2009

A Mixed-effects Height-Diameter Model for Pinus densiflora Trees in Gangwon Province, Korea

Young Jin Lee

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

Jung Kee Pyo

Sung Ho Kim

Woo Kyun Lee

See next page for additional authors

Authors
Young Jin Lee, Dean W. Coble, Jung Kee Pyo, Sung Ho Kim, Woo Kyun Lee, and Jung Kee Choi

Jour. Korean For. Soc. Vol. 98, No. 2, pp. 178~182 (2009)

JOURNAL OF KOREAN FOREST SOCIETY

A Mixed-effects Height-Diameter Model for *Pinus densiflora* Trees in Gangwon Province, Korea

Young Jin Lee[1]*, Dean W. Coble[2], Jung Kee Pyo[3], Sung Ho Kim[3], Woo Kyun Lee[4] and Jung Kee Choi[5]

[1]*Department of Forest Resources, Kongju National University, Yesan, Chungnam 340-802, Korea*
[2]*Arthur Temple College of Forestry and Agriculture, Stephen F. Austin State University Nacogdoches, Texas, 75962, USA*
[3]*Division of Forest Resource Information, Korea Forest Research Institute, Seoul 130-712, Korea*
[4]*Division of Environmental Science and Ecological Engineering, Korea University, Seoul 136-701, Korea*
[5]*Division of Forest Management, Kangwon National University, Chuncheon 200-701, Korea*

Abstract : A new mixed-effects model was developed that predicts individual-tree total height for *Pinus densiflora* trees in Gangwon province as a function of individual-tree diameter (cm). The mixed-effects model contains two random-effects parameters. Maximum likelihood estimation was used to fit the model to 560 height-diameter observations of individual trees measured throughout Gwangwon province in 2007 as part of the National Forest Inventory Program in Korea. The new model is an improvement over fixed-effects models because it can be calibrated to a local area, such as an inventory plot or individual stand. The new model also appears to be an improvement over the Forest Resources Evaluation and Prediction Program for the ten calibration trees used in this study. An example is provided that describes how to estimate the random-effects parameters using ten calibration trees.

Key words : *height-diameter prediction, random-effects, mixed-effects, missing heights, Pinus densiflora*

Introduction

A result of the third panel in the 5th Korean National Forest Inventory (NFI) Program reports that forests cover 6.38 million ha or about 64% of total land area in Korea. The coniferous forest comprises nearly 42.3% (2.70 million ha) of the total forest cover. The broadleaved forest and mixed forest cover 25.9% (1.66 million ha) and 29.3% (1.87 million ha), respectively. *Pinus densiflora,* which is the most abundant tree species in Korea, covers about 24% of the total forest area, and about 17% of total *Pinus densiflora* forest area is located in Gangwon Province (Korea Forest Service, 2008). The current growing stock volume in Gangwon province measured by the NFI in 2007 is about 149.65 cubic meters per hectare, which is the highest growing stock in Korea (Kim *et al.*, 2008). Among the native tree species in Korea, *Pinus densiflora* is one of the most important tree species in terms of high-value wood products and our cultural and historical meanings in Korea. Therefore, the management of *Pinus densiflora* and the reliability of the growth model used to forecast the future growth of these forests is very important to forest managers in Korea.

The Korea Forest Research Institute (2004, 2008) developed and updated the individual-tree level and stand level growth prediction system for major tree species, called the Forest Resources Evaluation and Prediction Program (FREPP), to predict the growth and yield for major tree species in Korea, including *Pinus densiflora*. However, the height-diameter relationship for *Pinus densiflora* has not been extensively studied for this important species. Total tree height is often a difficult and expensive variable to measure on trees. Many tree species exhibit a strong correlation between total height and diameter that can be exploited to reduce inventory costs with little loss in precision. In a typical forest inventory, such as the NFI in Korea (Korea Forest Research Institute, 2008), all sampled trees are measured for diameter while a sub-sample of trees is measured for total height and diameter. These sub-sampled trees are used to develop a statistical model that predicts total height as a function of diameter. Then, this model is used to predict total height for the other trees that were not measured for height. This approach produces a model that is often

*Corresponding author
E-mail: leeyj@kongju.ac.kr

only applicable to the localized area for which it was developed. In contrast, height-diameter models can be developed for large geographic regions rather than localized areas. These models represent the population average height which may or may not accurately represent a specific localized area. The recent development of mixed-effect estimation techniques (Budhathoki *et al.*, 2008; Calama and Montero, 2004; Lappi, 1997; Lynch *et al.,* 2005; Mehtatalo, 2004; Sharma and Parton, 2007; Trincado *et al.*, 2007) has made it possible to build height-diameter models from regional databases, like the national inventory, that can be calibrated to local stands with a sub-sample of trees measured for total height and diameter. This type of model has the potential to greatly reduce the effort to sample total height of *Pinus densiflora.*

The objective of this study was to develop a mixed-effects height-diameter model for *Pinus densiflora* trees in Gangwon province. No such model has been developed for this region. In this study, we examined a new mixed-effects model that contains two random-effects parameters to determine the best predictive model for *Pinus densiflora* trees in Gangwon province.

