A LAND USE INFORMATION SYSTEM BASED ON STATISTICAL INFERENCE

WILLIAM H. WIGTON
HAROLD F. HUDDLESTON

ECONOMICS, STATISTICS, AND COOPERATIVES SERVICE
U.S. DEPARTMENT OF AGRICULTURE
WASHINGTON, D.C.

APRIL 1978

A LAND USE INFORMATION SYSTEM BASED ON STATISTICAL INFERENCE*

William H. Wigton and Harold F. Huddleston

Economics, Statistics, and Cooperatives Service
U.S. Department of Agriculture

ABSTRACT

This paper presents an efficient, practical and accurate information system showing how to use some promising new technology. An agricultural information system can be built on objective ground-based data and combined with LANDSAT data to increase the precision of statistics for large areas and also generate local data. The results or statistics produced will be defensible and the accuracy can be stated if care is taken to follow strictly the laws of probability.

Resource requirements vary depending on accuracy levels needed. However, the system combines ground-based and LANDSAT data in a flexible manner depending on materials, equipment available, expertise and experience of personnel involved. The construction of an area sampling frame in the country is assumed to have preceded the time when sampling for making crop estimates will be undertaken.

* This was a talk given at the Twelfth International Symposium on Remote Sensing of Environment, April 20-26, 1978.

INTRODUCTION

The need for better agricultural production data on a worldwide basis is well known. An individual country needs to know the amount of available food and feed in order to make important marketing and policy decisions. In addition, the world community may need to know in order to provide assistance and give top priority to those commodities that fluctuate the most or fall into short supply.

This paper describes an agricultural data collection system that can produce accurate, timely, objective information, making effective use of people and the limited resources in many countries. We elaborate upon these ideas with respect to agricultural production.

An accurate estimate means that the value generated is close to the actual total quantity harvested. Usually. how close an estimator is to actual output is measured by the precision (the spread around the expected value of the estimator). In sample surveys generally we talk about sampling errors (precision) and nonsampling errors (bias). Sampling errors are reduced by improving the survey design and/or increasing the sample size. The nonsampling errors are controlled by concepts used, procedures, and training and measurement techniques. An important characteristic of a survey system is that it be possible to measure both sampling and nonsampling errors. Controlling these errors helps make the data useful in defending the results against other official and private estimates of crop production.

Timely data means that information is available when it is useful. Specifically, we mean that the statistics for current crops are available soon after the basic data are gathered either from the grower or from in-the-field counts of crops or animals. Timely data implies that both the data collection time and the office time to summarize the data are short.

Objective data implies that the results do not depend on vested interests of some group of individuals. Even more, it means that the results depend on the use of random numbers to select those areas used for data collection. Further, when both sampling and nonsampling errors are measured, the statistics will hold up in a court of law.

For example, suppose two areas are affected by drought. Funds will be distributed according to extent of damage. If personal biases are allowed to enter into the estimation process, people will not believe the statistics even if they are correct unless scientific sampling procedures and measurement concepts are used.

The effective use of resources in a data-collection system is usually a comparative criterion. That is, compared with other possible ways to obtain estimates, how much money does it take or how many people are required or what equipment is needed? Is there a cheaper way to get the same information? This criterion is simply a matter of practicality.

Flexibility of a system to new problems means that the techniques have been tested and found to be applicable. For example, data can be obtained on seed crops, vegetables, fruit and nuts, food crops, feed crops, acreage planted, acreage for harvest, remaining stocks, livestock, dairy and poultry products, including births, slaughter for cattle, sheep, pigs, milk, egg production, as well as prices paid by farmers for fertilizer, seeds, labor, and prices received for agricultural products. Also, can the system collect other types of data, such as forest inventories and timber volume or household data as they are needed? If an agricultural data collection system meets these stringent test criteria, it would be very useful to a country's economic and policy planners.

