RATIO METHODS FOR ESTIMATING FOREST BIOMASS

P. SNOWDON

CSIRO, Division of Forestry, P.O. Box 4008,
Queen Victoria Terrace, ACT 2600, Australia

(Received for publication 10 January 1992; revision 19 January 1993)

ABSTRACT

In a modification of the ratio method for estimating forest biomass, the relationship between the independent variable and the auxiliary variable is linearised prior to calculating the ratio. This is achieved by estimating an exponent for the auxiliary variable by applying a logarithmic regression to data obtained with a sample. This method is mathematically equivalent to the method of estimation based on logarithmic regression where the bias is corrected by a ratio method. A comparison was made of the simple ratio method and the modified method by using simulated samples from seven populations of *Pinus radiata* D. Don for which total biomass was known. There was little difference between the two methods when simple random samples or samples with probability proportional to the sum of sectional area (an unbiased method) were taken. With samples taken with probability proportional to size (PPS) the results with the simple ratio method were highly biased, thus counteracting gains in precision. With the modified method bias remained small while precision and accuracy were substantially improved. No consistent improvement over the modified method was obtained when Horvitz-Thompson estimators or the mean of ratios method were used. It is recommended that the modified method of estimation combined with sampling with probability proportional to size be used for future estimations of forest biomass.

Keywords: sampling; forest biomass; *Pinus radiata*

INTRODUCTION

The most common technique for estimating forest biomass makes use of a logarithmic regression estimator:

$$\sum Y = \sum \exp(a + b\ln(X)) \tag{1}$$

where Y is the biomass of some component in the population, X is an auxiliary variate such as diameter, exp denotes exponentiation, ln natural logarithms, and a, b are estimated by the least squares method in a sample (x, y) from the population.

Since the dependent variable is transformed prior to estimation there is an inherent negative bias in the method. The bias is not an arithmetic constant but a constant proportion of the estimated value (Wiant & Harner 1979). Several correction factors derived from the variance of the regression have been proposed (Flewelling & Pienaar 1981). Simulated sampling studies, however, show that the corrected estimates are often just as biased as the uncorrected estimates (Hepp & Brister 1982; Madgwick 1983; Snowdon 1985). Subsequently, Snowdon (1991) showed that the proportional bias in logarithmic regressions can be estimated from the ratio of the arithmetic mean of the sample and the mean of the back-transformed values predicted from the regression. Under the assumption of a lognormal distribution of errors, the conditions for application of this ratio estimator are optimal. Thus the corrected estimate for the population total is

$$\Sigma Y = \Sigma \exp(a + b\ln(X)) \cdot \Sigma y / \Sigma \exp(a + b\ln(x)). \tag{2}$$

An alternative approach is to use the ratio of means estimator:

$$\Sigma Y = \Sigma X \cdot \Sigma y / \Sigma x \tag{3}$$

is a best linear unbiased estimator if the relationship between the two variables, x and y, is a straight line through the origin and if the variance of y about this line is proportional to x (Cochran 1977). This method using basal area as the independent variable was used to estimate forest biomass by Ando (1962) and later Madgwick (1981). Results from simulated sampling of nine forest plots indicated that this method tends to be less biased and more precise than the logarithmic method, particularly when the latter has been adjusted by a correction factor based on the variance of the regression (Madgwick 1983). Madgwick (1981) noted that the exponents for diameter in power functions relating biomass to diameter are usually greater than 2, indicating that the relationship between biomass and basal area is curvilinear. He subsequently substituted diameter raised to the power of 2.3, 2.5, or 2.7 for basal area in (2) but obtained negligible improvement in estimates of total biomass. As an alternative to using arbitrary exponents for the independent variable Kotimaki & Cunia (1981) recommended that an exponent (b) be estimated from the sample by using Equation (1). This ensures that the dependent variable y has an approximately linear relationship with the auxillary variable x^b.

