SIMULATION OF SUBSAMPLING SELECTION RULES FOR HARDWOOD TREE HEIGHTS[1/]

Edwin J. Green and Charles T. Scott[2/]

Abstract.--Apart from diameter, height is often the most important tree dimension routinely measured in forest surveys. Because height is somewhat difficult and costly to measure, it is common to actually measure height on only a subsample of trees and to ocularly estimate height for the remaining sample trees. In this study we determine the "optimal" rule for selecting hardwood trees to subsample for measured total and merchantable heights. The data consisted of 20 fifth-acre plots in hardwood stands in New Jersey. On each plot, sawlog height and total height were estimated ocularly and measured on every tree. Distance and azimuth to each tree were also recorded. Six possible subsampling rules for fixed-area plots were established. Each rule was simulated 1000 times for varying subsampling intensities. Predicted heights were compared to known measured heights. The "optimal" subsampling rule was determined from the mean squared errors and cost data. Systematic subsampling rules are recommended.

INTRODUCTION

The USDA Forest Service, Northeastern Experiment Station's Forest Inventory, Analysis and Economics (FIA&E) unit is charged with, among other things, determining forest volumes in each state in its region during periodic surveys. This is accomplished through measurement of permanent and temporary sample plots of various shapes and sizes. Based on Scott (1986), FIA&E has determined that, for its purposes, ocular estimation of tree heights alone is not sufficient to produce accurate estimates of volume and volume growth, in particular. For the 1986-87 New Jersey forest survey, an estimated height is recorded for every sample tree. Trees are subsampled for height measurement in a multi-stage sampling framework (Murchison 1984). Measured heights are then regressed on estimated heights for the subsampled trees, and the resulting relationship is used to adjust the estimated heights for the remaining sample trees. These adjusted and measured heights are then combined with the measured diameters for volume determination. This study was undertaken in order to determine the optimal rule for selecting the subsample trees from fixed-area plots.

DATA

The data for this study came from 20 circular fifth-acre plots, randomly located in mixed hardwood stands of various stages of development. Two two-person crews performed the measurements. The stands were located in central New Jersey. On each plot, diameter at breast height, species, and cruiser were recorded for each tree. Total height was estimated for every tree, and for trees meeting sawlog requirements, sawlog height was estimated. A sawlog tree was one with a minimum dbh of 11 inches for hardwoods or 9 inches for softwoods, with a bole section at least 8 feet in length meeting minimum diameter requirements (9 inches for hardwoods, 7 inches for softwoods). Additionally, the distance and azimuth to each tree from plot center were recorded. After this process was completed, all trees were measured with a clinometer for total and sawlog heights. Over all 20 plots, there were 767 trees, with 421 of sawlog size.

[1/] Paper presented at the Sixth Central Hardwood Forest Conference, Knoxville, TN, February 24-26, 1987.

[2/] Assistant Professor, Forestry and Wildlife Section, Cook College - Rutgers University, P.O. Box 231, New Brunswick, NJ 08903, and Research Forester, Northeastern Forest Experiment Station, 370 Reed Rd., Broomall, PA 19008.

SUBSAMPLING RULES

The general type of estimator used in this study was a simple linear regression estimator. A preliminary study indicated weighted regression was unnecessary and that measured heights were best estimated by regressing measured heights on estimated heights for the subsample trees and applying the regression to the non-measured trees (Roesch et al. 1986). The preliminary study also indicated that a separate regression should be developed for each cruiser, but that separate regressions were not needed for each plot. Residual plots suggested that simple linear regressions were appropriate.

Six possible rules[2/] for selecting the subsample trees were identified and tested. For each subsampling rule 1000 simulations were performed for at least four subsampling intensities (ratio of total number of subsampled trees, across all plots to the total number of primary sample trees, across all plots). The subsampling intensities used were approximately 50, 70, 80, and 90 percent. The subsampling rules were to:

1. Subsample n trees randomly without replacement across all plots with equal probability per tree (EQPRB). Under this rule, each tree was visited, and a random number drawn from a Uniform (0,1) distribution. If the random number was less than a pre-specified, constant probability p, the tree was subsampled. If the random number was greater, it was not subsampled. The probability, p, was chosen to yield the desired subsampling intensity.

2. Subsample the first n trees in each two-inch diameter class, starting from a randomly selected tree (NDCLS). The n values were chosen to yield the desired sampling intensities.

3. Subsample every nth tree, starting from a randomly selected tree (EVRYN). The value of n was chosen to yield the desired sampling intensity.

