 tpa, 58 ft^2/ac, dominant height = 67 ft, Dq = 11.7 in., and D_{max} = 19 in.

Using Equation 2, the random effects for this stand can be predicted with the following equation presented in matrix format (Lappi 1991, Schabenberger and Pierce 2002):

$$\hat{\mathbf{u}} = (\mathbf{Z}'\hat{\mathbf{R}}^{-1}\mathbf{Z} + \hat{\mathbf{G}}^{-1})^{-1}\mathbf{Z}'\hat{\mathbf{R}}^{-1}(\mathbf{y} - \mathbf{Xb}), \qquad (4)$$

where $\hat{\mathbf{u}}$ is the vector of predicted random effects for Equation 2; $\mathbf{Z}$ is the derivative matrix of the fixed-effects design matrix with respect to the random effects evaluated at their expected value of zero (Sharma and Parton 2007); $\hat{\mathbf{R}}$ is the predicted variance–covariance matrix for the residual errors, $s^2(e)$, of individual trees in Table 3; $\hat{\mathbf{G}}$ is the predicted variance–covariance matrix of the random effects; $\mathbf{y}$ is the vector of observed individual tree heights; $\mathbf{X}$ is the vector of independent variables D_{ij}, Dq_i, and D_{max_i}, and $\mathbf{b}$ is the vector of estimated fixed-effects parameters, β, in Table 3.

For this example, these matrix values become

$$\hat{\mathbf{G}} = \begin{bmatrix} 0.001091 & 0.000510 & 0.000497 \\ 0.000510 & 0.03132 & -0.01655 \\ 0.000497 & -0.01655 & 0.01056 \end{bmatrix}$$

$$\hat{\mathbf{R}} = 0.009614X \underset{10\times 10}{\mathbf{I}}$$

$$(\mathbf{y} - \mathbf{Xb}) = \begin{bmatrix} -0.001715419 \\ 0.018151415 \\ 0.041705080 \\ 0.104006896 \\ 0.141192455 \\ 0.134086122 \\ 0.165757729 \\ 0.142749279 \\ 0.098355953 \\ 0.069759128 \end{bmatrix}$$

$$\mathbf{Z} = \begin{bmatrix} 1 & -0.546543706 & -1.344272802 \\ 1 & -0.641853886 & -1.578696656 \\ 1 & -0.747214402 & -1.837840205 \\ 1 & -0.99852883 & -2.455970369 \\ 1 & -0.410742165 & -1.010256846 \\ 1 & -0.546543706 & -1.344272802 \\ 1 & -0.747214402 & -1.837840205 \\ 1 & -0.30538165 & -0.751113298 \\ 1 & -0.379489622 & -0.933388439 \\ 1 & -0.087968773 & -0.216367012 \end{bmatrix}$$

These matrices along with equation (4) lead to the predicted random-effects:

$$\hat{\mathbf{u}} = \begin{bmatrix} 0.015371757 \\ 0.110078673 \\ -0.080127887 \end{bmatrix}.$$

These random effects were added to the estimated fixed-effects parameters, $\mathbf{b}$ (Table 3), to obtain stand-specific parameter estimates for this particular loblolly pine plantation,

$$\mathbf{b} + \hat{\mathbf{u}} = \begin{bmatrix} 0.02649 + 0.015371757 \\ 0.4490 + 0.110078673 \\ -0.01953 - 0.080127887 \end{bmatrix} = \begin{bmatrix} 0.04186176 \\ 0.55907867 \\ -0.09965789 \end{bmatrix}.$$

Thus, the final mixed-effects height–diameter equation for this loblolly stand is

$$\ln\left(\frac{H_{ij}}{Hd_i}\right) = (0.04186176) + (0.55907867) * \ln\left(\frac{D_{ij}}{D_{max_i}}\right) - (0.09965789) * \ln\left(\frac{D_{ij}}{D_{max_i}}\right) * \ln(D_{qi}). \qquad (5)$$

