A GUIDE TO AREA SAMPLING FRAME CONSTRUCTION UTILIZING SATELLITE IMAGERY

BY
WILLIAM H. WIGTON [1]
PETER BORMANN [2]

A paper submitted to the United Nations at the request of the United Nations Outer Space Affairs Division based on material used and experience gained at the Second International Training Course in Remote Sensing Applications for Agriculture: Crop Statistics and Agricultural Census.
(Rome: 25 April - 13 May 1977)

ACKNOWLEDGEMENTS

Preparation of this report would not have been possible without the assistance of Dr. H.G.S. Murthy, Expert on Space Applications, United Nations, New York, and Dr. John A. Howard, Head, Remote Sensing Unit, Food and Agriculture Organization, Rome, Italy. Special thanks go to Ms. Berna Spiers, Technical Officer, Remote Sensing Unit, Food and Agriculture Organization, for helping teach the course for which this was prepared.

1/ William H. Wigton, Mathematical Statistician, Consultant to the United Nations, U.S. Department of Agriculture; Economics, Statistics and Cooperatives Service, Washington, D.C. 20250

2/ Peter Bormann, Geophysicist, United Nations, Outer Space Affairs Division. New York, NY 10017

CONTENTS

SUMMARY

The area sampling frame is a basic technique for collecting agricultural statistics for a quick and comprehensive agricultural information system. It is used in a number of countries to estimate all types of agricultural products as well as economical parameters such as prices and labor for the current year. This methodology provides accurate information by taking representative samples from only a small part of the total land area. Estimates can be available five to six weeks after the beginning of the data collection. These estimates are based on an objective statistical method of data collection and evaluation.

The construction of the area sampling frame is carried out in several steps. The first step is the delineation of broad areas of homogeneous land use/land forms using all types of available data and maps of the most recent date such as satellite imagery, aerial photography, topographic and/or land use maps. Areas of the same land use type form a stratum. Once these strata have been formed, one must find boundaries for them that are identifiable on the ground, such as roads, footpaths, railways and rivers. These boundaries are then marked on the map in a unique way for each stratum and the areas within each stratum are labelled.

The next step is to divide these homogeneous strata into sample units. Normally, this is done in two steps: primary sampling units (PSU's) are delineated and a small sample of PSU's are selected to be further subdivided into sample units (SU's). Again, good boundaries must be obtained on the map at each level.

One sample unit is selected from each PSU and the selected SU is called a segment. The segments vary in size depending on stratum, land use, and

population density. The general rule is that they should be small enough to be enumerated in one day. In an agricultural area a typical size is 1 km^2.

The construction of the area sampling frame ends with the selection of segments that represent the total area. Again it must be ensured that these segments have clearly recognizable boundaries that will leave the field enumerator with no problems in deciding which area is inside and outside the segment.

The desired data are then collected from these segments, usually by interviewing the farmers, measuring crop acreages and making crop cuttings. Since the segments within each stratum are statistically representative of this stratum, the results collected from these segments can be expanded to the total area of the stratum. The desired production figures for a country are obtained by summing the results for the strata of that country.

I. INTRODUCTION

Each country needs accurate, timely information on its agricultural production for proper management of its food reserves, import and export planning and many other planning activities. There are different procedures to obtain this information. All countries have some kind of agricultural production data system but there is a general need to improve them. Some of the systems depend on compulsory reporting or complete enumeration, while others rely on statistical samples.

The first kind of system works effectively and accurately in centrally planned economies while the statistical sampling system is efficient and accurate in an economy where the individual farmer decides what he plants and may change his decisions corresponding to changing market conditions. A statistical sampling frame is effective where agricultural production estimates are needed for very large areas in a short time.

This report will deal exclusively with statistical sampling using area sampling frames. In order to be accurate, a statistical sampling frame must follow strict scientifically based procedures. A random sample will give an accurate estimate of the characteristics in a population provided the characteristics in the sample are representative of the characteristics in the population. Since the agricultural characteristics of interest are not homogeneously distributed across the whole country, the population must be divided into homogeneous areas with respect to the agricultural characteristics under study.

Imagery of the earth's surface taken from space provides a synoptic overlook of large areas and is, therefore, very useful to delineate these homogeneous areas. There are different satellite systems available. For more details on sensor characteristics see United Nations document A/AC.105/204[1].

Photographic systems provide imagery of very high spatial resolution (some 10 to 20 meters) but normally they do not provide complete and repetitive coverage of large territories. Scanning systems, as flown in the automatic LANDSAT satellites have a smaller spatial resolution (LANDSAT 1, 2 and C, some 80 meters and LANDSAT D some 30 meters), but they provide a global and repetitive data coverage every 18 days or 16 days, respectively. With more than one satellite in orbit at a time and a suitably chosen spacing between them, the time interval between repetitive coverage of the same area can be shortened significantly. Since the availability of LANDSAT data is extensive, this report will emphasize the use of LANDSAT data for the construction of the area sampling frame.

