APPLICATION OF REMOTE SENSING FOR RICE: UNITED STATES EXPERIENCE

Galen F. Hart, Paul W. Cook, and William H. Wigton
United States Department of Agriculture, Washington, DC, USA

1981

This paper describes procedures that integrate Landsat technology into the area sampling frame methodology to improve estimates of crop area. The methodology described is currently being used in four states in the United States, and it is planned that by 1985, ten states will be included. The orientation is statistical rather than "remote sensing" because estimating crops involves sampling, inference, and other principles of statistics. Land area probability sampling obtained data are combined with full frame digital Landsat data through an application of statistical regression. Results are presented for rice, soybeans and cotton area estimates in Arkansas. Although the procedures have been documented in other papers and reports, results of the Arkansas study have not been presented.

AREA SAMPLING FRAME

An area sampling frame is used by the Statistical Reporting Service (SRS) to make production estimates at both state and national levels. In addition, the use of the area sampling frame is crucial for the application of Landsat technology to improve these crop acreage estimates. Therefore, some time is devoted to explaining area sampling and its use.

The concepts of the area sampling frame are very simple:

1. Divide the total area to be surveyed into N small land areas (*sampling units*) without any overlaps or omissions.
2. Select a random sample of n sampling units (*segments*).

3. Obtain the desired data for reporting units of the population that are in the segments.

4. Estimate population totals by multiplying the sample totals by N/n (Houseman, 1975).

This procedure is used for crop acreage, livestock, and other agriculture parameter estimation, and is a dependable method. The use of random sampling in selecting units from the population accomplishes two things:

1. It gives a basis for making inference about the population of items included in a survey.
2. It provides a basis to determine the precision of estimates.

APPLICATION OF LANDSAT CLASSIFICATION

The satellite data discussed in this report are Landsat Multi-Spectral Scanner (MSS) data as described in Section 3 of the Landsat Data User's Handbook (1971).

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 (Baker and Mikhaul, 1975). The total radiance from an object is composed of reflected radiance represented at shorter wavelengths of the electromagnetic spectrum, and emissive radiance at the longer wavelengths. The combination of these two sources of energy represents a measurable spectral response of an object. This, then, is the *spectral signature* of an object and it is the differences between such signatures which allow the classification of objects by using the statistical techniques about to be discussed.

Classification Techniques

The object of the discussion that follows is to provide a description of how to classify a Landsat frame. The way this is done in the computer is by the use of *discriminant functions* to differentiate between crops on the basis of reflected energy. There must be two or more crops and a sample of individual pixels for each. A rule is established that uses a sample of pixels for each crop to allocate some unknown crop pixel, outside the sample, to a crop type. If all data in a Landsat frame were plotted in a scatter diagram, it might appear as in Figure 1.

A brief study of Figure 2 reveals some obvious but important features:

1. The location of the center of these concentric circles directly affects the type of rule to be followed in separating crops or land covers.
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. For example, soybeans have wide variability.

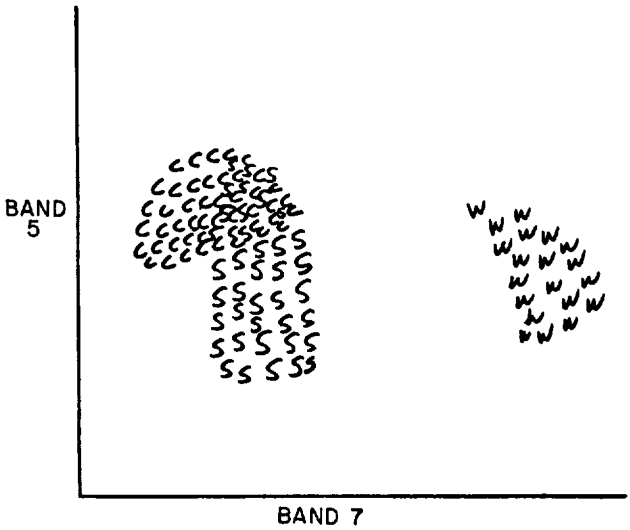

Figure 1. Scatter diagram of all values on one Landsat frame for three crops: C—corn, S—soybeans, and W—water.

