Reprinted from

Symposium on Machine Processing of Remotely Sensed Data

June 21 - 23, 1977

The Laboratory for Applications of
Remote Sensing

Purdue University
West Lafayette
Indiana

IEEE Catalog No.
77CH1218-7 MPRSD

CROP IDENTIFICATION AND AREA ESTIMATION BY COMPUTER-AIDED ANALYSIS OF LANDSAT DATA

MARVIN E. BAUER, MARILYN M. HIXSON,
BARBARA J. DAVIS, AND JEANNE B. ETHERIDGE
Purdue University

ABSTRACT

This report describes the results of a study involving the use of computer-aided analysis techniques applied to Landsat MSS data for identification and area estimation of winter wheat in Kansas and corn and soybeans in Indiana. Key elements of the approach included use of aerial photography for classifier training, stratification of Landsat data and extension of training statistics to areas without training data, and classification of a systematic sample of pixels from each county. Major results and conclusions are: (1) Landsat data was adequate for accurate identification and area estimation of winter wheat in Kansas, but corn and soybean estimates for Indiana were less accurate; (2) computer-aided analysis techniques can be effectively used to extract crop identification information from Landsat MSS data, and (3) systematic sampling of entire counties made possible by computer classification methods resulted in very precise area estimates at county as well as district and state levels.

I. INTRODUCTION

In 1972 the world food situation changed dramatically as world food production declined for the first time in many years at a time of rapidly expanding demand. World food reserve stocks were reduced to a historically low level of less than a 30-day supply.

As a result of these events, the importance of accurate and timely crop production information to rational planning and decision making by governments, agribusinesses, producers, and consumers has been increasingly recognized. Some benefits of improved crop production information are: (1) accurate estimates result in price stability; (2) timely and accurate forecasts of production allow governments to plan domestic and foreign policies and actions; and (3) accurate forecasts enable optimal utilization of storage, transportation, and processing facilities. Conversely, the socioeconomic costs of not having accurate and timely information available are substantial. Most countries forecast and estimate their crop production, but relatively few have reliable methods for gathering the necessary data. Recommendations to improve our capability to monitor crop production have been made by the National Academy of Sciences[12] and the United Nations World Food Conference.[17]

During the past decade considerable evidence has developed that multispectral remote sensing from aerospace platforms can provide quantitative data which can be effectively used to identify major crop species and determine their areal extent. A brief review of the development of the technology leading up to the study will help put this study in perspective and show the progress made. In 1964, multispectral photography was collected for the first time over agricultural fields, and the potential of the multispectral approach to crop identification was recognized.[5] After this approach was further defined, a crop classification was made from multispectral scanner data in 1967, using pattern recognition methods implemented on a digital computer.[9] One of the first investigations using satellite-acquired imagery to identify crops was performed by Anuta and MacDonald.[1] The Corn Blight Watch Experiment,[10] conducted in 1971 over seven Corn Belt states, provided a prototype remote sensing system which successfully integrated techniques of sampling, data acquisition, processing, analysis, and information dissemination in a quasi-operational system environment. The results showed that remote sensing from aircraft-mounted sensors could be used to quantitatively recognize corn leaf blight, as well as other agricultural crops and land uses over broad areas. Bauer and Cipra[3] used multivariate pattern recognition methods implemented on a digital computer to classify Landsat-1 data acquired over a three-county area in northern Illinois. Area estimates for corn and soybeans for the three-county area were within 1.5 and 1.1 percent, respectively, of those made by the U.S. Department of Agriculture. The conclusion from these as well

This research was sponsored by the National Aeronautics and Space Administration, Goddard Space Flight Center (Contract NAS5-20793).

as other studies is that remote sensing techniques may prove to be a more accurate, precise, timely, and/or cost effective method of acquiring crop production information than conventional surveys carried out on the ground. Remote sensing from satellites is particularly appropriate for crop surveys because of the capability to obtain repetitive coverage of wide areas.[2]

II. OBJECTIVES

The overall objective of the investigation was to develop and test procedures utilizing Landsat data to not only identify, but more importantly, determine the areal extent and distribution of earth surface features over large geographic areas. The specific application selected for investigation was crop identification and area estimation for two states in the Central United States.

The specific objectives of the study were:

- Using Landsat data and computer-implemented pattern recognition, classify the major crops from regions encompassing different climates, soils, and crops.
- Estimate crop areas for counties and states using the crop identification data obtained from the Landsat classifications.
- Evaluate the accuracy, precision, and timeliness of crop area estimates obtained from Landsat data.

Two important underlying premises tested in the investigation were:

- The synoptic view of Landsat provides the opportunity to obtain crop production information over large areas, e.g. states.
- By using computer-implemented data analysis to classify pixels distributed over entire counties, it is also possible to make accurate and precise estimates for local areas, e.g. counties.

An important distinction between this experiment and the Large Area Crop Inventory Experiment (LACIE) being conducted by the USDA, NASA, and NOAA is the method of sampling and estimation. LACIE has followed conventional sampling methods and, for example, its estimates for the United States are based on 638 segments 5x6 nautical miles in size.[11] On the other hand, the wide area coverage of Landsat, linked with computer processing as in this study, offers a unique opportunity to improve upon the sampling methods now used for making area estimates from ground-based systems.