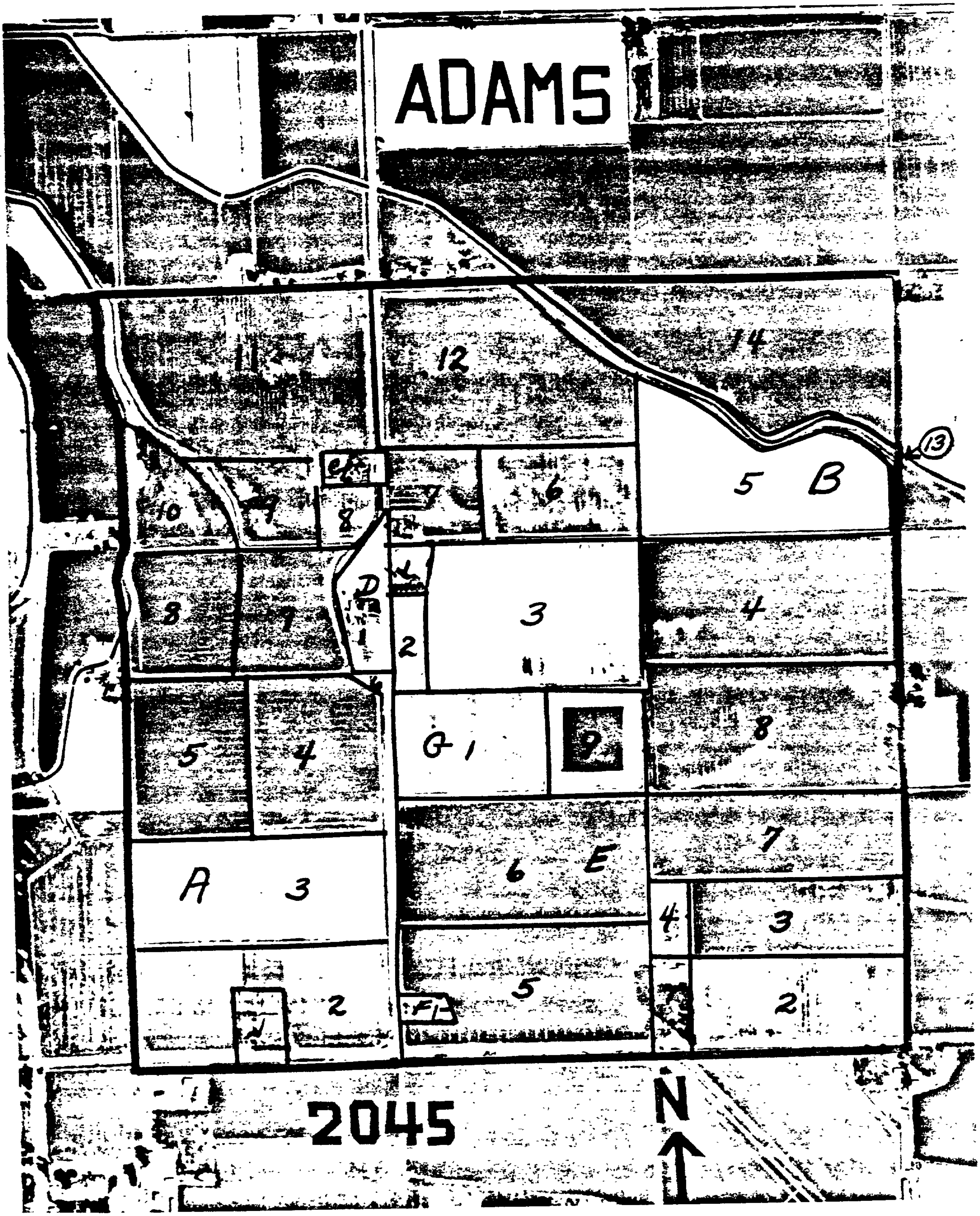
II. CONSTRUCTION OF THE AREA SAMPLING FRAME

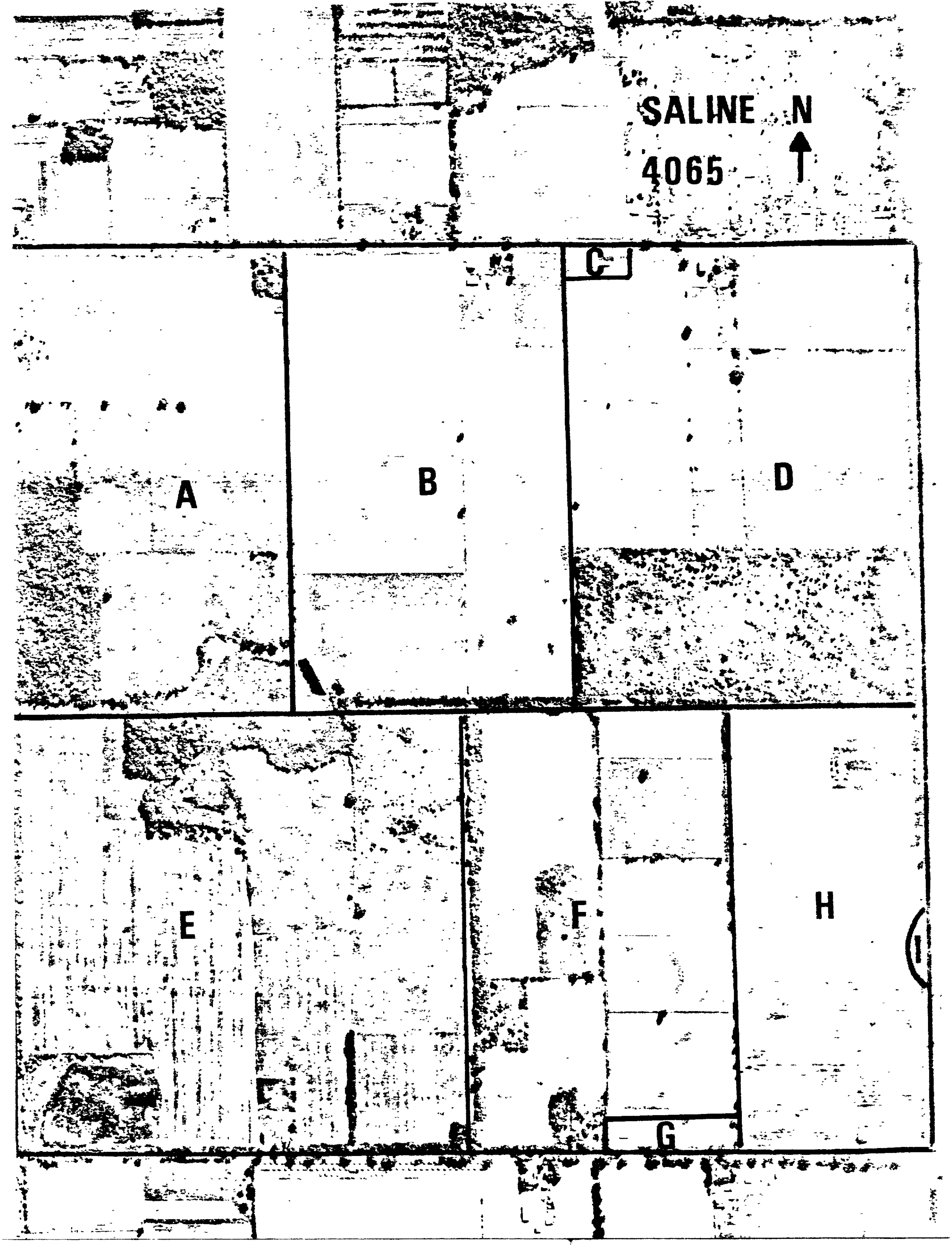
A. Explanation of terms used.

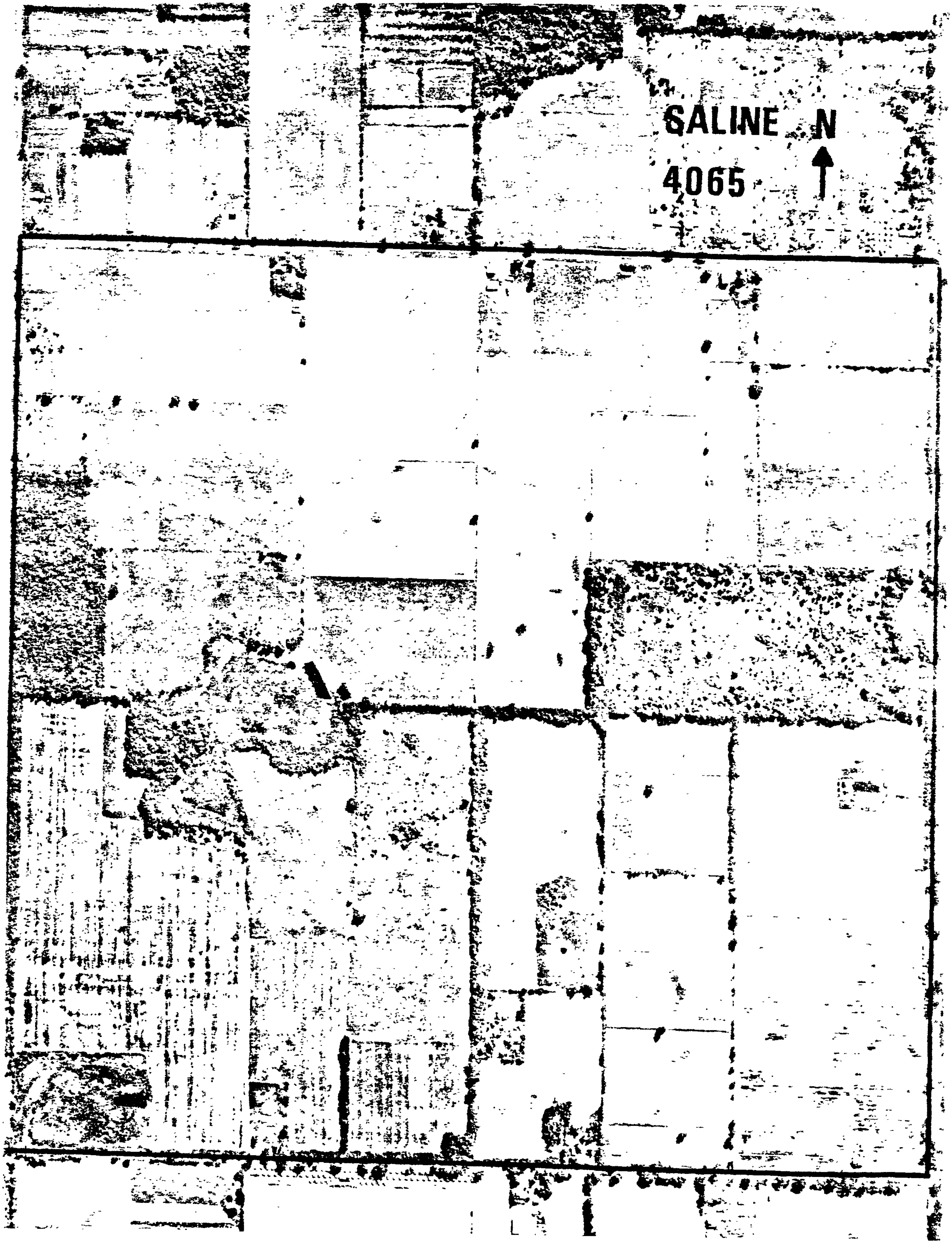
The area sampling frame (ASF) is the total land area of a country broken down into N small parts called the sampling units (SU's). Out of these N sampling units a number n will be randomly selected for enumeration. The selected sampling units are called segments. Figure 1 shows as an example Adams County segment No. 2045. These segments must be completely enumerated by personal interview. An interviewer must go to the area on the ground and locate the owners and operators of all land inside the boundaries of the segment. The data collected by the interviewer must be recorded in a suitable way. Questionnaires used by the Economics, Statistics, and Cooperatives Service of USDA in the United States may serve as an example. Copies are given in Appendix A.

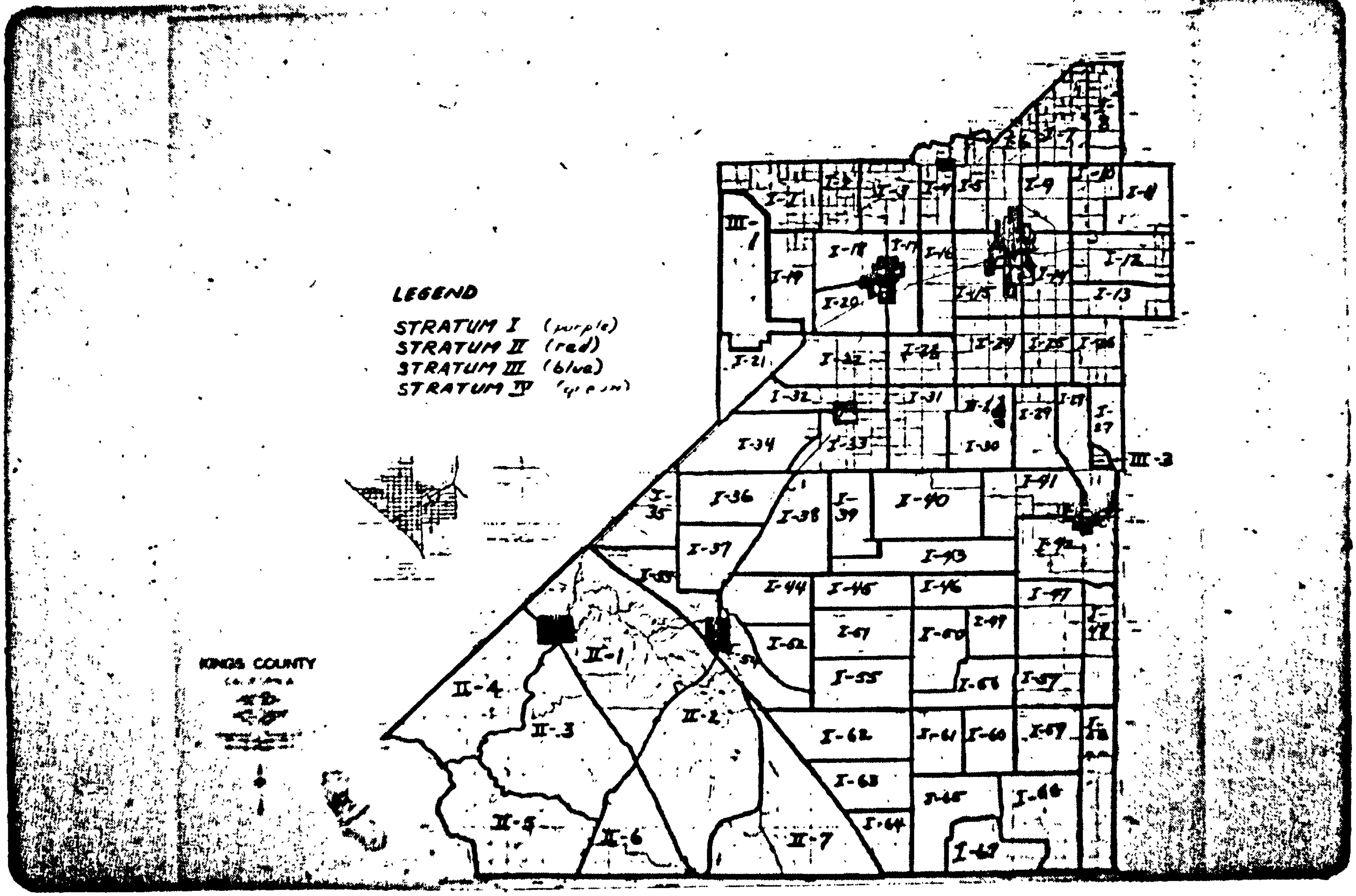
In the process of constructing the area sampling frame there are several steps which will be demonstrated in a more detailed way in the following chapters. The first step is to divide the total land into homogeneous areas with respect

FIGURE 1




SALINE
4065
N
A
B
C
D
E
F
G
H
I


SALINE N
4065


LEGEND
STRATUM I (purple)
STRATUM II (red)
STRATUM III (blue)
STRATUM IV
KINGS COUNTY

to the agricultural characteristics under study. These homogeneous areas are called strata. These strata will then be subdivided into intermediate parcels of land called primary sampling units (PSU's) to each of which a certain number of sampling units will be assigned. For example, the United States is broken down into approximately 3,000,000 sampling units and a sample of only 16,000 segments is selected to be interviewed. This means that only some 0.5% of the total land is sampled. Nevertheless, careful frame construction and interview procedures provide timely and sufficiently accurate estimates of the total agricultural production. The total is estimated by expanding the data collected from the n segments by the proper expansion factor N/n.

