Sampling Error of Cruises in the California Pine Region

A. A. Hasel[1]

California Forest and Range Experiment Station[2]

To organize cruises so as to steer a desirable middle course between high accuracy at too much cost and little accuracy at low cost is a problem many foresters have to contend with. It can only be done when the cruiser in charge has a real knowledge of the required standard of accuracy and of the variability existing in the timber stand. The study reported in this paper works out estimates of the average accuracy of cruises in California pine stands and outlines methods whereby the accuracy may be estimated for particular cruises.

THE usual method of evaluating the accuracy of a cruise requires data from plots, line plots, or strips that have been selected at random and independently. Such data are, however, almost nonexistent in the California pine region because both past and current practice has been to use regularly spaced sampling units. The fact that all cruises in current use are of the systematic pattern is taken to justify the first purpose of the present study, which is to contribute information as to the average reliability of such cruises with respect to sampling error. The second purpose of the study is to show some appropriate methods of deriving information about sampling error from random cruises.

In virgin stands, the cruise of two strips per 40 acres has long been the practice in the national forests of the California region, strips being a chain width. Until recently this method was also followed in surveys of cut-over land, but recent instructions[3] provide for a single 1- by 2-chain plot to be cruised in the center of each 20 acres (half of a 40, or 10 by 20 chains in dimensions), which reduces the intensity to 1 percent from 10. Intermediate percentages of intensity, when required, are to be obtained by increasing plot length.

The subdivision of area, or block unit, we shall be concerned with in this study is therefore 20 acres, and plot size will vary from 1 by 2 chains to 1 by 20 chains, inclusive.

Thirty-one 20-acre blocks, located within the pine and pine-fir region as shown in Figure 1, provide the basis for the study. For all these blocks the inventory data are complete and segregable by small units. The major timber types in the region are represented. Timber species include ponderosa pine (*Pinus ponderosa* Dougl.), Jeffrey pine (*P. jeffreyi* Oreg. Com.), sugar pine (*P. lambertiana* Dougl.), white fir (*Abies concolor* Lindley and Gordon), Douglas-fir (*Pseudotsuga taxifolia* Britt.), and California incense-cedar (*Libocedrus decurrens* Torr.). Site qualities I to IV, inclusive, in a five-site classification are included. The stands are uneven-aged. With the exception of four blocks, all the areas have been logged, the time of cutting ranging from 1907 to 1938. The different methods of cutting in use during this period are therefore represented.

Variance of Stand Volume

The variance of the mean of a random cruise, s^2_m, is a function of the variance of volume per plot, s^2, the number of plots cruised, n, and the total area in terms of number of plots, N. The equation is

$$s^2_m = \frac{s^2}{n} - \frac{s^2}{N} \qquad (1)$$

The estimate of s^2 within blocks is obtained by taking

$$s^2 = \frac{S\ (x_1 - x_2)^2}{n} \qquad (2)$$

in which x_1 and x_2 are the volumes of the two plots cruised in a block, S is the summation over all blocks, and n is the total number of plots cruised. For a fixed intensity of cruise, the size of plot which minimizes s^2 will give the best cruise estimate.

Sampling error is taken as 2 s_m expressed as percentage of the cruise mean and is the range within which the chances are 19 in 20 that the true mean lies. In a large number of trials the true mean will differ from the cruise mean by less than a single s_m approximately two-thirds of the time.

The values given for variance of plots of the sizes listed in Table 1 are those to which cruise

[1]Grateful acknowledgment is made to Duncan Dunning, of the California Forest and Range Experiment Station, and to Professor F. S. Baker, of the University of California, for helpful suggestions.

[2]Maintained by the U. S. Department of Agriculture at Berkeley, Calif., in cooperation with the University of California.

[3]Timber management handbook. California Region, Forest Service, U. S. Dept. Agric. Mimeographed. 1940.