4. Randomly subsample n trees per plot (NTREE) with equal probability and without replacement. Again, n was selected to achieve a given level of sampling intensity. Note that with this subsampling rule, the number of sample trees per plot was fixed (unless the number of trees on the plot was less than n). Also, in order to randomly choose n trees, we needed to know how many trees were on the plot so that random numbers could be drawn from a Uniform $(1,n_p)$ distribution, where n_p = number of trees at the particular plot. Thus the sample trees would be visited once to obtain the count. Then the subsample trees would be visited again for height measurement. This is a list sampling technique (e.g., see Avery and Burkhart 1983).

5. Subsample the first tree in each 360/n sector selected (SECTR). Here the sampling intensity was controlled by varying the number of sectors. Randomness was introduced by drawing a Uniform (0,360) random number before each simulation, and shifting the azimuth of each tree by that number of degrees.

6. Subsample trees with probability proportional to estimated height, with the selection procedure adjusted such that the desired subsampling intensity was obtained (PPEHT). The subsample was selected by drawing a random number for each tree. If the estimated height for that tree was greater than the random number, the tree was selected. The random number followed a Uniform $(0,K_c)$ distribution. The upper bound of the distribution was cruiser-specific and was determined as follows:

$$K_c = (M_c/p)\ 100 \qquad (1)$$

where: K_c = upper bound for cruiser c, c=1,2,3,4
M_c = mean estimated height (total or sawlog) over all trees observed by cruiser c.
p = desired percentage of trees subsampled, or subsampling intensity.

With this rule, the expected number of subsample trees is (767)p if subsampling for total height, and (421)p if subsampling for sawlog height. The probability of selection per tree is proportional to the height of the tree, so that tall trees will be sampled more intensively than short trees.

Of the above six subsampling rules, only rule 4 (NTREE) is a list sampling rule, i.e., a tree list is required on each plot before subsample trees can be selected. This should weigh heavily against rule 4 as implementation of this rule would incur an extra cost in the time used to count trees on each plot.

None of the six subsampling rules results in a fixed total sample size. Sample sizes for rule 3 (EVRYN) is fixed only when the total number of trees in the population is known in advance. Sample sizes for rules 1 (EQPROB) and 6 (PPEHT) depend on the sequence of random numbers drawn and on how well the population size is estimated.

2/A seventh subsampling rule, vertical point sampling, under which trees are selected proportional to their heights was identified but not included in this study due to difficulties in simulating this procedure without making dubious assumptions about the spacing of the sample trees.

Sample sizes for rules 2 (NDCLS), 4 (NTREE), and 5 (SECTR) are fixed only if each plot has a sufficient number and distribution of trees. However, an upper bound on the sample size could be computed for each of these rules.

COST PER SAMPLING RULE

Cost per subsampling rule can be considered as the time required to implement the rule since no other direct costs are involved. Based on Scott (1981) and other available data, a relationship was developed to estimate the time required to implement a subsampling rule. In general, this relationship was:

$$t_j = m\, t_1 + m\, N\, t_e + m\, n\, t_m + I_j(m\, n\, t_r) \qquad (2)$$

where: t_j = time (min.) required to implement subsampling rule j,

t_1 = time (min.) required to locate the sample plot,

t_e = time (min.) required to estimate the height of a tree,

t_m = time (min.) required to measure a tree for height,

t_r = time (min.) required to re-locate a tree when using a list subsampling technique,

m = number of plots,

N = average number of trees per plot,

n = average number of subsampled trees per plot,

I_j = 1 if rule j is a list sampling rule, 0 if not.

The time to estimate tree height, t_e, is a function of the time to perform the observation and the time to walk to the tree. The distance to be walked is a function of the square root of the area sampled. Based on Scott (1981) and interval timing studies, the relationship used was:

$$t_e = 2.3 + 0.777\, z^{1/2}, \qquad (3)$$

where:

z = plot size in acres.

The time to relocate is a function of the distance between trees to be relocated and depends on the plot size and the number of trees to relocate. More trees per unit area reduces the distance between trees. The relationship for t_r was developed the same way as that for t_e, and is:

$$t_r = 0.59\, f + 0.0341/z^{1/2}, \qquad (4)$$

where: f = subsampling intensity (n/N).

Assuming 160 trees per acre, N is:

$$N = 160\, z.$$

The other values used were:

$$t_1 = 15$$
$$t_m = 3.$$

Upon substitution, t_j becomes:

$$t_j = m(t_1 + 160\, z\, \{2.3 + 0.777\, z^{1/2} + 3.0\, f + I_j(0.59\, f + 0.0341/z^{1/2})\}). \qquad (5)$$

In the present study, m and z are constant (20 and 0.20, respectively), thus t_j is a function of only f and I_j.