Kotimaki & Cunia (1981) also examined the mean of ratios estimator:

$$\Sigma Y = \Sigma X \cdot \Sigma(y / x)/n \tag{4}$$

where n is the sample size. This is a best linear unbiased estimator when the relationship between x and y is linear through the origin, and the variance for y is proportional to x^2 (Jessen *et al.* 1947). Based on empirical evidence and subjective arguments, Kotimaki & Cunia (1981) concluded that the ratio of means estimator was probably the better method to use because it was a relatively robust estimator whereas the mean of ratios estimator could have rather erratic behaviour.

When sampling is from a finite population with unequal probabilities without replacement, then alternative ratio estimators can be used. For example, the Horvitz-Thompson ratio estimator is given by

$$\Sigma Y = \Sigma X \cdot \Sigma(y / p)/\Sigma(x / p) \tag{5}$$

where p is the probability that the ith unit from the population is in the sample (Horvitz & Thompson 1952; Cochran 1977). For large populations and sampling with probability proportional to size (i.e., $p_i = X_i / \Sigma X$) it can be shown that this method approximates closely to the mean of ratios method.

This paper uses simulated sampling to compare three methods of estimating forest biomass: (1) the simple ratio of means method; (2) the ratio method in which the exponent for diameter is estimated from a logarithmic regression on the sample; and (3) the logarithmic regression method using the ratio method to correct for bias. Since the ratio estimate has bias of order $1/n$ with simple random samples of size n (Cochran 1977) and regression estimators are affected by the method of sampling (Snowdon 1985), three methods of sampling were also compared: (A) simple random sampling; (B) sampling with probability of a given sample proportional to the sum of tree sizes ($\sum x_i$) which leads to an unbiased ratio estimate (Lahiri 1951); and (C) sampling with probability of selection of each tree proportional to its size. In the last, three further methods of estimating biomass were compared: (4) the Horvitz-Thompson estimator; (5) the Horvitz-Thompson estimator analogous to the proposed modification (2); and (6) the mean of ratios method.

METHODS

A procedure similar to that described by Snowdon (1985, 1991) was followed. Two data sets were used in the simulation study (Table 1). The first set comprised a stand of 100, 8-year-old, *Pinus radiata* trees which had been planted at 2 × 3 m spacing at Mt Stromlo, A.C.T. (Forrest 1969). Data used were sectional area at breast height (basal area) and the weights of bole, live branches, foliage, and roots. The second data set comprised 435, 3-year-old, *P. radiata* trees from an experiment located in Belanglo State Forest, NSW. This experiment, planted at 0.9 × 0.9 m spacing, tested the effects of six combinations of urea and

TABLE 1–Characteristics of *P. radiata* populations used in the simulated sampling study.

Stand No.	No. of stems	Mean sectional area (cm^2)	Biomass component	Mean weight (g)	Exponent*	Error* mean square
1	100	138.5	Stem	21 620	1.22	0.023
			Branches	13 630	1.46	0.067
			Foliage	8 462	1.33	0.052
			Roots	8 557	1.29	0.047
2	76	5.1	Stem	217	1.06	0.032
			Crown	249	1.18	0.136
3	76	5.0	Stem	191	0.98	0.019
			Crown	215	1.23	0.121
4	73	6.0	Stem	254	1.07	0.029
			Crown	252	1.23	0.084
5	84	8.1	Stem	381	1.05	0.066
			Crown	403	1.35	0.077
6	72	8.2	Stem	359	1.02	0.033
			Crown	404	1.38	0.065
7	54	7.1	Stem	297	1.10	0.034
			Crown	332	1.39	0.133

* Exponent (b) and error mean square for the regression
ln (biomass) = $a + b$ ln (sectional area) for the total population.

clover on establishment and growth (Waring & Snowdon 1985). Sectional area at 30 cm, stem dry weight, and the weight of combined branchwood and foliage (crown) were measured on these trees. For the purpose of the present study, data from the replicate plots for each treatment were pooled to provide data for six "stands". The weights of the biomass components were used as dependent variables while sectional area was used as the auxiliary variate and as a measure of tree size for sampling purposes.

