USDA United States Department of Agriculture

Forest Service

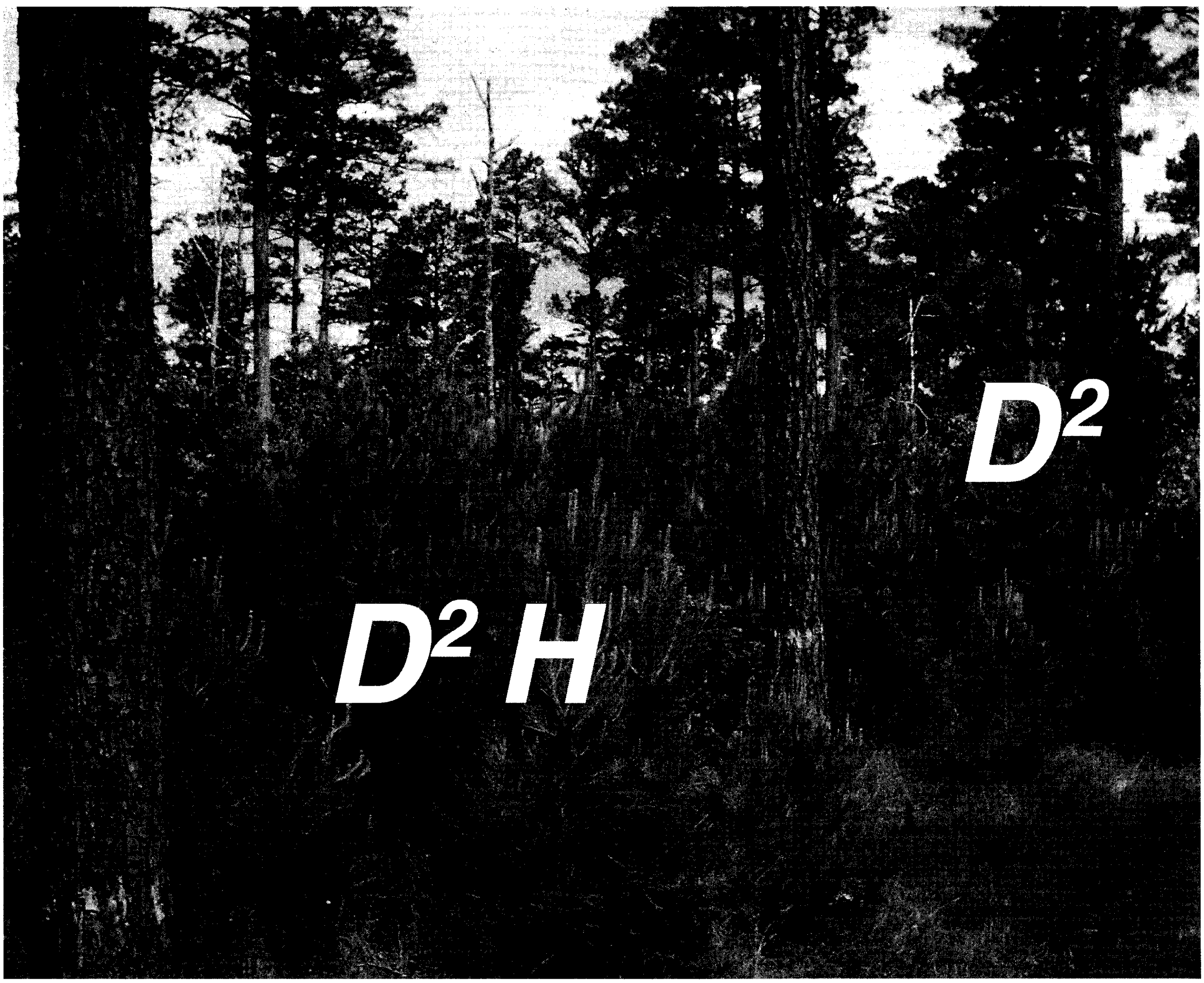
Rocky Mountain Research Station

Fort Collins, Colorado 80526

Research Paper RMRS–RP–6

Weighted Linear Regression Using D^2H and D^2 as the Independent Variables

Hans T. Schreuder and Michael S. Williams



Abstract

Schreuder, Hans T. and Michael S. Williams. 1998. **Weighted linear regression using D^2H and D^2 as the independent variables.** Research Paper RMRS-RP-6. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 10 p.

Several error structures for weighted regression equations used for predicting volume were examined for 2 large data sets of felled and standing loblolly pine trees (*Pinus taeda* L.). The generally accepted model with variance of error proportional to the value of the covariate squared (D^2H = diameter squared times height or D^2 = diameter squared) remains the best.

Although D^2H is a better covariate than D^2, we found no significant difference between them when testing model accuracy for felled trees, but there were significant differences for standing trees. When we predicted the total volume of a population using equations based on felled tree data, assuming known frequencies for diameter classes and using D^2H as the covariate, we obtained essentially the same estimate as that predicted using D^2 (0.1% difference). Using the conventional approach of D^2H for all trees (standing and felled) yielded an estimate of volume of 5.6% less than using the equation with D^2H for felled trees only.

Trees are more accurately measured for volume when felled, and total heights are often not measured accurately on standing trees. Therefore, we recommend that volume equations be based on felled tree data only and that when they are intended to be applied to standing trees, D^2 be used as the covariate in prediction.

