

Baldcypress height-diameter equations and their prediction confidence intervals

BERNARD R. PARRESOL

*Institute for Quantitative Studies, USDA Forest Service, Southern Forest Experiment Station,
701 Loyola Avenue, New Orleans, LA 70113, U.S.A.*

Received November 5, 1991

Accepted March 23, 1992

PARRESOL, B.R. 1992. Baldcypress height-diameter equations and their prediction confidence intervals. *Can. J. For. Res.* **22**: 1429-1434.

Height-diameter relationships are an important component in yield estimation, stand description, and damage appraisals. A nonlinear exponential function used extensively in the northwest United States was chosen for bald cypress (*Taxodium distichum* (L.) Rich.). Homogeneity and normality of residuals were examined, and the function as well as the mean and individual prediction confidence bands were plotted. The inclusion of stand basal area as an additional independent variable provided a better fit to the data. The paper is concluded with a section on construction and use of simple and joint confidence intervals about the mean and individual predictions from the nonlinear regression.

PARRESOL, B.R. 1992. Baldcypress height-diameter equations and their prediction confidence intervals. *Can. J. For. Res.* **22**: 1429-1434.

Les relations hauteur-diamètre constituent une composante importante pour l'estimation de la production forestière, la description des peuplements et l'évaluation des dommages. Un modèle exponentiel non linéaire fréquemment utilisé dans le nord-ouest des États-Unis a été appliqué à des mesures effectuées sur le cyprès chauve (*Taxodium distichum* (L.) Rich.). L'homogénéité et la normalité des résidus sont examinées et l'équation, de même que les intervalles de confiance pour la moyenne et les estimés individuels, sont représentées graphiquement. L'ajout de la surface terrière du peuplement comme variable indépendante assure un meilleur ajustement du modèle aux observations. L'article se termine avec une section portant sur la construction et l'utilisation des intervalles de confiance pour la moyenne et les prédictions individuelles à partir de régressions non linéaires.

[Traduit par la rédaction]

Introduction

Bald cypress (*Taxodium distichum* (L.) Rich.), extending across the Coastal Plain from southeastern Texas to southern Delaware, has slowly regained its importance as a commercial species. The volume of cypress growing stock on commercial forest land is estimated to be $5.5 \times 10^9 \text{ ft}^3$ ($155.7 \times 10^6 \text{ m}^3$) (Williston *et al.* 1980). Recently, tree-volume and stem-profile functions were developed for second-growth bald cypress (Hotvedt *et al.* 1985; Parresol *et al.* 1987). These functions used a fixed-height measurement point of 10 ft (3.0 m) above the ground for diameter as a superior choice over the variable-height measure called normal diameter, that is, diameter measured 1.5 ft (0.5 m) above pronounced butt swelling (Avery and Burkhart 1983). Diameter at 10 ft (3.0 m) is easily measured with a pole caliper (Fig. 1) such as described by Ferree (1946). In many survey and timber cruising operations, tree height is not measured or is measured only on a small subsample of trees. Therefore, a height-diameter equation would be a valuable addition to the prediction systems of Hotvedt *et al.* (1985) and Parresol *et al.* (1987). Where trees are damaged, such as occurred from hurricane Hugo in the Carolinas, a height-diameter equation would aid in estimating losses.

Perhaps more important than the height predictions themselves are the variances of the predictions. Too often research reports, such as those in scientific journals, do not provide the necessary information for the building of prediction confidence intervals! Yet this information is important to forest managers, damage appraisers, etc., who have to consider the potential volume and value involved. Full details and examples are provided in the section on Application and reliability.

