Agriculture Handbook No. 237

SAMPLING CODING and STORING FLOOD PLAIN DATA

Farm Economics Division • Economic Research Service
U.S. DEPARTMENT OF AGRICULTURE

FOREWORD

This handbook reports one phase of a study of the characteristics and use of rural flood plains conducted by the University of Chicago, under a Research and Marketing Act contract with the Economic Research Service, U. S. Department of Agriculture. Other phases include an analysis of types of rural occupance of flood plains, experimentation with classification of such occupance, and preliminary measurements of changes in occupance during the past 25 years. The project supervisor is Gilbert F. White, Chairman of the Department of Geography, University of Chicago. The designated representative of the Economic Research Service for administration of the contract is Robert C. Otte, Farm Economics Division.

Specifically, the report describes a unique field method for punchcard coding and storing of geographic sample data to be analyzed and compared with data from other sources. The method should be useful particularly to agencies concerned with flood-control planning and of general interest to researchers interested in coding and storing resource data.

CONTENTS

	Page
Summary	ii
Background and Objectives	1
Sampling Methods	2
Sampling units	4
Areal samples	4
Advantages of various areal samples	5
Tests of the relative efficiency of sampling systems	10
Sample size	11
Coding, Storage, and Comparative Analysis	14
Coding	15
Geographic coding	15
Storage	17
Analysis	20
Comparisons with previous studies	20
Analysis through time	21
Conclusion	24
Bibliography	26

August 1962

SUMMARY

This report attempts to develop practical methods of sampling, coding, and storing data relating to the agricultural occupancy of flood plains in the United States, given the problem that many objectives must be satisfied. It appears that the most efficient sample for analysis of areal distributions, for making estimates of the areal coverage of phenomena, and for comparative analysis is a systematic, stratified, and unaligned point sample.

The systematic feature of such a sample facilitates an unconventional use of punched card systems that has the invaluable property of maintaining geographic ordering of the data. Once prepared, the punched cards provide permanent storage in a compact format. A nominal scale is used for coding. The card system with nominal coding and geographic ordering facilitates comparative analysis of land-use, soil, slope, land-capability, flood-hazard, or other distributions by allowing easy preparation of contingency tables and rapid location of residuals in their geographic setting. Similarly, studies of changes through time are facilitated.

Since the systems and operations described are simple and readily applied, it appears that they could be undertaken directly in the field. This has the advantage of suggesting additional questions and supplemental observations while field workers are out in the study area, instead of later in the office.

SAMPLING, CODING, AND STORING FLOOD PLAIN DATA

By

Brian J. L. Berry, Assistant Professor, Department of Geography,
University of Chicago

BACKGROUND AND OBJECTIVES

This report is concerned with methods of sampling, coding, and storing data relating to the agricultural occupance of flood plains in the United States. 1/ Since a properly structured national sample must await the development of a classification of rural flood plains, we will focus on the analysis of individual flood plains.

Several aspects of the problem can be identified: (a) Overall measurements of land use on the flood plain, with a view to comparison with other areas and with the same area at other times; (b) variations in use of land within the flood plain and the adjacent area, together with changes in these uses through time; and (c) explanations of the variations in uses and their changes in terms of soil and slope characteristics, drainage, relative location, nature of the flood hazard, and so forth.

These objectives are not necessarily complementary. The nature and size of the sample and type of data storage system required are likely to differ, depending on which of the objectives is paramount. If all interests have to be served, as appears to be the case in studies of individual flood plains, the sampling system used must be a compromise.

For other reasons, however, the problem becomes one of maximizing utility for more than one purpose. Many different surveys are conducted at irregular intervals by a variety of public agencies and private individuals, working at different levels of inquiry and using contrasting units of observation. Examples include watershed work plans, soil surveys, national inventories, and censuses. The sampling and storage systems must also be designed so that ready comparisons between the findings of these diverse studies and the flood plain investigations are possible.

In this report we will attempt to devise methods that do no excessive violence to any of these ends. The first part of the discussion reviews

---

1/ Occupance, as used in this report, is defined as the process of occupying or living in an area and the transformations of the original landscape that result.

methods of sampling and proposes and exemplifies ways of collecting data relating to individual flood plains. These collection methods are designed to be directly translatable into handy storage systems. Likewise, the coding is arranged to facilitate analysis needed in both the office and the field, and to make subsequent comparative studies possible. Storage, coding, and analysis are the topics of the second half of the discussion.

Objectives of the flood plain sample are thus to (1) estimate land uses in the flood plain and its adjacent borderland (both of these will be referred to hereafter as the "flood plain"); (2) evaluate changes in these uses through time and space; and (3) study changes in relationships between these uses and associated geographic features of the flood plain, considering also its relative location.

SAMPLING METHODS

Sampling is a well-established method whereby part of a whole is selected to reduce the cost, increase the speed and scope, and improve the accuracy of estimates relating to the whole; see Cochran (4). 2/ For obvious reasons such estimates are subject to error. Whether or not a sample gives results which are representative of the whole depends on keeping sampling errors small and the sample unbiased.

Common causes of bias include deliberate selection of "typical" cases, "convenient" substitution of sampling units by observers, and failure to cover the whole of a chosen sample. The only certain way to avoid bias is to see that each member of the population under investigation has an equal chance of being included in the sample. There is one possible exception. In studies of change, bias will be removed from the estimates of change by using the same sample for both "before" and "after" studies, provided the bias is constant through time; see Goodall (7).

Given a proper sampling method, the probability of occurrence of errors of any magnitude can be calculated from information obtained in the sample. By extension, the relative efficiency of different sampling methods can be obtained. If random sampling errors are large, there are several ways in which they can be reduced. For example, the size of the sample can be increased, since, other things being equal, random sampling error is proportional to the square root of the number of observations. (See fig. 1.) Alternatively, restrictions such as stratification might be imposed, to eliminate from the sampling error differences between strata. These refinements need not be elaborated here, since excellent discussions of them are available in standard textbooks; see Cochran (4) and Yates (21). The present discussion must focus on special kinds of samples of observations arrayed in two-dimensional space.

The universe of study consists of all flood plains in the United States; the population which is sampled comprises all locations on some particular

---

2/ Underscored numbers in parentheses refer to Bibliography, page 27.

95% LIMITS OF PROPORTION FOR SIMPLE RANDOM SAMPLE

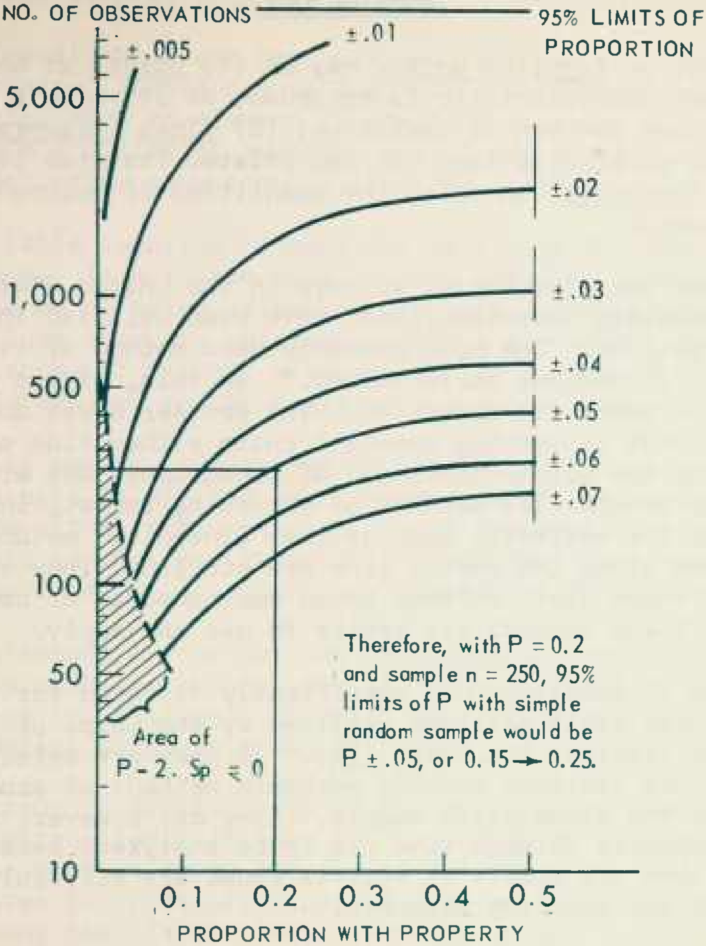

Figure 1.--This chart records the 95-percent confidence limits of a proportion p for different sample sizes n, assuming that the population N is large. Under such conditions these limits are given by p plus or minus 1.96 S_p, where $S_p = (p.1-p/n)^{\frac{1}{2}}$; Cochran, (4, chap. 3). Thus, if a random sample of size n=200 is taken, and a proportion of 0.4 of the observations has some particular characteristic, the 95-percent limits are plus or minus 0.07, and the true P is expected to lie between 0.33 and 0.47, although there is a 2.5-percent chance that P is less than 0.33 and another 2.5 percent chance that P exceeds 0.47. If the sample is increased to n=600 the limits are plus or minus 0.04 when p is 0.4, and even with n=2400 the limits are plus or minus 0.02.

