A MIXED-EFFECTS HEIGHT-DIAMETER MODEL FOR LONGLEAF PINE PLANTATIONS IN NORTHERN FLORIDA AND GEORGIA

Curtis L. VanderSchaaf and T. Eric McConnell

Abstract—A mixed-effects height-diameter model is presented for longleaf pine (*Pinus palustris* P. Mill.) plantations in northern Florida and Georgia. After obtaining height-diameter measurements from a plot/stand of interest, this mixed-effects model can be calibrated to produce localized individual tree height estimates. Based on model calibration of independent data from South Carolina, the use of three or four longleaf pine trees from a plot to calibrate the model provides a reasonable compromise between predictive ability and field sampling time. If calibrated at the stand-level, three trees could be used, but larger sample sizes of 10 or 15 would likely produce more accurate estimates. To calibrate these models for specific plots/stands, an Excel spreadsheet is available on request.

INTRODUCTION

Height (H)-diameter (D) models are an integral component of forest inventories used to reduce sampling times. Mixed-effects H-D models have been developed for many species. The advantages and rationale behind using mixed-effects H-D models is provided in this publication and others (Calama and Montero 2004; Lynch and others 2005; Trincado and others 2007; VanderSchaaf 2012, 2013, 2014). Longleaf pine (*Pinus palustris* P. Mill.) plantations are an important forest type in northern Florida and Georgia. There are 593,267 acres of longleaf pine plantations, fairly evenly distributed among both States. To increase the efficiency of forest inventories in longleaf pine plantations in northern Florida and Georgia, an individual tree mixed-effects H-D model was developed.

METHODOLOGY

Data Used in Model Fitting

The data used in model development were obtained from the U.S. Department of Agriculture, Forest Service, Forest Inventory and Analysis (FIA) annual surveys completed between 2011 and 2015. Survey data were obtained from all forested regions of Florida and Georgia. Data were obtained from the FIA database website (O'Connell and others 2017, USDA Forest Service 2017). Plots were clusters of four points arranged such that point 1 was central, with points 2 through 4 located 120 feet from point 1 at azimuths of 0, 120, and 240 degrees (Bechtold and Scott 2005). Each cluster point was surrounded by a 24-foot fixed-radius subplot where trees 5.0 inches in diameter at breast height (d.b.h.) and larger were measured. Combined, the four subplots totaled approximately 0.17 acres. Each subplot contained a 6.8-foot fixed-radius microplot where saplings (1.0-4.9 inches d.b.h.) were measured. The four microplots totaled approximately 0.01 acres. Each plot was only measured once; hence, there were no concerns about repeated measures or serial correlation.

Since trees with a d.b.h. of 5 inches or larger have a greater probability of being sampled, there is an increase in the number of trees relative to smaller diameters (fig. 1). Data from only those plots where longleaf pine comprised at least 60 percent of the total basal area were included in the model fitting dataset. A value of 60 percent basal area was selected to help reduce the inclusion of "wildlings," which are pine trees that were not planted, in the model fitting and validation analyses. Additionally, a value of 60 percent basal area reduces the impacts of hardwoods on the H-D relationships and hence produces more uniformity in H estimates when used operationally in appropriate plantations. Only H and D measurements of longleaf pine were modeled. Individual tree and plot-level summary data are presented in tables 1 and 2.

Author information: Curtis L. VanderSchaaf, Assistant Professor, Louisiana Tech University, Ruston, LA 71272; and T. Eric McConnell, Assistant Professor, Louisiana Tech University, Ruston, LA 71272.

Citation for proceedings: Bragg, Don C.; Koerth, Nancy E.; Holley, A. Gordon, eds. 2020. Proceedings of the 20th Biennial Southern Silvicultural Research Conference. e–Gen. Tech. Rep. SRS–253. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 338 p.

Figure 1—Height-diameter relationship for the model fitting dataset (n = 818 trees).