Data

The data are described in Hotvedt *et al.* (1985). Briefly, the data consist of 157 sample trees from 26 locations (6 trees per location except one where a 7th tree was measured) across the south Delta region of Louisiana (Fig. 2). Trees were felled, diameter at a fixed height of 10 ft (3.0 m), termed d_{10} , was measured to the nearest 0.1 in. (0.3 cm), and total height was measured to the nearest 0.1 ft (0.03 m). In addition, stand basal area (BA) around each sample tree was measured using a 10-factor prism. Surrounding trees were sighted through the prism at normal diameter for determination of BA. The range of BA was 30 to 300 ft^2/acre



FIG. 1. Forester measuring diameter at 10 ft (3.0 m) on a bald cypress tree using a pole caliper.

(6.89 to 68.88 m²/ha), with the median being 130 ft²/acre (29.85 m²/ha). The distribution of sample trees by d_{10} and height are listed in Table 1.

Height–diameter equation

A scatterplot of height over d_{10} (Fig. 3) shows a monotone increasing curvilinear relationship typical of height–diameter data. Curtis (1967), in his work with second-growth Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco) employed the following log-reciprocal model:

$$\ln(H) = \beta_0 + \frac{\beta_1}{d_{4.5}} + \varepsilon$$

where H is total tree height in feet, $d_{4.5}$ is tree diameter at 4.5 ft (1.4 m) measured in inches, \ln denotes the natural logarithm, β_i 's are model parameters, and ε is residual error. This equation possesses an asymptote, an inflection point, and passes through the origin. Curtis suggested two modifications of the above model. The first was to add a power term to the independent variable for increased flexibility and the second

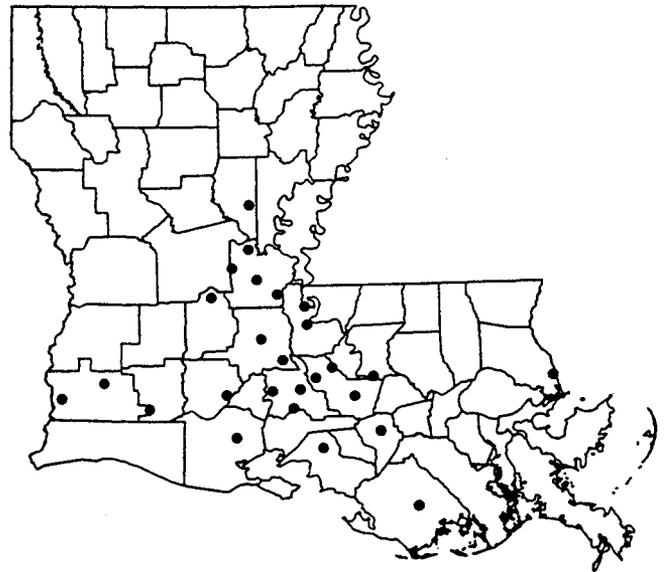


FIG. 2. Locations of the 26 bald cypress sample sites in Louisiana.

was to constrain the model to pass through the natural origin (0, 4.5). For this study the natural origin is (0, 10), that is, diameter is measured at 10 ft. Researchers such as Larsen and Hann (1987) found that the residuals of log-transformed height–diameter equations are not normally distributed. They preferred weighted regression over use of the logarithmic transformation. The result of these modifications is the following model that is nonlinear in the parameters:

$$[1] \quad H = C + \exp(\beta_0 + \beta_1 d_C^{\beta_2}) + \varepsilon$$

where C is the natural origin constraint and \exp is base of the natural logarithm. Larsen and Hann (1987) and Wang and Hann (1988) used this model (with $C = 4.5$ ft) for a variety of coniferous and hardwood species in Oregon.

Statistics used to examine the appropriateness of the regression were (i) Bartlett's χ^2 -test for homogeneity of variance, (ii) the Kolmogorov D -statistic for normality, (iii) variation in H explained, referred to as fit index or FI (Schlaegel 1981), and (iv) root mean square error for H . For Bartlett's test, the residuals were separated into three groups based on values of the independent variable: d_{10} less than 11, d_{10} from 11 to 16, and d_{10} greater than 16. This breakdown provides variances calculated on approximately equal-length diameter classes. Since residuals have a mean of zero, care must be taken to use the correct critical values in determining significance of the Kolmogorov test for normality. The appropriate percentage points for the case of μ known and σ^2 estimated are given in Stephens (1974). All tests of hypothesis were made using an α -level of 0.05.

