SAMPLING PROCEDURES FOR THE NATIONAL INVENTORY OF SOIL AND WATER CONSERVATION NEEDS[1]

BU-86-M | Janet Cassady and J. E. Dowd | June 1957

In order to obtain the basic soil information required for estimating conservation needs in each county or Soil Conservation district in the U.S., it was decided to use area sampling procedures in those counties where little or no soils mapping had been carried out. Where more extensive soils information is available as the result of a soils mapping program of the Soil Conservation Service, attempts will be made to combine these data with the estimates from the sampling procedure. In those counties where complete soils information is available this will be used as it stands. Samples will be drawn in these counties, however, and used to measure changes in land use either at the present time or at some future date.

The Biometrics Unit was given the responsibility under a joint agreement with the U.S.D.A. of working with State Soil Conservation officials in drawing samples for the thirteen northeastern states. The sample was drawn in the form of two independent subsamples, one to be mapped and measured in the current program, the other to be mapped and measured subsequently if time and resources allow.

Stratification

In all states but one (New Jersey) counties were stratified independently, the basis for stratification within each county being either a grid of compact land areas initially formed from the county, or an

[1] Biometrics Unit, Cornell University

initial division of each county into land resource areas (L.R.A.'s) followed by a division of each land resource area into compact land areas. The size of the resulting strata was in either case dependent upon the sampling rate, each stratum being of such a size that when two sampling units (one from each subsample) were drawn from it the designated over-all sampling rate for the county or L.R.A. was approximately achieved.

Thus if a 4% sample in the form of two 2% subsamples was designated for a county that county was divided into compact blocks of 49 sampling units each and two sampling units were then drawn (without replacement) from each stratum. Since a county or L.R.A. will rarely divide into strata all of equal size, one or more strata were either slightly greater or slightly less than the required size. This under-or-over representation of these strata can be compensated for by correct weighting of the estimates from these strata.

Most states specified that stratification be carried out by L.R.A. for all counties within the state. Certain counties in New York, Pennsylvania, and West Virginia were stratified on a straight grid basis ignoring L.R.A. separations since the State Soil Scientists for these states did not believe that there was significant enough differentiation in soil conditions betweel L.R.A.'s to make their use as strata boundaries worthwhile. In some cases counties were partially stratified by L.R.A. and partially by a straight grid system.

In New Jersey, the state was initially divided into L.R.A.'s and the L.R.A.'s further stratified into compact blocks, care being taken not to carry strata across county lines. Here the sampling rate was varied with the size and agricultural importance (as determined by the State Soil Scientist) of each L.R.A. Estimates will be expanded on to L.R.A. totals and county totals will be obtained for each soil separation by apportioning the acreage for each soil separation from each L.R.A. comprising the county in proportion to the amount of the L.R.A. occupied by the county. For example, if a county was composed of two L.R.A.'s and the county occupied one half of one L.R.A. and one third of the other, then given estimates x_1 and x_2 of soil separation x from the 1^{st} and 2^{nd} L.R.A.'s, the county estimate of x would be $\frac{1}{2} x_1 + \frac{1}{3} x_2$.

<u>Sampling Rate</u>

It was decided that a sampling rate of approximately 2% would give estimates of adequate precision for most soils information needed when the estimates were to be applied to a land area of between 300,000 and 500,000 acres. (The mean county size in the thirteen northeastern states is approximately 400,000 acres.) Thus, most counties in the survey were stratified into compact blocks of 4900 acres each and two separate subsamples of 2% each were drawn from each county. In the forest and mountainous areas of New Hampshire, New York and Vermont a reduced sampling rate of 1/2% was used because of the size of these land areas and because, due to the small agricultural importance of the areas, of a lesser need for detailed soil information. The standard sized sampling unit of 100 acres was used in the above areas. In the forest area of Maine a larger sized unit was used and to compensate for the lesser precision of estimates based upon larger sized units the sampling rate was raised from 1/2% for each subsample to .8% for each subsample. In New Jersey where the sampling was done by L.R.A. the sampling rate varied with both the size of L.R.A. and the agricultural importance of the L.R.A. Sampling rates vary from .3% for a large L.R.A. of little agricultural importance to 17% for a small L.R.A. of standard importance. (A short appendix to this report gives the rationale for choosing sampling rates for different sized land areas.)

