HILGARDIA

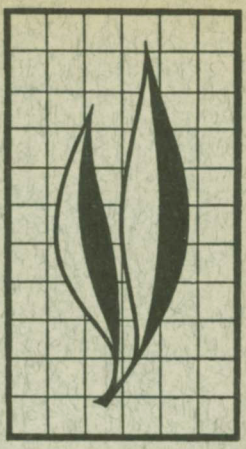
A JOURNAL OF AGRICULTURAL SCIENCE PUBLISHED BY THE CALIFORNIA AGRICULTURAL EXPERIMENT STATION

*Volume 38, Number 2 • March, 1967*

# Computer Simulation and Economic Efficiency in Forest Sampling

Loukas G. Arvanitis and William G. O'Regan

UNIVERSITY OF CALIFORNIA DIVISION OF AGRICULTURAL SCIENCES



In most forestry operations, estimates must be made. If those estimates are to be applied properly in making decisions, the forester must be able to determine their accuracy and their relative efficiency. Better methods are needed to help the decision-maker design forest sampling systems.

The study reported here was designed to demonstrate the use of the computer to simulate a forest sampling problem. Somewhat empirical in nature, the investigation sought to provide those engaged in forest sampling with a method of analysis to help in making decisions in a state of uncertainty.

In addition to providing limited answers to a specific problem—that of optimum combination of number and size of plots in sampling forests—the study suggests an approach to solving problems of this and a similar nature by introducing concepts from production economics and by using computer simulation.

---


THE AUTHORS:

Loukas G. Arvanitis was Assistant Specialist, School of Forestry, Berkeley, and is now Research Scientist, Forest Management Research and Services Institute, Canadian Department of Forestry, Ottawa.

William G. O'Regan is Mathematical Statistician, Pacific Southwest Forest and Range Experiment Station, Forest Service, U. S. Department of Agriculture, and Lecturer in Forestry, Berkeley.

*Loukas G. Arvanitis and*
*William G. O'Regan*

# Computer Simulation and Economic Efficiency in Forest Sampling[1]

## INTRODUCTION

IN ANY SAMPLING procedure, either the cost or the reliability of the information desired must be determined, the main objective being to obtain either maximum information about a given population for fixed expenditure, or a specified amount of information at minimum cost.

It is common knowledge that the larger the plot size and the greater the number of plots, the better the estimate (in terms of smaller variance). Reduction of variance, however, is not the complete solution. Since large samples are more costly, some accommodation between reliability and cost is required.

Much of the research in sampling, over the past 30 years, has concerned methods of determining the best plot size for obtaining estimates of population parameters. Many workers have conducted numerous experiments (mainly in agriculture or ecology) to provide guidelines for practical applications and to develop supporting theories (Mahalanobis, 1940, 1944, 1946; Yates, 1935; Sukhatme, 1946; Bormann, 1953; Cornfield, 1951).

Other work on sampling that has had a pronounced effect on forest estimation procedures includes: Neyman's (1934) study on optimum allocation in stratified sampling; Cochran's (1963) consideration of ratio and regression estimates; Hansen, Hurwitz, and Madow (1953) on the theory of sampling proportional to some measure of size; Mahalanobis' (1944) investigation of problems in surveys; Sukhatme and Seth (1952) on nonsampling errors; Bitterlich's (1948) idea of variable plot; Grosenbaugh's advancement of point (1958) and 3P sampling (1963); and the study by Palley and Horwitz (1961) on properties of some random and systematic point-sampling estimators.

Although sampling is an important factor in forest management, research on sampling as it relates to forestry has been fairly limited. Some of the work done over the past 10 years may be summarized as follows:

Mesavage and Grosenbaugh (1956) used the reciprocal of the product of the squared sampling error (as per cent) multiplied by the cost as a criterion to determine plot size for volume estimate.

Grosenbaugh and Stover (1957) compared sampling estimates of basal area, cubic-foot volumes, and board-foot volume derived from ¼-acre plots with those derived from points of basal-area factor[2] ten, in southeast Texas, and concluded that "it is merely a question of the magnitude of point sampling's superiority." However, since these workers had not secured cost data for the

[1] Submitted for publication October 25, 1965.

[2] Basal-area factor (for sq. ft. per acre) is defined as 43,560 × haversin of critical angle, i.e., the angle at the sampling point which is subtended by the diameter of a tree at breast height.

¼-acre plots, they were not able to calculate the relative efficiency between points and plots.

Strand (1959) showed that, among various methods of estimation, stratified sampling with allocation of sample trees in proportion to the sum of their basal areas gave better results (in terms of smaller variance) for basal-area growth which, in turn, was less accurately estimated than was volume growth.

Hall (1959), disregarding cost, reported the gain in precision for growth estimates obtained by using measurements from permanent sample plots.

Strand (1957) found that, for Norwegian conditions, the coefficient of variation of basal area or volume decreases approximately in proportion to the inverse of the square root of the plot size. Something similar was reported by Freese (1961) who suggested an approximation useful in many forests in the southern United States—namely, that the squared coefficient of variation is inversely proportional to the square root of the relative plot size. This approximation was tested against a variety of species by Boon (1962) who pointed out that "In most cases the formula as proposed by Freese does not seem to be justifiable, as not only may different tangent values be expected for the different species, but there is also a slight indication of curved forms" (of the coefficient of variation over the plot size when plotted on logarithmic paper). Boon (1962) also criticized Prodan's (1961) statement that, in rather irregular forest stands in which trees tend to group, "small plot sizes should have a greater rate of efficiency compared with larger plot sizes." Following the change in skewness caused by changes of plot size, Boon argued that what Prodan has reported is due to the asymmetric distribution of trees, and concluded that "apart from the influence of plot shape and strip orientation... the nature of the population will finally determine the plot size and variability relation."

Bickford (1962) stated that growth estimates derived from point sampling would be less efficient than those from fixed-area plots since points include fewer smaller trees than do plots.

Shiue (1962), on the other hand, pointed out that "since the point sampling system has the effect of reducing the sampling error caused by larger trees, net growth estimates of overmature, all aged stands will be better than those estimates obtained from a fixed size plot system."

In forest sampling procedures, use of very small plots often results in bias, either because border plants are included in the sample (Sukhatme, 1946), or because sampling units tend to be installed in areas of relatively good growth (Mahalanobis, 1946; Yates, 1935). Such practices can result in overestimation of the population parameters.

Very large plots, on the other hand, are free of border bias, and show lower variance per unit area. Furthermore, regardless of the distribution of the parent population, distributions of measurements based on large sampling units approach the normal distribution at a faster rate than do those derived from small plots (Evans and O'Regan, 1963). Large plots cost less per unit area than do small ones since plot location and establishment are independent of plot size. What this amounts to is that the properties of estimates are controllable by the number and size of random-sampling units. The optimum design for a given set of conditions is a function of the cost and variance functions.

Before 1947, the usual practice in forestry sampling studies was to establish a point on the ground, construct a sampling unit of fixed size in its neighborhood, and measure all elements inside the perimeter of that unit. Under

this sampling procedure, equal probabilities of selection are assigned to all elements of the sample, regardless of their location and size (except for border effect). Bitterlich (1948) initiated a new sampling rule—that of unequal probabilities of selection—which was further developed by Grosenbaugh (1958). This new rule has both advantages and disadvantages, depending on the particular objective of a given study.

In the present study, the authors used simulation techniques to obtain variance functions for various parameters and forests. Cost functions were devised for each estimator, and economically efficient sampling plans were discovered.

The effect of plot shape was not considered. We worked with circular plots of various radii and with variable plots with different basal-area factors (BAF). Although some difference of opinion concerning plot shape exists among investigators (Christidis, 1931; Sukhatme, 1947*a*, *b*), except for long, narrow plots, the effect of shape on the accuracy of the estimate is apparently less important than are size and number of sampling units. We hope to report soon on some aspects of the relationship between plot shape and variance.

## THE SAMPLING PROBLEM

A forest is an aggregation of trees fixed on a given land area. In this study, we have assumed that the area is rectangular. The position of any tree and the location of any sampling point can easily be given in Cartesian coordinates.

Sampling rules are chosen so that each tree has some (non-zero) probability of being observed in a random sample. This probability is obviously a function of the chosen rule, and may depend on the position of the tree, the tree dimensions, and the like.

Depending on the sampling procedure employed, the elementary sampling units could be defined as consisting of (1) only one tree—for example, when sampling is done with probability proportional to prediction; or (2) of one or more trees—for example, when the elementary sampling unit is a cluster of trees. Definition (2) is used in the present investigation. A rule to define clusters, and the number of clusters to be drawn (at random) are decisions to be made.

Trees of a given forest area can be partitioned into elementary sampling units (clusters) depending on the distribution of the trees on the ground and the sampling rule being used (O'Regan and Palley, 1965). Suppose that fixed-area circular plots are to be used in a survey. A sampling rule which takes a point at random (from a large number of possible points on the forest floor) and, with this point as a center, constructs a circle of radius, $R(1)$, generates a set of elementary sampling units (clusters) which is, in general, different from that created by some other radius. By the same token, the same sampling rule, $R(1)$, applied to an area with a different spatial distribution of trees, groups them into clusters in a different way. The same holds true for sampling rules which do not refer to a specified area of the sampling unit, such as point sampling (Bitterlich's procedure) where the clustering can be changed by altering the angle gauge.

In the present study, quantitative measurements of interest are: number of trees, sum of diameters (or of a function of diameters, such as basal area), and volume, all on a per acre basis. These measurements are to be taken either at a specified point of time or as increments during some period of time. The latter covers the growth of a forest.

That this procedure of sampling

points at random and clustering trees about points on the basis of a rule results in unbiased estimates can be demonstrated as follows:

Suppose that there are $M$ elements in the population (trees in a forest, for example), each of which has measurable characteristics $X_i$ and a probability of $P_i$ of being observed in a random drawing, where $0 < P_i \leq 1$.

Suppose that there are $N$ possible sampling points, and let $M_j$ be the number of the population elements associated with a single randomly drawn point, $j$.

Define:

$$T = \sum_{i=1}^{M} X_i$$

where $T$ is the sum of measurements over all $M$ elements in the population.

For the $j$th random point, define a measurement:

$$T_j = \sum_{i=1}^{M_j} X_i W_i$$

where $W_i$ is a weight associated with the $i$th tree.

Now,

$$E(T_j) = \frac{1}{N} \sum_{j=1}^{N} \sum_{i=1}^{M_j} X_i W_i$$

The double sum will be over only $M$ values of $X_i W_i$ because there are only $M$ elements.

Suppose that the $i$th element occurs $N_i$ times in the sum, then:

$$E(T_j) = \frac{1}{N} \sum_{i=1}^{M} N_i X_i W_i$$

For an unbiased estimate,

$$E(T_j) = T$$

and

$$\frac{1}{N} \sum_{i=1}^{M} N_i X_i W_i = \sum_{i=1}^{M} X_i$$

which implies,

$$\frac{1}{N} N_i W_i = 1$$

and

$$W_i = \frac{N}{N_i}$$

but,

$$\frac{N_i}{N} = P_i$$

therefore,

$$W_i = \frac{1}{P_i}$$

This derivation can easily be extended to more than one random point. In the case of circular plots, large $N$, and no edge effect, $P_i$ is approximately proportional to plot area.

It is apparent from the preceding discussion that every sampling rule generates a (conceptual) probability distribution of measurements that is a function of the number and the spatial distribution of the trees on the ground and of the rule selected.

Probability distributions can be described by means of a set of parameters. The central problem of sampling is to control the values of such parameters (mainly variances) of the resulting sampling distribution by proper sample design. Almost always, the interest is centered on the nature of the sampling distribution. Terms such as "unbiased," "minimum variance," and the like refer to the sampling distribution of the estimate.

Choice of a probability framework in sampling is important because it affects both the reliability and the economic efficiency of estimators. We shall show that economically efficient designs are determined by the interaction of variance and cost functions.

# VARIANCE FUNCTION

The relationship between variance and plot size was studied empirically by Smith (1938). He examined a large number of published uniformity trials, and found that for many field experiments this relationship can be described by the equation:

$$\log_e V_x = \log_e V_1 - b' \log_e X$$

where

$V_x$ = variance of yield expressed per unit area among plots of size $X$ units.

$V_1$ = variance of yield among plots of size unity.

$b'$ = regression coefficient, the magnitude of which reflects the relationship between measurements based on adjacent small areas of land.

This logarithmic relationship was linear in most cases. However, Smith found that several fields yielded curvilinear forms. He argued that since, in most populations, the elements of area making up the sampling unit (plot) have measurements that are positively correlated, plots of a given area will differ in their values more than would equal amounts of areas made up of randomly selected elements. That is, they will have larger variances. Smith refers to two papers by Harris (1915, 1920) in which he appears to have proposed using the intraclass correlation coefficient of yields from adjacent areas as a coefficient of heterogeneity.