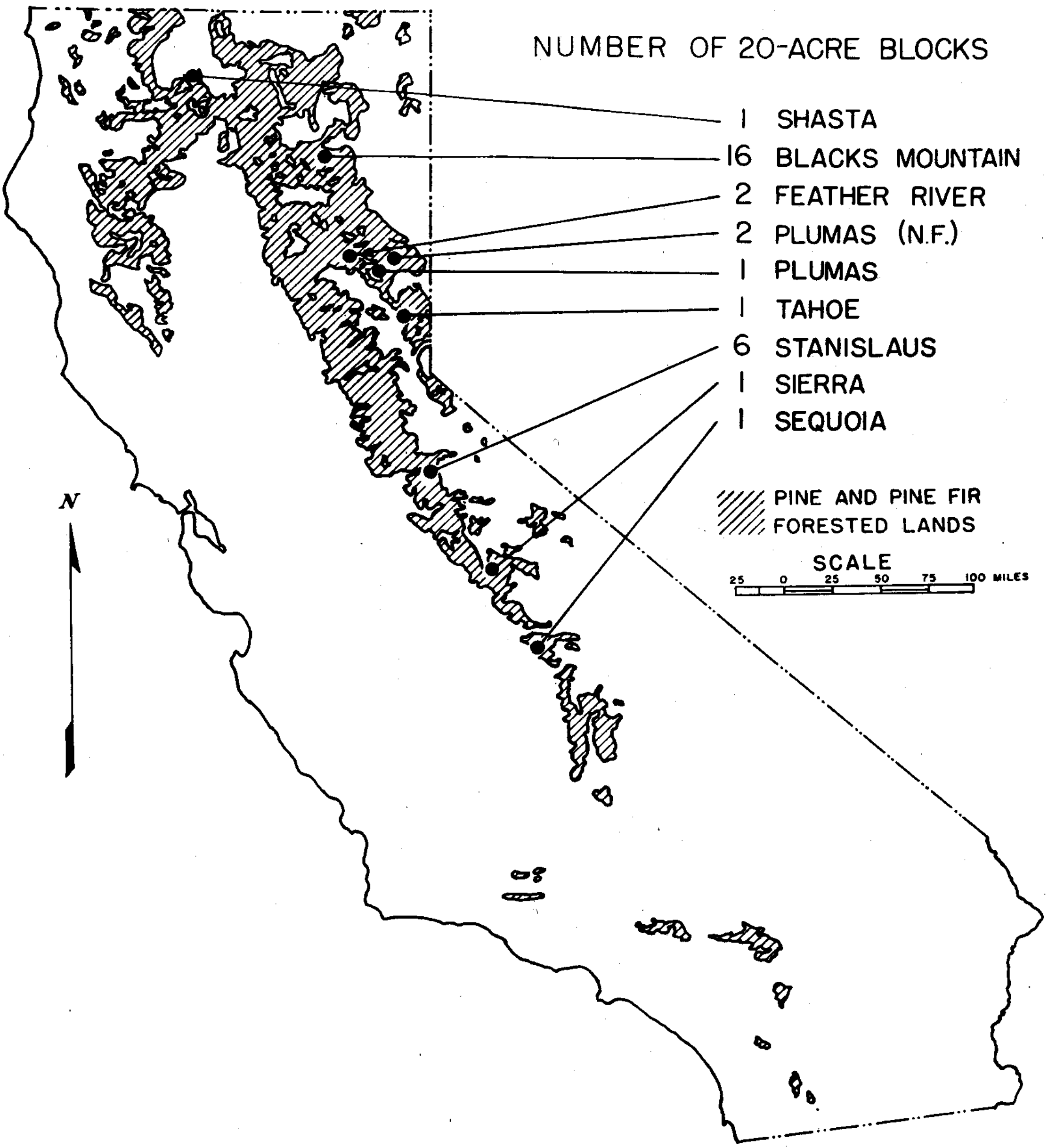

Fig. 1.—Location of blocks within the pine and pine-fir region (distribution of type based on data furnished by the Forest Survey Division of the California Forest and Range Experiment Station).

TABLE 1.—VARIANCE OF VOLUME WITHIN BLOCKS

Groups of blocks[1]	Plot size[2]	Number of plots	Variance[3] per plot
1	1 by 20	160	3.2382
	1 by 10	320	3.5936
	1 by 4	800	4.0702
	1 by 2	1,600	3.8519
2	1 by 12	25	26.9409
	1 by 4	75	18.4970
	1 by 2	150	15.6022
3	1 by 8	66	2.7210
	1 by 4	132	3.3554
	1 by 2	264	3.1782
4	1 by 6	14	4.3408
	1 by 2	42	4.1688

[1]Blocks made up of plots of the same size and shape within groups.
[2]Dimensions in chains.
[3]On 0.2-acre basis in terms of 1,000 feet board measure.

estimates obtained by formula (2) will tend. Blocks made up of plots of the same sizes and shapes are grouped together. Comparison of plot variance within such groups indicates that the effect of plot size within the range tested is not sufficiently important to outweigh the advantages which the use of any of these sizes may have in the field survey. It is obvious that the increased precision of estimate expected with increased intensity of cruise is assured about equally well by taking either more small plots or fewer large plots. The condition to bear in mind is that these values of variance apply within blocks of fixed size, and therefore with a low-intensity cruise the plots must be small to have at least one within each block. If the cruise is to provide an estimate of variance, two plots must be taken at random within each block to apply formula (2).

