# DETERMINATION OF THE ERRORS OF ESTIMATE OF A FOREST SURVEY, WITH SPECIAL REFERENCE TO THE BOTTOM-LAND HARDWOOD FOREST REGION[1]

By FRANCIS X. SCHUMACHER, *Silviculturist, Division of Silvics*, and HENRY BULL, *Assistant Silviculturist, Southern Forest Experiment Station, Branch of Research, Forest Service, United States Department of Agriculture*

## INTRODUCTION

The Forest Service, under the authorization of the McSweeney-McNary Forest Research Act, is at present undertaking an exhaustive survey of the country's forest lands, by comparatively small natural or political subdivisions, such as groups of counties, with the purpose of obtaining therefrom a complete and accurate picture of the forest situation and the forest needs of the Nation. Chief among the required results of this survey are the following: Area in each type of forest cover, subdivided according to present productive condition; volume of timber by species; rate of growth on existing forests and reforesting areas; rate of forest depletion by cutting, fire, insects, and other agencies; and present national timber requirements and probable future trends.

The diversity of timber types, forest conditions, and timber growth to be encountered in such a survey, together with the scanty knowledge we now have of these in quantitative terms, make it imperative, in order to carry out the project efficiently with the funds available, that methods of gathering and compiling the enormous amount of data be carefully worked out in advance for each major forest region. This is especially important in view of the fact that the bulk of the data must be collected at first hand in the field.

Of first importance is a method of estimating the extent of data that must actually be gathered in any one large unit of area, and the nature of the survey, in order to obtain a reliable basis for estimating accurately the condition of the entire unit. The object of this paper is to present, by means of the analysis of a comparatively small set of field data gathered for this purpose from the bottom-land hardwood forest region, a practical means of determining these particulars for any given region and of making certain that the probable error of the results, for groups as small as several counties, will not exceed a previously assigned maximum.

## THE GENERAL PROBLEM AND THE DATA

Fundamentally the problem is one of sampling. Its object, stated statistically, is to reduce the standard error of a mean to the practicable minimum. Theoretically this may be done either by eliminating all irrelevant variation from the standard deviation of the whole unit of data, or by increasing the number of variates or measurements, as is evident in the following formula:

$$\sigma_M = \frac{\sigma}{\sqrt{N}} \quad \text{-----------------------} \quad (1)$$

[1] Received for publication Mar. 18, 1932; issued December, 1932.

in which

$\sigma_M$ = the standard error of the mean

$\sigma$ = the standard deviation of the universe, and

$N$ = the number of variates or measurements

There are, however, practical difficulties in either of these procedures. On the one hand, the standard deviation is a characteristic of the universe dealt with. It is an aggregate of variation from an indefinite number of causes; and only that portion which is recognized by quantity and by source may be logically deducted, leaving the residual error of estimate. On the other hand, the number of variates that may be collected is limited by time and funds available for the field work of a survey. In order, therefore, to find out what reliability of the mean can be attained with given funds it is necessary to gather a certain amount of data from a given universe and, first of all, analyze the standard deviation in order to arrive at the proper value for $\sigma$, at the same time keeping cost account of the field work to establish the highest practicable numerical value of $N$.

With these ends in view, a preliminary survey was made in the spring of 1931 in East Carroll and West Carroll Parishes in northeastern Louisiana. These parishes lie wholly within the Mississippi River Delta region[2] and contain bottom-land hardwood forests typical of all but the southernmost portion of the Delta. The forest and nonforest types are arranged in irregular north and south belts paralleling the Mississippi River and smaller drainages. Immediately west of the Mississippi there is a belt of "batture" land that extends to the levee. It is partly to completely overflowed every year and is almost entirely forested, principally with willow (*Salix* sp.) and southern cottonwood (*Populus deltoides virginiana* (Castigl.) Sudw.). Just west of the levee is a belt of rich agricultural land, planted largely to cotton. Beyond this lie, in order, a belt of cut-over bottom-land hardwoods; a belt of agricultural land along Bayou Macon, which separates the two parishes, and on Macon Ridge; and finally a broad belt of bottom-land hardwoods which contains, in two large tracts, the bulk of the virgin timber. These so-called belts are, of course, by no means entirely of the classification given but are principally as described.

The sample survey was made by what is known as the line-plot method. Ten parallel survey lines, 3 miles apart, were run due east and west in order to cross the prevailing drainages approximately at right angles. At each 10-chain (660 feet) interval along the lines, a quarter-acre sample plot was established on which, if in the forest, measurements were made; if not, the nonforest type was noted. In this way, 240 miles of line were measured, and 1,918 plots were taken, of which 1,189 fell within the forest. Since the total area of the two parishes by the most reliable estimate is 503,000 acres, nearly 0.1 per cent of the area is actually in sample plots. Without entering into any detailed account of the manner in which the data were collected on each plot, or the exact nature of the many items of information, it is proposed to analyze the manner and extent of sampling to determine its adequacy with respect both to land area by forest and

[2] The Mississippi River Delta region extends from above Cape Girardeau, Mo., some 60 miles northwest of the mouth of the Ohio River, to the Gulf of Mexico, and comprises an area of about 29,000,000 acres.

nonforest classes and to volume of timber by species and forest-condition class.