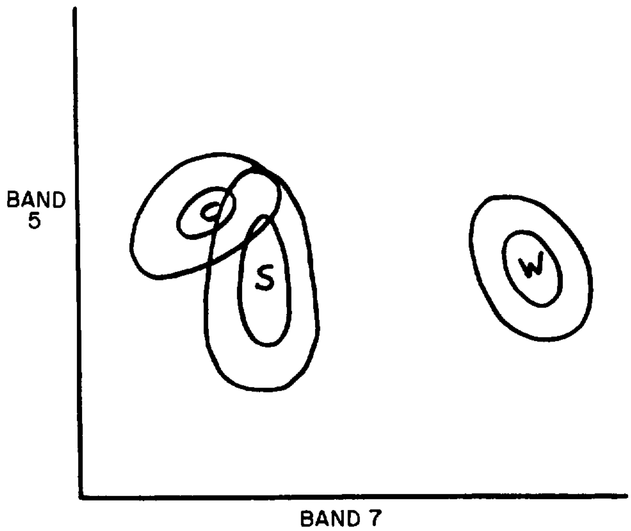

Figure 2. Confidence limits for data in Figure 1.

4. It will be impossible to distinguish the crop if the reflected energy coincides with the overlap region of corn and soybeans, because both are possible.

It would be ideal if the data for each crop were as far apart as water from corn, the spread were as small as the spread for water and elliptical in form, and there were no areas of overlap. These items cannot be controlled, since the sensor bands and band widths determine the location of the centers and 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, National Aeronautics and Space Administration (NASA) preprocessing, and many other circumstances. Overlap areas where mislabeling or misclassification is inevitable are a natural state. Some items reflect solar energy similarly and cannot be separated. The natural state of things cannot be changed and therefore must be dealt with as they are or not at all. However, from a sample, the scatter diagram of the population can be estimated; this is a statistical problem.

A valid statistical estimate requires a random sample from the population of interest. All parts of the population of interest must have a chance of selection and the sample size must be large enough to represent the population adequately. If estimates are needed that are quite accurate, then a fairly substantial sample size is required.

The area sampling frame is ideal because a valid statistical estimate can be made for the Landsat frame. In addition, a random sample of all possible sampling units is available and reflected energy for the crop types can be determined for the sample fields inside the selected sampling units or segments. These signatures are estimated for the scene they are in; therefore, 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 straightforward; they amount to drawing lines that partition the space. Figure 3 shows an example of this procedure.

All pixels that need crop labels should be plotted on the partitioned space. If they fall in partition C, the label is corn, even though some soybeans will be misclassified; obviously, water is separable.

The location, size, and shape of these population scatter diagrams shift relative to each other in different scenes and even in different parts of the same scene. Hence, using part of a Landsat scene to label pixels from another locale is hazardous. As an example, there are two cases; one is reasonable, the other is not. Figure 4 shows a possible division of a Landsat scene. Section A is divided into 600 small parts and a random sample of 60 parts is drawn from the 600. This sample may or may not be truly representative. If it is, the reflective and emitted energy (the signature) from these 60 segments adequately represents the reflected energy in all of section A. The use of signatures in the

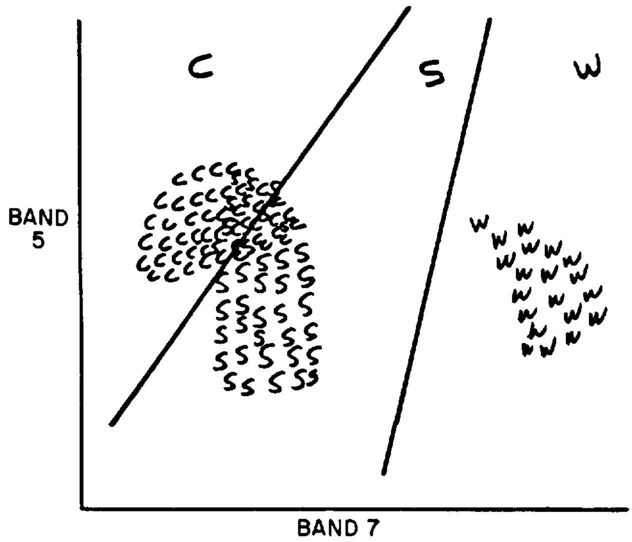

Figure 3. Partioned space showing population scatter diagram.

