Projecting native forest inventory estimates from public to private tenures

C.L. Brack

School of Resources, Environment and Society, The Australian National University, Canberra ACT 0200, Australia.
Email: Cris.Brack@anu.edu.au

Revised manuscript received 15 May 2004

Summary

Inventory information on privately managed forest areas tends to be more variable and less available than for equivalent publicly managed forests. This paper reports on an examination of the timber volume on Tasmanian private and public native forests and demonstrates that the differences between tenures in terms of total (entire stem) volume (m^3 ha^{-1}) are significant but relatively small. The paper also demonstrates that information from public forest inventories may be used to generate auxiliary information that can improve the efficiency of sampling on equivalent private forests. Regression and variable probability sampling using auxiliary information generated from public forest inventories can reduce the need for establishing sample points in private forests to only 25% of that required under simple random sampling for a given level of precision.

Keywords: forest inventories; volume determination; sampling; tenure systems; Tasmania

Introduction

The intensity of sampling across the forests of Australia is very irregular. Public service authorities managing commercial forest plantations tend to implement intensive sampling regimes with up to 1 sample point per 4 ha, while using a lower intensity (e.g. 100 plots / stratum) in their commercial native forest areas. Where inventories exist, sampling intensities on privately managed plantations and native forests tend to be much lower than in the equivalent publicly managed areas. This varying level of sampling and resulting quality of information creates difficulties when national or regional estimates of forest values are required. Often values derived from publicly managed forests are subjectively scaled down and applied to the privately managed forests. Resource Assessment Commission (1991, Appendix F.3.3), for example, scaled down the commercial timber volume available from private forests to as little as 10% of that available from public forests on the basis of subjective estimates by experienced forest managers. Industry and regional planning requires much better information as the privately managed resource becomes more important in areas like south-eastern Queensland. A subjective approach is also no longer sufficient, either to meet Australia's international reporting obligations (e.g. Montreal Indicators and Kyoto Protocol), or to demonstrate sustainability under Regional Forest Agreements.

This paper compares inventory estimates of entire (gross) tree volume across tenure boundaries and outlines the potential to use inventory information from publicly managed forests to predict information on privately managed forests.

Data

Forestry Tasmania maintains an extensive and reliable aerial photograph database of all native forests in Tasmania, both publicly and privately managed. These photographs are interpreted, and forests are classified into polygons (PI Types) delineating patches of relatively uniform forest. Each patch is described according to the broad species group of the forest, average crown height and density, etc. The minimum patch size ranges from 2 ha to 10 ha. Stone (1998) details this typing process for Tasmania. Tasmania is divided into 25 Inventory Areas (IA), and for State Forest sampling purposes the PI Types are grouped into Forest Classes by similar condition, height and crown density of the eucalypt component, and potential growth. Fixed-area plots (about 0.2 ha) are randomly placed within the gross State Forest (i.e. public forestland) area of each Forest Class. The DBH of all commercial trees over 10 cm is measured and a systematic sample of trees is also selected for measurement of height and bark thickness to allow the derivation of height : diameter relationships and bark thickness models. Tree shape, volume and growth are modelled from the data collected.

In 1982, Forestry Tasmania undertook an inventory over privately managed forests using their standard measurement techniques and models. However, plots were not randomly placed within each of the Forest Classes on the private lands. Instead, plots were subjectively selected to cover PI Types not adequately covered on public forestland or where there was an *a priori* expectation that there was a significant difference between public and private forests of a common Forest Class. This inventory data was used to model the Entire Stem Volume (ESV m^3 ha^{-1}) as at 1990.

For the current analysis, Forestry Tasmania provided the estimated ESV for sample plots on public forestland by IA, Forest Class and PI Type as at 1990. Types and IA that did not include samples for each tenure were excluded from analysis (Table 1). Each Forest Class is aggregated from up to 40 PI Types, but these PI Types were not all represented within both tenures. In total, 934 plots (96 private and 838 public) were available for analysis.