Methods

Avery and Burkhart (2002) and Clutter *et al.* (1983) present an equation that has been widely used to predict average individual tree to height as a function of individual tree diameter (D) and two constants (b_0 and b_1):

$$\ln(H) = b_0 + b_1\ D^{-1} \quad (1)$$

where,

H = total tree height (m),

D = tree diameter (cm) at breast height (dbh),

b_i = population average fixed-effects regression parameters, and

ln(·) = natural logarithm function.

Equation 1 provides predictions that represent the population average total height estimate for individual trees. The two parameters (b_0 and b_1) are assumed to be fixed and do not vary across sampling units (i.e., plot, stand, forest) within a population. However, trees are often sampled on plots and plot-specific site characteristics such as soils, nutrients, and competition vary from plot to plot. So, a more flexible model would allow the two parameters b_0 and b_1 to vary by plot. Since plots are sampled randomly, b_0 and b_1 can also be considered random. Thus, cluster-specific random-effects can be added to the two population average fixed-effects parameters in equation 1 to produce a model with cluster-specific parameters:

$$\ln(H) = (b_0 + u_0) + (b_1 + u_1)D^{-1} \quad (2)$$

where, u_i = cluster-specific random-effects regression parameters and all other variables are defined as before. For linear mixed-effects models such as equation 2, the fitting process estimates five parameters: b_0 and b_1, σ_0^2 and σ_1^2 =variances of u_0 and u_1, respectively, and σ_{01} = covariance of u_0 and u_1. These five parameters along with residual error are used in a Bayesian methodology that is implemented outside of the model fitting process to determine the cluster-specific random-effects, u_0 and u_1. This has the advantage of not only statistical efficiency, but also the ability to calibrate the mixed-effects model with independent data from specific plots (Vandershaaf, 2008).

Maximum likelihood estimation was used to fit equation 2 to 560 height-diameter observations measured on *Pinus densiflora* trees in Gangwon province (Table 1). Plots ranged in age from 12 to 110 years old, and represented a wide range of site quality and density. The plots have been measured in 2007 as part of the NFI Program in Korea. For comparison, equation 1 was also fit to the 560 observations using ordinary least squares estimation. Ten individual *Pinus densiflora* trees on two plots (plot numbers 27240481 and 2764484 in the NFI database) were randomly selected and removed from the dataset used to estimate the regression coefficients. These ten trees were used later to examine the differences in predictions from the mixed-effects model (equation 2), the ordinary least squares model (equation 1), and the FREPP. The Akaike Information Criterion (AIC, Akaike, 1974) was also used to select the best model (equation 1 or equation 2). AIC is a statistical tool for selecting a best fit model from a number of candidates, where a smaller value equates to a better fit. It is a measure of goodness of fit that considers both model bias and precision. The MIXED procedure of SAS was used to estimate the regression parameters in equation 2, and the

Table 1. Observed individual tree characteristics for *Pinus densiflora* trees in Gangwon province (n = 560 observations from NFI plots across Gangwon province).

Variables	Mean	Standard Deviation	Minimum	Maximum
Age (years)	41.5	15.2	12.0	110.0
Diameter at Breast Height (cm)	25.6	10.3	6.0	75.0
Dominant Height (m)	13.1	3.9	2.2	26.4
Crown Height (m)	6.8	2.8	1.1	17.0

REG procedure of SAS was used to estimate the regression parameters in equation 1.

Results and Discussion

Based on the AIC statistic, the mixed-effects model (equation 2) performed best for individual *Pinus densiflora* tree total height prediction (Table 2). The regression parameter estimates were both significantly different from zero ($p<0.0001$). The residual error and the variance-covariance matrix values of the random-effects (cluster-specific) regression parameters are presented in Table 3. An application example serves to show the benefits from using mixed-effects models to calibrate a population-level (fixed-effects) model to specific plots. The example also shows the superiority of using mixed-effects models versus ordinary least squares models.