Three sampling strategies, each without replacement, were studied:

Random - simple random sampling;

Unbiased - sampling with probability of a given sample proportional to the sum of tree sizes ($\sum x_i$) achieved by choosing the first member with probability proportional to size and the remainder as a simple random sample (Midzuno 1952);

PPS - sampling with probability of selection of each tree proportional to its size.

For each combination of stand and biomass component, 10 000 samples, each consisting of 10 observations, were simulated for each of the sampling methods. For each of these samples, estimates of total stand weight were calculated by three methods:

$RATIO_1$ - the ratio of means method (3) with sectional area as the auxiliary variate

$RATIO_2$ - a modified ratio method:

$$\sum Y = \sum X^b \cdot \sum y \;/\; \sum X^b \qquad (6)$$

where b is the least squares estimator of the slope in the logarithmic regression between y and x in the sample.

LOG - by logarithmic regression followed by estimation of bias from the ratio in the sample of observed values to the predicted values (2) (Snowdon 1991).

In sampling proportional to size, total stand weight was also calculated by three other methods:

HT_1 - the Horvitz-Thompson estimator (5);

HT_2 - the Horvitz-Thompson estimator corresponding to the modified ratio method (6);

$RATIO_3$ - the mean of ratios method (4).

Next, the ratios of the various estimates with the known stand totals and their means, standard deviations and other distributional properties were calculated. Three criteria were used for comparing the methods. "Bias" was estimated by the mean difference between the estimates of total stand biomass obtained by simulation and its known true value. "Precision", or the size of deviations from the biased mean obtained by repeated application of the sampling procedure, was estimated by the standard deviation. "Accuracy", a measure of the size of deviations from the true mean, was estimated by the formula for mean square error (Cochran 1977):

$$(\text{accuracy})^2 = (\text{bias})^2 + (\text{precision})^2. \qquad (7)$$

For convenience all three measures are expressed as percentages. Both precision and accuracy are described as increasing when their associated numerical values decline, and vice versa. The finite population correction was not applied to the estimates of either of these two measures.

Anaylses of variance were used to estimate standard errors for bias, precision, and accuracy. For each, two replicate estimates of the analytical variate were made using independent sets of simulations.

RESULTS

The results from the LOG and $RATIO_2$ methods of computation were identical. In retrospect, it can easily be shown that they are mathematically identical. Thus the estimate of total biomass in the LOG method is:

$$\begin{aligned}\Sigma Y &= \Sigma\exp(a + b\ln(X)).\ \Sigma y \ /\ \Sigma\exp(a + b\ln(x))\\ &= \exp(a).\ \Sigma\exp(b\ln(X)).\ \Sigma y /\{\exp(a)\ \Sigma\exp(b\ln(x))\}\\ &= \Sigma\exp(b\ln(X)).\ \Sigma y \ /\ \Sigma\exp(b\ln(x))\\ &= \Sigma X^{b}.\ \Sigma y \ /\ \Sigma x^{b}\end{aligned}$$

which last result is the $RATIO_2$ method of computation. Hereafter only the $RATIO_1$ and $RATIO_2$ methods will be considered.

With the simple random sampling strategy there was no consistent difference in bias (Table 2) between the two methods of computation. There was no significant bias for $RATIO_1$ when the unbiased sampling strategy was used and bias for $RATIO_2$ tended to be less than that for the random sampling strategy. A further reduction in bias tended to occur with $RATIO_2$ under PPS sampling strategy but bias with $RATIO_1$ was usually markedly increased.

In most calculations precision was numerically much greater than bias with the result that accuracy approximated precision. There was a tendency within sampling regimes for total

TABLE 2–Effects of sampling and calculation methods on bias in estimates of total stand biomass. Results are expressed as percentages of known stand totals.