## AREA ANALYSIS

The object of the area analysis is to arrive at a correct and feasible method of estimating the land area in the following land classes, together with a statement of the error of estimate of each: Total forest land; forest land in each of 5 forest sites—ridge, flat, swamp, batture, and upland;[3] forest land in each of 7 forest conditions—virgin, culled, cut-over and restocking, cut-over and not restocking, ruined, second growth, and old field;[3] total land not in forest (nonforest); and nonforest land in each of 11 nonforest conditions.

There are two sets of data both suitable for analysis. First is the number of plots counted in each land class in terms of the entire number of plots in the survey. Analysis of these data involves homograde statistics or the statistics of attributes. The other possibility is the linear distance traversed in each land class in terms of the entire length of line in the survey. Analysis of these data involves heterograde statistics or the statistics of variables.

TABLE 1.—*Summary of land class area data for East Carroll and West Carroll Parishes, La.*

| Land-area class | Line survey | | Plots | |
|---|---|---|---|---|
| | *Chains* | *Per cent* | *Number* | *Per cent* |
| Forest site: | | | | |
| Ridge | 4,708 | 24.56 | 469 | 24.45 |
| Flat | 6,198 | 32.33 | 616 | 32.12 |
| Swamp | 301 | 1.57 | 23 | 1.20 |
| Batture | 637 | 3.32 | 65 | 3.39 |
| Upland | 176 | .92 | 16 | .83 |
| Total in forest | 12,020 | 62.70 | 1,189 | 61.99 |
| Forest condition: | | | | |
| Virgin | 1,515 | 7.90 | 154 | 8.03 |
| Culled | 228 | 1.19 | 23 | 1.20 |
| Cut-over restocking | 7,654 | 39.93 | 758 | 39.52 |
| Cut-over nonrestocking | 1,032 | 5.38 | 93 | 4.85 |
| Ruined | 488 | 2.55 | 55 | 2.87 |
| Old field | 533 | 2.78 | 50 | 2.60 |
| Second growth | 570 | 2.97 | 56 | 2.92 |
| Total in forest | 12,020 | 62.70 | 1,189 | 61.99 |
| Nonforest type: | | | | |
| Cultivated farm land | 4,721 | 24.63 | 470 | 24.51 |
| Deadening | 1,034 | 5.39 | 114 | 5.94 |
| Pasture land | 541 | 2.82 | 59 | 3.08 |
| Wood-lot pasture | 317 | 1.65 | 33 | 1.72 |
| Abandoned farm land | 143 | .75 | 14 | .73 |
| Prairie | --- | --- | --- | --- |
| Marsh | 26 | .14 | 3 | .16 |
| Levee | 61 | .32 | 5 | .26 |
| Roads, railroads, etc | 138 | .72 | 9 | .47 |
| Towns, villages | 23 | .12 | 2 | .10 |
| Rivers, lakes, etc | 145 | .76 | 20 | 1.04 |
| Total, nonforest | 7,149 | 37.30 | 729 | 38.01 |
| Total area | 19,169 | 100.00 | 1,918 | 100.00 |

When the data are listed according to both methods, as in Table 1, it is evident that, on the basis of percentages of area, the two methods check satisfactorily. If the resulting errors of estimate were also to check, the use of the plot count would be preferred because of the

[3] For definitions of these sites and conditions, see the following publication: LENTZ, G. H. THE FOREST SURVEY IN THE BOTTOMLAND HARDWOODS OF THE MISSISSIPPI DELTA. Jour. Forestry 29: 1046–1055. 1931.

simplicity of its application, and because the linear measurement of area classes traversed in the field could then be omitted entirely.

## PLOT-COUNT METHOD

The plot-count method is useful in determining whether we are dealing with a true random sample of land-area classes; that is, whether the statistical sample—the number of plots by land-area class—has been selected in such a way that any plot has been as likely of selection from the statistical universe (the bottom-land area as represented by East Carroll and West Carroll Parishes) as any other. If we have such a sample, the standard deviation of the number of plots in a given land class is

$$\sigma = \sqrt{npq} \qquad (2)$$

in which

$\sigma$ = the standard deviation of the number of plots
$p$ = the chance that a plot is of a given land class
$q$ = the chance that it is not, where $p+q=1$
$n$ = the total number of plots

The standard error of the proportion of plots in a given land class may be written

$$\sigma_p = \frac{\sigma}{n} = \sqrt{\frac{pq}{n}} \qquad (3)$$

This is a more useful measure for the final expression because it gives, directly, the standard error of a particular land class as a percentage of the total land area. For comparison with actual dispersion, however, that measure will be used which is the handier for the particular purpose. In the following application of formula 3, the cut-over restocking forest condition is used because it contains the greatest number of plots. According to Table 1 there are:

758 plots in cut-over restocking condition
1, 160 plots in other conditions (both forest and nonforest)
1, 918 plots in all

hence

$$p = \frac{758}{1918} = 0.395$$

$$q = \frac{1160}{1918} = 0.605$$

and from equation 2

$$\sigma = \sqrt{1918 \times 0.395 \times 0.605}$$
$$= 21.5 \text{ plots}$$

This means that, provided the conditions of random sampling have been fulfilled, it is to be expected that in 2 out of 3 additional sets of 1,918 plots the number in the cut-over restocking conditions will be within 22 of the present count.