Table 1. Number of sample points by inventory area, type and tenure

Inventory Area	Tenure	Forest Class															
		5	6	7	8	9	10	12	19	21	26	31	32	34	35	37	40
4	Private	1			1										1		2
	Public	1			1										15		2
5	Private			3	4	23		2	3								
	Public			5	2	8		1	3								
7	Private				1	2	2	2	5	1		1	3				
	Public				23	28	2	2	29	21		2	12				
8	Private		1	1	5	1			2	1		3	3			2	
	Public		13	32	8	11			11	11		2	5			5	
10	Private								3								
	Public								5								
11	Private				1	2	1										
	Public				7	10	3										
13	Private					1	1		1	1							
	Public					10	2		50	6							
14	Private		1			2	2										
	Public		13			129	39										
15	Private								1								
	Public								169								
19	Private													1			
	Public													30			
21	Private										1			1			
	Public										12			98			

The ESV range is high, varying from 0 to 1500 m^3 ha^{-1} (Fig. 1). Three potential outliers with ESV exceeding 1200 m^3 ha^{-1} were retained for all analyses as there were no compelling reasons for excluding them.

Analysis and results

The unbalanced and sparse nature of the data made it impossible to undertake an ANOVA with Tenure, Forest Class and IA as

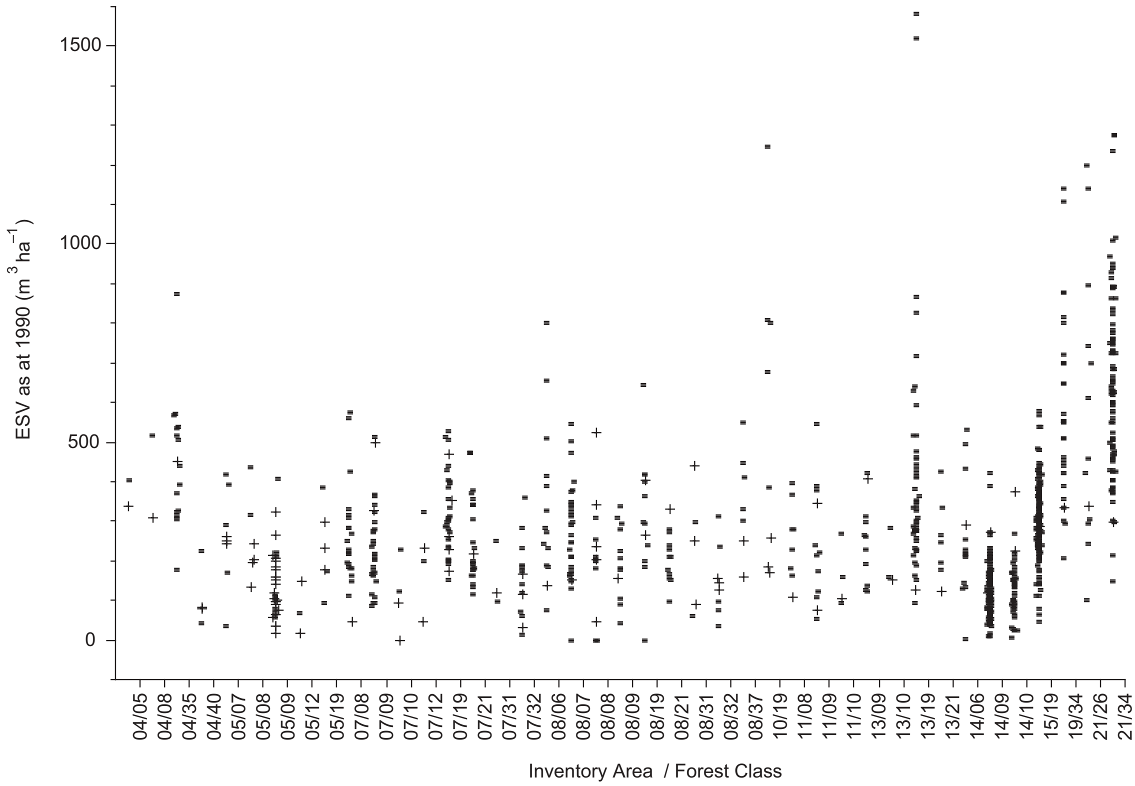

Figure 1. Plot of Entire Stem Volume (m^3 ha^{-1}) against Inventory Area and Forest Class for private (+) and public (■) forest samples

Table 2: ANOVA statistics for √ESV as dependent variable with tenure and Area / Class as independent variables

Source	DF	Sum of squares	*F* ratio	Prob > *F*
Tenure	1	164.59	10.90	0.0010
Area / Class	40	2306.19	3.81	<0.0001
Tenure x Area / Class	40	1107.75	1.83	0.0014
Error	850	12832.2		
Total	931	16410.73		<0.0001

(interacting) independent variables. However, as there appeared to be an interaction between IA and Forest Class a new variable (*Area / Class*) was defined as the combination of IA and Forest Class. A square root transformation of ESV was also necessary for further statistical analysis to meet assumptions about homogeneity of errors.

