E7.6-10145
CR-146154
II

Quarterly Progress Report

Period: October 1 to December 31, 1975

I. Title of Investigation: Crop Identification and Acreage Estimation over Large Geographic Areas Using LANDSAT MSS Data.

Contract No. NAS5-20793.
LANDSAT Investigation No. 21330.

II. Problems: None

III. Accomplishments: In preparation for classifying the Kansas LANDSAT data acquired during the spring and summer of 1975, a pilot study was conducted using frame 2072-16284 acquired April 5, 1975 over south central Kansas. The study included (1) an investigation of alternate methods of selecting training samples, (2) an evaluation of sampling procedures, and (3) definition and test of a method for obtaining county boundary coordinates from the LANDSAT data.

A. Selecting Training Samples

The objective of this portion of the pilot study was to determine the method of choosing training classes for classifying Kansas and Indiana data. The ground truth available was county maps and underflight photography covering a strip through Sumner County, Kansas. The LANDSAT data was from an April 5, 1975 pass, and the photography was from an April 20, 1975 flight. The general classes to be considered were wheat and non-wheat.

(E76-10145) CROP IDENTIFICATION AND ACREAGE ESTIMATION OVER LARGE GEOGRAPHIC AREAS USING LANDSAT MSS DATA Quarterly Progress Report, 1 Oct. - 31 Dec. 1975 (Purdue Univ.) 5 p HC $3.50 CSCL 02C G3/43

N76-17454

Unclas 00145

The first method used was hand-picking the training fields. Using six-class cluster maps of three frames in the photography, representative wheat and non-wheat fields were outlined. For the non-wheat fields, samples of bare soil, pasture, and what was thought to be either fallow or newly planted crops were identified. The next step for hand-picking fields was to obtain the line and column coordinates for them in the LANDSAT data. The wheat classes were divided into four subclasses using the clustering processor. In the same way, the non-wheat fields were separated into five subclasses.

The second method used to obtain training classes was to cluster separately the same three frames as above, requesting ten classes each time. Then, by comparing these cluster maps with the photography, the cluster classes were labeled wheat or non-wheat. The statistics decks from these clustered areas were merged and the separability processor was used to determine which classes should be pooled or deleted.

The entire aircraft flightline was classified four times, using the Guassian and the ECHO classifiers and both sets of training class statistics. Test fields, wheat and non-wheat, were chosen from the remaining frames in the underflight photography. The test field performance was very similar for the two classifiers with the same training class statistics. The test field performance though for the first method, using hand-picked training fields, was about 90% for both wheat and non-wheat while it was about 83% for the second method, using cluster classes. Based on

this study, it was decided to use hand-picked training fields for classifying the other counties.

B. Sampling Procedures

In the LANDSAT project, only a sample of the area of the state of Kansas will be classified as the basis for an estimate of the wheat acreage. This sampling will introduce a variance due to sampling error into the acreage estimate of wheat. The pilot study attempts to measure that error. The areas chosen for the pilot study were Rice County and Morton County, Kansas.

In order to measure only the effect of sampling, the error introduced in classification was ignored by comparing the various samples to a 100% sample. The results reported here are based on Rice County; the analysis on Morton County has not been completed.

A rectangular area of Rice County comprising 255012 LANDSAT resolution elements was classified. 75005 resolution elements were classified as wheat or 29.41% of the area.

If each resolution element is classified as either wheat or non-wheat, the binomial distribution can be applied to estimates of the total number of wheat pixels and of the percentage of wheat in the area. Theoretical estimates of the sampling error are then available [1].

Estimates of both the total number of wheat resolution elements and the percentage of wheat in the area were calculated for 50%, 33.3%, 25%, 11.1% 10%, 6.25%, 4% and 2.8% samples. These samples were taken systematically. For example, a 11.1% sample of the area was taken by tabulating the classification with both a line and

---

[1] Cochran, William G. (1963). Sampling Techniques. John Wiley and Sons, New York, second edition.

column interval of 3. Nine of these 11.1% samples were taken by selecting a different starting point for each sample. The theoretical variance of these sample estimates were calculated from the binomial distribution and compared to the variance among the repeated estimates of the same sample size. For example, the theoretical variance of a 11.1% sample was calculated and then compared to the variance of the 9 sample estimates.

In all cases the two variances were not significantly different, and in all but two cases (50% and 33.3% sampling rates), the calculated variance was less than the theoretical variance. This result indicates that the theoretical estimate of the sampling error based on the binomial distribution can be used as the basis for a decision on the sampling rate to use in the LANDSAT classifications.

C. Location of County Boundary Coordinates

In order to report classification results (acreage estimates) on a county and crop reporting district basis, it is necessary to locate the coordinates of the county boundaries in the LANDSAT data. The method defined for this task is as follows: (1) Digitize the x-y coordinates of county boundaries on a 1:250,000 scale map, (2) digitize the coordinates of the checkpoints such as highway intersections, airport runways, etc. from the map, (3) locate the LANDSAT coordinates of the checkpoints, and (4) apply a transformation to convert map coordinates to LANDSAT coordinates. The output of the procedure is punch cards with LANDSAT coordinates which may be used with the classification programs to tabulate results for given areas.

D. Indiana LANDSAT Data Selection

LANDSAT MSS data for Indiana has been ordered from the EROS data center. The data is from all portions of the growing season beginning May 2 and continuing through September 6. In total 27 frames were ordered including six LANDSAT-1 data sets.

E. Plans for Next Reporting Period

During the next three months, a substantial portion of the Kansas data will be analyzed. A system for extending statistics will be prepared so that all the counties for which data has been received can be classified and tabulated. It is intended that all the Kansas data be classified and tabulated prior to the start of the work on Indiana data.

IV. Significant Results: None

V. Publications: None

VI. Recommendations: None

VII. Funds Expended: November 30, 1975 - $35,897

VIII. Data Use:

	Value of Data Allowed	Value of Data Ordered	Value of Data Received
CCT	$23,800	$12,300	$5,000
Imagery	2,392	---	1,532

IX. Aircraft (NASA) Data: None