STATISTICAL CONSIDERATIONS IN SAMPLING BIOMASS AND SURFACE AREA OVER TIME FOR A *PINUS STROBUS* L. FOREST

H. T. Schreuder, Research Triangle Park, North Carolina, U.S.A.

And

W. T. Swank, Coweeta Hydrologic Laboratory, Franklin, North Carolina, U.S.A.

Southeastern Forest Experiment Station, Forest Service, U. S. Department of Agriculture, Asheville, North Carolina, U.S.A.

STATISTICAL CONSIDERATIONS IN SAMPLING BIOMASS AND SURFACE AREA OVER TIME FOR A PINUS STROBUS L. FOREST[1/]

Hans T. Schreuder and Wayne T. Swank

Biometrician, Research Triangle Park, N. C. and
Plant Ecologist, Coweeta Hydrologic Laboratory, Franklin, N. C.
Southeastern Forest Experiment Station, Forest Service,
U. S. Department of Agriculture, Asheville, N. C.

ABSTRACT

A 16.1 ha white pine plantation was sampled at ages 10, 12, and 15 years 'or biomass and surface area of aboveground tree components. Sampling was ›ased on a survey of 20, .08 ha permanent plots and the felling of 20, 6, and 3 trees at the respective ages. A weighted regression technique is recommended 'or estimating leaf, branch and stem biomass, and surface area with tree basal .rea as the independent variable. At least five sample trees in each of three :qually spaced diameter classes covering the range of diameters are recommended :o derive the weighting factor. The double log and double square root models .re alternate estimation models deserving serious consideration. For any of :hese models, data from different years can sometimes be combined to reduce :he number of trees that need to be felled. Stem area and weight can be ıeasured on additional sample trees every year by non-destructive dendrometry ampling and coring the trees for specific gravity. The additional information ained about these variables can also be used in improving estimation for the ther variables by increasing the probability of selecting the appropriate stimation model.

INTRODUCTION

Improved tree utilization and increased interest in the structure and 'unction of forest ecosystems in recent years has stimulated research on biomass roduction in forest types throughout the world. Concomitantly, a need has eveloped for improved methods of sampling forest stands to satisfy study or ıanagement objectives. Biomass studies for extensive forests are time consuming, ostly, and complex, even for monocultures. Consequently, it is important to rovide a sampling procedure which is efficient in terms of study objectives nd resources available to accomplish the task. In this paper, the sampling rocedures used in estimating forest biomass and surface area parameters will e briefly explained.

The objectives of the paper are to discuss some of the experience gained n sampling a young plantation of white pine regularly over a 6-year period and o recommend how the sampling procedure could be improved. The long term bjective of sampling the plantation is to describe changes in stand biomass nd surface area as related to nutrient cycling and hydrologic processes 2/.

1/ Contribution no. 99 from the Eastern Deciduous Forest Biome US-IBP.
2/ Swank, W. T. and H. T. Schreuder. 1973. Temporal Changes in Biomass, urface Area, and Net Production for a Pinus Strobus L. Forest. In preparation.

SAMPLING DESIGN

The population sampled is a white pine plantation, located on a steep, 16.1 ha watershed at the Coweeta Hydrologic Laboratory in the Southern Appalachians. Details of stand history and physical characteristics are fully described elsewhere (Swank and Miner, 1968). The stand has been sampled over a 6-year period to obtain estimates of biomass and surface area for foliage, branches, and stems.

The basic sampling design used in a given year was a two-stage sampling procedure. The first stage consisted of a complete tally of all tree diameters (d.b.h.) in 20 permanent, .08 ha plots in the fall of 1967, 1969, and 1972. The second stage consisted of the selection by stratified random sampling of 20, 6, and 13 trees in February of 1968, 1970, and 1972, respectively, where strata boundaries were based on tree diameters. These trees were felled and limbed and highly precise estimates of tree foliage, branch and stem biomass and surface area and heights were obtained 2/.

