

Prediction and Error of Baldcypress Stem Volume from Stump Diameter

Bernard R. Parresol, *USDA Forest Service, Southern Research Station, P.O. Box 2680, Asheville, NC 28802.*

ABSTRACT. *The need to estimate the volume of removals occurs for many reasons, such as in trespass cases, severance tax reports, and post-harvest assessments. A logarithmic model is presented for prediction of baldcypress total stem cubic foot volume using stump diameter as the independent variable. Because the error of prediction is as important as the volume estimate, I demonstrate construction and use of simple and joint confidence intervals about the mean and individual predictions. For completeness, I address prediction and error from inventory estimates of removals. South. J. Appl. For. 22(2):69–73.*

Many circumstances exist today where the volume, and ultimately the value, of trees must be assessed from stump measurements. These situations include: (1) assessing timber sales based on stump diameters, (2) checking harvesting practices following tree removals, (3) tracing the history of cutover lands, (4) assessing damage resulting from adverse environmental conditions, (5) determining volume loss resulting from trespass cutting, and (6) calculating growth on cut as part of a forest inventory (Bylin 1982, Wharton 1984). In all these situations only a stump diameter is known.

In terms of wetland tree species, baldcypress (*Taxodium distichum* [L.] Rich) ranks high in importance. Cypress has made a strong comeback since the last virgin stands were cut in the 1940s. Current growing stock is estimated at around 6 billion cubic feet in the southeastern United States (Beltz et al. 1992) and its distribution ranges from East Texas to southern Maryland. Its increasing volume, distribution, and desirability as a species for management, point to the need for the ability to assess removals. Currently no published equations exist for estimating cypress volume from stump diameter. In this paper I present a simple linear model for estimating bole cubic foot volume above stump, and I provide details and examples on the construction and use of simple and joint confidence intervals about the mean and individual predictions. Also, prediction and associated error in the context of forest inventories on removals is discussed.

Data

Taper data were collected on 157 trees from 26 sites (25 sites with 6 trees and 1 site with 7) located throughout the South Delta region of Louisiana (Figure 1). The typical stump height for

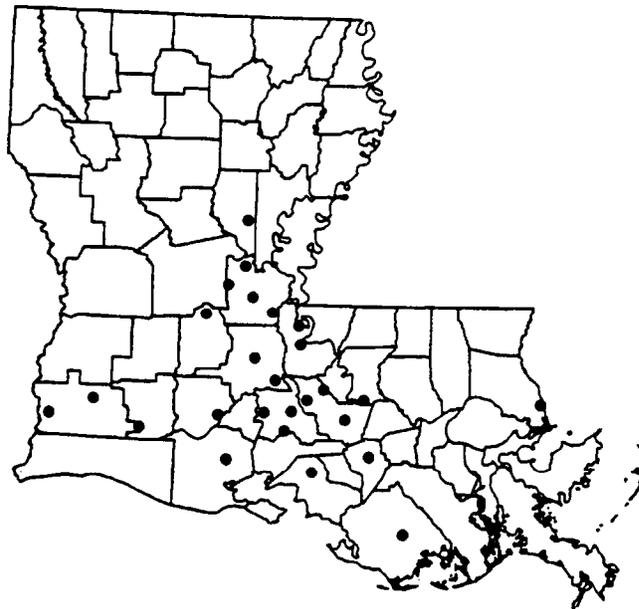


Figure 1. Locations in Louisiana of the 26 baldcypress sample sites used in development of stem volume equation based on stump diameter.

baldcypress trees is 3 ft due to their nature of forming fluted basal swells. Consequently, trees were felled leaving a 3 ft stump, and total bole length was measured to the nearest 0.1 ft. Diameters outside bark were measured to the nearest 0.1 in. at the stump and at 2 ft intervals for the first 14 ft of bole length and 4 ft intervals thereafter. Calipers were used to take two diameter readings at each measurement point, which were then geometrically averaged. Two bark thickness measurements were taken at each measurement point using a bark gauge so inside-bark diameter could be calculated.

While a 3 ft stump will capture most of the swell and fluting, some fluting might still exist on the bole above this point. What is desired is solid-wood diameter. The length of

NOTE: Manuscript received December 12, 1996, accepted April 2, 1997. Bernard Parresol may be reached at (828) 259-0500; Fax: (828) 257-4840; E-mail: bparresol/srs@fs.fed.us.

the bole exhibiting flutes was cross-sectioned at the appropriate points, and diameter readings were taken on the largest possible circle or ellipse which could be inscribed (with an expandable hoop) inside the flutes at each cross-section. This same procedure was used to measure stump diameter. Table 1 lists the distribution of sample trees by stump diameter inside bark.