The shaded area in the lower left section includes all points which, for the given sample size, lead to p-1.96 S_p of less than zero, an impossibility. This is therefore the area for which a sample is too small for meaningful estimates of the possible range of P, even with a 2.5-percent risk that P will be meaningless.

flood plain; and the sample will always consist of elements which are parts of this flood plain.

Sampling Units

The elements, or sampling units, may be (1) points at which the presence or absence of some characteristic is recorded, or at which a value is read from some continuous pattern of variation; (2) lines (traverses), the length of which lying on particular land use and related features is of interest; or (3) small areas (quadrats) in which the quantities of occupance characteristics present are measured.

Traverses are used for forest surveys in the United States and in the form of the "windshield investigations" have been utilized in agricultural land use investigations. The most commonly used method of timber cruising is the so-called "10 percent strip method." In this, strips 1 chain (66') wide and 10 chains apart are traversed. The cruiser moves down the middle of the strip and records everything one-half chain either side of his course. Strips are laid to run across the grain of topography, not with it. Cain and Castro (2) record alternative methods of surveying forest, including the 10 percent strip and its variants, such as the "line-plot" method in which blocks selected at random along the survey line are studied. They show a distinct preference for "random plot" methods based upon quadrat or point sample, however, because these methods are easier to use and apply.

Another type of sampling unit specifically designed for studies of flood plains is the sample cross section, utilized by the Corps of Engineers and the Soil Conservation Service, Schneider (16). If they are selected by a proper method, sample cross sections provide workable methods of achieving most of the objectives of the flood plain sample. They do, however, present certain difficulties if changes through time are to be analyzed, because coding and storage systems that are simple as well as sound are difficult to design with cross sections as the sampling elements.

Quadrat methods are used by plant ecologists, in particular. An excellent review of their advantages and limitations is to be found in Greig-Smith (8). In general, they may be used to achieve all the objectives of the flood plain sample, but results are subject to problems of "modifiable units." As Duncan (5) indicates, when sampling units vary in size, or when the choice of size is arbitrary, other sizes can always be taken and different results will be obtained. This is because local variability is averaged out as successively larger quadrats are used as sampling units.

The simplest of the alternative sampling units, points, have none of the above problems. If points are used as units of observation, all the objectives of the flood plain sample may be achieved, and both coding and storage are facilitated. Hence, the following discussion will deal exclusively with point samples.

Areal Samples

The required flood plain sample is of points in area. Thus, it may be referred to as an areal sample, or alternatively as a geographic, plane, or two-dimensional sample. Many such samples have been proposed; see Quenouille (15), Cochran (4), Yates (21), Krumbein (12), and Greig-Smith (8). The number of alternative areal samples is large because of the following considerations:

(a) A two-dimensional space has both latitudinal and longitudinal axes. Any element of an areal sample may be located by a coordinate reference to each axis. Usually, such an element is obtained by selecting sample pairs of coordinates. Since a different sampling method may be used to provide the coordinates on each axis, many combinations are possible.

(b) The available sampling methods for each axis include both random and systematic sampling. In a random sample every possible coordinate value is given an equal chance of being selected, and each value is chosen independently of other values that are obtained. With a systematic sample the first coordinate is selected at random, after which the locations of all other coordinates are determined on the basis of a regular interval.

(c) The sample may be independent or stratified. To obtain a stratified random sample the axis is divided into several parts, and coordinates are selected at random within these strata. For a systematic stratified sample, separate systematic samples are taken within strata. Such systematic samples may be aligned (as in a "square-grid" or "checkerboard" pattern) or unaligned.

(d) Sample elements may or may not be clustered, such that the location of a single sample point determines the location of a set of others, or they may be hierarchical (multistage), in which detailed sampling is done within fewer, larger sampling units selected by higher order samples.

Figures 2 through 6 depict a few of the more common combinations and describe their construction. In each of these figures point samples are presented, but the methods apply equally to selection of quadrat samples. Figure 7 illustrates two possible kinds of traverses, systematic and random, which are also among the alternatives discussed above.

Advantages of Various Areal Samples

Both theoretical and empirical analyses have been undertaken to determine which of the many possible types of areal samples is to be preferred. Quenouille (15), after providing expressions for the variance of random and systematic samples--both simple and stratified, and independent or aligned--concluded that provided the autocorrelation function of successive strata is concave upwards, not only do systematic samples provide the most precise estimates of the mean, but that least satisfactory variance estimates are given by simple random samples, followed by stratified random and systematic samples in that order. Moreover, unaligned samples are more efficient than aligned. The most efficient areal sample is, thus, the unaligned stratified systematic.

Osborne (14) provides empirical support for upward concavity of autocorrelation functions in land use surveys. Furthermore, in a review of the studies which have been completed to compare alternative areal sampling systems, Cochran (4) concludes that "systematic sampling shows a gain in

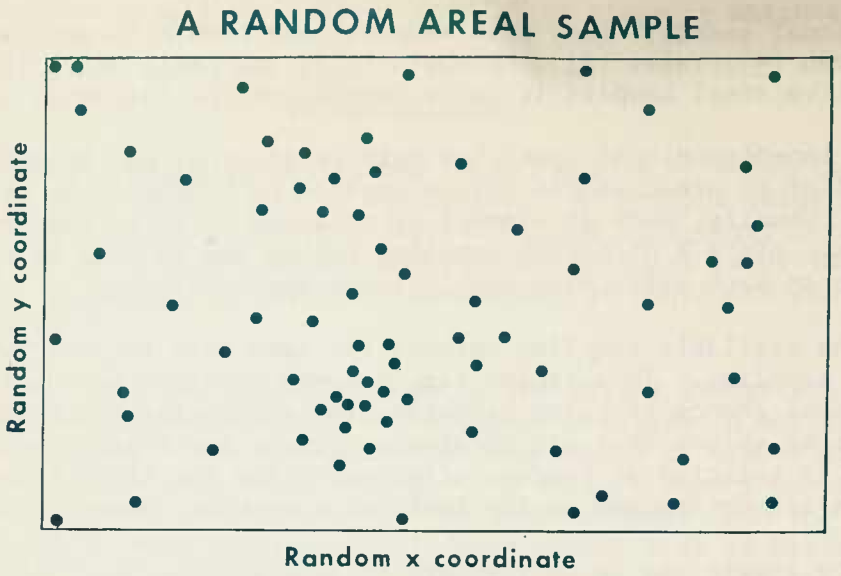

Figure 2.--In this sample each point is selected by entering a table of random numbers and selecting two such numbers, one within the range of coordinate values of the ordinate, the other within the range of coordinate values of the abscissa. The random ordinate and abscissa values thus obtained locate the sample element. Note the uneven areal coverage which can emerge.

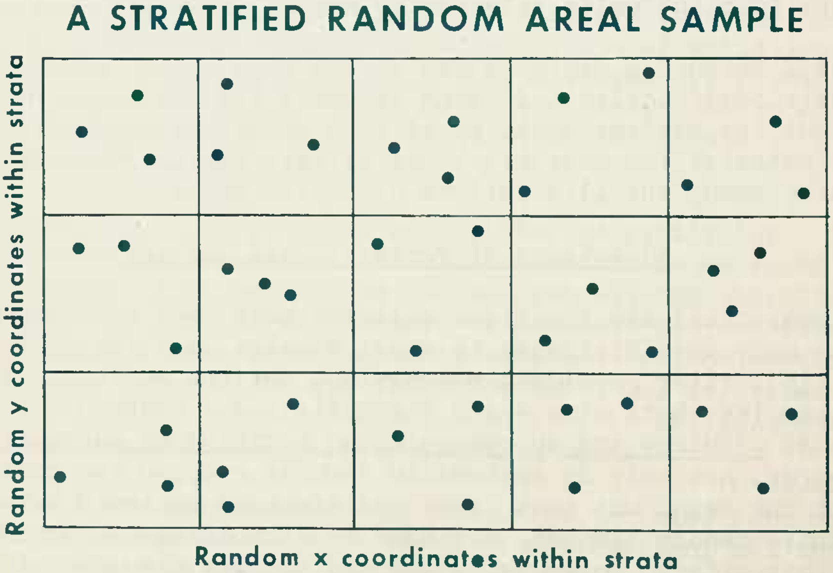

Figure 3.--In a stratified random areal sample, elements are located by selecting random pairs of coordinates within blocks of the larger area. The example has three points so located within each block, although variable sampling proportions could have been used such that the number of points is larger within more variable blocks, or within blocks which have more of the phenomena of interest. Note that areal coverage is better than that of simple random areal sample.