Upon fitting eq. 1 to the data with nonlinear least squares, the following function resulted¹:

$$[2] \quad \hat{H} = 10 + \exp(5.15907 - 2.65144 d_{10}^{-0.41589})$$

¹For height in meters use $\hat{H} = 3 + \exp(3.97097 - 3.90704 d_3^{0.41589})$, where d_3 is diameter measured in centimetres at a height of 3.0 m above the ground.

TABLE 1. Distribution of bald cypress sample trees by height and diameter at 10 ft

d_{10} (in.)	Total height (ft)												Total trees	
	45	50	55	60	65	70	75	80	85	90	95	100		
4	2													2
5		3	1	2	1									7
6	1	4	3		2		1							11
7			2	3	1	1			1					8
8			2	1	4	1	4							12
9					1	1	2		1	1				6
10						5	1	3	3	1				13
11					1	2	2		1	1		1		8
12					2	7	5	2	1					17
13						1	2	2	1	2				8
14				1	1		3	1	3	1				10
15							5	5	1	4	1			16
16							2	2	1	1	3			9
17				1			1	1		2	2	2		9
18									4	2	3			9
19								1	2		1	1		5
20									1		2	1		4
21											2			2
23												1		1
Total trees	3	7	8	8	13	18	28	17	20	15	14	6		157

This equation explained 64% (FI = 0.64) of the variation in observed height values, the root mean square error was 8.10 ft (2.47 m), Bartlett's test indicated homogeneous variance ($\chi^2 = 0.15$, $P = 0.93$), and the Kolmogorov test indicated normally distributed residuals ($D = 0.097$, $P = 0.09$). Weighted regression was deemed unnecessary based on the results of Bartlett's test and the Kolmogorov test. Equation 2 with its corresponding mean and individual prediction confidence bands is plotted in Fig. 3.

Height - diameter - basal area equation

To improve height predictions and to adjust for differences between stands, foresters have used additional independent variables such as age (Curtis 1967), site index, and basal area (Larsen and Hann 1987; Wang and Hann 1988) in their height-diameter equations. Equation 1 modified by adding basal area as an additional independent variable becomes

$$[3] \quad H = C + \exp(\beta_0 + \beta_1 d_C^{\beta_2} + \beta_3 BA) + \epsilon$$

No site index work has been done with bald cypress because it is difficult to determine tree age. The species has a habit of forming false rings. Hence I used stand basal area around the tree as a second independent variable. Fitting eq. 3 to the data using nonlinear least squares, the following function resulted²:

$$[4] \quad \hat{H} = 10 + \exp(5.45460 - 2.95441 d_{10}^{-0.30497} + 0.0010070 BA)$$

This equation explained 70% (FI = 0.70) of the variation in observed height values, the root mean square error was 7.42 ft

²For height in meters use $\hat{H} = 3 + \exp(4.26650 - 3.92584 d_3^{0.30497} + 0.0043859 BA)$, where BA is measured in m^2/ha .

(2.26 m) (a reduction of 8.4% over eq. 2), Bartlett's test indicated homogeneous variance ($\chi^2 = 0.52$, $P = 0.77$), and the Kolmogorov test indicated normally distributed residuals ($D = 0.046$, $P > 0.15$). Again, weighted regression was deemed unnecessary based on the results of Bartlett's test and the Kolmogorov test. Equation 4 is plotted in Fig. 4.