DATA ANALYSIS AND RESULTS

In a previous paper (Swank and Schreuder, 1973), a comparison was made between stratified sampling and ratio and regression estimation procedures using the 1967 and 1968 data collected with the two-stage sampling procedure. The stratified sampling estimation gave the most accurate and the ratio estimation procedure the poorest estimates of the population parameters. But, stratified sampling estimation has a major disadvantage in sampling over time because data from different years cannot be combined easily since strata boundaries change from year to year. Hence, regression estimation was selected as the preferred estimation procedure, although trees to be felled were still selected by a stratified sampling approach to ensure a desirable distribution of tree sizes. The regression estimates of biomass had standard errors of 10 to 16%, those for surface area of 7 to 11% (expressed as percent of the population estimates), depending upon the tree component of interest.

Our regression estimation procedure used simple linear regression between the biomass and surface area variables and basal area. This was done to simplify the calculation of the standard errors of estimates needed in the comparison of estimation methods. Subsequently, we made comparisons between several regression models involving transformations of the biomass and surface area variables and basal area. In addition, tree height times tree diameter (DH) and tree height times tree diameter squared (D^2H) were used as covariates for stem area and weight. The 1968 and 1970 data were combined for the comparisons, resulting in 26 trees for the 1968-70 combination and 13 trees for the 1972 data set. The specific models chosen were

$$Y = a_1 + b_1X + \varepsilon_1 \quad \text{where } \varepsilon_1 \sim N(0,\sigma_1^2) \tag{1}$$

$$Y = a_2 + b_2X + \varepsilon_2 \quad \text{where } \varepsilon_2 \sim N(0,\sigma_2^2X^k) \tag{2}$$

$$\ln Y = a_3 + b_3\ln X + \varepsilon_3 \quad \text{where } \varepsilon_3 \sim N(0,\sigma_3^2) \tag{3}$$

$$\sqrt{Y} = a_4 + b_4\sqrt{X} + \varepsilon_4 \quad \text{where } \varepsilon_4 \sim N(0,\sigma_4^2) \tag{4}$$

where Y represents biomass or surface area for tree components and X is basal area (DH or D^2H in the case of stem area and stem weight). Model (1) was selected because it is frequently used and simple; model (2) seemed intuitively most reasonable; model (3) is frequently used in the literature and our previous results had shown that model (4) was a promising alternative to the others.

The k values for model (2) were calculated from the data as follows: The felled trees were divided into three equally spaced diameter classes covering the range of diameters in the sample. The variances of the biomass and surface area variables within each class were calculated. Then a least squares fit was obtained for the model

$$\ln V_i = \alpha + \beta \ln D_i + \varepsilon_i$$

where

$$E\varepsilon_i = 0,$$

$$V(\varepsilon_i) = \sigma^2,$$

$$E\varepsilon_i\varepsilon_{i'} = 0 \quad i \neq i'.$$

V_i is the variance of a specific biomass or surface area variable within diameter class i and D_i is the midpoint diameter of class i. Because we used D_i rather than basal area in the above regression, k = b/2 is the calculated value in the weighted linear regression (Table 1). The results indicate that 4 to 5 trees per diameter class are enough to give results consistent with the idea that the variance of the variables of interest should increase with an increase in the size of the diameter class. Additional trees would provide better estimates of the k-values and would facilitate a more rigorous examination of the error structure.

The k values for all variables were essentially the same, except for stem surface area for which k was consistently lower. Surprisingly, k decreased for all variables from 1968 to 1972. The decreases may be due to sampling error in the estimates of k, however. A k value of .8 was used for stem area for both data sets. For all other variables, k = 1.5 was used for the combined 1968 and 1970 data and k = 1.2 was used for the 1972 data.

The comparison of models (1) - (4), using the likelihood index (see Schreuder and Swank, 1971 for a fuller discussion), yielded an unexpected result (Table 2). The double square root model, (4), was best for all 1972 data sets and was best or tied for best for stem weight, foliage and stem area for the 1968 plus 1970 data. In only three out of 12 cases was this model second or third and not top rated. The log-log model, (3), was best for foliage and branch biomass for the combined 1968 and 1970 data. The weighted regression model, (2), was best for branch area for the 1968 and 1970 data and tied for best with the double square root model for branch and stem area for the 1968 and 1970 data. The simple linear regression model, (1), ranked lowest in nine out of 12 cases.

Table 1.--Statistics for the relationship between the variance (V) of biomass and surface area variables with d.b.h. classes (D): $V = a D^b$.

