ON ESTIMATING THE ABOVEGROUND WEIGHTS OF TREE STANDS[1]

H. A. I. Madgwick[2] and T. Satoo[3]

*Department of Forestry and Forest Products, Virginia Polytechnic Institute
and State University, Blacksburg, Virginia 24061 USA*

Abstract. Simulated sampling of nine tree stands was used to compare estimated stand component weights with known weights based on complete harvest. On average, estimates based on regressions of logarithm weight on either logarithm tree diameter or on logarithm (diameter)^2 × height overestimated weights of stems, branches, and leaves by ≈ 3% when the inherent bias of such regressions was accounted for. Ninety-five percent confidence intervals based on random sampling encompassed the measured stand weights 79%–100% of the time. Similar confidence intervals using stratified random sampling based on tree diameter encompassed the measured stand weights only 44%–98% of the time. The small average bias in estimates of stand weights using logarithmic regressions is of minor importance compared with the variation in estimates among replicated samples.

Key words: *Estimation; forest biomass; natural regeneration; plantations; sampling.*

INTRODUCTION

During the past century interest in estimating the dry matter content of tree stands as a basis for determining dry matter production, nutrient cycling, and energy flow in forests has increased. Most current studies use the method of Kittredge (1944), in which dry weight of sample trees is related to tree size by a linear regression after logarithmic transformation without correcting for the inherent bias of such regressions (Baskerville 1972). Following the work of Ovington and Madgwick (1959) numerous authors have noted that this method tends to give larger estimated stand weights than does the method of multiplying the weight of trees of quadratic mean diameter by the number of trees per unit area, and they imply that the values obtained by the regression technique are more accurate.

Comparisons between stand values obtained by sampling and by weighing all trees in the stand are few. Satoo (1968, 1970) found that logarithmic regressions, uncorrected for bias due to transformation, overestimated stand values of foliage by between 1% and 9%. Ovington et al. (1968), using regressions of component weight on stem cross-sectional area, found on the basis of 10 replicated samplings of five trees that the weight of the major stand components (leaves, live branches, boles, and roots) could all be estimated within ± 21%. Except for roots the least accurate estimates were overestimates rather than underestimates of stand weights.

Madgwick (1971), using logarithmic transformations and either diameter or (diameter)^2 × height as an independent variable and 500 replicated random samples, found that regression estimates tended to overestimate needle weight when corrected for inherent bias. Stratified random sampling using diameter as a basis for stratification led to overestimates of live branch weight, while random sampling led to underestimates. Both sampling methods were satisfactory for boles, and stratified random sampling was satisfactory for needle weight.

Meyer (1941), noted that the use of logarithmic transformations leads to a systematic underestimate of the dependent variable, a bias which is related to the residual mean square of the regression. Meyer (1938) also considered the problem of obtaining a standard error of estimate for a variable predicted from a logarithmic regression. His predictors for mean (m) and error, expressed as a variance (v), are:

$$m = e^{\bar{x} + \frac{1}{2}s^2}$$

$$\text{and} \quad v = e^{\bar{x}^2 + s^2}$$

where $\bar{x}$ and s^2 are the mean and the variance of the transformed population. Mountford and Bunce (1973) (see Satchell et al. 1971) have also suggested the use of these forms.

Finney (1941) considered the problem further and noted that, when $\bar{x}$ and s^2 are estimates of the population parameters, approximations of efficient estimators are:

$$m = e^{\bar{x} + \frac{1}{2}s^2} \cdot$$

$$\left(1 - \frac{s^2(s^2 + 2)}{4n} + \frac{s^4(3s^4 + 44s^2 + 84)}{96n^2} + \cdots\right)$$

TABLE 1. Summary of sample plot data. † = at base of stem; * = data collected by the joint study group on forest productivity of 4 universities, Japan; ** = data collected by the joint study group on forest productivity of 5 universities, Japan; *** = data of Ovington et al. (1968) supplied by Dr. J. D. Ovington and Dr. W. G. Forrest