III. SELECTION AND DESCRIPTION OF TEST AREAS AND CROPS

Kansas and Indiana were selected as the test states; winter wheat in Kansas and corn and soybeans in Indiana were selected as the crops for which area estimates would be made from classifications of Landsat data. The test areas and crops were selected to sample the range of crop, soil, and management conditions which are present in the Great Plains and Corn Belt regions of the United States.

Winter wheat in Kansas is typically grown in relatively large fields and its crop calendar is quite different than any of the other crops or cover types. On the other hand, corn and soybeans in Indiana are grown in smaller fields, the soils are less uniform, and the crop calendars for corn and soybeans are similar to most other cover types in the state. Considering the spectral and spatial characteristics of Landsat data, corn and soybean identification and area estimation in Indiana is a more difficult problem than is winter wheat in Kansas.

IV. EXPERIMENTAL APPROACH AND PROCEDURES

The approach used was based on procedures developed and utilized in previous research at LARS with the objective of extending them to larger areas. The procedures were based upon five fundamentals determined early in the investigation:

- The classifier would be trained and tested using aerial photography as reference data.
- Counties without reference data would be classified using training statistics from an adjacent county having similar crops and soils and lying in the same Landsat frame.
- Area estimates would be made from a systematic random sample of pixels distributed over the entire county.
- Area estimates would be made on a county basis and aggregated to district and state levels.
- Estimates would be adjusted for classification bias.

The implementation of the basic steps is illustrated in Figure 1.

A. ACQUISITION AND SELECTION OF LANDSAT DATA

The selection of a Landsat scene to classify for a given county was based upon the date of the Landsat data, the location of ground truth, and the amount and location of cloud cover. The desired attributes were: the crops of interest were spectrally discriminable at the time of the

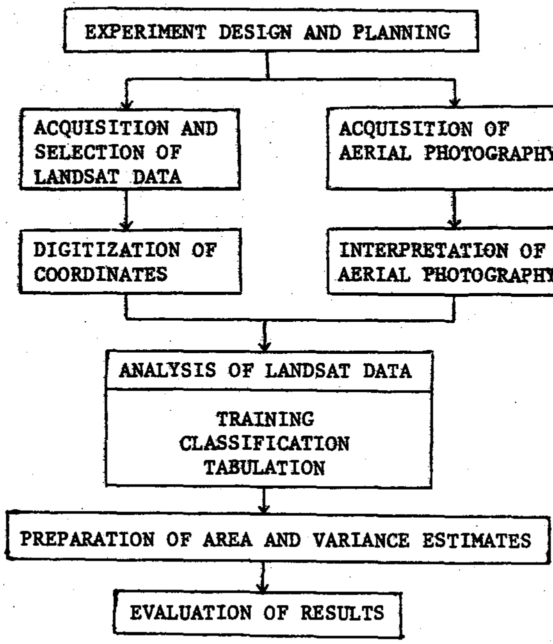

Figure 1. Implementation of Experimental Approach.

Landsat pass; aerial photography was available over areas lying in the same Landsat scene and having similar crops and soils; and both the county to be classified and the training areas were not obscured by clouds or bad data lines.

The amount of cloud cover created a serious problem in obtaining data for northeastern Kansas and much of Indiana. As a result, satisfactory data for classification was not available for the Northeast and East Central districts of Kansas. In Indiana, the only districts that had complete Landsat coverage were the Northwestern, West Central, Central, and East Central. Fifteen frames of Landsat data acquired over Kansas during March to June and six frames acquired during July, August, or September over Indiana were classified.

B. ACQUISITION OF AERIAL PHOTOGRAPHY

Multidate aerial photography was acquired for use as reference or "ground truth" data for training the classifier and evaluating classification accuracy. After studying soil, climatology, and land use maps, flightlines were selected throughout each state to sample the variation in soils, land use, and crops. Six flightlines in Kansas and five in Indiana were selected following major highways oriented north-south so that the photography and Landsat data could be coordinated easily. A 70 mm Hulcher two-camera system was used with color infrared and color transparency film. The average altitude for each flight mission was 3,000 meters yielding photography of approximately 1:80,000 scale. Each frame of photography covered an area roughly four km square (2.5 miles square). In Kansas, the photography was acquired on April 29-30 and June 26-27. In Indiana, photography was acquired in early May, early July, and mid-August to early September.

C. DIGITIZATION OF LANDSAT DATA COORDINATES

The Landsat coordinates of county boundaries were needed to make county crop estimates. Additional points were required along the flightline to assist the analyst in matching a computer map of Landsat data to the aerial photography. To find coordinates, the following procedure was used: (1) locate 25-30 checkpoints in the Landsat scene and digitize these checkpoints along with points defining county boundaries from a 1:250,000 scale USGS map; (2) for each county having aerial photography, digitize three to eight points along the flightline; (3) use a bivariate quadratic regression routine to fit coordinates of the checkpoints from the Landsat scene to the corresponding coordinates on the USGS maps. Then calculate and record on maps the Landsat coordinates for points defining county boundaries and checkpoints along the flightline.