Size of Sampling Unit

On the basis of a pilot study carried out by the Biometrics Unit it was determined that a sample unit of 100 acres would give close to optimum precision under conditions and costs of soils mapping currently found in representative counties in the Northeastern states. In only one instance was a different sized unit used. In L.R.A.B3a in Maine, a unit of 400 acres (approximately one day's work) was used since there are few roads in this area and travel costs in such rugged terrain would be considerably higher than average.

Mechanics of Drawing the Sample

The first step in the sampling procedure was to spread out a county highway map or aerial photo index and to place over it an overlay of thin paper marked off in square centimeters. On maps that have the scale 1":1 mile, 1 sq. cm. = 99.2025 acres. On maps that have the scale 1":62,500', 1 sq. cm. = 96.5281 acres. Thus, the square centimeter was used to approximate the sample unit for these. (On maps with scale 1":2 miles, one-fourth square centimeter equals 99.2025 acres.) Later, in transferring sample units to aerial photos, plastic templates were used to adjust the size of the sampling unit to 100 acres, according to the scale of the aerial photo.

The outline of the county was drawn on the graph paper overlay, then another line was drawn along the county boundary, following the edges of the square centimeters. A block was excluded if less than half of it was covered by county territory and included if more than half of it was covered by county territory.

In most cases the county was divided into strata of 49 sampling units. Different sized strata were used for forest land in Maine, New Hampshire, New York and Vermont and for each land resource area in New Jersey, according to the sampling rates determined. The strata were numbered. An A and a B sample unit were selected in each stratum by use of a random number table. By counting the squares within each stratum in serpentine fashion these numbers designated certain squares for the A and B samples. These units were transferred to the county map by placing red carbon paper between the overlay and the map and outlining the sample blocks on the overlay. Each unit was labeled according to its stratum number and sample letter.

Before this present procedure was adopted, a straight line grid was used, dividing the county into square strata of 49 100 acre blocks by use of a plastic overlay. Incomplete strata at the county edge were combined to form strata of approximately the same size, wherever possible.

In the beginning no attention was paid to exclusion of any water areas falling entirely within a county; estimates of the county under water were to be prepared from the sample acreage. This procedure has since been revised, though not before causing soil scientists some mental anguish by presenting them with sample plots in the middle of lakes or wide rivers.

The revised method is to exclude certain water areas before drawing the sample, so that the acreage sampled will approximate the Census Bureau figures for county land acreage. These figures exclude any acreage in ponds and lakes of more than 40 acres in extent, streams more than one-eighth mile in width, coastal waters, tidal flats below mean high tide or islands of less than 40 acres in any of the above-mentioned bodies of water. The figures do include as land acreage the acreage of bodies of water smaller than those mentioned above. In actual office practice, the procedure was to exclude a square centimeter from acreage sampled if more than half of it was covered by water of the category to be excluded.

Because of this procedure of approximating county boundaries and water boundaries when drawing the sample, sample units occasionally fell partially outside of county boundaries or partially in large bodies of water. The question was raised as to whether it was permissible to move those sampling units. A memorandum from J. E. Dowd stated as follows:

"From a statistical viewpoint it makes no difference if some units lie partially outside of the county. From the point of view of the soil scientist mapping or checking the soils in that county it might cause some inconvenience to have to map soils falling outside the county line. If this is the case it is permissable to move the sampling units in such a way that they fall just inside the county boundary.

"Instructions with respect to water areas are as follows: all rivers or streams greater than one-eighth mile in width and all ponds or lakes greater than forty acres in area are to be excluded from the area within each county to be sampled.

"Thus if less than half the area of a sampling unit falls in a body of water of the above category it should be moved until it is entirely on land within the county. If, on the other hand, more than one-half the area of the sampling unit falls in a body of water of the above category it should be excluded entirely from the sample. If this exclusion is made, the Biometrics Unit should be notified, designating the county and the sampling unit which has been excluded."

Bodies of water smaller than those excluded will be reported as they occur in the sample acreage, serving as a basis for estimates of total acreage of such bodies of water.

Land Resource Areas

In some states, and in some groups of counties within states, lines were drawn on the county highway maps dividing the county into land resource areas. A land resource area (sometimes called problem area) is a geographic area of land characterized by patterns of soils (including slope and erosion), climate, water resources, land use, and type of farming.