According to Whittle (1956), the variation of yield variance with size of plot may be explained by allowing for the possibility of spatial correlation between yield of any two points in the plot. This correlation decreases "relatively slowly with increasing distance between the two points as a power function of the distance rather than as an exponential." Whittle went on to point out that "for small regions (plots), the variance is proportional to the square of the area, while for large regions, the variance is proportional to the area."

Koch and Rigney (1951) pointed out that if larger plots were constructed by choosing at random $X$ units to form each plot, the variance of the plot measurement would be:

$$V_x = V_1/X$$

If the plots were formed (as they usually are) by choosing adjacent units, there would tend to be a correlation among these units in each plot, and the variance of plot means would be larger than if they were uncorrelated. In such cases, the formula above would be changed to:

$$V_x = V_1/X^b$$

Expressed in logs, this becomes:

$$\log_e V_x = \log_e V_1 - b \log_e X$$

which is identical with Smith's variance function.

For purposes of estimating optimum plot size, the coefficient $b'$ alone is of interest. Smith recommended that for uniformity data in estimating the regression coefficient $b'$, the variances of the different-sized plots should be weighted by their respective degrees of freedom. Koch and Rigney (1951) pointed out that this procedure is not feasible with experimental data since the variance estimates for different-sized plots are built up from common components, and are frequently highly correlated so that a simple weighting by the degrees of freedom is not accurate.

In order to calculate the regression coefficient $b'$, Wright and Freeland (1960) used weights equal to the re-

ciprocal of the variance of $N$-tree plot means, because "the reliability of a variance varies inversely with its size."

Hatheway and Williams (1958) developed efficient estimates of the constants of the relationship between plot size and variability, and worked out a general method of determining efficient linear estimates when the data are correlated and are of unequal variability. Their suggestion for estimating $b'$ by means of unweighted regression coefficients could sometimes result in values less than minus one—a result unacceptable on physical grounds. Instead, these workers used a minimum variance, unweighted estimate. However, the unacceptable values could also be attributed to sampling error of the variance.

Matern (1960) obtained a good approximation of Smith's formula by using the variance of an isotropic set function.

Recent investigations by the authors indicate that forest variance functions take the following form:

$$\log_e V_x = \log_e V_1 - b_1 \log_e X + b_2 (\log_e X)^2$$

There are two possible bases for this form of the variance function—first, the correlation between measurements on a plot and measurements on the increment to that plot usually increases with plot size (authors' unpublished data). Second, this result is deducible if the per plot measurement follows a negative binomial distribution.

In this study the very high correlation between log of variance, log of plot size, and square of log of plot size in the formula above seemed to make it unnecessary to consider the problem of weights.

For the same subject, Whittle (1956) indicated that "any model which is to provide a satisfactory explanation of the power law decay observed in agriculture must be non-linear. . . ."

Mahalanobis (1944) pointed out that there are two ways to study the variance functions: (1) to carry out an extensive and expensive field experiment using plots of various sizes, and (2) to prepare a complete inventory of an area and investigate the variance function in the laboratory by means of a model. The latter, which is cheaper than the first and permits a large number of trials, was used for the present investigation.

Size of field (or forest) is another aspect of the sampling problem, i.e., how stable is the variance–plot size relationship as field size increases. According to Smith (1938), if the observed field is assumed to be a block of some infinitely large field "the above law can be generalized (so as to be applicable to any size of field) by applying a certain adjustment to the regression coefficient $b'$ so as to give a modified coefficient $b$ applicable to an infinite field." Following Smith's discussion, in the case of an infinitely large field, the variance law could be expressed:

$$\log_e V_x = \log_e V_1 - b \log_e X$$

where $V_1$ means "variance of unit plot size for an infinite field."

A series of corrections for $b'$ (for finite field) were graphically determined by Smith to convert $b'$ to $b$ for infinite field.

As part of this study, a 20-acre forest was subdivided into smaller blocks for which separate regressions of variance on plot size were calculated. Although this forest does not approach Smith's infinite field, it is of interest to note that the curves for blocks of various sizes have the same general shape, and the regression coefficients differ only slightly (fig. 1).

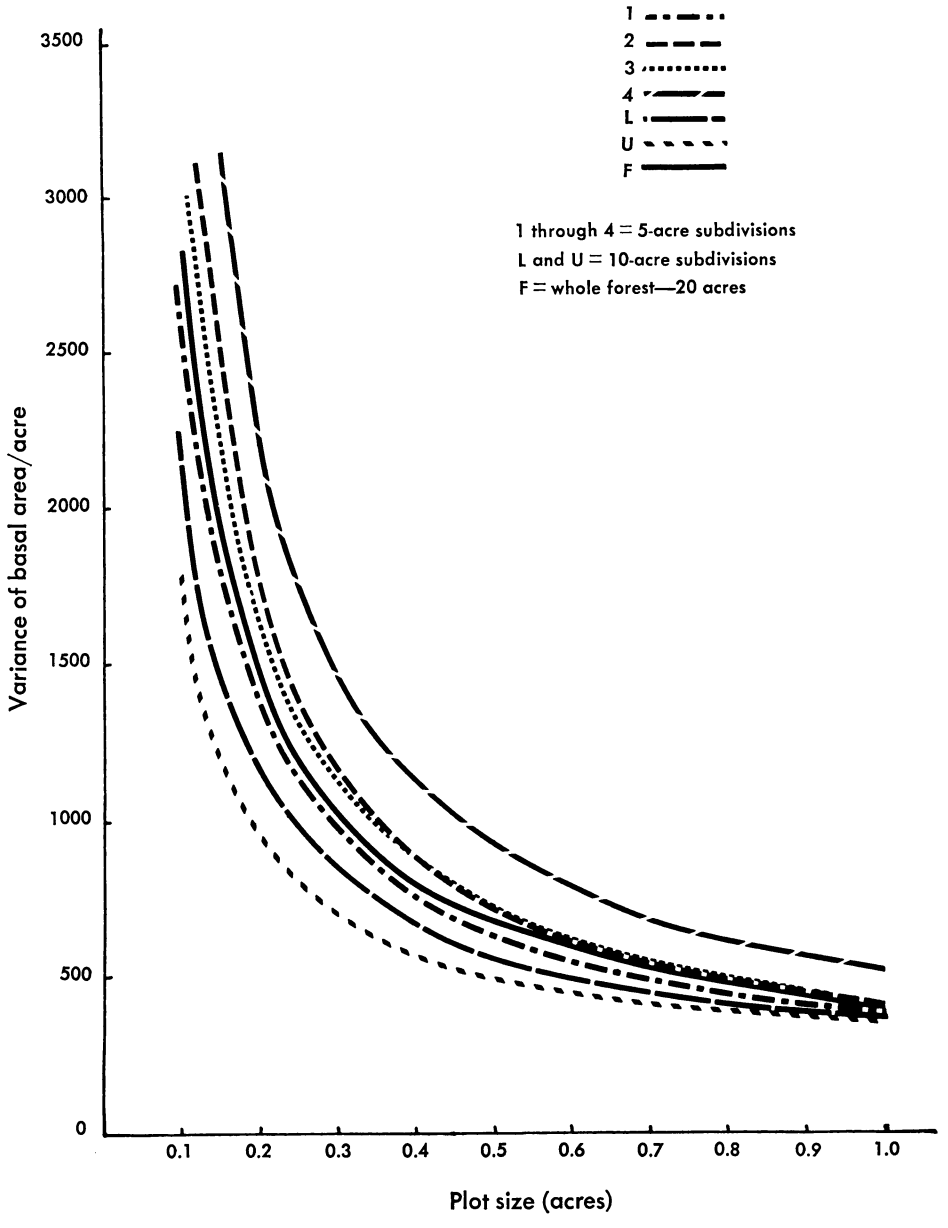



Fig. 1. Variance of measurements of basal area per acre over plot size for subdivisions of block 44(2) of the Blacks Mountain Experimental Forest.

## COST FUNCTION

Economic efficiency in sampling procedures implies a cost function to relate total expenditure to sampling activities. If the amount to be spent is fixed, the aim would be to minimize the variance of the estimate, thus obtaining the maximum possible information, under given conditions, about the quantity being estimated. However, if for some reason the expense is not limited, it might still be desirable to keep the cost as low as possible while still achieving a certain degree of reliability.

According to Hatheway and Williams (1958), departure from linearity of variance over plot size may create problems in some cases. However, if cost data are available, optimum plot size may be estimated with reasonable accuracy without the assumption that Smith's (1938) empirical relationship between variance and plot size holds. Hatheway and Williams stated that if the cost function is:

$$B = C_1 + C_2X$$

where $C_1$ is the cost of a plot (regardless of size), $C_2$ is the cost per unit of plot, and $X$ is the number of units per plot, then $V_x$ can be minimized for a given budget $B_o$. In this case,

$$B_o = n[C_1 + C_2X]$$

where $n$ is the number of plots, becomes a side condition.

In modern forest sampling problems, office expense can probably be taken as independent of the number, size, and pattern of field plots. In consequence, this investigation confined itself to the field component of the budget. We chose as a cost function:

$$C = C_1\sqrt{nA} + C_2n + C_3nTX$$

where

$C$ = the cost of survey
$C_1$ = cost of walking a unit distance, in crew-minutes
$n$ = total number of plots
$A$ = area to be sampled
$C_2$ = cost of plot establishment (identifying plot center, etc.)
$T$ = average number of elements per unit area
$X$ = plot "size"
$C_3$ = cost of measuring an element on the plot

In the case of point sampling, the third term of the budget equation becomes $C_3nT$, where $T$ is the average number of trees observed at a point. Magnitudes of $T$, corresponding to various basal-area factors, were counted and printed by computer. The following least-squares estimate of $T$ as a function of basal-area factors was then inserted into the third term of the budget equation:

$$T = 0.169436 + 53.8135\ (1/BAF),$$
$$\text{with } R^2 = 0.999$$

The reasoning behind the second and third term of the cost function is fairly obvious. In adopting the first term, the authors benefited from previous research work dealing with determination of travel distance through a number of points (see, for example, Hansen, Hurwitz, and Madow, 1953).

Numerical values for cost of walking a unit distance $C_1$, cost of establishment of a plot, $C_2$, and cost of measuring a tree in the plot, $C_3$, all expressed in crew-minutes (for a two-man crew) were based on data from Hornibrook (personal communication). Thus $C_1$ was estimated to be 32 crew-minutes per mile, $C_2$, 15 crew-minutes for plots and 12 crew-minutes for points, and $C_3$, 5 crew-minutes for measuring the diameter of a tree and 1 crew-minute for counting a tree. Although, in practice,

several measurements are usually taken at each random point, we have restricted ourselves to procedures designed to estimate one parameter. For discussion of the multiparameter problem, see Jagannathan (1965) and O'Regan and Arvanitis (1966).

Obviously, the formula is not universally applicable. Cost components may be added, changed, or dropped, depending on the purpose of the investigator. What is more important here is the fact that, given a reasonably good cost function and assuming that enough is known about the variance function, a relationship between variance, size of plots, number of plots, and budget would be secured for any given expenditure. (Tables could be used in lieu of formulas.)

In general, the effect of a cost function is not easy to see. A few large, but widely spaced plots decrease the total travel required, and increase the measuring time. On the other hand, and for the same degree of accuracy, many random plots of smaller size will cut down the travel time from plot center to plot center, but will increase the total travel time and the time required for plot establishment. Therefore, choosing the best combination of size and number of sampling units for a given budget is important in forest sampling design.

## ANALYTICAL PROCEDURE

A solution to the problem of sampling efficiency requires study of the relationship between input (number and size of plots) and output (the resulting variance or confidence-interval width of the estimate).

In production economics the mathematical relationship between output and variable inputs is termed the production function. Denoting the output by $Q$ and the $r$ variable inputs by $X_1$, $X_2, \ldots, X_r$, the production function is expressed as:

$$Q = f(X_1, X_2, \ldots, X_r)$$

In this study, the function takes the simple form:

$$Q = f(X_1, X_2)$$

The variable inputs are the number and size of the plots, and the output is a confidence-interval width.[3]

In production economics, an isoquant is defined as the locus of all combinations of variable inputs (in our case, plot size and number of plots) that yield a specified output level (confidence-interval width).

By setting

$$Q = Q_o = f(X_1, \ldots, X_r)$$

[3] Actually we have chosen:

$$W' = t\sigma/\sqrt{n}$$

where

$t$ = student's $t$
$\sigma^2$ = variance of plot measurements (dependent upon plot size)
$n$ = number of random points.
This is slightly different from the expected value of the usual confidence interval half-width:

$$E(W) = \frac{t}{\sqrt{n}} \sqrt{\sigma^2 - V(S_x)}$$

where $S_x$ is the estimated standard deviation of the plot measurement and $V(S_x)$ is the variance of the estimate. We could have used one of several measures of effectiveness. The choice of the half-width as a criterion also implies that the chooser can tolerate the random nature of the result. In a more sophisticated solution we could, of course, place probability limits on exceeding the criterion level.