AREA SAMPLING FRAMES

We wish to present some methodology at this point. We discuss area sampling because these frames can be related directly to LANDSAT. Other types of sampling frames exist but are not specifically related to the land.

An area sampling frame is a breakdown of the land area of a country into N small land units with no areas of overlap or omission. A relatively small number of units (n) are selected to be interviewed or observed by ground personnel.

If the area sampling frame is properly constructed and stratified, the small sample of selected units, hereafter called segments, will represent the total universe or land area of a country. It should be pointed out that the land area of a country includes its internal rivers and water bodies.

Area sampling frame concepts are simple to understand but extreme care must be taken to follow explicit details in the construction process. Any errors could produce substantial inaccuracies in the final results. A poster session has been presented to illustrate how one constructs a frame. Several documents on the subject also are available that give reasonable detail.1/2/3/ It does require substantial resources and highly trained people as well as some experienced help.

Since this topic has been covered in other reports, this paper is confined to showing how LANDSAT imagery is used.

COMPUTER CLASSIFICATION TECHNIQUES

The LANDSAT tapes for the scene are obtained. The four tapes are reformatted and the spectral values are interfolded for each pixel. Classification of land can be done in the computer by use of discriminant functions. The procedure 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 represents how those particular crops reflect energy. The problem is to set up rules, using samples of pixels for each crop. The procedure will enable us to allot some unknown land pixels to a crop or land cover type, given only the amount of reflected energy of that pixel.

The procedure can be formulated statistically. However, let us examine some of the concepts used.

If the reflected energy counts for channels 5 and 7 of a LANDSAT frame were plotted in a scatter diagram, it might appear as Figure 1.

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 because a quadratic classifier has been used based on the means and variance (often this is not the case for actual data).

3. The spread of data varies considerably for the crops. Soybeans have wide variability, for example.

4. If the reflected energy comes from the overlap region of corn and soybeans, it will be impossible to tell with certainty which crop is reflecting energy.

5. It would be ideal if the multivariate distributions for each crop were as far apart as water from corn. If the distributions were as compact and elliptical in form as water and there were no areas where distributions overlap, the conditions would be ideal.

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

Such factors as soil conditions, crop varieties, amounts of fertilizer used, planting dates, atmospheric conditions, and data preprocessing affect the spread of data.

In the overlapping areas of the distribution, some mislabeling or misclassification is inevitable. It is impossible to identify unambigously all types of land cover. Also, spectral reflectance of natural targets normally varies with time and environment. Often the differences between various land covers 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 multispectral scanner, these differences may no longer be recognizable.

Also, the relation between pixel and feature size must be taken into account. Most of the natural land covers fill a smaller area 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, the spectral signatures which overlap in a two-dimensional space <u>might be</u> completely separated in space 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 allow us to analyze the signature patterns in a four-dimensional color space.

In practice, a sample estimate of the scatter diagram of the population is obtained if the data are derived 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 of the sample must be large enough to adequately represent the population.

The area sampling frame is extremely effective for this purpose 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 entire scene they are in, so, it is valid to use these values for computer training of the discriminant function and avoid problems of signature extension to nonsampled areas or LANDSAT frames. After population scatter diagrams (or ellipsoids) have been estimated, rules are set up to allot pixels with known spectral values but unknown crop categories. Rules are simple: they amount to drawing lines or planes that partition the two-dimensional space. Figure 3 shows an example of the ellipses.

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 soybeans will creep in. Obviously, water will be no problem.

Incidentally, 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 may be hazardous and frequently two signatures are required for the same crop in a scene, and the estimates derived from each signature must be combined in the final estimation.

There are two quite different cases: one is reasonable, and the other is not. For illustration, the image is divided into two parts as shown in Figure 4. We assume that Section A has been divided into 600 small parts and a random sample of 60 parts is selected which is representative of the 600. A single random sample of 60 may not be truly representative but if the sampling process is repeated many times, sampling theory assures us that the procedure is representative. If so, the reflected energy (the signature) from these 60 segments adequately represents the reflected energy for all crops in Section A. An adequate sample size of segments stratified over the scene increases the likelihood of representative signatures for the scene. We do not need to consider the use of the signature extension because the procedure permits a valid statistical inference.