An analysis of variance (Table 2) found that the square root of ESV was dependent on Area / Class ($F = 3.81$; $P < 0.0001$), and the tenure x Area / Class interaction ($F = 1.83$; $P = 0.0014$).

Although tenure was a significant factor ($P < 0.001$) and on average public forest had a greater ESV than private forest, the significant tenure x Area / Class interaction term indicates that the difference in ESV between privately managed forest and the public forest was not consistent for all Area / Classes. Figure 2 shows that in some Forest Classes, the ESV on private tenure was greater than on public tenure.

Disaggregating Forest Classes into their original PI Types allowed the samples to be classified into condition (cutover, burnt, not recently disturbed) and the stand height and crown density of the mature eucalypt component. The ordinal height and density classes were then allocated average class heights and densities to create continuous numerical values. Analysis of variance (Table 3) indicates that the interaction between tenure and IA remains

Table 3. ANOVA statistics for √ESV as dependent with Tenure, IA and Condition as nominal variables and the stand height and density of the mature eucalypt component as continuous variables

Source	DF	Sum of squares	*F* ratio	Prob > *F*
Tenure	1	87.78	5.87	0.0156
Condition	2	271.51	9.08	0.0001
Height	1	1063.33	71.12	<0.0001
Density	1	570.32	38.14	<0.0001
IA	9	186.16	1.38	0.1909
IA x Tenure	9	285.96	2.12	0.0252
Error	900	13456.13		
Total	923	15921.19		<0.0001

Table 4: ANOVA statistics for reduced model with √ESV as the dependent variable and the stand height and density of the mature eucalypt component as continuous variables

Source	DF	Sum of squares	*F* ratio	Prob > *F*
Height	1	6370.97	384.13	<0.0001
Height^2	1	236.88	14.28	0.0002
Density	1	1848.18	111.43	<0.0001
Error	920	15258.38		
Total	923	23714.41		<0.0001

significant ($F = 2.12$; $P = 0.0252$). However, height and density alone explain almost as much of the error as the model that includes condition, tenure and IA (Table 4).

A linear regression model using height and density was fitted using only the public forest data ($R^2 = 0.52$):

$$\sqrt{ESV} = 5.983 + (0.02327 \times height) + (0.006819 \times height^2) + (0.07682 \times density) + E \quad (1)$$

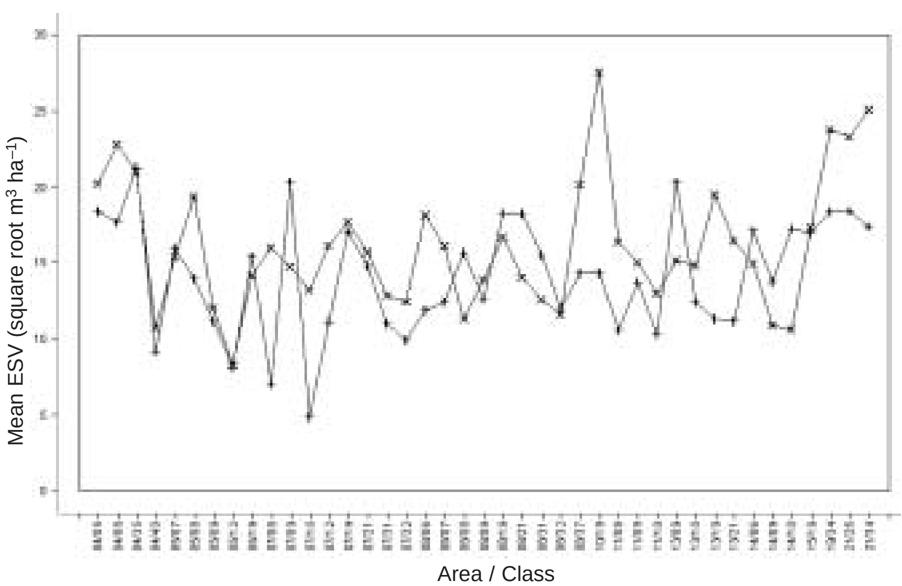

Figure 2. Entire Stem Volume (ESV; square root of m^3 ha^{-1}) for public (x) and private (+) tenures for each Area / Class, with least square mean values connected