Table 1—Tree-level summary statistics of trees used in model fitting and model calibration/validation

		Diameter (inches)				Height (feet)			
Model	*n*	Minimum	Mean	Maximum	SD	Minimum	Mean	Maximum	SD
Fit	818	1.1	6.7	17.9	2.04	10	44	88	13.9
Validation	938	1.2	6.5	20.1	1.84	9	40	83	9.9

SD is standard deviation, *n* is number of trees.

Table 2—Model fitting and validation plot-level means

		Species of interest			Other species		
Model	Number of plots	TPA	D_q	BAA	TPA	D_q	BAA
Fit	45	221	5.9	36	134	4.6	7
Validation	27	422	5.4	68	155	4.1	14

TPA = trees per acre, D_q = quadratic mean diameter (inches), and BAA = square feet of basal area per acre.

Species of interest refers to plot-level summary statistics for longleaf pine used in fitting the height-diameter equations. Other species refers to hardwoods as well as other conifers.

Model Development and Parameter Estimation

Total tree height was predicted as a function of diameter at breast height:

$$\ln H_{ki} = (\beta_0 + u_{0k}) + (\beta_1 + u_{1k})\ln D_{ki} + \varepsilon_{ki} \quad (1)$$

where

ln = natural logarithm,
H_{ki} = individual tree total height (feet) for tree *i* within plot (or stand) *k*,
D_{ki} = individual tree diameter at breast height (inches) for tree *i* within plot (or stand) *k*,
β_0, β_1 = parameters to be estimated,
u_{0k}, u_{1k} = plot/stand-specific random effects, assumed to be N(0, $\sigma^2{}_0$) and N(0, $\sigma^2{}_1$), respectively,
(β_0+u_{0k}) = plot/stand-specific intercept,
(β_1+u_{1k}) = plot/stand-specific slope, and
ε_{ki} = random error where it is assumed ε ~N(0, σ^2I).

Additionally, a covariance, σ_{01}, can be assumed to exist between u_{0k} and u_{1k}. Linear mixed-effects models, in this particular case, produce an efficient estimate of plot/stand-specific parameters because only six parameters are estimated using the model fitting algorithm ($\beta_0,\beta_1,\sigma^2{}_0,\sigma^2{}_1,\sigma_{01},\sigma^2$).

Based on the variance and covariance estimates, plot/stand-specific random effects (u_{0k}, u_{1k}) can be predicted and then added to the "population average" intercept and slope (β_0, β_1) estimates to obtain plot/stand-specific parameter estimates. Plot/stand-specific random parameters produce a more localized H-D equation since the random effects account for local site conditions such as soil type, genetic stock, site preparation, mid-rotation silvicultural practices, spatial and time-specific climatic conditions, etc. The prediction of plot/stand-specific random effects is conducted outside the model fitting algorithm and thus degrees of freedom are not lost due to specific plot/stand random effects. In terms of model fitting, a less efficient method of obtaining plot-specific parameter estimates would be to estimate parameters separately for each plot using ordinary least squares or weighted least squares.

Although the parameter estimation efficiency of mixed models is an advantage, the greatest advantage often is the ability to calibrate the model using data independent of those used in model fitting. Hence, for trees from plots/stands not in the model fitting dataset, plot/stand-specific H-D relationships can be produced if H and D observations have been collected from trees in that plot/stand.

When compared to the traditional means of developing local H-D equations, where H and D are measured and then a separate equation is fit for a stand (or in some cases a plot), a mixed-effects model analysis is efficient because a model can be calibrated without having to statistically fit a model and thus even small sample sizes can be used (Lynch and others 2005). In many cases, fixed-effects region-wide H-D equations are used to predict tree heights. To account for growing condition differences among stands for the same species, in addition to D, these models contain measures of site quality and/or stand density, among others. Studies have shown that a calibrated mixed-effects H-D model often produces better predictions than a region-wide H-D model containing stand-level regressors (Huang and others 2009, Temesgen and others 2008, Trincado and others 2007). Mixed-effects models fit using plot-level data can also be calibrated at the plot-level during operational inventories, providing more localized H-D relationships within a stand.