B. Survey priorities, resources and difficulties.

Before one begins the frame construction process, one should have a clear idea of which specific agricultural products are most important and must be emphasized. It is also important to know how accurate these estimates must be and how soon they must be available. Hopefully, once these questions are answered, the relative order of importance will remain fairly stable over time and the various crops and livestock items will not switch priority positions relative to each other.

Another item that needs to be determined is how much money will be available to develop the area sampling frame as well as to conduct each necessary survey. In general it is much easier to devise a frame for a single crop than it is for a general purpose survey where a long list of priority items are being estimated, but one usually does not have this luxury, so many compromises are necessary.

The total money available to conduct the survey determines the overall sample size n and the specific list of agricultural items that are to be emphasized determines how these n segments will have to be distributed among the individual strata. The allocation of segments to strata is done according to one of several schemes outlined in most sampling books.

Normally, in agricultural surveys, most of the available segments are allocated to the intensively cultivated strata so that most of the money is spent obtaining information in those areas which have the biggest share in the total agricultural production.

Also, important for deciding on the total number of segments to selected is the determination of the optimum segment size for each stratum. A "rule of thumb" which can be used is that the enumeration of a segment be accomplished in one day. The obvious points to consider when deciding on segment size are density and structure of the population, length of questionnaire and type of tranportation used. The structure of the population is a very subtle point is referred to in statistics books as the intra-class correlation coefficient.

In practice the careful execution of the survey plan will meet with a number of difficulties, the first being that the segments must be easy to locate in the field. This problem will be dealt with in more detail in Section II D.

Once the segment has been exactly located in the field, one has to be able to extract the wanted information. In a cattle census, for example, one would need to count the cattle in the segment. This may be possible although difficult. If, on the other hand, one is interested in an insect survey, one would have to count all the insects in the segment. This task would be impossible to implement.

Some other items that are difficult to estimate are the potential agricultural production of a country, the crop production when enumerating early in the growing season of that crop, the hectares of crops when farmers do not know their own hectarage, or the number of cattle when farmers are unwilling to tell (perhaps they must pay taxes on a per head basis). If one is confronted with problems of this type, one should do serious thinking before starting to construct the frame or the results will be disappointing.

Finally, be sure that the interviewers are qualified and willing to do a good job. Training and regular instruction of the interviewers as well as quality checks of their work are indispensible.

C. Material and facilities needed.

For the construction of the area sampling frame one needs topographic and/or land use maps or aerial photography or both. Satellite imagery can be useful supplementary material. Within reason, the more material one collects and uses in the construction of the area frame, the better the finished product. All materials used should be of the most recent date.

The scales of maps and photographic products used must be related to the size of the features on the ground. Not many countries in the world have sufficiently large field sizes to allow the recognition of single fields and their boundaries in LANDSAT imagery while, on the other hand, changes in land use pattern stand out clearly in that type of data. LANDSAT imagery can hardly be used at scales larger than 1:250,000 and fields smaller than 8 hectares are difficult to identify. LANDSAT imagery, therefore, will preferably be used in the stratification process while aerial photography at scales between 1:20,000 to 1:60,000 is particularly suitable for stratification as well as the delineation of the sampling units and their physical boundaries.

High resolution imagery from photographic satellite systems such as those flown in SOYUZ 22 and Salyut 6 can provide basic maps at scales as large as 1:50,000. Such imagery could therefore replace or supplement standard aerial photography and also be used in areas with small fields (down to 0.4 hectares).

Besides imagery and maps, light tables, magnifying glasses, various colored pencils and suitable maps storage space are needed for carrying out the construction of an area sampling frame.

D. Sequence of constructing the area sampling frame.

1. Stratification and boundary selection.

Construction of the area sampling frame is carried out in several steps. The first step is the delineation of broad areas of homogeneous land use/land forms using all types of available data as outlined in the previous section.

Areas of the same land use form a stratum. The number of strata into which the total population/land area can be or should be subdivided depends largely on the variety and distribution of various land use types in the areas under study, the significance of their visual differences in satellite imagery or aerial photography, the skill of the photointerpreter and the goal of the survey. If we wish to have 7 strata, we could, as an example, subdivide an area into water, forest, cities (inner city), urban agriculture (suburbs), intensively cultivated land, less intensively cultivated land, and non-farmland, such as recreational areas, deserts, high mountain areas and military bases.

Cities and towns are often difficult to delineate on satellite imagery and all types of administrative and political boundaries are normally not visible but the latter can easily be derived from maps.

Many times one crop may be so important and cover large contiguous areas that a separate stratum is set up for that particular crop. In principle, however, the land-use stratification should be more general so as to accomodate different types of surveys for a number of years. Land areas smaller than 5 km^2 should not be separated out as a different stratum even though they might not fit the stratum definition.

The delineation of the strata on LANDSAT imagery should be completed without regard to physical boundaries on the ground. This allows the photo-interpreter to concentrate on pattern recognition and differentiation. Another reason is that small physical boundaries such as country roads, footpaths, railroads, and small rivers can normally not be seen in imagery with a ground resolution coarser than some 30 m.

When transferring their strata from imagery onto maps of larger scales (scales of 1:20,000 to 1:50,000 are best suited) their initial boundaries might have to be changed slightly to coincide with physical boundaries on the ground that are easy to recognize and follow. Unique colors and roman numbers should be assigned to all strata and the strata boundaries colored correspondingly.

The need for good physical boundaries applies to all further subdivision of the strata into primary sampling units and sampling units. The importance of this cannot be overemphasized. Most of those sampling frames that do not work well fail becuase of poor enumeration. The key to quality enumeration is to have boundaries which can easily be located by both the interviewer and a supervisor carrying our quality control in the field.

2. Primary sampling unit construction.

The next intermediate step in the construction of the area frame is to subdivide the strata into primary sampling units (PSU's). They vary in size depending on the stratum and the country. Since in the final step a specific number of sampling units (usually some 6 to 20 sampling units) will be assigned to them, they should be small enough to permit subdivision in a short time. However, they should be large enough to be useful for a variety of surveys. A statistician may be needed to help decide the optimum size (see also Chapter II B).

Again, good boundaries must be obtained on the map and marked in the color of the stratum. Primary sampling units in non-contiguous parts of the same stratum must be grouped together and all PSU's be numbered in a unique way, separately for each stratum. Each primary sampling unit can then be identified on the map by the stratum number (roman numeral), its PSU number (arabic numeral) and the size of its area (in km^2). For example, I-3-16 means the PSU number is 3 in stratum I has an area of 16 km^2.

In numbering the PSU's one could begin in the northeast corner and number in serpentine fashion from east to west so as to guarantee that no PSU is left out. The area can easily be measured using a grid or, more accurately by using a planimeter.

After this is done, all the primary sampling units are listed on a PSU identification sheet (see Appendix C). A separate sheet is used for each stratum.

3. Sampling unit construction and primary sampling unit and segment selection.

In order to save time, not all the PSU's will actually be broken down into sampling units; rather, a certain number of sampling units will be assigned to all of them. Only a few PSU's will then be randomly selected for further subdivision into sampling units (SU's). The probability that a given PSU will be selected is proportional to the number of assigned sampling units in it. The most suitable number of PSU's to be selected (equal to the number of segments to be enumerated) depends on considerations as outlined in Chapter II B.

The number of sampling units assigned to a PSU depends on its size. The optimum SU size varies with the land use conditions in the survey area, the survey priorities and resources and the length of the questionnaire (see Chapter II B). It normally differs in different strata. As an example, the optimum SU size is given for two areas, Kings County, California and Salcedo Province in the Dominican Republic (Table 1 shown on page 11). All cities and

Table 1

Area		Stratum	Optimum size of the SU's (in km^2)	Range of tolerance
Kings County California	I	(Intensively cultivated agriculture)	2.5	1.3 - 5
	II	(Rangeland and desert)	15	10 - 31
	III	(Non-agricultural)	25	12.5 - 51
	IV	(Urban)	0.25	0.25 - 0.8
Salcedo Province, Dominican Republic	I	(Intensive agriculture)	2	1 - 3
	II	(Coffee)	2	1 - 3
	III	(Extensive agriculture)	4	3 - 5
	IV	(Non-agricultural land)	4	3 - 5
	V	(Urban)	1/2	1/4 - 3/4

towns, no matter how small, should have at least one sampling unit.

The actual number of SU's assigned to each PSU is determined by dividing the area of the PSU by the optimum SU size, then round the quotient down to the nearest whole number. The number of assigned sampling units in each PSU is then listed on the PSU identification sheet in the column marked "S.U." and their cumulative number for each stratum in the column "Cum" S.U. (see Appendix C).

The whole PSU and segment selection procedure can be summarized as follows:

(a) Pick the random number (see 4) from 1 to N where N is the total number of sampling units in the particular stratum. Compare the random number selected with the cumulative numbers given for each PSU in the PSU identification sheet. The PSU selected for further subdivision into sampling units is the nearest one containing the random number.

