This file was created by scanning the printed publication.
Errors identified by the software have been corrected;
however, some errors may remain.

Compatible taper equation for loblolly pine

J. P. McClure
USDA Forest Service, Southeastern Forest Experiment Station, Asheville, NC, U.S.A. 28804

AND

R. L. Czaplewski
USDA Forest Service, Rocky Mountain Forest and Range Experiment Station, Fort Collins, CO, U.S.A. 80526

Received February 26, 1986[1]

Accepted July 30, 1986

McClure, J. P., and R. L. Czaplewski. 1986. Compatible taper equation for loblolly pine. Can. J. For. Res. **16**: 1272–1277.

Cao's compatible, segmented polynomial taper equation (Q. V. Cao, H. E. Burkhart, and T. A. Max. For. Sci. 26: 71–80. 1980) is fitted to a large loblolly pine data set from the southeastern United States. Equations are presented that predict diameter at a given height, height to a given top diameter, and volume below a given position on the main stem. All estimates are inside bark. A condition is given that forces the Cao model to be exactly compatible with any total main stem volume equation. An exact volume estimation formula is derived. Twelve benchmarks, which represent realistic utilization criteria, are used to describe expected errors in actually applying the taper equation rather than the more common fit statistics that describe errors encountered when estimating model parameters. Errors in using the fitted model are very similar to errors using Cao's estimates.

McClure, J. P., et R. L. Czaplewski. 1986. Compatible taper equation for loblolly pine. Can. J. For. Res. **16** : 1272–1277.

Une équation de défilement polynomiale segmentée compatible de Cao (Q. V. Cao, H. E. Burkhart, and T. A. Max. For. Sci. 26 : 71–80. 1980) est ajustée à un ensemble important de données de *Pinus taeda L.* provenant du sud-est des États-Unis. Les équations qui sont présentées prédisent le diamètre, à une hauteur donnée, la hauteur pour un diamètre au sommet donné, et un volume au-dessous d'une position donnée de la tige principale. Toutes les estimations sont pour des mesures sous écorce. Une condition est donnée qui force le modèle de Cao à être exactement compatible avec une équation quelconque volume total d'une tige. Une formule pour l'estimation exacte du volume est dérivée. Douze repéres qui représentent des critères réalistes d'utilisation sont utilisés pour décrire les erreurs attendues en applicant l'équation de défilement actuelle au lieu de la méthode classique d'adjustement statistique qui décrit les erreurs obtenus lorsqu'on estime les paramètres du modèle. Les erreurs du modèle adjusté sont similaires aux erreurs en utilisant les estimations de Cao.

[Traduit par la revue]

Introduction

A taper equation is compatible if it predicts the same total main stem volume as a volume equation (Demaerschalk 1971). Compatibility avoids discrepancies between volume estimates from a taper equation and those from a commonly used volume equation (Demaerschalk 1972). Therefore, a compatible taper equation is ideal for calculating proportions of volume in various utilization categories in forest inventory or in optimizing log cutting pattern (Goulding and Murray 1976).

Cao et al. (1980) evaluated the ability of 12 models to predict bole diameters, volume below a specified height, and volume below a specified top diameter for loblolly pine (*Pinus taeda* L.). Included were the taper equations of Kozak et al. (1969), Demaerschalk (1973), Ormerod (1973), Goulding and Murray (1976), and Max and Burkhart (1976) and the volume ratio models of Honer (1967) and Burkhart (1977). No single model was best for all purposes. However, one taper equation estimated both volume and diameter consistently well. This model, like the segmented polynomial model of Max and Burkhart, joins three quadratic functions to represent the top, midsection, and butt portion of the main stem. It is made approximately compatible by using polynomials of the form in Goulding and Murray. Of the 12 models evaluated by Cao, Burkhart, and Max, this equation was the best estimator of volume below a given top diameter; it ranked third in predicting main stem diameters and fourth in predicting volume below a given height. Unlike many other taper models, it represents butt swell and diameters below breast height. It estimates diameters, which is not possible with volume ratio taper models. This compatible, segmented taper equation is the subject of our study and is referred to as the Cao model.