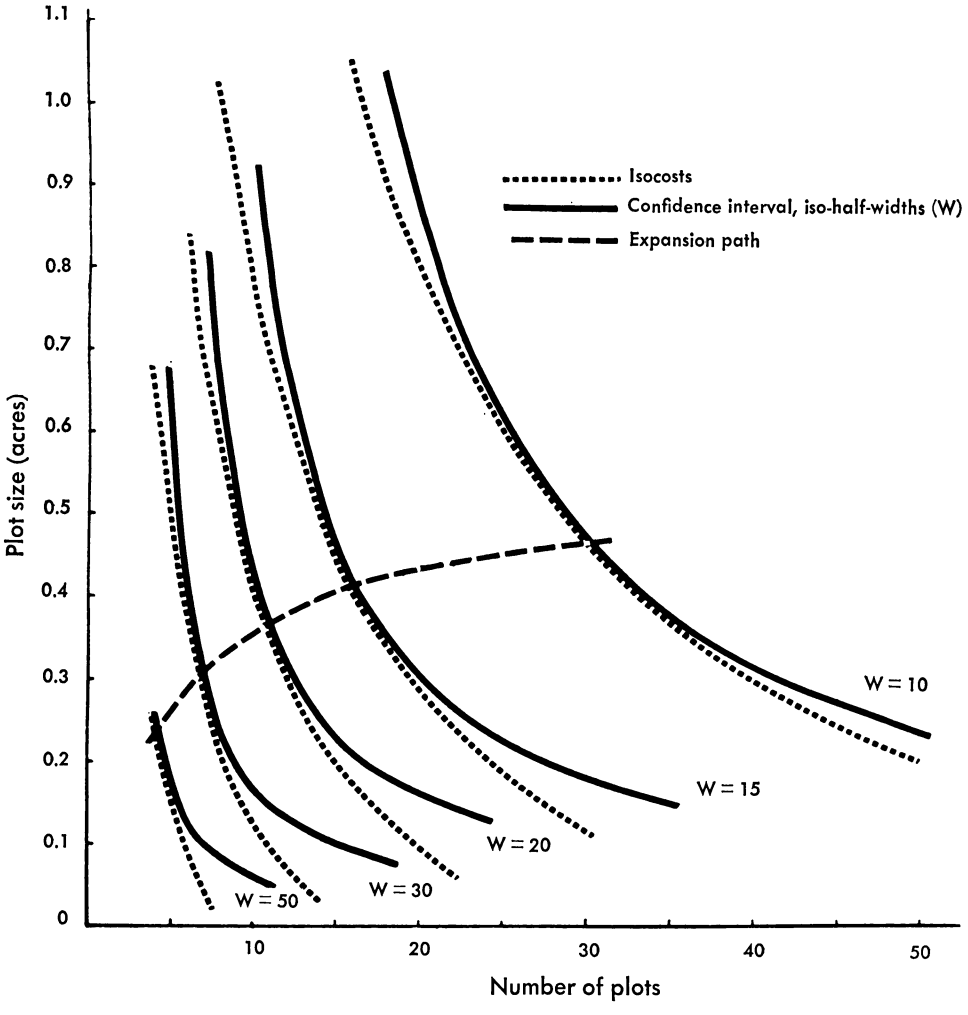



Fig. 2. Isocosts, confidence interval iso-half-widths for 0.95 confidence level, and expansion path of the estimate of basal area per acre from plot sampling.

we can solve for different sets of values $X_1, \ldots, X_r$ which combine to give the output $Q_o$. By setting different $Q$'s we could obtain a family of isoquants for any given production function (see figs. 2, 3).

In our case of two inputs, we have

$$Q_o = f(X_1, X_2),$$

and $X_2$ is obviously a function of $X_1$.

Taking the total differential

$$dQ_o = f_1 dX_1 + f_2 dX_2$$

where

$f_j$ = derivative of $f(X_1, X_2)$ with respect to $X_j$

$dX_j$ = differential of $X_j$

By the definition of isoquant,

$$dQ_o = 0;$$

therefore,

$$dX_2/dX_1 = -f_1/f_2$$

which gives the rate of change of $X_2$ on $X_1$ for the isoquant $Q_o$.

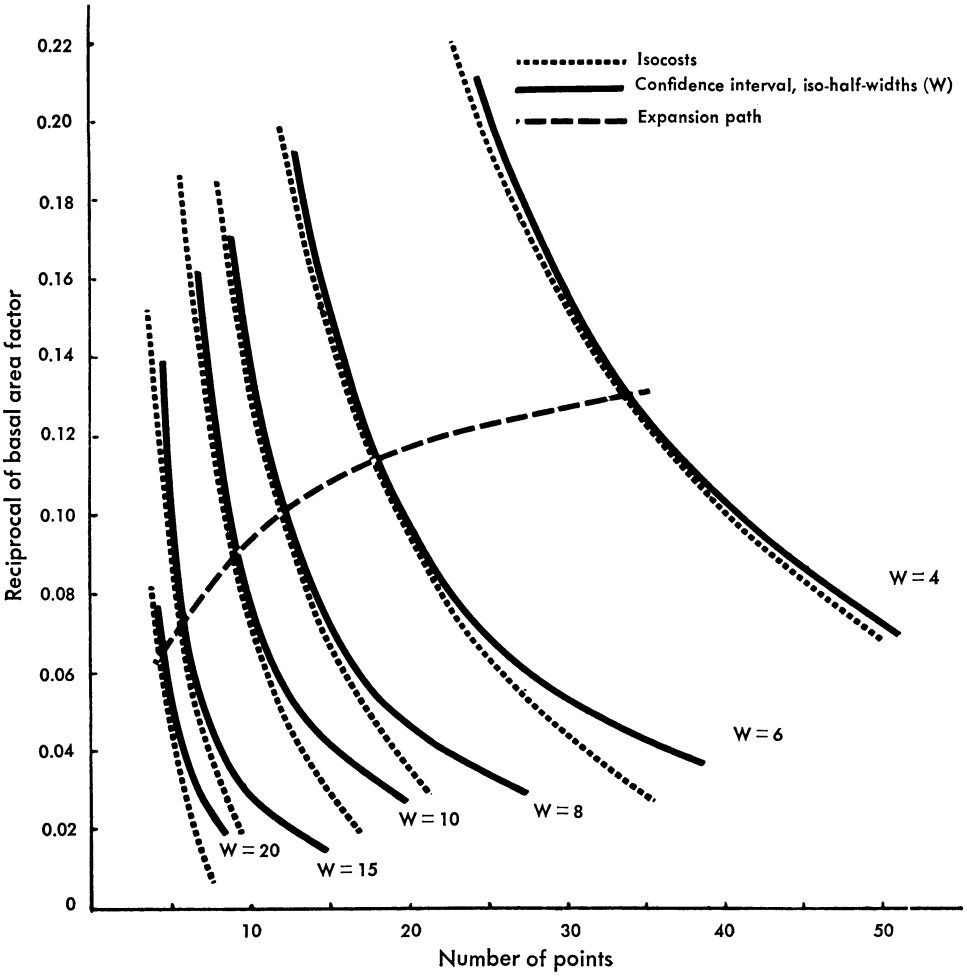



Fig. 3. Isocosts, confidence interval iso-half-widths for 0.95 confidence level, and expansion path of the estimate of number of trees per acre from point sampling.

According to Carlson (1956), this means that this slope "at any particular point is equal to the relationship between the marginal productivities of the services (inputs) with a minus sign."

The slope at a point on an isoquant is the rate at which one variable input must be substituted for the other (for example, number for size of plots) to maintain the corresponding output level (interval width) (Henderson and Quandt, 1958). This rate is called the "marginal rate of substitution" (Hitch, 1953).

Given a cost function:

$$B = g(X_1, X_2),$$

where $X_1$, $X_2$ are as above, a system of isocost lines can be derived. An isocost line (or curve) represents combinations of inputs obtainable at a given total cost.

The isocost expression

$$B_0 = g(X_1, X_2)$$

gives $X_2$ as a function of $X_1$. By differentiation as before, we find

$$dX_2/dX_1 = -g_1/g_2$$

where $g_j$ is the derivative of $g(X_1, X_2)$ with respect to $X_j$.

For linear cost functions

$$B = k + C_1X_1 + C_2X_2$$

". . . the isocosts are straight and parallel lines with negative slope proportional to $C_1$, $C_2$. . . ." (Carlson, 1956).

For nonlinear cost functions, the isocosts are convex or concave according to the sign (positive or negative) of the second cross derivative.

When these two systems of isoquant and isocost lines are combined, and the input along a given isoquant is varied, the minimum cost solution for a given output is found at the point where the given isoquant is tangent to the lowest possible isocost (Carlson, 1956). A given level of $X_1$ is optimally combined with $X_2$ at the point where

$$f_1/f_2 = g_1/g_2$$

If all minimum-cost points for different outputs are connected, a curve is obtained—the "expansion path." The most favorable input combination for any output is found along this line (figs. 2, 3).

In production economics, a criterion of maximum net revenue or minimum average cost determines the choice of a point, on the expansion path, that defines the amount to be produced, the quantity of each factor to be used, and the budget. In sampling, no such criterion is available. A point is determined by choice of the budget or the confidence-interval width.

In the present investigation, the cost function is not linear in the two variable inputs (size $X$ and number $n$ of the plots). The second (cross) derivative of the cost function is

$$dB/dX, dn = C_3T,$$

which is obviously positive, thus giving convex isocost curves.

Whenever convexity is present, a whole range of tangent points might exist between the isocost and isoquant lines. In such cases the expansion path may not be uniquely determined.

In addition to the above, we should keep in mind that nonsampling errors are always encountered in forest sampling work. These, unless properly controlled, can lessen the true meaning and importance of the resulting sampling information.

## A COMPUTER-SIMULATED PROBLEM

In the more rigorous sciences, mathematical models are used to explain and to study the phenomena of interest. When the variables in a system are not too numerous, when their interrelationships are reasonably simple and well known, and when the mathematician is sufficiently skillful, this procedure frequently is successful in achieving understanding, predictability, and control.

A forest, however, is a complex ecosystem involving numerous environmental factors acting in different directions and reacting in an unknown number of combinations and at various levels. This complexity usually creates difficulties in specifying variables and their relationships with sufficient exactness for use in mathematical solutions to forestry problems. In forest sample design problems in particular, the relationships between tree measurements, tree position, and sampling rules are not well enough known to allow the construction of a mathematical model of the variance-sampling rule relationship. A set of experiments could be carried out to allow inferences concerning this relationship, but such experiments would usually be quite costly. The authors felt that more information per dollar spent could be obtained by simulating the experiments.

In our study, a computer, provided with a stem map of the forest and with a sampling rule, identifies clusters of trees at randomly drawn points, and calculates measurements at those points and variances between point measurements.

In earlier work, Palley and O'Regan (1961) and O'Regan and Palley (1965) used a computer to identify clusters of trees, to approximate areas of regions (where a region is defined to be the set of all points that have a common cluster of associated trees), and to approximate the parameters of probability distributions of measurements associated with given sampling rules and forests. In their IBM 704 program, the area of the regions was approximated by placing a fine mesh (or grid) of points on the stand and counting the points that fell in each region. Such an approximation can be made as close as desired by controlling the number of points in the grid. Size of the grid openings is restricted to ensure that every tree has some non-zero probability of being measured.

From this earlier work we adopted and modified a program to allow us to obtain variances of the conceptual population discussed in the section "The Sampling Problem." We set the number of grid points at 8,712. For given plot sizes and measurements, the computer identified clusters of trees and probabilities, and calculated variances and covariances of the point measurements.

Defining, as we have,

$$P_i = N_i/N$$

automatically adjusts for edge. If, for example, the $k$th tree is near an edge, $N_k$ is smaller than it would be otherwise.

In the case of Bitterlich sampling, the trees were chosen on the basis of their diameter at breast height (d.b.h.) at the time of first measurement, for survivor and mortality trees, and on the basis of their d.b.h. at the time of second measurement, for ingrowth trees. This is one of several possible alternatives. The reader is asked to bear this choice in mind when evaluating our results.

Aside from certain limitations in storage capacity or existing computer units, the program can be modified to include any sampling rule. Increase in storage capacity and development of more accurate, rapid, and economical methods of constructing stem maps of large forests (by means of aerial photographs and electronic devices) would increase the efficiency of the computerized approach and would help to solve the difficult sampling problems currently facing the forest manager.