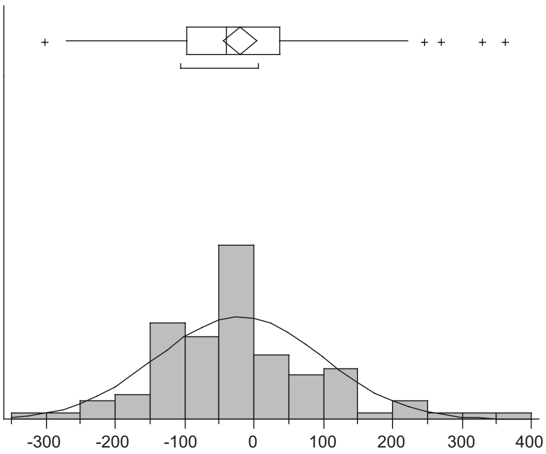

Figure 3. Quantile box plot and frequency distribution for errors in predicting private forest ESV, using model parameters fitted from public forest data. The central box in the quantile plot contains 50% of the observations with 25% extending above and below the box. The median location bisects the box with a horizontal line. The mean and standard errors form a diamond within the box, while the horizontal 'whiskers' contain all data within 1.5 times the interquartile range.

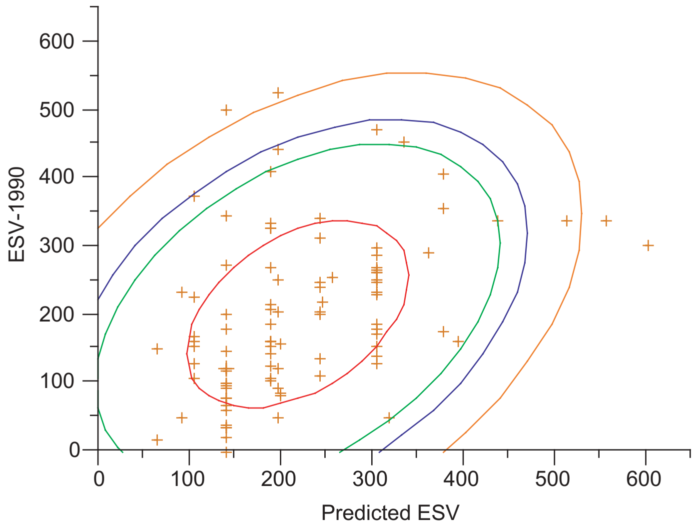

Figure 4. ESV measured in private forest compared to ESV predicted from the public forest model, with normal density ellipses (50%, 90%, 95%, 99%)

where ESV denotes the predicted entire stem volume (m^3 ha^{-1}), 'height' denotes the mean height of the dominant eucalypt component within the PI Type, 'density' denotes the mean stand crown density within the PI Type, and E denotes the error term (which is tested for homogeneity and normal distribution). The standard errors of the regression coefficients are 2.12, 0.0124, 0.00231 and 0.00779, respectively.

This equation was used to predict the $\sqrt{}$ESV for the privately-managed forests, which was transformed to normal units. The difference between the observed and predicted ESV (error distribution) for the private forest predictions was slightly skewed with a mean error of –19.6 m^3 ha^{-1} (std error = 12.2 m^3 ha^{-1}), which is not significantly different from 0 (P = 0.05) (Figs 3, 4).