In our study, a mathematical modification was made to the Cao model so that it is exactly compatible with any volume equation. The effect of this modification on model performance was investigated. Also, the data set used to estimate model parameters contains many more trees from a much larger geographic area than that used by Cao et al. (1980). Their published parameters were used to describe the effect of applying the Cao model to a different population of trees of the same species.

Performance of taper models is usually described by statistics such as standard error of estimate ($s_{y \cdot x}$) or the square multiple correlation coefficient (R^2). These statistics are useful in describing model fit during parameter estimation; however, they usually fail to describe the expected errors when the parameterized model is applied to actual utilization problems. The response variables used for most taper models in parameter estimation (e.g., volume / total volume, diameter / diameter at breast height) are not important utilization criteria. Twelve benchmark prediction problems are proposed to better describe prediction errors of a taper model when using realistic criteria for southern pines. These benchmarks can be applied to any taper model no matter what response variable is used to estimate model parameters.

Data

Data are available for 62 425 main stem sections representing 3828 loblolly pine trees (Cost 1978; McClure et al. 1983) from nonplantation plots. Data were collected as part of the USDA Forest Service extensive inventory of Florida, Georgia, North

[1]Revised manuscript received July 21, 1986.

Carolina, South Carolina, and Virginia conducted by the Forest Inventory and Analysis (FIA) Project at the Southeastern Forest Experiment Station in Asheville, North Carolina.

Forty-two percent of the trees are from randomly selected permanent plots, which were measured nondescructively using a McClure mirror caliper and section poles (Cost 1971). End diameters and length measurements are recorded for each tree section. Diameters inside bark (DIB) are estimated using prediction equations based on bark thickness at breast height, diameter at breast height, diameter outside bark (DOB) of section ends, and heights of the section ends (Cost 1978).

The remaining trees are from sampled logging operations from the same geographic area sampled using permanent plots. For each inventory cycle and state, 100 logging sites are selected proportional to the type of commercial timber removals occurring in the inventory unit. This is determined by observations made at local processing facilities for wood products. At each site, data are taken for as many trees as can be measured in 1 day. Inside bark diameters are measured directly for felled trees using a Swedish bark gauge.

Measurements are made at ground and stump levels, at 1.2-m intervals to an 18-cm DOB, at 1.5-m intervals to a 10-cm DOB, and at the tree top. Volume is determined for each section as reported in Cost (1978).

Only trees with an 18- to 50-cm diameter at breast height (DBH) were used to fit the taper equation. Smaller trees have limited potential for multiproduct utiliztion. Larger trees usually have a different stem profile compared with smaller, sawtimber-sized trees. Also, such larger trees are rare and often unavailable for commercial harvest in the study area. Trees with major defects, broken tops, or excessive limbing also were excluded because of their unmerchantable or atypical characteristics.

Half of the remaining trees were randomly selected and used to estimate model parameters. The other half were used as an independent validation sample to evaluate prediction errors. Standardized and orthonomalized cross-products matrices for the predicator variables from both data partitions were checked for balance using validation methods discussed by Draper and Smith (1981).

Model

The original linear multiple regression model used by Cao with a *a priori* estimates of a_1 and a_2 is

$$[1] \quad (d^2KH/V) - 2z = b_1(3z^2 - 2z) + b_2(z - a_1)^2 I_1 + b_3(z - a_1)^2 I_2$$

where

$$I_i = 1 \text{ for } z \geq a_i$$
$$I_i = 0 \text{ for } z \geq a_i$$
$$i = 1, 2$$

The notation is adapted from Cao et al. (1980): d is top diameter inside bark (DIB) at height h; K equals $\pi/[4(100^2)]$, a constant that when multiplied by $(\mathrm{DBH})^2$ in cm^2 equals basal area in m^2; H is total tree height; V is total main stem cubic volume from ground level to tree top, estimated using an existing volume equation; h is height above ground to top diameter d, or height to limit of utilization; z is $(H-h)/H$, i.e., relative tree height from top of tree to height above ground h; b_i are model coefficients estimated from sample data, i = 1, 2, 3; a_1 is the distance from the top of the tree to the upper join point divided by total tree height (H); a_2 is the distance from the top of the tree to the lower join point divided by total tree height (H); and I_i is a step function used to join each polynomial segment.