But this is the point to be investigated. We need to know whether the standard deviation calculated in this way from the 1,918 plots is identical with the standard deviation that would be obtained if all the quarter acres of land area in the two parishes could be thrown into an urn, thoroughly mixed, and a number of samples consisting of 1,918 plots each could be drawn out. Obviously such a test can not be made. However, the principle may be tested by means of two different schemes of setting up data for the calculation of the actual standard deviation. The first of these is a system of replications or mechanical groupings, in which, for example, group 1 is composed of plots Nos. 1, 21, 41, 61, etc., group 2 is composed of plots Nos. 2, 22, 42, 62, etc., making in all 20 groups with about 96 plots in each group. Another scheme compares the relative number of plots in cut-over restocking condition from one survey line to another, making 10 groups, one for each line.

### Test by Mechanical Grouping

From the distribution of the mechanical groupings as presented in Table 2, the mean and standard deviation are reckoned as follows:

$$\text{Mean} = 37.9 \pm 0.62 \text{ plots out of } 96$$
$$\text{Standard deviation} = 2.75 \pm 0.44 \text{ plots out of } 96$$

Table 2.—*Frequency of plots in cut-over restocking condition in 20 mechanically set up groups of 96 plots each*

| Group No. | Number of plots of cut-over restocking | Group No. | Number of plots of cut-over restocking |
|---|---|---|---|
| 1 | 35 | 12 | 37 |
| 2 | 38 | 13 | 34 |
| 3 | 33 | 14 | 37 |
| 4 | 33 | 15 | 37 |
| 5 | 41 | 16 | 38 |
| 6 | 38 | 17 | 42 |
| 7 | 41 | 18 | 40 |
| 8 | 40 | 19 | 39 |
| 9 | 43 | 20 | 39 |
| 10 | 36 | | |
| 11 | 37 | | 758 |

Now, the standard deviation to be expected from a random sample of 96 plots is, from equation 2:

$$\sigma = \sqrt{96 \times \frac{37.9}{96} \times \frac{58.1}{96}}$$
$$= 4.79$$

It is evident that $2.75 \pm 0.44$ can not be accepted as an accidental deviation from 4.79. However, familiarity with the region from which these samples are drawn helps to explain how this scheme of grouping plots might result in a subnormal dispersion. Each plot is but one-eighth mile distant from its immediate neighbors on the same line. Since areas of cut-over restocking condition, when encountered in the field at all, are usually extensive, it follows that wherever one plot in this condition occurs a number of consecutive ones are likely to be found. Since each consecutive plot has been put into a consecutive group, the variability from group to group is less than the variability of random samples.

### Test by Survey-Line Grouping

Table 3 gives the total number of plots on each line and the number and percentage of cut-over restocking plots on each, without regard to length of line. From these percentages the following calculations are made:

$$\text{Mean} = 39.0 \pm 2.42 \text{ per cent}$$
$$\text{Standard deviation} = 7.65 \pm 1.71 \text{ per cent}$$
$$\text{Average number of plots per line} = 192$$

TABLE 3.—*Frequency of plots in cut-over restocking condition, by survey lines*

| Line No. | All plots | Frequency of cut-over restocking | | Line No. | All plots | Frequency of cut-over restocking | |
|---|---|---|---|---|---|---|---|
| | *Number* | *Number* | *Per cent* | | *Number* | *Number* | *Per cent* |
| 1 | 115 | 46 | 40 | 7 | 260 | 118 | 45 |
| 2 | 150 | 53 | 35 | 8 | 210 | 96 | 46 |
| 3 | 201 | 71 | 35 | 9 | 209 | 117 | 56 |
| 4 | 197 | 72 | 37 | 10 | 171 | 57 | 33 |
| 5 | 200 | 71 | 35 | | | | |
| 6 | 205 | 57 | 28 | Total | 1,918 | 758 | ------ |

However, the theoretical standard error of the proportion is, by equation 3:

$$\sigma_p=\sqrt{\frac{0.395\times 0.605}{192}}$$
$$=0.0353 \text{ per plot or } 3.53 \text{ per cent}$$

The abnormal variation of the areas from line to line as compared with the theoretical dispersion demonstrates that the areas of cut-over restocking lands are by no means evenly distributed over the two parishes; that there is, rather, a relationship between percentage of cut-over restocking area and "place" in the two parishes.

It may be argued that since the actual variation may be made subnormal or abnormal depending upon the method of grouping, the theoretical standard deviation of the ungrouped data—21.5 plots out of the 1,918—may be a sufficiently satisfactory one, in that it should average out the high and low dispersions. The fallacy of this will be brought out in the test problem to be discussed later.

## LINEAR-MEASUREMENT METHOD

The linear-measurement method differs from the plot count in one important respect. Whereas the plot count of a land-area class is merely alternative, in that a given plot either is or is not of the class to be investigated, the linear measurement is qualitative, in that the distance traversed in the land class under investigation is measured. The problem, then, becomes one of analysis of variation of measurement on the hypothesis that the area of a given land class varies as the linear distance traversed within it. A discussion of this hypothesis will be taken up later.

To begin with, it is necessary to determine which unit length of line (e. g., one-half mile, 3 miles) gives the most stable measure of the dispersion of a land-area class. In this determination the cut-over restocking forest condition is again used.

### TESTS BY CONTINUOUS UNITS

The lengths analyzed are the ½-mile, 3-mile, and 24-mile units; for each of these the standard deviation and the standard error of the mean in number of chains are calculated and reduced for ready comparison to number of chains to the half mile. (Table 4.)