These areas may be continuous or divided into separate segments. Where land resource areas were deemed of importance, counties were divided by lines along edges of land resource areas, then the land resource areas were divided into strata of the desired number of blocks. Most strata were composed of 49 blocks. Some areas, primarily forest land, had strata large enough to permit the use of a reduced sampling rate and still give adequate representation of the area. In L.R.A.B3a in Maine (White Mountains and Maine Woods) strata were of 50,000 acres each from which two 400 acre sampling units were drawn giving two subsamples of .8% each. In L.R.A.B3a in Vermont and New Hampshire (White Mountains) and in L.R.A.B3b1 in New York (Adirondack Mountains) strata were of 19,600 acres from which two 100 acre sampling units were chosen, giving two subsamples of .5% each.

Ownership

According to Secretary Benson's original memorandum, May 10, 1956, "Data will be developed separately for privately owned and publicly owned land. The Soil Conservation Service will be responsible for collecting basic physical data on soil and water on non-federally owned lands." Later it was decided that the sample should be drawn with no consideration given to ownership of land, even though not expecting all the plots to be mapped subsequently. Samples falling on military reservations or within large blocks of natural forest will not be mapped at this time.

In New York the State Soil Scientist directed: "In cases where we are not permitted to get onto a reservation (this may be the case in some military installations) to map a sample, it will be necessary to accurately outline the reservation on the county map and then have Biometrics Unit redraw the strata surrounding it and to subtrace the area of the reservation from the acreage total for the county."

Counties Completely Surveyed

Where a county is completely surveyed and measured and there are no significant changes in land use from the time it was mapped, the entire county data will be used as a basis for the inventory and no new mapping or measuring will be done. However, sampling units will be drawn in all counties whether completely mapped or not. Thus it will be possible to check not only changes in land use that take place from the time of mapping until the present time, but also changes which might occur, say in five years' time. It is planned to record land use changes by mapping unit and not just as a total for each sample plot. In this way it will be possible to investigate changes in land use by land capability units, or any other grouping of soils that might be desired.

Where a county is partially surveyed in blocks and the mapping has been measured, the unsurveyed samples will be mapped and measured. The previously measured area will be outlined on a county map and submitted with the data for this area to the Biometrics Unit, together with the data from the new mapping for expansion to the county acreage.

Aerial Photos

For several states, the county maps with sample units delineated on them were returned to the state, one copy of each map being retained by the Biometrics Unit. A file folder was set up for each county, in which the sampling record was put.

For other states, the Biometrics Unit transferred the sample units to aerial photo field sheets. Thus, as each county was sampled, the Biometrics Unit requested the aerial photo maps needed, transferred the plots to the face of the maps, outlining the A samples in red and the B samples in blue, and labeling each plot according to its stratum number and sample letter. Plastic templates were used to fix the size of the plots at 100 acres, according to the scale of the aerial photos.

The aerial photos, together with the county map showing the sample units, were then returned to the Soil Conservation Service. One copy of each map was kept by Biometrics Unit, except Massachusetts.

State Notes

Connecticut: Sample drawn by land resource areas within counties for eight Connecticut counties. Sampling rate for A and B samples: one unit out of each 49. For 1" - 1 mi., 2%. County maps returned to state with sample units designated.

Delaware: Sample drawn by land resource areas within counties for three Delaware counties. Sampling rate for A and B samples: one unit out of each 49, 2%. County maps returned with sample units designated.

Maine: Sample drawn by land resource areas within counties. Maps (from Highway Atlas) at scale of 1":2 mi.; sampling rate for A and B samples:

one 100-acre unit out of each 49, 2% for part of state. For land resource areas B3a and B3b, reduced A and B samples of .8% were drawn from strata composed of 125 units of approximately 400 acres. This was done because of inacessibility of land areas, lack of photographic coverage in this area, and the fact that land in this area is designated to remain woodland for period of inventory.

The pages of the Atlas were returned to the state office after sampling.

Maryland: Sample drawn by land resource areas within counties for Maryland counties, on maps with scale 1":1 mi. Sampling rate for A and B samples: one unit out of each 49, 2%. County maps returned to state with sample units designated.

Massachusetts: Parts of two counties, Middlesex and Worcester, were sampled for the SuAsCo Watershed Study. This was in the form of one 4% sample. A straight grid was used for the whole area, dividing the highway map (1":2 mi.) into strata of 7x7 160 acre units (1/4 inch squares). Two units were drawn, using random numbers, from each stratum. The sample units were transferred to the aerial photos in the Biometrics Unit office. Those to the west of the major land resource boundary were outlined in orange on the photos; those to the east, in yellow.