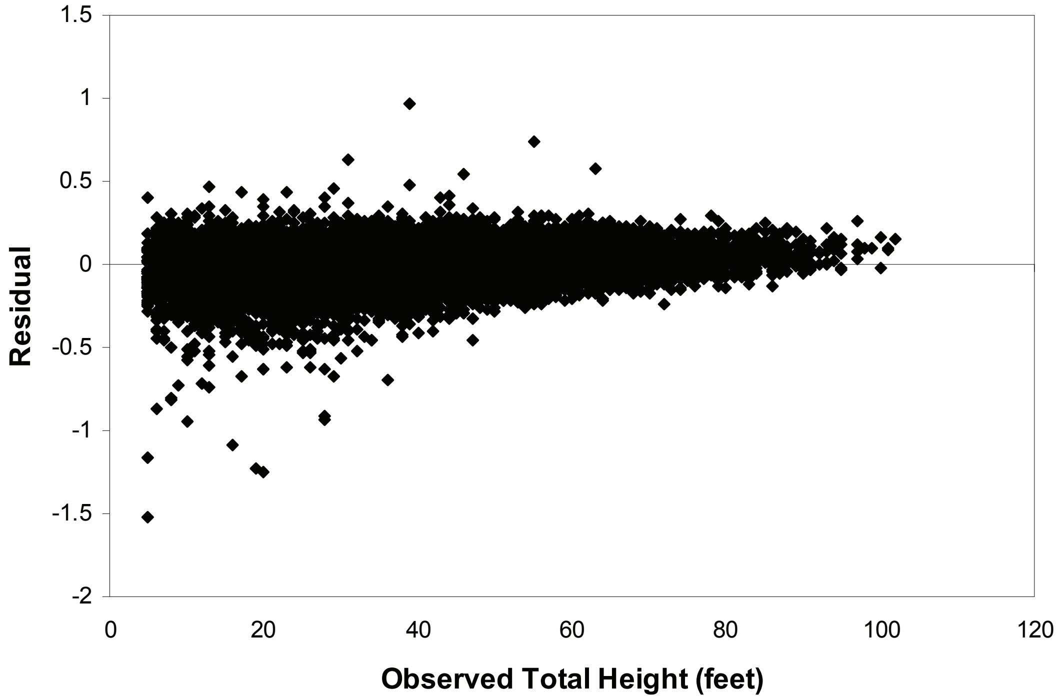

Figure 1. Plot of residuals for predicted heights of East Texas loblolly pine trees from Equation 2.