The random-effects parameters of equation 2 can be estimated with the following equation derived using Bayesian methodology that is presented in matrix format (Lappi, 1991; Schabenberger and Pierce, 2002):

$$u = (Z'\hat{R}^{-1}Z + \hat{D}^{-1})^{-1} Z'\hat{R}^{-1}(y - Xb) \quad (3)$$

where:

u = vector of predicted random-effects (cluster-specific) parameters,

Z = design matrix for the random-effects parameters,

$\hat{R}$ = predicted variance-covariance matrix for the residual errors of individual trees, e, in Table 2,

$\hat{D}$ = predicted variance-covariance matrix of the random-effects parameters using values in Table 3,

y = vector of observed individual tree heights,

X = vector of independent variables, D,

b = vector of estimated fixed-effects parameters in Table 2.

For this example, these matrix values become:

Table 2. Parameter estimates and fit statistics of the mixed-effects height-diameter equation for *Pinus densiflora* trees in Gangwon province.

Parameter	Parameter Estimate	Standard Error	Pr (Parameter=0)	AIC
b_0	2.8914	0.02478	< 0.0001	-356.0
b_1	-7.8385	0.4307	< 0.0001	

Table 3. Variance-covariance matrix for the random parameters (u_i) and residual error (e) of height-diameter equation (4) for *Pinus densiflora* trees in Gangwon province.

Parameter	u_0	u_1	e
u_0	0.03959		
u_1	0.02489	0.1583	
e	0	0	0.01485

$$\hat{D} = \begin{bmatrix} 0.03959 & 0.02489 \\ 0.02489 & 0.1583 \end{bmatrix}$$

$$\hat{R} = 0.01485 \text{ X } \underset{10\times 10}{\mathbf{I}}$$

$$(y\text{-}Xb) = \begin{bmatrix} 0.033934219 \\ 0.107705067 \\ 0.116684452 \\ -0.05422626 \\ 0.103241172 \\ 0.110723357 \\ 0.144718233 \\ 0.026991611 \\ 0.107922881 \\ 0.150687791 \end{bmatrix}$$

$$Z = \begin{bmatrix} 1 & 0.083333333 \\ 1 & 0.066666667 \\ 1 & 0.055555556 \\ 1 & 0.035714286 \\ 1 & 0.035714286 \\ 1 & 0.033333333 \\ 1 & 0.031250000 \\ 1 & 0.029411765 \\ 1 & 0.025000000 \\ 1 & 0.022222222 \end{bmatrix}$$

Using matrix algebra on equation 3 leads to the estimates of the cluster-specific random-effects parameters of equation 2 for these ten *Pinus densiflora* trees:

$$u = \begin{bmatrix} 0.080083327 \\ 0.040298019 \end{bmatrix}$$

These estimated random-effects parameters were added to the estimated fixed-effects estimates, **b** (Table 2), to obtain cluster-specific parameter estimates for these ten trees:

$$b + u = \begin{bmatrix} 2.8714 + 0.080083327 \\ -7.8385 + 0.040298019 \end{bmatrix} = \begin{bmatrix} 2.971483327 \\ -7.798201981 \end{bmatrix}$$

Thus, the final mixed-effects height-diameter equation for these ten trees is:

$$\ln(H) = (2.971483327) - (7.798201981)D^{-1} \quad (4)$$

The total height predictions from equation 4 were plotted along with predictions from equation 2 with only fixed-effects regression parameters, the ordinary least squares model (equation 1), FREPP, and observed values to show the improvement by using a mixed-effects height-diameter model (Figure 1). A log-transformation bias correction factor, c=residual error/2=e/2 (Baskerville, 1972), was also applied to equation 4 to convert the log units (i.e., ln(m) to absolute units (m)). The mixed-effects predictions are more aligned with the observed values than either the fixed-effects predictions or the ordinary least squares predictions. The FREPP predictions were over-

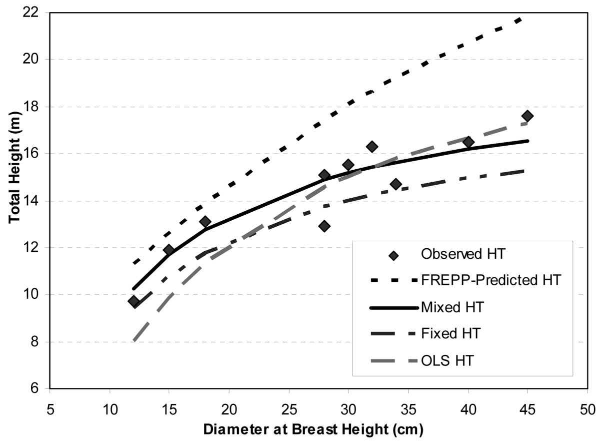

Figure 1. Predicted versus observed individual *Pinus densiflora* tree total heights for four predictions models: 1) Mixed-effects model (equation 2, fixed- plus random-effects parameters), 2) Maximum-likelihood fixed-effects model (equation 2, fixed-effects parameters only), 3) ordinary least squares fixed-effects model (equation 1), and FREPP.

estimated compared to the predictions of the other three models.

