N76-23668

JSC 09393

CROP IDENTIFICATION TECHNOLOGY ASSESSMENT FOR REMOTE SENSING (CITARS)

VOLUME X

INTERPRETATION OF RESULTS

National Aeronautics and Space Administration

LYNDON B. JOHNSON SPACE CENTER

Houston, Texas

December 1975

TECHNICAL REPORT INDEX/ABSTRACT

(See instructions on reverse side.)

1. TITLE AND SUBTITLE OF DOCUMENT CROP IDENTIFICATION TECHNOLOGY ASSESSMENT FOR REMOTE SENSING (CITARS), VOLUME X, INTERPRETATION OF RESULTS	2. JSC NO. JSC- 09393
3. CONTRACTOR/ORGANIZATION NAME Lockheed Electronics Company, Inc.	4. CONTRACT OR GRANT NO. NAS 9-12200
5. CONTRACTOR/ORIGINATOR DOCUMENT NO. LEC-4326H	6. PUBLICATION DATE (THIS ISSUE) December 1975
7. SECURITY CLASSIFICATION Unclassified	8. OPR (OFFICE OF PRIMARY RESPONSIBILITY) R. M. Bizzell, A. H. Feiveson, and F. G. Hall, NASA TF7
9. LIMITATIONS GOVERNMENT HAS UNLIMITED RIGHTS ☐ YES ☒ NO IF NO, STATE LIMITATIONS AND AUTHORITY	10. AUTHOR(S) R. M. Bizzell, A. H. Feiveson, and F. G. Hall, NASA; M. E. Bauer and B. J. Davis, Laboratory for Applications of Remote Sensing, Purdue University; and W. A. Malila and D. P. Rice, Environmental Research Institute of Michigan
11. DOCUMENT CONTRACT REFERENCES	12. HARDWARE CONFIGURATION
WORK BREAKDOWN STRUCTURE NO. Job Order 71-645	SYSTEM ERIPS, LARSYS
CONTRACT EXHIBIT NO.	SUBSYSTEM
DRL NO. AND REVISION	MAJOR EQUIPMENT GROUP
DRL LINE ITEM NO.	Univac 1108/1110

13. ABSTRACT

The CITARS was an experiment designed to quantitatively evaluate crop identification performance for corn and soybeans in various environments using a well-defined set of automatic data processing (ADP) techniques. These techniques differed mainly by the procedure used to obtain signatures from training data (e.g., clustering) and by the method of classification employed (e.g., linear or quadratic decision boundaries and equal or unequal class weights).

Each technique was applied to data acquired by the first Earth Resources Technology Satellite over six Indiana and Illinois test sites throughout the growing season in an attempt to recognize and estimate proportions of corn and soybeans. Both local and nonlocal (i.e., extended) recognition training statistics were used.

These analyses evaluated the significance of the various factors which contributed to the classification performance. This final volume of CITARS documentation summarizes, interprets, and discusses the crop identification performances obtained using (1) different ADP procedures; (2) a linear versus a quadratic classifier; (3) prior probability information derived from historic data; (4) local versus nonlocal recognition training statistics and the associated use of preprocessing; (5) multitemporal data; (6) classification bias and mixed pixels in proportion estimation; and (7) data with different site characteristics, including crop, soil, atmospheric effects, and stages of crop maturity.

14. SUBJECT TERMS

Classification algorithm	Multitemporal data	Statistical evaluation
Crop identification performance	Quantification	
Multispectral data	Radiometric preprocessing	

GLOSSARY

ADP — automatic data processing

ASCS — Agricultural Stabilization and Conservation Service of the U.S. Department of Agriculture

CCT — computer-compatible tape

CIP — crop identification performance, the quantitative assessment of crop inventories in specified areas using remote sensing, photointerpretation, and ADP techniques

CITARS — Crop Identification Technology Assessment for Remote Sensing

EOD — Earth Observations Division of the Lyndon B. Johnson Space Center, NASA

ERIM — Environmental Research Institute of Michigan

ERIPS — Earth Resources Interactive Processing System

ERTS-1 — the first Earth Resources Technology Satellite, which was launched in June 1972, orbits the Earth 14 times a day from an altitude of 915 kilometers and scans the same scene every 18 days (renamed Landsat-1 in January 1975)

IR — infrared

ISOCLS — Iterative Self-Organizing Clustering System, a computer program developed by the EOD, which uses a clustering algorithm to group homogeneous spectral data

JSC – Lyndon B. Johnson Space Center, NASA

Landsat-1 – the first Land Satellite launched in June 1972 (formerly called ERTS-1 and renamed in January 1975)