Equation [1] can be rearranged to produce an equation that predicts diameter at height h:

$$[2] \quad d = \{(V/KH)\,[2(1 - b_1)\,z + 3b_1z^2 + I_1b_2(z - a_1)^2 + I_2b_3(z - a_2)^2]\}^{1/2}$$

The estimated diameter at the top of the tree is zero (i.e., d=0 for h=H). Equation [2] is used to calculate d only when $0 \leq h < H$.

To predict the height h to a given top diameter d, [2] can be solved for h using the quadratic formula

$$[3] \quad h = [-B - (B^2 - 4AC)^{1/2}]\,/\,2A$$

where

$$A = (3b_1 + I_1b_2 + I_2b_3)\,/\,H^2$$
$$B = [(1 + 2b_1) + I_1b_2(1 - a_1) + I_2b_3(1 - a_2)]\,(-2/H)$$
$$C = 2 + b_1 + I_1b_2(1 - a_1)^2 + I_2b_3(1 - a_2)^2 - d^2KH/V$$
$$I_1 = 0 \text{ for } d \leq d_1$$
$$I_1 = 1 \text{ for } d > d_1$$
$$I_2 = 0 \text{ for } d \leq d_2$$
$$I_2 = 0 \text{ for } d > d_2$$
$$d_1 = (V/KH)^{1/2}\,[b_1(3a_1^2 - 2a_1) + 2a_1]^{1/2}$$
$$d_2 = (V/KH)^{1/2}\,[b_1(3a_2^2 - 2a_2) + b_2(a_2 - a_1)^2 + 2a_2]^{1/2}$$

Only one of two possible roots is given in [3]. The other root is usually negative or produces a height estimate greater than the total height. Otherwise it produces a height estimate that, when entered into [2], yields a different diameter than that originally specified in [3]. The diameter estimates at the join points d_1 and d_2 are determined by replacing $h = (1 - a_i)H$ for the join points in [2]. The values of I_1 and I_2 are determined by comparing the specified top diameter d to the diameters at the two join points. If diameter d is smaller than the diameter at a join point, then the predicted height to d must be greater than the height to the join point. This assumes that the parameterized taper equation always predicts smaller diameters as the height increases. This condition is met by the parameter estimates presented in this paper. If d = 0 in [3], then h = H (i.e., the estimated height to a zero diameter equals total tree height). By substituting h = 0 into [2], the diameter at ground level (d_G) is estimated to be

$$d_G = (V/KH)^{1/2}\,[b_1 + b_2(1 - a_1)^2 + b_3(1 - a_2)^2 + 2]^{1/2}$$

Equation [3] is only used for $0 < d < d_G$.

Cubic volume (V_m) between ground level and height h can be estimated by integrating d^2 with respect to height h (Goulding and Murray 1976). Using [2], it can be shown that

$$[4] \quad V_m = (V/3)\,[3 + (3b_1 - 3)z^2 - 3b_1z^3 + b_2(1 - a_1)^3 + b_3(1 - a_2)^3 + I_1b_2(a_1 - z)^3 + I_2b_3(a_2 - z)^3]$$

Volume below a given top diameter is estimated using [4], where z is determined using the height estimate from [3]. Section volumes are estimated by subtracting the estimated volume below the bottom end of the section from the predicted volume below the top end.