D. INTERPRETATION OF AERIAL PHOTOGRAPHY

Large scale aerial photography was used as reference data following the assumption that the crops of interest could be readily and accurately identified. Standard photointerpretation techniques were used to identify fields of wheat and nonwheat in Kansas and fields of corn, soybeans, and "other" in Indiana. The coordinates of the identified fields were then located in the Landsat data. Wheat was relatively easy to identify in Kansas; corn and soybeans were somewhat more difficult to identify in Indiana. Fields which were not positively identified were not included as either training or test fields. Problems in photointerpretation, therefore, resulted in smaller training sets rather than inaccurate identification. Two general problems, clouds or haze and improper film exposure, were occasionally encountered, but did not seriously affect the photointerpretation process.

E. ANALYSIS OF LANDSAT DATA

The Landsat data analysis involved computer-assisted techniques utilizing the LARSYS Version 3 multispectral data analysis system, a software system developed by Purdue/LARS which used pattern recognition for analyzing remote sensing data.[14,16]

The procedure (outlined in Figure 2) involves: (1) defining a group of spectral classes; (2) specifying these to a statistical algorithm which calculates a set of defined statistical parameters; (3) utilizing the calculated statistics to "train" a pattern recognition

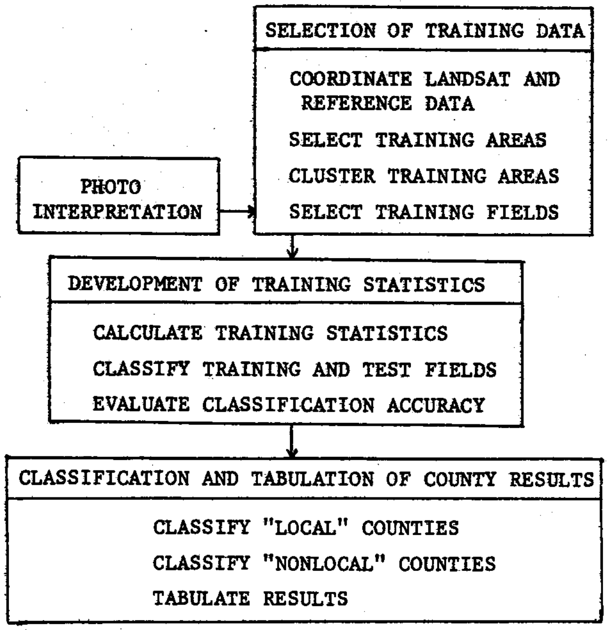

Figure 2. Flowchart of Procedures Used in Analysis of Landsat Data.

algorithm; (4) classifying each data point within the data set of interest into one of the training classes; and finally (5) displaying the classification results in map and/or tabular format.

Selection of Training Data. The accuracy of classification results is highly dependent upon the training data. Selection of training areas was based on two factors: first, the amount and quality of reference data (aerial photography) available, and second, the presence of a representative sample of cover types of the areas to be classified.

Training areas of 100 lines and 100 columns (approximately 8 x 5.5 km) of Landsat data were dispersed along the flightline throughout the county in order to adequately represent the variation present. To facilitate locating agricultural fields in the Landsat data, a spectral class map was produced by clustering each training area using all four wavebands. After matching the cluster maps with the corresponding frames of aerial photography, the boundaries and identities of fields were sketched on the cluster map.

Training fields had to meet three criteria: (1) the cover type of the fields selected for training had to be positively identified by the photo-interpreter; (2) the fields themselves must be of only one cover type; and (3) the training fields must adequately represent the variation present in the cover types throughout the area to be classified. The Landsat coordinates of field center (non-boundary) pixels were then obtained and field description cards prepared.

Development of Training Statistics. The field center samples of each of the major cover types were clustered separately to define subclasses by the natural groupings or spectral classes within the cover types. Each of these subclasses must be a unimodal distribution to satisfy the assumptions of the maximum likelihood Gaussian classifier. Statistics were calculated to represent each spectral class and the separability of classes was assessed using transformed divergence.

Test or training field classification results were used to evaluate the adequacy of the training statistics before the county was classified in order to allow for additional training if required.

Classification and Tabulation of County Results. The final training statistics were used to classify a systematic random sample of the Landsat pixels within each county. Either a one-fourth (every other line and column) or a one-sixteenth (every fourth line and column) sample was classified for each county. A sampling study showed that both sample sizes gave satisfactory precision.

When a county was classified with a training set at least partially trained with fields from that county, the classification was labelled "local". A "nonlocal" classification was one in which the training set did not contain any training fields from the county classified, but which came from a county in the same Landsat frame with similar soils and land use. In general, each training set was used to classify two to five counties.

F. PREPARATION OF AREA AND VARIANCE ESTIMATES

Following classification, crop area and proportion estimates were made. Estimates of the areal extent or proportion of a crop were desired for county, crop reporting district, and state levels. Steps in the area estimation procedure included: (1) calculation of the area and proportion estimates, (2) correction of the estimates for classification bias, and (3) calculation of variance estimates.