LARS – Laboratory for Applications of Remote Sensing of Purdue University

LARSYS – a system of classification programs developed at the LARS

MLA – mean level adjustment

MSS – multispectral scanner

MSP – multitemporal processing

NASA – National Aeronautics and Space Administration

Pixel – a picture element which refers to one instantaneous field of view as recorded by the ERTS-1 MSS and covers the equivalent of 0.44 hectare (1.09 acres)

PSP – preprocessing and standard processing

rms – root mean square

SP – standard processing

USDA – U.S. Department of Agriculture

CONTENTS

Section		Page
1.0	INTRODUCTION	1
2.0	OBJECTIVES	3
3.0	EXPERIMENTAL PROCEDURES	5
	3.1 TEST SITE SELECTION	5
	3.2 SELECTION OF SEGMENTS AND SECTIONS	5
	3.3 FIELD OBSERVATIONS FOR CROP IDENTIFICATION	6
	3.4 PHOTOINTERPRETATION FOR CROP IDENTIFICATION	6
	3.5 ERTS-1 DATA PREPARATION	7
	3.5.1 Data Quality Evaluation	7
	3.5.2 Geometric Correction and Registration	8
	3.5.3 Determination of Section and Field Coordinates	9
	3.6 ADP TECHNIQUES FOR MSS DATA PROCESSING	10
	3.6.1 LARS ADP Techniques	11
	3.6.2 ERIM ADP Techniques	12
	3.6.3 EOD ADP Techniques	14
	3.7 STATISTICAL ANALYSIS OF CROP IDENTIFICATION PERFORMANCE	15
	3.7.1 Analyses of Variance	16
	3.7.2 Nonparametric Tests	20

Section		Page
4.0	RESULTS AND DISCUSSION.	27
	4.1 COMPARISON OF ADP PROCEDURES FOR LOCAL RECOGNITION.	28
	4.1.1 Comparison of Standard Procedures.	28
	4.1.2 Use of Prior Probability Information	31
	4.1.3 Linear-Versus-Quadratic Decision Rule	33
	4.1.4 Multitemporal Analysis.	37
	4.2 NONLOCAL RECOGNITION	39
	4.2.1 Comparison of Local and Non-local Results Using Non-preprocessing ADP procedures. .	39
	4.2.2 Comparison of Results Obtained With and Without Preprocessing	40
	4.2.3 Interpretation of Nonlocal Recognition Results	42
	4.3 SEGMENT EFFECTS ON CROP IDENTIFICATION PERFORMANCE	45
	4.3.1 Site Characteristics.	46
	4.3.2 Atmospheric Effects	49
	4.4 EFFECT OF CROP MATURITY ON CROP IDENTIFICATION PERFORMANCE	50
	4.5 EFFECT OF DATA PREPARATION ON CROP IDENTIFICATION PERFORMANCE	52
	4.5.1 Effects of Multitemporal Registration.	52
	4.5.2 Effects of Training Set Selection	54

Section		Page
	4.6 CLASSIFICATION ACCURACY AND PROPORTION ESTIMATION.	56
	4.6.1 Comparison of Field Center Versus Whole Area	56
	4.6.2 Bias in Proportion Estimation.	58
5.0	SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS	81
	5.1 KEY TECHNICAL ACCOMPLISHMENTS.	82
	5.2 RESULTS AND DISCUSSION	84
	5.2.1 ADP Procedures.	85
	5.2.2 Nonlocal Recognition and Preprocessing	88
	5.2.3 Multitemporal Data Analysis . .	89
	5.2.4 Effect of Site and Crop Characteristics	89
	5.2.5 Relation of Crop and Sensor Characteristics	90
	5.3 CONCLUSIONS.	91
	5.4 RECOMMENDATIONS.	92
6.0	REFERENCES.	95