Application and reliability

Knowing the prediction interval is as important as being able to predict the height given d_{10} or d_{10} and BA. The construction of simple and joint confidence intervals about nonlinear regressions is analogous to that of linear regressions, using a matrix algebra approach. Two quantities are needed to construct the bounds on the predictions: (i) the standard errors of the predictions (se) and (ii) a t - or W -value, for simple or joint confidence intervals, respectively. The interval boundary points are obtained from

$$[5] \quad \hat{H}_i \pm se(t \text{ or } W)$$

where $W = \sqrt{pF(1 - \alpha; p, n - p)}$ is the Working-Hotelling value for confidence bands, p is number of parameters (3 or 4), F represents the F -statistic, and n is number of observations (157). If the user is interested in assessing limits for a single point on eqs. 2 or 4 then a confidence interval about that point is appropriate. If, however, as is more often the case, the user is interested in assessing limits about multiple points on eqs. 2 or 4, then joint confidence intervals (variously known as a confidence band, confidence region, or simultaneous confidence limits) are appropriate (Draper and Smith 1981; Neter *et al.* 1985).

Leverage and standard errors

To calculate standard errors we must first compute a value known as the leverage. Let F_n represent either eqs. 2 or 4.

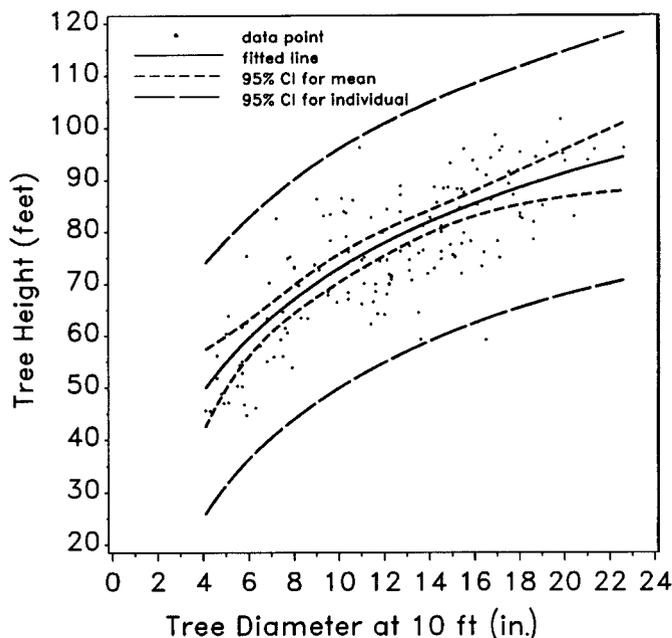


FIG. 3. Scatterplot and height-diameter response curve showing 95% confidence bands for the mean and individual estimates.

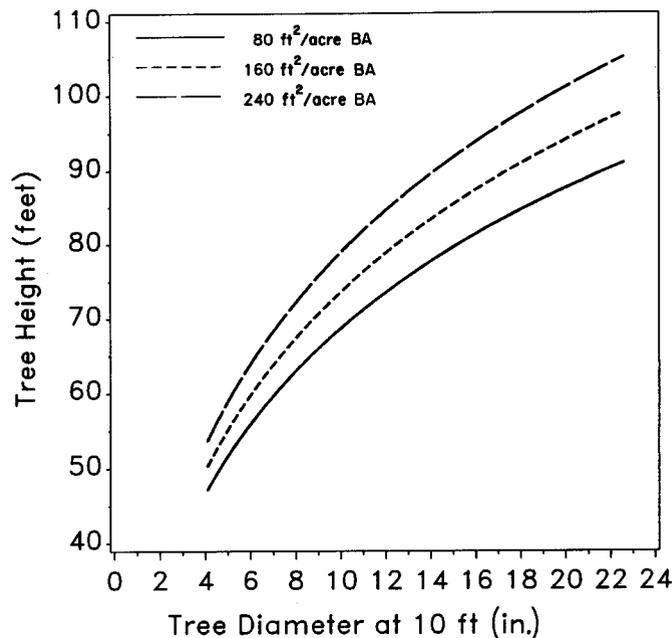


FIG. 4. Height-diameter-basal area response curve showing the height response across d_{10} for three levels of stand basal area.

