Sampling for tree-ring analysis

Oscar García
Forest Research Institute
Rotorua, New Zealand *

Abstract

Tree-ring analysis can be used to estimate growth in forest inventory, and to supplement or replace permanent sample plot data in the development of growth models. Costs can be reduced by measuring growth rings in a carefully selected subset of the trees in a plot, making use of the information on current tree size. Efficient sampling strategies for estimating past basal area were investigated by simulation with permanent sample plot data.

Combinations of simple random, stratified, and variable probability sampling with minimum variance unbiased, ratio, and regression estimators were tested, as well as a method based on order statistics from the Weibull distribution. Best results were obtained with a variable probability sampling procedure related to systematic sampling.

1 Introduction

Growth ring measurements are widely used in forest management. Measurements can be made at breast height, on increment cores or cut cross-sections, to determine past basal areas or diameters at breast height. A full stem analysis can provide information on taper, volumes, and heights.

In forest inventories, increment cores are often used to estimate growth. Because of the cost, cores are commonly taken only from a subset of the sample plot trees. This sub-sampling is usually random, or a uniform covering of the full diameter range is attempted (a kind of stratification).

Growth models for managed stands can be developed using pairs of consecutive measurements of stand variables such as basal area per hectare,

*Presented at the IUFRO Conference "Integrating Forest Information Over Space and Time", 13-17 January 1992, Canberra, Australia

stems per hectare, and mean top height. The availability of long series of permanent plot measurements, although desirable, is not essential (García 1988a, 1988b). Under certain circumstances, this kind of data can be reliably derived from tree-ring measurements (P. L. Wilcox, 1987 unpublished). Information obtained by this method has been used to supplement a permanent sample plot data base in a model for radiata pine in Canterbury, New Zealand (M. E. Lawrence 1988 unpublished, P. L. Wilcox 1990 unpublished, García 1988a). The extraction and analysis of several increment cores from each tree in a plot, however, was expensive and time-consuming.

Usually, individual tree diameters or basal areas at two points in time are highly correlated. This fact could be exploited to improve the estimation of past stand basal area from tree-ring samples. Knowledge of the current diameters can be used to select the sample trees, to improve the estimators, or both. A preliminary simulation study was carried out to compare estimation methods, and to assess the likely gains from using current diameter information.

Pairs of permanent sample plot measurements were used. The objective was to estimate the past basal area for the trees alive at the second measurement. To obtain the past plot basal area it would be necessary to add the basal area of any dead trees; in most situations their increment can be safely ignored. Strategies for the joint estimation of basal area and height or other variables were not considered.

The next section describes the data used. Then, several sampling/estimation methods from classical survey sampling theory, and one method based on order statistics are discussed. These correspond to “model independent” and “model dependent” methods, respectively (Hansen et al. 1983). Simulation details, results and conclusions follow.

2 Data

Measurements were obtained from the Forest Research Institute permanent sample plot (PSP) database (Pilaar and Dunlop 1990). Radiata pine plots established according to the specifications described by Tennent (1988) were used. These are variable-area plots, designed to include a similar number of final crop trees. Sixty pairs of measurements separated by 2 years, and sixty separated by 6 years (with no thinnings between them) were taken at random from the whole database. Since some of the plots had as little as 10 trees, the 10 plots with the smallest numbers of trees were then deleted

Table 1: Data ("current" values)

	2-yr intervals			6-yr intervals		
	Mean	Min.	Max.	Mean	Min.	Max.
Age (years)	17.1	7	33	22.2	13	38
Stems per hectare	0	0	0	0	0	0
Basal area (m^2/ha)	0	0	0	0	0	0
Mean dbh (cm)	36.1	8.2	57.7	44.5	24.9	65.4
Dbh increment (cm)	4.0	1.3	7.7	9.8	4.1	20.9
Plot size (m^2)	0	0	0	0	0	0
Trees per plot	35.8	17	111	37.9	21	83

(at random if necessary), leaving 50 plots for each interval length. Some characteristics of the selected data are given in Table 1.