In conclusion, the mixed-effects model (equation 2) best predicts total height for individual *Pinus densiflora* trees in Gangwon Province. Though additional computation is needed, we believe the inclusion of cluster-specific random-effects is justified considering that they improve total height predictions on specific plots. We compared the new mixed-effects model with a small sample of independent data to show how well mixed-effects models predict total height with very few (ten in this study) calibration trees. The new mixed-effects model performed better than fixed-effects models, including the Forest Resources Evaluation and Prediction Program (FREPP). We recommend that forest managers use equation 2 with ten calibration trees to predict missing total heights for *Pinus densiflora* trees growing in Gangwon Province. This methodology would be ideal for inventories that routinely sub-sample for total tree height but obtain diameters for all sampled trees.

Acknowledgements

This study was supported by the Division of Forest Resource Information in Korea Forest Research Institute.

Literature Cited

1. Akaike, H. 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19(6): 716-723.
2. Avery, T.E. and Burkhart, H.E. 2002. Forest measurements. 5th Ed. McGraw-Hill, Inc., New York. 456p.
3. Baskerville, G.L. 1972. Use of logarithmic regression in the estimation of plant biomass. Canadian Journal of Forest Research 2: 49-53.
4. Budhathoki, C.B., Lynch, T.B. and Guldin, J.M. 2008. A mixed-effects model for the dbh-height relationship of shortleaf pine (Pinus echinata Mill.). Southern Journal of Applied Forestry 32(1): 5-11.
5. Calama, R. and Montero, G. 2004. Interregional nonlinear height-diameter model with random coefficients for stone pine in Spain. Canadian Journal of Forest Research 34: 150-163.
6. Clutter, J.L., J.C. Fortson, L.V. Pienaar, G.H. Brister and Bailey, R.L. 1983. Timber Management: A Quantitative Approach. New York, NY: John Wiley and Sons. 333p.
7. Kim, D. H., D.M. Seol, D.K. Nor, C.H. Kim, H.T. Hwang, S.H. Kim, and Song, M.S. 2008. Forest resources based on national forest inventory in Korea. IUFRO Proceedings, Advances in Forest Management and Inventory, Kangwon National University, Korea. 125-127p.
8. Korea Forest Research Institute. 2008a. Forest Resources Evaluation and Prediction Program.
9. Korea Forest Research Institute. 2008b. Manual of the fifth National Forest Inventory (Version 1.2). 36p.
10. Korea Forest Research Institute. 2004. Forest Resources Evaluation and Prediction System for Major Species. 125p.
11. Korea Forest Service. 2008. Statistical yearbook of forestry (38).
12. Lappi, J. 1997. A longitudinal analysis of height/diameter curves. Forest Science 43(4): 555-570.
13. Lappi, J. 1991. Calibration of height and volume equations with random parameters. Forest Science 37(3): 781-801.
14. Lynch, T.B., Holley, A.G. and Stevenson, D.J. 2005. A random-parameter height-dbh model for cherrybark oak. Southern Journal of Applied Forestry 29: 22-26.
15. Mehtatalo, L. 2004. A longitudinal height-diameter model for Norway spruce in Finland. Canadian Journal of Forest Research 34: 131-140.
16. Schabenberger, O. and Pierce, F.J. 2002. Contemporary Statistical Models for the Plant and Soil Sciences. CRC Press, Boca Raton, FL. 738p.
17. Sharma, M. and Parton, J. 2007. Height-diameter equations for boreal tree species in Ontario using a mixed-effects modeling approach. Forest Ecology and Management 249: 187-198.
18. Trincado, G., VanderSchaaf, C.L. and Burkhart, H.E. 2007. Regional mixed-effects height-diameter models for loblolly pine (*Pinus taeda* L.) plantations. European Journal of Forest Research 126: 253-262.
19. Vandershaaf, C.L. 2008. Stand level height-diameter

mixed effects models: parameters fitted using loblolly pine but calibrated for sweetgum. In: Jacobs, D.F. and C.H. Michler, eds. Proceedings, 16th Central Hardwoods Forest Conference. Gen. Tech. Rep. NRS-P-24. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northern Research Station. 595p.

(Received February 5, 2009; Accepted March 4, 2009)