Discussion

Tenure appears to be a statistically significant effect for ESV of the native forests in Tasmania. However, with a few exceptions, the effect is relatively small and is not consistent across IA or Forest Classes. This small difference is in contrast to the large differences in standing wood volume reported by the Resources Assessment Commission (1991) between private and public managed forests in Tasmania (81 m^3 ha^{-1} and over 179 m^3 ha^{-1} respectively for wet eucalypt forests). Within a given height and density class, it is likely that management history will be different between privately and publicly managed forests — different grazing, fire and timber improvement regimes are to be expected. These differences may have a major impact on merchantability (merchantable volume) by introducing defects, damage or poor form without affecting the total volume. The interaction between tenure and Area / Class (Table 2) may be the result of different combinations of PI Type within Forest Classes for the two tenures.

Over 50% (R^2 = 0.52) of the total variation in ESV was explained by the simple height and density model. Introduction of the condition and location parameters increased the explanatory value to 58%: much greater than the explanatory value of the model developed in Hamilton and Brack (1999) for public native forests in Victoria. The Hamilton and Brack model used aerial photograph classification of tree cover, crown form, height and species composition, with elevation and a location index to predict merchantable volume (net D+ sawlogs m^3 ha^{-1}), but explained less than 20% of the variation (R^2 = 0.185). This low but significant R^2 may be due to their dependent variable being merchantable volume rather than total volume, and to the practice of using point samples (BAF = 3.0) to determine the on-ground volume. The point samples are likely to introduce more variability into the samples than the 0.2 ha plots used for the Tasmanian model.

Hamilton and Brack (1999) demonstrate that even the low explanatory power of their model is enough to improve the estimation of net volume when compared to a traditional (stratified random) sampling approach. The simple model developed from Tasmanian public forest data has much more explanatory power, and predicted the total volume of the private forest samples with a mean error that was not significantly different to 0 (P = 0.05) (Fig. 3). However, forest managers would be justifiably concerned about extrapolating estimates from one type of forest to another — especially if an ANOVA indicated that there were missing terms in the model (i.e. Table 3 indicates Tenure, IA and Condition are statistically significant (P < 0.05) and should be included in a model). This concern may be alleviated by either testing and 'proving' the extrapolation of the public forest model to a new population or using the model estimates as auxiliary variables in a two-phase or variable probability sampling scheme.

Brack and Marshall (1990, 1998) outline a sequential sampling approach that allows model predictions to be efficiently tested for applicability to a new population. The ESV for a relatively small number of random points within the private forest estate could be compared with the predictions from the model derived from public forest data. A series of disagreements (i.e. the predictions are significantly different or not usefully close to the ESV found at the sample point) would quickly indicate that the public forest model is inadequate and should not be used in privately managed forests. Brack and Marshall (1990) concluded that, depending on the error levels acceptable to the user, as few as three sample points could accept or reject a model as useful. If the test proves that the model is inadequate, the proposed approach allows the sample data to be reused to help determine an unbiased estimate of the new population without recourse to the invalidated model.

Two-phase sampling (ratio and regression sampling) is a powerful sampling approach that assumes a linear relationship between an auxiliary variable (e.g. ESV predicted by a model that uses aerial photography) and the dependent variable of interest (e.g. ESV measured on ground plots). Conditions such as whether the relationship goes through the origin and how the variance changes with increasing auxiliary value magnitude will determine what type of two-phase sampling is appropriate statistically. A relatively small number of points within the private forest estate could be chosen to parameterise an equation that relates the predicted ESV with the measured ESV, which can then be used to estimate the mean ESV measured on the ground from the mean of all the predicted ESV. Figure 4 indicates that the relationship between predicted and measured ESV of private land goes near the origin and the variance is approximately homogenous, which makes Regression Sampling most appropriate (Schreuder *et al.* 1993).

Alternatively, a probability proportional to prediction (3P) sampling frame requires only that the auxiliary variable be positively correlated to the dependent variable to provide an efficient estimate of the total volume. That is, 3P does not require the relationship between ESV predicted and measured on the ground to be linear or meet other normal regression assumptions, but simply that larger predicted ESV stands are generally associated with larger on-ground measurements of ESV. Again, Figure 4 confirms this positive relationship.

To simulate the efficiencies of these three potential sampling frames, samples were taken from the private forest data and used to predict the total volume from the 'estate' of 96 private sample plots. Samples were taken randomly (Control), systematically (Regression) and with variable probability (3P). The results (see Fig. 5) are based on 30 simulations of 10 samples for 3P and the Control. The Regression simulations ordered the plots by the auxiliary variable and systematically, beginning with each of the first 10 plots, selected every ninth plot to generate 10 sample regression lines.