Tree basal areas calculated from the diameters at breast height (dbh) over bark were used. This differs from ring measurements, where diameters under bark would be obtained and bark thickness would be estimated from regression equations. The errors in deriving dbh from radial measurements in increment cores or stem cross-sections are also different from those arising from diameter tape measuring. These differences, however, are unlikely to affect the comparison of sampling strategies.

3 Probability-sampling methods

Knowing the current dbh (and tree basal areas) of all the trees in a plot, we want to estimate their total basal area at some point in the past by ascertaining the past basal areas in a subset of them. Probability-sampling is a commonly-used approach to this type of problem (Cochran 1977, Sukhatme and Sukhatme 1970). Randomization is used to minimize any dependency of the inferences on models or assumptions about the population, although models and prior knowledge are used to achieve efficient sample designs (Hansen et al. 1983).

The sampling design has two parts, a sampling plan that determines the selection probabilities of the potential samples, and an inference procedure which here is a point estimator. Ancillary information, in this case the current dbh values, can be used in the sampling plan, in the estimator, or in both. I tested combinations of simple-random, stratified, and variable-probability sampling, with minimum variance unbiased, ratio, and regression estimators. In general, the methods were chosen as those likely to be the most efficient. The availability of sample-based error estimators was not a consideration, since this is not so important in this application, and error estimates based on very small samples would not be reliable anyway.

The estimators will be given as generalized expressions in terms of the sampling plan inclusion probabilities. This simplifies the presentation by treating all combinations of sampling plan and estimator in a unified way. Variables in the population (plot) are denoted by upper case letters, and variables in the sample by lower case.

3.1 Sampling plans

3.1.1 Simple random sampling (SRS)

A sample of n trees is selected with equal probabilities and without replacement from the N trees in the plot. The past basal areas $y_1, \dots, y_n$ are measured.

The probability of the i-th tree appearing in the sample (inclusion probability) is $\pi_i = n/N$. The joint inclusion probabilities, that is the probabilities of both trees i and j appearing in the sample, can be used to compute the variance of estimators. These are $\pi_{ij} = \frac{n(n-1)}{N(N-1)}$.

3.1.2 Stratification (STR)

The data shows that there is usually a close proportionality relationship between the past tree basal area Y_i and the current basal area X_i. One way of making use of this ancillary information is to divide the population into strata defined by intervals of the variable X, such that the aggregate values of the X variable are approximately the same for each stratum (Hansen et al. 1983). This corresponds to the Ekman rule for approximately optimal stratification, which is known to perform consistently well (Cochran 1977, Sukhatme and Sukhatme 1970). For maximum efficiency the stratification was carried out down to one observation per stratum.

The specific procedure was as follows. The N trees were sorted according to increasing values of X_i, and the values were accumulated. The range of the accumulated X was subdivided into n equal intervals. The strata were then formed by shifting the interval boundaries to the nearest accumulated X_i value. One tree was selected at random, with equal probabilities, within each stratum.

Trees with a value of X larger than the interval size $\sum X_i/n$ were placed in a separate stratum by themselves (i. e. sampled with probability 1), and the procedure just described was then applied to the remaining trees and sample size. This only happened in one plot, for the largest sample size.

The inclusion probabilities are clearly $\pi_i = 1/N_h$, where N_{h_i} is the number of trees in the stratum containing tree i. The joint selection probabilities are $\pi_{ij} = 1/(N_{h_i} N_{h_j})$ if trees i and j are in different strata, and zero otherwise.

3.1.3 Systematic with probability proportional to prediction (SP^3)

Another way of using ancillary information in the sampling plan is with variable probability sampling (Sukhatme and Sukhatme 1970, Chapter II). If the inclusion probabilities were proportional to the Y_i, the mean could be estimated with zero variance. Of course, this is not possible without knowing Y_i, but if it is predicted that the Y_i are approximately proportional to the X_i, sampling with probability proportional to X_i can be expected to be efficient.

There are a number of schemes that could be used for sampling with probability proportional to X. A simple one is to sample with replacement. Another is 3-P sampling, developed by Grosenbaugh for forest inventories (Grosenbaugh 1965). These two procedures have the inconvenient here that the number of trees to be measured would be a random variable. A method suggested by Madow produces samples of fixed size with specified inclusion probabilities, is likely to be more efficient, and is easy to use (Sukhatme and Sukhatme 1970, Cochran 1977, Schreuder et al. 1971).