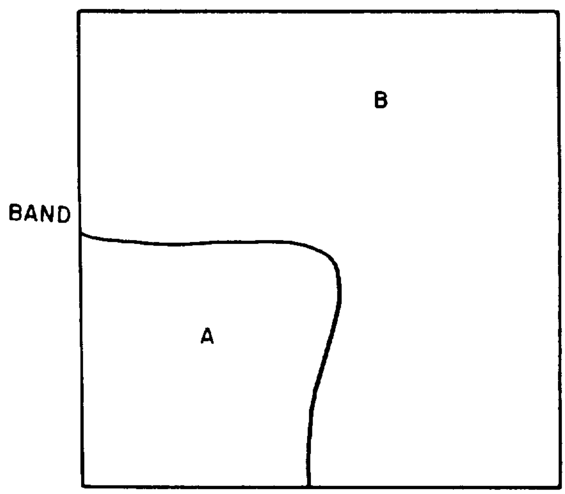

Figure 4. Landsat frame divided into two parts.

sample of 60 segments to classify the 600 is not an example of signature extension. This is simply a valid statistical inference. It is a commonly misunderstood notion that it is not necessary to draw a sample randomly from the population of interest to make an inference for that population; indeed a random sample is a necessity for proper statistical inference.

To classify crops in section B, it would be necessary to divide section B into sampling units and draw a random sample from these as representative of signatures in section B.

Model Utilizing Landsat

In order to make use of Landsat to reduce the sampling variation, a linear relationship between classified pixels for a crop and the area of the crop on a per sampling unit basis must be established. Figure 5 illustrates this relationship.

Again, these relationships are population relationships that are not known and therefore must be estimated from a sample. Data from area frame sample segments can be used to estimate this relationship. The sample observations for crop A are shown in Figures 6 and 7. This is a per segment relationship. Therefore, a segment can be located in Landsat, the segment classified, and a count made of the pixels in crop A. If the pixels for crop A turn out to be at point 1, then the corresponding value on the hectares axis is read. If the classified pixels for the segment turn out to be at point 2, then that value is read on the hectares axis.

This procedure could be completed for each segment in the population and all segments summed to get an estimate using satellite information across the whole area. However, all this is unnecessary. Since the total number of segments in the Landsat frame (N) is known, every pixel in the frame can be classified, and the total number of pixels in crop A divided by the number of segments in the frame. This is equal to the average number of pixels in crop A for the average segment.

The total number of pixels in crop A in sample segments (n) is known. With this information, the estimate from the area frame can be adjusted for the difference between the pixels in crop A for the sample (n) compared with the total number of pixels of the population (N).

Figure 7 illustrates how the adjustments would be made. 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 pixels is made on the hectares axis. The formula is:

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

where $\bar{Y}_{reg}$ is the adjusted number of hectares in the average segment. $\bar{Y}$ is the average hectares for crop A from the area sampling frame, $\bar{x}_{sample}$ is the corresponding average pixel values in the sample, and $\bar{X}_{total}$ is the average pixels