Whenever h = H, a compatible taper equation should yield the same main stem volume estimate as the volume equation (i.e., $V = V_m$). Substituting $V = V_m$, z = 0, and $I_1 = I_2 = 0$ into [4] produces the following compatibility constraint:

$$[5] \quad 0 = (1 - a_1)^3b_2 + (1 - a_2)^3b_3$$

This condition was not used by Cao et al. (1980). Their estimates of b_2 and b_3 yield values of −0.01 to −0.03 when

TABLE 1. Model coefficients and standard error of estimate

	Parameter estimates					
	b_1	b_2	b_3	a_1	a_2	$s_{y \cdot x}$
FIA data, exactly compatible	0.1526	−0.9538	1196.4	0.5366	0.9570	0.247
FIA data, approximately compatible	0.1493	−0.9367	1196.8	0.5379	0.9571	0.245
Cao et al. 1980	0.2403	−1.3614	250.3	0.5087	0.9159	0.273

substituted into [5], which makes their model only approximately compatible. Equation [5] ensures compatibility and reduces by one the number of coefficients that must be estimated from the data. The original linear version of the Cao model [1] is forced to be exactly compatible by solving [5] for b_3. The resulting linear form of the model is

$$[6] \quad (d^2\mathrm{KH/V}) - 2z = b_1(3z^2 - 2z) + b_2[(z - a_1)^2 I_1 - (z - a_2)^3 (1 - a_1)^3 I_2 \;/\; (1 - a_2)^3]$$

The only linear regression coefficients in [6] are b_1 and b_2. The value of b_3 is subsequently determined using [5] and regression estimates of a_1, a_2, and b_2. If [5] is applied when estimating the model parameters, then [4] simplifies to:

$$[7] \quad V_{\mathrm{m}} = V[1 + (b_1 - 1)z^2 - b_1 z^3 + I_1 b_2 (a_1 - z)^3/3 + I_2 b_3 (2 - a_2)^3/3]$$

Converting estimates to English (imperial) measures is easily performed. All estimated parameters in the Cao model are dimensionless. If English diameters and heights are entered into the model, then the basal area constant K is the only value that must be changed to produce English estimates.

Methods

A total main stem volume equation (inside bark) was fit to the estimation sample using weighted linear regression to stabilize variance of errors, with a weighting factor of $(D^2H)^{-0.75}$ as recommended by McClure et al. (1983). The taper equation was made compatible with the following volume equation.

$$V = -0.002833 + 0.00003156\, D^2H$$

where V is m^3 volume, D is DBH in cm, and H is total tree height in m.

The Cao model has two positions along the main stem where the polynomial segments abut. If estimates of the join points are available, multiple linear regression can be used to fit the Cao model. Initial estimates were obtained by varying the join points over a reasonable range ($0 < a_1 < a_2 < 1$). Multiple linear regression and 254 different combinations of the two join points were used. The join points that yielded the smallest mean squared error (MSE) were judged to be the best initial estimates. This step also indicated that the initial estimates were near a global minimum. No local minima were detected.

Parameter estimates were refined by using the nonlinear minimization routine ZXMIN from the IMSL (1972) statistical library[2] to vary join point values. Multiple linear regression was used to estimate the other coefficients. The MSE from the multiple linear regression model was minimized by ZXMIN. This procedure iterates between the nonlinear minimization routine and the multiple linear regression, which will converge to the least squares estimate of model parameters (Lawton and Sylvestre 1971).

Two sets of parameters for the Cao model were estimated using our data, one with and the other without the exact compatibility constraint [5]. This permits evaluation of the effect of [5] on model performance.

[2]International Mathematical and Statistical Library, GNB Building, 6th Floor, 7500 Bellaire Blvd., Houston, TX, U.S.A.

Also, the parameter estimates reported by Cao et al. (1980) were used to consider the effect on prediction error of using a different developmental data set.