The objective was to measure H on every tree but occasionally due to various reasons, H was not directly measured. Only trees whose heights were actually measured (as opposed to visually estimated or predicted using equations) were used in model fitting. Parameters of equation (1) were estimated using SAS Proc MIXED (Littell and others 1996) which assumes random errors are normally distributed and subsequently estimates parameters using maximum likelihood. To reduce complexity in model application, no attempt was made to include spatial correlation in the models. The random error covariance-variance matrix was assumed to be $\sigma^2 I_{nk}$.

Data Used in Model Calibration

To gain insight into the optimal calibration sample size, independent FIA survey data from South Carolina obtained from 2012 to 2016 were used. Individual tree and plot-level summary data are presented in tables 1 and 2.

A minimum sample size of 10 trees per plot was selected. Hence, a plot had to contain at least 12 trees of longleaf pine to be included, in this case 10 trees would be used to calibrate the model and the 11th and 12th trees would be predicted; at least two predicted trees allows for calculation of the variance. Calibration sample sizes of 1, 2, 3, 4, 5, 8, and 10 were examined.

For consistency, the tree used to calibrate the model for a sample size of one was also the first tree for a sample size of two, the two trees used to calibrate the model for a sample size of two were the first two trees for a sample size of three, etc. Calibrations were conducted 50 times for each sample size to avoid the dependence of the calibration results on one particular sample.

Validation analyses follow those presented in Trincado and others (2007). The difference between the observed (H_{obs}) and predicted height (H_{pred}) of all trees (*i*) whose Hs were predicted for each individual plot (*k*) and for each of the 50 replications (r), or each plot (k) and replication (r) combination, was calculated ($e_{kri} = H_{obs\ kri} - H_{pred\ kri}$), trees used in calibration for a particular plot and replication combination were not included in the validation statistic calculations. The mean residual ($\bar{e}$) and the sample variance (v) of residuals were computed and considered to be estimates of bias and precision; respectively. An estimate of mean square error (MSE) was obtained for each combination by combining the bias and precision measures using the following formula:

$$MSE_{kr} = \bar{e}^2_{kr} + v_{kr} \tag{2}$$

The 50 MSE, $\bar{e}$, and v values of a particular plot for a particular calibration sample size were then averaged to obtain an average value of the three calibration statistics for each plot and sample size. The average plot calibration statistics were then averaged for each calibration sample size to obtain final calibration statistics. To account for the transformation bias, the procedure recommended by Baskerville (1972) was used as described in the calibration example provided below.

RESULTS AND DISCUSSION

Based on the model fitting results (not presented for the sake of brevity), it is best to assume that both β_0 and β_1 are random (or, essentially, that each plot (or stand) has their own intercept and slope) and that a covariance (σ_{01}) exists between u_{0k} and u_{1k}. Table 3 contains parameter estimates and model-fitting statistics. In terms of choosing an optimal model calibration sample size at the plot- or stand-level, a reasonable trade-off between statistical measures (precision and accuracy) and sampling times (e.g., costs) is three or four trees (fig. 2).

Table 3—Population average (β_0 and β_1) and random effects variance (σ^2_0, σ^2_1) and covariance (σ_{01}) parameter estimates

Parameter	Estimate	Standard error
β_0	2.6575	0.08188
β_1	0.5666	0.03464
σ^2_0	0.2475	-
σ^2_1	0.03844	-
σ_{01}	-0.09177	-
-2LL	-1249.3	
AIC	-1241.3	
σ^2	0.009380	
n	45	

-2LL = twice the negative log-likelihood (smaller is better), AIC = Akaike's Information Criterion (smaller is better), σ^2 = estimated mean square error, and n = the number of clusters, or plots.