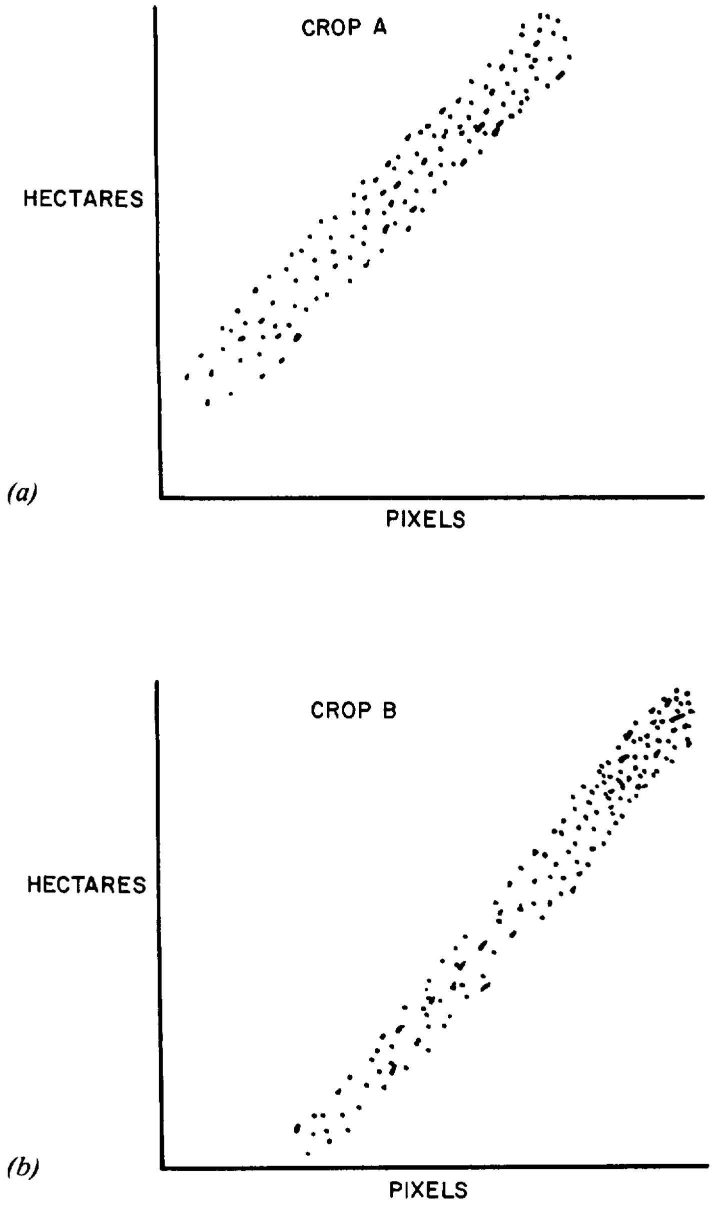

Figure 5. Population relationship between classification results and reported hectares of the same crop for one Landsat scene: Crop A and Crop B.

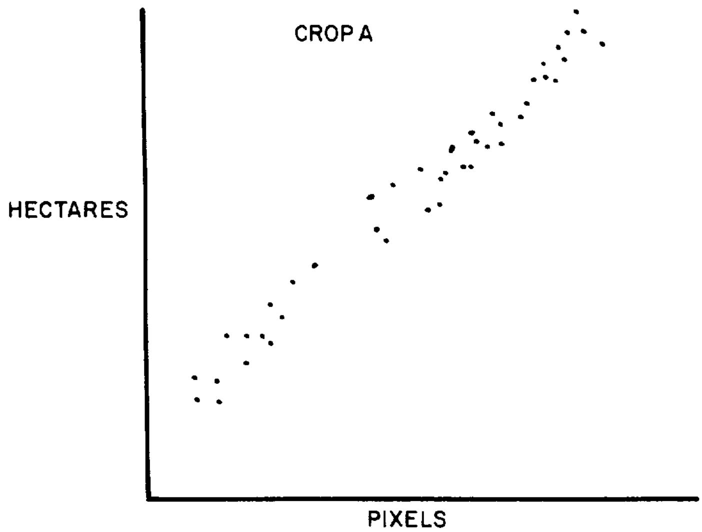

Figure 6. Sample data points for Crop A showing relationship between pixels and hectares.