The regression design matrix X ($157 \times p$) is formed by differentiating F_n with respect to the β 's, that is, $X = \partial F_n / \partial \beta$. For the i th observation, a scalar known as the leverage is computed as follows:

$$[6] \quad l_i = \mathbf{x}_i (X'X)^{-1} \mathbf{x}_i'$$

where \mathbf{x}_i is the i th row vector of X . There are three types of standard errors: (i) for the predicted mean value of H_i , $s(\hat{H}_i)$; (ii) for a predicted value of an individual (new) outcome drawn from the distribution of H_i , $s(\hat{H}_{i(\text{new})})$; and (iii) for the predicted mean of m new observations on H_i , $s(\hat{H}_{i(\text{new})})$. They

are calculated as

$$[7a] \quad s(\hat{H}_i) = \sqrt{l_i s^2}$$

$$[7b] \quad s(\hat{H}_{i(\text{new})}) = \sqrt{l_i s^2 + s^2}$$

$$[7c] \quad s(\hat{H}_{i(\text{new})}) = \sqrt{l_i s^2 + \frac{s^2}{m}}$$

where s^2 is the mean square error of the regression.

Vectors, matrices, and mean square errors

The vector \mathbf{x}_i for eq. 2 has the form

$$\left[(\hat{H}_i - 10) \quad (\hat{H}_i - 10)d_{10i}^{-0.41589} \quad -2.65144(\hat{H}_i - 10)d_{10i}^{-0.41589} \ln(d_{10i}) \right]$$

The $(X'X)^{-1}$ matrix from eq. 2 is

$$\begin{bmatrix} 0.007672833 & -0.000931329 & 0.003170803 \\ -0.000931329 & 0.000537888 & -0.000320559 \\ 0.003170803 & -0.000320559 & 0.001320337 \end{bmatrix}$$

The mean square error of eq. 2 is 65.65077. With these three pieces of information bounds can be computed for the height predictions from eq. 2.

The vector \mathbf{x}_i for eq. 4 has the form

$$\left[(\hat{H}_i - 10) \quad (\hat{H}_i - 10)d_{10i}^{-0.30497} \quad -2.95441(\hat{H}_i - 10)d_{10i}^{-0.30497} \ln(d_{10i}) \quad (\hat{H}_i - 10)BA_i \right]$$

The $(X'X)^{-1}$ matrix from eq. 4 is

$$\begin{bmatrix} 0.026\ 549\ 365 & -0.015\ 941\ 067 & 0.005\ 604\ 813 & 2.326\ 876\ 1 \times 10^{-7} \\ -0.015\ 941\ 067 & 0.009\ 866\ 133 & -0.003\ 331\ 052 & -2.662\ 829\ 0 \times 10^{-7} \\ 0.005\ 604\ 813 & -0.003\ 331\ 052 & 0.001\ 188\ 263 & 5.664\ 069\ 8 \times 10^{-8} \\ 2.326\ 876\ 1 \times 10^{-7} & -2.662\ 829 \times 10^{-7} & 5.664\ 069\ 8 \times 10^{-8} & 5.866\ 488\ 4 \times 10^{-10} \end{bmatrix}$$

The mean square error of eq. 4 is 55.041 33. With these three pieces of information bounds can be computed for the height predictions from eq. 4.

Calculations

The following examples serve to illustrate the use of eqs. 2 and 4 and the method of constructing confidence intervals. Consider a tree with a diameter measured at 10 ft above the ground of 15.4 in. Using eq. 2 the height is estimated as

$$\begin{aligned} \hat{H} &= 10 + \exp(5.159\ 07 - 2.651\ 44(15.4)^{-0.415\ 89}) \\ &= 84.3\ \text{ft} \end{aligned}$$