Should one wish to classify crops in Section B, it would be necessary to divide Section B into segments and select a random sample of segments to represent signatures in Section B. The critical point is that the population of interest be sampled.

ACREAGE ESTIMATES USING CLASSIFIED LANDSAT DATA

In order to use LANDSAT for statistical inferences and reduce the sampling variation, one must first estimate the relation, generally linear, between classified pixels for a crop and acres of that crop. That is, the relation must be estimated for each scene by use of the area segment sample. Figure 5 illustrates this relation. Again, these relations are unknown and must be estimated from a sample.

The crop cover data from the sampled segments can be used to estimate this relation. For example, sample observations for Crop A are shown in Figure 6 and Figure 7. Figure 7 illustrates the relation that is needed in order to use classified LANDSAT "Computer Compatible Tape" (CCT) results.

The relations in Figures 6 and 7 are based on total acres and pixel counts per segment. 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 on the x-axis, we read the corresponding acres on the y-axis. If, on the other hand, the number of classified pixels for the segment turns out to be at point 2 on the x-axis, 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 entire area. However, all this is unnecessary because of the additive property of a linear estimator.

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 result would equal the average number of pixels in Crop A for all segments.

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

Figure 7 illustrates how the adjustments would be made. The average number of pixels for Crop A for the sample is at point 1 and the average for the universe (or entire image) is at point 2. The adjustment is made on the y-axis. The formula is:

$$\bar{Y}_{reg} = \bar{Y} - \hat{b}\,(\bar{x}_{sample} - \bar{X}_{univ.})$$

$\bar{Y}$ is the number of acres in the n segments. $\bar{Y}_{reg}$, the average of all N segments, is then multiplied by N to get an estimate of the number of acres in the crop.

The variance for $\bar{Y}_{reg}$ is $\frac{n-1}{n-2}(1 - r^2)$ times the variance of the direct expansion.

The regression model reduces the spread of the sampling error distribution by a factor of $(1 - r^2)$.

In summary, ground data for a properly selected statistical sample as well as computer classification for the same areas are used to adjust the classified pixel counts. Thus, the necessary information is available to adjust a full-frame classification of all linear relations existing between ground data and what the computer classifies as being on the ground. The sampling error will be substantially reduced as compared with not having remotely sensed data.

DERIVING CROP PRODUCTION STATISTICS

In the previous section we discussed the statistical methodology to be employed. Now we focus on the actual context in which the statistical techniques are used to make inferences. The techniques are appropriate in operational setting where the officials of the country are responsible for their production statistics.

Crop production statistics are expected to constitute a major component in meeting domestic and world food and feed needs. The primary goal is to derive agricultural production statistics for large geographic units, such as countries and major subdivisions of countries (provinces, states, etc.). At the same time, it would be desirable and beneficial if the system were capable of providing production statistics for small geographic areas or political units when needed.

The analysis approach to production statistics is based on obtaining area (acreage) and yield estimates for each geographic area based on the area sample and LANDSAT. Crop production is derived for the same area as a product of the two components, area and yield for each crop. First, we examine some of the implications of the operational constraints with respect to inferences on crop production and then illustrate these techniques.

A. Some of the factors to be considered in deriving current year statistics are:

1. Timing: The crop development stage has an effect on both the acreage classification and the yield estimation procedures. The same timing may not be optimum for both components for the same crop and may vary substantially for different crops.

2. LANDSAT Frame or Image: A LANDSAT frame serves as a basis for the poststratification of the area frame units and as an analysis unit. The sample of ground area units is selected independently of the image used for crop classification and yield estimation. If the area frame sampling units are selected from land-use strata, the area of the LANDSAT image poststratifies these strata thus creating two stratifications.