EVALUATION PROCEDURE

Each subsampling rule was implemented 1000 times for each desired subsampling intensity over the twenty fifth-acre plots by cruiser for both sawlog height and total height. A simple linear regression of measured height on estimated height was developed. These regressions were used to predict measured heights for the non-subsampled trees. The regression used for each non-subsampled tree was the one corresponding to the cruiser who estimated the height for that tree. A mean square difference between predicted heights for non-subsampled trees and actual measured heights for these trees was computed on each simulation trial. These mean squared differences were averaged over 1000 simulations. This statistic was used to assess the precision of the subsampling rule.

After the simulations were completed, the field time required for each rule was estimated using equation (5) and the average subsampling intensity[3/] over the 1000 simulations.

RESULTS

For both heights, a graph of cost (time) vs. precision was established, and the results from the six rules were super-imposed. The results are presented in Figures 1 and 2. From the figures it is apparent that the differences between the various subsampling rules are small. Because of the cost involved in counting the trees before the sample selection, the list sampling rule (NTREE) yielded the highest cost per level of precision and was thus the least efficient procedure. Of the remaining procedures, selecting the first tree in each 360/n sector (SECTR) performed best for total heights. This rule yielded a cost-precision relationship that was closest to the origin (Fig. 1). SECTR was not such a clear choice based on sawlog heights, where EVRYN might be preferred because of its lower cost at higher (less precise) levels of precision. The rule to choose would depend on the relative importance of sawlog heights versus total heights. In this case, because of its almost complete domination of the other rules for total height, and because it appeared competitive on sawlog heights, we would recommend the SECTR rule.

[3/]Because of the random nature of the simulations, the desired subsampling intensity was a target figure. The subsampling intensity achieved, while usually close to the target, could not be constrained to equal the target.

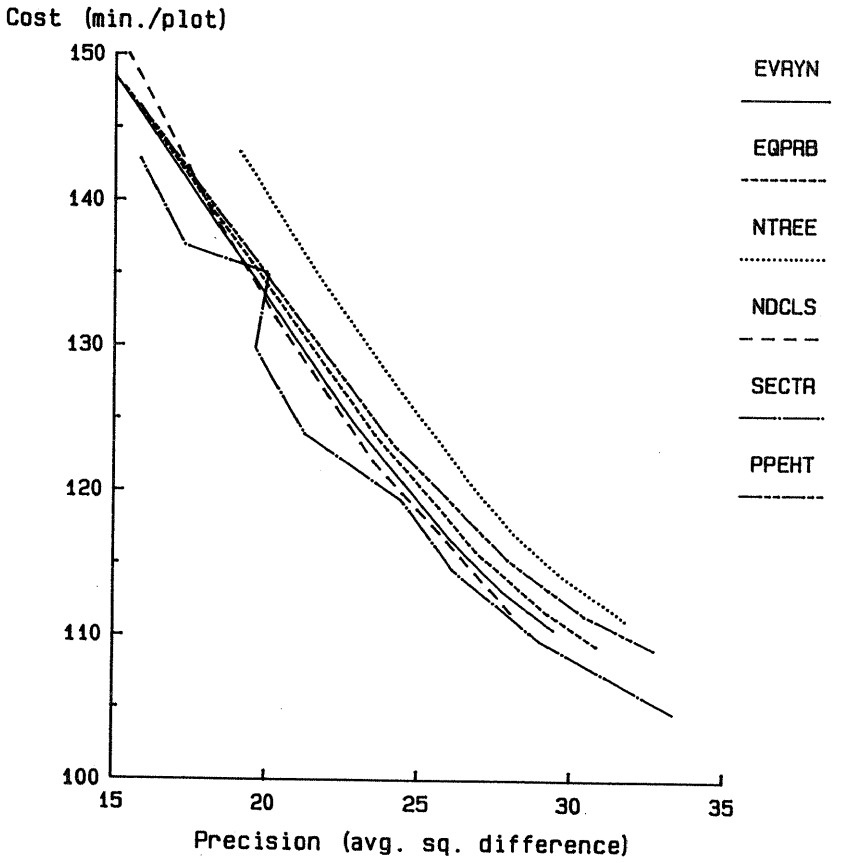

Figure 1.--Cost vs. precision for 6 subsampling rules for total height (plot size is 1/5-acre).