This area was not included in the subsequent sample for conservation needs. The needs sample was drawn by land resource areas within counties, on maps with scale 1":2 mi. It was in the form of two 2% samples, drawn from strata of 49 blocks. Only one set of maps was received by Biometrics Unit and this set was returned to Massachusetts with sample units designated on it.

New Hampshire: Sample drawn by land resource areas within counties. Maps 1":1 mi. Regular 2% A and B samples drawn except for land resource area B3a, for which 1/2% samples were drawn. Sample units were plotted on photos; photos and maps returned to state.

New Jersey: Sample drawn by land resource areas, without regard to county lines. Sampling rates within land resource areas determined according to size of land resource areas and intensity of agriculture. For area having low agricultural development, as specified by the State Soil Scientist, there was a lowered sampling rate. The rate was varied by varying the stratum size, selecting A and B sample units in each stratum. Map scale 1:62,500.

The tidal marsh and urban areas within sample counties, as well as three counties almost wholly taken over by urbanization, were excluded from the sample.

Maps returned to state.

New York: One of the first states sampled. Some of the counties were sampled on a straight grid, 2% A and B samples, with no attention paid to land resource areas. Because no water areas falling entirely within a county were excluded at first, some had sample plots in lakes. Later samples were drawn according to land resource areas within counties, 2% A and B samples. 1"-1 mi.

For most counties, Biometrics Unit plotted samples on aerial photographs. Maps and photographs were returned to New York officials.

Pennsylvania: Sample was drawn with stratification by land resource areas for 21 counties and 4 parts of counties in the ridge and valley section. These four counties had slightly different sized strata in the parts sampled differently. Forty-one other counties have been sampled on a straight grid without regard to land resource areas. Sampling rate: 2% A and B samples, 1":1 mi.

For some counties samples were plotted on field sheets or contact prints.

Rhode Island: Sample drawn by land resource areas within counties. 2% A and B samples. 1":1 mi. Maps returned.

Vermont: Sample drawn by land resource areas within counties. 2% A and B samples, except for B3a, which had 1/2% A and B samples. Maps at scale 1:62,500. Sample units plotted on photos by Biometrics Unit. Photos and map returned.

Virginia: Sample drawn by land resource areas within counties. 2% A and B samples. Maps at scale 1":1 mi. Sample units plotted on photos for some counties.

West Virginia: Sample drawn by land resource areas within specified counties. 2% A and B samples. Maps scale 1":1 mi. Maps returned with sample units designated.

Appendix

Suppose V_p is a measure of variation in the amount of soil of a particular separation found on each of the 100-acre sampling units in a county or land resource area of size p (i.e. the land area under consideration contains p 100-acre sampling units). Further suppose that for this amount of variation it is decided that estimates of sufficient relative precision can be derived if a sampling rate of r_p is used. Let n_p be the sample size derived from this rate. ($n_p = r_p \cdot p$). Then a measure of the relative precision with which the soil separation can be estimated by a sample of size n_p is

$$\text{Rel. prec.} = \frac{\frac{p(p-n_p)}{n_p(n_p-1)} \sum_{i=1}^{n_p} (X_i - \bar{x}_p)^2}{(p\bar{x}_p)^2} = \frac{(p-n_p)}{pn_p} \; \frac{V_p}{\bar{x}_p^2}$$

where $$V_p = \frac{1}{n_p-1} \sum_{i=1}^{n_p} (X_i - \bar{x}_p)^2 ,$$

X_i = the amount of the particular soil separation found on the i^{th} 100-acre unit of the land area under consideration

$\bar{x}_p$ = the average amount of the particular soil separation found on 100-acre units in the land area of size p.

Now if a land area is less than p units in size (say $k < p$) and if the sampling rate, r, is used, the relative variance associated with the estimate of the same soil separation will then be

$$\frac{k - n_k}{kn_k} \; \frac{V_k}{\bar{x}_k^2} \quad \text{where } n_k = r_k \cdot k$$

$$V_k = \frac{1}{n_k-1} \sum_{i=1}^{n_k} (X_i - \bar{x}_k)^2$$

and $\bar{x}_k$ = the average amount of the particular soil separation found on 100-acre limits in a land area of size k.

Note that if $k < p$ then $n_k < n_p$ if $r_k = r_p$.