3. Classification Procedures: The classification for acreages attempts to maximize the correlation between the objective data for the area segments and the classification pixels. However yield estimation is based on the universe of classified pixels; consequently, the total number of classified

pixels needs to provide a relatively unbiased estimate of acreage. This requirement may affect the acreage classification routine in subtle ways, like requiring representative prior probability to be used for each crop.

4. Geographic Areas: The techniques illustrated are potentially more cost-effective for estimating crop production because LANDSAT covers every acre for large geographic areas and the regressions for acreage and yield (i.e., one for each component) are applicable to all geographic areas (irrespective of size) within the LANDSAT image.

B. Acreage Component Estimation

All pixels in the image for a specific analysis district (e.g., a particular set of geographic areas or counties wholly contained within a LANDSAT frame) are classified by crop types. Then, a classified pixel total for each crop type is aggregated to obtain individual totals for all segments sampled (ground units observed or enumerated) within the analysis district.

An estimator of the total acreage for a particular crop in a particular analysis district i with its sampling error is then computed as follows. The total acreage is estimated using the double sampling methodology for each stratum or analysis district as

$$\hat{A}_i = N_i\,(\bar{a}_i - \hat{B}_i(\bar{x}_i - \bar{X}_i))$$

and, assuming a sufficiently large sample of segments the variance is

$$V(\hat{A}_i) = N_i^2\, V(\bar{a}_i)(1 - r_i^2)\left(\frac{n_i - 1}{n_i - 2}\right) .$$

For individual analysis districts, the normal approximation for small samples is used; that is, $V(\hat{A}_i)$ for a large sample is multiplied by

$$\left(1 + \frac{1}{n_i - 3}\right)$$

where A_i = total acres of the crop in all segments contained within the i^{th} analysis district

N_i = total number of segments contained within the i^{th} analysis district (known from sample frame)

n_i = the number of ground area segments sampled in the i^{th} district

$\bar{a}_i$ = average number of acres of the crop reported per area segment for all n_i area segments sampled in the i^{th} district

$\bar{x}_i$ = average number of pixels per segment classified as corresponding to the crop, for the n_i area segments sampled in the i^{th} district

$\bar{X}_i$ = average number of pixels per segment classified as corresponding to the crop over all N_i segments for the i^{th} district

a_{ij} = number of acres of the crop enumerated for j^{th} segment sampled in the i^{th} district

x_{ij} = number of pixels classified as corresponding to the crop for the j^{th} segment sampled in the i^{th} district

$\hat{B}_i$ = the regression coefficient between a_{ij} and x_{ij} based on the n_i area segments smapled in the i^{th} district

$\hat{r}_i$ = correlation coefficient squared between a_{ij} and x_{ij} for the i^{th} district

$$V(\bar{a}_i) = \left\{ \sum_{j=1}^{n_i} a_{ij}^2 - \left[\left(\sum_{j=1}^{n_i} a_{ij} \right)^2 / n_i \right] \right\} / n_i(n_i - 1) \quad .$$

Based on 33 segments falling in 10 western Illinois counties with LANDSAT image ID#2194016042 of August 4, 1975, estimates of crop acreages and estimate errors were computed for several crops or land cover types. The estimates are shown in Table I and their squared sampling errors in Table II. The rectangular LANDSAT data window containing the 10 counties included 4,887,960 pixels and required less than 80 seconds for classification by the ILLIAC IV. Column 3 of Table II indicates the gain or loss of information by using the remote sensing information in conjunction with the conventional area sample ground data. Column 3 in Table I shows the raw total of pixels classified and converted to acres by the factor 1.114, the acreage of one pixel. This type of direct estimate can lead to serious biases in the estimates for individual crops, with the extent of bias varying among crops. For an individual county within the analysis district the same type of estimation employing the regression parameters for the large geographic area can be used to the extent that the regional landscape may be considered homogeneous for the complete set of counties.