Madow's method is a kind of systematic sampling from the cumulated X_i. A sampling interval $L = \sum_1^N X_i/n$ was calculated, and a starting point s was selected at random between 0 and L. The N values of X_i were accumulated, and the first trees whose accumulated X values exceeded each of the numbers $s, s{+}L, s{+}2L, \ldots, s{+}(n{-}1)L$ were taken as the sample. This procedure produces a sample of size n with probabilities proportional to X without replacement, provided that the largest of the X_i does not exceed L. As in the STR sampling plan, and as suggested by Schreuder et al. (1971), trees larger than L were placed in a separate stratum and sampled with probability 1.

The individual inclusion probabilities (for $X_i \leq L$) are $\pi_i = nX_i/\sum X_j$. The exact joint inclusion probabilities are complicated, and have been given by Connor (1966). The theoretical analysis is easier for a variant of the method, where the units are arranged at random before the accumulation. However, if the trees are considered in the order in which they are visited in the plot it is possible to approximate this sampling plan without having to list and number the trees (Schreuder et al. 1971). Although tree sizes are related to their locations (García 1991), the effect of the random ordering would probably be small. In the simulations the trees were taken in the order of their PSP numbers, which is typical of the order in which they might be visited in the field.

3.1.4 Ordered systematic with probability proportional to prediction (OSP^3)

This is similar to SP^3, except that the trees are sorted in increasing order of X before sampling. The method has been analyzed by Hartley (1966), and usually leads to more efficient estimates. Listing and numbering of all the trees before selection is required, however.

3.2 Estimators

I discuss estimation of the mean past basal area $\bar{Y} = \sum_1^N Y_i/N$. The plot basal area estimate is obtained multiplying by the known number of trees N.

The estimator equations are given here in terms of the inclusion probabilities π_i. Thus they can be applied in a uniform way to any sampling plan.

3.2.1 Minimum variance unbiased (MVU)

The minimum variance unbiased estimator for the mean with variable probability sampling is the Horwitz-Thompson estimator:

$$\hat{\bar{Y}} = \frac{1}{N} \sum_1^n \frac{y_i}{\pi_i}.$$

With SRS and STR this gives the usual unbiased estimators after substituting the inclusion probabilities given previously.

Expressions for the sampling variance in terms of the π_{ij} are given by Sukhatme and Sukhatme (1970) and Cochran (1977). Sample-based unbiased estimates of the variance are available when none of the π_{ij} is zero (this is not the case for STR, SP^3 or OSP^3).

3.2.2 Ratio (RAT)

Ratio estimators attempt to improve estimation efficiency by making use of an assumed proportionality relationship between the variable of interest (Y) and some ancillary variable (X). They are slightly biased, but can be much more efficient than the UMV estimator. Only large-sample variance approximations are available.

For sampling with variable probabilities, a generalized ratio estimator based on "π-inverse weighting" is

$$\hat{\bar{Y}} = \frac{\sum y_i/\pi_i}{\sum x_i/\pi_i}\bar{X},$$

where $\bar{X} = \sum X_i/N$ is the population mean (Särndal 1980).

With the SRS and STR sampling plans this estimator reduces to the ordinary ratio estimator, and to the combined ratio estimator, respectively. With SP^3 and OSP^3 it coincides with the UMV estimator.

3.2.3 Regression (REG)

If there is an approximately linear relationship between X and Y, but with a non-zero intercept, a regression estimator can be preferable. From Särndal (1980), a π-inverse weighted regression estimator can be written as

$$\hat{\bar{Y}} = \tilde{y} + \hat{\beta}(\hat{X} - \tilde{x}),$$

where

$$\tilde{x} = \frac{\sum x_i/\pi_i}{\sum 1/\pi_i}, \quad \tilde{y} = \frac{\sum y_i/\pi_i}{\sum 1/\pi_i},$$

and

$$\hat{\beta} = \frac{\sum(x_i - \tilde{x})(y_i - \tilde{y})/\pi_i}{\sum(x_i - \tilde{x})^2/\pi_i}.$$

With SRS this gives the usual regression estimator. For stratified sampling this differs from the classical regression estimators given by Cochran (1977) and Sukhatme and Sukhatme (1970), but (Särndal 1980) points out that it is in line with the approach of other authors. The classical estimators cannot be used with one sample tree per stratum, as in STR.