## INPUT-OUTPUT INFORMATION

The empirical results of this study are based on an analysis of data from the Blacks Mountain Experimental Forest, located within the Lassen National Forest, in northeastern California. The cover type is what is commonly called "eastside" pine, consisting of ponderosa pine (*Pinus ponderosa* Dougl.) and Jeffrey pine (*Pinus jeffreyi* Murr.) with a mixture of incense cedar (*Libocedrus decurrens* Torr.) and white fir (*Abies concolor* Lindley and Gordon). At the time of the first inventory (1933–1934), the percentage distribution (board-feet volume) by species was: pines, 90; white fir, 7; and incense cedar, 3.

The forest covers an area of approximately 10,000 acres, at elevations ranging from 5,600 to 6,800 feet. Experimental work related to management was begun in 1910, and was intensified following establishment of the forest as an experimental area in 1934. A num-

ber of permanent experimental units were established, and detailed tree and block records are kept. The primary objective was to develop new theories of management, silviculture, and insect control, resulting from early research, "... into a system of management and to test, demonstrate, and improve the system through continuous operation of a timber tract on a commercial scale" (Hallin, 1959).

Although the forest may not have been ideally suitable for the present investigation, it has the advantages of size, availability of stem maps, and a consistent, long-range procedure for measurement and recording. For purposes of the present study, measurements from several 20-acre experimental units (blocks) were analyzed. Only results from block 44 (2) are presented here.

This block, a rectangle $660 \times 1{,}320$ feet, was established in 1944. At that time, all trees larger than 11 inches d.b.h. were measured to the nearest 0.1 inch and recorded, along with certain qualitative characteristics (log grades, crown class, susceptibility, etc.). Each tree was given an identification number, and its relative location was marked on a fairly accurate stem map prepared on the spot. The block itself is part of a larger unit of 80 acres divided into four 20-acre blocks, all established at the same time but treated differently afterward.

Following establishment and first treatment, a certain number of trees were removed by logging, and the block was remeasured in 1946, 1949, and 1964. The last two measurements served as a basis for this study.

In 1949, 354 trees were present in the block. In 1964, 141 new trees (which had grown beyond the minimum diameter limit) were measured, and their position was marked on the stem map. For the 15-year period, 341 trees survived (present in both measurements), 13 died, and 141 were ingrowth (new) trees (present only at the last measurement, fig. 4).

For each tree in the block, a separate IBM card was prepared with the following entries: tree number, species code, rectangular coordinates, and d.b.h. and volume for 1949 and 1964. An indicator variable was also necessary to distinguish the ingrowth trees in point sampling. Data about tree height were not available although some inference could be drawn from the number of logs recorded. Tree number and d.b.h. were determined from the plot sheets. Board-feet volumes were estimated from local volume tables, and cubic-feet volumes were calculated from the board-feet measurements. The coordinates of each tree were measured to the nearest 0.1 foot by means of a special electronic chart reader (the Gerber Digital X, Y Reader) connected with an IBM key-punch machine. The cards were sorted by X coordinates, and cross-checked for punching and other errors before they were fed into the computer.

If necessary, additional input tree variables, such as basal area and growth (in number of trees, diameter, basal area, volume, etc.), both for totals and on a per unit area basis, can be calculated and punched or printed out by the computer on the basis of the initial input information.

The computer provided output information for eight forest parameters—number of trees, d.b.h., basal area, volume, and their corresponding increments. Results for four parameters are presented and discussed: (1) number of trees; (2) basal area; (3) ingrowth in number of trees; and (4) growth in basal area.

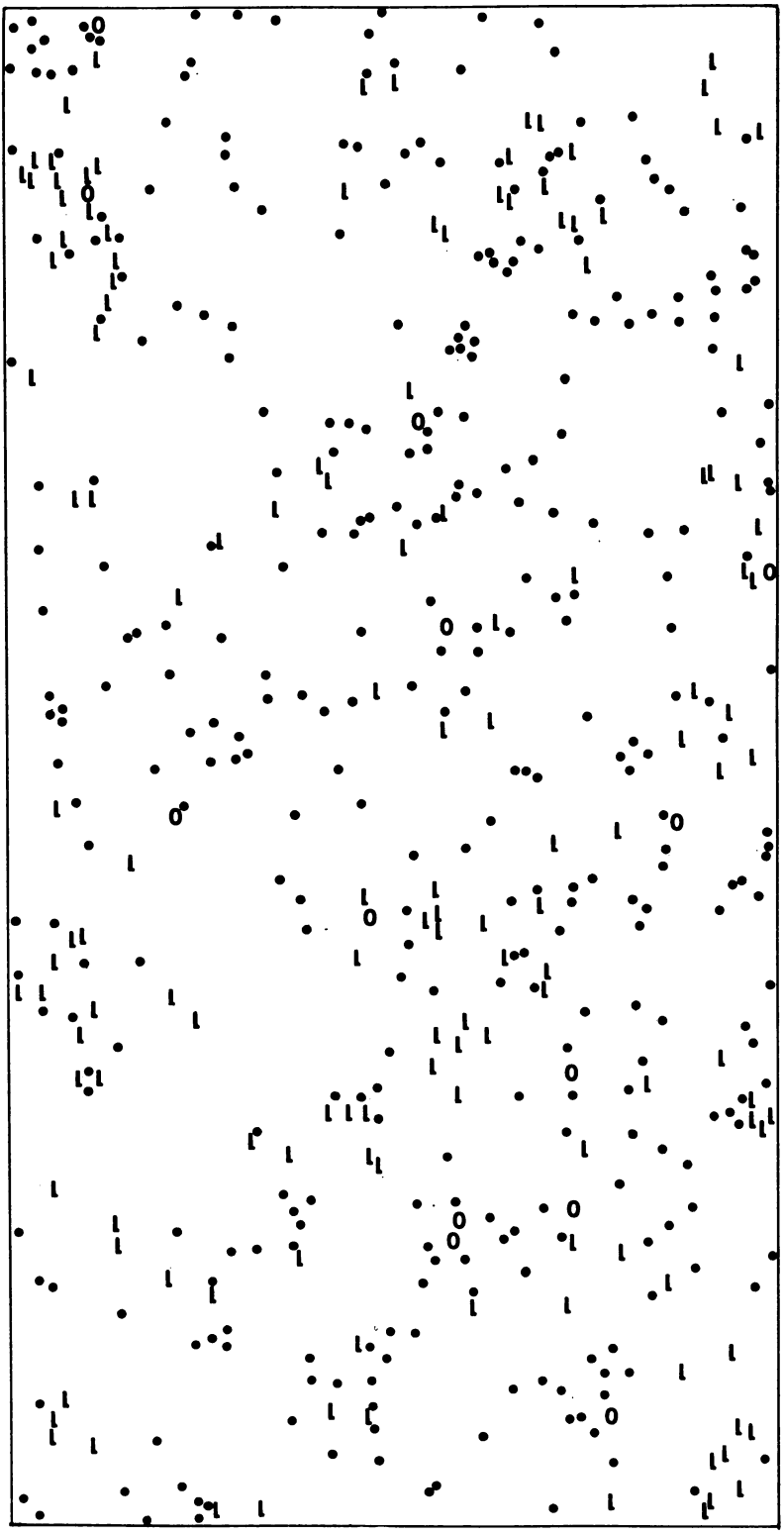


Fig. 4. Blacks Mountain Experimental Forest block 44(2). Spatial distribution of survivor (•), ingrowth (L), and mortality (0) trees for the period 1949–1964.

# RESULTS AND DISCUSSION

Results are cited and discussed for two sampling rules, fixed-area circular plots, and variable, or Bitterlich plots. In the literature, the terms "point sampling," "variable plot," and "Bitterlich plots" are used interchangeably to mean the same thing, namely, varying probabilities of tree selection in constructing the elementary sampling units. In the following discussion, "point sampling" will refer to Bitterlich procedures, in which the probability of a tree's being observed at a randomly drawn plot is a function of some measurement of the tree. "Plot sampling" will refer to the use of circular plots in which the probability of observation is constant for given plot size.

As anticipated, increase in plot size resulted in a decreased variance of per unit area measurements (tables 1, 2). This inversely proportional relationship between variance and plot size was more pronounced in the left third of the range of plot sizes than in the right third (fig. 5). Note that, in the case of point sampling, increase in plot size is achieved when the basal-area factor is reduced.

In general, the findings confirm Smith's variance function. The regression equations

$$\log_e V_x = \log_e V_1 + b \log_e X$$

where $X$ is the plot size, fit the data fairly well (tables 3, 4). However, the inclusion of a quadratic expression,

$$\log_e V_x = \log_e V_1 + b_1 \log_e X + b_2 (\log_e X)^2$$

resulted in an even better fit in most of the cases (tables 5, 6).

In order to compare the relative economic efficiency of the two sampling rules, it is necessary to consider the cost of the work performed. Since the cost of traveling from plot center to plot center is equal for point and plot sampling, the average number of trees included at a point or plot might be used as a measure of the effort expended at a location. The comparison then is made at points of equal expenditure of effort (O'Regan and Palley, 1965). Following this approach, the coefficient of variation, plotted over the average number of trees at a point, can serve as a measure of the relative economic efficiency in the use of the two rules (figs. 6, 7, 8, 9).

Under the conditions studied, circular plots had lower plot-to-plot variability for number of trees per unit area than did points (fig. 6). And point sampling showed smaller variability for basal

TABLE 1

COMPUTER APPROXIMATIONS TO VARIANCES, BY CIRCULAR PLOT SIZE, OF FOUR FOREST MEASUREMENTS

| Plot size (acres) | Variance of per acre forest measurements | | | |
|---|---|---|---|---|
| | Number of trees | Ingrowth (trees) | Basal area* | Growth (basal area)* |
| 0.1..... | 253.9 | 122.4 | 2,869 | 145.5 |
| 0.2..... | 138.9 | 76.2 | 1,469 | 85.5 |
| 0.3..... | 99.1 | 58.6 | 1,003 | 61.1 |
| 0.4..... | 78.9 | 45.3 | 790 | 47.2 |
| 0.7..... | 53.8 | 27.4 | 531 | 27.7 |
| 1.0..... | 43.2 | 18.9 | 417 | 19.4 |

* Basal-area measurements are in square feet.

TABLE 2

COMPUTER APPROXIMATIONS TO VARIANCES, BY BASAL-AREA FACTOR, OF FOUR FOREST MEASUREMENTS

| Basal-area factor | Variance of per acre forest measurements | | | |
|---|---|---|---|---|
| | Number of trees | Ingrowth (trees) | Basal area* | Growth (basal area)* |
| 5..... | 95.0 | 82.8 | 492 | 94.0 |
| 10..... | 159.0 | 131.6 | 775 | 153.5 |
| 15..... | 221.1 | 180.5 | 1,063 | 211.4 |
| 20..... | 277.3 | 219.6 | 1,334 | 261.4 |
| 25..... | 335.7 | 263.6 | 1,630 | 318.8 |
| 30..... | 394.8 | 306.0 | 1,925 | 371.3 |
| 40..... | ..... | ..... | 2,491 | 472.9 |

* Basal-area measurements are in square feet.

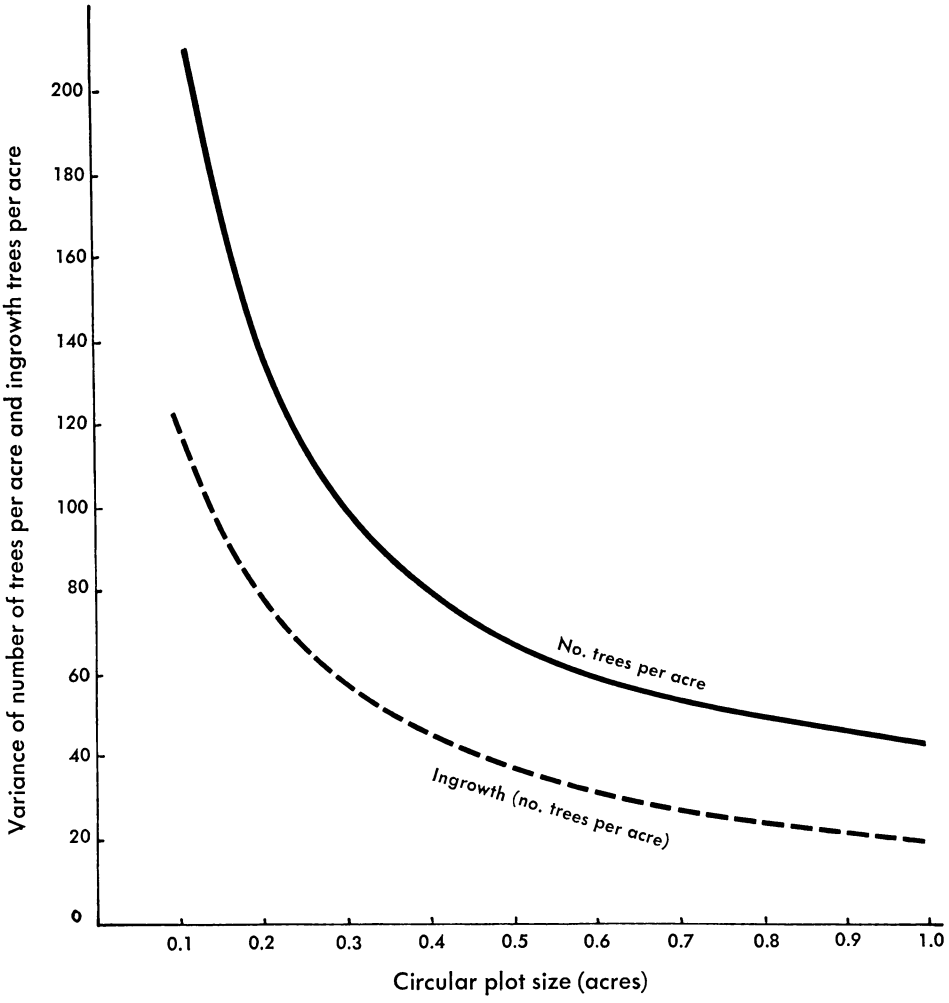



Fig. 5. Variance of measurements, on circular plots, of number of trees per acre and ingrowth in number of trees per acre.

area than did circular-plot sampling (fig. 7).