Keywords: best volume equations, felled trees, standing trees, measured heights, basal area

The Authors

Hans T. Schreuder and **Michael S. Williams** are mathematical statisticians at the Rocky Mountain Research Station in Fort Collins.

Publisher

Rocky Mountain Research Station

Fort Collins, Colorado

June 1998

You may order additional copies of this publication by sending your mailing information in label form through one of the following media. Please send the publication title and number.

Telephone	(970) 498-1719
E-mail	rschneider/rmrs@fs.fed.us
FAX	(970) 498-1660
Mailing Address	Publications Distribution Rocky Mountain Research Station 3825 E. Mulberry Street Fort Collins, CO 80524-8597

Weighted Linear Regression Using D^2H and D^2 as the Independent Variables

Hans T. Schreuder and Michael S. Williams

Contents

Introduction

Weighted linear regression is widely used to estimate total tree volume (V) as a function of tree diameter at breast height squared times total tree height. One volume model commonly used is:

$$V_i = \alpha + \beta D_i^2 H_i + e_i \quad [1]$$

with $Ee_ie_j = \sigma^2 v_i$ for $i = j$, and $Ee_ie_j = 0$ for $i \neq j$, where D_i is the diameter at breast height and H_i is the total tree height. This is the best model to use if V_i, D_i, and H_i are measured essentially without error in model development, and if D_i and H_i are measured equally reliably for volume predicted trees. Because identifying the top of hardwoods and seeing the tops of conifers in many natural forest stands and closely-spaced plantations is difficult, errors often occur when measuring or estimating standing tree height. In addition, measuring height is time consuming. Consequently, alternative models are appealing:

$$V_i = \alpha + \beta D_i^2 + e_i, \quad [2]$$

and:

$$V_i = \alpha + \beta D_i + \gamma D_i^2 + e_i, \quad [3]$$

where the error structure is assumed to be similar to that in [1].

We determined the best weight functions for $\sigma^2 v_i$ with models [1]-[3] based on felled tree data. We also compared the efficiency and bias of models [1]-[3] in predicting the total volume of a forest. Model [1], with heights and diameters measured on felled trees and volumes derived, is the standard. This standard is compared with the best of models [2] and [3], with diameters measured on standing trees and with model [1] with diameter and height measured on standing and felled trees.

Literature Review

Cunia (1964), McClure et al. (1983), and Gregoire and Dyer (1989) have studied the error model, $v_i = (D_i^2 H_i)^{k_1}$, for various tree species and suggest k_1 values ranging from 1.01 - 2.07.

Meng and Tsai (1986) proposed using the error model $v_i = (D_i^{\lambda})^2$ in conjunction with model [3]. They studied red pine (N = 92, *Pinus resinosa* Ait.) and white pine (N = 14, *Pinus strobus* L.) using a method described by Box and Cox (1964). They estimated 95% confidence intervals of $0.55 < \lambda < 1.80$ for red pine and $1.90 < \lambda < 2.18$ for white pine.

Williams et al. (1993) studied model [1] with 4 different weight functions

$$v_i = x_i^{k_1}, \quad [4]$$

$$v_i = (x_i + k_2 x_i^{k_3}), \quad [5]$$

$$v_i = (1 + k_4 x_i + k_5 x_i^{k_6}), \quad [6]$$

and

$$v_i = (1 + k_7 x_i)^2, \quad [7]$$

where $x_i = D_i^2 H_i$. They found no significant improvement when using the more complex weight functions and suggested using $v_i = x_i^{k_i}$ with $k_1 = 2$. Williams and Gregoire (1993) tested the function $v_i = D_i^{k_1} H_i^{k_2}$ and found it to be a significant improvement over $v_i = (D_i^2 H_i)^{k_3}$. We did not consider the model $v_i = D_i^{k_1} H_i^{k_2}$ when comparing D^2H to D^2 models because the basal area weighting function is no different and $v_i = (D_i^2 H_i)^k$ is widely used.

Williams et al. (1994) studied the performance of 5 instruments. They found that for trees less than 40 ft tall, the average measurement error was generally less than 5% but as height increased, error also increased. In January 1991, a large group of forest biometricians and mensurationists in Fort Collins, Colorado, agreed that heights of tall standing trees cannot be measured reliably.

Gregoire and Williams (1992) identified and evaluated the errors in tree volume associated with estimating tree height. They found the mean square error of volume prediction increased 35% to 38% when estimated heights were used in place of actual heights.

Data Description

Since 1963, the Forest Inventory and Analysis (FIA) unit in the southeastern US has measured the volumes of individual standing and felled trees on a subsample of all regular FIA sample plots in Virginia, North Carolina, South Carolina, Georgia, and Florida. Of the trees in this data set, 3748 were felled before measurement to ensure accuracy. The remaining 3133 trees were measured standing using height poles and calipers. All these trees were measured uniformly as a series of tapering sections (diameter outside bark at both ends of each section and section length) by highly trained inventory specialists. Measurements were taken from ground level to the tip of the main stem and from the base to the tip of each fork (Cost 1978).

