

N73 - 28232

Paper A 25

IDENTIFICATION OF AGRICULTURAL CROPS BY COMPUTER PROCESSING OF ERTS MSS DATA

Marvin E. Bauer and Jan E. Cipra, *Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana 47907*

ABSTRACT

Quantitative evaluation of computer-processed ERTS MSS data classifications has shown that major crop species (corn and soybeans) can be accurately identified. The classifications of satellite data over a 2000 square mile area not only covered more than 100 times the area previously covered using aircraft, but also yielded improved results through the use of temporal and spatial data in addition to the spectral information. Furthermore, training sets could be extended over far larger areas than was ever possible with aircraft scanner data. And, preliminary comparisons of acreage estimates from ERTS data and ground-based systems agreed well. The results demonstrate the potential utility of this technology for obtaining crop production information.

1. INTRODUCTION

Crop identification has long been recognized as one of the potential applications of remote sensing. In fact, accurate identification of crop species is a prerequisite to deriving crop production estimates (acres x yield per acre) from remotely sensed multispectral measurements, and the value of this type of information is substantial. It is used in managing production, storage, and transportation and in determining the market price of crops. Additionally, the USDA uses crop production information in designing national farm programs and establishing import/export policies.

Any improvement in the quality of this vital information could have far-reaching benefits. Eisgruber, ("Potential Benefits of Remote Sensing: The Actual Framework and Empirical Estimate", LARS Information Note 03C872) shows that reduction from three to two percent in the error of estimate for corn, soybean, and wheat production estimates would result in 14 million dollars net social benefit to the country. Reduction to one percent would add another nine million dol-

lars. More frequent and timely estimates alone, even without an accompanying improvement in accuracy, would result in additional benefits (Ewart, "Effect of Information on Market Behavior", Ph.D. dissertation, Purdue Univ., 1972).

The wide area coverage of ERTS, combined with the capabilities of computer data processing, offers a unique opportunity to reduce the above-mentioned error of estimate through reduction in sampling error. Furthermore, the sequential coverage capabilities of ERTS may lead to benefits arising from improvements in the frequency and timeliness of estimates.

Because of these potential benefits, the overall objective of this research has been to quantitatively evaluate the utility of the machine analysis of ERTS MSS data in identifying crop species. Future investigations will verify the utility of ERTS MSS data in identifying crops over a wide range of environments with differing soils, climates, and cultural practices and make crop acreage estimates for larger geographic areas.

2. PROCEDURES

The area which was classified on the basis of ERTS data (DeKalb, Lee and Ogle Counties in northern Illinois) contains highly productive, level to gently rolling prairie soils and is intensively cropped. In 1972, about two-thirds of the total farm land was planted to corn and soybeans (Table 1). ERTS data analyzed came from Frame 1017-16093, which was acquired on August 9, 1972.

Ground truth data used to support the analysis consisted of identification of the crop or use of over 500 fields in four different areas of the three counties. Crops identified were corn and soybeans, sorghum and alfalfa, and other cover types or land used identified were hay, pasture, and small grain stubble. However, for the purposes of this analysis, only three classes were considered: corn, soybeans, and "other" (all cover other than corn and soybeans, including towns and highways). These 500 fields were used for training the maximum likelihood classifier and testing the accuracy of classifications.

Most fields could be accurately located in the ERTS data using a computer printout image generated on the basis of statistics from the nonsupervised classifier. Following the locating procedure, a random selection of training fields was made from each crop or cover type (all available fields not used for training were included in the test set). The number

of corn and soybean training fields varied from 3 to 12 in order to allow evaluation of the number of training fields on classification performance. For the "other" classes, two to four fields of each cover type were included in the training set.

Statistics for the selected set of spectral training fields were then computed and evaluated. The "other" class displayed much more variation than either the corn and soybeans which were found to be uniform spectrally (as indicated by their small variances). Additionally, the feature selection program that was run indicated that bands 5 and 6 should give results as good as either a band 5, 6, and 7 combination or all four MSS channels. Classification results subsequently confirmed this.

Classifications of each of the three counties were next carried out utilizing a training set comprised of fields from that county. This led up to a classification of the entire three-county area which was made using a subset of training fields previously used for classifying the individual counties. Several different analyses were also conducted in order to answer basic questions concerning the utility of the machine-processing of ERTS MSS for crop species identification.

3. RESULTS

The DeKalb County area was classified first, using a training set with 12 corn fields, 12 soybean fields and two to three fields for each of the "other" classes. An overall test performance (total test points correctly classified/total test points classified) of 83 percent was achieved for the three classes of corn, soybeans, and "other". Quantitative results from this classification are shown in Table 2 (results in the other two counties were similar.)

A classification map of corn, soybeans, and "other" present in part of the three-county area is shown in Figure 1. While useful in qualitatively evaluating a classification, such maps are not necessary to successfully determine crop acreages. Nor will it be necessary to classify entire areas since estimates with very small sampling error can be obtained from classifying a random selection of sample units.