For both growth variables—number of trees and basal area— circular plots showed less variability when compared with points (figs. 8, 9). Measurement of ingrowth in number of trees is better, in terms of smaller variance, when plots are employed. Basal-area growth consists of two components—the change in diameters of existing trees during a certain time interval, and the basal area of the ingrowth trees. In forests similar to the one studied, growth in basal area is more highly correlated with number of trees than with basal area itself when estimates are based on point sampling. The correlation coefficients between basal area and basal-area growth ranged from 0.018 to 0.063, while the correlation between number of trees and basal-area growth was of the order 0.70 to 0.80. This may be attributed to the fact that, in forests of this kind, and for the basal-area growth that corresponds to the site and the stage of the stand devel-

TABLE 3

LINEAR REGRESSIONS OF LOGARITHM OF VARIANCE OVER LOGARITHM OF PLOT SIZE*

| Per acre measurements | $a$, intercepts for: | | $b$, regression coefficients for: | | $R^2$, coefficient of multiple determination for: | |
|---|---|---|---|---|---|---|
| | Plot sampling | Point sampling | Plot sampling | Point sampling | Plot sampling | Point sampling |
| No. of trees | 3.7118 | 3.2616 | −0.7699 | −0.7928 | 0.995 | 0.999 |
| Ingrowth (trees) | 3.1036 | 3.2252 | −0.8079 | −0.7286 | 0.993 | 0.999 |
| Basal area† | 5.9697 | 4.8874 | −0.8356 | −0.7821 | 0.993 | 0.995 |
| Growth (basal area)† | 3.0123 | 3.2625 | −0.8782 | −0.7779 | 0.997 | 0.998 |

* For plots: $\log_e V_x = a + b \log X$. For points: $\log_e V_{BAF} = a + b \log_e \left(\frac{1}{BAF}\right)$.

† Basal-area measurements are in square feet.

TABLE 4

LINEAR REGRESSIONS OF LOGARITHM OF COEFFICIENT OF VARIATION OVER LOGARITHM OF PLOT SIZE*

| Per acre measurement | $a$, intercepts for: | | $b$, regression coefficient for: | | $R^2$, coefficient of multiple determination for: | |
|---|---|---|---|---|---|---|
| | Plot sampling | Point sampling | Plot sampling | Point sampling | Plot sampling | Point sampling |
| No. of trees | −1.0176 | −1.2432 | −0.3848 | −0.3966 | 0.995 | 0.999 |
| Ingrowth (trees) | −0.4460 | −0.8709 | −0.4038 | −0.5388 | 0.993 | 0.957 |
| Basal area† | −1.0075 | −1.5471 | −0.4177 | −0.3908 | 0.993 | 0.995 |
| Growth (basal area)† | −1.0485 | −0.7578 | −0.4394 | −0.3890 | 0.997 | 0.998 |

* For plots: $\log_e CV_x = a + b \log X$. For points: $\log_e CV_{BAF} = a + b \log_e \left(\frac{1}{BAF}\right)$.

† Basal-area measurements are in square feet.

opment, the peripheral ring of the new growth added to existing trees becomes smaller and smaller with the increase of the diameter of the tree. Thus it does not appreciably affect the probability of selection in the case of point sampling. On the other hand, the newly qualified ingrowth trees are of relatively smaller size and, unless they happen to be close to the point center, are not taken into account, at least within the range of basal-area factors encountered in practice. Therefore, on the basis of the variance of point-to-point (or plot-to-plot) measurements, sampling schemes better suited for number of trees are also shown to be preferable for estimation of basal-area growth, when cost is measured by number of trees at a point.

It is customary, in practice, to estimate growth parameters from observations taken in sampling units considered to be of a size appropriate (in some sense) for estimation of other forest parameters. That this is not always a good decision is shown in table 7, where the same sampling rule and plot size for the forest studied yield estimates of ingrowth (as number of trees) with larger relative variability than estimates for number of trees. Likewise, in point sampling, and for the same basal-area factor, estimates of basal-area growth are less accurate than estimates of basal area itself.

The authors are aware of possible changes in the relative efficiency of plot and point sampling for growth estimates had trees smaller than 11 inches d.b.h. been included in the sample. The somewhat artificial truncation of small trees was unavoidable because of existing field measurements for this forest.

TABLE 5

QUADRATIC REGRESSIONS OF LOGARITHM OF VARIANCE OVER LOGARITHM OF CIRCULAR PLOT SIZE*

| Per acre measurement | $a$ | $b_1$ | $b_2$ | $R^2$ |
|---|---|---|---|---|
| Number of trees | 3.76249 | −0.6035 | 0.0733 | 0.999 |
| Ingrowth (trees) | 2.94807 | −1.0206 | −0.0937 | 0.999 |
| Basal area† | 6.03462 | −0.6185 | 0.0956 | 0.999 |
| Growth (basal area)† | 2.96383 | −1.0351 | −0.0693 | 0.999 |

* $\log_e V_x = a + b_1 \log_e X + b_2 (\log_e X)^2$.
† Basal-area measurements are in square feet.

TABLE 6

QUADRATIC REGRESSIONS OF LOGARITHM OF VARIANCE OVER LOGARITHM OF PLOT SIZE IN POINT SAMPLING*

| Per acre measurement | $a$ | $b_1$ | $b_2$ | $R^2$ |
|---|---|---|---|---|
| Number of trees | 3.52055 | −0.5715 | 0.0442 | 0.999 |
| Ingrowth (trees) | 3.48862 | −0.5036 | −0.0449 | 0.999 |
| Basal area† | 5.44138 | −0.3324 | 0.0851 | 0.999 |
| Growth (basal area)† | 3.56975 | −0.5285 | 0.0472 | 0.999 |

* $\log_e V_{BAF} = a + b_1 [\log_e \left(\frac{1}{BAF}\right)] + b_2 [\log_e \left(\frac{1}{BAF}\right)]^2$.
† Basal-area measurements are in square feet.

In this study, we were not primarily interested in the variability of plot measurements, but rather in a wise allocation of funds to travel time and measuring time and a wise choice of a sampling rule. Taking the half-width of a 0.95 confidence statement as a measure of production, we wanted to compare points with plots in the estimation of various parameters. When this width is plotted over the minimum budget required to obtain that half-width, another set of curves is obtained. These curves, for a given variance and cost function, illustrate the relative efficiency of fixed-area and variable plots for estimating a parameter. For example, figures 10, 11, and 12 show that number of trees, ingrowth (in number of trees), and basal-area growth were estimated more efficiently with fixed-area plots than with variable-area plots over the full range of budgets. The reverse appears to be true for the estimate of basal area (fig. 13) because point sampling yielded lower variance, and the cost of measurement was much lower than in the case of fixed-area plots. (Figures 10, 11, 12 and 13 are based on data from table 8.) However, we should not lose sight of the fact that the outcome for growth and variable-area plots is a function of the probability framework. Choice of another probability scheme might have led to a different result.

These findings indicate that each parameter requires a separate analysis in order to find an optimum combination of size and number of plots. In practice, a number of parameters are estimated with a given sample of points and trees. Optimum procedures in these cases are beyond the scope of this paper.

As indicated earlier, the whole block 44 (2) was subdivided into smaller forests for which separate regressions of variance over plot size were fitted by the least-squares method (table 9).

It is worth noticing that the shapes of the curves of the logarithm of variance over the logarithm of plot size, which correspond to various subdivisions of the forest, are almost identical (fig. 1, p. 139), and that the regression coefficients are centered on that of

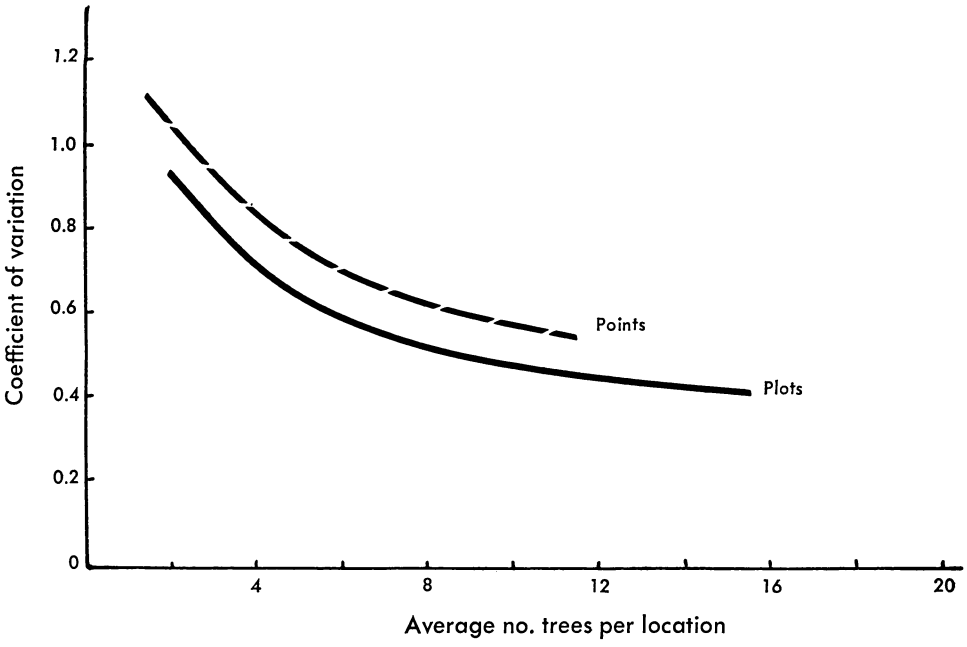



Fig. 6. Coefficient of variation of number of trees per acre versus average number of trees measured at a location. Plots compared with points.

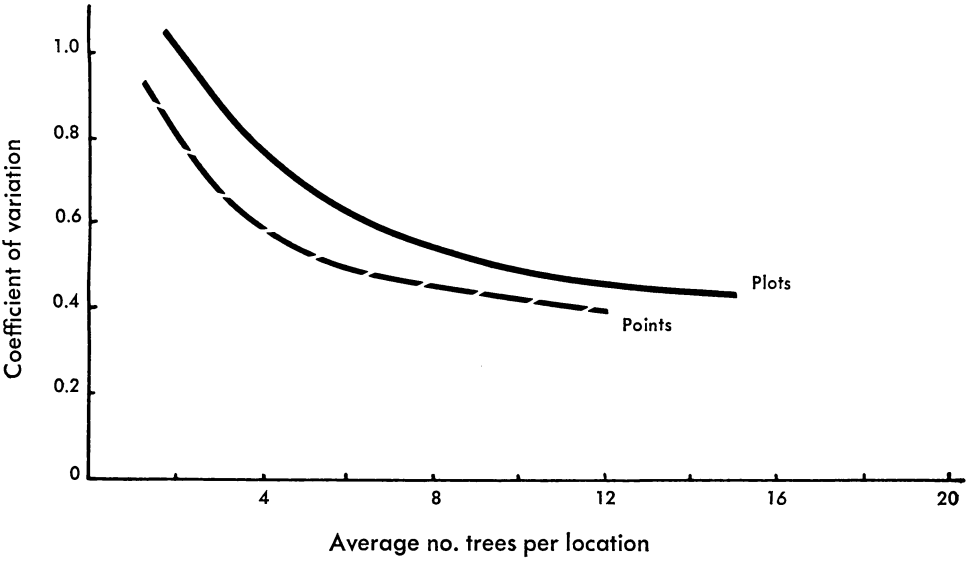



Fig. 7. Coefficient of variation of basal area per acre versus average number of trees measured at a location. Plots compared with points.