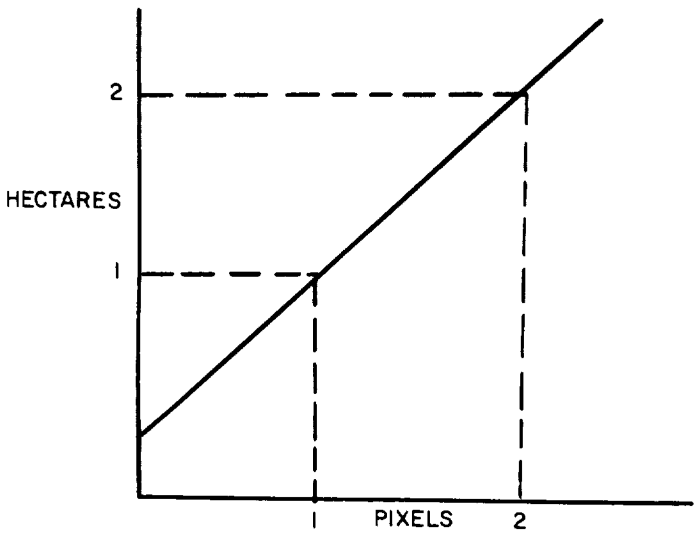

Figure 7. Estimated population linear relationship based on sample data in Figure 6.

per segment for the entire population (or stratum). $\bar{Y}_{reg}$ is then multiplied by N to get an estimate for the total. The variance for Y_{reg} is $[(n-1)/(n-2)](1-r^2)$ times the variance of $N \cdot \bar{Y}$. This regression model reduces the spread of the sampling error distribution by a factor of $(1-r^2)$.

In summary, with ground data for a properly selected statistical sample, as well as the computer classification for the same, the necessary information is available to adjust a full frame classification for all systematic errors. If there is a good linear relationship between ground data and what the computer classifies as being on the ground, the sampling error will be substantially reduced as compared with the results without remotely sensed data (Wigton, 1976).

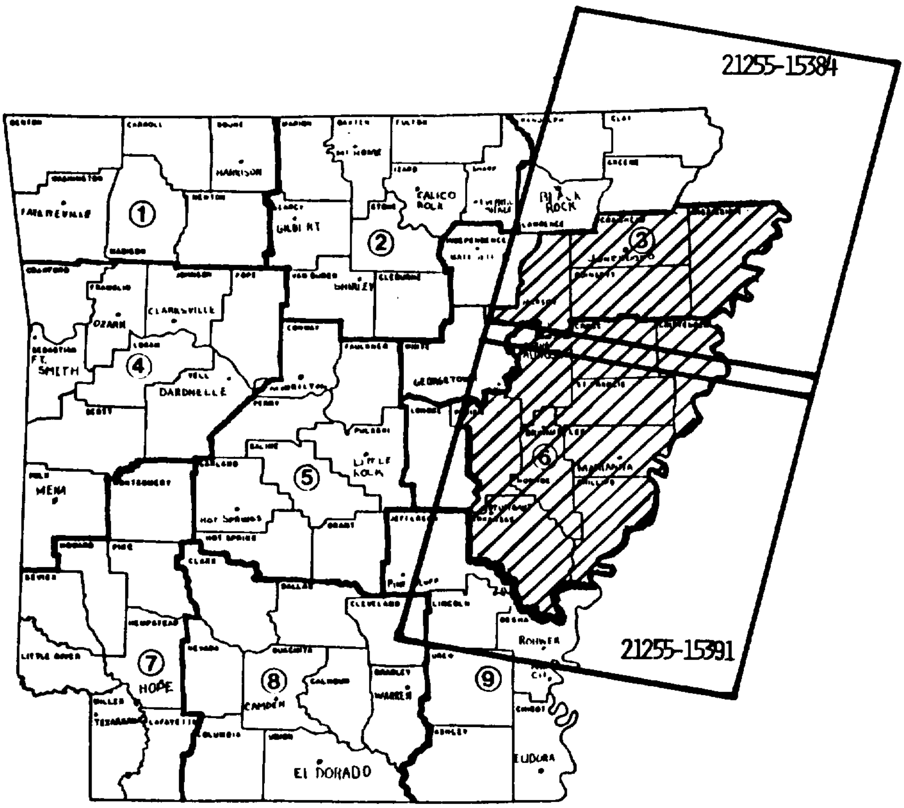

Figure 8. Arkansas study counties and Landsat coverage.