Inserting the values 84.3 for \hat{H} and 15.4 for d_{10} into x_i we obtain

$$[74.344\ 38\ 23.843\ 51\ -172.865\ 70]$$

The leverage is calculated from eq. 6 and is 0.010 213. From eq. 7 the standard errors are (using $m = 5$ in eq. 7c)

$$\begin{aligned} s(\hat{H}_i) &= \sqrt{(0.010\ 213)(65.650\ 77)} = 0.818\ 84 \\ s(\hat{H}_{i(\text{new})}) &= \sqrt{(0.010\ 213)(65.650\ 77) + 65.650\ 77} \\ &= 8.143\ 79 \\ s(\hat{H}_{i(\text{new})}) &= \sqrt{(0.010\ 213)(65.650\ 77) + 65.650\ 77/5} \\ &= 3.714\ 92 \end{aligned}$$

The 95% t -value is 1.975. From eq. 5 the overall mean CI, individual CI, and mean CI of five trees, respectively, are

$$\begin{aligned} 84.3 \pm 0.818\ 84(1.975) &= 82.7 \leq \hat{H} \leq 85.9 \\ 84.3 \pm 8.143\ 79(1.975) &= 68.2 \leq \hat{H}_{(\text{new})} \leq 100.4 \\ 84.3 \pm 3.714\ 92(1.975) &= 77.0 \leq \hat{H}_{(\text{new})} \leq 91.6 \end{aligned}$$

Suppose from a damage appraisal we have three trees with d_{10} and surrounding BA values of 12.6 in. and 200 ft²/acre, 16.2 in. and 160 ft²/acre, and 20.1 in. and 130 ft²/acre. Using eq. 4 heights are estimated as

$$\begin{aligned} \hat{H}_1 &= 10 + \exp(5.454\ 60 - 2.954\ 41(12.6)^{-0.304\ 97} \\ &\quad + 0.001\ 007\ 0(200)) = 83.1\ \text{ft} \\ \hat{H}_2 &= 10 + \exp(5.454\ 60 - 2.954\ 41(16.2)^{-0.304\ 97} \\ &\quad + 0.001\ 007\ 0(160)) = 87.6\ \text{ft} \\ \hat{H}_3 &= 10 + \exp(5.454\ 60 - 2.954\ 41(20.1)^{-0.304\ 97} \\ &\quad + 0.001\ 007\ 0(130)) = 91.6\ \text{ft} \end{aligned}$$

Inserting the appropriate values into x_i , for $i = 1$ to 3, and stacking the three row vectors into a matrix, we obtain

$$\begin{bmatrix} 73.095\ 01 & 33.752\ 64 & -252.658\ 04 & 14\ 619.001\ 41 \\ 77.643\ 94 & 33.207\ 93 & -273.236\ 99 & 12\ 423.030\ 59 \\ 81.644\ 38 & 32.695\ 69 & -289.858\ 99 & 10\ 613.768\ 77 \end{bmatrix}$$

By using this matrix in eq. 6 instead of the individual vectors x_i , a leverage matrix is calculated whose diagonal elements are the leverage values. This averts the need for repetitive calculation of eq. 6 for each observation. The leverage matrix is

$$l_{\text{matrix}} = \begin{bmatrix} 0.021\ 266 & 0.011\ 298 & -0.000\ 536 \\ 0.011\ 298 & 0.014\ 484 & 0.018\ 387 \\ -0.000\ 536 & 0.018\ 387 & 0.042\ 754 \end{bmatrix}$$

The individual leverage values are as follows: $l_1 = 0.021\ 266$; $l_2 = 0.014\ 484$; $l_3 = 0.042\ 754$. From eq. 7b, the standard errors for each individual prediction are

$$\begin{aligned} s(\hat{H}_{1(\text{new})}) &= \sqrt{(0.021\ 266)(55.041\ 13) + 55.041\ 13} \\ &= 7.497\ 44 \\ s(\hat{H}_{2(\text{new})}) &= \sqrt{(0.014\ 484)(55.041\ 13) + 55.041\ 13} \\ &= 7.472\ 51 \\ s(\hat{H}_{3(\text{new})}) &= \sqrt{(0.042\ 754)(55.041\ 13) + 55.041\ 13} \\ &= 7.575\ 91 \end{aligned}$$