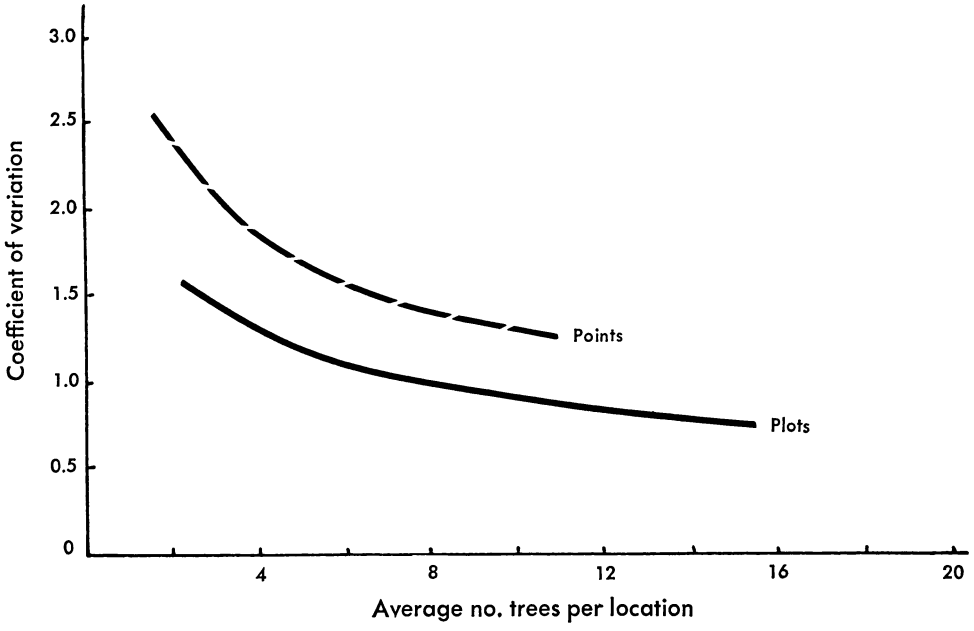



Fig. 8. Coefficient of variation of ingrowth in number of trees per acre versus average number of trees measured at a location. Plots compared with points.

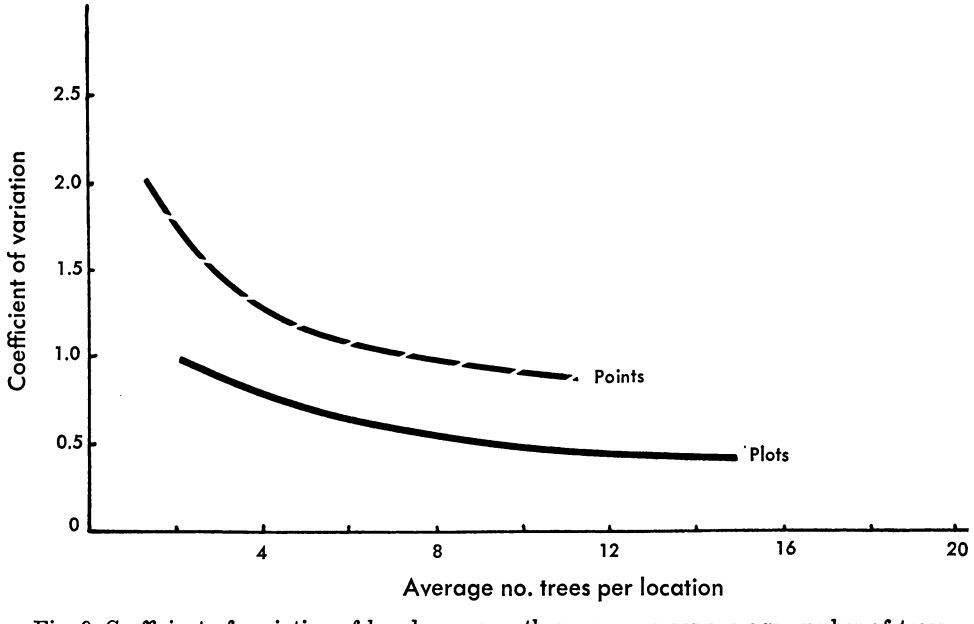



Fig. 9. Coefficient of variation of basal area growth per acre versus average number of trees measured at a location. Plots compared with points.

TABLE 7

COEFFICIENTS OF VARIATION FOR PLOT AND POINT SAMPLING

| Factor | Coefficient of variation for: | | | |
|---|---|---|---|---|
| | No. of trees | Ingrowth (trees) | Basal area | Growth (basal area) |
| Plot size (acres): | | | | |
| 0.1 | 0.9003 | 1.5691 | 0.9893 | 0.9384 |
| 0.2 | 0.6659 | 1.2380 | 0.7078 | 0.7191 |
| 0.3 | 0.5626 | 1.0691 | 0.5849 | 0.6079 |
| 0.4 | 0.5017 | 0.9546 | 0.5190 | 0.5347 |
| 0.7 | 0.4146 | 0.7427 | 0.4258 | 0.4090 |
| 1.0 | 0.3715 | 0.6167 | 0.3773 | 0.3424 |
| Basal area factor: | | | | |
| 5 | 0.5509 | 1.2910 | 0.4097 | 0.8892 |
| 10 | 0.7123 | 1.6275 | 0.5142 | 1.1366 |
| 15 | 0.8409 | 1.9056 | 0.6023 | 1.3337 |
| 20 | 0.9408 | 2.1020 | 0.6747 | 1.4832 |
| 25 | 1.0352 | 2.3030 | 0.7457 | 1.6379 |
| 30 | 1.1225 | 2.4812 | 0.8014 | 1.7675 |
| 40 | ..... | ..... | 0.9218 | 1.9948 |

TABLE 8

EXPANSION PATHS FOR ESTIMATING NUMBER OF TREES AND INGROWTH IN NUMBER OF TREES ON THE BASIS OF PLOT AND POINT SAMPLING

| Per acre measurement | Interval half-width | Number of plots for: | | Plot size for: | | Minimum budget for: | |
|---|---|---|---|---|---|---|---|
| | | Plot sampling | Point sampling | Plot sampling | Point sampling | Plot sampling | Point sampling |
| | | | | *acres* | *BAF* | *crew min.* | *crew min.* |
| No. of trees | 4 | 21.9 | 33.6 | 0.392 | 7.634 | 1,239 | 1,801 |
| | 6 | 12.1 | 17.7 | 0.341 | 8.627 | 650 | 914 |
| | 8 | 8.5 | 11.9 | 0.296 | 9.751 | 443 | 592 |
| | 10 | 6.6 | 9.1 | 0.267 | 11.022 | 337 | 436 |
| | 15 | 4.5 | 5.9 | 0.216 | 13.757 | 221 | 269 |
| | 20 | 3.6 | 4.5 | 0.183 | 15.829 | 173 | 202 |
| Ingrowth (trees) | 4 | 11.1 | 32.7 | 0.519 | 9.305 | 783 | 1,549 |
| | 6 | 8.3 | 17.1 | 0.316 | 10.440 | 449 | 793 |
| | 8 | 6.6 | 11.4 | 0.232 | 11.666 | 317 | 517 |
| | 10 | 5.5 | 8.6 | 0.186 | 12.886 | 248 | 384 |
| | 15 | 4.1 | 5.6 | 0.121 | 16.176 | 170 | 241 |
| | 20 | 3.3 | 4.3 | 0.099 | 18.782 | 136 | 183 |
| Basal area* | 6 | 73.8 | 77.7 | 0.502 | 8.711 | 4,660 | 3,680 |
| | 10 | 29.6 | 30.7 | 0.468 | 9.213 | 1,844 | 1,468 |
| | 15 | 15.7 | 16.0 | 0.417 | 10.180 | 942 | 756 |
| | 20 | 10.7 | 10.7 | 0.368 | 11.302 | 614 | 497 |
| | 30 | 6.7 | 6.6 | 0.304 | 13.507 | 364 | 298 |
| | 50 | 4.2 | 4.1 | 0.233 | 17.580 | 215 | 180 |
| Growth (basal area)* | 4 | 10.0 | 34.8 | 0.622 | 8.536 | 801 | 1,759 |
| | 6 | 7.5 | 18.1 | 0.400 | 9.304 | 466 | 892 |
| | 8 | 6.1 | 12.1 | 0.297 | 10.485 | 331 | 577 |
| | 10 | 5.2 | 9.1 | 0.239 | 11.570 | 261 | 425 |
| | 15 | 3.9 | 5.9 | 0.168 | 14.536 | 180 | 263 |
| | 20 | 3.2 | 4.5 | 0.138 | 16.790 | 144 | 198 |

* Basal-area measurements are in square feet.

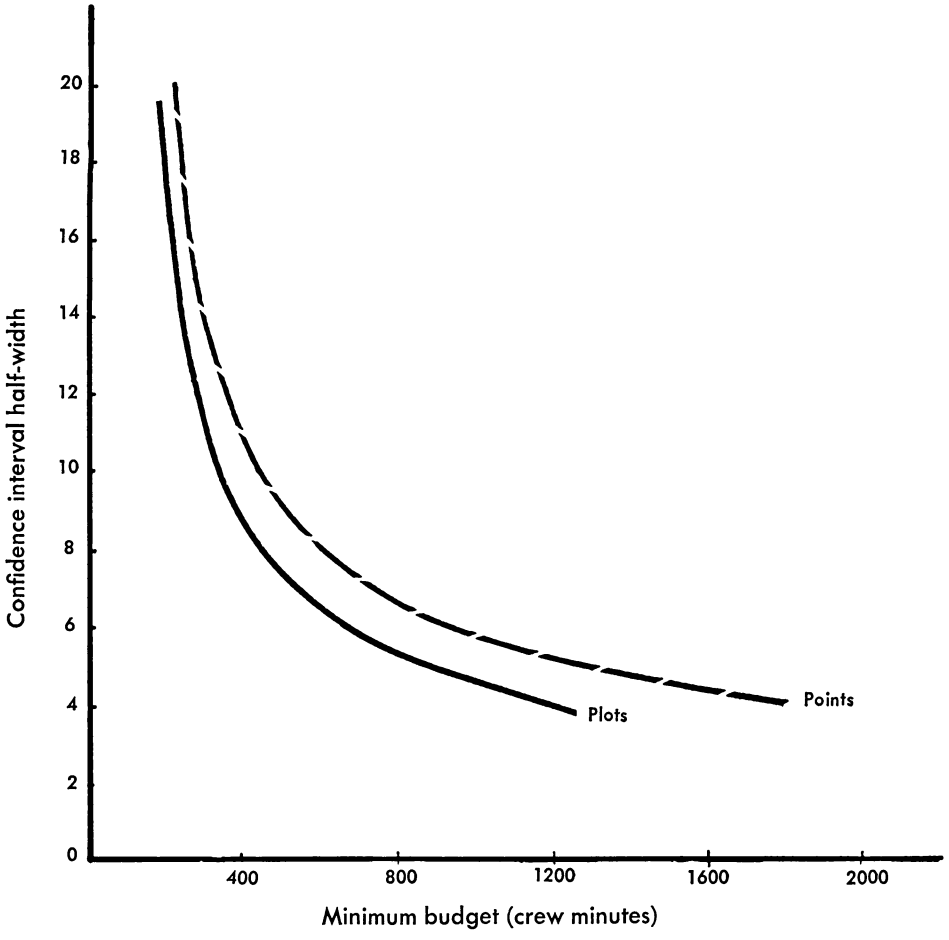



Fig. 10. Confidence interval half-width (for confidence level 0.95) over minimum budget corresponding to the best combination of number and "size" of plots for estimating number of trees per acre.

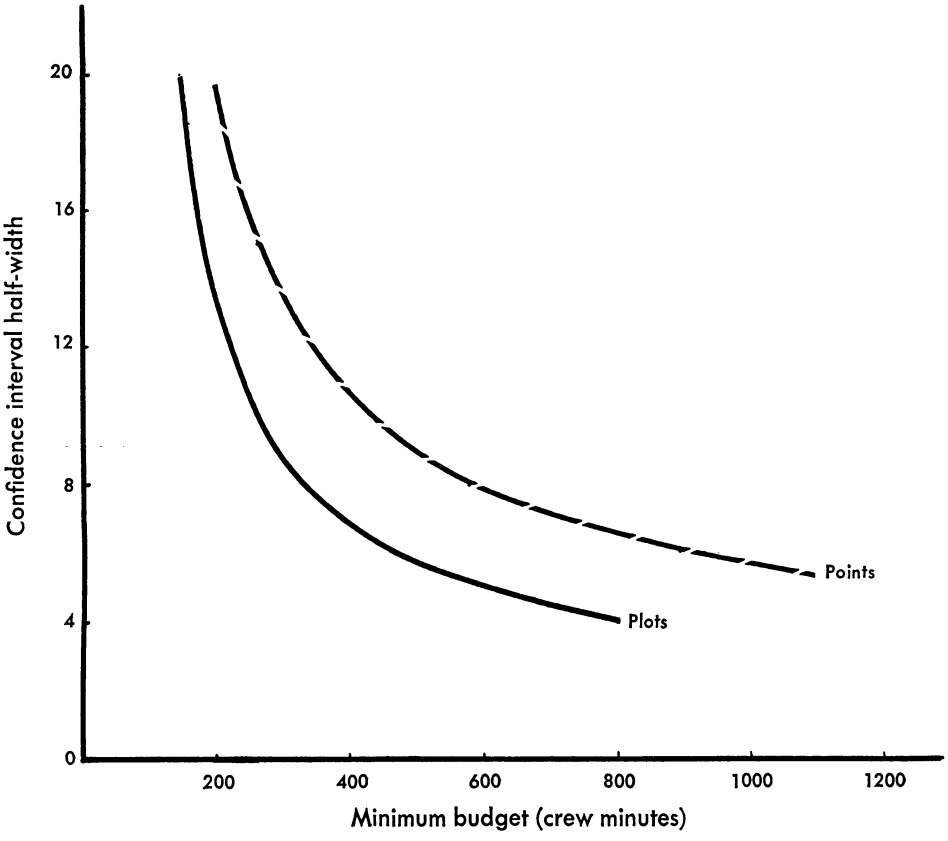



Fig. 11. Confidence interval half-width (for 0.95 confidence level) over minimum budget corresponding to the best combination of number and "size" of plots for estimating ingrowth in number of trees per acre.