ARKANSAS EXAMPLE

Background

The primary objective in this project was the determination of area statistics for all major crops in the 13 counties falling within a single strip of two Landsat scenes along the eastern edge of Arkansas during the summer of 1978 (Fig. 8). The major rice- and soybean-producing counties are located in this area. Both rice and soybeans are major crops for Arkansas; an improvement in crop area estimates for these crops would be beneficial.

The crop rotation plan used in Arkansas ensures that both soybeans and rice will be grown within the same region because rice cultivation practices in Arkansas consist of a 3 or 4 year crop rotation of rice, soybeans, soybeans and rice (or oats, then rice), and sometimes will include a year of fish farming. A substantial investment in wells and an associated irrigation network of water pipes is also made to provide a sufficient water supply for proper rice growth. This outlay of funds usually means that once rice is grown in a given field, it will continue to be grown as part of the rotation plan of crops. Levees built to contain water are less than a foot high, since only 3 to 4 inches of water are needed within the field. Levees are leveled when the crop is mature to facilitate harvest. This also permits the producer to grow other crops on the same land.

The decision to restrict the areal extent of this study to two Landsat scenes made possible a thorough examination of a relatively extensive rice- and soybean-growing region. The degree of correlation between classified Landsat data and area sample frame survey data, collected in late May and early June, determined how well the Landsat data could be used for making more accurate acreage estimates for rice, soybeans, and cotton.

Landsat Data Acquisition

During the summer of 1978, Landsat II was operating very well and atmospheric conditions were excellent. There were three different dates in late June, middle July, and late August when Landsat scenes of high quality were available. Because of the Mississippi River and other waterways within the region, it is quite easy to determine data quality by examination of the Landsat transparencies.

By mid-July rice is starting to head so that the spectral appearance of flooded fields is less pronounced than earlier in the growth period, and thereby makes rice less distinguishable from other crops. Soybeans are planted early in this area to avoid heat stress during the early growth and reproductive periods. The 30 June date was chosen for analysis since the presence of water with a lower stand of rice would likely improve the chances of developing a unique signature for rice. Also, cotton should be recognizable at this time because of early planting and rapid vegetative growth.

Data Analysis

After choosing the Landsat scenes for analysis, registration of CCT digital data to a map base proceeded. A total of 32 control points were found within scene 21255-15391 on 7½ minute United States Geological Survey (USGS) quadrangle maps. After analysis using a full third-order linear polynomial and some correction of point locations, 31 points were accepted with an overall scene registration accuracy of 93 meters root mean square (rms) (Hanuschak et al., 1979).

Field boundaries from each area sampling frame segment were digitized by using a coordinate digitizer registering to the map base (Gleason et al., 1977). A registration file of coefficients was used to predict the location of each of the 109 segments contained in the two scenes.

Plots of the segment and field outlines were made to overlay grey-scale printouts of the locations for each segment (Hanuschak et al., 1979). Movements of each segment were made by using the lightness and darkness values of the printout that correspond to crop types for precision local registration. These shifts were, in general, one or two pixels in CCT row or column and in many cases were less than one pixel. No movements of greater than three pixels were encountered.

By using a masking operation within the computer, Landsat pixels corresponding only to the interior of each field were chosen (Hanuschak et al., 1979). These pixels were then clustered and signatures for each of the crops were obtained. In the total, there were 42 usable segments on scene 21255-15391 and 45 usable segments on scene 21255-15384. There were two segments that were located on both scenes within the overlap area.