TABLE 4.—*Comparison of standard deviations and standard errors of the mean for various unit lengths for survey line in cut-over restocking condition*

| Length of unit | All units | Cut-over restocking statistics per one-half mile | | |
|---|---|---|---|---|
| | | Mean | Standard deviation | Standard error of the mean |
| | *Number* | *Chains* | *Chains* | *Chains* |
| ½ mile | 480 | 16 | 15. 9 | 0. 73 |
| 3 miles | 80 | 16 | 8. 76 | . 98 |
| 24 miles | 10 | 16 | 4. 78 | 1. 51 |

The increasing standard error of the means for units greater than one-half mile in length, as shown in Table 4, clearly indicates that the variation in the data from the longer units, like the abnormal variation of the plot count when full line units were used, is correlated with "place."

### TEST BY RANDOMLY GROUPED ½-MILE UNITS

Although Table 4 shows that the error of land-class estimate is less with the ½-mile unit than with the longer unit lengths, the standard deviation—15.9 chains to the half mile—is practically as great as the mean of 16 chains. This follows from the fact that the distribution is U-shaped; of the four hundred and eighty ½-mile units, 155 contain no chainage at all of cut-over restocking and 88 are entirely cut-over restocking.

On this account the standard deviation is subject to an error greater than that of unimodal distributions; and as the standard error of the mean of any distribution varies directly as the standard deviation, this also is subject to greater fluctuation in this type of distribution. Therefore, rather than to rely upon a single standard deviation from which the error of the mean is to be calculated, actual variations of means from groups or randomly selected ½-mile units were compared with the theoretical. For this purpose, the number of chains of land in cut-over restocking condition in each of the four hundred and eighty ½-mile units was tallied upon small, metal-rimmed, circular tags, such as are used for price tags. The 480 tags were thoroughly mixed in a box, drawn out one at a time, and the values noted in the order of draw. Four groupings were thus made. Their standard deviations are given in Table 5.

TABLE 5.—*Comparison of standard deviations of groups of random-selected ½-mile units in cut-over restocking condition*

| Units in group | Groups | Standard deviation per one-half mile | Units in group | Groups | Standard deviation per one-half mile |
|---|---|---|---|---|---|
| *Number* | *Number* | *Chains* | *Number* | *Number* | *Chains* |
| 12 | 40 | 4. 71 | 24 | 20 | 2. 61 |
| 20 | 24 | 3. 82 | 40 | 12 | 2. 72 |

Figure 1 summarizes graphically the work thus far on the linear-measurement method. It shows the standard deviation of ½-mile units and the standard deviation of the 3-mile and 24-mile continuous units, all from Table 4; the standard deviations of distribution of the randomly grouped ½-mile units, from Table 5; and the curve of the theoretical standard deviation of randomly grouped ½-mile units. Figure 1 indicates that unit lengths greater than one-half mile introduce systematic errors of the mean, the result of correlation of cut-over restocking areas with "place" in the two parishes. On the other hand, the actual error of the groups, after the effect of "place"

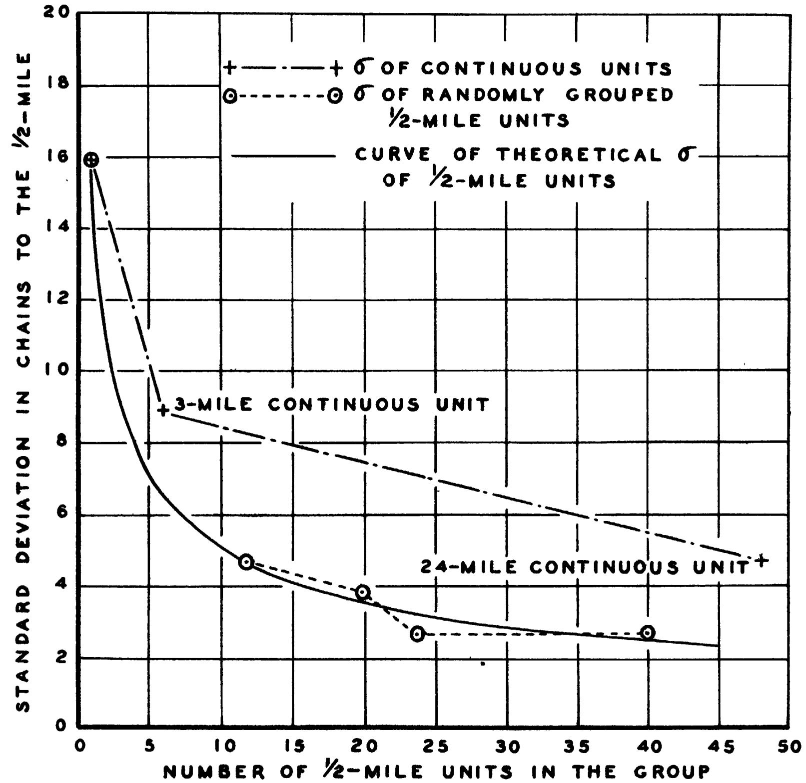


FIGURE 1.—Comparison of the standard deviations resulting from the various methods of grouping with the curve of the theoretical standard deviation. Cut-over restocking condition

is removed by random selection of ½-mile units, checks very well with the theoretical.