The variance per plot in the second group in Table 1 is much larger than in the other groups. It is also known that the volume per plot is larger in this group. This relationship of variance to volume is useful in deriving the average variance applicable to existing cruises. Instead of a single average based on all blocks, it is better to use a regression line expressing this relationship as an average varying with volume. Such a regression line could be derived for each plot size, but the spread between lines is so small that this amount of detail is not justified. Therefore the 0.4-acre plot size is used, as it shows either average or slightly greater values of variance compared with the other sizes.

The relationship between variance, s^2, and mean volume per plot, $\bar{x}$, is satisfactorily represented by a straight line when plotted on logarithmic scale as in Figure 2. The equation is

$$\log s^2 = 0.202409 + 1.171241 \log \bar{x} \quad (3)$$

In deriving equation (3), partial volumes such as volume in Dunning tree class 1 and volume of sugar pine or white fir, or both, were used as well as total volume, provided that these volumes were represented on all or almost all plots in a block.

In fitting the regression line in Figure 2 the sum of squares of deviations of logarithms of the observations was minimized by the method of least squares, and on this basis the curve accounts for 96 percent of the variation of s^2 in the data. Other curves fitted directly to the observed values did not give appreciably different estimates of sampling error.

With plots of a constant size, if the coefficient of variability were the same for different average volumes, variance would quadruple with doubling of the volume. The regression coefficient in that case would be 2.00, which is significantly greater than the value obtained in equation (3). Consequently it is apparent that for a fixed accuracy of estimate, expressed in terms of percentage of the mean, less intensive cruises are needed in heavy stands than in light stands. Likewise for a given stand, the estimate of total volume is usually better than the estimates for partial volumes.

Application of the information given in Figure 2 to existing systematic surveys requires that the cruise data be segregated by timber type, site quality, and condition class—i.e., virgin or cut-over land, land cut over by similar methods, and land subjected to similar treatment since cutting. The basis for subdividing into these strata is met by current methods of mapping and recording of cruise data, or it can be obtained from aerial photographs. For each stratum the mean volume is calculated on a 0.4-acre basis and the corresponding variance is read from Figure 2. Application of formula (1), taking n as the number of 0.4 acres in the stratum, gives the estimate of the standard error, s_m, and 200 $s_m \div \bar{x}$ gives the estimated sampling error. If the estimate is desired for all strata combined, it is only necessary to calculate the weighted average s^2_m, take its square root, and proceed as above to determine the sampling error.

Examples of the accuracy of volume estimates are shown in Table 2 for stands of different volumes, cruises of varying intensity, and for different sizes of area to which the cruise estimates apply. The sampling errors apply to separate strata as usually mapped. The size of block is constant, being 20 acres, with dimensions 10 by 20 chains. Plot length is always taken parallel to the long side of the block, and the single plot per block is taken in the center. Plot size varies from 1 by 2 chains in a 1-percent cruise to 1 by 20 chains in a 10-percent cruise. Sampling error will vary from the values shown, the degree depending upon the size of trees and the distribution of stems on the ground. A random cruise would be required to obtain a more accurate estimate of sampling error for a particular cruise. The values in Table 2 are the averages to which a large number of cruises will tend according to the data available. Where volumes run consistently high or low at the center of blocks, systematic cruises will give biased results. Otherwise there is no reason to believe that systematic cruises will estimate volume less accurately than random cruises.

The effect of size of area and average volume upon sampling error is apparent in Table 2. In cruising sizable areas the estimates are frequently segregated by single sections or quarter sections; in that case the accuracy required for these units will be the appropriate basis for judging adequacy of sampling. Except for stands of 10,000 board feet per acre and more, a 10-percent cruise does not provide sufficiently accurate estimates by sections for most purposes. On the other hand, if estimates are to apply to a minimum area of 15 sections, a 1- or 2-percent cruise may be sufficient.