(b) Find the selected PSU on the map and divide it into the assigned number of sampling units using the best available boundaries. The actual size of the sampling units to be constructed may vary within the tolerance range.

(c) Number the sampling units in the selected PSU beginning in the northeast corner and proceeding in serpentine fashion as before, select one at random and identify it with the segment number.

(d) Record the segment number, the stratum number, the PSU number, and the number of sampling units in the PSU on a segment location sheet (see Appendix C). The final column on this sheet may be used to record the name of the cities/towns for segments in the stratum "urban" or any other pertinent information.

Since only one sampling unit is selected within each selected PSU, the sample selection procedure may be thought of as two-step single stage rather than two stage cluster sampling.

4. The use of random number tables in the selection of primary sampling units and segments.

(a) Divide the random number sheet (Appendix B) into columns of the size needed.

(b) Count the number of one-digit columns, if any, and number them.

(c) Using another random number table, decide which column to begin with.

(d) Again using a random number table, decide whether to begin at the top or bottom of the column. Mark the start on the random number sheet.

(e) Again using a random number table, decide which column to go to next. Draw an arrow from the first column to the second. If you began at the top of the first column, draw an arrow from the bottom of the first to the bottom of the second.

(f) Randomly select the third column and draw an arrow from the second to the third as before.

(g) Proceed until all the one-digit columns are used up.

(h) Go to the two-digit columns and proceed as above. Continue until all the page is in order.

It may be helpful where you have a number of arrows crossing each other to use a different color for each set of different size columns.

To use the random number table, decide how many digits are in the highest possible number to be selected and use columns of the size. For numbers between one and ten, use a one-digit column (0 is ten). For numbers between eleven and one hundred, use a two-digit column (00 is 100); and so forth. Thus, if you need to select a random number between 1 and 11, go to the two-digit column and select the first number which falls between one and eleven.

As you go down the column, cross off each number considered even though these were not actually selected. In the above example, any two-digit numbers which did not fall between 01 and 11 should be marked off until a number is found. These numbers had a chance of selection and should not be used again. Don't start over at the beginning each time but begin where you left off the previous time. Random numbers are commodities to be used up and discarded.

E. Practical Example.

We have included an example to illustrate the various steps involved in the construction of an area sampling frame.

A more detailed exercise of this type can be found in a paper by Huddleston.[5]

The exercise presented here in Appendix C makes use of a LANDSAT scene for stratification and of maps for boundary selection. A more detailed treatise of this exercise can be found in a paper by Hanuschak and Morrissey.[3]

III. COST AND TIME REQUIREMENTS FOR CONSTRUCTION OF AN AREA SAMPLING FRAME

This section will need to be updated as the technique develops and as we become more experienced. As for materials, the costs vary depending on what is available. LANDSAT scenes can cost between $20.00 and $1,000.00 U.S. dollars depending on the amount of preprocessing and where the images are purchased. Costs of maps and aerial photography vary so much that it would be meaningless to try to generalize that cost. Those particular materials must be priced in each country.

People resources in terms of man/months will be presented as guidelines to follow. No country has ever constructed a sampling frame, selected the segments, developed a questionnaire and trained interviewers in less than a year and more commonly it has taken 2-3 years for a country to run regular surveys and producing agricultural estimates. It takes some time to obtain a trained staff to support an agricultural information system.

There are many variables. Obviously, countries with large land areas require proportionately more time to construct frames for than small countries. Also, intensively cultivated land requires more time than desert land. As a general rule, for the important agricultural areas, allow 1 to 2 man/months per LANDSAT image (32,000 km^2). This allows some time for training also.

Once a trained staff has its materials in hand and is working, the frame construction is quite fast. Consider Table 2 as a guide on page 16.

IV. USE OF LANDSAT DIGITAL DATA FOR CROP ACREAGE ESTIMATION IN NEAR REAL TIME

A. Characteristics.

The satellite data used in this report is the 4-channel LANDSAT Multi-Spectral Scanner (MSS) data. It is described in Section 3 of Data User's Handbook.

The MSS is a passive electro-optical system that can record radiant energy from the scene being sensed. All energy coming to earth from the sun is either reflected, scattered, or absorbed, and subsequently, emitted by objects on earth. The total radiance from an object is composed of two components, reflected radiance and emitted radiance. In general, it may be stated that the reflected radiance forms a dominant portion of the total radiance from an object at shorter wavelengths of the electromagnetic spectrum, while the emissive radiance becomes greater at the longer wavelengths. The combination of these two sources of energy would represent the total spectral response of the object. At the wavelengths of energy LANDSAT monitors, reflected energy completely dominates the picture. This, then, is the "spectral signature" of an object and it is the differences between such signatures which allow the classification of objects using the statistical techniques about to be discussed.

B. Computer Classification Techniques.

Suppose that LANDSAT digital data is available to classify in a computer. This can be done in the computer by use of discriminant functions. Computers must differentiate between crops on the basis of reflected energy. Before starting, a sample of data from two or more crops must be available that represent how those particular crops reflect energy. The problem is to set up a rule using the sample pixels for each crop, which will enable us to allot

Table 2

Estimated Time Required to Construct a Land Use Area
Sampling Frame for Various Countries

Area	Total Area	Agriculture Area	Time to construct area frame
Africa			Man-years
Egypt	1,000,000	30,000	1
Ethiopia	1,220,000	800,000	2
Morocco	446,600	150,000	2
South Africa	1,221,000	1,100,000	4
North America			
Canada	9,970,000	700,000	4
Mexico	2,000,000	1,000,000	4
South America			
Argentina	2,780,000	1,700,000	4
Bolivia	1,098,600	300,000	2
Brazil	8,500,000	1,300,000	4
Asia			
China	9,600,000	3,270,000	4
India	3,300,000	2,000,000	4
Indonesia	1,910,000	300,000	2
Europe			
Bulgaria	110,000	50,000	1
Czechoslovakia	120,000	70,000	1
France	550,000	320,000	2
Hungary	93,000	70,000	1
Italy	300,000	170,000	1
Poland	310,000	200,000	1
Spain	505,000	340,000	2
Yugoslavia	256,000	150,000	1

some unknown crop pixel outside the sample to the correct crop type given only the amount of reflected energy of that pixel.

This can be formulated statistically, but let me introduce some notation.

If all data in a LANDSAT frame were plotted in a scatter diagram it might appear as Figure 2.

Figure 2. Scatter Diagram of All Values in One LANDSAT Frame for Three Crops. C-Corn, S-Soybeans, W-Water.

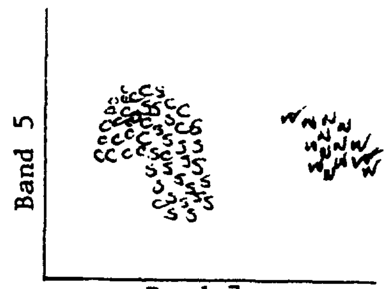
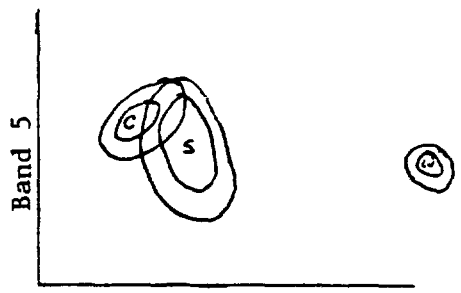


Figure 3 shows confidence limits for above data.

Figure 3. Confidence Limits for Data in Figure 2.



If one studies Figure 2, the following observations can be made:

1. The location of the center of these concentric circles has an impact on how easy it is to set up the rules.
2. The data looks quite elliptical (often this is not the case for actual data).

3. The spread of the data varies considerably for the crops. Soybeans have wide variability for example.
4. It will be impossible to tell with certainty which crops we have, if the reflected energy comes from the overlap region of corn with soybeans, because both are possible.
5. It would be ideal if the data for each crop were as far apart as water from corn and if the spread were as small as water and elliptical in form and there were not areas of overlay.

However, it appears that these items are not under our control. The position of sensor bands and their band width determines the locations of the centers of the spread of points.

The spread of the data and its contour are determined by factors such as soil conditions, varieties of crops, amount of fertilizer used, planting dates, atmospheric conditions, and data preprocessing.

As far as the overlapping areas are concerned where mislabeling or misclassification is inevitable, the complexity of nature herself is the reason. It is impossible to identify unambiguously all types of targets on the ground on the basis of their reflectance characteristics in the visible and near infrared part of the spectrum alone without considering additional criteria such as shape, size, texture, pattern and association. This the more so, since spectral reflectance of natural targets normally varies with time and environment. Often the differences between various targets are significant only in very narrow wavelength bands, therefore, when comparing the reflectance values of different targets in the broad spectral bands of the LANDSAT multi-spectral scanner, these differences may no longer be recognizable.

One has also to take into account the relation between pixel and feature size. Most of the natural targets on the ground are smaller than a LANDSAT picture element. The spectral radiation value of a given pixel, therefore, represents normally a mixed signature. Only when a sufficiently large area of the surface is densely and homogeneously covered by the feature under study can one expect to have a predominant signature representative of this feature.

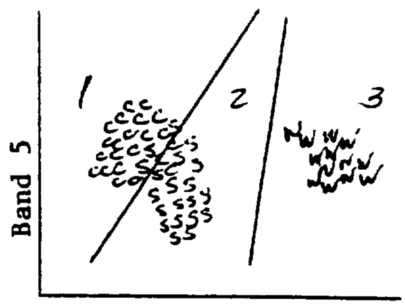
Finally, one should mention that spectral signature clusters overlapping in a two-dimensional color space might be completely separated in color spaces of more than 2 dimensions. In general, it can be said that the more spectral bands that are compared and the narrower they are, the better the separability of the targets. Digital LANDSAT MSS data principally allows to analyze the signature patterns in a four-dimensional color space.

The best that can be hoped for in practice is to estimate from a sample the scatter diagram of the population and this we know how to do if it is dealt with using scientific sampling procedures.