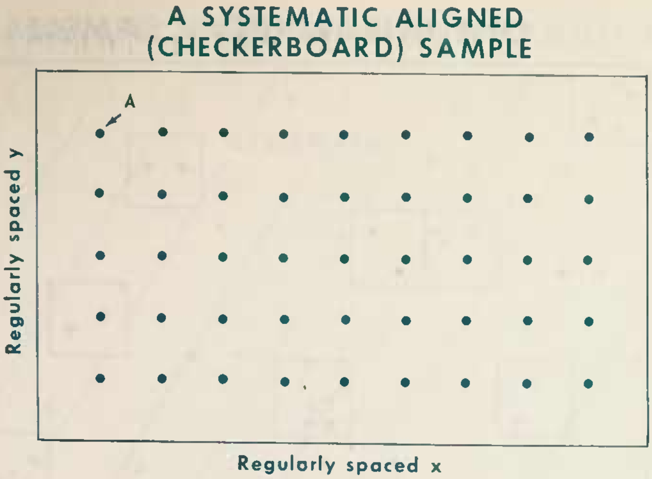

Figure 4.--A checkerboard sample has a perfectly even spread of points, with regular spacing on both abscissa and ordinate after point A has been located at random. But such a selection procedure implies that all parts of the study area do not have an equal chance of being included in the sample. Furthermore, if there are periodicities in the data being collected, the regularly spaced points could hit the same point on a cycle time and again, and give completely biased pictures of the spatial variations of phenomena under study.

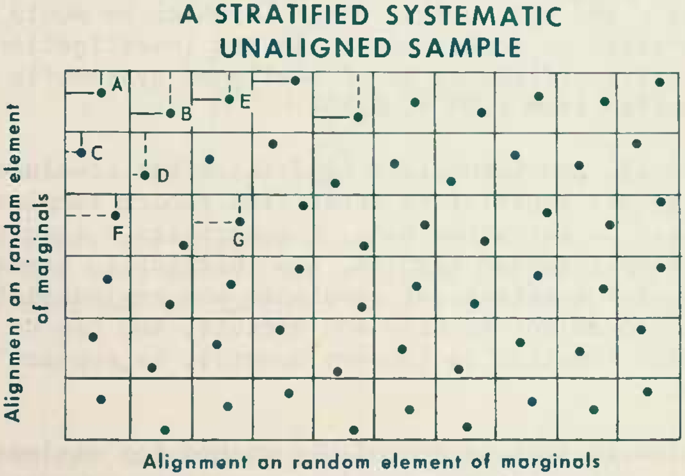

Figure 5.--The preferred areal sample is the stratified systematic unaligned sample. It is constructed as follows: First, point A is selected at random. The x coordinate of A is then used with a new random y coordinate to locate B, a second random y coordinate to locate E, and so on across the top row of strata. By a similar process the y coordinate of A is used in combination with random x coordinates to locate point C and all successive points in the first column of strata. The random x coordinate of C and y coordinate of B are then used to locate D, of E and F to locate G, and so on until all strata have sample elements. The resulting sample combines the advantages of randomization and stratification with the useful aspects of systematic samples, while avoiding possibilities of bias because periodicities are present.

ONE HEIRARCHICAL SAMPLE

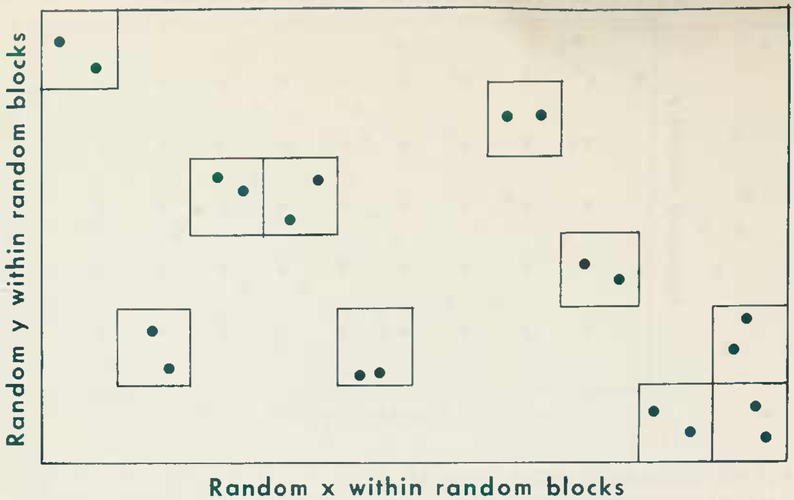

Figure 6.--Blocks are selected at random with random pairs of coordinates, and then points are selected at random within the blocks so obtained.

precision which, although modest, is worth having ... the gains are largest for the types of data [land use, soils, etc.] in which we would guess the variations would be nearest to continuous." In the investigations reviewed by Cochran, the relative efficiency 3/ of unaligned systematic over stratified random samples varied from 1.83 to 5.83.

Both Cochran (4) and Quenouille (15) therefore conclude that unaligned systematic samples are superior to stratified random samples in studies of two dimensional space. On the other hand, checkerboard systematic samples may be only as good as simple random samples, and inferior to stratified random samples, because of the effects of gradients and periodicities. Systematic samples are both convenient to draw and execute, and can be recommended when the autocorrelation function is concave upwards, as appears to be the case in land use surveys.

A disadvantage is that no dependable method for estimating the variance of the means from systematic samples is known, because systematization implies lack of equality of opportunity of places being included in the sample. This has led, frequently, to the assertion that systematic samples should not be used. For example, "if sampling is systematic an estimate of the mean is available which may ... deviate less from the true value than that given by

---

3/ Relative efficiency is defined as variance of the stratified random sample divided by variance of the systematic unaligned sample. Thus, if the relative efficiency is 2.0, the variance estimate of the stratified is twice that of the systematic, or the systematic is twice as efficient.

TRAVERSES: SYSTEMATIC AND RANDOM

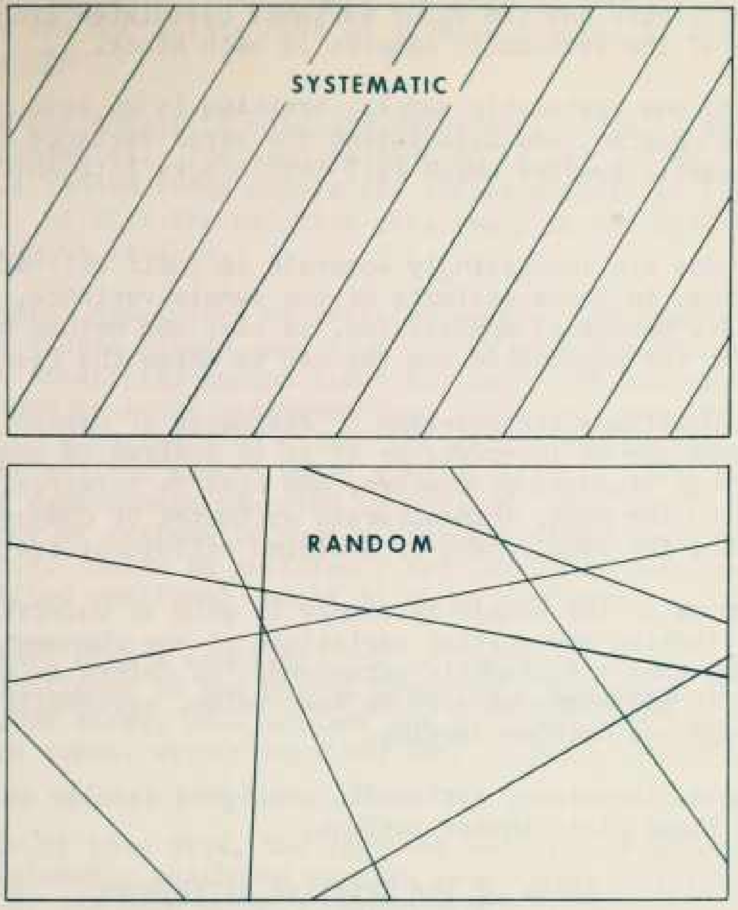

Figure 7.--Traverses are another variety of areal sample, often used in forest surveys. It has been shown that unbiased estimates of the percent coverage of various uses may be obtained from a traverse, as the percentage of length of line.

random samples, but there is no indication of its precision and no possibility of assessing the significance of its difference from the mean of another area," according to Greig-Smith (8, p. 21).

These considerations may not be absolutely insurmountable, however. Expressions for the variance of systematic samples are suggested by both Yates (21) and Cochran (4) and, after his theoretical analysis, Quenouille (15) proposed three different methods of estimating the variance:

(a) By using sets of systematic samples randomly placed with respect to each other, with the error variance calculated from the variances of the systematic samples in each block.

(b) By using one set of systematic samples randomly placed, with the area then broken into blocks and the error variance calculated from the variances of the portions of the systematic samples in each block.

(c) By using one systematic sample, breaking it up into several systematic samples of wider spacing, and calculating the error variance from the portions of the sub-systematic samples which fall into blocks into which the area is divided.

These three methods are increasingly accurate in their estimate of the mean, increasingly biased in their estimate of the sample variance, and decreasingly difficult in their practical application, so that the method of sampling may vary according to the population and the use to which the results are to be put.