4 A model-dependent approach

Given a probability distribution, good location and scale parameter estimates can often be obtained from just a few properly chosen order statistics (Zacks 1971, Chan and Cheng 1988). For example, for a family of continuous distributions of the form $F[(x-\mu)/\sigma]$, the location (μ) and scale (σ) parameters may be estimated by linear combinations of any subset of order statistics. Coefficients can be obtained to produce estimates that are best linear unbiased (BLUE) or "asymptotically best linear" (ABLE, almost BLUE for large samples). For any subset size, the ranks of the selected order statistics can be chosen to maximize the efficiency of the estimators (Chan and Cheng 1988).

Assume that the past tree basal area distribution can be modelled by some continuous probability function, so that the past N basal areas are a random sample from it. Assume also that growth preserves the dbh rankings, so that the ordering of the trees according to the current basal areas coincides with the one for the past basal areas. Then, we can chose n trees corresponding to any specified set of order statistics (sample quantiles) from the sample of N, and use their past basal areas to estimate the distribution parameters. The mean or total past basal area can be obtained from the estimated distribution.

A commonly used model for tree basal area (or, equivalently, dbh) distributions is the two-parameter Weibull, with distribution function

$$F(y) = 1 - \exp[-(y/b)^c].$$

The logarithm $z = \ln y$ has the smallest extreme value distribution

$$G(z) = 1 - \exp[-exp(\frac{z-\mu}{\sigma})],$$

with

$$\mu = \ln b, \quad \sigma = 1/c,$$

which is of the location-scale type. Hassanein (1972) gives the ABLE coefficients and optimal ranks for this distribution and n between 2 and 10. These values were used to estimate the past basal area with samples of 2, 4, and 8 trees.

The procedure is as follows. The trees in a plot are sorted in decreasing order of dbh, receiving ranks $1, \ldots, N$. Select the trees with ranks $[\lambda_1 N + 1], \ldots, [\lambda_n N + 1]$, where $[x]$ denotes the integer part of x (i. e. the largest integer not exceeding x), and the λ_i are the optimal "spacings", from Hassainen's Table 2. Estimate the Weibull parameters by

$$
\begin{aligned}
b &= \exp(\sum_{i=1}^{n} b_i \ln y_i), \\
c &= 1/(\sum_{i=1}^{n} c_i \ln y_i),
\end{aligned}
$$

where y_i are the past tree basal areas for the selected trees, and the coefficients are from Table 2 in Hassanein (1972). The past mean basal area estimate is given by

$$\hat{\hat{Y}} = b\Gamma(1 + 1/c).$$

The assumptions on which this method is based can be regarded only as rough approximations (Sutton 1973, García 1991). In addition, "best" estimates of distribution parameters do not result necessarily in "best" estimates for the mean. Nevertheless, the assumptions are probably not too far from reality, and the high efficiency of the estimators when the model is "true" originally suggested the possibility of obtaining acceptable estimates with a small number of tree-ring measurements. Incidentally, Weibull distributions and no changes in ranking imply a linear relationship between the logarithms of the tree diameters at two points in time (Bailey 1980, but note that equation (11) and following results are incorrect). Data plotting shows reasonably good log-log trends.

5 Simulation

The 10 distinct combinations of probability sampling plan and estimator were applied to the 100 plots (fifty 2-year intervals and fifty 6-year intervals), with sample sizes of 2, 4, and 8 trees. Although in a few cases the exact variances could have been calculated, repeated sampling was used to compare the performance of the various methods. In each instance, 100 samples were simulated, and mean errors (bias) and mean squared errors were computed.

All the programming and computations were done in APL, using STSC's APL*PLUS interpreter. The APL built-in random number generator, known to be of good quality, was used. Computing time for the 300 000 samples was approximately 3.7 hours on a 16 MHz 386SX personal computer.