Area and Proportion Estimates. The Landsat estimated proportion of the ith crop in the jth county was calculated using the equation

$$\hat{p}_{ij} = \frac{n_{ij}}{n_j}$$

where n_{ij} is the number of pixels classified as crop i and n_j is the total number of pixels in an irregular polygon representing the county. The crop estimates were adjusted for large cities and nonagricultural areas. Area and proportion estimates for the crop reporting districts and the entire state were aggregated from the county estimates.

Correction for Classification Bias. Since it is inevitable that some pixels are incorrectly identified by the maximum likelihood classifier, the resulting area estimates may be biased. However, if the error rates are known, the area estimates can be unbiased after the classification has been performed.

An estimate of the classification error rates is the matrix of training or test field classification performance,

$$E = \begin{bmatrix} e_{11} & e_{12} \\ e_{21} & e_{22} \end{bmatrix}$$

where e_{ij} is the proportion of samples of type i classified as type j. If $\hat{P}$ is the vector of proportions estimated from the Landsat data and P is the vector of true proportions, then the adjusted estimates can be found by solving

$$P = (E^t)^{-1} \hat{P}$$

subject to the constraint $0 \leq p_i \leq 1$ for all p_i, elements of the vector P, or equivalently by solving

$$\min_{0 \leq p_i \leq 1} \quad || P - (E^t)^{-1} \hat{P} || .$$

The discussion of bias correction generalizes to n cover types of interest with E being an n x n matrix and the vectors having n components.

The corrected estimates will be unbiased if the error matrix found from the test or training field performance is the true error matrix. It may not be truly unbiased because of photointerpretation difficulties or because the flightline might not be representative of the entire area classified.

Calculation of Variance Estimates. Since each pixel either is or is not classified as crop i, the binomial distribution can be used to obtain the variance of the bias-corrected proportion estimates. A sampling study showed that the binomial theory gave a variance not significantly different from the true sample variance, so, for the ith crop in the jth county, an estimate of the variance is given by:

$$v(\hat{p}_{ij}) = \frac{\hat{\hat{p}}_{ij} \quad (1 - \hat{\hat{p}}_{ij})}{n-1} \quad (1-f_j)$$

where f_j is the county sampling fraction.[4] For individual county estimates, the sampling fraction can be ignored (though it is not negligible) to give a conservative estimate of the variance. The variance for a crop reporting district was calculated considering each county as a stratum and is given by:

$$\Sigma W_j^{\,2} \frac{\hat{\hat{p}}_{ij} \quad (1 - \hat{\hat{p}}_{ij})}{n_j} \quad (1-f_j)$$

where the summation is taken over all counties in the crop reporting district.[4]

G. EVALUATION OF RESULTS

Two quantitative evaluation techniques were used to judge the accuracy of crop classification and area estimates. One evaluation involved statistical sampling of individual areas of known cover types (designated as test fields). This offers an effective method of examining inclusive and exclusive classification errors for the various crops or cover types. Areas with a known cover type which were not used for training were chosen as test fields. These were then classified and the accuracy of the classifier determined by the proportions of pixels which were correctly identified. If these fields have been randomly selected and their classification accuracy is high, then the classification of the entire area should be accurate.

The second quantitative technique used for evaluating classification accuracy was comparison of area estimates from the computer classification and area estimates obtained by conventional methods. In this case, the standard of comparison for the Landsat estimates was the USDA/SRS estimate of acres harvested.[6,8,15] To avoid accepting the hypothesis that SRS and Landsat estimates were the same when they were, in fact, different, a large value of α, usually 0.25, was used in testing.

Tests were also made to identify and assess factors which might affect the accuracy of the area and proportion estimates including: date of the Landsat coverage, date of the aerial photography (Indiana only), effect of the data analyst (Kansas only), the effect of local versus nonlocal recognition, and the effect of geographic location (crop reporting districts).

V. WHEAT IDENTIFICATION AND AREA ESTIMATION IN KANSAS

In this section the results of the Landsat data analysis for winter wheat identification and area estimation in Kansas are presented and evaluated.

A. ANALYSIS OF FACTORS AFFECTING CLASSIFICATION ACCURACY

Several analyses to assess factors which might have influenced classification results were performed in order to more fully understand and interpret the results. Statistical tests showed that the date of Landsat coverage was not a major factor influencing the classification performance and that all counties regardless of the date of Landsat data can be considered together. Since there was no significant date effect, the effect of analysts on the classification performance could be considered. Because all analysts used similar methods, no inferences could be made about methodology; but it was concluded that individual analysts did not introduce a bias in the results.

One of the major problems encountered in the LACIE has been to develop a means for successfully extending training statistics from a training segment to "recognition" segments.[11] A test to determine if the stratification method employed in this investigation was satisfactory showed that there was some difference in accuracy between estimates for local and nonlocal counties, but that it did not have a strong influence on the overall results.

B. LANDSAT CLASSIFICATION RESULTS

Classification accuracy was determined by test field or training field performances. The overall classification performances were generally 85% or higher, an indication that the classification should result in accurate area estimates.