A valid statistical estimate is needed and this requires a random sample from the population of interest. All parts of the population of interest must have a chance of selection and the size must be large enough to adequately represent the population. If the population structure is as complicated as water in Figure 2, or if estimates are needed that are quite accurate, as for corn and soybeans, then, a fairly substantial sample size is required.

The area sampling sampling frame is perfect because a valid statistical estimate can be made for the LANDSAT frame since a random sample of all possible segments is available and reflected energy for the crops can be determined for the fields inside the segments. These signatures are estimates for the scene they are in, so, it is valid to use these values for computer training of the discriminant functions. After population scatter diagrams have been estimated, rules are set up to allot pixels with known energy readings but unknown crop labels to crop categories. Rules are simple; they amount to drawing lines that partition the space. Figure 4 shows an example of this.

Figure 4. Partitioned Space Showing Population Scatter Diagram.

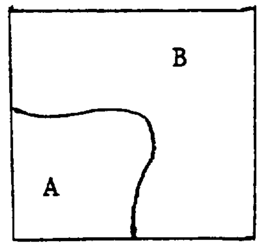


All pixels that need crop labels should then be plotted on the partitioned space. If they fall in partition one, give it a label of corn, even though some soybeans will creep in obviously, water will be no problem.

Incidently, it turns out that the location, size and shape of these population scatter diagrams shift relative to each other in different scenes and even different parts of the same scene. Hence, using a partition developed on one locale of a LANDSAT scene to label pixels from another locale is hazardous.

There are two cases, both are quite different. One is reasonable, and the other is not. For illustration we divide LANDSAT image into two parts as shown in Figure 5.

Figure 5. LANDSAT Frame Divided Into Two Parts.



We further assume that Section A has been divided into 600 small parts and draw a random sample of 60 parts representative of the 600. This may or may not be truly representative. If it is, then, the reflective energy (the signature) from these 60 segments adequately represents the reflected energy for all crops in Section A. We do not consider the use of the signature extension. This is simply a valid statistical inference. It is a commonly misunderstood notion that one does not have a sample from the population of interest to make an inference, for that population.

Should one wish to classify crops in Section B, it would be necessary to divide the Section B into segments and draw a random sample from these segments as representative for signatures in Section B. One must sample the population of interest or the inference will be erroneous.

C. Acreage Estimates Using Classified LANDSAT Data.

In order to make use of LANDSAT to reduce the sampling variation one must first estimate the linear relationship between classified pixels for a crop and acres of that crop. Figure 6 illustrates this relationship.

Figure 6. Population Relationship Between Classification Results and Reported Acres of the Same Crop for One LANDSAT Scene.

Again, these relationships are population relationships that are unknown, so they must be estimated from a sample.

Our area frame sample segments can be used to estimate this relationship. For example, sample observations for Crop A are shown in Figure 7 and Figure 8.

Figure 7. Sample Data Points for Crop A Showing Relationship Between Pixels and Acres.

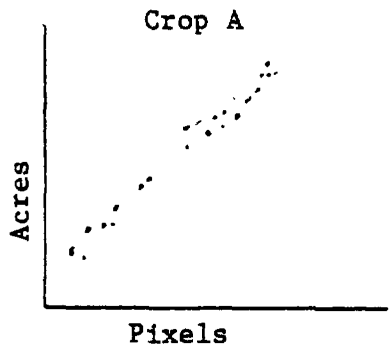


Figure 8. Estimated Population Linear Relationship Based on Sample Data in Figure 7.

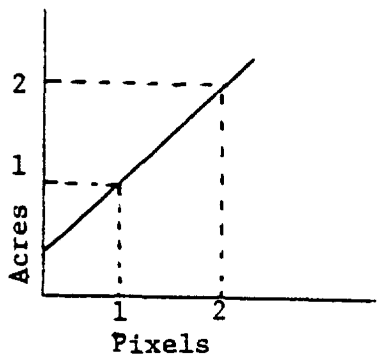


Figure 8 illustrates the relationship that is needed in order to use classified LANDSAT CCT results. (CCT stands for "Computer Compatible Tape").

This is on a per segment relationship. Therefore, we can locate a segment in LANDSAT, classify the segment and count the pixels of Crop A. If the number of pixels for Crop A turns out to be at point 1 then we read the corresponding acres on the y-axis. If on the other hand, the number of classified pixels for the segment turn out to be at point 2 then we read that value on the y-axis.

This procedure could be completed for each segment in the population and we could sum up all the segments to get an estimate using satellite information across the whole areas. However, all this is unnecessary.

Since we know N, the total number of segments in the LANDSAT frame, we can classify every pixel in the frame and divide the total number of pixels in Crop A by the number of segments in the frame. This then would equal the average number of pixels in Crop A for the average segment.

Also, the total number of pixels of Crop A in sample segments (n) is known. With this information we can adjust the direct expansion estimate for the difference between the pixels in Crop A for the sample (n) versus the total of the population (N).

Figure 8 illustrates how the adjustments would be made. Say a difference between the average pixels for Crop A for the sample is at point 1 and the average for the universe is at point 2. The adjustment in acres is made on the y-axis. The formula is:

$$\hat{Y}_{reg} = \bar{Y} + b\ (\bar{X}_{total} - \bar{x}_{sample})$$

$\hat{Y}_{reg}$ is the adjusted number of acres in the average segment. $\hat{Y}_{reg}$ is then multiplied by N to get an estimate for the total.

The variance for $\hat{Y}_{reg}$ is $\frac{n - 1}{n - 2}$ $(1 - r^2)$ times the variance of the direct expansion. This regression model reduces the spread of the sampling error distribution by a factor of $(1 - r^2)$.

In summary, one needs ground data for a properly selected statistical sample, as well as the computer classification for the same areas. Thus, the necessary information is available to adjust a full frame classification for all linear relationships between ground data and what the computer classifies as being on the ground, the sampling error will be substantially reduced as compared to not having remotely sensed data.

V. CONCLUSIONS

The area sampling frame is a basic means used in a number of countries for collecting statistical data in agriculture. It allows one to derive estimates of economical parameters and of all types of agricultural products from samples that cover only a small part of the land under survey. A statistical sampling frame is therefore fast and cost-effective, particularly in very large areas or in cases where a complete enumeration is not practicable or economically feasible due to other reasons such as socio-economic conditions or lack of infrastructure. The area sampling frame technique may therefore help to establish a quick and comprehensive agricultural information system in developed and developing countries alike.

The accuracy of the estimates depends on whether the characteristics in the sample are representative of the characteristics in the whole population. In order to ensure this, the construction of the area sampling frame must follow scientifically based procedures. The subdivision of the total land into homogeneous areas (strata) with respect to their agricultural characteristics, as well as the further subdivision of these strata into sampling units with clearly recognizable physical boundaries are two decisive steps in the construction of the area sampling frame. Topographic and/or land use maps or aerial photography or both are needed for this purpose. Satellite imagery is a useful supplementary material in the construction of the area sampling frame particularly for the stratification of large areas. High resolution imagery from photographic satellite systems is also suitable for further subdivision of the strata into sampling units and the identification and delineation of their physical boundaries. This might be of particular importance for developing countries which are still lacking aerial photography and accurate small scale maps. If satellite data are available also in digital form on computer compatible tapes (CCT's) they can additionally be used for crop classification and the improvement of acreage estimates.

Philosophy and procedure of the construction of the area sampling frame are outlined in this paper and demonstrated by example.

Appendix A

Questionnaires

Crop Reporting Board
Economics, Statistics, & Cooperatives Service
U.S. Department of Agriculture

Form Approved
O.M.B. Number 40-R2766
Approval Expires 4-30-80
C. E. 12-0029i
PART A – 5
Non-MF, Eastern
LA, MD, DE, MI, N. Eng.
NJ, SC, VA, WV

JUNE 1979
ACREAGE & LIVESTOCK
Enumerative Survey

State	District	Segment	Tract
__	____	00000 ____	__

Response to this survey is voluntary and not required by law. However, cooperation is very important in order to establish crop acreage planted this spring and current livestock numbers. Facts about your farm or ranch will be kept CONFIDENTIAL and used only in combination with similar reports from other producers.

1. Segment Number: ______________ Tract Letter: __________

County: ______________________

Cattle E. O.	Hog E. O.	Rice	Potato	Labor	Optional
402	403	409	406	410 . 000	407
NOL ☐ POL ☐ OL ☐	NOL ☐ POL ☐ OL ☐			NOL ☐ OL ☐	408

2. I need to make sure that we have your *(the operator's)* name and address complete and correct.

Name of Farm, Ranch or Operation: ______________________

Name of Operator: ______________________
(Last) *(First)* *(Middle)*

Address: ______________________
(Route or Street)

____
(City) *(State)* *(Zip)*

Phone No: (______) ______________
Area Code

3. Is the operation named above: *Individually operated* - 1 ☐ *Partnership or joint* - 2 ☐ *Managed Land* - 3 ☐ . . . . . ENTER CODE | 845 |

4. Does the operator of this tract live INSIDE or OUTSIDE the segment?

INSIDE ☐ - 5 *Enter 5 in Code Box and continue.*
OUTSIDE ☐ - 6 *Enter 6 in Code Box and go to Page 2.* . . . . | 81 __ 1 |

5. Are there any other persons living in this household who operate a farm or ranch?

NO ☐ - *Continue* YES ☐ - *Enter Name* ______________________
(Assign tract on Part ID, go to item 6.)

6. Do you operate land under any other name or land arrangement other than the one listed above?

NO ☐ - *Continue.* YES ☐ - *Assign another tract letter for other arrangement.*

SECTION A – ACREAGES OF FIELDS AND CROPS INSIDE BLUE TRACT BOUNDARY

How many acres are inside this blue tract boundary drawn on the photo *(or map)*? . . . . Acres [.]

Now I would like to ask about each field inside this blue tract boundary and its use in 1979.