In a sense, using one tree in calibration allows for the "population average" curve to be moved up and down based on that tree's relationship to the "population average" curve. Figure 3 helps to provide context. The use of two trees in calibration essentially allows for the curve to be moved up and down, as well as a reshaping of the curve, whereas the use of three or more trees will in a sense refine the movement and reshaping of the "population average" curve.

Studies have shown that even the use of only one tree in calibration substantially improves height estimates (Calama and Montero 2004; Huang and others 2009; Temesgen and others 2008; Trincado and others 2007; VanderSchaaf 2012, 2013, 2014). Thus, when calibrating at the plot-level, one tree may suffice. However, if calibrating at the stand-level it may be feasible to use larger sample sizes of 10 or 15 trees.

EXAMPLE OF MODEL CALIBRATION

For clarity and ease of application, the methodology to predict random effects is presented. Nomenclature is based on Schabenberger and Pierce (2002: 431). The expression used to predict random effects, Estimated Best Linear Unbiased Predictors (EBLUPS), is:

$$\hat{\boldsymbol{\mu}}_k = \hat{\mathbf{D}}\mathbf{Z}_k^{\mathbf{T}}(\hat{\mathbf{R}}_k + \mathbf{Z}_k\hat{\mathbf{D}}\mathbf{Z}_k^{\mathbf{T}})^{-1}(\mathbf{y}_k - \mathbf{X}_k\hat{\boldsymbol{\beta}}) \tag{3}$$

where

$\hat{\boldsymbol{\mu}}_k$ = predicted random effects of plot/stand k, a 2 x 1 vector where 2 is the number of predicted random effects,

$\hat{\mathbf{D}}$ = estimated covariance-variance matrix (2 x 2) of the random effects,

$\mathbf{Z}_k$ = random effects matrix (n_k x 2) containing observed values of tree diameter (natural log transformed) from plot/stand k, and a column of 1s,

$\hat{\mathbf{R}}_k$ = estimated covariance-variance random error matrix expressed as $\sigma^2\mathbf{I}_{nk}$ since the variance is assumed constant across all plots/stands and tree heights are assumed temporally and spatially uncorrelated within a plot/stand,

$\mathbf{y}_k$ = n_k x 1 vector of observed heights (natural log transformed) from plot/stand k,

$\mathbf{X}_k$ = regressor matrix (n_k x 2) consisting of a column of 1s and the observed values of tree diameter (natural log transformed) from plot/stand k, and

$\hat{\boldsymbol{\beta}}$ = 2 x 1 vector of estimated fixed-effects parameters.

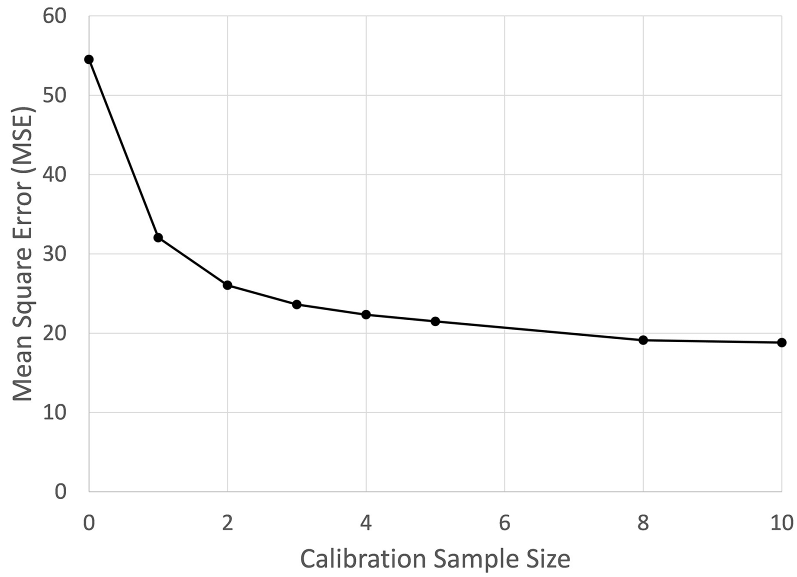

Figure 2—Model calibration mean square error (MSE) results. Calibration sample sizes are the number of trees used in calibration (n = 27 plots).