The quantile-based estimator was also applied to the 100 plots, using 2, 4, and 8 quantiles. The method does not involve randomization, so that there were no replications.

6 Results

6.1 Probability sampling

There was no consistent pattern in the mean errors for the unbiased (UMV) compared to the ratio and regression estimators (RAT and REG). Therefore, any biases in RAT and REG are likely to be small. By using mean square errors, bias effects are included in the performance comparisons below.

There are clear relationships between the mean square errors and the means, with the form of these trends varying across methods. The number of trees in a plot had no appreciable effect. For presentation purposes, the plots have been grouped according to their mean basal areas into three classes of 17, 16, and 17 trees.

It was found that a logarithmic transformation gave a more symmetrical distribution and stabilized the variances of the mean square errors. For this reason, the statistical analysis was based on the logarithms of the relative root-mean-square (RMS) errors.

Tables 2 and 3 show the percentage relative RMS for the 2-year interval and 6-year interval plots, respectively. These are geometric means within each class (corresponding to the means of the logarithms). An analysis of variance of the logarithmic relative RMS was used to calculate least significant differences between methods (LSDs) at the 95% confidence level. The individual-class LSDs were all fairly similar, with a clear trend only across sample sizes, so that pooled LSDs for each sample size were used. In the tables, vertical bars join methods differing by less than the LSD.

Table 2: Relative RMS error of past basal area estimates (%). 2-year intervals

	Mean diameter, cm					
	3 – 28		29 – 37		38 – 56	
$n = 2$	OSP^3/UMV	7.35	OSP^3/UMV	3.30	STR/RAT	1.95
	OSP^3/REG	7.37	OSP^3/REG	3.30	STR/REG	1.95
	STR/RAT	7.58	STR/RAT	3.45	OSP^3/UMV	2.01
	STR/REG	7.61	STR/REG	3.46	OSP^3/REG	2.01
	SP^3/UMV	7.89	SP^3/UMV	3.50	SP^3/UMV	2.21
	SP^3/REG	7.90	SP^3/REG	3.51	SP^3/REG	2.22
	SRS/RAT	8.49	SRS/REG	3.66	SRS/RAT	2.30
	SRS/REG	8.49	SRS/RAT	3.66	SRS/REG	2.30
	STR/UMV	16.80	STR/UMV	12.44	STR/UMV	12.57
	SRS/UMV	24.03	SRS/UMV	18.96	SRS/UMV	19.90
$n = 4$	OSP^3/UMV	4.50	OSP^3/UMV	2.16	OSP^3/UMV	1.25
	OSP^3/REG	4.52	OSP^3/REG	2.16	OSP^3/REG	1.26
	STR/RAT	4.86	SP^3/UMV	2.30	STR/RAT	1.29
	STR/REG	4.89	SP^3/REG	2.30	STR/REG	1.29
	SRS/RAT	5.46	STR/RAT	2.43	SP^3/UMV	1.41
	SRS/REG	5.47	STR/REG	2.43	SP^3/REG	1.41
	SP^3/UMV	5.52	SRS/REG	2.48	SRS/RAT	1.45
	SP^3/REG	5.53	SRS/RAT	2.49	SRS/REG	1.45
	STR/UMV	8.56	STR/UMV	6.13	STR/UMV	5.61
	SRS/UMV	16.49	SRS/UMV	13.17	SRS/UMV	13.46
$n = 8$	OSP^3/UMV	3.07	OSP^3/UMV	1.30	OSP^3/UMV	0.77
	OSP^3/REG	3.09	OSP^3/REG	1.31	OSP^3/REG	0.77
	STR/RAT	3.11	SP^3/UMV	1.38	STR/RAT	0.84
	STR/REG	3.13	SP^3/REG	1.39	STR/REG	0.84
	SP^3/UMV	3.51	STR/RAT	1.45	SRS/RAT	0.97
	SP^3/REG	3.53	STR/REG	1.46	SRS/REG	0.97
	SRS/RAT	3.54	SRS/REG	1.61	SP^3/UMV	0.98
	SRS/REG	3.54	SRS/RAT	1.61	SP^3/REG	0.98
	STR/UMV	4.71	STR/UMV	2.87	STR/UMV	2.59
	SRS/UMV	10.66	SRS/UMV	8.06	SRS/UMV	8.50

It is clear that large gains are possible by using fully the information on current diameters instead of ignoring it (SRS/UMV) or of using it just for stratification (STR/UMV). Simple random sampling with ratio or regression estimators was generally inferior to stratification and variable probability sampling.