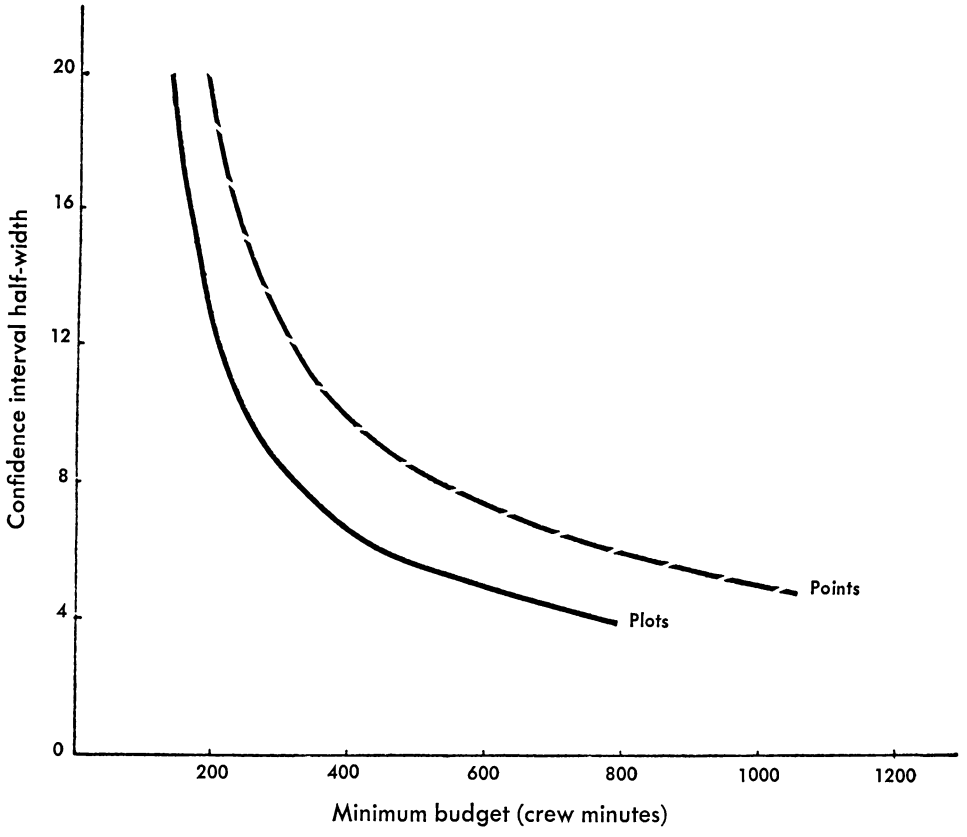



Fig. 12. Confidence interval half-width (for 0.95 confidence level) over minimum budget corresponding to the best combination of number and "size" of plots for estimating basal-area growth per acre.

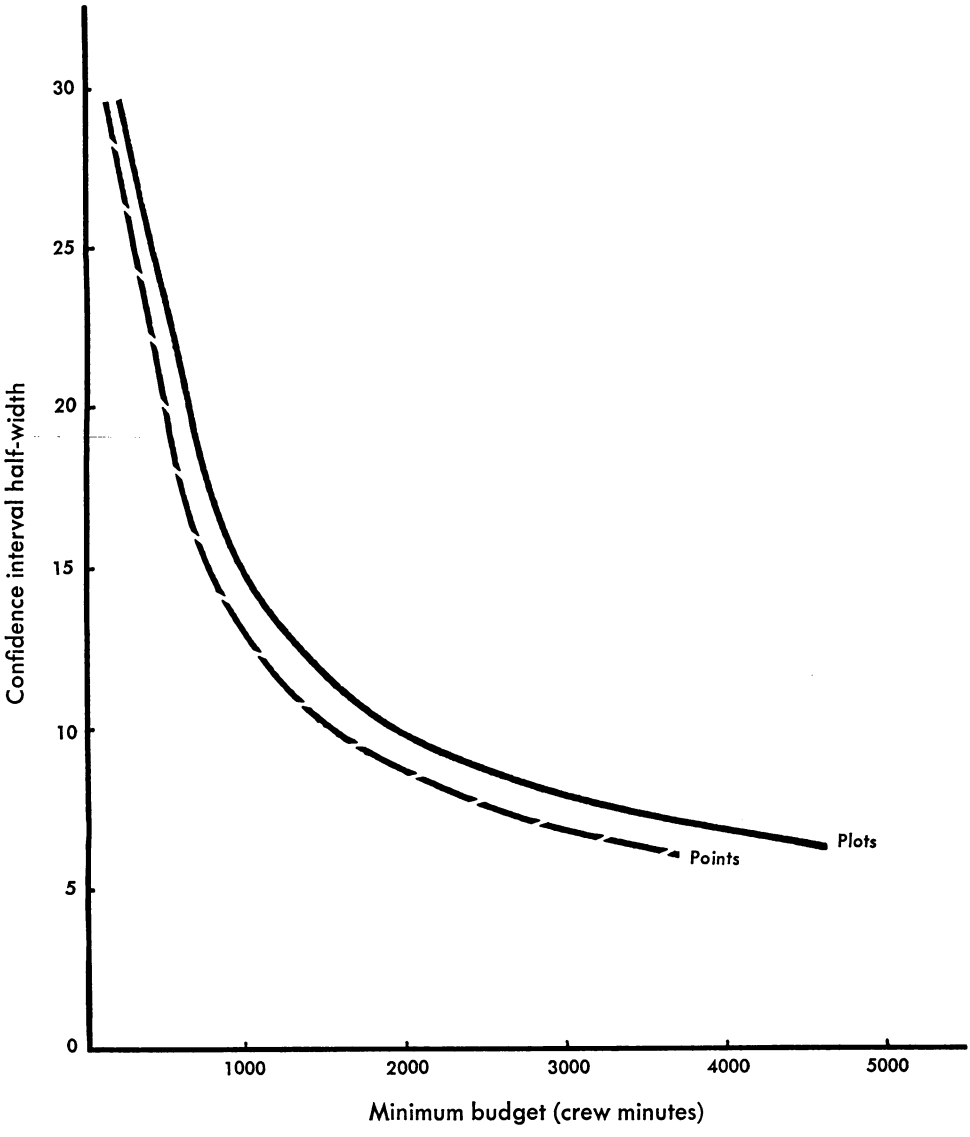



Fig. 13. Confidence interval half-width (for 0.95 confidence level) over minimum budget corresponding to the best combination of number and "size" of plots for estimating basal-area per acre.

TABLE 9

LINEAR REGRESSIONS OF VARIANCE OF BASAL AREA PER ACRE OVER PLOT SIZE FOR SUBDIVISIONS OF BLOCK 44(2), BLACKS MOUNTAIN EXPERIMENTAL FOREST

| Subdivision* | Size | *a*, intercept | *b*, regression coefficient | $R^2$ |
|---|---|---|---|---|
| | *acres* | | | |
| 1 | 5 | 5.8799 | −0.8532 | 0.98 |
| 2 | 5 | 5.7651 | −0.7023 | 0.98 |
| 3 | 5 | 5.9831 | −0.8702 | 0.99 |
| 4 | 5 | 6.1976 | −0.9219 | 0.99 |
| L | 10 | 5.8407 | −0.7878 | 0.99 |
| U | 10 | 5.9080 | −0.9225 | 0.99 |
| F | 20 | 5.9697 | −0.8356 | 0.99 |

* 1, 2, 3, 4 = quarters (5 acres each).
L, U = lower and upper halves (10 acres each).
F = whole forest (total, 20 acres).

variance over plot size for the whole forest. This might be the one further indication of what Smith (1938) has found—that when the regression coefficient is smaller than −0.7, the correction needed for infinite fields is not very important over a certain range of the ratio of plot size to the size of the field. This point, which needs further investigation, is of particular interest for populations covering extensive areas, as in forestry, where inferences are made from a portion of the population to the whole.

For optimum solutions among number of plots, plot size, and budget, corresponding to each subdivision of the forest studied, see table 10.

TABLE 10

OPTIMUM SOLUTIONS FOR ESTIMATION OF BASAL AREA PER ACRE FOR VARIOUS SUBDIVISIONS OF BLOCK 44(2), BLACKS MOUNTAIN EXPERIMENTAL FOREST

| Interval half-width (sq. ft./acre) | Subdivision* | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | L | U | F |
| | Number of plots (rounded) | | | | | | |
| 6 | 69 | 67 | 57 | 78 | 70 | 50 | 74 |
| 10 | 28 | 27 | 25 | 32 | 28 | 22 | 30 |
| 15 | 15 | 14 | 15 | 17 | 15 | 13 | 16 |
| 20 | 10 | 9 | 10 | 12 | 10 | 9 | 11 |
| 30 | 6 | 6 | 7 | 7 | 6 | 6 | 7 |
| 50 | 4 | 4 | 4 | 5 | 4 | 4 | 4 |
| | Plot size (acres) | | | | | | |
| 6 | 0.488 | 0.353 | 0.744 | 0.669 | 0.429 | 0.843 | 0.502 |
| 10 | 0.458 | 0.337 | 0.627 | 0.611 | 0.403 | 0.710 | 0.468 |
| 15 | 0.412 | 0.309 | 0.509 | 0.534 | 0.363 | 0.580 | 0.417 |
| 20 | 0.368 | 0.279 | 0.427 | 0.468 | 0.328 | 0.487 | 0.368 |
| 30 | 0.307 | 0.235 | 0.332 | 0.376 | 0.276 | 0.379 | 0.304 |
| 50 | 0.237 | 0.189 | 0.236 | 0.285 | 0.216 | 0.269 | 0.233 |
| | Minimum budget (crew minutes) | | | | | | |
| 6 | 4,290 | 3,371 | 4,875 | 6,073 | 3,954 | 4,669 | 4,660 |
| 10 | 1,707 | 1,357 | 1,944 | 2,377 | 1,578 | 1,880 | 1,844 |
| 15 | 879 | 708 | 1,000 | 1,197 | 815 | 978 | 942 |
| 20 | 577 | 471 | 653 | 769 | 537 | 645 | 614 |
| 30 | 346 | 289 | 386 | 445 | 324 | 385 | 364 |
| 50 | 208 | 179 | 225 | 254 | 196 | 229 | 215 |

* 1, 2, 3, 4 = quarters (5 acres each).
L, U = lower and upper halves (10 acres each).
F = whole forest (total, 20 acres).

## SUMMARY

The data indicate that although Smith's (1938) variance function of linear relationship between the logarithm of variance and the logarithm of plot size holds, the addition of a quadratic term to the variance equation resulted in an even better fit.

One explanation of the curvilinear form of the variance law could be the existing intraplot correlation between measurements on the plot and measurements on increments (or decrements) of the plot size. Another could be the shape (form) of the distribution of the measurements themselves, from which the quadratic expression of the variance is deducible.

On the basis of the cost of measuring a number of trees at a point, and for the forest studied, a better estimate (in terms of smaller variance) for number of trees, ingrowth (as number of trees), and growth in basal area was obtained from circular plots, while point sampling yielded a better estimate of basal area.

On the basis of optimum combinations of variance and cost functions, plots also proved better for estimating number of trees, ingrowth (as number of trees), and basal-area growth, while points were superior for estimating basal area.

The usefulness of the values corresponding to the optimum solution (plot size, number of plots, minimum budget) depends on how far the variance and cost functions being used deviate from reality. However, given any pair of variance and cost functions, an optimum solution can be found by following the method adopted in the present investigation.