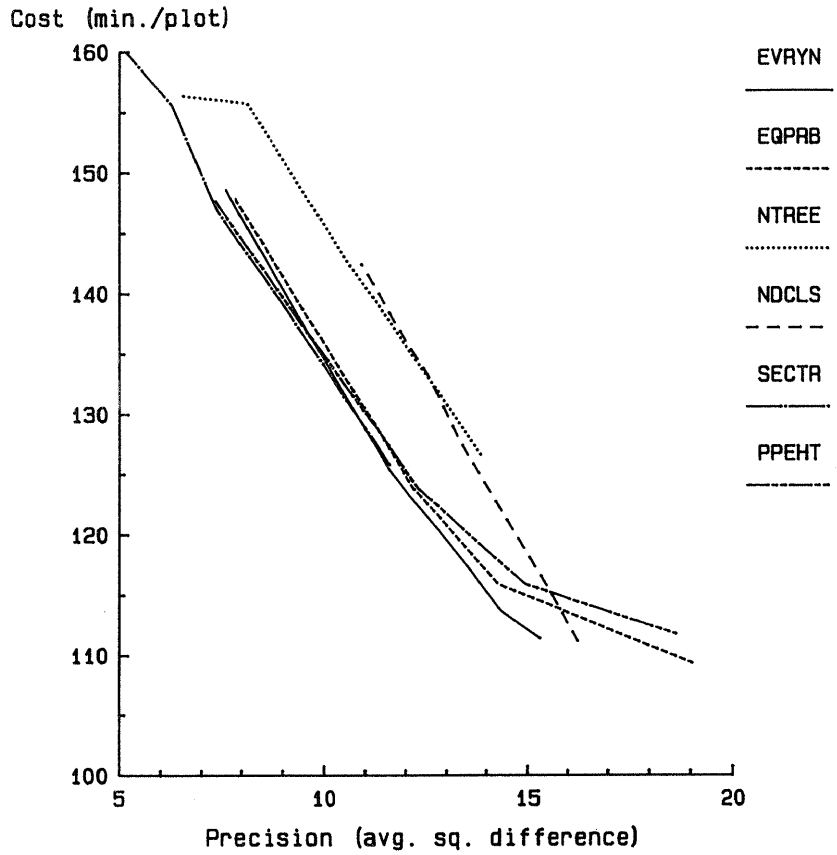

Figure 2.--Cost vs. precision for 6 subsampling rules for sawlog height (plot size is 1/5-acre).

CONCLUSIONS

As might have been expected, there were no major differences among the various subsampling selection rules, except for the inefficiency of the list sampling rule. It is also interesting to note that the systematic rules (SECTR, EVRYN and NDCLS) generally performed as well as, if not better than, the random selection rules (EQPRB, NTREE, and PPEHT). The one unequal probability selection rule (PPEHT) did not perform as well as might have been expected. Perhaps its performance would improve if estimation of mean height was the objective rather than selecting trees to develop a good prediction equation.

Based on these results and the well-known cost-effectiveness of systematic sampling (compared to random sampling in non-periodic populations), combined with the ease of systematic sampling in the field, we recommend that the first tree per (360/n) degree sector subsampling rule be used for hardwood tree heights on fixed area plots. This is the design currently in use in the Northeastern Forest Survey of New Jersey except that the last tree per sector is selected. This was done to ensure that the cruiser would not alter his/her estimates based on the measured information from the first tree. Thus the prediction equation is equally applicable to all trees.

LITERATURE CITED

Avery, T. E. and H. E. Burkhart.
1983. Forest Measurements. 3rd ed. McGraw Hill, NY, 331 p.

Murchison, H. G.
1984. Efficiency of multi-phase and multi-stage sampling for tree heights in forest inventory. Ph.D. thesis, Univ. Minnesota, St. Paul. 158 p.

Roesch, F. A., Jr., E. J. Green, and C. T. Scott.
1986. A model for predicting measured heights. U.S.D.A. For. Serv. Northeast. For. Exp. Sta. Internal report.

Scott, C. T.
1981. Design of optimal two-stage multi-resource surveys. Ph.D. thesis, Univ. Minnesota, St. Paul 138 p.

Scott, C. T.
1986. A test of ocular estimation of merchantable tree heights. Program and Abstracts, Joint Meeting of Midwest Forest Economists and Midwest Forest Mensurationists. Aug. 13-14. Dept. of Forestry, Univ. of Illinois, Urbana, IL. 32 p.