These considerations are relevant if estimates of percent of area occupied by particular uses are of interest, or if it is desired to compare areas. When comparisons through time are of concern, and bias is relatively constant from one time period to the next, then accurate estimates of change can be obtained without satisfying the requirement of unbiased estimates; see Goodall (7).

If the purpose of the sample is simply to gain an understanding of the geographic distribution and spatial variations of any phenomenon over an area, there is no doubt that a systematic arrangement of points is preferable; Krumbein (12). In addition, as will be seen below, systematic patterns facilitate storage and machine mapping.

On all counts, therefore, systematic unaligned samples appear to be most advantageous in flood plain investigations.

Tests of the Relative Efficiency of Sampling Systems

Tests of the relative efficiency of sampling systems were undertaken to ascertain whether the above conclusions could be supported in practice, in flood plain studies. One check was made in the Coon Creek intensive study area, 4/ and another in the Montfort area, previously studied by Finch and Platt (6).

In the Coon Creek area (about 10 square miles) a map of land use was available, and is shown in figure 8. Planimetered estimates of the proportions of total area occupied by different types of land use were 40.8 percent woodland, 32.5 percent cropland, 22.5 percent pasture, 2.0 percent gallery, and 2.2 percent other uses. Four stratified systematic unaligned samples were taken, randomly oriented with respect to each other, as recommended by Quenouille (15). One of

---

4/ Analyzed in detail by Burton, in a companion study (Burton, Ian. Types of Agricultural Occupance of Flood Plains in the United States. Research Paper No. 75. Department of Geography, University of Chicago, 1962).

these is shown in figure 9. Each sample contains an average of one point per 10 acres, so that the 660 points have an average spacing of one-eighth of a mile or 660 feet. 5/

Estimates of percentage of woodland cover were 40.49, 40.96, 40.24, and 41.07 in the four samples. The mean is thus 40.69 percent and the variance is 0.17. In a simple random areal sample the expected variance for a sample of this size is 3.66, so that the relative efficiency of the systematic sample over the simple random is 21.5.

For a meaningful comparison of relative errors, the area was stratified into quarter-mile square blocks, and four points were located at random in each block to create a stratified random areal sample. This was repeated four times. The percent woodland coverage estimates from these four samples yielded a mean of 41.4 percent and a variance of 0.96. The relative efficiency of the systematic over stratified random areal samples is thus 5.65, much the same as in the results of experiments reported by Cochran. Similarly, for pasture the relative efficiency of systematic over stratified random samples was 2.3 and for cropland it was 3.4. Gains in efficiency are consistent and useful. A systematic stratified unaligned sample is to be preferred.

In the Montfort area a complete detailed land use survey was undertaken, and the results published in the volume Geographic Surveys in 1933 (6). Of the total area of 29,396 acres, 55.4 percent was in cultivated land. Relatively small samples were taken, with n equaling 184, or about one point to every 160 acres.

With a sample of this size, the expected variance of simple random samples is 13.4. Four systematic unaligned samples were taken, and they yielded an overall estimate of the mean of 54.7 percent, with a variance of 10.2. The variance of four stratified random samples was 11.3, of four checkerboard systematic samples 12.8, of systematic traverses 13.5 (50.5 percent, 48.7 percent, 51.4 percent, 57.2 percent), and of random traverses 11.0. Although gains in efficiency are less impressive than in the Coon Creek case, the systematic unaligned sample again provides the most efficient of the sampling methods, even when compared with traverses.

Sample Size

The relative efficiency of systematic unaligned samples over stratified random samples is greater than one, and often as high as five. This implies that fewer observations are needed with an unaligned systematic sample than with other methods to obtain estimates with a given variance. Let us assume that to obtain estimates which have similar variance the systematic sample has to be only one-fifth the size of a simple random sample. Then for a p of 0.1

5/ 660 feet was chosen as an average spacing because, as will be seen in fig. 9, a grid of that density was placed over the land use map. However, such a spacing approximates the 1,000 feet suggested as somewhere near the optimum for soil surveys by Youden and Mehlich (22).