Knowledge of the relationship between variance and volume should prove useful to cruisers in setting up the initial intensity of coverage for random cruises of timberland where prior information regarding degree of variability is lacking. After sufficient plots are taken to provide an improved estimate of variance, the intensity and method of cruise can be adjusted accordingly. However, where strata are small or so irregular in shape that the use of rectangular blocks of equal size is precluded, other methods of estimating both volume and variance are more appropriate than the formulae given in this study. The total number of rectangular blocks possible should provide a good basis for estimating variance, but need not include all or even most of the area. Once variance is satisfactorily estimated, n and N for the total area cruised may be used in applying formula (1).

TABLE 2.—ACCURACY OF CRUISE ESTIMATES OF VOLUME

Volume[1] per acre	Acreage to which cruise estimate applies: 160	640	3,200	6,400	9,600	19,200
	Sampling error[2] of 1-percent cruise					
1					23.7	16.8
2				21.8	17.8	12.6
5			21.1	14.9	12.2	8.6
10			15.8	11.2	9.1	6.4
15			13.4	9.4	7.7	5.5
20		26.5	11.9	8.4	6.8	4.8
	Sampling error[2] of 2-percent cruise					
1			28.9	20.4	16.7	11.8
2			21.7	15.3	12.5	9.0
5			14.8	10.5	8.6	6.0
10		24.9	11.1	7.8	6.4	4.5
15		21.0	9.4	6.6	5.4	3.8
20		18.7	8.4	5.9	4.8	3.4
	Sampling error[2] of 5-percent cruise					
1			18.0	12.7	10.4	7.3
2		30.2	13.4	9.5	7.8	5.5
5		20.6	9.2	6.5	5.3	3.8
10		15.5	6.9	4.9	4.0	2.8
15		13.1	5.9	4.1	3.4	2.4
20	23.2	11.6	5.2	3.7	3.0	2.1
	Sampling error[2] of 10-percent cruise					
1		27.7	12.4	8.8	7.1	5.0
2		20.8	9.3	6.6	5.4	3.8
5		14.2	6.3	4.5	3.7	2.6
10	21.3	10.7	4.8	3.4	2.8	1.9
15	18.0	9.0	4.0	2.8	2.3	1.6
20	16.0	8.0	3.6	2.5	2.0	1.5

[1]1,000 feet board measure.

[2]Percentage of mean. Range for 19 in 20 trials; 2 out of 3 trials will tend to fall within one-half the values given.

VARIANCE OF NUMBER OF TREES

Variability between number of trees on plots was analyzed in the same manner as volume, and the effect of size of plot upon sampling precision was found to be essentially the same, that is, so small that the choice within the range of size of plot studied should depend upon convenience in the field survey.

The relationship between variance per plot and average number of trees is shown as linear on semilogarithmic scale in Figure 3. On this scale the regression line accounts for 83 percent of the variation in the dependent variable for the data represented. The slope of the line is such that the coefficient of variability decreases significantly with increase in number of trees, indicating that as in estimates of volume, heavier and denser stands require less sampling than lighter stands to yield the same degree of accuracy.

The equation of the regression line in Figure 3 is

$$\log s^2 = 0.087726\,\bar{x} + 0.598552 \qquad (4)$$

in which s^2 is variance of number of trees per 0.4-acre plot and $\bar{x}$ is the average number of trees.

The accuracy of estimates of number of trees tends to be as good as or better than volume estimates. For blocks representing the better sites on the west slope of the Sierra Nevada, estimates of number of trees were more accurate than estimates of volume in the same stands. On the poorer sites, such as at Blacks Mountain, the estimates of number of trees were of about the same accuracy as estimates of volume in the same stands, as is shown in Table 3.

The sampling errors shown in Table 3 are based on plots corresponding in size to those used previously for varying intensities of cruises in which a single plot is taken per block, but they are based on true variance of these plots within the Blacks Mountain blocks, rather than upon the curved estimates from Figures 2 and 3. Variability tends to be less in this part of the region than on the west slope of the Sierra Nevada, and consequently the sampling errors are smaller than the curved estimates of variance would show. These are the only locality, site quality, and type for which the data are considered extensive enough to improve upon the curved estimates of variance.

Table 4 tests the assumption that estimates of sampling error may be derived from tree count data by the same methods as for volume measurements. The results demonstrate that variation between cruise means can be considered as a random sampling deviate of variance within blocks. Variability between blocks is significantly greater than variability within blocks. These results are typical of similar tests that were made for other stands and degrees of cutting.