Sectional cubic foot volumes were calculated using Smalian's formula. The top section was treated as a cone. Inside- and outside-bark total stem cubic foot volumes (above stump) were obtained by summing the inside- and outside-bark sectional volumes, respectively. The range of inside-bark volume was 2.3 to 109.0 ft³, with the mean being 32.0 ft³. The range of outside-bark volume was 2.7 to 119.9 ft³, with the mean being 35.3 ft³.

Model

After examining a scatter plot of the data the following logarithmic model was hypothesized:

$$\ln V = \beta_1 + \beta_2 \ln D + \epsilon \quad (1)$$

where V is bole volume above stump, D is stump diameter inside bark, \ln represents natural logarithms, the β_i 's are model parameters, and ϵ is residual error. This model was fitted to the data using ordinary least squares regression. An examination of the residuals revealed no trends. A nonlogarithmic model was also tried (fitted with weighted least squares) but model (1) fit the data very nicely and is easy to work with. The final fitted equations are

$$\widehat{\ln V_{ib}} = -3.2994 + 2.4495 \ln D; R^2 = 0.87, \hat{\sigma}^2 = 0.1184 \quad (2)$$

$$\widehat{\ln V_{ob}} = -3.0658 + 2.4019 \ln D; R^2 = 0.87, \hat{\sigma}^2 = 0.1160 \quad (3)$$

Table 1. Distribution of baldcypress sample trees by inside bark stump diameter.

Stump diameter (in.)	No. of trees
6	4
7	9
8	8
9	8
10	10
11	7
12	5
13	14
14	7
15	12
16	15
17	10
18	12
19	14
20	9
21	5
22	2
23	2
24	1
26	3
Total	157

where ib and ob stand for inside bark and outside bark, respectively. It is immediately obvious that the slope of Equation (3) is slightly smaller than the slope of Equation (2); however, predicted V_{ob} is greater than predicted V_{ib} for D less than 135.3 in. (nearly 12 ft), well beyond the application range of the functions, so constrained equations were deemed unnecessary.

Application and Reliability

Transforming Estimates to Original Units

When the logarithmic transformation is used, it is usually desired to express estimated values of V in arithmetic (i.e., untransformed) units. However, the conversion of the unbiased logarithmic estimate of the mean to arithmetic units is not direct. The antilogarithm of $\ln V$ yields the median of the skewed arithmetic distribution rather than the mean. If $\hat{Y} = \widehat{\ln V}$ and $\hat{\sigma}^2 =$ sample variance of the logarithmic equation, then

$$\hat{V} = \exp(\hat{Y} + \hat{\sigma}^2 / 2) \quad (4)$$

where \hat{V} is the estimated value in arithmetic units (Yandle and Wiant 1981). Uncertainty limits (i.e., confidence intervals) about the $\widehat{\ln V}$ can be converted to arithmetic units in the same manner, using Equation (4); these limits will be asymmetric about the regression line.

Uncertainty Limits

Knowing the prediction interval is as important as being able to predict the volume given D . The construction of simple and joint confidence intervals is straightforward. Two quantities are needed to construct the bounds on the predictions: (1) the standard errors of the predictions (se) and (2) a t - or W -value, for simple or joint confidence intervals, respectively. The interval boundary points are obtained from:

$$\widehat{\ln V}_i \pm se(t \text{ or } W) \quad (5)$$

where

$$W = \sqrt{pF(1 - \alpha; p, n - p)}$$

is the Working-Hotelling value for confidence bands, p is number of parameters (2), and n is number of observations (157). If the user is interested in assessing limits for a single volume using Equations (2) or (3), then a confidence interval about that volume is appropriate. If, however, as is more often the case, the user is interested in assessing limits about multiple volumes on Equations (2) or (3), then joint confidence intervals (variously known as a confidence band or confidence region or simultaneous confidence limits) are appropriate (Draper and Smith 1981, Neter et al. 1985).