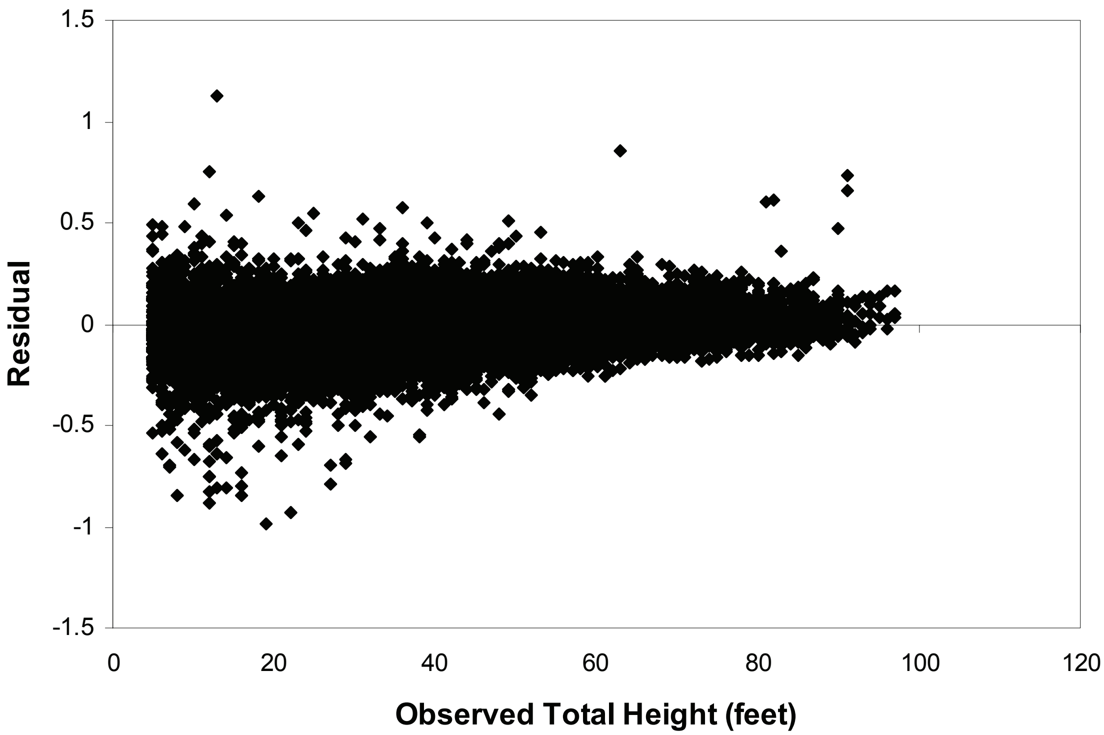

Figure 2. Plot of residuals for predicted heights of East Texas slash pine trees from Equation 2.

The total height predictions from Equation 5 were plotted with the fixed-effects predictions and observed values for the 10 calibration trees to show the improvement by using a mixed-effects height–diameter model (Figure 3). A similar plot was created for themeasured trees not used for calibration to show that the predicted heights from the mixed-effects model versus the fixed-effects model are still more closely aligned to the observed heights (Figure 4). A log-transformation bias correction factor, $c = \sigma^2/2$ (Baskerville 1972), was also applied to Equation 5 to account for conversion from log space to absolute units (ft). The mixed-effects total height predictions are more aligned with the observed values than the fixed-effects total height predictions, which shows the benefits of local calibration when observed height–diameter pairs are available. The sums of squares error (SSE) for total height predictions also illustrate

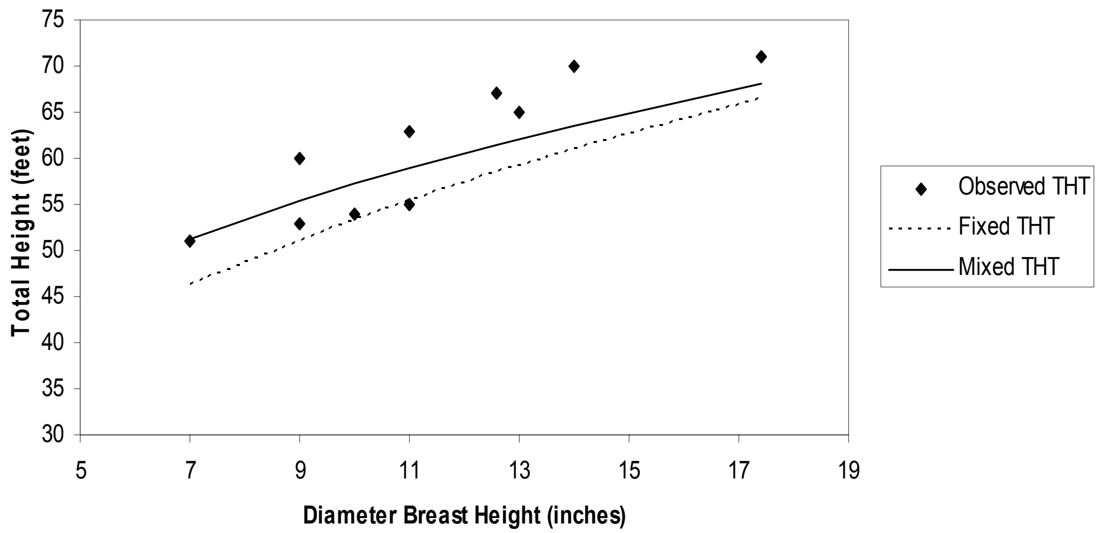

Figure 3. Predicted versus observed East Texas loblolly pine tree total heights for calibration trees only with and without calibration of random-effects parameters for Equation 2.