Component	Stand	Random sampling $RATIO_1$	Random sampling $RATIO_2$	Unbiased sampling $RATIO_1$	Unbiased sampling $RATIO_2$	PPS sampling $RATIO_1$	PPS sampling $RATIO_2$
Stem	1	−0.07	0.97	0.03	0.84	0.11	0.86
Branches	1	−0.56	0.00	−0.08	−0.09	4.45	−0.17
Foliage	1	−0.35	0.31	0.04	0.23	3.21	0.09
Root	1	−0.35	0.52	−0.04	0.43	2.77	0.54
Stem	2	−0.16	0.33	0.02	0.26	0.87	0.08
	3	−0.17	−0.43	−0.01	−0.27	1.25	0.20
	4	−0.71	−0.66	−0.04	−0.43	3.92	0.07
	5	−0.09	−0.17	0.00	−0.19	1.46	−0.18
	6	−0.54	−0.67	0.08	−0.19	2.69	−0.14
	7	−0.30	1.12	0.07	0.73	1.91	−0.59
Crown	2	0.61	0.18	0.05	0.08	4.88	0.08
	3	−0.88	1.55	0.09	1.05	6.60	0.72
	4	−1.42	0.29	0.12	0.59	8.68	0.51
	5	−0.44	0.46	0.02	0.12	3.89	−0.23
	6	−1.66	0.51	0.05	0.26	9.06	0.04
	7	−1.87	3.30	0.12	2.65	10.30	1.21

Standard error for difference of two means: 0.085

stem biomass to be estimated most precisely by $RATIO_1$ and the other components by $RATIO_2$ (Table 3). For all components the most **precise** estimates were obtained with PPS sampling. The most **accurate** estimates were also obtained with PPS sampling, usually in combination with $RATIO_2$ (Table 4). The improvement in accuracy compared to that obtained with simple random sampling and the simple ratio method was in the range 3–43%.

TABLE 3–Effects of sampling and calculation methods on precision of estimates of total stand biomass. Results are expressed as percentages of known stand totals.

Component	Stand	Random sampling		Unbiased sampling		PPS sampling	
		$RATIO_1$	$RATIO_2$	$RATIO_1$	$RATIO_2$	$RATIO_1$	$RATIO_2$
Stem	1	4.25	5.09	4.32	5.09	4.11	4.86
Branches	1	8.23	6.97	8.23	6.94	7.75	6.55
Foliage	1	6.24	5.90	6.22	5.81	5.85	5.42
Root	1	6.52	6.53	6.42	6.45	5.90	6.01
Stem	2	5.64	6.13	5.58	5.90	5.10	5.22
	3	5.34	5.77	5.34	5.60	4.95	4.38
	4	6.11	5.83	6.14	5.65	5.26	4.80
	5	7.91	8.25	7.86	8.07	7.23	7.63
	6	5.96	6.55	5.79	6.08	4.68	4.70
	7	6.05	7.79	5.82	7.38	4.23	5.20
Crown	2	11.51	11.91	11.39	11.41	10.52	9.14
	3	11.96	13.59	12.00	12.90	10.75	8.88
	4	10.55	9.77	10.32	9.51	8.43	7.78
	5	7.54	7.57	7.45	7.37	6.58	6.47
	6	11.11	9.60	10.94	8.92	8.84	6.36
	7	10.88	13.71	10.29	13.53	6.84	8.11

Standard error for difference of two means: 0.0635

TABLE 4–Effects of sampling and calculation methods on accuracy of estimates of total stand biomass. Results are expressed as percentages of known stand totals.

Component	Stand	Random sampling		Unbiased sampling		PPS sampling	
		$RATIO_1$	$RATIO_2$	$RATIO_1$	$RATIO_2$	$RATIO_1$	$RATIO_2$
Stem	1	4.25	5.09	4.32	5.15	4.11	4.86
Branches	1	8.25	6.97	8.23	6.94	8.93	6.55
Foliage	1	6.25	5.91	6.22	5.81	6.67	5.42
Root	1	6.52	6.55	6.42	6.46	6.52	6.03
Stem	2	5.64	6.13	5.58	5.91	5.17	5.22
	3	5.34	5.79	5.35	5.60	5.10	4.38
	4	6.15	5.87	6.14	5.67	6.56	4.80
	5	7.91	8.25	7.86	8.07	7.73	7.63
	6	5.98	6.58	5.79	6.08	5.39	4.70
	7	6.05	7.87	5.82	7.41	4.64	5.23
Crown	2	11.52	11.91	11.39	11.41	11.59	9.14
	3	11.99	13.68	12.00	12.94	12.33	8.90
	4	10.64	9.77	10.32	9.52	12.10	7.80
	5	7.55	7.58	7.45	7.37	7.64	6.47
	6	11.23	9.61	10.94	8.92	12.66	6.36
	7	11.03	14.10	10.29	13.78	12.36	8.20