It is concluded, therefore, that when areas in land units similar to these two Louisiana parishes are determined by the linear-measurement method, the error of sampling of a land class may be deduced directly from the distribution of ungrouped ½-mile units.

## PLOT COUNT VERSUS LINEAR MEASUREMENT

Up to this point it has been satisfactorily established (1) that the plot-count method exposes systematic variation, if present, but does not give stable errors of estimate of land-class area; and (2) that if the linear-measurement method is employed, the use of the ½-mile unit results in errors of estimate free from systematic variation over the territory sampled. But this gives little indication of which method is to be preferred. It may be, for instance, that the error of the ungrouped plot-count data—the 21.5 plots out of the 1,918, or about 1 per cent of the land area—is sufficiently close to the truth in that it averages the subnormal and abnormal variations of the grouping schemes used. The choice of method rests upon the agreement of estimated areas with the true areas. But such comparisons can not be made for the land classes as defined because we have no true areas upon which to base them.

Recourse was had in this instance to a soil-type map of East Carroll and West Carroll Parishes upon a scale of 1 inch to the mile.[4] As the different soil types roughly parallel the Mississippi River and interior drainages just as do the forest sites and conditions, this served very well as a medium for the comparison of method, in the following manner:

The 10 parallel east and west survey lines actually determined were drawn on the map, 3 miles apart, establishing as line universes the area within 1½ miles on either side of each survey line.

The area of each soil type was planimetered by line universe. Seven soil types occurred on nine of the line universes and three soil types on the remaining one. The planimetered areas of these 66 cases were taken as the true areas of the soil types, the base for comparison of sampling method.

The sampling process on the map resembled the sampling of land-area classes in the field—(1) by counting plots at every 10 chains along the lines and noting the soil type of each, and (2) by scaling the distance in chains across each soil type traversed by the survey lines.

### Test of the Plot-Count Method

The difference between the actual areas and the areas estimated by plot count were calculated in terms of the standard error of the latter for each of the 66 cases. A frequency distribution of these ratios—if the method has fulfilled the requirement of random sampling—should have a standard deviation of 1, and its departure from 1 may be considered a measure of the failure of the sampling method. The following example of Sharkey clay soil type on line universe 1 serves as an illustration:

Planimetered area of Sharkey clay = 3,456 acres

Planimetered area of line universe = 26,132 acres

Relative area of Sharkey clay $= \frac{3456}{26132} = 0.132$

Number of Sharkey clay plots on line 1 = 18

Total number of plots on line 1 = 112

Relative number of Sharkey clay plots $= \frac{18}{112} = 0.161$

Standard error of relative number of Sharkey clay plots:

$$\sigma_p = \sqrt{\frac{0.161 \times 0.839}{112}} = 0.035$$

[4] Worthen, E. L., and Belden, H. L. soil survey of east carroll and west carroll parishes, louisiana. U. S. Dept. Agr., Bur. Soils Field Oper. 1908, Rpt. 10: 875–898, illus. 1911.

Difference between actual and estimated areas in terms of the standard error of the latter is

$$\frac{0.132-0.161}{0.035}=-0.8\sigma$$

The distribution of the ratios is shown in Table 6.

TABLE 6.—*Distribution of the ratios of actual errors of soil-type area to estimated errors by the plot-count method*

| Error [a] | Frequency | Error [a] | Frequency | Error [a] | Frequency | Error [a] | Frequency |
|---|---|---|---|---|---|---|---|
| +5.5 | -------- | +2.0 | 6 | −1.5 | 5 | −4.5 | -------- |
| +5.0 | 1 | +1.5 | 5 | −2.0 | 2 | −5.0 | -------- |
| +4.5 | 1 | +1.0 | 6 | −2.5 | 1 | −5.5 | 1 |
| +4.0 | -------- | +0.5 | 6 | −3.0 | 2 | | |
| +3.5 | -------- | 0 | 8 | −3.5 | -------- | | 66 |
| +3.0 | 3 | −0.5 | 11 | −4.0 | -------- | | |
| +2.5 | 1 | −1.0 | 7 | | | | |

[a] The difference between the actual area and the estimated area in terms of the standard error of the estimated area.

Their standard deviation is

$$\sigma=1.72\pm0.149$$

Had we assumed the sampling method to have fulfilled the requirements of random sampling we would have taken for granted that the standard error of the plot count would give the range within which the true error lies two times out of three, or with odds of 2 to 1. We find, however, that the actual error is, most probably, 72 per cent higher; that the estimated standard error is in reality $\frac{1}{1.72}$, which, by reference to the normal curve of error, is equivalent to odds of but 0.8 to 1.

## TEST OF THE LINEAR-MEASUREMENT METHOD

In testing the linear-measurement method essentially the same kind of comparison was made as in the case of the plot-count method. The following is an illustration of the calculation in one case, employing the Sharkey clay on line universe 1:

Relative area of Sharkey clay = 0.132 (as above).