There was no gain from using regression over the ratio or UMV estimators in stratified and variable probability sampling, respectively. Regression could be useful, however, in estimating variables not so directly related to current basal area, such as height or volumes.

The most efficient methods were OSP^3/UMV or STR/RAT, with the first slightly better for the 2-year intervals, and the second for the 6-year intervals. Implementing the stratification procedure is somewhat more complicated than OSP^3 sampling. Together with the theoretical bias of the ratio estimator, this might suggest OSP^3/UMV as the better method in general.

SP^3/UMV was not far behind, and the fact that it can be approximated without having to list and number all the plot trees (i. e., without a sampling frame) would make it preferable in many circumstances. The approximation involves having to guess the current basal area before measuring the trees, and can cause some variation in the intended sample size. Incidentally, as indicated by Schreuder et al. (1971), this seems an attractive alternative to the popular 3-P sampling in forest inventories, especially since the advent of intelligent portable data-loggers. Sample size is more predictable, and precision can be expected to be higher.

As expected, the RMS error is higher for 6-year than for 2-year intervals. Data plotting shows much weaker relationships between current and past basal areas in very young stands compared with more mature ones. This contributes to the differences in performance with tree size observed for the methods based on these relationships.

To evaluate the magnitude of the sampling errors and choosing a sample size, it would be necessary to compare them to measurement and other error sources. For example, Kinashi (1953) found a standard deviation of 0.44 cm in random callipering of trees having a mean dbh of 20 cm. This is about 4.4% for a tree basal area, or 0.88% for the mean of 25 trees. More research on the precision of tree-ring data is needed.

Although the relative basal area errors are smaller for larger trees and shorter time intervals, the basal area increments are also smaller. For the best methods it was found that the relative increment errors are not strongly dependent on mean tree or increment size. The averages (arithmetic means) of the relative RMS errors for basal area increments are given in Table 4.

Table 3: Relative RMS error of past basal area estimates (%). 6-year intervals

	Mean diameter, cm					
	3 – 28		29 – 37		38 – 56	
$n=2$	OSP^3/UMV	10.43	STR/RAT	6.98	STR/RAT	5.39
	OSP^3/REG	10.49	STR/REG	7.00	STR/REG	5.40
	STR/RAT	10.58	SP^3/UMV	7.44	OSP^3/UMV	5.74
	STR/REG	10.62	SP^3/REG	7.45	OSP^3/REG	5.75
	SP^3/UMV	11.06	OSP^3/UMV	7.53	SP^3/UMV	5.86
	SP^3/REG	11.09	OSP^3/REG	7.54	SP^3/REG	5.86
	SRS/RAT	12.03	SRS/RAT	7.71	SRS/REG	6.07
	SRS/REG	12.04	SRS/REG	7.72	SRS/RAT	6.07
	STR/UMV	16.30	STR/UMV	14.61	STR/UMV	13.18
	SRS/UMV	23.25	SRS/UMV	20.81	SRS/UMV	19.32
$n=4$	OSP^3/UMV	6.82	STR/RAT	4.55	STR/RAT	3.59
	OSP^3/REG	6.85	STR/REG	4.57	STR/REG	3.60
	SP^3/UMV	6.94	OSP^3/UMV	4.69	SP^3/UMV	3.77
	SP^3/REG	6.98	OSP^3/REG	4.71	SP^3/REG	3.77
	STR/RAT	7.09	SP^3/UMV	5.07	OSP^3/UMV	4.03
	STR/REG	7.12	SRS/RAT	5.07	OSP^3/REG	4.04
	SRS/RAT	7.82	SRS/REG	5.08	SRS/REG	4.17
	SRS/REG	7.83	SP^3/REG	5.09	SRS/RAT	4.17
	STR/UMV	9.32	STR/UMV	7.82	STR/UMV	6.46
	SRS/UMV	15.53	SRS/UMV	13.89	SRS/UMV	13.75
$n=8$	SP^3/UMV	4.04	STR/RAT	2.96	STR/RAT	2.28
	SP^3/REG	4.05	STR/REG	2.98	STR/REG	2.28
	OSP^3/UMV	4.39	OSP^3/UMV	3.14	OSP^3/UMV	2.46
	OSP^3/REG	4.42	OSP^3/REG	3.15	OSP^3/REG	2.47
	STR/RAT	4.69	SP^3/UMV	3.29	SP^3/UMV	2.59
	STR/REG	4.71	SP^3/REG	3.30	SP^3/REG	2.61
	SRS/RAT	4.99	SRS/RAT	3.45	SRS/REG	2.67
	SRS/REG	5.00	SRS/REG	3.45	SRS/RAT	2.67
	STR/UMV	5.46	STR/UMV	4.12	STR/UMV	3.61
	SRS/UMV	9.93	SRS/UMV	9.20	SRS/UMV	8.78