COON CREEK: LAND USE

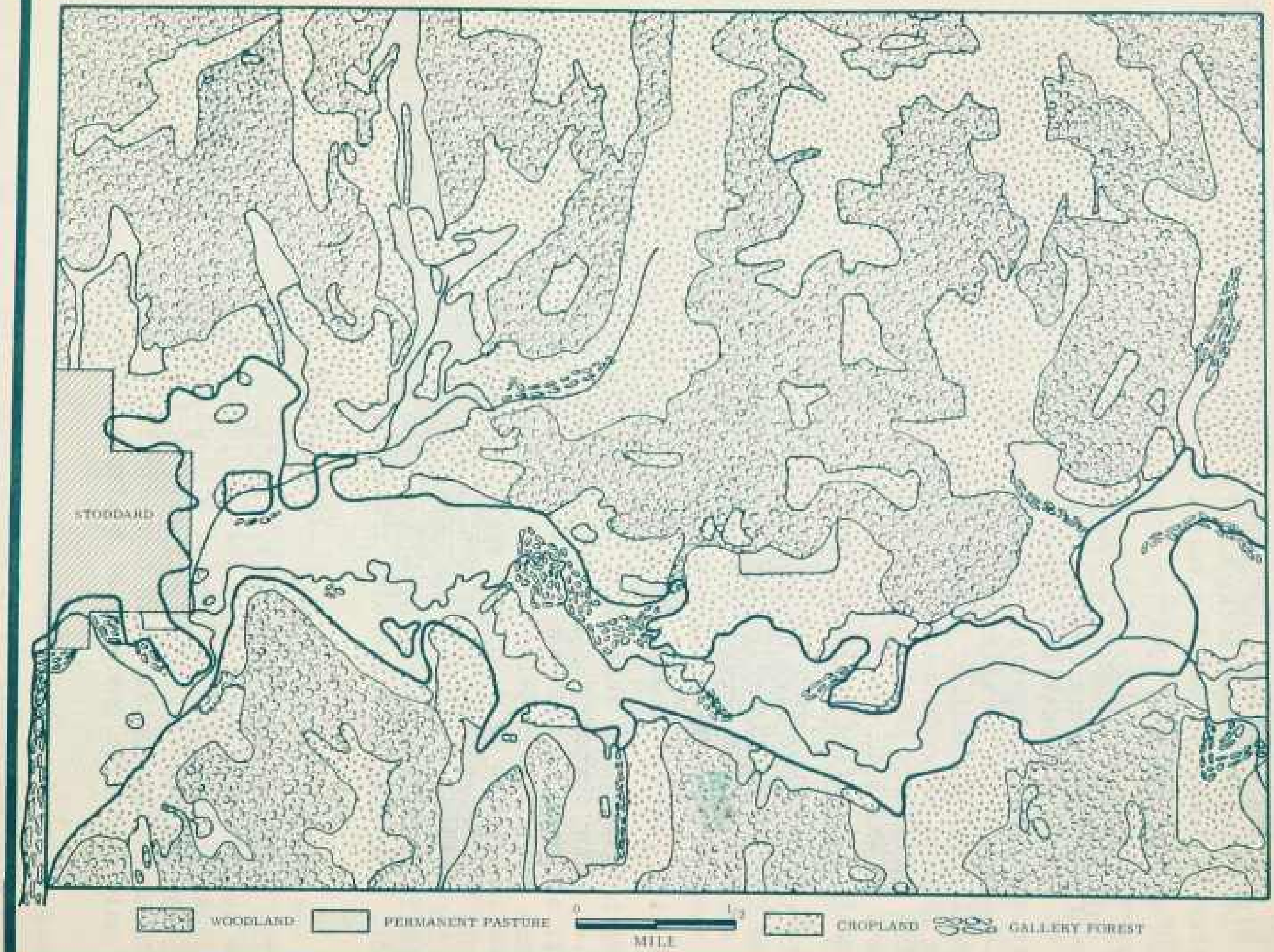

Figure 8

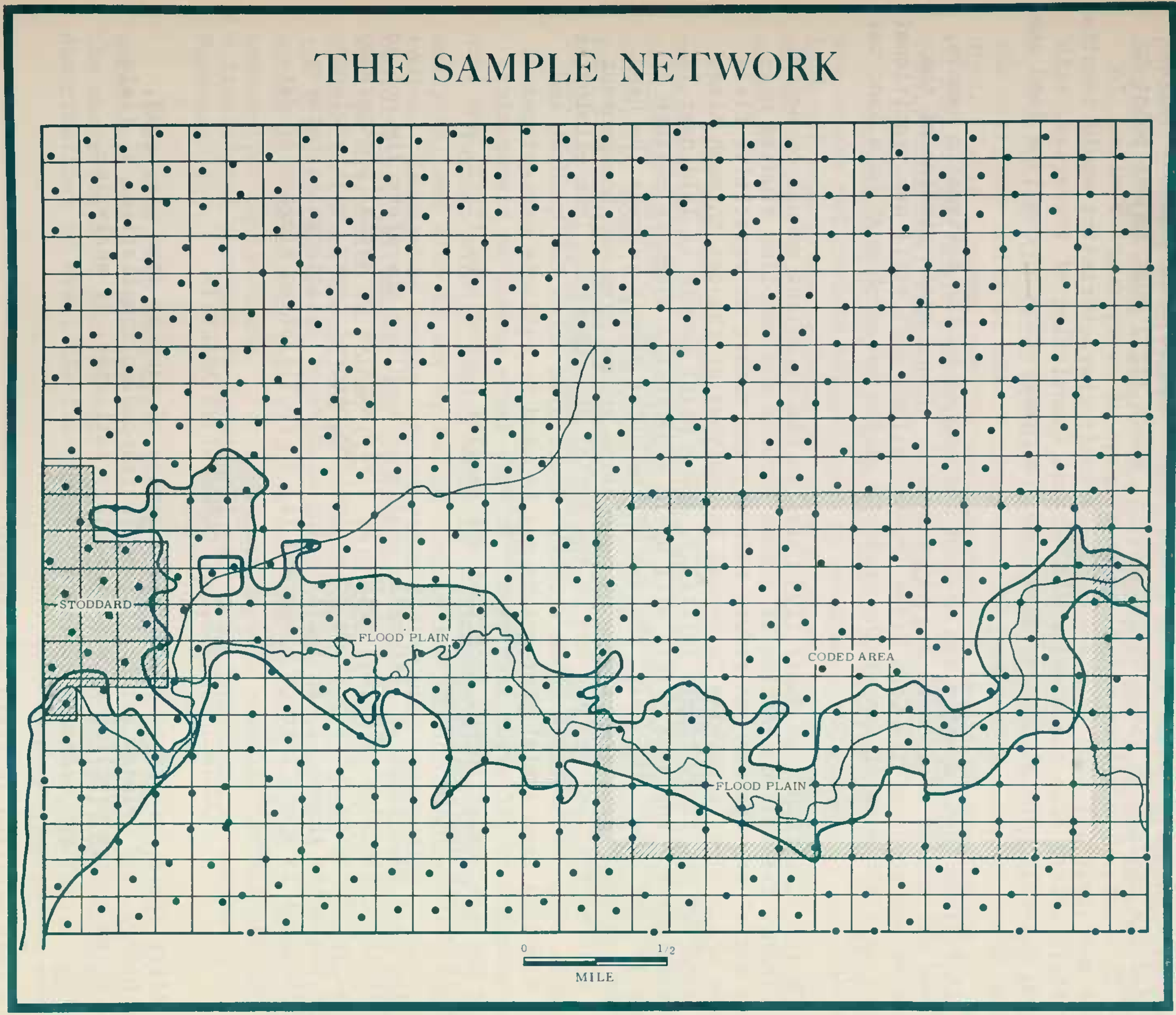
THE SAMPLE NETWORK
STODDARD
FLOOD PLAIN
CODED AREA
FLOOD PLAIN
0
1/2
MILE

Figure 9

to have 95 percent limits of plus or minus 0.02, the systematic sample can be of 180 observations rather than the 900 of the simple random sample (fig. 1).

In the Coon Creek area of 10 square miles this means 18 observations per square mile rather than the 90 the simple random sample would demand. While considerable savings in field effort result, there still must be one point to every 35 acres, with points spaced at an average of every quarter mile. To locate and visit all such points and record their characteristics would require more field effort than the usual field methods (consisting of traverses, with mapping directly onto air photographs, and the preparation of complete land use maps). 6/

It follows that, despite the savings introduced by the systematic sample, improved methods of field work are not provided. Flood plain studies of the character and scope of the Coon Creek investigations must still use traditional methods of field work if one of the products of the research must be a land use map. 7/

On the other hand, as Spurr (17) notes in his excellent manual on aerial photography, careful combination of ground surveys and sampling studies utilizing aerial photographs can reduce markedly the number of repetitive field studies which might otherwise be needed in analyses of changes through time. If a land use map is available initially, and a point sample is laid over it, then readings for the same points will yield a perfectly adequate measure of change in land uses. Moreover, the use of the point sample removes the need for the hazardous practice of drawing boundaries around presumed homogeneous areas both on the land use map and in the aerial photographs. This eliminates one whole range of possible human errors. Spurr also notes how quadrat and point sampling from aerial photographs may be used to estimate percentages of areas occupied by different land uses. In very large areas, an hierarchical sample can also be arranged, such that one samples photographs first, and then elements within the photographs.

The real advantages of systematic unaligned sampling therefore lie not so much in the collection of data, whether by field work or by aerial photography but (a) in providing quick methods of estimating percent cover, with variance of the estimates known; (b) in facilitating studies of relationships between distributions in space through time; and (c) in facilitating storage of data and mapping by machine.

CODING, STORAGE, AND COMPARATIVE ANALYSIS

After field workers have provided a series of maps of land use, soil, flood hazard, and related variables, and corresponding compilations of field data, there emerge problems of making summary estimates, of analyzing

---

6/ A fact verified in field experiments.

7/ Of course, this is not to imply that other kinds of studies, for example where farms have to be visited and interviews made, will not benefit from use of samples. These remarks apply to flood plain studies specifically, and to similar intensive land use analyses in general.

distributions, of evaluating relationships between distributions, and of storing the mapped data for subsequent comparative analysis.

A series of systematic unaligned point samples may be taken, being certain that the grids in each case are randomly oriented. From these samples, the percent coverage of different phenomena, with their related variances, may be obtained in the usual manner. Coding, storage, and analysis problems remain.

Coding

Coding is the first problem to solve and depends in turn on the nature of the data, the units of observation, and the level of measurement which has been used. Table 1 summarizes the various possible levels of measurement or types of scaling and their characteristics. In flood plain studies two of these levels occur most frequently, nominal and ratio scaling.

Nominal scaling is used when the units of observation are points which are or are not certain things (such as woodland or flood plain) and have or have not some property (such as slope in excess of fifteen degrees). It is normal to code such points 1 for presence and 0 for absence of the property in question.

Ratio scaling is used when the units of observation are quadrats or management units, or when points have associated with them readings from some continuous surface. Various quantitative statements may be made about such units, for example area devoted to different land uses, outputs, and inputs. Such statements, on ratio scales, are coded in their full numeric terms. Once coded, both types of data may be analyzed in a variety of ways.

Geographic Coding

A further kind of coding which is of fundamental importance to any spatial analysis, and also to any system of storage and machine mapping, is geographic coding. Geographic coding implies attaching to each observation a pair of coordinates, latitudinal and longitudinal, that assign it to a unique location; see the discussion in Tobler (19). This kind of coding is now being introduced to Swedish census operations on an experimental basis, using a 10-meter grid on the 1:10,000 map series. 8/ All houses, farms, fields, etc., may be located within five meters on this grid, in addition to being located by their usual census-type designations. For a discussion with examples see Hagerstrand (9); a traditional approach to locational coding is described, by contrast, in Houseman (10).

Once data have been coded geographically both the conventional and unconventional storage systems described below may be used. Just as important, the whole field of machine mapping and automatized analysis of geographic distributions is available for use by the research worker, as Tobler (20) shows

---

8/ A special subcommittee of the Census Advisory Committee of the Association of American Geographers is beginning preliminary feasibility studies for the United States as well.

Table 1.- A classification of scales of measurement

Scale	Basic empirical operations	Permissible statistics	Examples
Nominal	Determination of equality	Number of cases Mode Phi and chi-square	Assignment of type, as "Woodland" "Flood plain"
Ordinal	Determination of greater or less	Median Percentiles Order correlation	Hardness Grades of severity
Interval	Determination of the equality of intervals or differences	Mean Variance Product-moment Correlation	Temperature (F or C) Potential energy Calendar time) no absolute zeros
Ratio	Determination of the equality of ratios	Geometric mean Harmonic mean Percent variation	Length Numerosity Density Time Kelvin temperature) have absolute zero

Source: S. S. Stevens, "The Psychophysics of Sensory Function," _The American Scientist_, 48 (1960), 226-53.

so clearly, with only one problem remaining. If latitude and longitude are used as coordinates, accurate maps of large areas are difficult to prepare on machine, because of the usual problems of representing the curved surface of the earth on flat paper. It is therefore desirable to have orthogonal two-dimensional grids available for purposes of geographic coding, similar to the military coordinate systems or the British Ordnance Survey grid. In the United States, the State Plane Coordinates appear to provide such a required framework for geographic coding of large areas, enabling local study areas to be fitted into large regions; see Mitchell and Simmons (13). On a world scale, other grid systems may be constructed to achieve the desired mappings; see Bailey (1).

Storage

Several methods of storing data are feasible once these data have been collected and coded. The data can be left in the form of maps or on appropriate aerial photographs, with point sample and grid system superimposed. Such a method has many advantages, but subsequent quantitative analysis is not possible, and most research workers are aware of the dangers attendant upon simple visual comparisons of maps in lieu of more sophisticated methods of analysis.