Discussion of Results

The thirty-one 20-acre blocks used as the basis for this study were established during the past 30 years, primarily for periodic determination of growth after different systems of cutting. All the major timber types, site classes, and condition classes in all-aged stands of the pine and pine-fir region of California are represented.

The range of volume and number of trees by blocks was sufficient to show that a well-defined relationship existed between variance and means. The variance did not quadruple with doubling of the mean, however, a fact which indicated that the coefficient of variability decreased, and therefore that heavier stands generally require less sampling than lighter stands for the same degree of accuracy. Thus a virgin stand requires less cruising than the residual stand after logging, and estimates of total volume tend to be more accurate than separate estimates by species, tree classes, and size classes.

Variation along any ordinate of the regression curves of variance over means is greater than variation resulting simply from random sampling of a homogenous timber stand. The curves will not give nearly as accurate estimates of sampling error for particular cruises as random cruise data will give. The curves are intended mainly for use with systematic cruises already

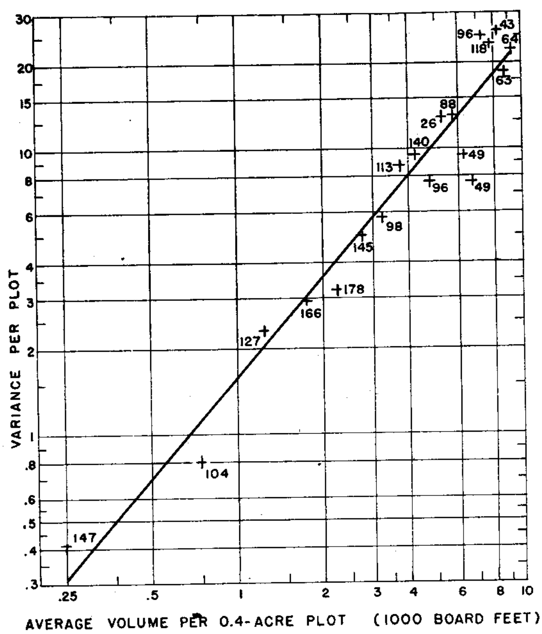

Fig. 2.—Relationship of variance per plot to average volume per plot.

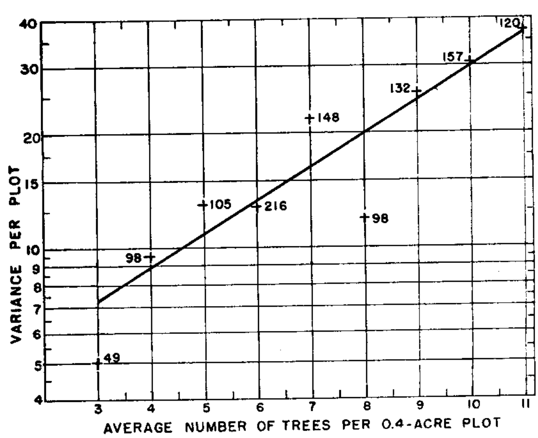

Fig. 3.—Relationship of variance per plot to average number of trees per plot.

TABLE 3.—SAMPLING ERROR OF CRUISE ESTIMATES OF VOLUME AND NUMBER OF TREES IN EASTSIDE PINE[1]

Variable	Mean[2] 1.0 acre	Mean[2] 0.4 acre	1,000 acres 1 percent	1,000 acres 2 percent	1,000 acres 5 percent	1,000 acres 10 percent	10,000 acres 1 percent	10,000 acres 2 percent	10,000 acres 5 percent	10,000 acres 10 percent
			Percent of mean[3]				Percent of mean[3]			
			Virgin stand, 80 acres as basis							
Total volume	20.32	8.13	17.2	12.6	6.7	4.6	5.4	4.0	2.1	1.4
Number of trees	22.15	8.86	17.5	12.8	8.1	5.2	5.5	4.0	2.6	1.6
Volume in tree class 1	1.15	0.46	59.1	41.3	25.7	16.7	18.3	13.0	8.0	5.2
			Lightly cut stand, 80 acres as basis							
Total volume	17.15	6.86	19.4	13.9	8.2	5.3	6.1	4.4	2.6	1.7
Number of trees	19.50	7.80	18.6	13.7	8.7	5.4	5.9	4.3	2.7	1.7
			Moderately cut stand, 80 acres as basis							
Total volume	9.92	3.97	23.0	16.5	10.9	6.8	7.2	5.2	3.5	2.1
Number of trees	14.32	5.73	22.7	16.8	11.0	7.0	7.2	5.3	3.5	2.2
			Heavily cut stand, 80 acres as basis							
Total volume	5.02	2.01	28.6	19.4	13.4	8.4	9.2	6.2	4.2	2.7
Number of trees	10.90	4.36	28.1	20.3	12.8	8.5	8.9	6.4	4.0	2.7

[1]Based on true variance within Blacks Mountain blocks.
[2]Volume in 1,000 feet board measure.
[3]Twice the standard error.