It is interesting that there are no large differences either between the 2-year and the 6-year intervals, although some interaction between method and interval length is again apparent.

Table 4: Relative RMS error of basal area increment estimates (%)

	Method	2 years	6 years
	STR/RAT	14.5	12.9
$n = 2$	OSP^3/UMV	14.5	13.7
	SP^3/UMV	15.8	13.9
	STR/RAT	9.7	8.5
$n = 4$	OSP^3/UMV	9.2	9.1
	SP^3/UMV	10.5	9.2
	STR/RAT	6.2	5.5
$n = 8$	OSP^3/UMV	5.8	6.2
	SP^3/UMV	7.3	6.2

6.2 Quantile-based estimation

For the quantiles method there is only one error value from each plot. RMS errors were computed within each of the three tree-size groups. These are compared in Table 5 with RMS errors for OSP^3/UMV calculated by pooling the square errors of the 100 replications. In the table, the first number of each pair corresponds to the quantiles method, and the second to OSP^3.

The variability of the quantile RMS errors for the size groups is large, since they are based on just 16 or 17 values. The OSP^3 method appears

Table 5: Mean tree basal area RMS errors (cm^2) for the quantiles and OSP^3/UMV methods

		Tree-size groups			
Interval	n	Small	Medium	Large	All
2-yr	2	22.5 19.6	35.5 32.4	64.1 33.8	44.4 29.3
2-yr	4	15.8 11.9	34.4 21.5	27.6 21.8	26.9 18.9
2-yr	8	13.1 8.1	20.6 13.0	27.8 13.8	21.4 11.9
6-yr	2	67.3 58.6	96.2 83.9	63.4 98.9	76.6 82.1
6-yr	4	35.8 40.8	52.9 52.6	75.8 66.5	57.3 54.3
6-yr	8	21.8 31.0	45.1 36.9	35.2 44.9	35.1 38.0

superior for the 2-year intervals, but for the 6-year intervals the comparison is inconclusive. With model-dependent methods there is a risk of bias causing spurious relationships between variables, which could be especially troublesome in growth modelling applications. Therefore, on the base of these results the quantile-based method cannot be recommended.

It seems plausible, however, that related approaches, such as using BLUE instead of ABLE estimators, could result in better performance. It was found that with $n = 8$ and $N \leq 30$, for example, Hassainen's spacings produce samples with less than 8 distinct trees, so that taking into account the finite nature of the population would certainly improve the estimates in those instances. Methods could be designed also aiming at minimizing the variance of estimators for the mean instead of for the distribution parameters. Another possibility would be to introduce randomization and develop a hybrid method, with a sampling plan near-optimal under the assumed order statistics model and a model-independent estimator.

7 Conclusions

To estimate past basal area, methods that exploit the correlation between current and past diameters were much more efficient than unbiased estimation with simple or stratified random sampling.

The best methods were a near-optimal stratification with ratio estimation, and an "ordered systematic with probability proportional to prediction" (OSP^3) sampling plan with the Horwitz-Thompson estimator. Regression estimators did not reduced the mean square errors. The OSP^3 method is somewhat easier to implement and is free of the theoretical bias of the ratio estimator, so that it may be recommended in preference to the stratification approach.