To test extendability of training sets, the training set from each county was used to classify the other two counties also. Results (Figure 2) show that equally good performance was achieved by using any of the three training sets for classifying the three, which are 15 to 25 miles apart. In further tests, there was little difference in classification

performance for two different training sets selected in the same manner (Table 3).

The next test was designed to determine if a reduction in the number of fields in the training set would bring about a decrease in classification accuracy. This test, carried out with the Dekalb County data, evaluated training sets composed of 12, nine, six, and three fields each of corn and soybeans. Although there was a decrease in the accuracy of corn identification in the reduction from 12 to nine fields, the remainder of the training sets for corn and the three-, six-, and nine-soybean training field-sets performed as well as the 12 (Figure 3).

The extendability, variability, and size limitations of training sets required to achieve accurate classifications have a large impact on the design and requirements of a ground observation system to use with remote sensing.

Preliminary comparisons between estimates from ERTS data and estimates by USDA of the percentage of the area in the three counties covered by corn and soybeans show that the estimates agree well, particularly for corn (Table 4). It should be noted that the estimates from ERTS data have been corrected for classification bias, (non-corn called corn, etc.) using information from test class performance tables such as in Table 1. There has not yet, however, been any adjustment for the highways, which are over-represented and included in the tabulation of "other".

Through use of a sample classifier which includes spatial information in the classification process, recognition accuracy of corn and soybeans went up about five percent (Table 5). This classifier requires input of field coordinates which would make it difficult to use for making acreage estimates.

One of the important advantages of computer processing of multispectral scanner data is that data from two or more dates can be included in the same analysis. This is accomplished by spatially registering (overlying) the data. In many instances the addition of temporal information in this manner can be expected to improve classification performance. And, the capability to register data for successive dates also means that field coordinates need be manually located. Temporal analyses with August, September, and October data have been conducted. Recognition of "other" was markedly improved, while corn and soybean recognition was unchanged. These preliminary results, however, do not represent the full potential of temporal information. Data from earlier

in the growing season should give greater improvement in crop recognition.

4. SUMMARY AND CONCLUSIONS

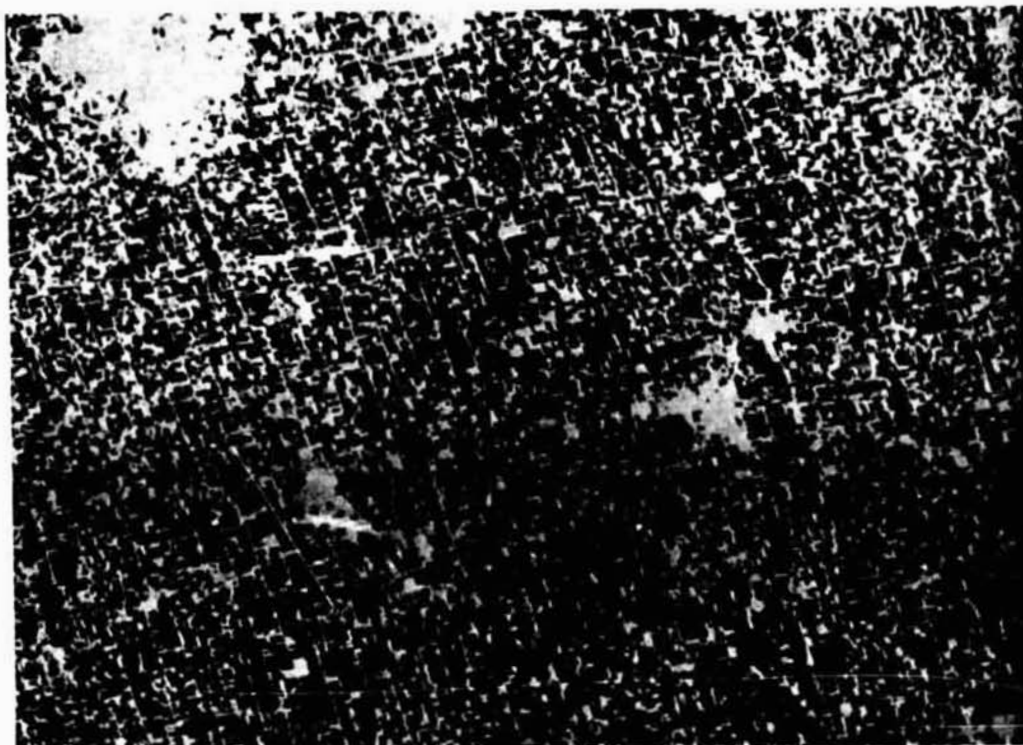
Computer analysis of ERTS MSS data was shown to be an effective method for identifying agricultural crops over large areas. When fully developed, this capability should lead to improvements in the precision and timeliness of crop production estimates.

5. ACKNOWLEDGEMENTS

The contribution of the LARS staff to this work is gratefully acknowledged. The research was sponsored under NASA Contract NAS5-21773 and NASA Grant #NGL 15-005-112.