A second way to store the observations is by making conventional use of punched cards. In such a system each observation is assigned to one or more data cards. Recorded on the cards are the location of the observation, both traditional (State, county, sub-county unit, etc.) and geographic (latitude and longitude), followed by the various bits of information (scores on variables) available about the observation. To compare distributions when data are coded in this manner requires a pass of the entire deck of cards through an appropriate tabulating or computing machine. These machines may be programmed so that they will use the locational coding information to produce a map of any required distribution in addition to comparing distributions, although the row-column arrangement of the printing devices attached to computing machines means that data have to be mapped in a square-grid pattern; see Tobler (19). 9/

The most efficient sample for estimating percent cover is one of points in a systematic unaligned pattern. This sample also may be combined with <u>unconventional</u> use of punched cards to provide an alternative and extremely

---

9/ Tobler (20) shows how, if one has observations on altitudes, for example, these may be mapped by the machine, and then contoured in the usual way. However, some simple processing of the altitudes will produce an <u>equation</u> which represents the contoured surface, and this may be stored even more completely. The first derivative of this equation, when mapped, yields a map of slope; the second derivative yields curvature. Hence, once stored in compact form, elementary mathematical operations produce maps of interesting characteristics of distributions available only at the expense of considerable effort if appropriate compact storage methods are not used.

useful storage system that maintains the spatial properties of the map, yet provides quick methods for comparative analysis of distributions both in the office and the field.

A conventional storage system allocates each observation to a card, and information relating to the standing of the observation on many variables is stored in that card. But the card also provides a square grid (conventionally of 12 rows--Y, X, O, 1, 2, 3, ... 9--and 80 columns). This could well represent 960 points in a systematic point sample spread over an area.

Let us assume that each cell of the card does represent an observation. Then if a characteristic is present at that point, the hole is punched; if not, the cell remains untouched. Only one card is needed for each characteristic, rather than one for each observation, with real advantages of compactness. Observations occupy the same position on each of these cards. Of course, the conventional system must still be used when data have been collected on something other than a nominal scale.

One variable might be "woodland." The woodland card will have holes punched in it when the points in the sample fall on woodland, and will be left untouched where points fall on something other than woodland. If the card is held up to the light, then the pattern of holes is the geographic pattern of woodland in the study area. For example, a sample section of the Coon Creek study area has been enclosed in figure 9. This area contains 10 x 14, or 140 sample points. The land use at these points is tabulated in figure 10.

Figure 10 shows four cards punched in the unconventional manner described above, one of the map of woodland, one of cropland, one of pasture, and one of flood plain. A comparison of figure 9 with figure 10 will reveal that the only distortion introduced is that the unaligned points of the stratified systematic sample are forced into the square grid shape of the card. The cards are not the usual 12 x 80 cards, but 10 x 40 cards, especially scored so that they may be punched on a simple inexpensive hand punching board 10/ while in the field (hence only 40 columns, to maintain strength of the scored cards). They are also labeled so that use (variable) and areas (locational identification for the standard card) may be entered at the top of the card.

It is entirely possible to prepare these storage cards while in the field. Prior to the field survey, the coordinate reference system and the preselected systematic point sample should be drawn on the aerial photographs and maps. The area should then be broken into a master set of 10 x 40 units, representing cards. These units should be given special grid designations, such as area A (latitude) 1 (longitude) in the top left hand corner, followed by A2 to the right of it and B1 below it, etc. Stable base information which can be derived from published sources, such as slope and soil type, should be mapped and punched into the 10 x 40 cards, which will be provided for the field workers.

10/ The punch board slips into a jacket pocket, is light, and costs only $7.50. Cards cost only $3 per thousand after initial costs of the plate have been met. Cards in figure 10 were prepared especially for the Department of Geography, University of Chicago, to be used in field work and field training.

LAND USE: COON CREEK SAMPLE

Sample Observations

	2	4	6	8	10	12	14	16	18	20	22	24	26	28
0	W	W	W	W	W	W	C	C	C	W	W	W	P	C
1	C	W	W	C	W	C	C	C	C	P	W	C	C	P
2	W	C	W	W	W	W	W	W	W	G	P	P	P	P
3	W	C	W	P	C	W	W	W	C	C	P	P	P	P
4	W	C	C	C	C	C	W	W	C	C	C	P	P	P
5	C	P	P	C	P	P	P	P	P	C	P	P	P	P
6	C	P	P	P	P	C	P	P	P	P	P	P	P	P
7	P	P	P	P	P	P	P	P	P	P	P	W	W	C
8	W	W	P	P	P	P	P	C	C	C	W	W	W	C
9	W	W	P	P	P	P	C	C	W	W	W	W	C	C

C = Cropland G = Gallery P = Pasture W = Woodland = Flood Plain

Data Storage, Punched Cards

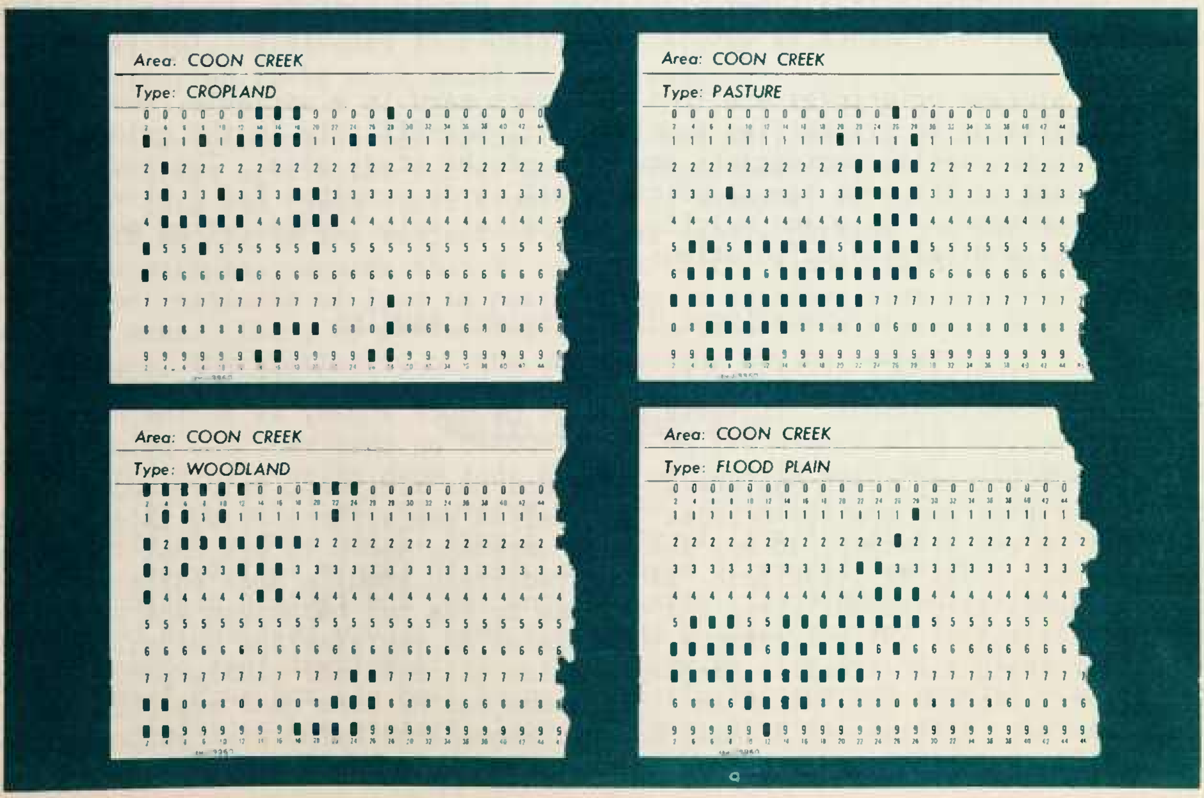

Figure 10

The cards, placed in envelopes and mounted on a board in the A1, B1, A2, ... pattern, constitute the initial storage system.

Field work proceeds in the usual fashion, and maps are prepared. Then, additional cards may be made up by the field workers, one for each type of data collected, and the cards filed. The master maps and aerial photographs, with coordinates, card-grid, and point sample on them, plus the board and envelopes of cards, constitute the storage system.

Analysis

The stage is now set for cross-sectional analysis. To compare any two geographic distributions, all that is required is to pull out the appropriate cards and lay one over the other. A two-by-two contingency table may be prepared in which coincident punches, coincident lack of punches, and punches on one card but not on the other are indicated. Phi (ϕ) then provides the coefficient of correlation, indicating the degree of association of the distributions, and chi-square (χ^2) provides a test of whether the association is significantly greater than could have arisen by chance.

For example in figure 10, woodland, pasture, cropland, and the flood plain have been punched into appropriate storage cards. In an analysis by optical coincidence (as suggested by Campbell and Caron (3) table 2 was produced. Forty-one out of 140 points, for example, are both pasture and flood plain. Phi between pasture and flood plain is 0.694, which is significant at the 0.001 level. Very little flood plain land is not in pasture, so the fact that correlation is not higher is due to the presence of pasture off the flood plain.