TABLES

Table		Page
I	ERTS-1 COVERAGE SCHEDULE FOR TEST SEGMENTS. . .	22
II	COMPARISON OF CROP IDENTIFICATIONS MADE BY ASCS AND BY PHOTOINTERPRETATION	23
III	ERTS-1 MSS ACQUISITION RESULTS.	24
IV	MEAN CLASSIFICATION ACCURACY FOR THREE STANDARD PROCEDURES AVERAGED OVER 15 DATA SETS . . .	61
V	OVERALL BIAS AND RMS ERROR IN PROPORTION ESTIMATION FOR THREE STANDARD PROCEDURES AVERAGED OVER 15 DATA SETS	61
VI	AVERAGE CONDITIONAL CLASSIFICATION ACCURACY AND RANKINGS OF THREE STANDARD PROCEDURES . . .	62
VII	RMS ERRORS IN PROPORTION ESTIMATION AND RANKINGS OF THREE STANDARD PROCEDURES	63
VIII	SUMMARY OF PLOT AND PIXEL COUNTS.	64
IX	COMPARISON OF AVERAGE CONDITIONAL CLASSIFICATION ACCURACY AND RMS ERRORS IN PROPORTION ESTIMATES FOR SINGLE "BEST" DATES AND MULTITEMPORAL	65
X	COMPARISON OF LOCAL AND NONLOCAL RECOGNITION PERFORMANCES.	66
XI	AVERAGE CONDITIONAL CLASSIFICATION ACCURACY AS A FUNCTION OF SEGMENT AND DATE.	67
XII	RMS ERRORS FOR PROPORTION ESTIMATES AS A FUNCTION OF SEGMENT AND PERIOD.	68
XIII	OPTICAL DEPTH MEASUREMENTS MADE AT THE TIME OF ERTS-1 OVERPASSES	69
XIV	COMPARISON OF DIFFERENT SAMPLES OF TRAINING AND TEST FIELDS	70

Table		Page
XV	COMPARISON OF FIELD-CENTER CLASSIFICATION ACCURACIES AND WHOLE-AREA PROPORTION ESTIMATES. . .	71
XVI	MEAN VALUES OF V FOR CORRECTED WHOLE-AREA PROPORTION ESTIMATES.	72
XVII	MEAN VALUES OF V FOR CORRECTED WHOLE-AREA PROPORTION ESTIMATES — FIELD-CENTER PILOT DATA.	72
XVIII	COMPARISON OF RMS ERROR FOR UNCORRECTED AND CORRECTED OVERALL-SEGMENT-PROPORTION ESTIMATES	73

FIGURES

Figure		Page
1	Technology assessment data set, May through September 1973	25
2	Effect of number of signatures used on full-section proportion estimation by ERIM/SP1 on CITARS data.	74
3	Illustration of a situation where a linear decision rule would outperform a quadratic decision rule on test data	75
4	Comparison of local and nonlocal recognition performance on field-center pixels	76
5	Correlation of field size and crop identification performance	77
6	Graph showing relation of optical depth to local classification accuracy for main procedures	78
7	Comparison of nonlocal classification accuracy as a function of the difference in optical depth between training and classification segments	79

1.0 INTRODUCTION

In 1973, the Crop Identification Technology Assessment for Remote Sensing (CITARS) was undertaken by the Earth Observations Division (EOD) of the Lyndon B. Johnson Space Center (JSC) of the National Aeronautics and Space Administration (NASA), the Environmental Research Institute of Michigan (ERIM), the Laboratory for Applications of Remote Sensing of Purdue University (LARS), and the Agricultural Stabilization and Conservation Service (ASCS) of the U.S. Department of Agriculture (USDA). The primary goal of this joint task was to quantify crop identification performances (CIP's) resulting from the identification of corn, soybeans, and wheat by remote sensing, using automatic data processing (ADP) techniques developed at ERIM, LARS, and EOD. The ADP techniques were automatic in the sense that subjective human interactions with the classification algorithms were minimized by specifying the steps required for an analyst to convert a multispectral data tape to a classification result.

The remotely sensed data were acquired by multispectral scanner (MSS) onboard NASA's Earth Resources Technology Satellite (ERTS-1)[1] and high-altitude aircraft. Six 8- by 32-kilometer segments in Illinois and Indiana were selected for data gathering during six periods from early June through early September 1973. Concomitantly with the spacecraft and aircraft data, ground truth was acquired by a combination of ASCS field visits and interpretation of large-scale aerial photographs. The major crops of corn,

---

[1]The ERTS-1 was redesignated the first Land Satellite (Landsat-1) in January 1975 and now bears the name Landsat-1.

soybeans, and wheat were classified for the six periods during the growing season for both of the following conditions:

1. Local recognition: Crop signatures for classifier training were obtained from the geographic region in which the crops were identified.
2. Nonlocal recognition: Crop signatures for classifier training were obtained from a geographic region other than the one in which the crops were identified.

An additional category "other" was established for the classification of ground features other than the three major crops.

The classification results from the MSS data and ADP techniques were compared to the ground-truth data to establish the CIP. The CIP's resulting from several basic types of ADP techniques were then compared and examined for significant differences. Once the CIP was established for each of the ADP techniques for local and nonlocal recognition, differences in the performances of these techniques were examined as functions of geographic location, time of the year, and other pertinent factors.