Each tree's location, species, diameter at breast height (dbh), double bark thickness at dbh, and total height were also recorded. Double-bark thicknesses at various points on the main stem were measured and recorded for all felled tree samples. Volumes of sections were computed using Grosenbaugh's (1952) cubic-foot volume equation, which allows for form and taper. Total-stem volume and total height were obtained by summing all the section volumes and lengths respectively.

Objectives and Criteria for Evaluation

There are 2 advantages to accurate weight functions in linear regression. First, estimates of the variance-covariance matrix are biased if the error structure is incorrectly specified. Second, a better fitting weight function leads to improved precision in estimation and is more likely to produce reliable confidence intervals. The objectives of this study were to:

1. establish which of the 15 models discussed below, using D^2H and D^2 as covariates, is the best for predicting volume;
2. show the effect of different models on estimation of volume, and determine if only diameters should be used, as in models [2] and [3], to predict volumes.

Methods

The 15 models studied in a comparison of their performance comprised the 3 volume models (equations 1-3) combined with each of 5 error models. These latter included the 4 weighting functions used by Williams et al. (1993) and the following additional exponential weighting function:

$$EM1{:}\sigma^2 v_i = \sigma^2 x_i^{k_1},$$

which has been studied by Cunia (1964), McClure et al. (1983), Meng and Tsai (1986), Kelly and Beltz (1987), and Gregoire and Dyer (1989). Model:

$$EM2{:}\sigma^2 v_i = \sigma^2 (x_i + k_2 x_i^{k_3}),$$

which was suggested by Scott et al. (1978). Model:

$$EM3{:}\sigma^2 v_i = \sigma^2 (1 + k_4 x_i + k_5 x_i^{k_6})$$

is a generalization of $EM2$. Model:

$$EM4{:}\sigma^2 v_i = \sigma^2 e^{k_7 x_i},$$

is an exponential error model used by Williams (1994). In addition, we considered a 5th model:

$$EM5{:}\sigma^2 v_i = \sigma^2 (1 + k_8 x_i)^2,$$

which is appealing since it is non-negative for any value of k_8. For models to be fully specified, we combined one of models [1]-[3] with one of $EM1 - EM5$. Hence, 15 fully specified volume models were fitted, where $x_i = D_i^2 H_i$ or D_i^2, depending on the covariate used in the 3 volume models. Volume (V_i) is expressed in m^3, diameter (D_i) in centimeters, and height (H_i) in meters.

The log likelihood function for either model [1] or [2] is:

$$\ln L = -\frac{N}{2}\ln(2\pi) + \frac{1}{2}\ln|\Omega^{-1}| - \frac{1}{2}\underline{\in}^T \Omega^{-1} \underline{\in}, \qquad [8]$$

where $\underline{\in}$ is a ($Nx1$) column vector of residuals $V_i - \alpha - \beta x_i^2$ and $\Omega = E[\underline{\in}\underline{\in}^T] = \sigma^2$ diag (v_i), where x_i is either $D_i^2 H_i$ or D_i^2 and the v_i's are calculated using models $EM1 - EM5$.. For volume equation [3] the log-likelihood is given by [8] with residual $V_i - a - bD_i - gD_i^2$. Ordinary least squares (OLS) estimates of α, β, and γ were used as starting values.

If $\phi_i = \phi_2 = (\alpha, \beta)$ for volume equations [1] and [2] and $\phi_3 = (\alpha, \beta, \gamma)$ for [3], then the parameter vectors for maximum likelihood estimation (MLE) are $\underline{\theta}_{aj} = (\underline{\phi}_j, \sigma^2, k_1)$, $\underline{\theta}_{bj} = (\underline{\phi}_j, \beta, \sigma^2, k_2, k_3)$, $\underline{\theta}_{cj} = (\underline{\phi}_j, \sigma^2, k_4, k_5, k_6)$, $\underline{\theta}_{dj} = (\underline{\phi}_j, \sigma^2, k_7)$, and $\underline{\theta}_{ej} = (\underline{\phi}_j, \sigma^2, k_8)$. The solutions to the MLE are the $\underline{\theta}_{kj}$ that minimize $-\ln L$ for the k values in $EM1 - EM5, j = 1,2,3.$. The minimum of $-\ln L$ with respect to $\underline{\theta}_{lj}, l = [a],...[e], j = 1,2,3$ was found using IMSL[1] routine DBCOAH. This constrained optimization routine uses a modified Newton's method to find min ($-\ln L$) with respect to $\underline{\theta}_{lj}$ subject to the constraint $\sigma^2 > 0$.

[1] *International Mathematical Statistical Laboratory, 2500 Park West Tower One, 2500 City West Boulevard, Houston, TX 77402*

A consistent estimator of the asymptotic variance-covariance matrix is given by the inverse of the information matrix (Mood et al. 1974), which can be expressed as the expectation (E) of the information matrix ($I(\underline{\theta}_{lj})$):

$$E[I(\underline{\theta}_{lj})] = (E[\frac{\partial^2(-\ln L)}{\partial\underline{\theta}_{lj}^2}]), \qquad [9]$$

with the k values in $EM1-EM5$, $j = 1,2,3$. Using the information matrix, confidence intervals of all parameters can be obtained.