		FIELD NUMBER	1	2	3	4
1.	TOTAL ACRES IN FIELD		.	.	.	.
2.	CROP OR LAND USE *(Specify)*					
3.	WOODS, WASTE, IDLE LAND, ROADS, DITCHES, ETC.		.	.	.	.
4.	OCCUPIED FARMSTEAD OR DWELLING		843 .	843 .	843 .	843 .
5.	PASTURE		842 .	842 .	842 .	842 .
6.	TWO CROPS PLANTED IN THIS FIELD for harvest this year or two uses of the same crop?		NO ☐ YES ____	NO ☐ YES ____	NO ☐ YES ____	NO ☐ YES ____
			844 .	844 .	844 .	844 .
7.	ACRES LEFT TO BE PLANTED?		61_ .	61_ .	61_ .	61_ .
11.	WINTER WHEAT	Planted	540 .	540 .	540 .	540 .
12.		For Grain	541 .	541 .	541 .	541 .
13.	RYE	Planted and to be planted	547 .	547 .	547 .	547 .
14.		For Grain	548 .	548 .	548 .	548 .
15.	OATS	Planted and to be planted	533 .	533 .	533 .	533 .
16.		For Grain	534 .	534 .	534 .	534 .
17.	BARLEY	Planted and to be planted	535 .	535 .	535 .	535 .
18.		For Grain	536 .	536 .	536 .	536 .
19.	CORN	Planted and to be planted	530 .	530 _ .	530 .	530 .
20.		For Grain	531 .	531 .	531 .	531 .
21.	SORGHUM *(Excl. crosses)*	Planted and to be planted	570 .	570 .	570 .	570 .
22.		For Grain	571 .	571 .	571 .	571 .
23.	OTHER USES OF GRAINS PLANTED .	Use				
	Acres abandoned, cut for hay, silage, etc.	Acres	.	.	.	.
24.	HAY Cut and to be cut	ALFALFA and ALFALFA MIXTURES	653 .	653 .	653 .	653 .
25.		OTHER HAY Kind				
		Acres	65_ .	65_ .	65_ .	65_ .
26.	SOYBEANS	Planted and to be planted	600 .	600 .	600 .	600 .
27.	TOBACCO	Class *(Specify* ____________)	67_ .	67_ .	67_ .	67_ .
28.	PEANUTS	Planted and to be planted	690 .	690 .	690 .	690 .
29.	RICE	Planted and to be planted	605 .	605 .	605 .	605 .
30.	COTTON UPLAND	Planted and to be planted	524 .	524 .	524 .	524 .
31.		Abandoned	523 .	523 .	523 .	523 .
32.	DRY EDIBLE BEANS	Planted and to be planted	607 .	607 .	607 .	607 .
33.	SUGAR BEETS	Planted and to be planted	691 .	691 .	691 .	691 .
36.	IRISH POTATOES	Planted and to be planted	552 .	552 .	552 .	552 .
38.	OTHER CROPS	Acres planted or in use	--- .	--- .	--- .	--- .

SECTION A – ACREAGES OF FIELDS AND CROPS INSIDE BLUE TRACT BOUNDARY (Cont'd)

ENUMERATOR ENTERED TRACT ACRES .

FIELD NUMBER . . .	5	6	7	8	9	OFFICE USE
						TOTAL ACRES
1. Total Acres in Field	.	.	.	.	.	840 .
2. Land Use Name						SUM OF WOODS, ETC.
3. Woods, Waste, etc.	.	.	.	.	.	841 .
4. Occupied Farmstead	843 .	843 .	843 .	843 .	843 .	
5. Pasture	842 .	842 .	842 .	842 .	842 .	
6. Two Crops Planted?	NO ☐ YES ____ 844 .	NO ☐ YES ____ 844 .	NO ☐ YES ____ 844 .	NO ☐ YES ____ 844 .	NO ☐ YES ____ 844 .	
7. Left To Be Planted?	61_ .	61_ .	61_ .	61_ .	61_ .	
11. Winter Wheat Planted	540 .	540 .	540 .	540 .	540 .	
12. Winter Wheat For Grain	541 .	541 .	541 .	541 .	541 .	
13. Rye Planted	547 .	547 .	547 .	547 .	547 .	
14. Rye For Grain	548 .	548 .	548 .	548 .	548 .	
15. Oats Planted	533 .	533 .	533 .	533 .	533 .	
16. Oats For Grain	534 .	534 .	534 .	534 .	534 .	
17. Barley Planted	535 .	535 .	535 .	535 .	535 .	
18. Barley For Grain	536 .	536 .	536 .	536 .	536 .	
19. Corn Planted	530 .	530 .	530 .	530 .	530 .	
20. Corn For Grain	531 .	531 .	531 .	531 .	531 .	
21. Sorghum Planted	570 .	570 .	570 .	570 .	570 .	
22. Sorghum For Grain	571 .	571 .	571 .	571 .	571 .	
23. Grain Other Uses						
Acres	.	.	.	.	.	
24. Alfalfa Hay	653 .	653 .	653 .	653 .	653 .	
25. Other Hay – Kind						SOYBEANS FOLLOWING ANOTHER CROP
Other Hay – Acres	65_ .	65_ .	65_ .	65_ .	65_ .	
26. Soybeans	600 .	600 .	600 .	600 .	600 .	602 .
27. Tobacco	67_ .	67_ .	67_ .	67_ .	67_ .	
28. Peanuts	690 .	690 .	690 .	690 .	690 .	
29. Rice	605 .	605 .	605 .	605 .	605 .	
30. Cotton Upland – Planted	524 .	524 .	524 .	524 .	524 .	
31. Cotton Upland – Abandoned	523 .	523 .	523 .	523 .	523 .	
32. Dry Edible Beans	607 .	607 .	607 .	607 .	607 .	
33. Sugar Beets	691 .	691 .	691 .	691 .	691 .	
36. Irish Potatoes	552 .	552 .	552 .	552 .	552 .	
38. Other Crops – Acres	--- .	--- .	--- .	--- .	--- .	

SECTION A – ACREAGES OF FIELDS AND CROPS IN TRACT (Cont'd)

WINTER WHEAT INTENTIONS:

40. Do you intend to seed any winter wheat inside this blue tract boundary this fall?	YES ☐ = 1 DON'T KNOW ☐ = 1 NO ☐ = 2	... Enter Code	543

Appendix B

Random Number Table

A TABLE OF 14,000 RANDOM UNITS

Line/Col.	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
1	10480	15011	01536	02011	81647	91646	69179	14194	62590	36207	20969	99570	91291	90700
2	22368	46573	25595	85393	30995	89198	27982	53402	93965	34095	52666	19174	39615	99505
3	24130	48360	22527	97265	76393	64809	15179	24830	49340	32081	30680	19655	63348	58629
4	42167	93093	06243	61680	07856	16376	39440	53537	71341	57004	00849	74917	97758	16379
5	37570	39975	81837	16656	06121	91782	60468	81305	49684	60672	14110	06927	01263	54613
6	77921	06907	11008	42751	27756	53498	18602	70659	90655	15053	21916	81825	44394	42880
7	99562	72905	56420	69994	98872	31016	71194	18738	44013	48840	63213	21069	10634	12952
8	96301	91977	05463	07972	18876	20922	94595	56869	69014	60045	18425	84903	42508	32307
9	89579	14342	63661	10281	17453	18103	57740	84378	25331	12566	58678	44947	05585	56941
10	85475	36857	43342	53988	53060	59533	38867	62300	08158	17983	16439	11458	18593	64952
11	28918	69578	88231	33276	70997	79936	56865	05859	90106	31595	01547	85590	91610	78188
12	63553	40961	48235	03427	49626	69445	18663	72695	52180	20847	12234	90511	33703	90322
13	09429	93969	52636	92737	88974	33488	36320	17617	30015	08272	84115	27156	30613	74952
14	10365	61129	87529	85689	48237	52267	67689	93394	01511	26358	85104	20285	29975	89868
15	07119	97336	71048	08178	77233	13916	47564	81056	97735	85977	29372	74461	28551	90707
16	51085	12765	51821	51259	77452	16308	60756	92144	49442	53900	70960	63990	75601	40719
17	02368	21382	52404	60268	89368	19885	55322	44819	01188	65255	64835	44919	05944	55157
18	01011	54092	33362	94904	31273	04146	18594	29852	71585	85030	51132	01915	92747	64951
19	52162	53916	46369	58586	23216	14513	83149	98736	23495	64350	94738	17752	35156	35749
20	07056	97628	33787	09998	42698	06691	76988	13602	51851	46104	88916	19509	25625	58104
21	48663	91245	85828	14346	09172	30168	90229	04734	59193	22178	30421	61666	99904	32812
22	54164	58492	22421	74103	47070	25306	76468	26384	58151	06646	21524	15227	96909	44592
23	32639	32363	05597	24200	13363	38005	94342	28728	35806	06912	17012	64161	18296	22851
24	29334	27001	87637	87308	58731	00256	45834	15398	46557	41135	10367	07684	36188	18510
25	02488	33062	28834	07351	19731	92420	60952	61280	50001	67658	32586	86679	50720	94953
26	81525	72295	04839	96423	24878	82651	66566	14778	76797	14780	13300	87074	79666	95725
27	29676	20591	68086	26432	46901	20849	89768	81536	86645	12659	92259	57102	80428	25280
28	00742	57392	39064	66432	84673	40027	32832	61362	98947	96067	64760	64584	96096	98253
29	05366	04213	25669	26422	44407	44048	37937	63904	45766	66134	75470	66520	34693	90449
30	91921	26418	64117	94305	26766	25940	39972	22209	71500	64568	91402	42416	07844	69618
31	00582	04711	87917	77341	42206	35126	74087	99547	81817	42607	43808	76655	62028	76630
32	00725	69884	62797	56170	86324	88072	76222	36086	84637	93161	76038	65855	77919	88006
33	69011	65797	95876	55293	18988	27354	26575	08625	40801	59920	29841	80150	12777	48501
34	25976	57948	29888	88604	67917	48708	18912	82271	65424	69774	33611	54262	85963	03547
35	09763	83473	73577	12908	30883	18317	28290	35797	05998	41688	34952	37888	38917	88050
36	91567	42595	27958	30134	04024	86385	29880	99730	55536	84855	29080	09250	79656	73211
37	17955	56349	90999	49127	20044	59931	06115	20542	18059	02008	73708	83517	36103	42791
38	46503	18584	18845	49618	02304	51038	20655	58727	28168	15475	56942	53389	20562	87338
39	92157	89634	94824	78171	84610	82834	09922	25417	44137	48413	25555	21246	35509	20468
40	14577	62765	35605	81263	39667	47358	56873	56307	61607	49518	89656	20103	77490	18062
41	98427	07523	33362	64270	01638	92477	66969	98420	04880	45585	46565	04102	46880	45709
42	34914	63976	88720	82765	34476	17032	87589	40836	32427	70002	70663	88863	77775	69348
43	70060	28277	39475	46473	23219	53416	94970	25832	69975	94884	19661	72828	00102	66794
44	53976	54914	06990	67245	68350	82948	11398	42878	80287	88267	47363	46634	06541	97809
45	76072	29515	40980	07391	58745	25774	22987	80059	39911	96189	41151	14222	60697	59583
46	90725	52210	83974	29992	65831	38857	50490	83765	55657	14361	31720	57375	56228	41546
47	64364	67412	33339	31926	14883	24413	59744	92351	97473	89286	35931	04110	23726	51900
48	08962	00358	31662	25388	61642	34072	81249	35648	56891	69352	48373	45578	78547	81788
49	95012	68379	93526	70765	10593	04542	76463	54328	02349	17247	28865	14777	62730	92277
50	15664	10493	20492	38391	91132	21999	59516	81652	27195	48223	46751	22923	32261	85653