Methods described by Reynolds (1984) were used to compare performance among these three sets of coefficients for the Cao model using independent test data. A confidence interval was used to test the hypothesis of unbiased error ($\mu = 0$), which assumes that the errors are normally distributed (with an estimated mean and variance). This assumption was tested using the Kolmogorov–Smirnov and Cramer – von Mises statistics (Stephens 1974). All α levels were 0.05.

In one-fourth of the cases, the hypothesis of normally distributed errors had to be rejected, and the hypothesis test described above could not be validly applied. However, quantiles can be used if the errors are not normally distributed. Therefore, the 0.025, 0.500 (median), and 0.975 quantiles were also computed. Blom's ideal data depth (Hoaglin 1983) was used to interpolate between order statistics to estimate an exact probability of 0.95. Quantiles are not as efficient in estimating descriptors of the true error distribution as the mean and confidence interval if the errors are normally distributed; however, the large number of trees in the test sample compensates for this suboptimal efficiency, and quantiles can be applied even if the errors are not normally distributed.

Results

Table 1 contains coefficients estimated from the FIA data. One set was estimated with, and another without, the exact compatibility constraint [5]. Also, the coefficients reported by Cao et al. (1980) are provided for comparison. Relatively good fit was obtained during initial parameter estimation in which a wide range of join point estimates were tested. Refining estimates of join points using the nonlinear minimization routine decreased residual mean square error by less than 5%.

Normalized prediction errors for these three sets of coefficients are displayed in Fig. 1. Model error is described as observed minus predicted diameter inside bark divided by DBH (outside bark) using the independent validation sample. Differences in this error are given as a function of height divided by total height. Cao's parameter estimates result in underestimates of diameter at heights between 20 and 65% of the main stem height compared with the models fit to FIA data (Fig. 1). Virtually no difference in prediction error exists between the two models fit to independent FIA data. The exact compatibility constraint does not noticeably affect model performance in this study. Bias of these two models is under 1% of DBH below 70% of total height; 95% of the predictions for inside bark diameter are within 15% of DBH for most heights. The normality hypotheses for the transformed errors in Fig. 1 are rejected using the procedures recommended by Reynolds (1984).

To better evaluate model performance, 12 prediction problems are used to study model errors in more familiar terms. The independent validation sample is again used. Dimensions of the first three 5-m log sections and 9-, 15-, and 20-cm diameters are used, with errors partitioned by 2.5-cm diameter classes. The model coefficients fit to FIA data under the exact compatibility