Initially we tested which of the models [1]-[3] gave the best estimates of volume. For the volume-basal area models, the hypothesis tested was:

$$H_0: V_i = \alpha + \beta D_i^2$$

versus:

$$H_1: V_i = \alpha + \beta D_i + \gamma D_i^2.$$

Rejecting the null hypothesis would indicate that volume equation [3] was superior to [2]. To test this hypothesis, the asymptotic distribution of the generalized likelihood-ratio (Mood et al. 1974) was used. This test was applicable because volume equation [2] is a special case of [3], where $\beta = 0$ in [3]. The test with approximate size η is given by the following:

$$\text{Reject } H_0 \text{ if and only if } -2\ln\frac{\sup_{\theta_{l2}} L}{\sup_{\theta_{l3}} L} > \chi^2_{1-\eta,1}$$

with the k values in $EM1-EM5$, where $\sup_{\theta_{l3}} L$ is the maximum likelihood values generated using equation [3] and $\sup_{\theta_{l2}} L$ is the maximum likelihood values generated using equation [2] with weight function k values in $EM1-EM5$, respectively.

Comparing the volume equation [1] against [2] or [3] was not as straight-forward, because the latter were not special cases of model [1]. To test model [3] against model [1] the hypothesis used was:

$$H_0: V_i = \alpha + \beta D_i + \gamma D_i^2$$

versus:

$$H_1: V_i = \alpha + \beta D_i^2 H_i,$$

and the test was:

$$\text{Reject } H_0\text{: if and only if } -2\ln\frac{\sup_{\theta_{l3}} L}{\sup_{\theta_{l1}} L} > \chi^2_{1-\eta,1}. \qquad [10]$$

This gives a conservative test of level η, since the denominator $\sup_{\theta_{l1}} L$ is necessarily smaller than the denominator traditionally used in the generalized likelihood ratio test. The conservative test can be applied to compare models [2] and [1] but is unnecessary if the null hypothesis in [10] is rejected.

To determine which weight function best describes the actual variance, an index derived by Furnival (1961) and Cox (1961) was used. This index compares the fit of different weight functions and can be written as:

$$I = [anti\log\frac{\sum_{i=1}^{n}\log_{10}\sqrt{v_i}}{n}]^{-1}S,$$

where S is the standard error about the model. Smaller values of the index indicate a better fit.The maximized log likelihood values were also given. Larger values generally indicate a better fit.

Results and Discussion

Objective 1 — Best Prediction Model

The results in tables 1 and 2 show that inclusion of the linear diameter term in [3] gave a statistically superior estimate of volume relative to [2] regardless of the weight function used. Given a particular weight function from $EM1-EM5$, model [1] with D^2H as covariate outperformed models [2] and [3]. Differences in the likelihoods for model [1] were so much better than for [3] for every weight function that any concerns about the testing method used are irrelevant. Due to the poor performance of [2], further discussion is limited to models [1] and [3] only. For both of these volume equations, no solution was found for weight function $EM3$. This was due to numerical problems, where the change in the log-likelihood function was less than machine precision. Furnival's index of fit consistently indicated that weight function $EM2$ produced the best results, generally followed by $EM5$ and $EM1$. Weight function $EM4$, with the smallest log likelihoods was clearly unsatisfactory.

Further study revealed some problems with using weight function $EM2$. The relationship between the linear and nonlinear term was inconsistent, with the k_2 parameter ranging in value from 4.167 to 15560.0. This great variability combined with very wide parameter confidence intervals made a general weight function based on $EM2$ impractical and indicated overparameterization of the function. Comparing the felled tree and standing tree results indicated further problems with $EM2$, with k_2 generally being approximately an order of magnitude larger for the standing tree data set. This was true for k_8 in $EM5$, too. The most robust weight function was $EM1$, which supports the finding of Williams et al. (1993). For both data sets and volume equations [1] and [3], $v_i = x_i^2$ was a reasonable recommendation.

Objective 2 — Effect of Different Models on Estimation of Volume, and Determination of Using Only Diameter to Predict Volumes

The Forest and Inventory Analysis units, including the unit in Asheville, North Carolina, have a good estimate of the tree frequency in each state by species and diameter class (2 inch, 5 cm). For purposes of comparison, we used frequency estimates by diameter class (except for last 2) for 1989 in Georgia (Sheffield and Johnson 1993). If there are N_D such classes and f_i is the estimated frequency of class $i(i = 1, \ldots, N_D)$, an estimate of total volume (V_T) using the regression coefficients of $EM1$ for the felled tree data (table 1) is:

$$\hat{V}_T = \sum_{l=1}^{N_D} f_l[-.003918 + 0.3350\overline{D}_l^2\overline{H}_l]$$

We used $\overline{H}_l = \sum_{i=1}^{n_l} H_{il} / n_l$ and $\overline{D}_l^2 = \sum_{i=1}^{n_l} \overline{D}_{il}^2 / n_l$, as the average D^2 and H in each diameter class, where n_l is the number of trees in diameter class $l = 1, \ldots, N_D$ (table 3).