Statistics from frequency distribution of linear measurements, shown in Table 7:

TABLE 7.—*Frequency distribution of scaled distances of Sharkey clay land on line 1*

| Chains of Sharkey clay to the one-half mile | ½-mile units | Chains of Sharkey clay to the one-half mile | ½-mile units | Chains of Sharkey clay to the one-half mile | ½-mile units | Chains of Sharkey clay to the one-half mile | ½-mile units |
|---|---|---|---|---|---|---|---|
| | *Number* | | *Number* | | *Number* | | *Number* |
| 0 | 23 | 15 | ---------- | 30 | ---------- | 40 | 3 |
| 5 | ---------- | 20 | 2 | 35 | ---------- | | |
| 10 | 1 | 25 | ---------- | | | Total---- | 29 |

Mean = 5.17 chains to the half mile
Standard deviation = 10.7 chains to the half mile
Standard error of mean = 1.99 chains to the half mile

Hence, relative area of Sharkey clay $= \frac{5.17}{40} = 0.129$

Standard error or relative area of Sharkey clay:

$$\sigma_p = \frac{1.99}{40} = 0.050$$

Difference between actual and estimated areas in terms of the standard error of the latter:

$$\frac{0.136 - 0.129}{0.050} = +0.1\sigma$$

The distribution of the ratios are shown in Table 8.

TABLE 8.—*Distribution of the ratios of actual errors of soil-type area to estimated errors by the linear-measurement method*

| Error [a] | Frequency | Error [a] | Frequency | Error [a] | Frequency | Error [a] | Frequency |
|---|---|---|---|---|---|---|---|
| +4.2 | 1 | +2.1 | -------- | 0 | 9 | −2.1 | -------- |
| +3.9 | | +1.8 | 3 | −.3 | 15 | −2.4 | -------- |
| +3.6 | | +1.5 | 1 | −.6 | 4 | −2.7 | -------- |
| +3.3 | 1 | +1.2 | 2 | −.9 | 2 | −3.0 | 1 |
| +3.0 | -------- | +.9 | 7 | −1.2 | 4 | | |
| +2.7 | -------- | +.6 | 8 | −1.5 | -------- | | 66 |
| +2.4 | -------- | +.3 | 7 | −1.8 | 1 | | |

[a] The difference between the actual area and the estimated area in terms of the standard error of the estimated area.

The standard deviation is:

$$\sigma = 1.05 \pm 0.093$$

This does not differ significantly from 1 and therefore satisfies the requirement that the estimated errors be the true errors of estimate. Furthermore, since no correlation is found between the ratios and the size of area estimated, it is proved that the area of a land class, in a universe represented by these two parishes, varies as the linear distance traversed within it.[5]

### COMPARISON OF METHODS

The final results of the two sampling methods are compared in Table 9, the last three columns of which show the errors of estimate. On the assumption that plot counts at spacings of 10 chains by 3 miles satisfy the conditions of random sampling, the resultant error by this method is the lesser. As has been brought out, however, this assumption is erroneous, as it requires, for the soil-type data at hand, a correction factor of 1.72. This correction has been applied in the table, the two right-hand columns bringing out the properly comparable errors.

[5] At first thought it seemed more logical to one of the authors that the area of a land class should vary as the sums of the squares of the linear distances traversed within it. Walter H. Meyer, of the Pacific Northwest Forest Experiment Station, tested this hypothesis and found that it leads to estimates that are too high whenever a single land class occurs as an extensive, unbroken area traversed by two or more survey lines.

TABLE 9.—*Relative areas of soil types and comparison of their standard errors of estimate by different sampling methods*

| Soil type [a] | Relative areas of soil types | | | Standard error of area as a percentage of corresponding area | | |
|---|---|---|---|---|---|---|
| | | | | By plot count | | |
| | By planimeter | By plot count | By linear measurement | Random sampling | Corrected for sampling [b] | By linear measurement |
| | *Per cent* | *Per cent* | *Per cent* | *Per cent* | *Per cent* | *Per cent* |
| YS | 5.5 | 5.3 | 4.2 | 10.0 | 17.2 | 24.0 |
| YL | 6.4 | 5.8 | 6.5 | 9.5 | 16.3 | 15.0 |
| WC | 17.2 | 19.6 | 18.1 | 4.7 | 8.1 | 9.3 |
| SC | 25.6 | 24.2 | 24.5 | 4.1 | 7.1 | 7.9 |
| WSL | 8.9 | 9.8 | 9.9 | 7.1 | 12.2 | 14.1 |
| RL | 17.8 | 17.3 | 18.7 | 5.1 | 8.8 | 9.0 |
| CS | 18.3 | 18.0 | 16.6 | 5.0 | 8.6 | 9.8 |
| Total | [c] 99.7 | 100.0 | [d] 98.5 | | | |

[a] Symbols for soil types are those given on the map mentioned in text.
[b] This correction is 1.72 times the value obtained on hypothesis of random sampling.
[c] Three-tenths of 1 per cent of the total area is made up of 3 soil types which are not included because they are not encountered in the plot-count method.
[d] The 3 soil types not given in the table account for the remaining 1.5 per cent.

In six of the seven soil types, the corrected error of the plot-count method is still less than the error of the linear-measurement method. It is very doubtful, however, whether the differences in its favor are to be considered as a recommendation, for the correction is an empirical one, and may safely be applied only to this soil-type test. The error of linear measurement, on the other hand, is the outcome of variation actually encountered; it calls for no correction, because by comparison with the universe from which the data were drawn, it satisfies the limitations of random sampling.

## VOLUME ANALYSIS

The volume data consist of the board-foot contents, by species, of the timber on 1,208 sample plots [6] taken in the forest. From these the following values for all species together have been calculated:

Mean = 659 board feet
Standard deviation = 854 board feet
Standard error of mean = 24.5 board feet

But the standard error of the mean varies directly as the standard deviation of the distribution, and the latter measure is an aggregate of variation from many causes. One of these readily recognized in the field is the forest condition.