TABLE 4.—ANALYSIS OF VARIANCE OF RANDOM 10-PERCENT CRUISES OF NUMBER OF TREES, BLACKS MOUNTAIN STANDS

Source of variation[1]	Degrees of freedom	Sum of squares	Mean square	F[2]	Significance[3]
		Based on 80 acres, 7.8 trees per 0.4 acre			
Between cruises	19	229.45	12.08	-------	n.s.
10 by 20 within cruise	60	1,309.15	21.82	1.54	++
1 by 4 within 10 by 20	320	4,526.80	14.15		
Subtotal	380	5,835.95			
Total	399	6,065.40			
		Based on 80 acres, 4.4 trees per 0.4 acre			
Between cruises	19	267.85	14.10	1.33	n.s.
10 by 20 within cruise	60	983.55	16.39	1.54	++
1 by 4 within 10 by 20	320	3,399.20	10.62		
Subtotal	380	4,382.75			
Total	399	4,650.60			

[1]Dimensions of blocks and plots in chains.
[2]Ratio of mean square to mean square of 1 by 4 within 10 by 20.
[3]n.s. = not significantly greater; ++ = highly significant.

made and in current use, in which all 20-acre tracts were sampled and segregated by strata. Used in this way the curves estimate the average accuracy which may be expected but from which individual cruises will fluctuate more or less, depending on the distribution of stems on the ground. A second suggested use of the curves is in setting up the intensity to use in starting a forest inventory. Any prior knowlege of average volume, based either on ocular inspection of the tract or on a small preliminary sample will then be helpful in planning the most economical cruise.

The marked effect upon sampling error of the minimum size of area to which cruise estimates are to be applied and the effect of volume per acre on variability indicate that these factors must be carefully considered in aiming at the economical expenditure of time and money in timber surveys. Fixed intensities and methods applied uniformly over a wide range of conditions, even though they may strike a correct average procedure, are likely to give either better estimates or less adequate estimates than are needed for many areas. Very close correspondence between estimate and actual cut may not be entirely satisfactory, because overintensive cruising may be indicated. For any contemplated cruise the standard of accuracy required should be carefully determined, with consideration of the fact that doubling of the intensity will be required to reduce the standard error by less than one-third and quadrupling the intensity will reduce the standard error by one-half. If the estimates are to be applied to quarter sections a much more intensive cruise is needed than if applied to sections. If applied to more extensive areas the amount of biased error caused by faulty measurements, by the use of volume tables not well adapted to the timber, or by error in estimating defect may be so much larger than sampling error that a less intensive but more careful cruise is indicated. All these factors are so variable that no one procedure or only a few cruising procedures would appear to be economical for all conditions and requirements. Flexibility in procedure should be allowed for specific jobs.

It is sometimes felt that a given intensity of cruise is justified for the results obtained in mapping of types and topography, rather than for the timber estimate. This is not true to the same degree now as in the past because many areas are already mapped, either from previous cruises or from aerial photographs. In other cases it would be well to know how much more time and money are spent in cruising than are needed for an adequate sample and judge whether the additional lines needed for mapping alone should be left uncruised and charged wholly to mapping. Sometimes, too, simply taking linear measurements of distance through types will provide the estimate of type acreage required and rapidly made sketch maps of boundaries will suffice for administrative purposes.

The results with respect to the effect of size of plot differ from results in most sampling studies. Over a range of plot size from 1 by 2 chains to 1 by 20 chains, inclusive, a greater effect on sampling precision would be expected. A suggested explanation of this lies in the fact that length of plot extended generally across the contours, just as in regular cruising practice. Variability tends to be greater in this direction, and it happens with these data that variation between plots adjacent end to end in this direction is about the same as variation between plots taken at random. Use of such strip segments as random plots would not, however, give an unequivocal basis for estimating variance.