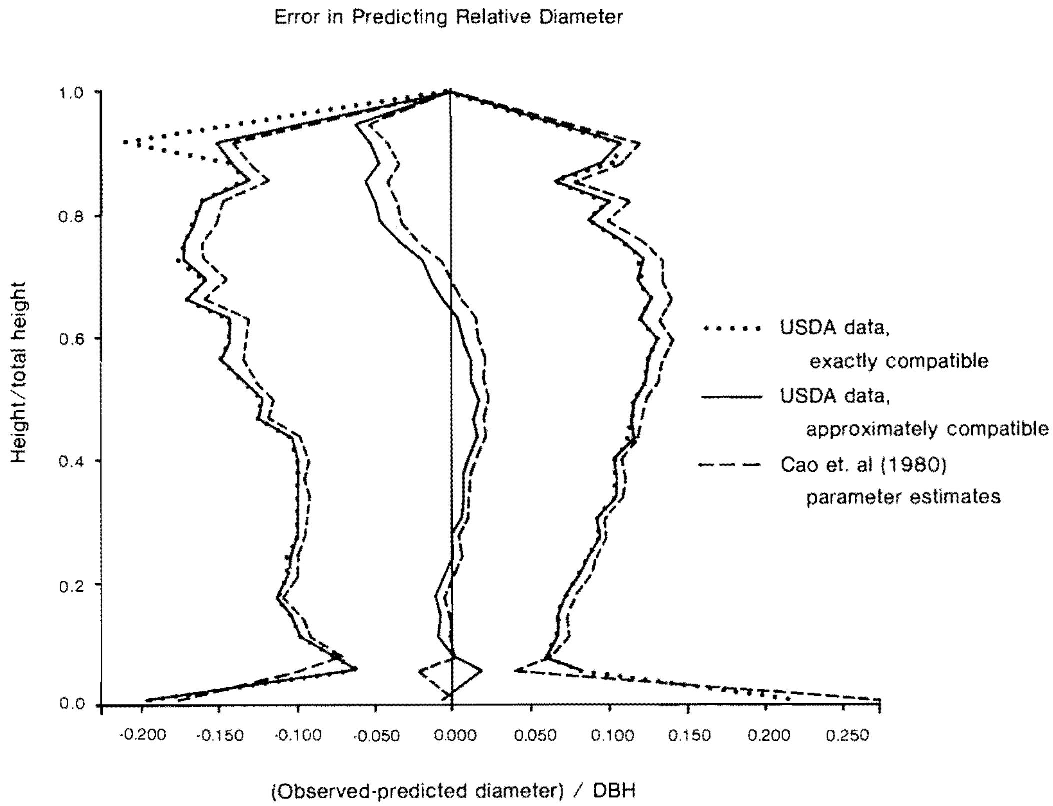

FIG. 1. Comparison among estimated parameter sets of relative error in estimating diameter inside bark. Residual error (observed−predicted diameter) is divided by DBH and height is divided by total tree height so that prediction errors from trees of all sizes could be plotted together. The middle curves are median errors and the curves to the right and left are the 2.5 and 97.5 percentiles. Almost no difference exists between the approximately compatible and completely compatible models estimated using USDA data. The parameters estimates reported by Cao et al. (1980) produce very similar results except that predicted diameters are smaller than the models fit to USDA data by a factor of 1% DBH. Sample sizes are very small near the top of the trees (relative height near 1.0).

condition [5] are used. Errors in inside bark predictions for diameters at 5.3, 10.3, and 15.3 m; heights to 20-, 15-, and 9-cm diameters; 1.2-m section volumes between 0.3 and 1.5, 5.3 and 6.5, and 10.3 and 11.5 m above ground; section volumes between 9- to 15- and 15- to 20-cm DIB; and volume between a 0.3-m stump height and 20-cm DIB are presented in Fig. 2. In addition to the median and the 95% prediction interval (using the 2.5 and 97.5 percentiles for the residuals), the horizontal bars in Fig. 2 contain results of the normality tests (Reynolds 1984). The hypotheses of normally distributed errors are accepted about 75% of the time at the 0.05 α level. Sample sizes range from 45 to 260 and in most cases exceed 125 trees for each 2.5-cm diameter class.

Estimates of inside bark diameter at 5.3 m were generally unbiased (Fig. 2). Errors in predicting diameter at 10.3 and 15.3 m were not normally distributed, and tests for bias were not performed. Median error for both diameter estimates were less than 1 cm, with overestimates of diameter at 15.3 m common for trees 35-cm DBH and smaller. Errors in predicting height to a given top diameter are usually normally distributed but biased. Heights tended to be overestimated for the larger trees, especially heights to a 9-cm DIB.

Most errors in predicting volume are normally distributed (Fig. 2). The least precise estimator was for volume between a 0.3-m high stump and a 20-cm DIB. Note the different scale for this variable in Fig. 2. This imprecision is caused by the major effect on predicted volume caused by small errors in estimating height to a 20-cm top diameter. Despite this high variability, the estimates are unbiased. Efficiency in predicting volume between 15- to 20- and 9- to 15- cm DIB was much better, but significant bias was detected (overestimates of volume). Volume between 0.3 and 1.5 m above ground was significantly underestimated. However, estimates of the other two 1.2-m sections were unbiased.