Appendix C

Practical Exercise

Contents

1. LANDSAT scene 2537-17480 (July 12, 1976) - San Joaquin Valley, California.
2. Blank transparent sheet.
3. Check sheet with strata boundaries.
4. Instructions for this exercise.
5. Land use strata definitions.
6. Kings County, California map.
7. Completed Kings County, California map.
8. Blank listing sheets for primary sampling units.
9. Primary sampling unit listing sheets filled out.

Instructions for Exercise

1. Obtain thin leaded pencils or pens that will write on transparencies.

2. Tape blank transparent sheet on LANDSAT image and delineate strata using enclosed definitions of land use.

3. Check with transparent overlay showing land use strata. This is only a guide - there are many right answers.

4. Locate Kings County and transfer boundaries onto Kings County map. Change boundaries to correspond to nearest physical boundaries from Kings County map.

5. Check with transparent overlay showing outline of Kings County on LANDSAT.

6. Outline cities on Kings County map. These areas should be large enough to show some street patterns. Follow the outside boundaries. Use a green colored pencil or pen for this.

7. Check to make sure all boundaries used to delineate strata are physical.

8. Construct primary sampling units (PSU's) for each stratum. Divide the strata into PSU's by using the following guide:

   Primary Sampling Unit Guide

Stratum Number		Size Range of PSU
I	Cultivated	6 - 20 sq. miles*
II	Rangeland	15 - 50
III	Non-agricultural	15 - 100
IV	Urban	.01 - 1.00

   *Miles have been used because they can be counted directly on the map.

   Draw PSU's in color of each stratum. Use physical boundaries and try to stay within size range, however, always use good boundaries regardless of size.

9. Check your finished product with one provided in packet.

10. Number PSU's in serpentine fashion beginning at one corner and continuing until all PSU's have been numbered in each stratum. Start at 1 again for each stratum.

11. Estimate the size of each PSU in each stratum. Count the number of square miles in each PSU and enter this number on the map. Use grid or planimeter to obtain this area if necessary.

    Write the stratum number, PSU number, and size in each PSU. For example, II-3-16. This is the identification for the 3rd PSU in strata II which is 16 sq. miles.

12. List PSU's on PSU identification sheets.

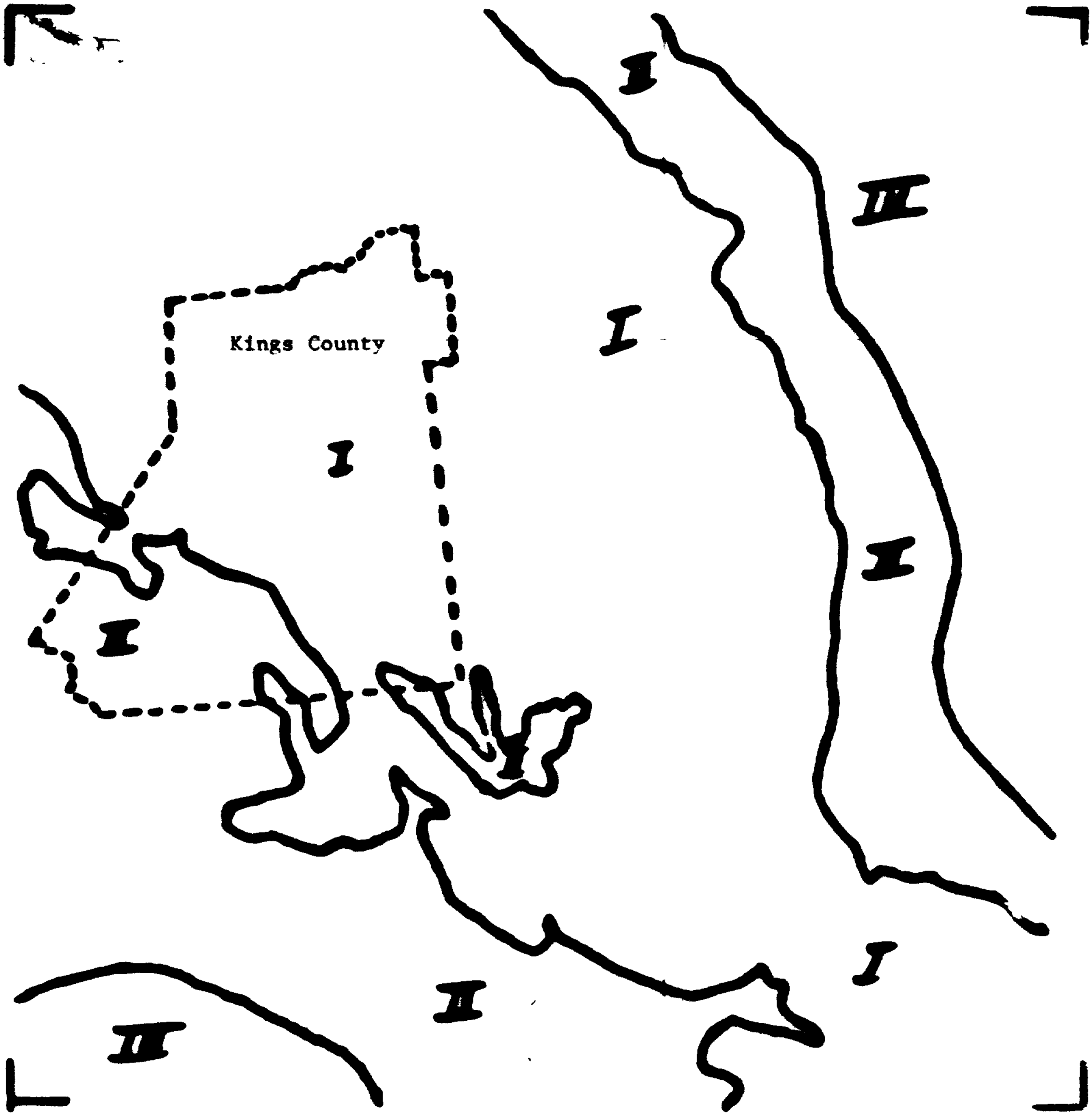
13. Round size down to nearest whole number (if necessary). This is the number of assigned sample units (SU's). Cumulate these assigned numbers in the "CUM" S.U. column.

14. Check your PSU listing sheet with one in packet.

Land Use Strata Definitions

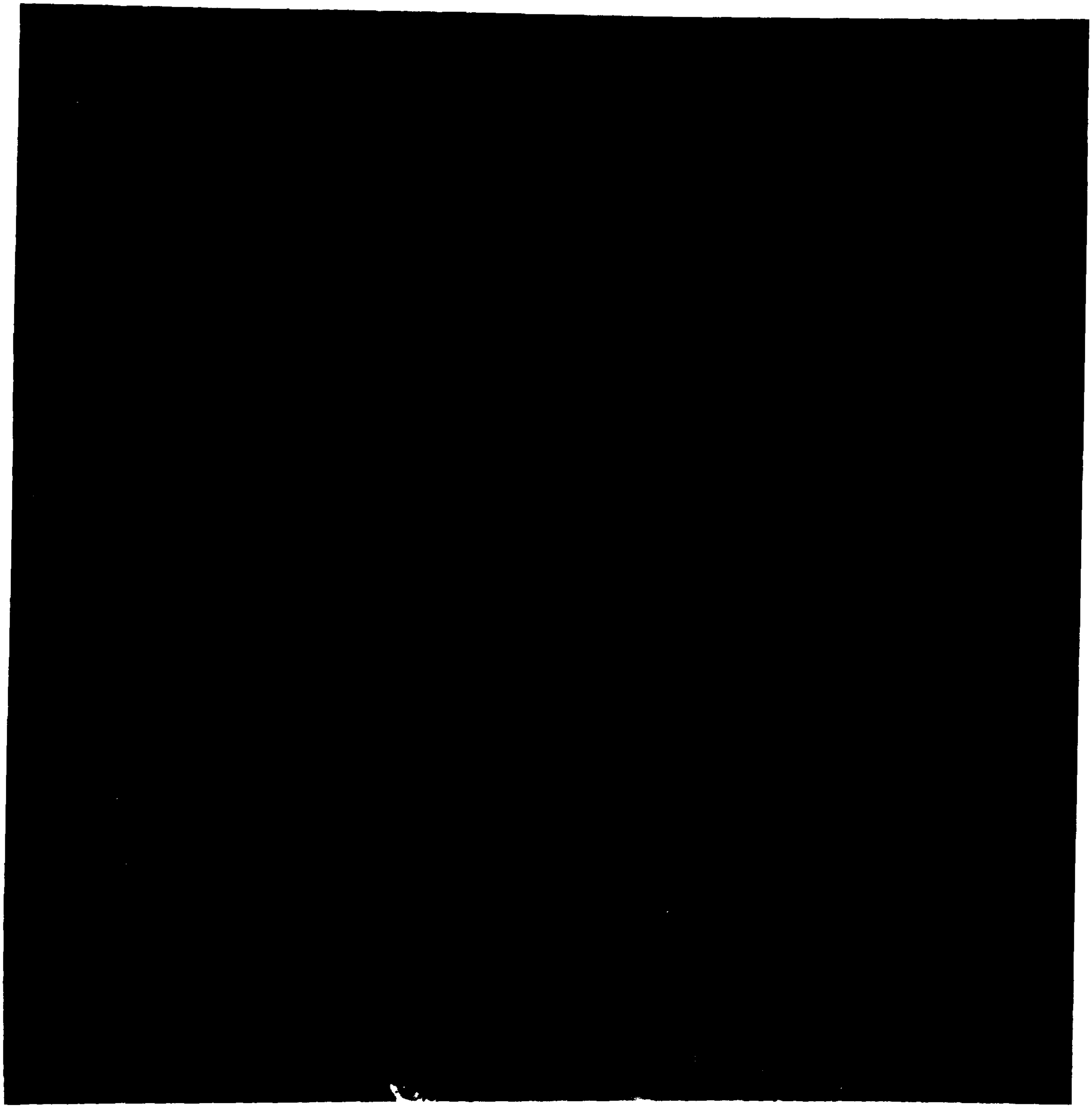
Stratum I (purple)	Intensively Cultivated	This stratum is cultivated land 75% - 100%.
Stratum II (brown)	Rangeland and Desert	This stratum is low land non-cultivated or with a very small percentage of land cultivated.
Stratum III (blue)	Non-agricultural Land - Mountains, Resorts	This stratum land appears non-productive for agriculture. Also, could be unused for reasons such as resort or military base.
Stratum IV (green)	Urban	This stratum is for cities. Generally, these areas cannot be delineated on the LANDSAT.