The plots were accordingly sorted by condition. The mean volume and standard deviation of each are as given in Table 10. The standard deviations are, in fact, standard errors of estimate of plot volume free from the influence of forest condition.

[6] This does not agree with the 1,189 forest plots of Table 1. The field instructions called for the measurement of a forest plot whenever a nonforest plot happened to fall within 2 chains of the forest, the purpose being simply to strengthen volume data. The added 19 plots were not used in area analyses.

TABLE 10.—*Means and standard deviations of plot volumes by forest condition*

| Forest condition | Basis, plots | Mean volume per plot | Standard deviation of volume per plot |
|---|---|---|---|
| | *Number* | *Board feet* | *Board feet* |
| Virgin | 155 | 1, 886 | 1, 232 |
| Culled | 22 | 1, 779 | 792 |
| Cut-over restocking | 761 | 478 | 533 |
| Cut-over not restocking | 100 | 354 | 576 |
| Ruined | 59 | 96 | 200 |
| Old field | 55 | 360 | 450 |
| Second growth | 56 | 700 | 976 |

The average of these, weighted by number of plots, may be expressed:

$$\bar{\sigma}=\sqrt{\frac{N_1\sigma_1{}^2+N_2\sigma_2{}^2+\;.\;.\;.\;.\;.\;.\;.\;.\;.\;+N_n\sigma_n{}^2}{N_1\;\;+\;\;N_2+\;.\;.\;.\;.\;.\;.\;.\;.\;.\;+N_n}} \qquad (4)$$

in which

$\bar{\sigma}$ = the weighted mean value of standard deviations from $n$ frequency distributions

$\sigma$ = the standard deviation of each distribution

$N$ = the number of plots in each distribution

and subscripts 1, 2, . . . . $n$ refer to the individual distributions.

The weighted mean value of the standard deviations of Table 10 thus becomes 684 board feet.[7] Hence, by equation 1, the standard error of the mean, free from the influence of forest condition, becomes 19.7 board feet.

The data were further scrutinized for evidence of relationship between the remaining variation of the mean and other factors, particularly site and locality, but nothing of significance was found.[8] Hence it is concluded that 19.7 board feet to the quarter acre is the error of sampling. This is 2.99 per cent of the mean volume.

The error of the total volume of the two parishes, however, must be greater than 2.99 per cent because it is affected not only by volume variation from plot to plot, but also by the error of estimate of total forest area. In short, the correct estimate of volume on a large forest area is

$$(A\pm\sigma_A)\;(M\pm\sigma_M)$$

in which

$A$ = the total forest area, in this case in quarter acres

$M$ = the mean volume to the quarter acre

$\sigma_A$ = the standard error of the total forest area

$\sigma_M$ = the standard error of the mean volume

This equation has been shown to be:

$$AM\pm\sqrt{A^2\sigma_M{}^2+M^2\sigma_A{}^2} \qquad (5)$$

[7] We may calculate a measure of the effect of forest condition upon the volume of hardwoods in the two parishes. For the squared variation of volume free from the influence of condition, [$=(684)^2$]+the squared variation of volume associated with condition=the total squared variation [$=(854)^2$]. From this equality we find that 35.8 per cent of the total squared variation of volume is associated with forest condition. Hence the correlation ratio, which is calculated directly from the square root of this value, is 0.598 ±0.018.

[8] A standard deviation below the 684 board feet obtained by stratifying according to condition may be arrived at by grouping the plots either systematically or at random and finding the standard deviation of the means of the groups. This value times the square root of the number of plots per group gives the residual standard deviation of plot distribution. With 20 or less groups a standard deviation of 566 board feet has been obtained. However, there is no justification for the reduction as it is impossible to state just what factor or factors of correlated variation have been averaged out in the process of grouping.

149774—32——4

The best estimate of the total area of the two parishes is 503,000 acres,[9] and this is taken as without error. The total forest area, calculated by the linear-measurement method, is $62.7 \pm 2.95$ per cent of this, or $1,261,524 \pm 37,215$ quarter acres. Since the mean volume to the quarter acre is $659 \pm 19.7$ board feet, equation 5 gives a total volume in the two parishes of $831 \pm 34.9$ million board feet and the standard error is 4.20 per cent of the total.

We now have a working basis for calculating the total number of miles of survey line that must be run across land units similar to East Carroll and West Carroll Parishes in order to determine the total volume of merchantable timber within any given limit of accuracy. The total number of miles, with a plot taken every 10 chains, is one-eighth of the total number of plots and, in the present case, almost exactly one-fifth of the number of forest plots (for forest area equals 62.7 per cent of the total area). From equation 1 we have

$$N = \left(\frac{\sigma}{\sigma_M}\right)^2 \qquad (6)$$

where

$N$ = the number of forest plots

The numerator to use for ready calculation of equation 6 should be that standard deviation which measures variation of plot volume and forest area, but which is free from the influence of variation in forest condition. We have this measure, but it is expressed as the standard error of the mean (or total)—the 4.20 per cent arrived at above. The corresponding standard deviation is $\sqrt{1,208}$ plots times this value. The denominator should be the standard error expressed as a percentage of the mean (or total)—that is, the measure of precision desired.