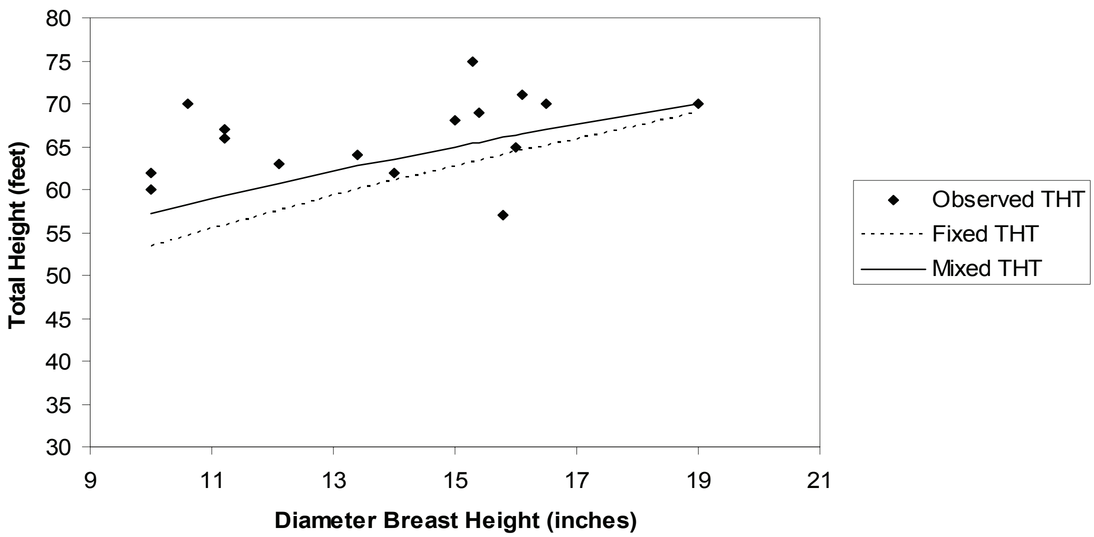

Figure 4. Predicted versus observed East Texas loblolly pine tree total heights for noncalibration trees only with and without calibration of random-effects parameters for Equation 2.

the benefits of using a mixed-effects model. The SSE for the fixed-effects predictions was 373.9 versus 251.1 for the mixed-effects predictions. This represents a 33% improvement for using the 10 calibration trees to localize the model.

The random selection of 10 trees for this independent timber cruise was arbitrary. Lynch et al. (2005) used 10 calibration trees selected randomly from a stand. Trincado et al. (2007) randomly selected one to three calibration trees per sample plot, although they found that one calibration tree was sufficient. Foresters often measure one tree for height per cruise plot when subsampling for heights in a forest inventory. Additional trees may be measured to ensure that the range of diameters is represented in the subsample. These subsampled trees are then typically pooled to develop a local height–diameter equation for the stand. If a mixed-effects model is available, then a height–diameter equation localized for each cruise plot can be estimated. On the other hand, if calibration trees are selected within diameter classes rather than cruise plots, then separate mixed-effects models can be developed for each diameter class. Furthermore, if no individual tree measurements are available, then the average diameter and height could conceivably be used to calibrate a mixed height–diameter model. We were unable to find any objective guidelines for the selection of calibration trees in the literature, which is why we arbitrarily chose 10 calibration trees for our example. Further research is needed to develop guidelines for calibration tree selection for use with mixed-effects height–diameter models.

In conclusion, the mixed-effects model that includes stand-level independent variables (Equation 2) best predicts total height in East Texas loblolly and slash pine trees. We believe the inclusion of stand-level variables (dominant height, quadratic mean diameter, and maximum diameter) is justified considering that they are readily available from any timber cruise. We compared the new mixed-effects equations with independently cruised stands to show how well mixed-effects models predict total height with few (10 in this study) calibration trees. This methodology would be ideal for timber cruises that routinely subsample for total height but obtain diameters for all sampled trees.