C. Yield Component Estimation

In order to derive production, the yield for the same area or group of counties is needed. This can be accomplished in any one of several ways. Here we describe one procedure based on data collected during the growing season for the same set of segments used for acreage. The correlation of satellite spectral reflectance information with appropriate ground-survey plant-yield data has been investigated to obtain statistical estimators with measurable standard errors for small areas contained in a single LANDSAT frame. The method is illustrated in terms of corn and soybean yields in Illinois during the 1975 crop year.

All sample fields of corn and soybeans that were selected corresponded to the objective yield ground surveys and were located on the LANDSAT images. The digital values for the four spectral channels based on all pixels within each field were used to compute the mean channel values. Mean vectors for corn and soybean were obtained from the LANDSAT imagery of August 4, 1975, while the plant field data relate to a 10-day period centered on August 28, 1975, (for the September forecast) and at harvest. Specifically, the mean vectors for 27 corn and 17 soybean fields with 16 or more pixels were computed.

In addition to the data for individual fields which were used to derive the regression equation for predicting the yield for the analysis district, the mean vector for the four LANDSAT channels would be needed for all pixels classified as corn and soybeans. That is, the entire population of classified pixels for the area of the LANDSAT scene must be examined to identify and summarize the corn and soybean pixels to employ double sampling. The frame or portion of the LANDSAT frame classified for the area must be matched to corresponding spectral channel values for the same unclassified pixels on a second tape (i.e., the tape used to derive the classification) to derive information needed to estimate yield. The double sampling model(s) is the same as given earlier except the independent variables are now a vector of four channel values. However, the estimation of the yield is achieved through a double-sampling regression estimator using the classified LANDSAT data for the larger geographic area collated with the ground data for corn and soybean fields in the sample of segments. In addition, it is possible to use the same double-sampling regression estimator to obtain estimates for any smaller area within the larger area classified for which the regression relationship is appropriate. One could be used to <u>forecast</u> final yield based on ear counts and size of ears on September 1. A second regression related the actual grain harvested (i.e., pounds per acre) to the means of the four channel spectral values from LANDSAT based on the August 4 image.

While a number of different variables or combinations of variables based on the field mean vectors and variance vectors were investigated using the August 1975 imagery in western Illinois, only two sets of variables were consistently significant: (1) means of channel 2 and channel 4, and (2) means of channels 2 and 4 plus variances of channels 2 and 4. The regressions based on data set (1) for September 1 and final harvest for the 10-county area within the LANDSAT frame of August 4 are as follows:

$$\text{September 1:} \quad \hat{Y}_s = \bar{y}_s - B_2(\bar{x}_2 - \bar{X}_2) - B_4(\bar{x}_4 - \bar{X}_4)$$

for corn R = .56 and for soybeans R = .48. The subscript i to identify the analysis district has been omitted from each variable.

Harvest: $\hat{Y}_h = \bar{y}_h - B_2'(\bar{x}_2 - \bar{X}_2) - B_4'(\bar{x}_4 - \bar{X}_4)$

for corn $R = .49$ and for soybeans $R = .58$.

where $\hat{Y}_s$ = forecasted yield per acre for the geographic area

$\hat{Y}_h$ = harvested yield per acre for the geographic area

$\bar{y}_s$ = average forecasted yield per acre for sample of fields on September 1

$\bar{y}_h$ = average yield per acre for the sample fields at harvest

$\bar{x}_2$ = mean spectral value for channel 2 on August 4 for sample fields

$\bar{x}_4$ = mean spectral value for channel 4 on August 4 for sample fields

$\bar{X}_2$ = mean spectral value for channel 2 for classified corn (or soybean) pixels on August 4 for the entire geographic area

$\bar{X}_4$ = mean spectral value for channel 4 for classified corn (or soybeans) pixels on August 4 for the entire geographic area

B_2 and B_4 = regression coefficients (also B_2' and B_4')

R = multiple correlation coefficients

The information gained by using spectral data to estimate yield may be computed in a manner similar to column 3 of Table II based on the ratio of variances. For corn and soybeans these information gains are in the range of 1.27 to 1.42. Based on these data sets for western Illinois in 1975, the potential information gain is obviously much less than that for acreage estimation.