Standard error for difference of two means: 0.064

The Horvitz-Thompson estimators did not consistently improve accuracy of the estimates of total biomass (Table 5). The $RATIO_2$ and HT_2 methods gave similar results to each other and tended to give the highest accuracy for crown and root components. The $RATIO_3$ and HT_1 methods gave similar results to each other but tended to give the most accurate results for stems.

TABLE 5–Effects of four methods of calculation on accuracy for total stand biomass when samples were chosen with probability proportional to size

Component	Stand	Method of Calculation*			
		$RATIO_2$	HT_2	$RATIO_3$	HT_1
Stem	1	5.00	4.84	4.23	4.23
Branches	1	6.54	6.56	7.75	7.77
Foliage	1	5.40	5.35	5.99	6.01
Roots	1	6.01	5.92	6.23	6.23
Stem	2	5.25	5.20	4.86	4.97
	3	4.45	4.29	4.02	3.97
	4	4.84	4.77	4.80	4.87
	5	7.71	7.68	7.10	7.11
	6	4.70	5.01	4.83	4.97
	7	5.21	5.64	5.18	5.42
Crown	2	9.23	9.04	9.25	9.38
	3	8.87	9.33	8.92	8.98
	4	7.86	7.25	8.26	8.36
	5	6.46	6.53	6.97	6.95
	6	6.41	6.41	8.14	8.26
	7	8.32	7.37	8.24	8.53

* $RATIO_2$ – modified ratio method
$RATIO_3$ – mean of ratio method
HT_1 – Horvitz-Thompson estimator for the simple ratio method
HT_2 – Horvitz-Thompson estimator for the modified ratio method

DISCUSSION

The choice of the best combination of sampling strategy and estimation procedure is a complex problem which depends in part on whether bias, precision, or accuracy is the criterion of choice. In the examples given above the average bias obtained with the simple ratio method combined with PPS sampling was usually high. For the other combinations average bias was small and consequently, being less than one-fifth of the magnitude of precision, had little effect on reducing accuracy. Although bias was negligible using an unbiased sampling the simple ratio method, precision was little affected. As a consequence, the range for bias obtained with individual samples remained high. For example, about 95% of biases from simulated samples of branches in Stand 1 were in the range ± 16% while the minimum and maximum were –29% and 28% respectively. Thus, there is no practical advantage in using unbiased sampling.

The wide range in bias obtained with the simple ratio method can be attributed to non-linearity in the relationship between the dependent and independent variables and to lack of balance or representativeness of individual samples with respect to the distribution of the independent variable and (or) its square (Royall & Cumberland 1981). This is illustrated by the high average biases obtained with PPS sampling which results in unbalanced samples. Some degree of balance can be achieved by using stratified sampling. Madgwick (1981) found in a study of nine stands that this method, on average, improved precision by 18%, 2%, and 5% for stems, branches, and foliage respectively. His results for Stand 1 studied here were 6%, –2%, and –2%.

On average, PPS sampling improved precision by 23% compared to simple random sampling. Except for the stem component in a few stands, the average bias obtained by applying the simple ratio method after PPS sampling was high with the consequence that the modified ratio method ($RATIO_2$) gave the most accurate results. This is in accord with the results of Snowdon (1985) who found that estimates of biomass using the logarithmic regression method were more accurate when samples were chosen with probability proportional to size rather than in a simple random manner.

The mean of ratios method and the closely related simple Horvitz-Thompson ratio method combined with the PPS sampling strategy tended to give the most accurate results for stems. This can be attributed to the near-to-linear relationships between stem weight and sectional area in the test populations (Table 1). These methods were less satisfactory for populations for which the relationship was more curvilinear and more variable. For these populations, methods which made a correction for curvilinearity in the sample tended to give the most accurate results.