Species	Age (yr)	Plot area (m²)	Stems (n)	Mean diam (cm)	Mean ht (m)	Origin
Abies sachalinensis*	9–30	1.5	45	1.66†	1.12	natural regeneration
Abies sachalinensis*	17–30	2	34	2.27†	1.38	natural regeneration
Betula ermanii*	18	24	25	4.93	7.00	natural regeneration
Cryptomeria japonica*	10	37.2	16	7.97	5.24	plantation
Cryptomeria japonica*	43	32	14	15.18	14.85	plantation
Larix leptolepis**	18	100	14	11.05	9.11	plantation
Pinus densiflora	15	20	13	7.13	6.61	natural regeneration
Pinus radiata***	8	810	100	13.28	7.91	plantation
Pinus virginiana	19	237	136	7.54	8.65	natural regeneration

and

$$v = e^{\bar{x}^2 + s^2}.$$

$$\left[e^{s^2}\left(1 - \frac{2s^2(2s^2+1)}{n} + \frac{2s^4(12s^4 + 44s^2 + 21)}{3n^2} \right) \right.$$
$$\left. - \left(1 - \frac{s^2(s^2+2)}{n} + \frac{s^4(3s^4 + 28s^2 + 42)}{6n^2} \right) \cdots \right]$$

where n is the number of observations. Finney concluded that "without evaluating further terms . . . it seems impossible to say for what range of values of n these last two equations provide satisfactory approximations." However, for $s^2 = 0.69$ he suggested $n > 50$ and $n > 100$ would be safe limits for practical purposes of estimating the mean and variance, respectively. In estimating the biomass of a given stand using logarithmic regression the relevant estimate of s^2 is the mean square deviation from the regression (Snedecor 1956). Typical values of s^2 vary between 0.02 and 0.30, indicating a biased underestimate of stand biomass of the order of 1%–16% if correction is omitted and the underlying model is correct. This information appears to have escaped the notice of forest ecologists for 30 yr (Madgwick 1970) but has been the basis of a number of recent publications (Baskerville 1972, Beauchamp and Olson 1973, Baskerville 1974, Munro 1974).

Recently, Mountford and Bunce (1973) published a method for estimating confidence intervals of stand biomass using the Pearson system of curves (Pearson and Hartley 1966).

For a comparison with workers who have not considered the influence of bias in logarithmic transformation, the above ideas should be compared with those of Whittaker and Woodwell (1968). They purport to use e^s as an estimate of relative error.

They calculated s as $[\Sigma d^2/(n-1)]^{\frac{1}{2}}$ where Σd^2 is the sum of squares of deviations from regressions and n is the number of observations (Whittaker and Woodwell 1968:15). In fact the correct value of s is given by $[\Sigma d^2/(n-2)]^{\frac{1}{2}}$ (Snedecor 1956:125).

Their estimates of relative error can be converted to obtain approximate corrections for their stand data assuming that the logarithmic regression model is valid.

METHODS

Complete stand data for foliage, branches, and boles were available for nine stands encompassing a wide range of tree sizes, species, and stand types (Table 1). For each stand, regressions were calculated relating component weights (stems, branches, and leaves) of individual trees to stem diameter and (diameter)² × height. For two stands of small trees, diameter was that at the base of the stem, while for all other stands, diameter was that at breast height. Stand estimates were calculated using regressions based on different combinations of sample trees.

The alternatives and the reason for choosing them were:

(1) Basing the regression on all trees in the stand and summing the estimates for each tree, to test the overall validity of the logarithmic regression model; and (2) using 10 trees in the stand including 2 trees of each of five diameter classes, smallest, intermediate small, average basal area, intermediate large, and largest. The sample trees were chosen by ordering the sample trees by diameter and choosing the two trees most nearly meeting the required constraints on diameter. We calculated 32 separate regressions using each combination of one tree from each of the five pairs. This method of sampling follows closely that used by many investigators. To compare random *versus* stratified random sampling, two further series of estimates were derived using the data for three stands and 500 replicated samplings per stand. The stands of *Betula ermanii, Pinus radiata,* and *Pinus virginiana* were chosen for this study since they contained the most stems among plots with trees large enough to measure at breast height. (3) Regressions were based on one-third of the trees in the plot chosen at random, or (4) regressions were based on five trees with one randomly sampled tree from each of five diameter classes.