Variable	ln a	b	r	k	n
Foliage Weight 1968	5.34	3.22	.83	1.61	20
Foliage Weight 1972	7.84	2.25	.64	1.12	13
Branch Weight 1968	7.00	3.50	.99	1.75	20
Branch Weight 1972	10.30	2.44	.70	1.22	13
Stem Weight 1968	-.51	3.66	.92	1.83	20
Stem Weight 1972	-1.78	2.49	.99	1.24	13
Foliage Area 1968	4.91	2.89	.82	1.44	20
Foliage Area 1972	-.28	2.41	.62	1.20	13
Branch Area 1968	3.39	2.93	.99	1.46	20
Branch Area 1972	-3.43	2.27	.43	1.13	13
Stem Area 1968	-.22	1.64	.92	.82	20
Stem Area 1972	4.75	1.39	.55	.69	13

Table 2.--Likelihoods ($\log_e$) for models of biomass and surface area variables with basal area*.

Variable	(1) simple linear	(2) weighted linear	MODEL (3) log-log	(4) double square-root	BEST MODEL
FW - 1968,1970	-204	-197	-196	-198	(3)
FW - 1972	-105	-102	-96	-90	(4)
BW - 1968,1970	-242	-225	-213	-222	(3)
BW - 1972	-120	-118	-116	-108	(4)
SW - 1968,1970	-48	-38	-35	-33	(4)
SW - 1972	-37	-33	-35	-23	(4)
FA - 1968,1970	-98	-91	-100	-91	(2) and (4)
FA - 1972	-56	-52	-46	-40	(4)
BrA - 1968,1970	-55	-37	-38	-40	(2)
BrA - 1972	-26	-24	-30	-17	(4)
SA - 1968,1970	-82	-78	-82	-78	(2) and (4)
SA - 1972	-46	-44	-49	-36	(4)

Notation: FW = Foliage weight; BW = Branch weight; SW = Stem weight; FA = Foliage area; BrA = Branch area; SA = Stem area.

* D^2H and DH were more strongly related than basal area to stem weight and stem area, respectively, and were used as independent variables for those variables.

The superiority of the double square root model is somewhat mitigated by the fact that its superiority was less consistent for the larger data sets, i.e., for the 1968 and 1970 data sets. This was true even if we omitted the 1970 sample trees (n = 6) and compared results for the models using only the 1968 and 1972 data sets. Thus, we can reject the hypothesis that the lack of consistent superiority was due to the combination of data from the 1968 and 1970 populations. The superiority of the double square root model needs to be established more fully before we use it as our basic estimation model. The fact that the double square root model is not readily interpretable biologically may mean that the surprisingly good performance of this model is due to the fact that it mimics a model not considered here, that can biologically be interpreted. For example, a meaningful hypothesis may be that the true underlying relationship between surface area and biomass variables with basal area is between linearized forms of the variables.

The comparison of models (1) - (4), using the squared correlation criteria (Table 3), indicated that the log-log model, (3), was best or tied for best for foliage and branch weight, for foliage area and for the 1968 plus 1970 branch area data. The double square root model was best or tied for best for stem weight, for the 1968 plus 1970 foliage area, and for the 1968 plus 1970 branch and stem area data. The weighted regression model was best or tied for best for the 1972 stem weight and area, and branch area data. The simple linear model tied for best for the 1968 plus 1970 stem area data. These results are somewhat more erratic than those obtained with the likelihood criteria in choosing a best general model with any degree of consistency. The squared correlation criteria results are somewhat more consistent than those for the likelihood criteria in that the 1968 plus 1970 and 1972 data sets indicate the same best model for the same variables more often under the correlation criteria (four out of six versus three out of six cases).

As indicated in Schreuder and Swank (1971), we feel that the likelihood criterion is the preferred method in selecting the best estimation model (see also Sclove, 1972). Under this criterion, the double square root model is generally the preferred estimation model. From a biological point of view, we still feel uncomfortable with this model since it cannot be readily interpreted. We decided to use the weighted regression model as the basic estimation model for all 12 variables because it performed well under the likelihood criterion (if the double square root model is ignored), is biologically interpretable, and because we wanted to be consistent throughout. (The log-log model also performed quite well under the same conditions.) Since the models were fairly close in performance, as indicated by the closeness of the actual parameter estimates, we felt that the small loss in precision from using possibly the wrong model was justifiable for the sake of consistency.