Table 1. Crop acreages and percent agricultural land in three-county test site.

| Area | Total Area (000 Acres) | 1971 Corn Acreage (000) | 1971 Soybean Acreage (000) | Percent Farmland in 1971 |
|------------|---------------------------|-------------------------------|----------------------------------|--------------------------------|
| Dekalb Co. | 415 | 194 | 80 | 90 |
| Ogle Co. | 487 | 206 | 47 | 93 |
| Lee Co. | <u>437</u> | <u>185</u> | <u>90</u> | <u>97</u> |
| Total | 1339 | 585 | 217 | 93 |



black: Corn gray: Soybeans white: "Other"

Figure 1. Computer classification map of corn, soybeans, and "other" for Dekalb, Ogle, and Lee Counties, Illinois.

Table 2. Classification of corn, soybean, and "other" test fields by computer analysis of MSS data.

| Crop | No. Points | No. Points Classified As | | | Percent Correctly Classified |
|----------------------|------------|--------------------------|-----------|------------|------------------------------|
| | | Corn | Soybeans | "Other" | |
| Corn | 3968 | 3367 | 357 | 244 | 85 |
| Soybeans | 1113 | 115 | 855 | 133 | 77 |
| "Other" | <u>295</u> | <u>16</u> | <u>50</u> | <u>234</u> | 79 |
| Total | 5376 | 3498 | 1262 | 611 | |
| Overall Performance: | | 83 percent | | | |

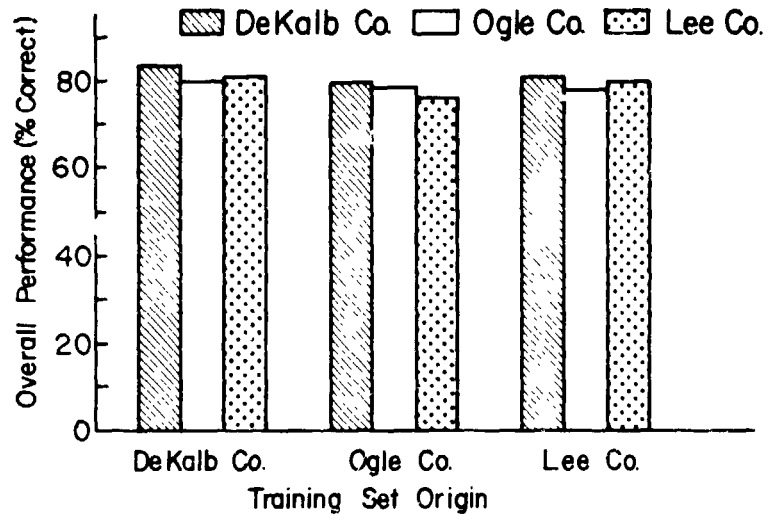


Figure 2. Test of the extendability of training sets.

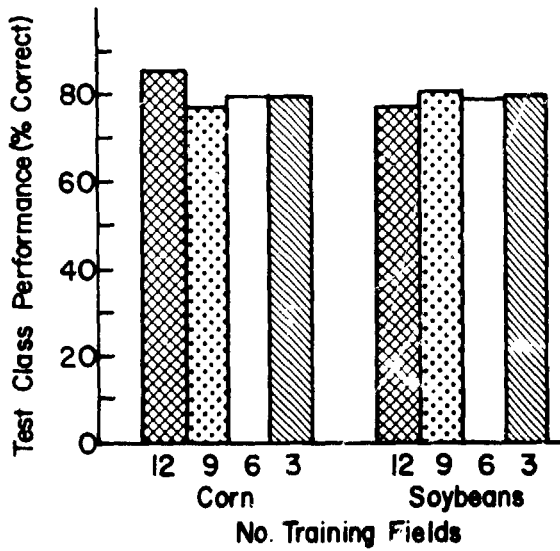


Figure 3. Influence of training set size on classification performance.

Table 3. Comparison of test field classification accuracy for two random selections of training fields

| Area Classified | Overall Classification Performance (% Correct) | |
|-----------------|--|----------------|
| | Training Set 1 | Training Set 2 |
| Dekalb Co. | 83 | 81 |
| Ogle Co. | 79 | 78 |
| Lee Co. | 81 | 75 |

Table 4. Comparison of USDA acreage estimates with estimates derived from computer analysis of ERTS data (Dekalb, Ogle, and Lee Counties, Illinois)

| | USDA | ERTS |
|----------|-------------------------|------|
| | (Percent of Total Area) | |
| Corn | 40 | 40 |
| Soybeans | 17 | 23 |
| "Other" | 43 | 37 |

Table 5. Use of spatial information (sample classifier) in addition to only spectral information (point classifier)

| | Point Classifier | Sample Classifier |
|----------|------------------|-------------------|
| | % Correct | |
| Corn | 85 | 90 |
| Soybeans | 77 | 81 |
| "Other" | 79 | 71 |
| Overall | 83 | 84 |