Because we had to use estimates like $\overline{H}_l$ and $\overline{D}_l^2$, we could not compute meaningful standard errors for our estimates. The more conventional estimate of V_T would be based on an equation using both standing and felled trees volume (not shown in the tables):

$$\hat{V}_i = 0.0008378 + 0.3309 D_i^2 H_i \cdot$$

Using this equation, our estimate of V_T was:

$$\hat{V}_{T1} = \sum_{l=1}^{N_D} f_l[0.0008378 + 0.3309\overline{D}_l^2\overline{H}_l].$$

A reasonable alternative, which ignores the information on heights and uses the regression coefficients of $EM1$ for felled tree data (table 1), was:

$$\hat{V}_{T2} = \sum_{l=1}^{N_D} f_l[0.09743 - 1.861\overline{D}_l + 12.52\overline{D}_l^2].$$

This estimator avoids the information on heights in the felled-tree sample and in the large-scale sample. This may be an advantage because it is unknown whether this is a representative sample of heights for a given diameter class, and these heights are measured or estimated with possibly large errors.

For comparison, we also computed:

$$\hat{V}_{T3} = \sum_{l=1}^{N_D} f_l[0.01943 - 0.8525\overline{D}_l + 9.543\overline{D}_l^2]$$

using regression coefficients computed from all trees but ignoring heights.

We tested the difference in model accuracy for the felled and standing trees using D^2H and D^2 as covariates (table 4). As noted, the felled trees were assumed measured without error relative to the standing trees. Using the accuracy test developed by Reynolds (1984) and implemented by Gribko and Wiant (1992) and Wiant (1993) , mean error was used as a measure of bias. This test involves constructing confidence intervals (CI) around the mean error to determine if the CI contains 0. If it does, bias is insignificant.

There was significant difference in model accuracy for the felled trees if either D^2H or D^2 was used as the covariate, even though D^2H was more reliable. However, there was a significant difference in accuracy between standing and felled trees when D^2H was used as the covariate. Of less practical use was the significant difference in model accuracy when D^2H or D^2 was used with standing trees.

Using D^2H as a covariate with both felled and standing trees resulted in an estimate ($\hat{V}_{T1}$) yielding 5.6% less than our best estimate of total volume, ($\hat{V}_T$) (table 5). Using only D^2 for felled trees yielded an estimate ($\hat{V}_{T2}$) identical to $\hat{V}_T$. $\hat{V}_{T3}$, using D^2 for both felled and standing trees yielded an estimate only 0.1% less than $\hat{V}_T$. This result was unexpected because it was difficult to measure the volume of standing trees as accurately as that of felled trees. Apparently, the errors in measurement of the standing trees were compensatory.

Conclusions

Based on Furnival's index, it was clear that D^2H was a better covariate than D^2 for predicting volume of both felled and standing trees. The new error structures tested for both covariates were less satisfactory than the traditional one, $V(e_i) = \sigma^2 x_i^{k_1}$. As in other studies, a value of $k_1 = 1.5, 2.0$, or some value in between, seems reasonable.

In terms of model accuracy, there was no difference in bias when using D^2H or D^2 as the covariate for the felled trees, but there was for standing trees. Similarly the use of D^2H for felled vs. standing trees gave significant differences in accuracy as did felled D^2H vs. all D^2. This is important because the use of the prediction equation with D^2H for all trees yielded an estimate 5.6% of that obtained using the more accurate prediction equation with D^2H for felled trees only. In contrast, for felled trees the prediction equation with D^2 gave an estimate identical to that with D^2H. We recommend that volume prediction be based on felled tree data only and that D^2 be used as the covariate if predictions will also be made for standing trees. This should provide more precise estimates of total wood volume in a stand for lower cost since tree heights are expensive to measure.

Literature Cited

Box, G.E.P., and D. R. Cox, 1964. An analysis of transformation. J. R. Stat. Soc. B, 26: 211-243.

Cox, D. R. 1961. Tests of separate families of hypotheses. 4th Berkeley Symp. Proc. 1: 101-123.

Cost, N. D. 1978. Multiresource inventories: a technique for measuring volumes in standing trees. SE Forest Exp. Sta. USDA For. Serv. Res. Pap. SE-196. 18 p.

Cunia, T. 1964. Weighted least squares method and construction of volume tables. For. Sci. 10: 180-191.

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

Gregoire, T. G., and M. E. Dyer. 1989. Model fitting under patterned heterogeneity of variance. For. Sci. 35: 105-125.

Gregoire, T. G., and M. Williams. 1992. Identifying and evaluating the components of non-measurement error in the application of standard volume equations. The Statistician. 41:509-518.

Gribko, L. S., and Wiant, H. V. 1992. A SAS Template Program for the Accuracy Test. The Compiler 10(1): 48-51.

Grosenbaugh, L. R. 1952. Shortcuts for cruisers and scalers. Southern For. Exp. Sta. USDA For. Serv. Occas. Pap. 126. 24 pp.

Kelly, J. F., and Beltz, R. C. 1987. A comparison of tree volume estimation models for forest inventory. USDA For. Serv. So. For. Exp. Sta. Res. Paper SO-233. 9 pp.

McClure, J. P., Schreuder, H. T., and Wilson, R. L. 1983. A comparison of several volume table equations for loblolly pine and white oak. USDA For. Serv. SE For. Exp. Sta. Res. Paper SE-240. 8 pp.

Meng, C. H., W. Y. Tsai. 1986. The selection of weights for a weighted regression of tree volume. Can. J. For. Res. 16:671- 673.