Classification bias correction was carried out on all proportion estimates because a study showed that: (1) the accuracy achieved by estimates which used training field performance matrices to calculate the bias was not significantly different from that achieved when test field performance matrices were used, (2) error matrices can be extended to nonlocal recognition counties, and (3) correction for the bias increased the accuracy of the estimates by decreasing the difference from the SRS estimates.

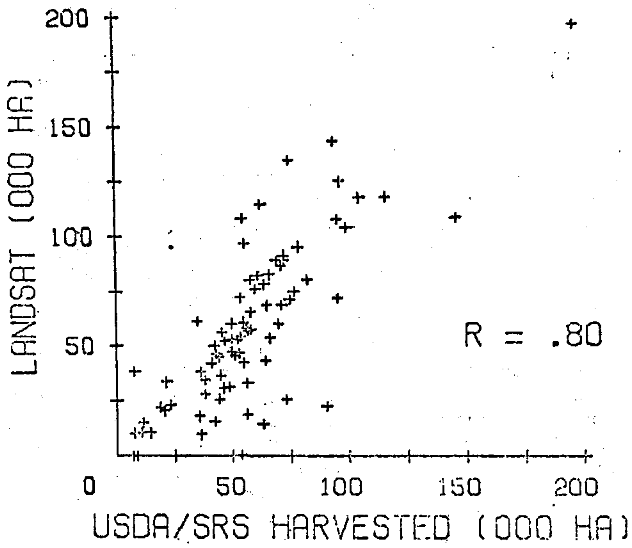

Figure 3. The Correlation of Landsat and USDA/SRS Estimates of the Area of Winter Wheat in Kansas Counties.

C. ACCURACY AND PRECISION OF WHEAT AREA AND PROPORTION ESTIMATES

Landsat estimates were calculated for 80 counties in Kansas and were compared to the corresponding USDA/SRS estimates. The two estimates were highly correlated with $r = 0.80 \pm 0.04$ for area estimates (Figure 3).

Table 1. Summary of USDA/SRS and Landsat Estimates of Area and Proportion of Wheat in Kansas.

Region	Area			Proportion			Relative Difference
	USDA/SRS	Landsat	Difference	USDA/SRS	Landsat	Difference	
	(000 Hectares)				(%)		(%)
State	4555	4613	58	26.2	26.6	0.4	1.3
District							
Northwest	470	387	- 83	23.3	19.2	- 4.1	- 21.5
North Central	578	575	- 3	25.1	25.0	- 0.1	- 0.5
West Central	522	579	57	25.2	28.0	2.8	9.9
Central	770	956	187	33.1	41.2	8.1	19.5
Southwest	784	715	- 68	25.6	23.3	- 2.3	- 9.6
South Central	1164	1158	- 6	40.2	40.0	- 0.2	- 0.5
Southeast	267	242	- 25	10.0	9.1	- 0.9	- 10.2
Counties							
(Median)	55.0	53.4	0.6	24.85	26.25	0.4	1.3

The accuracy of Landsat estimates of the area and proportion of wheat were assessed at three levels: state, district, and county (Table 1). At the state level, there was no difference at the 25% significance level in the proportion or area of wheat when comparing Landsat and SRS estimates. In all except one crop reporting district, there was also no significant difference between the two estimates. In the Central district, wheat was overestimated for every county compared to the USDA/SRS estimates, creating a significant bias in the Landsat estimates. However, all except two county estimates (which accounted for most of the difference) were close to the SRS estimates.

No statistical tests could be performed for differences from SRS estimates on a county-by-county basis because SRS does not calculate county variance estimates. Similarly, confidence limits cannot be placed around the SRS estimates. However, if the standard deviation of the SRS proportion estimates is assumed to be 10% at the county level, then 89% of the Landsat estimates were within a 90% confidence interval. For further comparison of Landsat and SRS county estimates, 49% of the counties were within ±5% (absolute difference) of SRS, 81% were within ±10%, and 88% were within ±15%.

The second measure of the quality of an estimate is its precision which refers to the size of the deviations from its expected value obtained by repeated application of the sampling procedure. The standard deviations and coefficients of variation (CV) of the Landsat estimates are extremely small even at the county level. The CV of the SRS estimate of wheat acreage in the state of Kansas is approximately 4%, compared to the CV of 0.06% for the Landsat estimate. The median CV of the Landsat county estimates is 0.60% which is smaller even than the 1.5% CV of the SRS national estimate of wheat acreage. Clearly the combined technologies of Landsat MSS data and computer-aided classification methods provide a means to make very precise crop area estimates.

VI. CORN AND SOYBEAN IDENTIFICATION AND AREA ESTIMATION IN INDIANA

The second state selected for analysis was Indiana; corn and soybeans, the two major grain crops in the state, were selected for study. As for Kansas, the factors affecting classification performance, comparisons of USDA/SRS and Landsat estimates of the area and proportions of the crops, and evaluations of the accuracy and precision of the Landsat estimates are discussed.