In addition to the two sets of variables which were derived from LANDSAT channels 2 and 4, a vegetative index variable was also investigated. The vegetative index

$$I^2 = \frac{x_4 - x_2}{x_4 + x_2} + .5$$

was investigated in each case and was significantly correlated in most instances but the correlations were less than those reported above.

D. Derived Production Statistics

The crop production estimates are derived by multiplying the acreage and yield component estimates for the analysis district and each geographic subarea within the district. That is, the September 1 forecast of production for the analysis district would be

$$\hat{P}_{is} = \hat{A}_{is} \cdot \hat{Y}_{is}, \text{ and}$$

the production estimator at harvest would be

$$\hat{P}_{ih} = \hat{A}_{ih} \cdot \hat{Y}_{ih} \ .$$

However, the use of the LANDSAT spectral data for both acreage and yield would result in an information gain of approximately 7.0 x 1.3 = 9.1 for estimation of corn and soybean production.

This factor of almost 10 for the high quality image used would probably translate at the state level into a factor of 3 or 4. In general, correlations and information gains are reduced over larger areas due to the improbability of high-quality images. A gain in information by a factor of 4 would be roughly equivalent to $65,000 of additional resources (for harvested production) which would be needed based on the current area sampling system to reduce the sampling errors of corn and soybean production at the state level by one-half. The benefits of reduced errors in Illinois have not been estimated but the marginal increase in costs of the LANDSAT analysis should be much less than $65,000 provided real-time imagery is available and assuming that all initial fixed costs of establishing an operational information system have already been provided. The marginal costs of providing forecasts in addition to the estimate of harvested production would increase these dollar figures to about $80,000.

Table III summarizes the derived production estimates for the analysis district (i.e., one LANDSAT scene).

CONCLUSIONS

A technology utilizing LANDSAT that generates statistical estimates of crop production is available. LANDSAT is combined with a probability sample of ground data to provide statistics at the scene or small-area level. Ground data for land-area units from an area sampling frame are poststratified by LANDSAT frames to obtain estimates of crop areas and production which can be added to get estimates for the entire country. This technology is available to any country that wishes to improve its crop statistics. It may not be useful to estimate crops in a country where highly selective samples of ground data or no ground data are available. The potential benefit from employing LANDSAT technology with area sampling is that the data needs and resources for large areas can be used to obtain results for small areas simultaneously.

REFERENCES

1/ "Pilot Study of the Potential Contributions of LANDSAT Data in the Construction of Area Sampling Frames," George A. Hanuschak and Kathleen M. Morrissey, SRS, U.S. Department of Agriculture, October 1977.

2/ "A Training Course in Sampling Concepts for Agricultural Surveys," Harold F. Huddleston, SRS, U.S. Department of Agriculture, April 1976.

3/ "A Guide to Area Sampling Frame Construction Utilizing Satellite Imagery," William H. Wigton, ESCS, U.S. Department of Agriculture, and Peter Bormann, United Nations Outer Space Affairs Division, New York, N.Y. 10017.