In the following definitions let $\hat{Y} = \widehat{\ln V}$ and $X = \ln D$. There are three types of standard errors one can utilize: (1) for the predicted mean value of $\ln V_i, s(Y_i)$; (2) for a predicted value of an individual (new) outcome drawn

from the distribution of $\ln V_i, s(\hat{Y}_{i(new)})$; and (3) for the predicted mean of m new observations on $\ln V_i, s(\hat{Y}_{i(new)})$. They are calculated as:

$$s(\hat{Y}_i) = \sqrt{\hat{\sigma}^2 \left(\frac{1}{n} + \frac{(X_i - \bar{X})^2}{CSS} \right)} \quad (6a)$$

$$s(\hat{Y}_{i(new)}) = \sqrt{\hat{\sigma}^2 \left(1 + \frac{1}{n} + \frac{(X_i - \bar{X})^2}{CSS} \right)} \quad (6b)$$

$$s(\hat{Y}_{i(new)}) = \sqrt{\hat{\sigma}^2 \left(\frac{1}{m} + \frac{1}{n} + \frac{(X_i - \bar{X})^2}{CSS} \right)} \quad (6c)$$

where \bar{X} is the mean of the logarithmic transformed values of stump diameter of our sample of $n = 157$ trees ($\bar{X} = 2.6179$), and CSS is the corrected sum of squares of X ($CSS = \sum(X_i - \bar{X})^2 = 20.061$).

Examples

The following examples illustrate the use of Equations (2) through (6). Consider a stump with a measured solid-wood diameter inside bark of 16.0 in. Using Equation (2) we obtain

$$\hat{Y}_{ib} = -3.2994 + 2.4495(2.7726) = 3.492$$

From Equations (6a), (6b), and (6c), the standard errors are [using $m = 5$ in Equation (6c)]:

$$s(\hat{Y}_{ib}) = \sqrt{0.1184 \left(\frac{1}{157} + \frac{(2.7726 - 2.6179)^2}{20.061} \right)} = 0.0299$$

$$s(\hat{Y}_{ib(new)}) = \sqrt{0.1184 \left(1 + \frac{1}{157} + \frac{(2.7726 - 2.6179)^2}{20.061} \right)} = 0.3454$$

$$s(\hat{Y}_{ib(new)}) = \sqrt{0.1184 \left(\frac{1}{5} + \frac{1}{157} + \frac{(2.7726 - 2.6179)^2}{20.061} \right)} = 0.1568$$

The 95% t -value is 1.975. From Equation (5) the overall mean confidence interval, individual confidence interval, and mean confidence interval of five trees, respectively, are:

$$3.492 \pm 0.0299(1.975) = 3.433 \leq \hat{Y}_{ib} \leq 3.551$$

$$3.492 \pm 0.3454(1.975) = 2.810 \leq \hat{Y}_{ib(new)} \leq 4.174$$

$$3.492 \pm 0.1568(1.975) = 3.182 \leq \hat{Y}_{ib(new)} \leq 3.802$$

To convert the logarithmic values to arithmetic units, we apply Equation (4). For the predicted mean tree volume inside bark from a 16 in. stump diameter we obtain:

$$\hat{V}_{ib} = \exp(3.492 + 0.1184 / 2) = 34.9 \text{ ft}^3$$

Applying Equation (4) to the uncertainty limit values we obtain the following intervals:

$$32.9 \leq \hat{V}_{ib} \leq 37.0$$

$$17.6 \leq \hat{V}_{ib(new)} \leq 68.9$$

$$25.6 \leq \hat{V}_{ib(new)} \leq 47.5$$

The large confidence interval bounding $\hat{V}_{ib(new)}$ reflects the considerable uncertainty surrounding individual tree volume when tree height is unknown.

Suppose from a trespass case there are three stumps with diameters of 12.8, 18.5, and 23.6 in. Using Equation (3), we first estimate the logarithm of outside bark volumes:

$$\hat{Y}_{ob_1} = -3.0658 + 2.4019(2.5494) = 3.058$$

$$\hat{Y}_{ob_2} = -3.0658 + 2.4019(2.9178) = 3.942$$

$$\hat{Y}_{ob_3} = -3.0658 + 2.4019(3.1612) = 4.527$$

Inserting the appropriate values into Equation (6a), the mean value standard errors are:

$$s(\hat{Y}_{ob_1}) = \sqrt{0.116 \left(\frac{1}{157} + \frac{(2.5494 - 2.6179)^2}{20.061} \right)} = 0.0277$$

$$s(\hat{Y}_{ob_2}) = \sqrt{0.116 \left(\frac{1}{157} + \frac{(2.9178 - 2.6179)^2}{20.061} \right)} = 0.0355$$

$$s(\hat{Y}_{ob_3}) = \sqrt{0.116 \left(\frac{1}{157} + \frac{(3.1612 - 2.6179)^2}{20.061} \right)} = 0.0495$$