TABLE 2.   The ratio of estimated to actual stand weights and leaf area. Based on logarithmic regressions using either all trees in the stand or stratified random sampling. Data from 9 stands for weights and one stand (*Betula ermanii*) for leaf area. D = diam, D^2H = diam^2 × ht

Sampling method	Independent variable	Replicates/ stand		Stand weights			Leaf area
				Stems	Branches	Leaves	
All trees							
	D	1	Mean	1.04	1.01	1.06	1.04
			Range	1.00–1.13	0.92–1.09	1.00–1.21	
	D^2H	1	Mean	1.01	0.98	1.02	0.99
			Range	0.99–1.04	0.90–1.02	0.97–1.08	
Stratified							
	D	32	Mean	1.04	1.04	1.05	1.07
			Range	0.90–1.59	0.45–1.72	0.81–1.55	1.03–1.13
	D^2H	32	Mean	1.04	1.05	1.06	1.08
			Range	0.90–1.67	0.46–1.82	0.82–1.80	1.03–1.11

It should be noted that the replicated samples are not independent since any particular tree could appear in more than one sample. However, the number of samples taken was small compared with the total number of possible samples in all cases.

Where partial sampling of the stand was used, total stand weight was calculated using the actual weights of the sample trees plus the expected weight, using Finney's (1941) method for the remaining trees. The variance of the estimated weight was obtained using Finney's formula and also the Mountford and Bunce (1973) method applied to the trees for which weights were estimated.

RESULTS

Estimated stem weights of the stands, using logarithmic regressions including all trees in the stands, ranged from 99% to 113% of measured weights (Table 2). For branches and leaf weight the corresponding ranges were 90%–109% and 97%–121%. For stem weight and leaf weight, the means of all estimates based on logarithmic regressions involving D^2H were closer to the true means than those based on D alone. Maximum errors were 4% and 8%, respectively. For branch weight estimates, D and D^2H were approximately of equal accuracy when all stands were considered, though in any particular stand, estimates based on D were at least as large as or larger than those based on D^2H.

Just as for the estimates based on all trees, there was a tendency for the stratified samples to give overestimates, as opposed to underestimates, of stand weights. The range of individual estimates was large, namely from 90% to 167% for bolewood, 45% to 182% for branches, 81% to 180% for leaf weight, and 103% to 113% for leaf area. In 80% of the cases examined the error in the largest estimate was greater than in the smallest estimate. Neither independent variable was clearly superior. Variation

among estimates was greater for naturally regenerated stands (45%–182%) than for plantations (83%–135%).

The 500 replicated random samples from the stands of *Betula ermanii*, *Pinus radiata*, and *Pinus virginiana* yielded mean estimates of stand weights and leaf area very similar to those based on all trees (Table 3). Except for branch estimates in *P. virginiana*, the 95% confidence interval included the measured stand value 87%–100% of the time with almost no difference depending on whether the confidence interval was based on the method of Finney (1941) or that of Mountford and Bunce (1973).

On average the 500 replicated stratified random samples for the same stands yielded results which were almost as accurate as the corresponding strictly random sampling using a much larger number of trees. However, the confidence intervals based on stratified random sampling were, in all but one instance, less inclusive than the corresponding intervals based on strictly random sampling (Table 3). There was no marked difference in the results using D or D^2H as independent variable and only data based on diameter have been presented.

DISCUSSION

The model used by Kittredge (1944) and many subsequent investigators appears reasonably satisfactory for estimating the branch, foliage, and stem weights of tree stands. Regression estimates tend to overestimate stand weights and therefore cannot act as the best values against which to test other methods.