Mood, A. M., Graybill, F. A., and Boes, D. C. 1974. Introduction to the Theory of Statistics. 3rd edition, McGraw Hill, New York. 564 pp.

Reynolds, M. R., Jr. 1984. Estimating the error in model predictions. For. Sci. 30: 454-469.

Scott, A. J., Brewer, K. R. W., and Ho, E. W. H. 1978. Finite population sampling and robust estimation. J. Amer. Stat. Assoc. 73: 359-361.

Sheffield, R. M. and Johnson, T. G. 1993. Georgia's Forests, 1989. USDA Forest Service Resource Bulletin SE-133. 97 p.

Wiant, H. V., Jr. 1993. DOSATEST for testing accuracy. The Compiler 11(3): 28-29.

Williams, M. S. 1994. Evaluating the performance of weight functions with no constraints on function types. For Sci. 40: 787-793.

Williams, M. S., Bechtold, W. E., and LaBau, J. 1994. Height measurement evaluation. South. J. Appl. For. 18: 76-82.

Williams, M. S., and Gregoire T. G. 1993. Estimating weights when fitting linear regression models for tree volume. Can. J. For. Res. 23:1725-1731.

Williams, M. S., Schreuder H. T., Gregoire T. G., and Bechtold, W. 1993. Estimating variance functions for weighted linear regression. In: 'Applied Statistics in Agriculture', G.A. Milliken and J.R. Schwenke (eds.). Proc. Conf. April 26-28, 1992, Manhattan, KS, Manhattan State Univ. p. 153-160

Tables 1 – 5 follow

Table 1. Estimates of parameters, 95% confidence interval limits, Furnival's index, and log likelihood values for various weighting factors in the felled loblolly pine tree data set.

Parameter	Volume equation $V_i = \alpha + \beta D_i^2 H_i$	$V_i = \alpha + \beta D_i^2$	$V_i = \alpha + \beta D_i + \gamma D_i^2$
EM_i $\sigma^2(x_i)^{k_1}$			
α	$-.3918x10^{-2} \pm 0.8519x10^{-3}$	$-.6955x10^{-1} \pm 0.2515x10^{-2}$	$0.9743x10^{-1} \pm 0.7874x10^{-2}$
β	$0.3350 \pm 0.1573x10^{-2}$	$0.8160x10 \pm 0.7418x10^{-1}$	$-.1861x10 \pm 0.9094x10^{-1}$
γ			$0.1252x10^2 \pm 0.2379$
σ^2	$0.1631x10^{-2} \pm 0.7587x10^{-4}$	$0.3338x10 \pm 0.4206$	$0.5444x10 \pm 0.6859$
k_1	$0.1859x10 \pm 0.6346x10^{-1}$	$0.2120x10 \pm 0.5397$	$0.2427x10 \pm 0.5397x10^{-1}$
Furnival's Index	$0.5345x10^{-1}$	0.1065	0.1501
Log Likelihood	5659.9	3073.5	3701.2
EM_2 $\sigma^2(x_1 + k_2 x_1^{k_3})$			
α	$-.5242x10^{-2} \pm 0.9679x10^{-3}$	$-.8031x10^{-1} \pm 0.3008x10^{-2}$	$0.9795x10^{-1} \pm 0.7974x10^{-2}$
β	$0.3359 \pm 0.1621x10^{-2}$	$0.8240x10 \pm 0.7750x10^{-1}$	$-.1866x10 \pm 0.9177x10^{-1}$
γ			$0.1253x10^2 \pm 0.2394$
σ^2	$0.2818x10^{-3} \pm 0.1091x10^{-3}$	$0.4113x10^{-1} \pm 0.6211x10^{-2}$	$0.3539x10^{-3} \pm 0.3340x10^{-2}$
k_2	$0.4167x10 \pm 0.2200x10$	$0.4245x10^3 \pm 0.1221x10^3$	$0.1556x10^5 \pm 0.1445x10^6$
k_3	$0.2098x10 \pm 0.1143x10$	$0.3008x10 \pm 0.1766$	$0.2433x10 \pm 0.9712x10^{-1}$
Furnival's Index	$0.5326x10^{-1}$	0.1034	$0.9015x10^{-1}$
Log Likelihood	5673.9	3188.1	3700.9
EM_3 $\sigma^2(1 + k_4 x_i + k_5 x_i^{k_6})$	No Useful Results		
EM_4 $\sigma^2 e^{k_7 x_i}$			
α	$-.1021x10^{-1} \pm 0.1906x10^{-2}$	$-.9564x10^{-1} \pm 0.3953x10^{-2}$	$0.1191 \pm 0.1818x10^{-1}$
β	$0.3404 \pm 0.1810x10^{-2}$	$0.8490x10 \pm 0.8050x10^{-1}$	$-.2089x10 \pm 0.1744$
γ			$0.1304x10^2 \pm 0.3914$
σ^2	$0.5041x10^{-3} \pm 0.2378x10^{-4}$	$0.1123x10^{-2} \pm 0.1792x10^{-3}$	$0.8850x10^{-3} \pm 0.1413x10^{-3}$
k_8	$0.9401 \pm 0.4398x10^{-1}$	$0.2566x10^2 \pm 0.5709x10^{-1}$	$0.2679x10^2 \pm 0.5709x10^{-1}$

Table 1. Cont'd.