For 90% joint confidence intervals about these three predictions, the W -value is

$$\sqrt{2F(0.90; 2, 155)} = \sqrt{2(2.337)} = 2.162$$

Using Equation (5), the joint confidence intervals are:

$$3.058 \pm 0.0277(2.162) = 2.998 \leq \hat{Y}_{ob_1} \leq 3.118$$

$$3.942 \pm 0.0355(2.162) = 3.865 \leq \hat{Y}_{ob_2} \leq 4.019$$

$$4.527 \pm 0.0495(2.162) = 4.420 \leq \hat{Y}_{ob_3} \leq 4.634$$

As before, to convert the logarithmic values to arithmetic units, we apply Equation (4). For the three example trees we have:

$$\hat{V}_{ob_1} = \exp(3.058 + 0.116 / 2) = 22.6 \text{ ft}^3$$

$$\hat{V}_{ob_2} = \exp(3.942 + 0.116 / 2) = 54.6 \text{ ft}^3$$

$$\hat{V}_{ob_3} = \exp(4.527 + 0.116 / 2) = 98.0 \text{ ft}^3$$

Applying Equation (4) to the uncertainty limit values we obtain the following intervals:

$$21.2 \leq \hat{V}_{ob_1} \leq 24.0$$

$$50.6 \leq \hat{V}_{ob_2} \leq 59.0$$

$$88.1 \leq \hat{V}_{ob_3} \leq 109.1$$

Prediction and Error from Inventories

In cases where it is desired to estimate removals from a large area where it is impractical to tally all the stumps, a sample-based estimate must be employed. For example, the periodic inventories conducted nationwide by the USDA Forest Service are used to report timber removals to all counties and states. In general, the error of forest inventory estimates has two main components. First is the component due to the random selection of sample plots based on the inventory design, be it systematic, stratified, etc. The second component is associated with the error of the regression function. These two components constitute what is known as the **sampling error**.

An approach proposed by Cunia (1965, 1987a) can be used to combine the error from the sample plots with the error from a regression function. This approach requires that the estimators be of the linear form:

$$\hat{y} = b_1 z_1 + b_2 z_2 + \dots + b_m z_m = \mathbf{b}'\mathbf{z} \quad (7)$$

where \mathbf{b} is the coefficient vector from a regression function and \mathbf{z} is a vector of statistics calculated from the data of the sample points or plots. The variance of y is calculated as:

$$S_{yy} = \mathbf{b}'S_{zz}\mathbf{b} + \mathbf{z}'S_{bb}\mathbf{z} \quad (8)$$

where S_{zz} and S_{bb} are the covariance matrices of \mathbf{z} and \mathbf{b} . The first term of S_{yy} is the variance component associated with the error of the sample plots, and the second term is the variance component associated with the regression function.

It is quite easy to imagine \hat{y} represents $\widehat{\ln V}$ in Equation (7) and to have z_1 be a statistic based on a vector of l 's and

z_2 be a statistic based on $\ln D$. For use in Equation (8), the S_{bb} matrices for Equations (2) and (3) are listed in the Appendix. Cunia (1987a,b,c,d,e,f), in a series of papers, describes in detail the steps of the above approach for making inventory-based predictions, and combining the two error components and constructing confidence intervals, when the sampling designs are: (1) simple random sampling, (2) stratified sampling, (3) two-stage sampling, (4) double sampling, (5) continuous forest inventory (CFI) without sampling with partial replacement (SPR), and (6) CFI with SPR.

Discussion

Predicting stem volume directly from stump diameter is useful in many situations. Volume equations were developed using cubic foot volumes because cubic foot volumes are more accurate estimates of solid-wood volume in a tree stem. Board foot equivalents would depend upon the local merchantability limits and log rules used, while actual lumber recovery would depend on the mill technology employed in the conversion process. For those interested, baldcypress board foot equivalents, or board foot/cubic foot ratios for two sets of merchantability limits for both the Doyle and International 1/4 in. log rules can be found in Hotvedt et al. (1985).

Resource professionals should bear in mind that regression functions like Equations (2) and (3) provide point estimates which have a variance. When evaluating a large group of trees with the same stump diameter, constructing a confidence interval on V_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 $V_{i(\text{new})}$ will provide a range of values that should contain the true volume of the individual. When evaluating a few trees with the same D value, constructing a confidence interval on $\bar{V}_{i(\text{new})}$ 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 values, then joint confidence intervals are necessary.

The ability to estimate removals from forest inventories is a critical function. Fortunately, formulas for many different inventory designs are available to build proper error terms and put bounds on the estimates.