Only Forrest and Ovington (1970) provide variance estimates of stand weights using a method which was not described in their paper. The present studies indicate that the methods of Finney (1941) or of Mountford and Bunce (1973) can provide reasonable confidence intervals for stand estimates. Three

TABLE 3. Summary of replicated sampling of 3 stands. The ratio of estimated to measured stand weights and the fraction of replicates for which measured stand values were within the 95% confidence intervals based on the method of Finney (1941) (F) and Mountford and Bunce (1973) (MB). Based on 500 replicates using either one-third of the trees sampled at random or 5 trees sampled in a stratified random method. Independent variable logarithm diameter breast height. Sample plots are identified by species

Species	Sampling method		Stand weights			Leaf area
			Stems	Branches	Leaves	
Betula ermanii	Random	Ratio	1.06	1.03	1.15	1.04
		F	0.96	0.99	0.88	0.99
		MB	0.96	0.99	0.87	0.99
	Stratified random	Ratio	1.04	1.03	1.09	1.03
		F	0.98	0.96	0.80	0.98
		MB	0.98	0.96	0.79	0.98
Pinus radiata	Random	Ratio	1.02	1.01	1.01	....
		F	0.99	1.00	1.00	....
		MB	0.99	1.00	1.00	....
	Stratified random	Ratio	0.99	1.02	1.00	....
		F	0.76	0.60	0.78	....
		MB	0.76	0.61	0.78	....
Pinus virginiana	Random	Ratio	1.01	0.95	1.04	....
		F	1.00	0.79	0.99	....
		MB	1.00	0.81	0.98	....
	Stratified random	Ratio	1.00	1.11	1.07	....
		F	0.73	0.45	0.59	....
		MB	0.74	0.44	0.59	....

factors appear to combine to cause these confidence intervals to contain the measured stand value less often than expected from theory. First, random sampling and, to a greater extent, stratified random sampling gave residual mean squares around the regression which, on average, were smaller than those found using all trees (Table 4). Second, the bias of estimated means caused a displacement of the confidence interval. (Such displacement excludes many more estimates from one tail of the distribution curve than it includes from the other.) Third, estimated stand weights from stratified sampling were skewed by a few large overestimates (Table 2).

The residual mean squares of regressions relating major tree components to tree size are sufficiently small that, even with only five sample trees per

plot, the estimates of mean and variance using the formulae of Finney (1941) and of Meyer (1938, 1941) give approximately the same results. However, with the widespread use of digital computers the extra effort required in using the more accurate estimates of Finney is negligible.

Estimates of confidence intervals using the method of Mountford and Bunce (1973) gave no improvement over those based on Finney (1941). Mountford and Bunce used the tables of Pearson and Hartley (1966) in estimating confidence intervals. However, in both the Betula and Pinus virginiana stands some replicate samples yielded values of skewness (β_1 and β_2) which necessitated the use of the more extensive tables of Johnson et al. (1963).

Attempts to devise a suitable sampling scheme

TABLE 4. Estimates of the mean square deviation from regressions (Snedecor 1956), relating weight to size. Independent variable diameter

Species	Regression based on	Weight			Leaf area
		Stem	Branches	Leaves	
Betula ermanii	All trees	0.0292	0.0535	0.2644	0.0457
	Random sample	0.0245	0.0447	0.2681	0.0463
	Stratified random sample	0.0217	0.0323	0.2576	0.0448
Pinus radiata	All trees	0.0227	0.0666	0.0520	....
	Random sample	0.0213	0.0636	0.0502	....
	Stratified random sample	0.0194	0.0721	0.0443	....
Pinus virginiana	All trees	0.0689	0.2069	0.2141	....
	Random sample	0.0720	0.1895	0.1923	....
	Stratified random sample	0.0228	0.2058	0.1741	....

for estimating stand weights involves a basic conflict. The use of stratified random sampling results in a considerable reduction of the number of sample trees required to obtain a given level of precision (Madgwick 1971). However, stratification apparently results in erroneous estimates of confidence limits. The type of sampling used by an investigator should, therefore, reflect the purpose of the study.