Conclusions

Performance of the model and coefficients originally reported by Cao et al. (1980) is very similar to the same model fitted to FIA data, although there is somewhat less bias in the lower main stem with the latter parameter estimates. Application of the exact compatibility constraint to the Cao model, which is derived in this paper, does not detrimentally affect prediction error. This will reduce anomalies between volume estimates using the modified Cao taper model and those made using any existing table or equation for total main stem cubic volume. The equation for estimating section volumes, also derived in this paper, will facilitate applications of the Cao model to utilization problems.

Precison of height estimates to given top diameter and volume estimates between stump height and a 20-cm top diameter is low, but other diameter and volume estimates are much better. The majority of the estimators are unbiased, and application of the model to many trees in an extensive inventory would reduce the prediction error for the entire survey area. In a majority of cases, errors encountered in application of common merchantibility standards are normally distributed, permitting

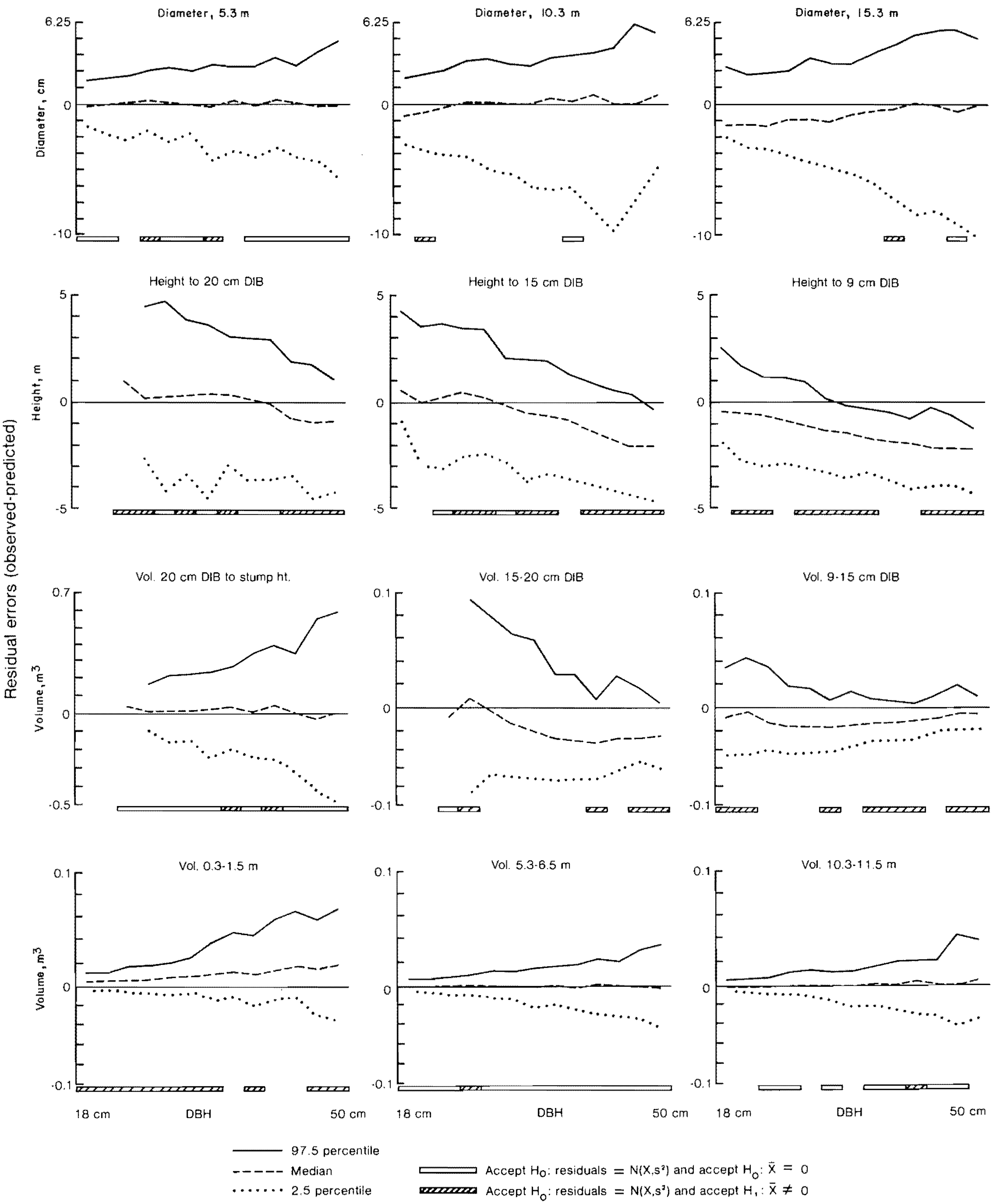

FIG. 2. Prediction error by DBH class (18–50 cm) for 12 benchmark estimation problems showing expected performance of the taper model fit to FIA data under the exact compatibility condition. An independent test sample is used. Results of tests for unbiased predictions are also included except when the hypothesis of normally distributed errors was rejected (both mean and variance estimated).

application of a wide range of efficient statistical evaluation techniques.

BURKHART, H. E. 1977. Cubic-foot volume of loblolly pine to any merchantable top limit. South. J. Appl. **1**: 7–9.

CAO, Q. V., H. E. BURKHART, and T. A. MAX. 1980. Evaluation of two methods for cubic-volume prediction of loblolly pine to any merchantable limit. For. Sci. **26**: 71–80.