The central box in the quantile plot contains 50% of the observations with 25% extending above and below the box. The median location bisects the box with a horizontal line. The mean and standard errors form a diamond within the box, while the horizontal 'whiskers' contain all data within 1.5 times the interquartile range.

The Control was unbiased, but simulated total ESV ranged from 12 600 to 27 700 m^3 (about ± 40% of the observed total), with a Coefficient of Variation (CV) of more than 20%. This relatively poor level of precision is expected due to the small sample size and the high variation of the ESV measured on the individual ground plots (CV ≈ 60%). The observed range of estimated values agrees with the expected sampling error (E%) (P = 0.05) for 10 randomly-chosen samples over a population with a CV of about 60%:

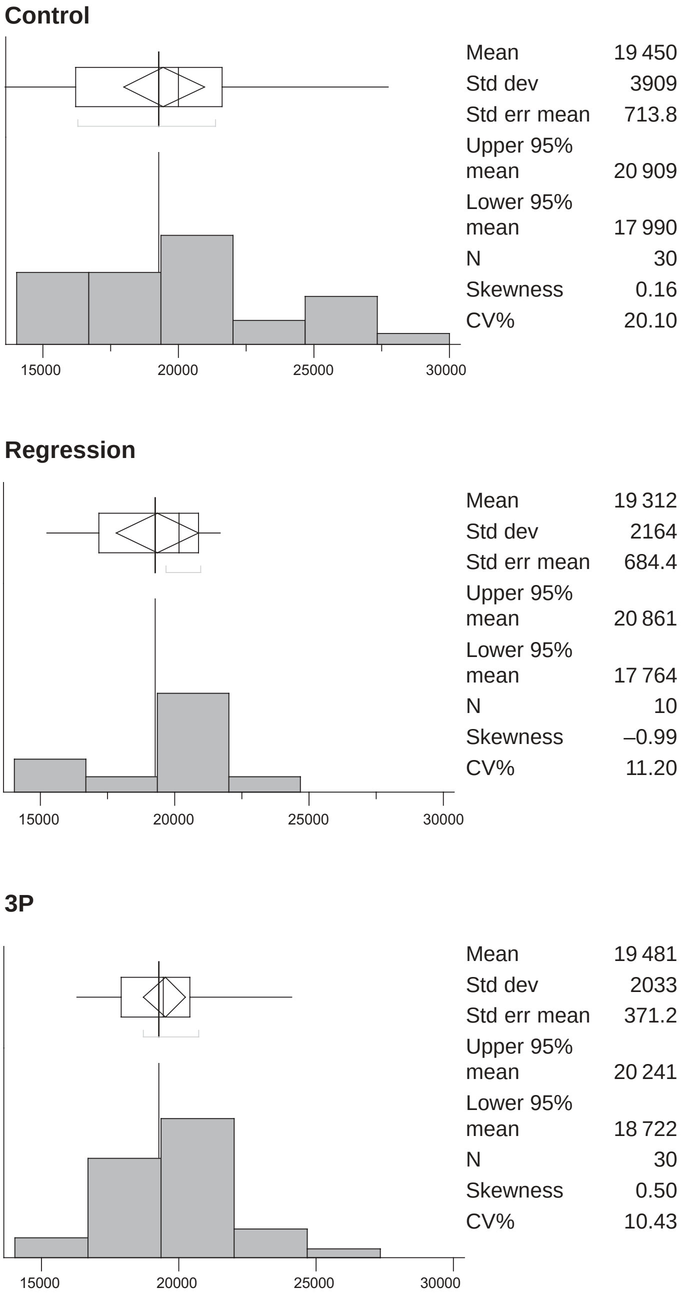

Figure 5. Distribution of estimated total ESV for the 96 private forest plots using three sampling frames. Observed total ESV (19 300 m^3) is marked by the vertical line.

$$E\% = \frac{CV\% \times t_{0.05,\,9}}{\sqrt{n}} = \frac{60\% \times 2.26}{3.16} = 43\% \,. \qquad (2)$$

