Stratified sampling based on photo-interpreted forest types reduces the cost of strategic forest inventory

D.N. Thompson
Forestry Tasmania

Abstract

This study assesses the effectiveness of stratification, based on photo-interpreted forest types, for strategic inventory. It uses results (from stratified sampling) for Total Merchantable Volume (TMV) in 18 Inventory Areas in Tasmanian State forest. The variance of the TMV as determined from stratified sampling is compared with the estimated variance of a simple random sample using the same number of plots. The numbers of additional plots required to obtain, without stratification, estimates with the same variance as the stratified estimates are then computed. Annual savings due to reduced plot numbers are estimated at about $75 000.

Introduction

Forestry Tasmania has historically placed great importance on accurate mapping of the native forest resource for which it is responsible. This mapping relies heavily on classification of the forest into forest types by means of photo-interpretation (PI). Unlike the other Australian States, where forest type is based mainly on species composition, Tasmanian photo-interpreted forest types (PI types) are based primarily on forest structure, particularly age class, stand height and stand density. The resultant type maps are used for a multitude of purposes in the management of the forests. A comprehensive account of the history and current status of photogrammetric mapping in Forestry Tasmania can be found in the article by Stone (1998).

One prominent use is in the stratification of the forest for inventory purposes. This study provides an indication of the gain in efficiency in forest inventory obtained through the use of stratified random sampling, with strata based on photo-interpreted forest types. An earlier study in Tasmania (Lawrence 1957) used analysis of variance to show that photo-interpreted strata generally differ in volume as expected. However, the analysis covered only a few areas, and some types of plots were excluded from the analysis. The present analysis takes a pragmatic approach, comparing the cost of the stratified sample with the estimated cost of an unstratified random sample that would produce volume estimates with the same statistical precision. It does so using well-known mathematical results from sampling theory.

e-mail: don.thompson@forestrytas.com.au

Notation

V_{ran} = Variance of the mean for a simple random sample of size n;
N = number of units in the population;
n = total number of units in the sample;
N_h = number of units in stratum h;
n_h = number of stratum h units in the sample;
L = number of strata in the population;
y_{hj} = observation j in stratum h;

$\bar{y}_{st}$	=	unbiased estimate of population mean from stratified sample;
$v(\bar{y}_{st})$	=	unbiased estimate of variance of $\bar{y}_{st}$
$\bar{y}_h$	=	sample mean for stratum h;
s_h^2	=	sample variance for stratum h;
A	=	total area covered by the sample;
A_h	=	area of stratum h;
a_{hj}	=	area of plot j in stratum h;
a_h	=	total area of plots in stratum h;
$\bar{a}_h$	=	mean area of plots in stratum h;
deff	=	design effect.

Method

Given a stratified random sample from a population, the variance, V_{ran}, of a simple random sample from the same population can be estimated as follows (Cochran 1977, p. 136):

$$V_{ran} = \frac{(N-n)}{n(N-1)N}\left[\sum_h^L \frac{N_h}{n_h}\sum_j^{n_h} y_{hj}^2 - N\left[\bar{y}_{st}^2 + v(\bar{y}_{st})\right]\right] \quad \text{........(1a).}$$

The above equation includes the finite population correction factor, *fpc* = $(N-n)/N$ that compensates for the effect on the estimated variance of the mean of samples that are a significant fraction of the total population. According to Cochran (1977, p. 25), the *fpc* can be ignored in practice whenever the sampling fraction, $f = n/N$, does not exceed 0.05. In the following analysis, no f exceeds 0.01; therefore, the *fpc* is omitted in this analysis and Equation (1a) becomes:

$$V_{ran} = \frac{1}{n(N-1)}\left[\sum_h^L \frac{N_h}{n_h}\sum_j^{n_h} y_{hj}^2 - N\left(\bar{y}_{st}^2 - v(\bar{y}_{st})\right)\right] \quad \text{........(1b).}$$

The mean and variance of the stratified sample are computed by standard formulae as follows (again omitting the *fpc*):

$$\bar{y}_{st} = \frac{1}{N}\sum_{h=1}^{L} N_h \bar{y}_h \quad \text{........(2a),}$$

$$v(\bar{y}_{st}) = \frac{1}{N^2}\sum_{h=1}^{L} N_h^2 \frac{s_h^2}{n_h} \quad \text{........(3a).}$$

As they stand, these formulae are appropriate for populations of discrete individuals. However, sampling in forest management is often conducted on an area basis, with a plot as the basic sampling unit. When this is so, it is necessary, especially when plots vary in size, to substitute areas for the counts (N, N_h, n, n_h) in the above equations where they represent stratum weights or expansion factors. The modified equations actually used in this analysis are:

$$V_{ran} = \frac{1}{n(N-1)}\left[\sum_h^L \frac{A_h}{a_h}\sum_j^{n_h} y_{hj}^2 - N\left(\bar{y}_{st}^2 - v(\bar{y}_{st})\right)\right] \quad \text{........(1c),}$$

$$\bar{y}_{st} = \frac{1}{A}\sum_{h=1}^{L} A_h \bar{y}_h \quad \text{........(2b),}$$