For 90% joint confidence intervals about these three predictions, the W -value is $\sqrt{4F(0.90;4,153)} = \sqrt{4(1.982)} = 2.816$. Using eq. 5 the joint confidence intervals are

$$\begin{aligned} 83.1 \pm 7.497\ 44(2.816) &= 62.0 \leq \hat{H}_{1(\text{new})} \leq 104.2 \\ 87.6 \pm 7.472\ 51(2.816) &= 66.6 \leq \hat{H}_{2(\text{new})} \leq 108.6 \\ 91.6 \pm 7.575\ 91(2.816) &= 70.3 \leq \hat{H}_{3(\text{new})} \leq 112.9 \end{aligned}$$

Summary

Predicting height directly from diameter is useful in many situations. Resource professionals should bear in mind that regression functions like eqs. 2 and 4 provide point estimates that have a variance. When evaluating a large group of trees with the same d_{10} or d_{10} and surrounding BA, constructing a confidence interval on \hat{H}_i will provide a range of values that should contain the true mean of that group. When evaluating one tree, constructing a confidence interval on $\hat{H}_{i(\text{new})}$ will

provide a range of values that should contain the true height of the individual. When evaluating a few trees with the same d_{10} or d_{10} and surrounding BA, constructing a confidence interval on $\bar{H}_{i(\text{new})}$ will provide a range of values that should contain the true mean of this small group. If, however, as is more often the case, one is interested in evaluating a number of observations across a range of d_{10} or d_{10} and surrounding BA values, then joint confidence intervals are necessary.

Acknowledgements

Research for this paper was supported by Louisiana State University and USDA Forest Service forest management research funding. The author wishes to thank Dr. Jim Hotvedt and Dr. Quang Van Cao for guidance and teaching and to thank the two anonymous reviewers.

- Avery, T.E., and Burkhardt, H.E. 1983. Forest measurements. 3rd ed. McGraw-Hill Book Co., New York.
- Curtis, R.O. 1967. Height-diameter and height-diameter-age equations for second-growth Douglas-fir. *For. Sci.* **13**: 365-375.
- Draper, N.R., and Smith, H. 1981. Applied regression analysis. 2nd ed. John Wiley & Sons, New York.

- Ferree, M.J. 1946. The pole caliper. *J. For.* **44**: 594-595.
- Hotvedt, J.E., Cao, Q.V., and Parresol, B.R. 1985. Tree-volume and stem-profile functions for baldcypress. *South. J. Appl. For.* **9**(10): 227-232.
- Larsen, D.R., and Hann, D.W. 1987. Height-diameter equations for seventeen tree species in southwest Oregon. *Oreg. State Univ. For. Res. Lab. Res. Bull.* 49.
- Neter, J., Wasserman, W., and Kutner, M.H. 1985. Applied linear statistical models. 2nd ed. Richard D. Irwin, Inc., Homewood, Ill.
- Parresol, B.R., Hotvedt, J.E., and Cao, Q.V. 1987. A volume and taper prediction system for baldcypress. *Can. J. For. Res.* **17**: 250-259.
- Schlaegel, B.E. 1981. Testing, reporting, and using biomass estimation models. *In Proceedings of the 1981 Southern Forest Biomass Workshop*, 11-12 June 1981, Georgetown, S.C. Clemson University, Clemson, S.C. pp. 95-112.
- Stephens, M.A. 1974. EDF statistics for goodness of fit and some comparisons. *J. Am. Stat. Assoc.* **69**: 730-737.
- Wang, C., and Hann, D.W. 1988. Height-diameter equations for sixteen tree species in the central western Willamette Valley of Oregon. *Oreg. State Univ. For. Res. Lab. Res. Pap.* 51.
- Williston, H.L., Shropshire, F.W., and Balmer, W.E. 1980. Cypress management: a forgotten opportunity. *USDA For. Serv. For. Rep. Southeastern Area - Forestry Report* 8.