By using the digitized segment files to extract and pack Landsat data into individual files by crop, it was possible to cluster each crop into many individual groups. This process provided greater separability and accuracy for crop classification on both Landsat scenes. The procedure used for clustering was a modified ISODATA program which allowed for reseeding the clusters after initial clusters were determined.[1]

Initially as many as 32 clusters were determined for each crop depending upon the number of pixels available. This was further reduced by means of a computer program which calculated the information matrix for the clusters and recommended which clusters should be combined.[1] These new cluster centers were used to recluster to obtain final clusters for that crop.

[1]Cook, P.: 1976 Follow-On Study, United States Department of Agriculture, SRS, 1978 (unpublished).

This method of clustering requires a final statistics file that contains numberous clusters.[1] For the two Arkansas scenes, the final file for 21255-15391 contained 60 categories (water was added) and for 21255-15384, there were 64 categories. The following list shows the breakdown in the number of clusters used for each scene for each crop.

Crop	Number of clusters from scene	
	21255-15384	21255-15391
Rice	14	7
Pasture	6	8
Wasteland	10	7
Soybeans	22	27
Cotton	10	8
Oats		1
Other hay		1
Water	2	1
	64	60

Both equal priors and prior probabilities, determined by the historic proportion of each crop within the sample, were used to classify the segment data (Gleason et al., 1977). Relative efficiencies and r^2's calculated from both of these samples indicated better results were obtained by using equal priors for the major crops (Hanuschak et al., 1979). Therefore, equal priors were used in the final classification of all the Landsat data.

Relative efficiencies (RE) and r^2's on one scene 15384 were as follows:

	RE	r^2
Rice	5.14	0.69
Pasture	3.45	0.58
Wasteland	1.47	0.19
Soybeans	2.94	0.45
Cotton	1.53	0.39

All pixels within the 13 county area for both scenes were classified by using the appropriate statistics file. Since the county strata were also digitized in a like manner to the segment digitization, aggregation files for each county by stratum were made so that regression estimates could be made at the county level.

[1]See footnote on preceding page.

Crop Acreage Estimation

The regression estimator was used to make estimates for all strata where an adequate number of segments were available (Gleason et al., 1977). For one agricultural stratum, of lesser importance, this was not possible since only two or three segments had been selected and so the variances were large. Consequently, only the area sample was used to estimate those areas where there were too few segments for the regression estimator to be used.

The formulas used to establish "swiss-cheese" estimates were as follows (Gleason et al., 1977):

$$\frac{N_c}{N_s} E_s = \text{County Estimate of Total}$$

and

$$\frac{N_c^2}{N_s^2} V_s = \text{County Variance}$$

where N_s is the number of frame units for a stratum, E_s is the area sampling estimate of a total at the stratum level, V_s is the variance of a total at the stratum level, and N_c is the number of frame units in a county to be pro-rated. This formulation permitted Landsat regression estimates to be made even in those areas where the variance for the regression estimator were too large because of an insufficient number of selected segments.

Results of this estimation procedure are presented in Tables 1, 2 and 3. A substantial increase in estimating precision was achieved for the two dominant crops, rice and soybeans. The regression estimates for all three crops compare favorably with the independently determined state estimate. It should be noted that the precision of some estimates at the county level are above maximum standards determined by SRS as acceptable for publication without supporting independent data.

CONCLUSIONS

Results from this study are encouraging for future use of this procedure in estimating rice areas in multiple county contiguous areas in the United States by using Landsat data within 3 weeks after final planting. The reductions in coefficients of variance (CVs) and the favorable relative efficiencies for the group of counties within a scene (values in the three or greater range), are comparable with those experienced in other unitemporal analysis conducted by the Statistical Reporting Service.