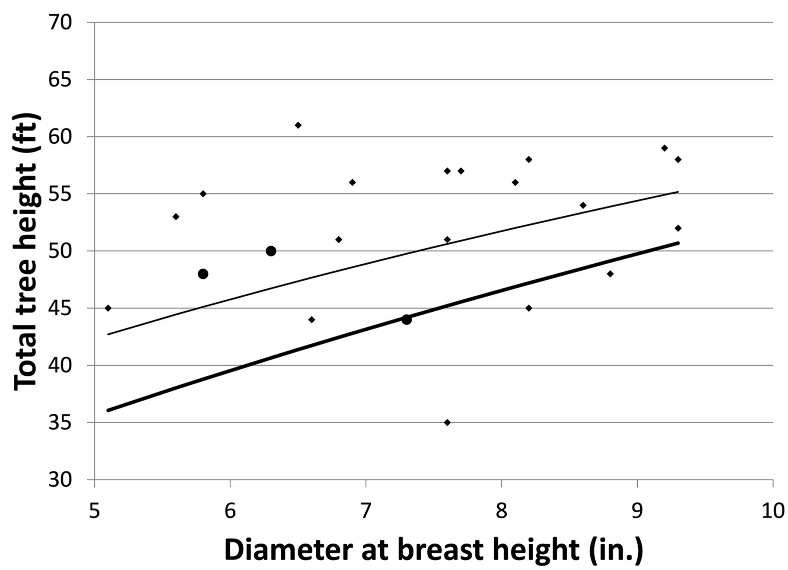

Figure 3—Observed total tree height and diameter at breast height for n = 22 trees. This plot resides in Charleston County, SC. The bold solid line is the "population average" estimate obtained by not calibrating equation (1), and the lighter solid line is the estimate following calibration of equation (1) producing equation (6). Black diamonds and black circles are paired height-diameter observations. Black circles were trees used to calibrate equation (1).

To demonstrate model calibration, three individual trees were randomly selected from an individual plot (or stand) to be used in calibration. Heights for the three trees were 48, 50, and 44 feet and the corresponding Ds were 5.8, 6.3, and 7.3 inches, respectively.

$$\mathbf{Z} = \begin{bmatrix} 1 & \ln(5.8) \\ 1 & \ln(6.3) \\ 1 & \ln(7.3) \end{bmatrix} = \begin{bmatrix} 1 & 1.757858 \\ 1 & 1.840550 \\ 1 & 1.987874 \end{bmatrix},$$

$$\mathbf{X} = \begin{bmatrix} 1 & \ln(5.8) \\ 1 & \ln(6.3) \\ 1 & \ln(7.3) \end{bmatrix} = \begin{bmatrix} 1 & 1.757858 \\ 1 & 1.840550 \\ 1 & 1.987874 \end{bmatrix}$$

Besides the column of 1s, all numerical values are lnD (naturally log transformed value of D in inches) for the three observations randomly selected from this particular plot (or stand).

$$\mathbf{y} = \begin{bmatrix} \ln(48) \\ \ln(50) \\ \ln(44) \end{bmatrix} = \begin{bmatrix} 3.871201 \\ 3.912023 \\ 3.784190 \end{bmatrix}$$