For the forest studied, the variance function was not materially affected by the size of the forest. Differences in minimum budgets were more pronounced in the smaller confidence-interval half-widths than in the larger ones.

The diversity of recommendations concerning sampling optimums for estimation of various forest parameters could be attributed to differences in the size and spatial distribution of the trees on the ground, the sample rule employed, the field arrangement of the sampling units, and the cost functions, when used. Thus generalizations should be avoided unless there is enough evidence to justify them.

## SUGGESTIONS FOR FURTHER RESEARCH

Results presented in this study were derived, essentially, from one budget equation. Other budget equations were considered, however—for example, the inclusion of a cost component for preparatory work in the field, the cost of moving the camp after a certain number of plots were measured (or a given area in square miles was covered by sampling), and the like. This problem should be studied further in order to approximate reality as closely as possible and to provide the necessary information for more efficient sampling designs.

There remains also the problem of finding the shortest path through a number of points included in the area to be sampled. Matern (1960) pointed out some aspects of the problem and worked out a table for the "average length per sample point of a path passing through all sample points in a plane region." The same problem was treated

by several other investigators (Verblunsky, 1951; Few, 1955; Danzig, 1963; Kaufman, 1963).

Another problem that needs further treatment is the effect of the forest edge on the variability of plot measurements. Edge effect has two aspects—one ecological and one statistical. The ecological effect results from the fact that, for favorable edges, border trees usually benefit more than the inner ones from abundance of light, space, and nutrients, and thus produce more per unit area. The opposite is of course true for trees along edges with unfavorable orientation for growth.

The statistical aspect of the edge effect is due to the fact that edge affects the probability framework by changing the probabilities associated with trees close to the edge. If this change in probability is neglected, a bias will be introduced into the estimation procedure. The nature of the ecological edge and the relative extent of the edge affect the seriousness of the bias. In the present study, it was possible to obtain unbiased estimates by our choice of sampling procedures. More than one procedure is available for avoiding bias due to edge effect. There is still some argument among investigators as to the effect of these procedures on plot-to-plot variability. Several research workers have dealt with it. (See, for example, Finney and Palca, 1949; Masuyama, 1953, 1954; Grosenbaugh, 1958; Haga and Maezawa, 1959; Zarkovic, 1960; Barrett, 1965.)

The study of "intraplot correlation" (where the term is defined as the correlation between measurements on the plot and measurements on increments to the plot as the plot size changes) is of special interest in forest sampling. When clusters of elements are the sampling units, the intraplot correlation will influence the sampling error. Reduction in variability due to increase of plot size is greater when adjacent units are not correlated. This was realized by early research workers (Harris, 1915: Smith, 1938), and several methods of estimating the magnitude and effect of intraplot correlation were proposed. Hansen and Hurwitz (1942) pointed out that the sign (positive or negative) of the intraplot correlation could serve as a guide for choosing single elements or clusters of elements as elementary sampling units. A computer approach now seems appropriate for study of the whole problem of intraplot correlation and its impact on sampling designs.

The problem of variance behavior over large forests, as it could be inferred from smaller ones, is of both theoretical and practical interest. Results from previous and similar research should be expanded with the aid of computers and the development of fast, accurate, and economical methods of constructing stem maps (possibly from aerial photographs and electronic devices).

Finally, the use of mathematical programming as a means of finding the best combination of size and number of sampling units in multiparametric cases also deserves additional study.

## ACKNOWLEDGMENTS

The authors wish to thank the School of Forestry, University of California, Berkeley, the Pacific Southwest Forest and Range Experiment Station, U. S. Forest Service, and the California Division of Forestry for financial and other assistance. Appreciation is extended to the University Computer Center, Berkeley, for generously providing free use of its facilities.

# LITERATURE CITED

BARRETT, J. R.
1965. Correction for edge effect bias in point sampling. Forest Sci. 10(1):52–55.

BICKFORD, C. A.
1962. Methods of measuring the growth of trees as individuals and in stands. *In:* Tree growth. New York: The Roland Press Co. 452 pp.

BITTERLICH, W.
1948. Die Winkelzahlprobe (The angle count sample). Allg. Forst. Holzwirts. Ztg. 59:4–5.

BOON, D. A.
1962. Plot size and variability. I. C. T. Information Series B, No. 17. Delft, The Netherlands.

BORMANN, F. H.
1953. The statistical efficiency of sample plot size and shape in forest ecology. Ecology 34: 474.

CARLSON, S.
1956. A study on the pure theory of production. (2nd. ed.) New York: Kelley and Millan, Inc. 128 pp.

CHRISTIDIS, B. G.
1931. The importance of the shape of plots in field experimentation. Jour. Agr. Sci. 21: 14–37.

COCHRAN, W. G.
1963. Sampling techniques. (2nd. ed.) New York: John Wiley and Sons, Inc. 413 pp.

CORNFIELD, J.
1951. Modern methods in the sampling of human populations. The determination of sample size. Amer. Jour. Public Health 41(6): 654–61.

DANZIG, G. B.
1963. Linear programming and extensions. Princeton, New Jersey: Princeton University Press. 625 pp.

EVANS, T. C., and W. G. O'REGAN
1963. Sampling problems in measurement of range vegetation. Range Res. Methods, U. S. Dept. Agr., Forest Serv., Misc. Publ. 940: 54–60.

FEW, L.
1955. The shortest path and the shortest road through N points. Mathematika 2(4): 141–44.

FINNEY, D. J., and H. PALCA
1949. The elimination of bias due to edge effects in forest sampling. Forestry 23: 31–37.

FREESE, F.
1961. Relation of plot size to variability. An approximation. Jour. Forestry 59(9): 679.

GROSENBAUGH, L. R.
1958. Point sampling and line sampling. Probability theory, geometric implications, synthesis. U. S. Forest Serv., Southern Forest Exp. Sta., Occas. Paper No. 160, 33 pp.
1963. Some suggestions for better sample tree measurement. Proc. Soc. Amer. Foresters: 36–42.

GROSENBAUGH, L. R., and W. F. STOVER
1957. Point sampling compared with plot sampling in southeast Texas. Forest Sci. 3(1): 2–14.

HAGA, T., and K. MAEZAWA
1959. Bias due to edge effect in using the Bitterlich method. Forest Sci. 5: 370–76.

HALL, O. F.
1959. The contribution of remeasured sample plots to the precision of growth estimates. Jour. Forestry 57(1): 807–11.

HALLIN, W. E.
1959. The application of unit area control in the management of ponderosa-Jeffrey pine at Blacks Mountain Experimental Forest. U. S. Dept. Agr. Tech. Bul. 1191, 96 pp.

HANSEN, M. H., and W. N. HURWITZ
1942. Relative efficiencies of various sampling units in population inquiries. Jour. Amer. Stat. Assoc. 37: 89–94.

HANSEN, M. H., W. N. HURWITZ, and W. G. MADOW
1953. Sample survey methods and theory (vols. 1–2). New York: John Wiley and Sons.

HARRIS, J. A.
1915. On a criterion of substratum homogeneity (or heterogeneity) in field experiments. Amer. Nat. **49**: 430–54.
1920. Practical universality of field heterogeneity as a factor influencing plot yields. Jour. Agr. Res. **19**: 279–314.

HATHEWAY, W. H., and E. J. WILLIAMS
1958. Efficient estimation of the relationship between plot size and variability of crop yields. Biometrics **14**(2): 207–22.

HENDERSON, J. M., and R. E. QUANDT
1958. Microeconomic theory, a mathematical approach. New York: The McGraw-Hill Book Co. 291 pp.

HITCH, C.
1953. Sub-optimization in operations problems. Jour. Oper. Res. Soc. Amer. **1**(3): 87–99.

JAGANNATHAN, R.
1965. The programming approach in multiple character studies. Econometrica **33**(1): 236–37.

KAUFMAN, A.
1963. Methods and models of operations research. New York: Prentice-Hall, Inc. 510 pp.

KOCH, E. J., and J. A. RIGNEY
1951. A method of estimating optimum plot size from experimental data. Agron. Jour. **43**:17–21.

MAHALANOBIS, P. C.
1940. A sample survey of the acreage under jute in Bengal. Sankhya **4**: 511–30.
1944. On large scale sample surveys. Philosophical Transactions **231**(584) Series B: 329–451.
1946. Use of small size plots in sample surveys of crop yields. Nature **158**: 798.

MASUYAMA, M.
1953. A rapid method of estimating basal area in timber survey. An application of integral geometry to aerial sampling problems. Sankhya **12**(3): 291–302.
1954. On the error in crop cutting experiment due to the bias on the border of the grid. Sankhya **14**: 181–86.

MATERN, B.
1960. Spatial variation. Stochastic models and their application to some problems in forest surveys and other sampling investigations. Meddelanden Fran, Statens Skogsforskningsinstitut, Band **49**(5): 144 pp.

MESAVAGE, C., and L. R. GROSENBAUGH
1956. Efficiency of several cruising designs on small tracts in north Arkansas. Jour. Forestry **54**(9): 569–76.

NEYMAN, J.
1934. On the two different aspects of the representative method. The method of stratified sampling and the method of purposive selection. Jour. Royal Stat. Soc. **97**: 558–625.

O'REGAN, W. G., and M. N. PALLEY
1965. A computer technique for the study of forest sampling methods. Forest Sci. **11**(1): 99–114.

O'REGAN, W. G., and L. G. ARVANITIS
1966. Cost effectiveness in forest sampling. Forest Sci. **12**(4):406–14.

PALLEY, M. N., and L. G. HORWITZ
1961. Properties of some random and systematic point sampling estimators. Forest Sci. **7**(1): 52–65.

PALLEY, M. N., and W. G. O'REGAN
1961. A computer technique for the study of forest sampling methods. 1. Point sampling compared with line sampling. Forest Sci. **7**(3):282–94.

PRODAN, M.
1961. Forstliche Biometrie. München, Bonn, Vien: Blv. Verlagsgesellschaft. 432 pp.

SHIUE, C.
1962. Measuring and predicting growth for all-aged stands. *In:* Tree growth. New York: The Ronald Press Co. 452 pp.

SMITH, F. H.
1938. An empirical law describing heterogeneity in the yields of agricultural crops. Jour. Agr. Sci. **28**: 1–23.

STRAND, L.
1957. The effect of plot size on the accuracy of forest surveys. (In Norwegian with English summary.) Norweg. Forest Res. Inst. Rep. **14**: 621.
1959. Noyaktigheten ved noen Metoder til Bestemmelse av Kubikk-og Tilvekstmassen pa Proferlater. (Accuracy of some methods for estimating volume and increment on sample plots.) (In Norwegian with English summary.) Saertrykk Av, Meddelelser fra det Norske Skogforsoksvesen **52**: 392.

SUKHATME, P. V.
1946. Bias in the use of small-size plots in sample surveys for yield. Nature **157**: 630.
1947*a*. Use of small plots in yield surveys. Nature **160**: 542.
1947*b*. The problem of plot size in large-scale yield surveys. Jour. Amer. Stat. Assoc. **42**: 297–310.

SUKHATME, P. V., and G. R. SETH
1952. Non-sampling errors in surveys. Jour. Indian Soc. Agr. Stat. **4**: 5–41.

VERBLUNSKY, S.
1951. On the shortest path through a number of points. Proc. Amer. Math. Soc. **2**(6): 904–13.

WHITTLE, P.
1956. On the variation of yield variance with plot size. Biometrika **43**: 337–42.

WRIGHT, J. W., and F. D. FREELAND
1960. Plot size and experimental efficiency in forest genetic research. Michigan State Univ. Tech. Bul. **280**: 28.

YATES, F.
1935. Some examples of biased sampling. Annals of Eugenics **6**: 202–13.

ZARKOVIC, S. S.
1960. On the efficiency of sampling with varying probabilities and the selection of units with replacement. Metrika **3**: 53–60.

To simplify this information, it is sometimes necessary to use trade names of products or equipment. No endorsement of named products is intended nor is criticism implied of similar products not mentioned.

4*m*-3,'67 (G9035) LL

**The journal HILGARDIA is published at irregular intervals, in volumes of about 650 to 700 pages. The number of issues per volume varies.**

**Single copies of any issue may be obtained free, as long as the supply lasts; please request by volume and issue number from:**

**Agricultural Publications**
**University Hall**
**University of California**
**Berkeley, California 94720**

**The limit to nonresidents of California is 10 separate titles. The limit to California residents is 20 separate titles.**

**The journal will be sent regularly to libraries, schools, or institutions in one of the following ways:**

1. **In exchange for similar published material on research.**
2. **As a gift to qualified repository libraries only.**
3. **On a subscription basis—$7.50 a year paid in advance. All subscriptions will be started with the first number issued during a calendar year. Subscribers starting during any given year will be sent back numbers to the first of that year and will be billed for the ensuing year the following January. Make checks or money orders payable to The Regents of The University of California; send payment with order to Agricultural Publications at above address.**