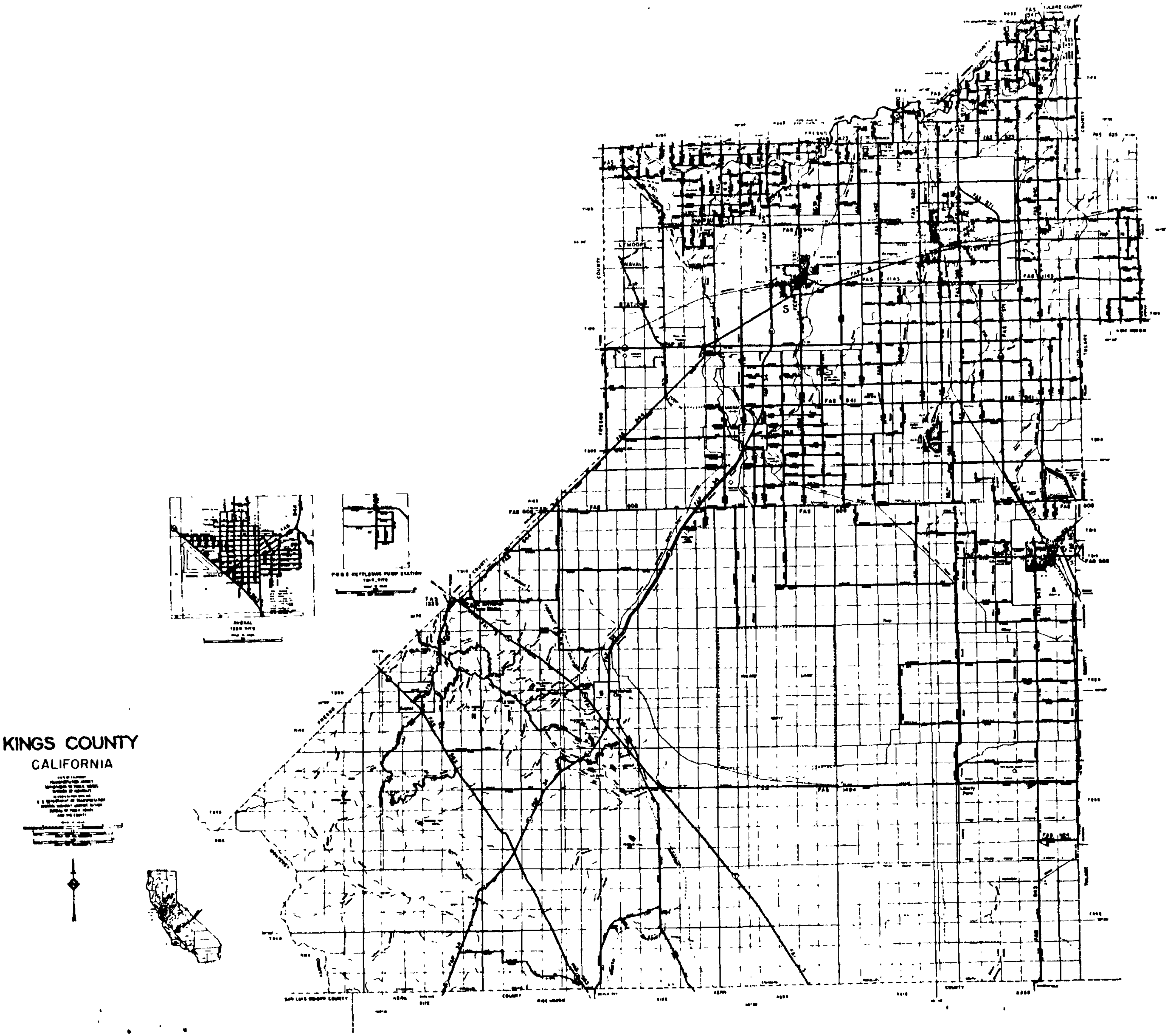
Check Sheet

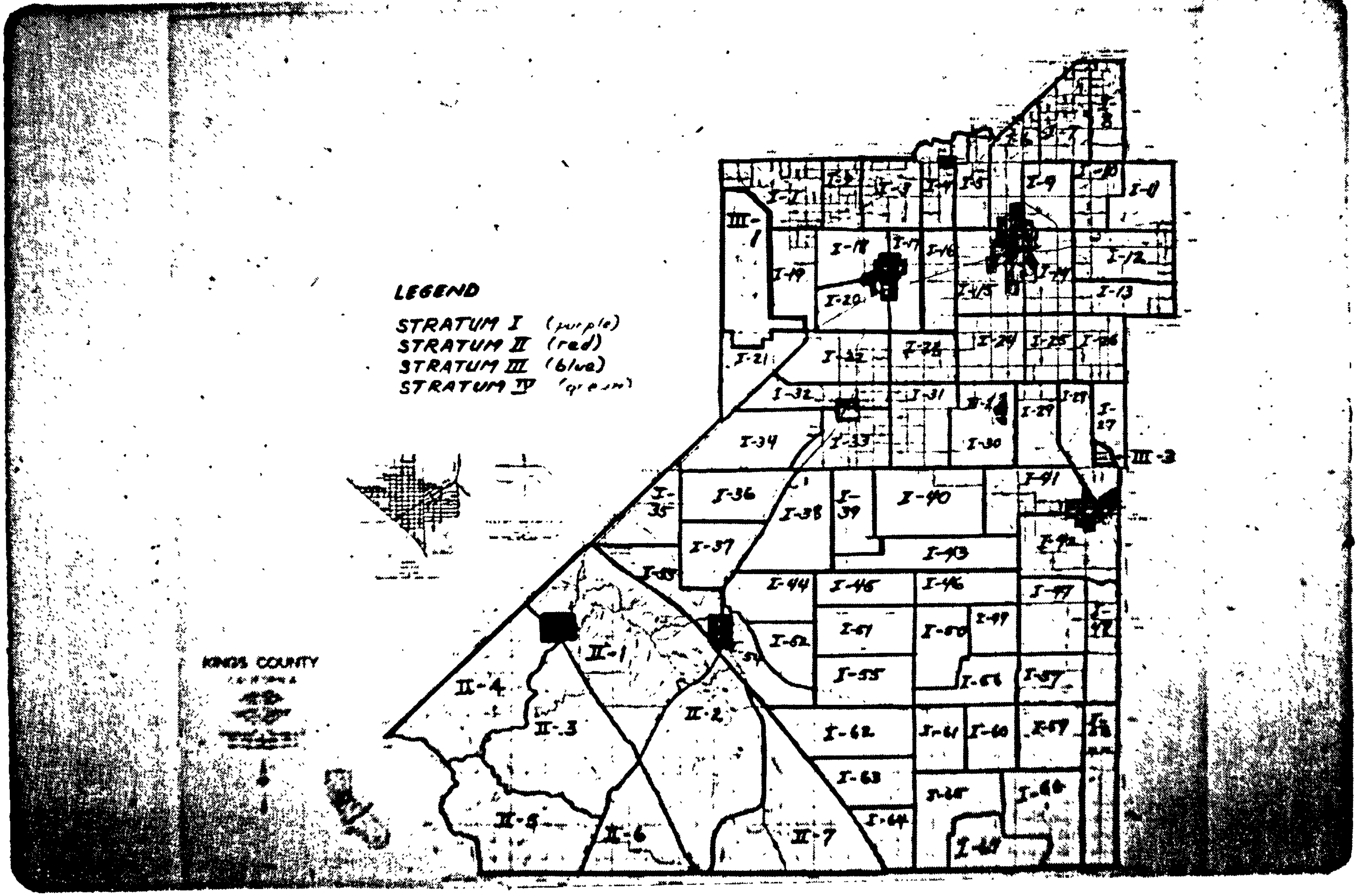


Stratum Boundaries (solid line)

Kings County Boundary (dashed line)




KINGS COUNTY
CALIFORNIA
TULARE COUNTY
LEMOORE
NAVAL
AIR
STATION


LEGEND
STRATUM I (purple)
STRATUM II (red)
STRATUM III (blue)
STRATUM IV (green)
KINGS COUNTY

PRIMARY SAMPLING UNIT IDENTIFICATION SHEET

COUNTY Kings

MAP DESCRIPTION	STRATUM	PSU NUMBER	AREA MEASURE-MENT	NO. OF ASSIGNED S.U.	"CUM" S.U.	SELECTED PSU (S.U. NO.)	NOTES
Kings	II	1	66				
	II	2	55				