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 2.6575 \\ 0.5666 \end{bmatrix}$$

$$(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \begin{bmatrix} 3.871201 - (2.6575 + 0.5666[\ln(5.8)]) \\ 3.912023 - (2.6575 + 0\ 5666[\ln(6.3)]) \\ 3.784190 - (2.6575 + 0.5666[\ln(7.3)]) \end{bmatrix} =$$

$$\begin{bmatrix} 3.871201 - 3.653502 \\ 3.912023 - 3.700355 \\ 3.784190 - 3.783830 \end{bmatrix} = \begin{bmatrix} 0.217699 \\ 0.211668 \\ 0.000360 \end{bmatrix}$$

$$\hat{\mathbf{D}} = \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{bmatrix} = \begin{bmatrix} 0.2475 & -0.09177 \\ -0.09177 & 0.03844 \end{bmatrix}$$

$$\hat{\mathbf{R}} = \begin{bmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{bmatrix} = \begin{bmatrix} 0.00938 & 0 & 0 \\ 0 & 0.00938 & 0 \\ 0 & 0 & 0.00938 \end{bmatrix}$$

All numerical values for $\hat{\mathbf{D}}$, $\hat{\mathbf{R}}$, and $\hat{\boldsymbol{\beta}}$ were obtained from table 3 and were obtained from the model fitting results of equation (1). The dimensions of $\hat{\mathbf{R}}$, **Z**, **X**, **y**, and ($\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$) will change based on the number of observations used in model calibration.

When performing matrix operations as seen in equation (3), the following predictions of the random effects, EBLUPS, for β_0 and β_1 of this particular plot (or stand) were obtained:

$$\hat{\boldsymbol{\mu}} = \begin{bmatrix} \mu_{0k} \\ \mu_{1k} \end{bmatrix} = \begin{bmatrix} 0.3977 \\ -0.1404 \end{bmatrix}$$

These predicted random effects for this plot (or stand) are added to the "population average" parameter estimates, $\hat{\boldsymbol{\beta}}$, to obtain plot/stand-specific parameter estimates for this particular plot/stand, $\hat{\boldsymbol{\beta}}_{Calibrated}$:

$$\hat{\boldsymbol{\beta}}_{Calibrated} = \begin{bmatrix} \beta_0 + \mu_{0k} \\ \beta_1 + \mu_{1k} \end{bmatrix} = \begin{bmatrix} 2.6575 + 0.3977 \\ 0.5666 - 0.1404 \end{bmatrix} = \begin{bmatrix} 3.0552 \\ 0.4262 \end{bmatrix}$$

It is well known that logarithmic transformations often linearize data and produce homogeneity of variances; however, a transformation bias occurs since additive errors in log-log models become multiplicative when transformed back to the original scale. To account for the transformation bias, the procedure recommended by Baskerville (1972) should be used when predicting heights:

$$\ln H_{ki} = (\beta_0 + u_{0k}) + (\beta_1 + u_{1k})\ln D_{ki} + \sigma^2/2 \qquad (4)$$

where

σ^2 = mean square error (or residual variance) from the model fit (see table 3).

To obtain height predictions in the original units (feet), equation (5) should be used:

$$H_{ki} = \exp^{[\beta 0k + \beta 1k \ln D_{ki} + \sigma 2/2]} \qquad (5)$$

where

$\beta_{0k} = (\beta_0 + u_{0k})$ = plot/stand-specific intercept,
$\beta_{1k} = (\beta_1 + u_{1k})$ = plot/stand-specific slope, and
all other variables as previously defined.

For the example above, equation (5) would be expressed as:

$$H_{ki} = \exp^{[3.0552 + 0.4262 \ln D_{ki} + 0.00938/2]} \qquad (6)$$

Assuming a tree has a D of 7.1 inches, the predicted H would be 49 feet. Use of the simplified, "population average" curve, on the other hand, provides a predicted height of 43 feet, a 12 percent difference.

Figure 3 clearly demonstrates that calibrating equation (1) using observed Hs and Ds from a particular plot (producing equation (6)) vastly improved predictive ability, consistent with many other studies (e.g., Calama and Montero 2004; Huang and others 2009; Lynch and others 2005; Temesgen and others 2008; Trincado and others 2007; VanderSchaaf 2012, 2013, 2014). Relative to the "population average" curve, for a given D, most Hs for this particular plot were greater (perhaps due to genetic stock, stand-level species composition, aspect, soil type, site productivity, etc.). Hence, equation (6), through the calibration process, provided an H-D curve that reflected this behavior. However, using the so-called "population average" trend failed to recognize the behavior of trees in this individual plot relative to the average behavior across all plots.