A. ANALYSIS OF FACTORS AFFECTING CLASSIFICATION ACCURACY

The effects of several factors likely to influence the accuracy of the Landsat area and proportion estimates were investigated. September was found to be a significantly worse time for acquisition of Landsat data and aerial photography for corn estimation than either July or August. July soybean estimates were slightly closer to SRS than those made from August data. There was some effect of local versus nonlocal classifications for corn estimation, but soybean estimates were equally accurate. Many additional factors such as field size, number of crops and cover types present, uniformity of soils, and production practices may have also influenced the results, but were beyond the scope of this investigation to pursue.

B. LANDSAT CLASSIFICATION RESULTS

Classification accuracy was determined for Indiana by the training field performance matrices. The training field classification performances were typically 75 to 85%. Although these accuracies were about 10% lower than those obtained in Kansas, they would generally be considered adequate for making satisfactory area estimates provided a consistent bias was not present. The area and proportion estimates, however, particularly on a county basis, were not as accurate as might have been predicted from the training field classification performances. This is believed to have been caused by a combination of two factors: (1) the proportion of pure pixels for Indiana fields which average only about 10 hectares in size is typically no more than 50%, but training statistics are calculated only on the basis of pure pixels and (2) since there was some difficulty in accurately identifying all fields and since positive identification of a field was required in order to use it for training, several spectral classes were omitted from training, biasing the classification performance upward.

All crop estimates were corrected for the classification bias because, on the average, this operation brought them closer to SRS estimates. For soybeans, there was no significant difference at any reasonable α level in the accuracy of corrected and uncorrected estimates. For corn estimates, however, corrected estimates were closer to SRS at the 20% significance level.

C. ACCURACY AND PRECISION OF CORN AND SOYBEAN AREA AND PROPORTION ESTIMATES

Plots comparing the Landsat and SRS county estimates of corn and soybean area, along with correlation estimates, are shown in Figures 4 and 5. The two sets of estimates are not as highly correlated as were the Kansas estimates; three counties, however, accounted for much of the lack of correlation of the corn estimates. The Landsat estimates for corn are consistently greater than the SRS estimates. On the other hand, the Landsat soybean estimates do not appear biased, but are clearly more variable than either the corn or Kansas wheat estimates.

Estimates were made for four Indiana districts using Landsat classification methods; these four districts together make up a "pseudo" state estimate which was tested against the SRS estimate

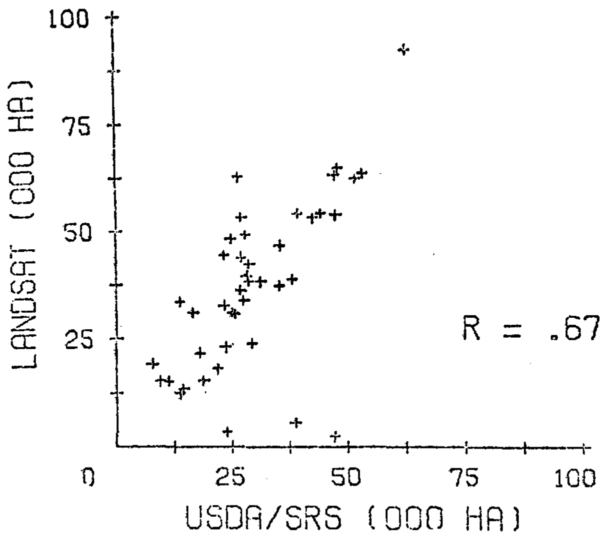

Figure 4. The Correlation of Landsat and USDA/SRS Estimates of the Area of Corn in Indiana Counties.

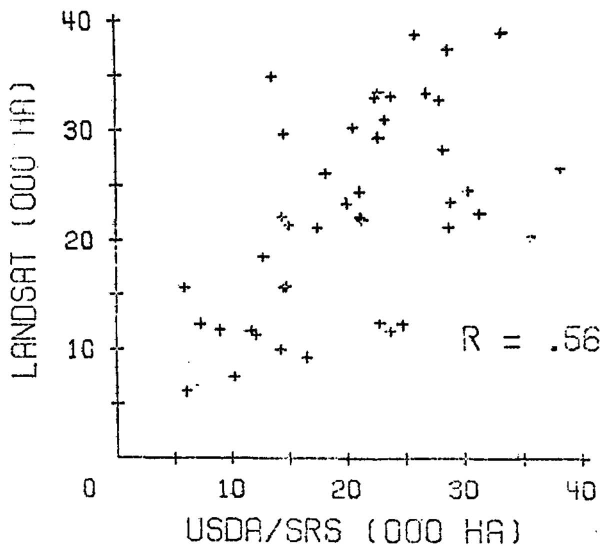

Figure 5. The Correlation of Landsat and USDA/SRS Estimates of the Area of Soybeans in Indiana Counties.

for the same area. Both Landsat corn and soybean proportion and area estimates were significantly different from the SRS estimates. Assuming that the SRS estimates were unbiased in these districts, the estimates derived from the Landsat classification were not as accurate as the SRS estimates. Corn estimates differed from SRS in three of the four crop reporting districts while soybean estimates differed in two of the four districts at the 25% significance level. Summaries of these results are presented in Tables 2 and 3.