Table 3.--Squared correlation coefficients for models of biomass and surface area variables with basal area*.

Variable	(1) simple linear	(2) weighted linear	MODEL (3) log-log	(4) double square-root	BEST MODEL
FW - 1968,1970	.960	.940	.965	.960	(3)
FW - 1972	.946	.942	.972	.955	(3)
BW - 1968,1970	.947	.959	.987	.976	(3)
BW - 1972	.969	.966	.975	.969	(3)
SW - 1968,1970	.983	.967	.987	.991	(4)
SW - 1972	.969	.979	.971	.979	(2) and (4)
FA - 1968,1970	.969	.954	.971	.971	(3) and (4)
FA - 1972	.939	.934	.971	.952	(3)
BrA - 1968,1970	.957	.967	.976	.976	(3) and (4)
BrA - 1972	.976	.982	.943	.969	(2)
SA - 1968,1970	.997	.996	.996	.997	(1) and (4)
SA - 1972	.998	.999	.995	.998	(2)

Notation: FW = Foliage weight; BW = Branch weight; SW = Stem weight; FA = Foliage area; BrA = Branch area; SA = Stem area.

* D^2H and DH were used as covariates for stem weight and stem area, respectively, because they were more strongly related to these variables than basal area.

RECOMMENDATIONS FOR FUTURE SAMPLING

In sampling a pine population over time, a sample of 20 sample trees ›led with a survey of 20 permanent, .08 ha plots gave estimates with isfactory precision for a specific point in time. We recommend a ıtified sample selection design in conjunction with a regression estimation :edure. The same regression model may hold for more than one year so that sample size required for a given year can be reduced. Changes in the values ·egression coefficients over time may be of interest.

Basal area was used as the covariate in the estimation of all components. :he case of stem weight and area, D^2H and DH would have been statistically :er covariates than basal area, but their use as covariates would have meant : tree heights had to be measured on a larger sample of trees than was ;ible. Basal area performed almost as well as D^2H and DH for stem weight area, so little precision was lost in estimation.

In selecting the best available model, it is clear that large sample sizes ld be preferred to make a clear choice possible. The weighted regression ɘl requires at least five trees in three equally spaced basal area classes : cover the entire range of basal areas encountered in the field to estimate weighting coefficient k. At least 20 sample trees are required if standard ›rs of estimate of around 10% are desired. Stem area and weight can be ;ured on additional sample trees every year by non-destructive sampling, with use of dendrometers (Grosenbaugh, 1954) or by climbing trees and by coring n to determine specific gravity. The additional information gained on these iables could be used to improve estimation for the other variables, too. ·e is some indication that the best model for stem biomass is also best for iage and branch biomass. Similarly, the best model for stem surface area also be best for foliage and branch surface area.

We were surprised to find that the double square root transformation performed vell under the likelihood criteria. Since we still have difficulty attaching biological meaning to this model, we omitted it from consideration. We ided to use the weighted regression model to estimate all parameters because ›erformed best (ignoring the double square root) in six out of 12 cases and ld be implemented easily in the estimation procedure set-up for our sampling ign. However, a valid argument can be made for using the log-log model in estimation of the biomass parameters.

REFERENCES

Grosenbaugh, L. R. 1954. New Tree-Measurement Concepts: Height Accumulation, Giant Tree, Taper and Shape. Occ. Pap. 134, Southeastern For. Exp. Sta., For. Serv., USDA.

Schreuder, H. T. and W. T. Swank. 1971. A Comparison of Several Statistical Models in Forest Biomass and Surface Area Estimation. Misc. Publ. 132. Life Science and Agriculture Exp. Sta., Univ. of Maine, Orono, pp. 125-136.

Sclove, S. L. 1972. (Y vs. X) or (log Y vs. X)? Technometrics 14(2): 391-403.

Swank, W. T. and N. H. Miner. 1968. Conversion of Hardwood Covered Watersheds to White Pine Reduces Water Yield. Water Resources Res. 4: 947-954.

Swank, W. T. and H. T. Schreuder. 1973. Comparison of Three Methods of Estimating Surface Area and Biomass for a Forest of Young Eastern White Pine. In press. Forest Science.