A stratified random sample based on only five trees per stand may result in estimated stand weights as much as 55% below or 80% above measured stand weights. However, detailed analysis of three stands indicated that, on average, estimates of bole weights were within 20% of measured stand values at least 98% of the time. Estimates for branches and leaves were within 20% of the measured values 53%–99% of the time. In the light of this variation the small average bias in estimates of stand weights using logarithmic regressions is of minor importance.

LITERATURE CITED

Baskerville, G. L. 1972. Use of logarithmic regression in the estimation of plant biomass. Can. J. For. Res. 2:49–53.

———. 1974. Use of logarithmic regression in the estimation of plant biomass: Reply. Can. J. For. Res. 4:149.

Beauchamp, J. J., and J. S. Olson. 1973. Corrections for bias in regression estimates after logarithmic transformation. Ecology 54:1403–1407.

Finney, D. J. 1941. On the distribution of a variate whose logarithm is normally distributed. J. R. Stat. Soc. Ser. B 7:155–161.

Forrest, W. G., and J. D. Ovington. 1970. Organic matter changes in an age series of Pinus radiata plantations. J. Appl. Ecol. 7:177–186.

Johnson, N. L., E. Nixon, D. E. Amos, and E. S. Pearson. 1963. Table of percentage points of Pearson curves, for given $\sqrt{\beta_1}$ and β_2 expressed in standard measure. Biometrika 50:459–498.

Kittredge, J. 1944. Estimation of the amount of foliage on trees and stands. J. For. 42:905–912.

Madgwick, H. A. I. 1970. Biomass and productivity models of forest canopies, p. 47–54. In D. E. Reichle [ed.] Analysis of temperate forest ecosystems. Springer-Verlag, Berlin.

———. 1971. The accuracy and precision of estimates of the dry matter in stems, branches and foliage in an old-field Pinus virginiana stand, p. 105–112. In H. E. Young [ed.] Forest biomass studies. Univ. Maine Press, Orono.

Meyer, H. A. 1938. The standard error of estimate of tree volume from the logarithmic volume equation. J. For. 36:340–342.

———. 1941. A correction for a systematic error occurring in the application of the logarithmic volume equation. Pa. State For. School Res. Pap. 7. 3 p.

Mountford, M. D., and R. G. H. Bunce. 1973. Regression sampling with allometrically related variables, with particular reference to production studies. Forestry 46:203–212.

Munro, D. D. 1974. Use of logarithmic regression in the estimation of plant biomass: Discussion. Can. J. For. Res. 4:149.

Ovington, J. D., W. G. Forrest, and J. E. Armstrong. 1968. Tree biomass estimation, p. 4–31. In H. E. Young [ed.] Symposium on primary productivity and mineral cycling in natural ecosystems. Univ. Maine Press, Orono.

Ovington, J. D., and H. A. I. Madgwick. 1959. Distribution of organic matter and plant nutrients in a plantation of Scots pine. For. Sci. 5:344–355.

Pearson, E. S., and H. O. Hartley, ed. 1966. Biometrika tables for statisticians, 3rd ed. Vol. 1. Cambridge Univ. Press.

Satchell, J. E., D. K. Lindley, and J. K. Hibbard. 1971. Computation of confidence limits for estimates of net primary production, p. 75–82. In T. Rosswall [ed.] Systems analysis in northern coniferous forests. Swed. Natl. Sci. Res. Counc. Bull. 14.

Satoo, T. 1968. Primary production relations in woodlands of Pinus densiflora, p. 52–80. In H. E. Young [ed.] Symposium on primary productivity and mineral cycling in ecosystems. Univ. Maine Press, Orono.

———. 1970. A synthesis of studies by the harvest method: Primary production relations in the temperate deciduous forests of Japan, p. 55–72. In D. E. Reichle [ed.] Analysis of temperate forest ecosystems. Springer-Verlag, Berlin.

Snedecor, G. W. 1956. Statistical methods, 5th ed. Iowa State Coll. Press, Ames. 534 p.

Whittaker, R. H., and G. M. Woodwell. 1968. Dimension and production relations of trees and shrubs in the Brookhaven Forest, New York. J. Ecol. 56:1–25.