Now if it is desirable to estimate this particular soil separation with the same relative precision in both land areas it is necessary that

$$\frac{(k-n_k)}{k \cdot n_k} \frac{V_k}{\bar{x}_k^2} = \frac{(p-n_p)}{p \cdot n_p} \frac{V_p}{\bar{x}_p^2}$$

or that
$$n_k = \frac{1}{\frac{1}{k} + \frac{(p-n_p)}{p \cdot n_p} \frac{V_p}{V_k} \cdot \frac{\bar{x}_k^2}{\bar{x}_p^2}}$$

If we assume that the average amount of the soil separation does not change as the size of the land area changes we get

$$n_k = \frac{1}{\frac{1}{k} + \frac{p-n_p}{p \cdot n_p} \frac{V_p}{V_k}}$$

or
$$r_k = \frac{1}{1 + \frac{(p-n_p)}{n_p} \frac{V_p}{V_k} \cdot \frac{k}{p}}$$

If the decrease in variance were proportional to the decrease in land area, i.e. $\frac{k}{p} = \frac{V_k}{V_p}$ then $r_k = r_p$ and the same sampling rate could be used for a land area no matter what the size. It is strongly suspected however (and has been observed empirically for other characteristics) that the decrease in variance is somewhat less than proportional to the decrease in size of land area. If it is possible to establish empirically a relation between V_k and V_p in terms of size of the land areas p and k, i.e. $V_k/V_p = f(p,k)$ then it is possible to determine that sampling rate r_k which will give the same precision for a land area of size k as would the rate r_p for a land area of size p.

It should be noted that an implicit assumption in the foregoing argument is that the soil separation under consideration has the same form of distribution throughout both land areas with different scale parameters (variances in this case) and the same location parameters (means in this case).

Unfortunately, in order to empirically determine the form of the function f(p,k) information is needed on the ratio V_k/V_p for several values of k. This information is not available. The pilot study carried out by the Biometrics Unit involved three counties of very nearly the same land area so that the results were of little use in establishing a relationship between V_p and V_k.

It was hypothesized, however, that the relationship between V_p and V_k could be approximated by $f(p,k) = (k/p)^{1/2}$. Although there was no way of checking the correctness of this hypothesis, this particular function gave "reasonable" results in that it damped the reduction in variance by an intermediate amount when the size of the sample was multiplied. That a function of the type $(k/p)^b$ may be appropriate is also suggested by the empirical relationship between size of cluster and the variance within clusters when the cluster size is assumed to be that of a county. This relationship as developed by Jessen[1] states that the variance V_k between units within a cluster of size k is a monotone increasing function of the cluster size given by

$$V_k = ak^b \quad b > 0$$

Thus, if we could assume this relationship will hold for very large cluster sizes we get

$$\frac{V_k}{V_p} = \frac{ak^b}{ap^b} = (\frac{k}{p})^b \quad b > 0$$

where we choose b = 1/2 as an arbitrary value.

In the following table we will assume that the precision given by a 4% sample in a county of 400,000 acres will give adequate average precision. Thus, we choose p = 400,000 and let k range in value from 4,000,000 acres to 5,000 acres.

---

[1] Jessen, R. J. (1942) Statistical Investigation of a Sample Survey for Obtaining Farm Facts, Iowa Agricultural Experiment Station, Research Bulletin 304.

Relationship Between Size of Land Area and Sampling Rate

Land area (acres)	$\frac{k_i}{p}$	$(\frac{k_i}{p})^{1/2}$	Sampling rate (%)
4,000,000	10.0	3.162	1.300
3,000,000	7.5	2.739	1.498
2,000,000	5.0	2.236	1.829
1,000,000	2.5	1.581	2.564
950,000	2.375	1.541	2.633
900,000	2.25	1.500	2.703
850,000	2.125	1.458	2.778
800,000	2.0	1.414	2.863
750,000	1.875	1.369	2.953
700,000	1.75	1.323	3.053
650,000	1.625	1.275	3.167
600,000	1.50	1.225	3.289
550,000	1.375	1.173	3.430
500,000	1.25	1.118	3.593
450,000	1.125	1.061	3.779
400,000	1.000	1.000	4.000
350,000	0.875	.935	4.266
300,000	0.750	.866	4.591
250,000	0.625	.791	5.004
200,000	.500	.707	5.565
150,000	.375	.612	6.374
100,000	.250	.500	7.692
75,000	.1875	.433	8.778
50,000	.125	.353	10.557
25,000	.0625	.250	14.286
15,000	.0375	.194	17.680
10,000	.0250	.158	20.868
5,000	.0125	.112	27.115