Compared to SRS, the Landsat estimates of corn area and proportion were consistently overestimated. This is attributed in part to the spectral similarity of corn to other cover types, particularly trees, as well as to factors mentioned earlier such as boundary pixels. The soybean estimates, on the other hand, have a large variation but, when aggregated, were reasonably close to the SRS estimates.

The variances of the corn and soybean estimates were calculated from the binomial assumptions. As in Kansas, the sampling errors of the state, district, and county crop area estimates are very small. The coefficients of variation for the state estimates of corn and soybeans are 0.15 and 0.22%, respectively. The CVs for districts range from 0.23 to 0.56% and almost all county estimates have coefficients of variation less than 3%.

The generally lower level of performance in Indiana compared to Kansas is attributed to the greater number of crops and spectral classes to discriminate among; smaller, less homogeneous fields; less optimal timing of Landsat data acquisition; and less adequate reference or training data. A major difference between winter wheat identification in Kansas and corn and soybean identification in Indiana is that the crop calendar of winter wheat is different than most other cover types; whereas, corn and soybeans, both summer crops, have crop calendars similar to (i.e. are green at the same time as) other cover types present such as oats, hay, pasture, and trees. In summary, the identification of corn and soybeans in Indiana is a much more difficult problem than winter wheat identification in Kansas.

It should, however, be pointed out that accurate crop classifications have previously been achieved using aircraft scanner data.[10] Two particular limitations of Landsat MSS data are its spectral bands and spatial resolution. Work with aircraft data and more recently with Skylab data has clearly shown the importance of the middle infrared and thermal infrared portions of the spectrum for crop identification. Because the Landsat scanner does not obtain data in these important wavelength regions, we believe that the classification accuracies achieved are not as high as would be possible. Addition of at least one wavelength band in the middle infrared portion of the spectrum (1.3-2.6μm) and at least one channel in the 8-13.5μm thermal infrared region in future satellite scanner systems will unquestionably allow significant improvements in many of the results obtained, and in the utility of this type of satellite data. Further, the narrower and more optimally placed visible and near infrared bands of the proposed thematic mapper sensor on Landsat D will be a substantial improvement.[13]

Table 2. Comparison of USDA/SRS and Landsat Estimates of Area and Proportion of Corn in Indiana.

Region	Area USDA/SRS	Area Landsat	Area Difference	Proportion USDA/SRS	Proportion Landsat	Proportion Difference	Relative Difference
	(000 Hectares)				(%)		(%)
State	1285	1595	310	29.2	36.2	7.0	24.1
District							
Northwest	386	545	159	36	50	15	41.0
West Central	262	366	104	24	34	10	39.7
Central	474	472	- 2	30	30	0	- 0.4
East Central	162	212	49	24	31	7	30.3
Counties							
(Median)	27.3	37.3	9.3	28.4	38.9	8.8	23.8

Table 3. Comparison of USDA/SRS and Landsat Estimates of Area and Proportion of Soybeans in Indiana.

Region	Area USDA/SRS	Area Landsat	Area Difference	Proportion USDA/SRS	Proportion Landsat	Proportion Difference	Relative Difference
	(000 Hectares)				(%)		(%)
State	884	964	81	20.1	21.9	1.8	9.1
District							
Northwest	221	209	- 12	20	19	- 1	- 5.3
West Central	191	181	- 10	18	17	- 1	- 5.3
Central	328	405	77	21	26	5	23.6
East Central	144	170	25	22	25	4	17.5
Counties							
(Median)	21.1	22.1	3.1	21.5	20.9	3.0	16.4

The 80 meter IFOV of the current Landsat MSS appears generally adequate for areas having relatively large fields, but it is definitely a limitation in working in areas with field sizes of 10 hectares or less. The 30 meter IFOV of the proposed thematic mapper sensor would be a major improvement in that it would greatly reduce the proportion of "mixed" field boundary pixels and facilitate locating field boundaries.

VII. SIGNIFICANT RESULTS AND CONCLUSIONS

Many different phases of our investigation have produced results which we believe are significant in the development of remote sensing technology, particularly for crop surveys. The overall conclusions of the investigation are:

- Landsat MSS data was adequate to accurately identify wheat in Kansas; corn and soybean estimates for Indiana were less accurate.
- Computer-aided analysis techniques can be effectively used to extract crop identification information from Landsat data and make area estimates.
- Systematic sampling of entire counties made possible by computer classification methods resulted in very precise area estimates at county, district, and state levels.

- Training statistics can be successfully extended from one county to other counties having similar crops and soils if the training areas sampled the total variation of the area to be classified.

The synoptic view of Landsat provides the opportunity to obtain crop production information over very large areas, e.g. states and countries. By using computer processing techniques to classify pixels distributed over entire counties, it is also possible to make accurate and precise estimates for local areas, e.g. counties. These capabilities combining satellite, sensor, and computer make a worldwide, and at the same time, a local crop production information system possible.