COST, N. D. 1971. Accuracy of standing-tree volume estimates based on McClure mirror caliper measurements. USDA For. Serv. Res. Note SE-152.

——— 1978. Multiresource inventories — a technique for measuring volumes in standing trees. USDA For. Serv. Res. Paper SE-196.

DEMAERSCHALK, J. P. 1971. Taper equations can be converted to volume equations and point sampling factors. For. Chron. **47**: 352–354.

——— 1972. Converting volume equations to compatible taper equations. For. Sci. **18**: 241–245.

——— 1973. Integrated systems for the estimation of tree taper and volume. Can. J. For. Res. **3**: 90–94.

DRAPER, N. R., and H. SMITH. 1981. Applied regression analysis. John Wiley & Sons Inc., New York.

GOULDING, C. J., and J. C. MURRAY. 1976. Polynomial taper equations that are compatible with tree volume equations. N.Z. J. For. Sci. **5**: 313–322.

HOAGLIN, D. C. 1983. Letter values, a set of selected order statistics. *In* Understanding robust and exploratory data analysis. *Edited by* D. C. Hoaglin, F. Mosteller, and J. W. Tukey. John Wiley & Sons Inc., pp. 33–57.

HONER, T. G. 1967. Standard volume tables and merchantable conversion factors for the commercial tree species of central and eastern Canada. For. Manage. Res. Serv. Inst. (Ottawa, Ont.), Inf. Rep. FMR-X-5.

KOZAK, A., D. D. MUNRO, and J. H. G. SMITH. 1969. Taper functions and their application in forest inventory. For. Chron. **45**: 278–283.

LAWTON, W. H., and E. A. SYLVESTRE. 1971. Elimination of linear parameters in nonlinear regression. Technometrics, **13**: 461–467.

MAX, T. A., and H. E. BURKHART. 1976. Segmented polynomial regression applied to taper equations. For. Sci. **22**: 283–289.

MCCLURE, J. P., H. T. SCHREUDER, and R. L. WILSON. 1983. A comparison of several volume table equations for loblolly pine and white oak. USDA For. Serv. Res. Pap. SE-240.

ORMEROD, D. W. 1973. A simple bole model. For. Chron. **49**: 136–138.

REYNOLDS, M. 1984. Estimating the error in model predictions. For. Sci. **30**(2): 454–469.

STEPHENS, M. A. 1974. EDF statistics for goodness of fit and some comparisons. J. Am. Stat. Assoc. **69**: 730–737.