Literature Cited

- BELTZ, R.C., N.D. COST, N.P. KINGSLEY, AND J.R. PETERS. 1992. Timber volume distribution maps for the eastern United States. USDA For. Serv. Gen. Tech. Rep. WO-60. 59 p.
- BYLIN, C.V. 1982. Estimating dbh from stump diameter for 15 southern species. USDA For. Serv. Res. Note SO-286. 3 p.
- CUNIA, T. 1965. Some theory on the reliability of volume estimates in a forest inventory sample. For. Sci. 11:115-128.
- CUNIA, T. 1987a. Error of forest inventory estimates: Its main components. P. 1-13 in Estimating tree biomass regressions and their error. Proc. of the Workshop on Tree Biomass Regression Functions and their Contribution to the Error of Forest Inventory Estimates, Wharton, E.H., and T. Cunia (comps.). USDA For. Serv. Gen. Tech. Rep. NE-1 17. 303 p.
- CUNIA, T. 1987b. On the error of forest inventory estimates: Stratified sampling and double sampling for stratification. P. 63-70 in Estimating tree biomass regressions and their error. Proc. of the Workshop on Tree Biomass Regression Functions and their Contribution to the Error of Forest Inventory Estimates, Wharton, E.H., and T. Cunia (comps.). USDA For. Serv. Gen. Tech. Rep. NE-1 17.

- CUNIA, T. 1987c. On the error of forest inventory estimates: two-stage sampling of plots. P. 71-77 in *Estimating tree biomass regressions and their error. Proc. of the Workshop on Tree Biomass Regression Functions and their Contribution to the Error of Forest Inventory Estimates*, Wharton, E.H., and T. Cunia (comps.). USDA For. Serv. Gen. Tech. Rep. NE-1 17.
- CUNIA, T. 1987d. On the **error of** forest inventory estimates: Double sampling with regression. P. 79-87 in *Estimating tree biomass regressions and their error. Proc. of the Workshop on Tree Biomass Regression Functions and their Contribution to the Error of Forest Inventory Estimates*, Wharton, E.H., and T. Cunia (comps.). USDA For. Serv. Gen. Tech. Rep. NE-1 17.
- CUNIA, T. 1987e. On the **error of** forest inventory estimates: Continuous forest inventory without SPR. P. 89-98 in *Estimating tree biomass regressions and their error. Proc. of the Workshop on Tree Biomass Regression Functions and their Contribution to the Error of Forest Inventory Estimates*, Wharton, E.H., and T. Cunia (comps.). USDA For. Serv. Gen. Tech. Rep. NE-1 17.
- CUNIA, T. 1987f. On the **error of** forest inventory estimates: Continuous forest inventory with SPR. P. 99-106 in *Estimating tree biomass regressions and their error. Proc. of the Workshop on Tree Biomass Regression Functions and their Contribution to the Error of Forest Inventory Estimates*, Wharton, E.H., and T. Cunia (comps.). USDA For. Serv. Gen. Tech. Rep. NE-1 17.
- DRAPER, N.R., AND H. SMITH. 1981 *Applied regression analysis*. Ed. 2. Wiley, New York. 709 p.
- HOTVEDT, J.E., Q.V. CAO, AND B.R. PARRESOL. 1985. Tree-volume and stem-profile functions for baldcypress. *South. J. Appl. For.* 9(10):227-232.
- NETER, J., W. WASSERMAN, AND M.H. KUTNER. 1985. *Applied linear statistical models*. Ed. 2. Richard D. Irwin, Homewood, IL. 1127 p.
- WHARTON, E.H. 1984. Predicting diameter at breast height from stump diameters for northeastern tree species. USDA For. Serv. Res. Note NE-322.4 p.
- YANDLE, D.O., AND H.V. WILSON. 1981. Biomass estimates based on the simple power function-some calculation considerations. P. 350-354 in *Proc. In-place Resource Inventories: Principle and Practices Workshop*. Soc. Am. For., Bethesda, MD. 1 101 p.

APPENDIX

The covariance matrix of the coefficient vector \mathbf{b} for Equation (2), the volume inside bark estimator, is

$$S_{bb} = \begin{bmatrix} 0.04119 & -0.01545 \\ -0.01545 & 0.00590 \end{bmatrix}$$

The covariance matrix of the coefficient vector \mathbf{b} for Equation (3), the volume outside bark estimator, is

$$S_{bb} = \begin{bmatrix} 0.04037 & -0.01514 \\ -0.01514 & 0.00578 \end{bmatrix}$$