FIGURES

Figure 1. Scatter Diagram for Three Types of Cover. C-Corn, S-Soybeans, W-Water

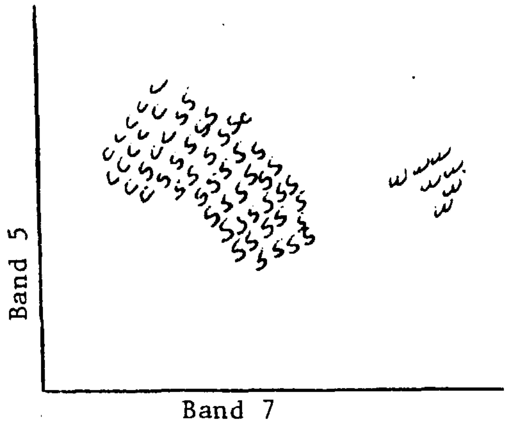


Figure 2. Confidence Limits for Data in Figure 1

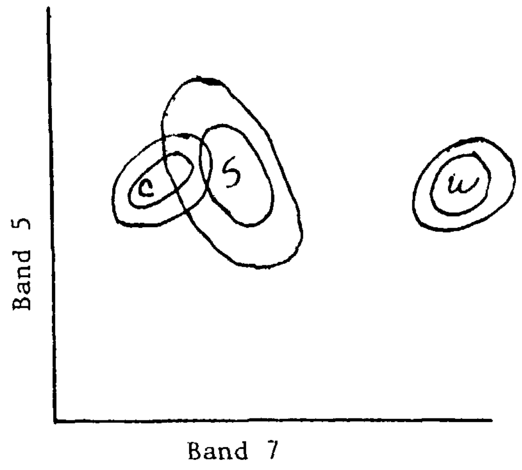


Figure 3. Partitioned Space Showing Population Scatter Diagram

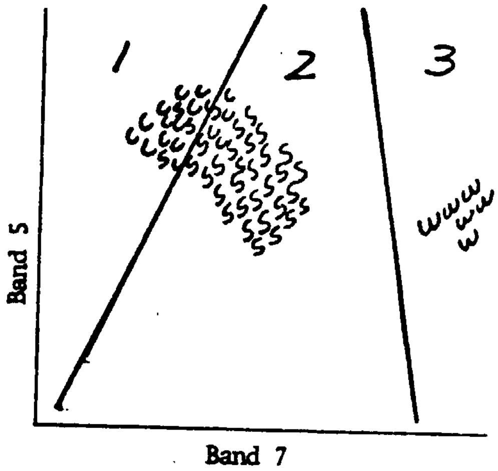


Figure 4. LANDSAT Frame Divided Into Two Parts

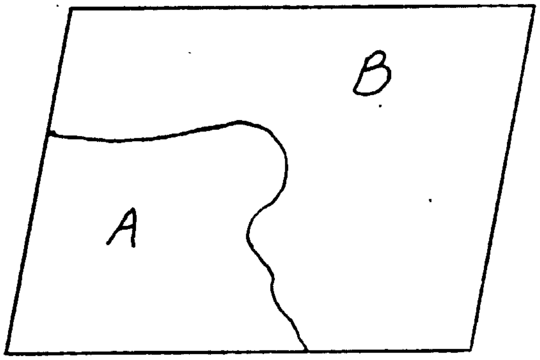


Figure 5. Population Relation Between Classified Pixels and Reported Acres

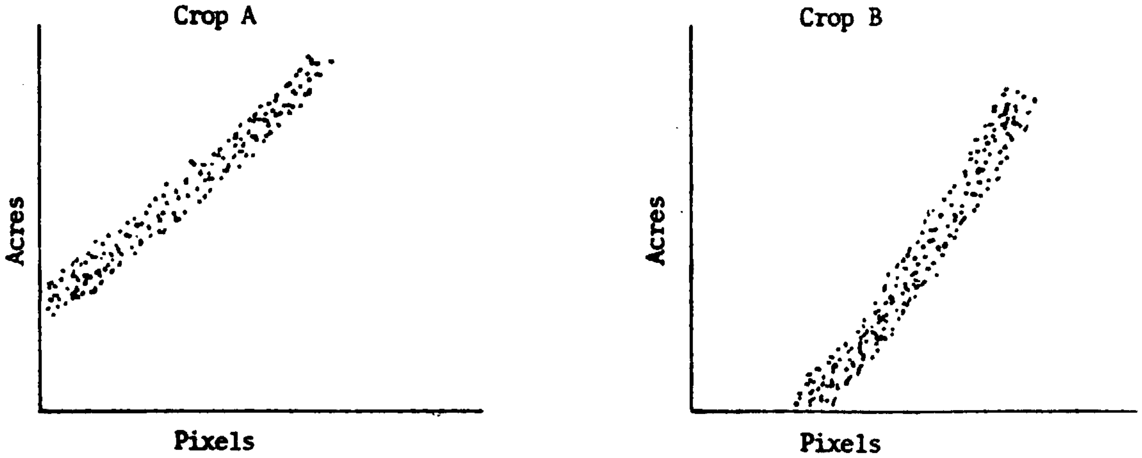


Figure 6. Sample Data Relation Between Pixels and Acres

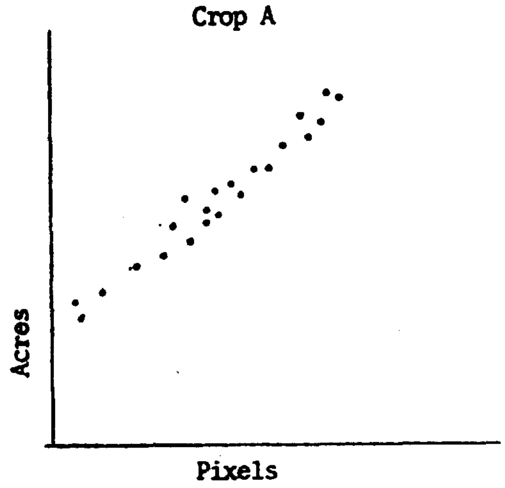
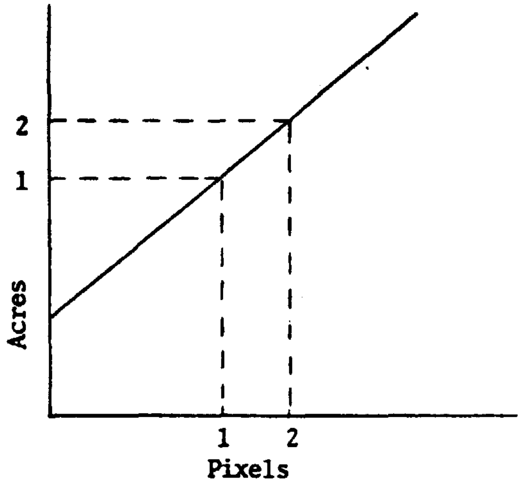


Figure 7. Estimated Population Relation



TABLES

TABLE I. ESTIMATES OF AGRICULTURAL COVER TYPES

Crop or Cover Type	Reported Acres July 27	Regression Estimate	Pixel Count x 1.114
		-- (000 acres) --	
Corn	1,286	1,390	2,105
Soybeans	631	701	610
Perm. pasture	533	434	678
Hay	179	154	104
Alfalfa	69	71	14

TABLE II. VARIANCES OF AGRICULTURAL COVER TYPES

Crop or Cover Type	Variance Reported (10^6 acres^2)	Variance Regression Estimate (10^6 acres^2)	Information Gain or Loss (1) ÷ (2)
	(1)	(2)	(3)
Corn	17,202	2,459	7.00
Soybeans	5,880	847	6.94
Perm. pasture	4,489	1,096	4.09
Hay	630	376	1.67
Alfalfa	155	135	1.14

TABLE III. PRODUCTION STATISTICS DERIVED FROM ACREAGE AND YIELD

Crop	Area (1,000 Acres)	Yield per Acre (bu)	Production (1,000 bu)
Corn			
Sept. 1, 1975	1,390	119.9	166,661
Harvest 1975	(1,390)1/	123.8	172,082
Soybeans			
Sept. 1, 1975	701	38.2	26,778
Harvest 1975	(701)1/	39.9	27,970

1/ Unchanged from September 1.