A similar method without pre-ordering (SP^3) was almost as good. It can be approximated without the need for numbering and listing all the plot trees, which would make it the preferred alternative in many circumstances.

A model-dependent approach based on order statistics of the Weibull distribution seems reasonably precise, but not as good as OSP^3. Further refinements might make this and related methods an attractive option.

Definite decisions about sample sizes would require more information on measurement and other non-sampling errors. Strategies for the joint estimation of basal area and height also need to be investigated.

Acknowledgments

Thanks to Mark Kimberley for discussions on statistical analysis of the results.

References

Bailey, R. L. (1980). Individual tree growth derived from diameter distribution models. *Forest Science*, *26*(4), 626–632.

Chan, L. K. and Cheng, S. W. (1988). Linear estimation of the location and scale parameters based on selected order statistics. *Commun. Statist. -Theory Meth.*, *17*(7), 2259–2278.

Cochran, W. G. (1977). *Sampling Techniques* (Third ed.). Wiley.

Connor, W. S. (1966). An exact formula for the probability that two specified sampling units will occur in a sample drawn with unequal probabilities and without replacement. *Journal of the American Statistical Association*, *61*, 348–390.

García, O. (1988a). Experience with an advanced growth modelling methodology. In Ek, A. R., Shifley, S. R., and Burke, T. E. (Eds.), *Forest Growth Modelling and Prediction*, pages 668–675. USDA Forest Service, General Technical Report NC-120.

García, O. (1988b). Growth modelling — a (re)view. *New Zealand Forestry*, *33*(3), 14–17.

García, O. (1991). What is a diameter distribution? *Presented at the IUFRO Symposium on Integrated Forest Management Information Systems, Tsukuba, Japan, October 13–18, 1991.*

Grosenbaugh, L. R. (1965). Three-pee sampling theory and program "THRP" for computer generation of selection criteria. Research Paper PSW–21, USDA Forest Service.

Hansen, M. H., Madow, W. G., and Tepping, B. J. (1983). An evaluation of model-dependent and probability-sampling inferences in sample surveys. *Journal of the American Statistical Association*, *78*, 776–793.

Hartley, H. O. (1966). Systematic sampling with unequal probability and without replacement. *Journal of the American Statistical Association*, *61*, 739–748.

Hassanein, K. M. (1972). Simultaneous estimation of the parameters of the extreme value distribution by sample quantiles. *Technometrics*, *14*(1), 63–70.

Kinashi, K. (1953). Forest inventory by sampling methods. Bulletin of the Kyūshū University Forests 23, pages 1–153.

Pilaar, C. H. and Dunlop, J. D. (1990). The permanent sample plot system of the New Zealand Ministry of Forestry. In Adlard, P. and Rondeux, J. (Eds.), *Forest Growth Data: Capture, Retrieval and Dissemination. Proceedings of the Joint IUFRO Workshop S4.02.03-S4.02.04 held on 3-5 April 1989 at Gembloux, Belgium*, pages 5–17. Faculty of Agriculture of Gembloux, Belgium.

Särndal, C. E. (1980). On π-inverse weighting versus best linear unbiased weighting in probability sampling. *Biometrika*, *67*(3), 639–650.

Schreuder, H. T., Sedransk, J., Ware, K. D., and Hamilton, D. A. (1971). 3-P sampling and some alternatives, II. *Forest Science*, *17*(1), 103–118.

Sukhatme, P. V. and Sukhatme, B. V. (1970). *Sampling Theory of Surveys with Applications*. Iowa State University Press.

Sutton, W. R. J. (1973). Changes in tree dominance and form in a young radiata pine stand. *New Zealand Journal of Forestry Science*, *3*(3), 323–330.

Tennent, R. B. (1988). A New Zealand data collection procedure for growth modelling. In Leech, J. W. et al. (Eds.), *Modelling Trees, Stands, and Forests*. School of Forestry, The University of Melbourne.

Zacks, S. (1971). *The Theory of Statistical Inference*. Wiley.