A curvilinear relationship between biomass and its predicting variable is expected to occur in biomass studies; for this reason, a method which generally provides more accurate results by taking curvilinearity into account should be preferred to other methods. In practice it can be difficult to assign precise probabilities for the inclusion of specific trees into the sample taken for biomass estimation because the choice of the sample trees is often partially subjective or constrained by other considerations. Consequently, Horvitz-Thompson estimators would not be applicable.

The modified ratio method takes account of curvilinearity in the sample and is not dependent upon the probability that particular members of the population occur in the sample. When it is combined with PPS sampling it is superior to the simple ratio estimate for the estimation of forest biomass. Of the two possible formulations of the modified method (LOG, $RATIO_2$) the approach used here ($RATIO_2$) is preferable because computation of variance for the estimated population totals is more straightforward. Since most estimates will be based on small samples, a robust estimator of variance, such as the jackknife estimate, should be used (Cochran 1977; Royall & Cumberland 1981).

ACKNOWLEDGMENTS

The author wishes to thank Dr W.G. Forrest for permission to use his data, and M.L. Benson (Division of Forestry, CSIRO) and W.J. Müller (Biometrics Unit, Institute of Natural Resources and Environment, CSIRO) for critical comment on the manuscript.

REFERENCES

ANDO, T. 1962: Growth analysis on the natural stands of Japanese red pine (*Pinus densiflora* Sieb. et Zucc.). II. Analysis of stand density and growth. *Government Forest Experiment Station Bulletin 147*: 45–77.

COCHRAN, W.G. 1977: "Sampling Techniques", 3rd ed. John Wiley and Sons, New York.

FLEWELLING, J.W.; PIENAAR, L.V. 1981: Multiplicative regression with lognormal errors. *Forest Science 27*: 281–9.

FORREST, W.G. 1969: Variations in the accumulation, distribution and movement of mineral nutrients in radiata pine plantations. Ph.D. Thesis, Australian National University, Canberra.

HEPP, T.E.; BRISTER, G.H. 1982: Estimating crown biomass in loblolly pine plantations in the Carolina flatwoods. *Forest Science 28*: 115–27.

HORVITZ, D.G.; THOMPSON, D.J. 1952: A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association 47*: 663–85.

JESSEN, R.J.; BLYTHE, R.H.; KEMPTHORNE, O.; EDWARDS DEMING, W. 1947: On a population sample for Greece. *Journal of the American Statistical Association 42*: 357–84.

KOTIMAKI, T.; CUNIA, T. 1981: Effect of cluster sampling in biomass tables construction : ratio estimator models. *Canadian Journal of Forest Research 11*: 475–86.

LAHIRI, D.B. 1951: A method for sample selection providing unbiased ratio estimates. *Bulletin of the International Statistical Institute 33*: 133–40.

MADGWICK, H.A.I. 1981: Estimating the above-ground weight of forest plots using the basal area ratio method. *New Zealand Journal of Forestry Science 11*: 278–86.

—— 1983. Above-ground weight of forest plots - comparison of seven methods of estimation. *New Zealand Journal of Forestry Science 13*: 100–7.

MIDZUNO, H. 1952: On the sampling system with probability proportional to the sum of sizes. *Annals of the Institute of Statistical Mathematics 2*: 99–108.

ROYALL, R.M.; CUMBERLAND, W.G. 1981: An empirical study of the ratio estimator and estimators of its variance. *Journal of the American Statistical Association 76*: 66–77.

SNOWDON, P. 1985: Alternative sampling strategies and regression models for estimating forest biomass. *Australian Forest Research 15*: 353–66.

SNOWDON, P. 1991: A ratio estimator for bias correction in logarithmic regressions. *Canadian Journal of Forest Research 21*: 720–4.

WARING, H.D.; SNOWDON, P. 1985: Clover and urea as sources of nitrogen for the establishment of *Pinus radiata. Australian Forest Research 15*: 1115–24.

WIANT, H.V.; HARNER, E.J. 1979: Percent bias and standard error in logarithmic regression. *Forest Science 20*: 167–8.