CONCLUSIONS

A mixed-effects H-D model is presented for longleaf pine plantations in northern Florida and Georgia. By obtaining H-D measurements from plots/stands of interest, equation (1) can be calibrated to local site conditions. In most cases, model calibration will result in better H predictions relative to a fixed-effects region-wide model (Huang and others 2009; Temesgen and others 2008; Trincado and others 2007; VanderSchaaf 2012, 2013, 2014). Calibration of equation (1) will often be advantageous to fitting local H-D equations (or plot- or stand-specific equations) because calibration can be appropriate and have acceptable model fit statistics with small sample sizes (Lynch and others 2005). To calibrate these models for specific plots/stands, an Excel spreadsheet is available on request.

LITERATURE CITED

Baskerville, G.L. 1972. Use of logarithmic regression in the estimation of plant biomass. Canadian Journal of Forest Research. 2(1): 49-53.

Bechtold, W.A.; Scott, C.T. 2005. The Forest Inventory and Analysis plot design. In: Bechtold, W.A.; Patterson, P.L., eds. The enhanced Forest Inventory and Analysis program—national sampling design and estimation procedures. Gen. Tech. Rep. SRS-80. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station: 27-42.

Calama, R.; Montero, G. 2004. Interregional nonlinear height-diameter model with random coefficients for stone pine in Spain. Canadian Journal of Forest Research. 34(1): 150-163.

Huang, S.; Wiens, D.P.; Yang, Y. [and others]. 2009. Assessing the impacts of species composition, top height and density on individual tree height prediction of quaking aspen in boreal mixedwoods. Forest Ecology and Management. 258(7): 1235-1247.

Littell, R.C.; Milliken, G.A.; Stroup, W.W.; Wolfinger, R.D. 1996. SAS® system for mixed models. Cary, NC: SAS Institute Inc. 633 p.

Lynch, T.B.; Holley, A.G.; Stevenson, D.J. 2005. A random-parameter height-dbh model for cherrybark oak. Southern Journal of Applied Forestry. 29(1): 22-26.

O'Connell, B.M.; Conkling, B.L.; Wilson, A.M. [and others]. 2017. The Forest Inventory and Analysis database: database description and user guide for Phase 2 (version 7.0). U.S. Department of Agriculture, Forest Service. 830 p. https://www.fia.fs.fed.us/library/database-documentation/index.php. [date accessed month day, 2017].

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

Temesgen, H.; Monleon, V.J.; Hann, D.W. 2008. Analysis and comparison of nonlinear tree height prediction strategies for Douglas-fir forests. Canadian Journal of Forest Research. 38(3): 553-565.

Trincado, G.; VanderSchaaf, C.L.; Burkhart, H.E. 2007. Regional mixed-effects height-diameter models for loblolly pine (*Pinus taeda* L.) plantations. European Journal of Forest Research. 126(2): 253-262.

U.S. Department of Agriculture, Forest Service [USDA Forest Service]. 2017. FIA DataMart 1.6.1. St. Paul, MN: U.S. Department of Agriculture, Forest Service, Northern Research Station. https://apps.fs.usda.gov/fia/datamart/CSV/datamart_csv.html. [date accessed August 10, 2017].

VanderSchaaf, C.L. 2012. Mixed-effects height-diameter models for commercially and ecologically important conifers in Minnesota. Northern Journal of Applied Forestry. 29(1): 15-20.

VanderSchaaf, C.L. 2013. Mixed-effects height-diameter models for commercially and ecologically important hardwoods in Minnesota. Northern Journal of Applied Forestry. 30(1): 37-42.

VanderSchaaf, C.L. 2014. Mixed-effects height-diameter models for ten conifers in the inland Northwest, USA. Southern Forests: a Journal of Forest Science. 76(1): 1-9.