Recommendations are made for increasing the number and placement of spectral bands, spatial resolution, and frequency of coverage for data acquired by future satellite systems, along with preprocessing to geometrically correct and register data sets. It is recommended that continued attention be given to developing more effective methods of scene stratification and large area training and classification methods.

In closing, we believe considerable progress toward an operational crop survey system was made as a result of this investigation. The results conclusively demonstrated the efficiency and applicability of computer-aided analysis techniques for estimating crop areas. Many of the techniques used in the investigation could be transferred to an operational system capable of producing accurate and precise crop area estimates for local areas such as counties, as well as for larger areas such as states or countries.

VIII. REFERENCES

1. Anuta, P.E. and R.B. MacDonald. 1971. Crop surveys from multiband satellite photography. Remote Sensing of Environment, 2:53-67.

2. Bauer, M.E. 1975. The role of remote sensing in determining the distribution and yield of crops. Advances in Agronomy, N.C. Brady, ed., Vol. 27, pp. 271-304, Academic Press, New York.

3. Bauer, M.E. and J.E. Cipra. 1973. Identification of agricultural crops by computer processing of ERTS MSS data. Proc. Symp. on Significant Results from ERTS-1, pp. 205-212, NASA SP-327, Washington, D.C.

4. Cochran, W.G. 1963. Sampling Techniques. John Wiley and Sons, Inc., New York.

5. Hoffer, R.M. 1967. Interpretation of remote multispectral imagery of agricultural crops. Purdue University Agricultural Experiment Station Research Bulletin 831.

6. Indiana Crop and Livestock Reporting Service. 1976. Indiana Annual Crop and Livestock Summary, 1975. Purdue University, West Lafayette, Indiana.

7. Johnson, M.E. Personal communication. August 23, 1976.

8. Kansas State Board of Agriculture. 1976. Kansas Farm Facts, 1975-1976. Topeka, Kansas.

9. Laboratory for Agricultural Remote Sensing. 1968. Remote multispectral sensing in agriculture. Purdue University Agricultural Experiment Station Research Bulletin 844.

10. MacDonald, R.B., M.E. Bauer, R.D. Allen, J.W. Clifton, J.D. Erickson, and D.A. Landgrebe. 1972. Results of the 1971 Corn Blight Watch Experiment. Proc. Eighth Int'l Symp. on Remote Sensing of Environment, pp. 157-190, Ann Arbor, Michigan.

11. MacDonald, R.B., F.G. Hall, and R.B. Erb. 1975. The use of Landsat data in a Large Area Crop Inventory Experiment (LACIE). Proc. Symp. on Machine Processing of Remotely Sensed Data, pp. 1B, 1-23, Purdue University, West Lafayette, Indiana. (IEEE Catalog No. 75 CH 1009-0-C).

12. National Academy of Sciences. 1977. World Food and Nutrition Study: Information Systems for World Food and Nutrition. Washington, D.C.

13. National Aeronautics and Space Administration. 1975. Landsat-D Thematic Mapper Technical Working Group Recommendations, Final Report. National Aeronautics and Space Administration, Johnson Space Center, Houston, Texas, JSC-09797.

14. Phillips, T.L., ed. 1973. LARSYS User's Manual. Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana.

15. Statistical Reporting Service. 1975. Scope and Methods of the Statistical Reporting Service. Miscellaneous Publication No. 1308, USDA, SRS.

16. Swain, P.H. 1972. Pattern Recognition: A Basis for Remote Sensing Data Analysis. Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana. Information Note 111572.

17. United Nations. 1974. Report of the World Food Conference. Rome, Italy.

Marvin E. Bauer, research agronomist and program leader, Crop Inventory Systems Research, B.S.A., and M.S., Purdue University, and Ph.D., University of Illinois, has had key roles in the design, implementation, and analysis phases of several major remote sensing projects including the 1971 Corn Blight Watch Experiment and the Crop Identification Technology Assessment for Remote Sensing Project. He has been the principal investigator of a LANDSAT investigation for crop area estimation survey. Currently, he is the technical leader of the NASA agricultural field measurements program.

Barbara J. Davis, statistician/analyst, Data Processing and Analysis Research, B.S. mathematics, Michigan State University with High Honor, joined the LARS staff in 1973. Her work at LARS has included algorithm development, crop inventory surveys, and the application of statistical methods to remote sensing problems; in particular, the design of experiments and the statistical evaluation of analysis results.

Marilyn M. Hixson, research statistician, Crop Inventory Systems Research; B.S., mathematics, Miami University; M.S., mathematical statistics, Purdue University, joined the LARS staff in 1976. Her area of specialization is the design and analysis of statistical experiments. Particular projects in which she has been involved include studies of sampling methodology for crop inventories and analysis of Landsat, helicopter, and field spectrometer data for crop identification and estimation.

Jeanne B. Etheridge, programmer/analyst, B.S., mathematics, University of Illinois; M.S., mathematics, North Dakota State University, joined the LARS staff in 1972; and has since been included in applications programming for remote sensing, computer systems programming, and analysis of remote sensing data. She managed the LANDSAT investigation for crop area estimation survey on which this paper is based and has also investigated the feasibility of multitemporal data analysis.