$$v(\bar{y}_{st}) = \frac{1}{A^2}\sum_{h=1}^{L} A_h^2 \frac{s_h^2}{n_h} \quad \text{........(3b),}$$

$$a_h = \sum_{i=1}^{n_h} a_{hj} \quad \text{..........(4),}$$

$$\bar{a}_h = \frac{a_h}{n_h} \quad \text{..........(5),}$$

$$N = \sum_{h=1}^{L} \frac{A_h}{\bar{a}_h} \quad \text{..........(6).}$$

Equation 6 gives an estimate of population size as the total number of non-overlapping plots possible over the whole area based on the average plot size in each stratum. The terms A_h/A in Equations 2b and 3b are stratum weights but are based on stratum areas rather than numbers of units. The term A_h/a_h in Equation 1c is, on the other hand, an expansion factor that converts the sum of squares for the sample to an estimate of the sum of squares for the entire stratum. The sum of the stratum estimates provides an estimate of the sum of squares for the population.

The measure of effectiveness used in this study is the design effect, *deff* (Kish 1965, cited in Cochran 1977). The *deff* measures the effectiveness of a complex sampling design, in this case stratified random sampling, relative to a simple random sample of the same size. It can also be used to estimate the size of a simple random sample required to produce an estimate with the same variance as that obtained from the more complex design. The *deff* is computed as follows:

$$deff = v(\bar{y}_{st}) / V_{ran} \quad \text{..........(7).}$$

Data

The data used in this study includes nearly all of the currently available strategic inventory data for areas of State forest in Tasmania zoned for Multiple Use. For the purposes of strategic inventory, the State is subdivided into 25 Inventory Areas. Five of these areas (Numbers 17 and 22–25) are either on the Bass Strait islands or in the Southwest Conservation Area and contain little or no forest classified for multiple use. They are not sampled and were not included in this analysis. Because there has been almost no strategic inventory in the dry forests of the east coast, Inventory Areas 5 and 6, which cover that region, were also omitted.

Stratification for inventory purposes is based on forest classes, which are amalgamations of photo-interpreted types using general forest type, stand height, oldgrowth height potential, stand density, and year of regeneration. Both the set of forest class definitions and the collection of inventory plots used in the study are the same as those in use by Planning Branch for strategic planning purposes at the time of the study, except that only native eucalypt forest classes were included. Only plots measured in 1970 or later are included in this sample. All data were processed by Forestry Tasmania's Forest Inventory and Projection System, which computes or estimates various characteristics of both trees and plots. It was also used to grow all plots on to a common date, in this case 1998. Growing on the plots is necessary because, for practical reasons, plot measurement is phased over a number of years. Even though some additional variablility is introduced by the growing-on process, the grown-on volumes more accurately reflect the actual volumes at the time of the analysis than would the volumes at measurement. For purposes of this study, growing on the volumes was assumed to have similar proportional effects across the strata.

Each combination of Inventory Area and Forest Class will be referred to as a cell in the remainder of this article. In cases where a cell contains fewer than 10 plots, the standard sample for the cell is normally augmented by 'importing' plots from the same Forest Class in nearby or similar Inventory Areas. Using such imported data would distort this analysis; therefore, each plot was used only in the cell where it is physically located.

Only cells that contained two or more plots and had an area of at least 250 ha[1] were included in the analysis. Under these criteria, a total of 18 Inventory Areas were analysed. The number of forest classes per

[1] It can be shown that, at the overall average sampling rate (one plot per 155 ha), an area of 255.46 ha has 50% probability of containing at least two plots. That area was rounded down to 250 ha and used as the minimum area for a cell to be included in the analysis.

inventory area included in the analysis ranged from 5 to 24. Analyses were based on the mean Total Merchantable Volume (TMV) per hectare for the entire Inventory Area; volumes were not subdivided by species group, age class or form class.

Results

Results for TMV are summarised in Table 1. As stated earlier, the design effect (*deff*, Column 6) is the primary measure of effectiveness used. It indicates the relative reduction in variance achieved by use of stratified sampling rather than simple random sampling. A *deff* of 0.80 indicates that the variance of the mean from stratified sampling is 80% of the variance that would have been obtained with a simple random sample of the same size. The *deff* can also be used to estimate the number of additional plots required to give the same precision with simple random sampling; this is shown in Column 7. The last column shows an estimate of the annual cost of the additional plots, based on the estimate of $65 / plot from Scenario 6, Table 11 in Baalman (1999). Unfortunately, there appears to be no straightforward way to assess the reliability of these estimates, which may be an interesting topic for further research.

Discussion

The substantial overall cost saving of $75 000 represents the additional expenditure on measurement that would have been required to produce estimates of the same precision using simple random sampling. The overall reduction in plots required amounts to 37% of the total. Nine of the 18 Inventory Areas show relative gains of 20% (*deff* = 0.80) or better. Five others show smaller relative gains, but four Inventory Areas show no statistical benefit from stratification, although the operational benefits from mapping remain. More than half of the overall saving occurs in just three Inventory Areas (9, 12 and 21) which, not coincidentally, are among those with the highest sampling intensity (averaging 1 plot per 80, 129 and 100 ha respectively, compared with the 1 plot per 155 ha overall rate).