The storage principles are thus that each card is a variable, and each one of its positions is a specified sample point, identifiable in the grid system overlaying the aerial photographs and maps of the study area. The same position belongs to the same sample point on all cards prepared for a given area. Analysis proceeds simply and quickly, by optical coincidence, using contingency tabulations and related calculations.

Comparisons With Previous Studies

Other studies may provide data which can be compared with the results of the field investigations. If these studies produce a map, as in the case of a soil survey, the grid system may be superimposed on this map, and point sample data extracted. (We have already suggested that much of this information could be provided in a preliminary storage system to the field worker). Figure 11 is such a map of soil types. In figure 12 the sample points which were used to construct figure 10 have been used to sample soil types. Table 3 lists the soil types and their letter designation, and table 4 presents all the characteristics of these soils and the areas they occupy that could be extracted from descriptions in the soil survey report. It is evident that comparisons can be made between the spatial distribution of land use and such items as physiography, terrain, drainage, depth of soil, texture and materials of soil, and degree of acidity.

Since there is no reason why such data could not be coded and stored before embarking on field work, field workers could be the ones to compare

Table 2.- Correlation of pasture and flood plain land

Item	Pasture present	Pasture absent	Totals
	Points	Points	Points
Flood plain present--------	41 a	4 b	45 C
Flood plain land absent----	17 c	78 d	95 D
Totals-------------------	58 A	82 B	140 N

$$\phi = \frac{ad - bc}{\sqrt{A\ B\ C\ D}} = 0.694$$

$$\chi^2 = N\phi^2 = 67.4, \text{ significant at beyond 0.001 level}$$

results of their own work with such stable or base data. Comparisons of this kind inevitably raise questions which can only be answered in the field; hence more penetrating field work would be fostered.

If other studies do not produce maps, but only totals or averages relating to census areas or similar survey units, comparisons may still be made between the coded data and these other studies. For example, if the area mapped in figures 10 and 12 is the subject of another study, perhaps one for which a watershed workplan records that 95 percent of the flood plain is in pasture, the percent estimate of land in pasture may be obtained from the point sample for that area. Then, by performing a simple test for difference between the percentages, the similarity of the observations may be compared.

Analysis Through Time

If the stored materials and the grid system are preserved through time, then similar point samples may be taken at later dates and the nature and degree of land use or other changes, whatever are of interest, may be compared by exactly the same methods described above. Similarly, if later aggregate estimates are presented for all or part of the area, the correct point estimate for comparative analysis may be obtained by summarizing the information for the points lying within the area for which aggregate estimates are presented. Comparisons are then made in the usual way. As Spurr (17) suggests, real savings occur in studies of changes through time if appropriate combinations of field survey, sample field check, aerial photography, and sample analysis

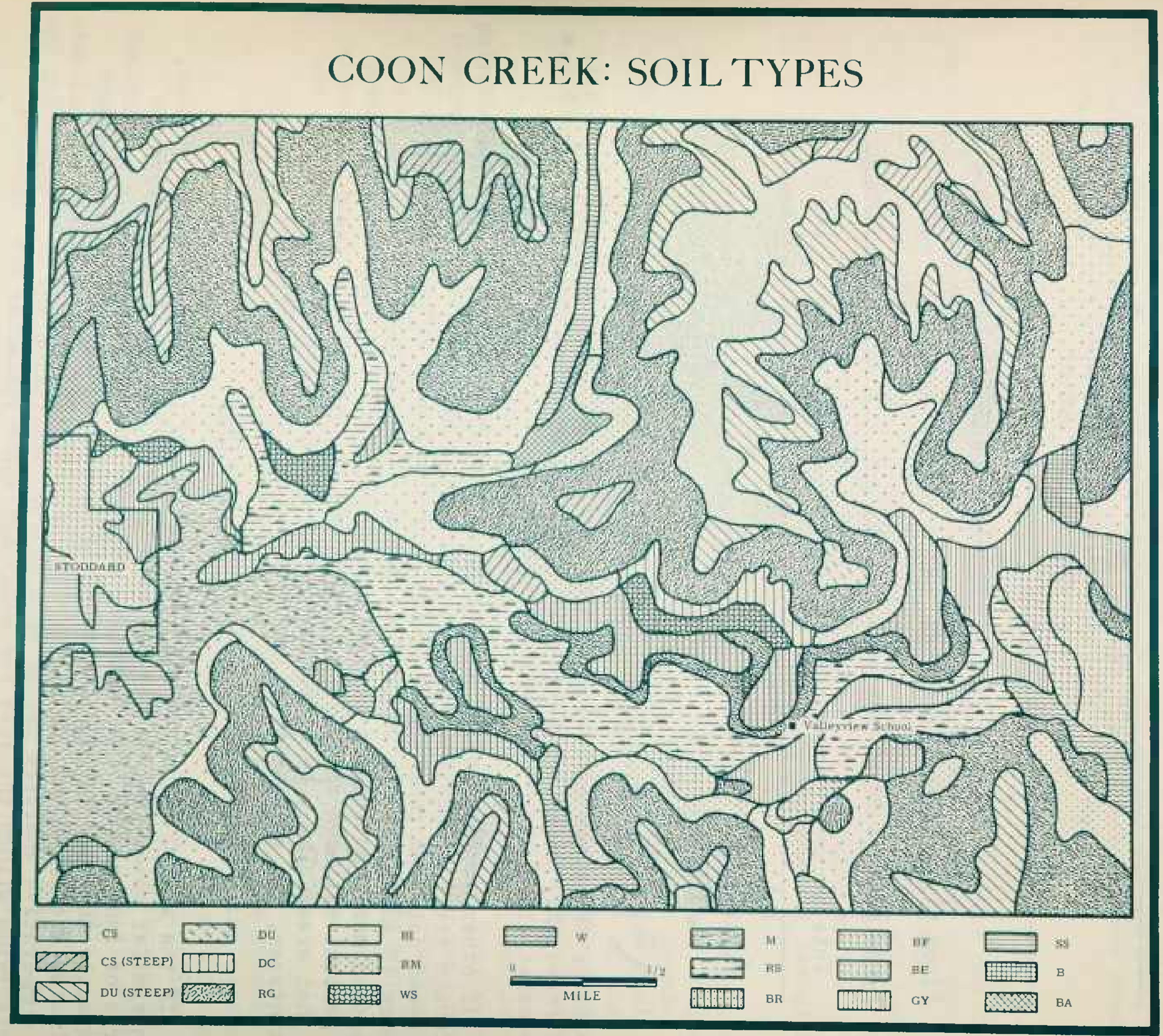
COON CREEK: SOIL TYPES
STODDARD
Valleyview School
CS
CS (STEEP)
DU (STEEP)
DU
DC
RG
BM
WS
W
0
1/2
MILE
M
RS
BR
BF
BE
GY
SS
B
BA

Figure 11

SOIL TYPES: COON CREEK SAMPLE

Sample Observations

	2	4	6	8	10	12	14	16	18	20	22	24	26	28
0	Rb	Rb	Rb	Cs	Du st	Rb	Bm	Bl	Bl	Rb	Rb	Bl	B	Rb
1	Cs st	Rb	Rb	Rb	Du st	Cs	Cs	Rb	Bm	Br	Bm	Bl	Gy	Gy
2	Rb	Ws	Rb	Du	Rb	Rb	Rb	Du	Rb	Br	Gy	Bf	Gy	Be
3	Bm	Ws	Rb	Rb	Dl	Dl	Rb	Rb	Bm	Br	Gy	Rs	Bf	Be
4	Br	Br	Ws	Rs	Rb	Ws	Rb	Rb	Rb	Br	Rb	Rs	Rs	Bf
5	Br	Br	Rb	Rs	Br	Br	Rs	Gy	Rs	Rb	Rs	Rs	Gy	Bf
6	Rs	Rs	Rs	Rs	Br	Br	Gy	Rs	Rs	Gy	Rs	Rb	Ws	Ws
7	Rs	Bm	Rs	Rs	Rs	Rs	Gy	Rs	Rs	Rb	Rs	Rb	Rb	Bm
8	Bm	Rb	Bm	Bf	Gy	Gy	Bf	Bf	Bl	Br	Bm	Rb	Du st	Rb
9	Bm	Rb	Du st	Bm	Bm	Bl	Bf	Rb	Bl	Rb	Rb	Du st	Rb	Bm

SEE TABLE 3 FOR EXPLANATION OF CODES.