Parameter	Volume equation $V_i = \alpha + \beta D_i^2 H_i$	$V_i = \alpha + \beta D_i^2$	$V_i = \alpha + \beta D_i + \gamma D_i^2$
Furnivall's Index	$0.6097x10^{-1}$	0.1078	0.6555
Log Likelihood	5166.9	3032.1	3286.6
EM_5 $\sigma^2(1+k_8 x_i)^2$			
α	$-.5427x10^{-2} \pm 0.9864x10^{-3}$	$-.8124x10^{-1} \pm 0.2988x10^{-2}$	$0.1019 \pm 0.1007x10^{-1}$
β	$0.3362x10 \pm 0.1612x10^{-2}$	$0.8429x10 \pm 0.7471x10^{-1}$	$-.1911x10 \pm 0.1066$
γ			$0.1264x10^{2} \pm 0.2515$
σ^2	$0.7667x10^{-5} \pm 0.3863x10^{-5}$	$0.2253x10^{-4} \pm 0.2006x10^{-4}$	$0.6277x10^{-8} \pm 0.2616x10^{-6}$
k_8	$0.1301x10^{2} \pm 0.3552x10$	$0.3073x10^{3} \pm 0.1452x10^{3}$	$0.1704x10^{5} \pm 0.3556x10^{6}$
Furnival's Index	$0.5326x10^{-2}$	0.1063	$0.9258x10^{-1}$
Log Likelihood	5674.0	3082.8	3601.2

Table 2. Estimates of parameters, 95% confidence interval limits, Furnival's index, and log likelihood values for various weighting factors in the standing loblolly pine.

Parameter	Volume equation: $V_i = \alpha + \beta D_i^2 H_i$	Volume equation: $V_i = \alpha + \beta D_i^2$	Volume equation: $V_i = \alpha + \beta D_i + \gamma D_i^2$
$EM_1\ \sigma^2(x)^{k_1}$			
α	$0.8742x10^{-3} \pm 0.8656x10^{-4}$	$-.4278x10^{-2} \pm 0.2839x10^{-3}$	$0.1131x10^{-1} \pm 0.6603x10^{-3}$
β	$0.3310 \pm 0.1605x10^{-2}$	$0.5438x10 \pm 0.7418x10^{-1}$	$-.5472 \pm 0.2285x10^{-1}$
γ			$0.7794x10 \pm 0.1162$
σ^2	$0.2001x10^{-2} \pm 0.1023x10^{-3}$	$0.6220x10 \pm 0.7636$	$0.5042x10 \pm 0.6190$
k_1	$0.1701x10 \pm 0.4199x10^{-1}$	$0.2215x10 \pm 0.4389x10^{-1}$	$0.2319x10 \pm 0.4389x10^{-1}$
Furnival's Index	$0.3330x10^{-1}$	$0.8402x10^{-1}$	0.1404
Log Likelihood	6213.6	3314.4.5	4142.5
$EM_2\ \sigma^2(x_i + k^2 x_i^{k_3})$			
α	$0.1013x10^{-2} \pm 0.1295x10^{-3}$	$-.9257x10^{-2} \pm 0.8244x10^{-3}$	$0.1652x10^{-1} \pm 0.1057x10^{-2}$
β	$0.3298 \pm 0.1612x10^{-2}$	$0.5429x10 \pm 0.6556x10^{-1}$	$-.6874 \pm 0.2963x10^{-1}$
γ			$0.8279x10^{1} \pm 0.1365$
σ^2	$0.8499x10^{-4} \pm 0.2698x10^{-4}$	$0.1685x10^{-1} \pm 0.2987x10^{-2}$	$0.2808x10^{-2} \pm 0.7951x10^{-3}$
k_2	$0.2118x10^{2} \pm 0.7388x10$	$0.1656x10^{4} \pm 0.4124x10^{3}$	$0.2870x10^{4} \pm 0.8175x10^{3}$
k_3	$0.1921x10 \pm 0.6205x10^{-1}$	$0.2886x10 \pm 0.1068$	$0.2530x10 \pm 0.7539x10^{-1}$
Furnival's Index	$0.3288x10^{-1}$	$0.8134x10^{-1}$	$0.1636x10^{-1}$
Log Likelihood	6254.6	3416.4	4187.9
$EM_3\ \sigma^2(1 + k_4 x_i + k_5 x_i^{k_6})$	**No useful results**		

Table 2. Cont'd.