Although the estimated overall benefit from stratification is encouraging, the fact that there are relatively small or no gains in many inventory areas is somewhat surprising. However, gains from stratified sampling are known to be greatest when the differences between strata are large relative to the variance of the stratum means. The forest class definitions used by Forestry Tasmania are based primarily on the requirements of strategic planning, not on what is best for inventory purposes. It is not considered realistic to have a different classification system for inventory. However, the system is such that it is common for several forest classes in an inventory area to have very similar mean volumes. This may often be true of, for example, Forest Classes 1, 3, 13, 15, 25, and perhaps 34, since all are tall, relatively dense forest on very good sites, differing only in the relative amounts of oldgrowth and regrowth. This can also happen with other groups of forest classes. Stratified sampling can still result in large savings, even when some strata are very similar, if (1) there are substantial differences among groups of strata or there are individual strata with means very different from the groups, (2) sampling rates are sufficient to produce small variances of the stratum means and (3) strata with relatively large areas are given priority in sampling. The reasons for these points can be seen by examining the formula for the estimate of stratified variance, reproduced here for convenience:

$$v(\bar{y}_{st}) = \frac{1}{A^2}\sum_{h=1}^{L} A_h^2 \frac{s_h^2}{n_h} \qquad \text{........(3b).}$$

In this formula, the three variables specific to stratum *h* are of interest. Consider first s_h^2 which is the estimate of the within stratum variance. It represents the natural spatial variation in volume for the forest class; it will stabilise at something near its

Table 1. Summary of results based on Total Merchantable Volume (TMV).

Inventory Area number	Number of classes	Total area (ha) of included forest classes	Number of plots	Sampling intensity (ha / plot)	Design effect	Notional savings in: plots required	annual cost[1] ($)
1	13	12 371	171	72	0.93	12	780
2	5	5 430	27	201	0.66	13	845
3	20	68 183	288	237	0.94	18	1 170
4	13	46 439	94	494	0.84	17	1 105
7	11	18 847	153	123	1.23		
8	19	47 396	157	302	0.64	88	5 720
9	21	23 214	291	80	0.50	291	18 915
10	9	19 740	49	403	0.99		
11	18	35 845	218	164	0.90	18	1 170
12	24	36 246	282	129	0.59	195	12 675
13	20	27 014	221	122	0.70	94	6 110
14	13	32 423	163	199	0.68	76	4 940
15	17	25 525	167	153	1.12		
16	10	7 872	78	101	1.01		
18	8	13 215	137	96	0.88	18	1 170
19	14	17 323	170	102	0.73	62	4 030
20	6	6 098	52	117	0.31	115	7 475
21	20	39 384	394	100	0.74	138	8 970
Total		**482 565**	**3 112**	**155**		**1 155**	**75 075**

[1] At $65/plot/year (Scenario 6, Table 11 in Baalman 1999).

true value at some sample size and cannot be further reduced by additional sampling. Next consider n_h the number of plots in the stratum. Division of s_h^2 by n_h gives an estimate of the variance of the stratum mean. Since it is in the denominator of the equation, increasing this number reduces the contribution of the stratum to the estimate of the overall variance and correspondingly reduces that estimate. That, clearly, is how increasing the sample size reduces the overall variance. Lastly, consider A_h, the area of the stratum. In computing the overall stratified variance, the variance of the mean is multiplied by the square of the stratum area. Consequently, the larger strata make very large relative contributions to the estimate of the overall variance. Therefore, low variances of the stratum means are considerably more important for strata with large areas, even if they have relatively small per hectare volumes. Fortuitously, it is common for the variance of forest volumes to increase as volume increases. Therefore, in most cases, comparable variances of the stratum means can be achieved with less intensive sampling in low volume strata than is required in high volume strata, providing the sample is large enough to reliably estimate the variance.

Conclusion

Mapping of State forest by photo-interpretation serves a variety of strategic and operational purposes, one of which is to stratify the forest for inventory purposes. The foregoing analysis demonstrates that its use for that purpose provides substantial cost savings over inventory conducted without stratification. Looked at in another way, it gives significantly more precise volume estimates for any given investment in field inventory. The analysis also suggests that even greater gains are possible with more carefully targeted inventory along the lines recommended in Baalman (1999) and Stone (1999).

References

Baalman, E. (1999). Costed options for timber inventory. Internal report, Forestry Tasmania, Hobart.

Cochran, W.G. (1977). *Sampling Techniques*. 3rd edn. Wiley, New York.

Kish, L. (1965). *Survey Sampling*. Wiley, New York.

Lawrence, P.R. (1957). Testing the efficiency of photo-interpretation as an aid to forest inventory. Paper presented to the 7th British Commonwealth Forestry Conference, Australia and New Zealand, 1957.

Stone, M.G. (1998). Forest-type mapping by photo-interpretation: A multi-purpose base for Tasmania's forest management. *Tasforests* 10: 15–32.

Stone, M.G. (1999). Wood inventory strategy. Internal report, Forestry Tasmania, Hobart.