The results in number of plots or number of miles of line must be coordinated with the results similarly obtained from other analyses, such as volume of certain important species or volume by log scale, before any final estimate of the amount of data needed can be rendered.

## COORDINATION AND APPLICATION OF ANALYSES

In conclusion, it remains only to show the practical use or application of the calculated errors of sampling and the way in which the errors for the different items of area and volume are coordinated. The vital question, indicating the raison d'être for the foregoing analyses, is: How many miles of survey lines must be run to obtain acceptable limits of error in all important phases of the survey? From the number of miles, the distance between lines is of course readily calculated when the area of the land unit to be surveyed is approximately known.

To answer this question, it is necessary first to coordinate the different errors for the same mileage of line by putting them in such form that they can readily be compared. This is conveniently done by making a table that shows the standard error (in percentage of

[9] This figure has been arrived at after comparing estimates of the area concerned as established by the Bureau of Chemistry and Soils and the Bureau of Agricultural Economics, United States Department of Agriculture, and the county surveyor.

the mean) for each important item for several selected mileages of line. The mileages selected should be within the range of practical possibility as determined by the size of the land unit to be surveyed, the cost per mile of line, the funds available, the approximate permissible errors, etc. Table 11 is a sample portion of such a tabulation.

TABLE 11.—*Standard errors of important items for different total lengths of line*

[A sample portion of a tabulation from which the required mileage of survey lines can most easily be determined. Eight plots taken per mile. Each standard error expressed as a percentage of the mean of the item concerned]

| Item to be evaluated | Percentage of the mean for the indicated total number of miles of line run | | | | |
|---|---|---|---|---|---|
| | 200 | 300 | 500 | 1,000 | 1,200 |
| | At standard error of— | | | | |
| | $x\sqrt{5}$ | $x\sqrt{3.33}$ | $x\sqrt{2}$ | $x$ | $\frac{x}{\sqrt{1.2}}$ |
| Area: | | | | | |
| Virgin forest | 15.4 | 12.6 | 9.8 | 6.9 | 6.3 |
| Cut-over-restocking forest | 4.7 | 3.8 | 3.0 | 2.1 | 1.9 |
| Total forest | 3.2 | 2.6 | 2.0 | 1.4 | 1.3 |
| Volume (including error of area): | | | | | |
| Virgin forest | 16.6 | 13.5 | 10.5 | 7.4 | 6.7 |
| Cut-over-restocking forest | 6.5 | 5.3 | 4.1 | 2.9 | 2.6 |
| Red gum | 12.3 | 10.1 | 7.8 | 5.5 | 5.0 |
| Total forest | 4.5 | 3.6 | 2.8 | 2.0 | 1.8 |
| (Etc.) | | | | | |

The greatest acceptable error in the most important item (in this case, and usually, total merchantable volume) may then be decided upon and the corresponding mileage read directly or interpolated from the table. The errors in other items for this mileage are then noted and if any of them are unsatisfactory the mileage may have to be increased. Many compromises and concessions will invariably be necessary before the errors in most or all of the items are declared satisfactory for a financially practicable mileage.

In the bottom-land hardwood region the forest survey proper will be started in a land unit of about 5,500 square miles in northern Louisiana. After a detailed study of a tabulation of expected errors for given mileages, of which Table 11 is a small part, it is estimated that the results for this unit will be sufficiently close to the truth if the total mileage run is 550. This gives a line interval of 10 miles. The standard errors expected for the several items of Table 11 are about 95 per cent of those tabulated therein in the 500-mile column.

## SUMMARY

Samples of land-area measurements and timber-volume data were taken in two parishes in northeastern Louisiana by the line-plot system, for the purpose of determining the probable errors of estimate by land-area classifications and timber volumes to be expected from a projected forest survey of similar nature of 29,000,000 acres in the bottom-land forest region.

The land-area analyses disclosed that the number of plots in the various land classes, while correctly establishing their proportional areas, can not be considered as random samples of the land classes.

This is because neither the standard deviation as calculated from 20 mechanically set-up groups nor the standard deviation as calculated from the variation between the survey lines, checked with the standard deviation to be expected upon the hypothesis of random sampling.

The linear-measurement method of area by land classification was analyzed as an alternative to the plot-count method. The error of estimate was tested by using unit lengths of one-half mile, 3 miles, and 24 miles, and was found to increase with unit lengths longer than the one-half mile.

The comparison of the errors of estimate as between the plot-count and the linear-measurement methods was based upon a test problem from the soil-type map of the two parishes. The true areas in each soil type were arrived at by planimetering the map. The soil types were then sampled on the map in a manner analogous to that by which land-area classes were sampled in the field. The comparison of sampling method is based upon the agreement of the actual with the estimated sampling errors. It was found that the actual errors of the linear-measurement method agreed with the estimated errors.

In the volume analyses, the standard error of the mean volume in board feet was found to be about 3 per cent of the mean volume per ¼-acre plot; and this value is the residual after taking out that portion of the error which is due to variation in forest-condition classes. Since the total volume in the forests of the two parishes is the product of the mean volume per ¼-acre plot and the number of quarter acres in forest area, with their respective errors of estimate, the standard error of estimate of the total volume becomes slightly higher than 4 per cent of the total volume.

The results of the area and volume analyses were used as the basis for the recommendation that in the forest survey proper the line interval be 10 miles.