Literature Cited

Avery, T.E., and H.E. Burkhart. 2002. *Forest measurements*, 5th Ed. McGraw-Hill, Inc., New York. 456 p.

Baskerville, G.L. 1972. Use of logarithmic regression in the estimation of plant biomass. *Can. J. For. Res.* 2:49–53.

Budhathoki, C.B., T.M. Lynch, and J.M. Guldin. 2008. A mixed-effects model for the dbh-height relationship of shortleaf pine (*Pinus echinata* Mill.). *South. J. Appl. For.* 32(1):5–11.

Calama, R., and G. Montero. 2004. Interregional nonlinear height–diameter model with random coefficients for stone pine in Spain. *Can. J. For. Res.* 34:150–163.

Clutter, J.L., J.C. Fortson, L.V. Pienaar, G.H. Brister, and R.L. Bailey. 1983. *Timber management: A quantitative approach*. John Wiley and Sons, New York. 333 p.

Coble, D.W., and Y.J. Lee. 2008. A new diameter distribution model for unmanaged slash pine plantations in east Texas. *South. J. Appl. For.* 32(2):89–94.

Furnival, G.M. 1961. An index for comparing equations used in constructing volume tables. *For. Sci.* 7:337–341.

Huang, S., S.X. Meng, and Y. Yang. 2009. Prediction implications in nonlinear mixed-effects forest biometric models estimated with a generalized error structure. P. 1174–1188 in *Proc. of Joint Statistical Meetings, Section on Statistics and the Environment*, Washington, DC, Aug. 1–6, 2009. American Statistical Association, Alexandria, VA.

Lappi, J. 1991. Calibration of height and volume equations with random parameters. *For. Sci.* 37(3):781–801

Lappi, J. 1997. A longitudinal analysis of height/diameter curves. *For. Sci.* 43(4):555–570.

Lee, Y.J., and D.W. Coble. 2006. A new diameter distribution model for unmanaged loblolly pine plantations in east Texas. *South. J. Appl. For.* 30(1):13–20.

Lenhart, J.D.1968. *Yields of old-field loblolly pine plantations in the Georgia Piedmont.* Unpublished PhD dissertation, Univ. of Georgia, Athens, GA. 98 p.

Lenhart, J.D., E.V. Hunt, Jr., and J.A. Blackard. 1985. Establishment of permanent growth and yield plots in loblolly and slash pine plantations in east Texas. P. 436–437 in *Proc. of the 3rd Biennial Southern Silvicultural Research Conf.*, 1985, Shoulders, E. (ed.). US For. Serv. Gen. Tech. Rep. SO-54.

Lynch, T.B., A.G. Holley, and D.J. Stevenson. 2005. A random-parameter height–dbh model for cherrybark oak. South. *J. Appl. For.* 29:22–26.

Mehtatalo, L. 2004. A longitudinal height–diameter model for Norway spruce in Finland. *Can. J. For. Res*. 34:131–140.

SAS Institute, Inc. 2000–2004. *SAS 9.1.3 Help and documentation*. SAS Institute, Inc., Cary, NC.

Schabenberger, O., and F.J. Pierce. 2002. *Contemporary statistical models for the plant and soil sciences.* CRC Press, Boca Raton, FL. 738 p.

Sharma, M., and J. Parton. 2007. Height–diameter equations for boreal tree species in Ontario using a mixed-effects modeling approach. *For. Ecol. Manag.* 249:187–198.

Trincado, G., and H.E. Burkhart. 2006. A generalized approach for modeling and localizing stem profile curves. *For. Sci.* 52(6):670–682.

Trincado, G., C.L. VanderSchaaf, and H.E. Burkhart. 2007. Regional mixed-effects height–diameter models for loblolly pine (*Pinus taeda* L.) plantations. *Eur. J. For. Res.* 126: 253–262.

Zeide, B. 1989. Accuracy of equations describing diameter growth. *Can. J. For. Res.* 19(9):1283–1286.

Zeide, B. 1993. Analysis of growth equations. *For. Sci.* 39(3):594–616.