The regression samples were also unbiased and estimated total ESV ranged from 15 200 to 21 600 m^3 (about ±15% of the observed total) with a CV only slightly more than half the Control. This improved precision is not surprising because the possible lines that can be drawn through systematically selected sample plots can not vary too far from the 'true' line that represents the relationship of predicted against observed ESV for all the plots. If tenure were not a significant factor in predicting ESV from aerial photography then the 'true' line relating ground measurements with predictions would have a slope of 1.0 and an intercept of 0. However, the intercepts for the regression model were significantly greater than 0 and the slopes were significantly less than 1.0 ($P < 0.05$), which indicates that individual plots with a small predicted ESV were underestimated, while plots with a large predicted ESV were overestimated by the model developed from the public forest data (Fig. 4). A deviation from the 1:1 relationship is in agreement with the earlier ANOVA (Table 3) that indicates tenure does have a significant influence on ESV. Note however that despite deviation from the public forest model predictions, the regression sample estimates of the total ESV are precise and unbiased (Fig. 5). An alternative two-phase sample approach could re-parameterise equation 1 for private tenure, i.e. use the public forest data simply to define the best equation form, then sample within the private forest to estimate the coefficients. However, a sample size much greater than 10 systematically selected plots would be needed to reliably estimate the coefficients and attempts to do so with only 10 samples resulted in imprecise, biased and illogical estimates.

The 3P estimates of total ESV ranged from 16 200 to 24 900 m^3 (about ±20% of the observed total) with a CV marginally smaller than the regression. As with regression sampling, the 3P estimates were unbiased even though the predicted ESV used as the auxiliary variable was biased.

The distributions for all three sampling frame simulations were centred on the mean and not significantly skewed. However, the variation in the Control sampling frame was much greater than in the frames that used the model predictions as auxiliary data. Equation (2) indicates that halving of the CV for regression and 3P sampling when compared to that for the Control means that only one quarter as many samples would be required to achieve any given level of sampling precision. However both regression and 3P samples require the auxiliary variable to be available for all points in the forest, which is possible in the Tasmanian situation because of the aerial photography and interpretation carried out over the entire forested estate. Satellite imagery or existing maps potentially provide other sources of auxiliary data that could lead to similar improvements in the CV for modelled estimates. Figure 5 demonstrates that regression or 3P frames can utilise even biased auxiliary information like model predictions different populations to predict the ESV without significant bias.

Conclusion

Private forests have significantly different standing volume to public forests of equivalent eucalypt height and density in Tasmania. With few exceptions, however, the difference is small and there are situations where the differences in standing volume are positive as well as negative. There is no evidence in this study to support the Resource Assessment Commission (1991) average estimate of private forest containing less than 50% of the standing volume in public forests.

Statistical relationships derived from well-sampled public forests can be used to generate auxiliary information about private forests. Providing the auxiliary data cover all the private forest area and there is a positive correlation with a variable of real interest (like biomass or volume), this information can be used in appropriately designed multi-phase or variable probability sampling approaches to substantially reduce sample size, as compared to simple random sampling, to achieve any given level of precision. Biases that may be introduced by extrapolating from public to private forests can be corrected by regression sampling approaches. Variable probability sampling systems will also be able to use a biased model to produce auxiliary variables without resulting in a biased estimate of the population total.

Acknowledgements

Forestry Tasmania and Private Forests Tasmania made their data available for this analysis. The Australian Greenhouse Office sponsored the original analysis of these data as part of their National Carbon Accounting System research.

References

Brack, C.L. and Marshall, P. (1990) Sequential sampling and modelling for mean dominant height estimation. *Australian Forestry* **53,** 41–46.

Brack, C.L. and Marshall, P. (1998) Sequential sampling with systematic selection for estimating mean dominant height. *Australian Forestry* **61,** 253–257.

Hamilton, F. and Brack, C.L. (1999) Stand volume estimates for modelling inventory data. *Australian Forestry* **62,** 360–367.

Resource Assessment Commission (1991) *Forest and Timber Inquiry Draft Report*. Volumes 1 and 2. Commonwealth of Australia, Canberra.

Schreuder, H.T., Gregoire, T.G. and Wood, G.B. (1993) *Sampling Methods for Multiresource Forest Inventory*. John Wiley and Sons, Inc., New York, 446 pp.

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