Parameter	Volume equation $V_i = \alpha + \beta D_i^2 H_i$	$V_i = \alpha + \beta D_i^2$	$V_i = \alpha + \beta D_i + \gamma D_i^2$
$EM_4\ \sigma^2 e^{k_7 x_i}$			
α	$-.1707x10^{-2} \pm 0.1363x10^{-2}$	$-.4419x10^{-1} \pm 0.3028x10^{2}$	$0.3922x10^{-1} \pm 0.6318x10^{-2}$
β	$0.3334 \pm 0.3057x10^{-2}$	$0.6608x10 \pm -.8209x10^{-1}$	$-.1148x10 \pm 0.8213x10^{-1}$
γ			$0.9849x10 \pm 0.2501$
σ^2	$0.3005x10^{-3} \pm 0.1529x10^{-4}$	$0.8834x10^{-3} \pm 0.1296x10^{3}$	$0.6054x10^{-3} \pm 0.8881x10^{-4}$
k_8	$0.1275x10 \pm 0.3359x10^{-1}$	$0.2933x10^{2} \pm 0.4510x10^{1}$	$0.3163x10^{2} \pm 0.4510x10^{-1}$
Furnival's Index	$0.4316x10^{-1}$	$0.8632x10^{-1}$	0.3147
Log Likelihood	5401.5	3229.9	3560.6
$EM_5\ \sigma^2(1+k_8x_i)^2$			
α	$0.1182x10^{-2} \pm 0.1510x10^{-3}$	$-.9469x10^{-2} \pm 0.8168x10^{3}$	$0.1773x10^{-1} \pm 0.1113x10^{-2}$
β	$0.3290 \pm 0.1640x10^{-2}$	$0.5911x10 \pm 0.6791x10^{1}$	$-.7509 \pm 0.3220x10^{-1}$
γ			$0.8659x10 \pm 0.1342x10^{-1}$
σ^2	$0.3572x10^{-6} \pm 0.1276x10^{-6}$	$0.5079x10^{5} \pm -.2603x10^{5}$	$0.1839x10^{6} \pm 0.2489x10^{-6}$
k_8	$0.7227x10^{2} \pm 0.1353x10^{2}$	$0.7522x10^{3} \pm 0.2022x10^{3}$	$0.3204x10^{4} \pm 0.2200x10^{4}$
Furnival's Index	$0.3295x10^{-1}$	$0.8424x10^{1}$	$0.6538x10^{1}$
Log Likelihood	6247.4	3306.5	4100.0

Table 3. Frequency in Forest Inventory and Analysis samples. (N), mean diameter $(\overline{D}_F)$, mean height $(\overline{H}_F)$, frequency of felled trees (N_F), mean diameter $(\overline{D}_a)$, mean height $(\overline{H}_a)$, and frequency of all volume trees (N_a) by diameter class (DBH class).

DBH class (cm)	N (in thousands)	(D_F) (m)	(H_F) (m)	(N_F) (m)	(D_a) (m)	(H_a)	(N_a)
2.5-7.5	999,774	.07	8.2	1	.05	5.27	174
7.5-12.5	473,529	.11	12.1	120	.10	10.50	248
12.5-17.5	275,981	.15	14.4	606	.15	13.41	1234
17.5-22.5	161,006	.20	17.5	598	.20	16.24	1223
22.5-27.5	101,320	.25	20.0	588	.25	19.04	1063
27.5-32.5	61,759	.30	22.1	600	.30	20.98	1002
32.5-37.5	34,929	.35	23.8	542	.35	22.76	850
37.5-42.5	18,470	.40	25.4	356	.40	24.17	559
42.5-47.5	9,161	.45	26.2	183	.45	25.12	290
47.5-52.5	4,588	.50	27.5	93	.50	25.91	143
52.5-72.5	2,999	.57	28.0	58	.57	26.46	89
>72.5	83	.86	30.8	3	.84	29.57	6

Table 4. Test for model accuracy using DOSATEST.

Models	Bias	CI [a]	Significant sample size
$D^2H_{Felled}\,vs.\,D^2{}_{Felled}$	$+2.22x10^{-3}$	$\pm 4.22x10^{-3}$	3748
$D^2H_{Felled}\,vs.\,D^2H_{Stand}$	$+2.06x10^{-4}$	$\pm 2.26x10^{-4}$	3133
$D^2H_{Stand.}\,vs.\,D^2_{Stand.}$	$-1.40x10^{-2}$	$\pm 4.61x10^{-3}$	3133
$D^2H_{Felled}\,vs.\,D^2_{All}$	$-3.84x10^{-2}$	$\pm 5.30x10^{-3}$	3748

[a] *confidence interval*

Table 5. Percent of total volume for a large frequency sample using the 4 estimators assuming $\hat{V}_T$ is correct.

Estimators	Description	Percent of standard
$\hat{V}_T$	Felled tree, D^2H	100.0
$\hat{V}_{T1}$	All trees, D^2H	94.4
$\hat{V}_{T2}$	Felled tree, D^2	100.0
$\hat{V}_{T3}$	All trees, D^2	99.9

RMRS
ROCKY MOUNTAIN RESEARCH STATION

The Rocky Mountain Research Station develops scientific information and technology to improve management, protection, and use of the forests and rangelands. Research is designed to meet the needs of National Forest managers, Federal and State agencies, public and private organizations, academic institutions, industry, and individuals.

Studies accelerate solutions to problems involving ecosystems, range, forests, water, recreation, fire, resource inventory, land reclamation, community sustainability, forest engineering technology, multiple use economics, wildlife and fish habitat, and forest insects and diseases. Studies are conducted cooperatively, and applications may be found worldwide.

Research Locations

Flagstaff, Arizona
Fort Collins, Colorado*
Boise, Idaho
Moscow, Idaho
Bozeman, Montana
Missoula, Montana
Lincoln, Nebraska
Reno, Nevada
Albuquerque, New Mexico
Rapid City, South Dakota
Logan, Utah
Ogden, Utah
Provo, Utah
Laramie, Wyoming

*Station Headquarters, 240 West Prospect Road, Fort Collins, CO 80526