Data Storage, Punched Cards

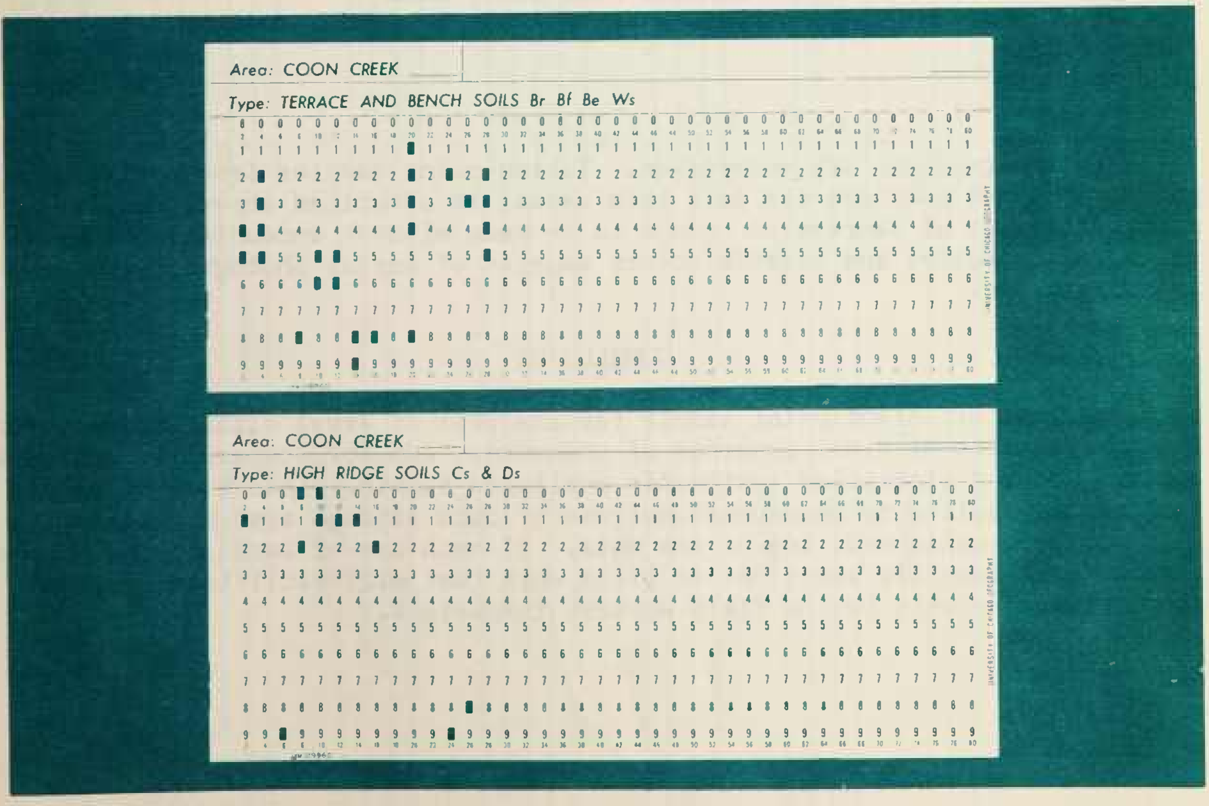

Figure 12

Table 3.- Soil Types

Code	
Cs	Clinton Silt Loam
Cs steep	Clinton Silt Loam, steep phase
Du	Dubuque Silt Loam
Du steep	Dubuque Silt Loam, steep phase
Rb	Rough Broken Land
Bm steep	Boone Loam, steep phase
Bl steep	Boone Silt Loam, steep phase
B	Boone Fine Sandy Loam
Ba	Bates Silt Loam
Br	Bertrand Silt Loam
Bf	Bertrand Fine Sandy Loam
Be	Bertrand Loam
Ss	Sparta Sand, Brown Phase
Ws	Waukesha Silt Loam
Rs	Ray Silt Loam
W	Wabash Silt Loam
Gy	Genesee Fine Sandy Loam
M	Meadow

of aerial photographs are undertaken. Taking point samples from aerial photographs and punching the data into cards, as described above, enables ready estimates of degrees of change to be made without encountering some of the more subjective and unsatisfactory aspects of aerial photograph interpretation.

CONCLUSION

The most efficient point sample for analysis of areal distributions, for making estimates of areal coverage of phenomena, and for comparative analysis, appears to be a stratified systematic unaligned sample. This kind of sample is admirably suited to unconventional use of card storage systems which maintain geographic ordering of the data. If such use of cards is made there appear to be feasible additions to field work that will facilitate immediate comparative analysis by the field workers themselves.

Table 4.- Soil characteristics: Coon Creek "intensive" survey 1/

Soil description	Cs	Cs steep phase	Du	Du steep phase	Rb	Bm steep phase	Bl steep phase	B	Ba	Br	Bf	Be	Ss	Ws	Rs	W	Gy	M
<u>Location</u>																		
High ridges	x	x	x	x														
Low ridges						x	x											
Valley slopes					x	x	x	x	x									
Terraces and benches										x	x	x	x	x				
First bottom lands, along streams and drainageways															x	x	x	
Bottom lands																		x
<u>Slope and Terrain</u>																		
Broken land, slopes often 30 percent					x													
Steep phase, 15-30 percent		x		x		x	x											
Gently undulating or rolling	x		x					x	x									
Level to gentle undulating										x	x	x	x	x	x	x	x	x
<u>Drainage</u>																		
Excessive, low water holding capacity, subject to drought			x	x	x								x					
Good to fair	x	x				x	x	x	x	x	x	x		x				
Imperfect																	x	
Poor															x	x		x
"Subject to overflow"															x			x
<u>Depth</u>																		
Shallow			x	x	x		x											
To 36 inches						x		x		x	x	x	x				x	
To 5 feet														x	x	x		
To more than 5 feet	x	x							x									
<u>Texture</u>																		
Floury soft	x	x		x														
Fine, friable			x					x			x	x	x		x	x		
Medium																		
Heavy						x	x			x				x			x	x
<u>Materials</u>																		
Stony					x													
Sandy						x		x			x		x				x	x
Silty	x	x	x	x		x	x	x	x	x	x	x			x		x	x
Clayey	x	x	x	x			x								x			
<u>Acidity</u>																		
Alkali																		
Neutral																x		
Acid	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x

1/ See table 3 for explanation of soil type codes.

BIBLIOGRAPHY

(1) Bailey, H. P.
1956. Two Grid Systems That Divide the Entire Surface of the Earth Into Quadrilaterals of Equal Area. Amer. Geophys. Union Trans. 37, 628-35.

(2) Cain, S. A., and G. M. de Oliveira Castro.
1959. Manual of Vegetation Analysis. New York: Harper.

(3) Campbell, J. D., and H. S. Caron.
1961. Data Processing by Optical Coincidence. Science, 133, 1933-8.

(4) Cochran, W. S.
1953. Sampling Techniques. New York: Wiley.

(5) Duncan, O. D., and others.
1960. Statistical Geography. Glencoe: The Free Press.

(6) Finch, V. C., and R. S. Platt.
1933. Geographic Surveys. Bul. 9, Chicago Geog. Soc.

(7) Goodall, D. W.
1952. Some Considerations in the Use of Point Quadrats for the Analysis of Vegetation. Austral. Sci. Jour., Res. Ser. B, 5, 1-61.

(8) Greig-Smith, P.
1957. Quantitative Plant Ecology. New York: Academic Press.

(9) Hägerstrand, T.
1955. Statistika Primäruppgifter, Flykartering och Data Processing Maskiner. Meddelanden Fran Lunds Geog. Inst. 344, 233-55.

(10) Houseman, E. E., and T. J. Read.
1954. Application of Probability Area Sampling to Farm Surveys. U. S. Dept. Agr., Agr. Handb. 67.

(11) King, A. J., and R. J. Jessen.
1945. The Master Sample of Agriculture. Amer. Statis. Assoc. Jour. 40, 38-56.

(12) Krumbein, W. C.
1955. Experimental Design in the Earth Sciences. Amer. Geophys. Union Trans. 36, 1-11.

(13) Mitchell, H., and L. Simmons.
1945. The State Plane Coordinate Systems. U. S. Coast and Geod. Survey.

(14) Osborne, J. G.
1942. Sampling Errors of Systematic and Random Surveys of Cover-Type Areas. Amer. Statis. Assoc. Jour. 37, 256-64.

(15) Quenouille, M. H.
1949. Problems in Plane Sampling. Ann. Math. Statis. 20, 355-75.

(16) Schneider, R.
1960. A Preliminary Appraisal of Flood Plain Data. Unpublished report to U. S. Dept. Agr., Agr. Res. Serv.

(17) Spurr, S. H.
1960. Photogrammetry and Photo Interpretation. New York: The Ronald Press.

(18) Tanner, J. C.
1959. Sample Survey of the Road System of Great Britain: Method of Selecting the 1,000 Sample Points. D.S.I.R. Road Res. Lab. RN/3544?JCT.

(19) Tobler, W.
1959. Automation and Cartography. Geog. Rev. 44, 536-44.

(20) ________
1960. Geographic Ordering of Information: An Exploratory Examination. Discussion Paper 38, Geog. Dept. Univ. of Wash.

(21) Yates, F.
1953. Sampling Methods for Censuses and Surveys. London: Griffin.

(22) Youden, W., and A. Mehlich.
1937. Selection of Efficient Methods of Soil Sampling. Boyce Thompson Inst. Proc. 9, 59-70.

U. S. GOVERNMENT PRINTING OFFICE : 1962 O - 654768

UNITED STATES DEPARTMENT OF AGRICULTURE
1862 1962

Growth Through Agricultural Progress