Table 1. Rice (1978 Planted Hectares—Arkansas Study Area)

County	Regression estimate, ha	CV(%)	Independent state estimate, ha
Craighead	29,340	6.9	27,920
Crittenden	14,000	5.0	4,860
Cross	21,080	8.3	27,920
Jackson	29,750	5.7	29,540
Mississippi	19,950	14.5	3,640
Poinsett	35,010	6.4	43,300
Woodruff	19,510	13.0	19,430
(Ark 15384)	(168,640)	(9.0)	(156,610)
Arkansas	44,400	7.7	45,330
Lee	14,330	11.8	10,120
Monroe	18,210	26.8	14,970
Phillips	18,900	15.9	2,830
Prairie	24,810	9.1	27,110
St. Francis	18,050	16.0	15,380
(Ark 15391)	(138,700)	(12.2)	(115,740)
Regression estimate	307,340	7.4	
Independent state estimate	272,350		
Area sampling estimate	321,000	12.9	

Table 2. Cotton (1978 Planted Hectares—Arkansas Study Area)

County	Regression estimate, ha	CV(%)*	Independent state estimate, ha
Craighead	20,520	*	29,460
Crittenden	13,800	*	12,750
Cross	8,740	*	1,620
Jackson	17,280	*	3,890
Mississippi	19,790	*	42,650
Poinsett	18,540	*	23,070
Woodruff	13,680	*	8,300
(Ark 15384)	(112,350)	(16.9)	(121,740)
Arkansas	200	*	200
Lee	9,350	*	12,950
Monroe	7,080	*	13,760
Phillips	13,840	*	17,000
Prairie	6,600	*	1,050
St. Francis	4,780	*	7,490
(Ark 15391)	(41,850)	(27.6)	(52,450)
Regression estimate	154,200	14.4	
Independent state estimate	174,190		
Area sampling estimate	149,050	19.4	

*Not computed at county level.

Table 3. Soybeans (1978 Planted Hectares—Arkansas Study Area)

County	Regression estimate, ha	CV(%)	Independent state estimate, ha
Craighead	86,850	4.4	70,820
Crittenden	80,660	4.1	97,130
Cross	81,990	4.1	84,990
Jackson	71,190	5.0	82,960
Mississippi	139,540	3.6	151,760
Poinsett	89,680	4.4	89,030
Woodruff	64,910	4.9	76,890
(Ark 15384)	(614,820)	(4.1)	(653,580)
Arkansas	54,920	6.2	91,060
Lee	65,930	3.7	72,850
Monroe	47,510	2.2	56,660
Phillips	78,070	3.6	105,220
Prairie	47,750	5.4	70,820
St. Francis	63,660	4.0	80,940
(Ark 15391)	(357,840)	(5.6)	(477,550)
Regression estimate	972,660	3.3	
Independent state estimate	1,131,130		
Area sampling estimate	1,094,010	4.6	

REFERENCES

Anon. (1971): LANDSAT Multi-Spectral Scanner Data User's Handbook. Goddard Space Flight Center, Greenbelt, Maryland.

Baker, J. R., and E. M. Mikhaul (1975): Geometric Analysis and Restitution of Digital Multispectral Scanner Data Arrays. LARS Information Note 052875.

Gleason, C., R. Starbuck, R. Sigman, G. Hanuschak, M. Craig, P. Cook, and R. Allen (1977): The Auxiliary Use of LANDSAT Data in Estimating Crop Acreages: Results of the 1975 Illinois Crop-Acreage Experiment. United States Department of Agriculture, SRS, SRS-21, October 1977.

Hanuschak, G., R. Sigman, M. Craig, M. Ozga, R. Luebbe, P. Cook, D. Kleweno, and C. Miiller (1979): Obtaining Timely Area Estimates Using Ground-Gathered and LANDSAT Data. United States Department of Agriculture, SRS, Tech. Bull. No. 1609, August 1979.

Houseman, E. E. (1975): Area Frame Sampling in Agriculture. United States Department of Agriculture, Washington, D.C.

Kleweno, D., and C. Miller (1980): 1980 AgRISTARS DC/LC Project Summary Crop Area Estimates for Kansas and Iowa. United States Department of Agriculture, SRS, Report No. AGESS-810414.

Wigton, W. H. (1976): Use of LANDSAT Technology by the Statistical Reporting Service, Purdue LARS Symposium on Machine Processing of Remotely Sensed Data, June 1976.