This concluding volume of the final report presents the major results of CITARS and discusses their significance. The first results from the experiment were described in an earlier paper (ref. 1); and the final report, which includes all volumes of the CITARS (ref. 2), contains complete descriptions of the various aspects of the experiment and their results.

2.0 OBJECTIVES

The overall objective of CITARS was to quantify the CIP resulting from the remote identification of corn, soybeans, and wheat using ADP techniques developed at EOD, ERIM, and LARS. The ADP techniques were evaluated on this data set for local and nonlocal recognition. Specific objectives included performance comparisons to answer the following questions.

1. How do corn, soybeans, and wheat CIP's vary with time during the growing season?
2. How does the CIP vary among different geographic locations having different soils, weather, management practices, crop distributions, and field sizes?
3. Can statistics acquired from one time or location be used to identify crops at other locations and/or times?
4. How much variation in CIP is observed among different data analysis techniques?
5. Does use of radiometric preprocessing extend the use of training statistics and/or increase the CIP?
6. Does use of multitemporal data increase the CIP?

3.0 EXPERIMENTAL PROCEDURES

3.1 TEST SITE SELECTION

The test sites were chosen over a large geographic area in order to include a wide variety of conditions. It was recognized that much variation in soils, weather, agricultural practices, and crop distribution occurs in the Corn Belt and that all of these factors are related to its geographic location. The goal, then, was to obtain the best measure of the effects of these factors by including as many test sites as possible over as large an area as possible.

To increase the probability of obtaining cloud-free data for each ERTS-1 cycle, six test sites were selected within the four overlap zones of five passes over Indiana and Illinois. The areas shown in figure 1 included many of the different conditions which could be expected to be encountered in the Corn Belt. (See section 4.3.1.)

3.2 SELECTION OF SEGMENTS AND SECTIONS

The segments, 8 by 32 kilometers in size, were chosen at random within each of the six selected counties. This segment size provided a limited area for field visits and yet an adequately large area for a representative sample of agriculture within a county. Within each segment, 20 sections and 20 quarter sections were chosen at random in a manner such that the selected quarter sections were spatially separated from the selected sections (ref. 2, vol. I, pp. 11-13).

3.3 FIELD OBSERVATIONS FOR CROP IDENTIFICATION

From May to September 1973, every 18 days and coincident with ERTS-1 passes (table I), ASCS personnel visited the 20 quarter sections in each segment and recorded the crop type and other descriptive information for each field. Atmospheric optical depth measurements and subjective assessments of cloud cover and weather were also recorded during the ERTS-1 overpasses.

3.4 PHOTOINTERPRETATION FOR CROP IDENTIFICATION

To obtain a more accurate estimate of the CIP, the field observation data from the 20 quarter sections were supplemented by photointerpretation of the 20 sections chosen in each segment.

The photointerpretation effort used large-scale color infrared (IR) aerial photography acquired three to five times and large-scale metric photography acquired two times during the growing season. In this manner, proportions of ground-cover classes and other agricultural parameters were established within each of the 20 sections in each segment (ref. 2, vol. IV). The crop type and other information collected by the ASCS from 16 of the 20 quarter sections were used by the three interpreters for training. The information from the other 4 quarter sections was concealed from the interpreters while they interpreted the 20 sections and the 4 sections which contained the concealed quarter sections. At the conclusion of the photointerpretation subtask, the crop identifications and area measurements for the four concealed quarter sections were compared to the data collected by the ASCS. A summary of these results is shown in table II.

While the photointerpretation accuracy was not 100 percent, it was considered sufficiently high to allow evaluation of ADP classifications. The accuracy for acres was higher than for individual fields, indicating that interpretation accuracy was lower for small fields. Since pixels from small fields were not included, this was not a problem in evaluating field-center classification results. The effect of evaluating small fields, however, would have been included in the proportion estimates for entire sections. Yet, errors in the crop types used to evaluate the classification results probably accounted for 5 percent or less of any misclassification.

3.5 ERTS-1 DATA PREPARATION

The ERTS-1 data preparation for CITARS consisted of (1) data quality evaluation, (2) geometric correction and registration, and (3) section and field coordinate location.

3.5.1 Data Quality Evaluation

The ERTS-1 data quality evaluation included examination for cloud cover and electronic data quality checks (ref. 2, vol. III). The satellite passed over each of the six test segments twice (on successive days) during each 18-day period. Since seven periods were of interest, from early June to late September 1973, a total of 84 data sets were available for potential processing and analysis. Cloud cover problems were identified by reference to the ERTS-1 data catalog and by visual inspection of available imagery. Of the 84 possibilities: cloud cover on 53 sets was severe enough to cause their outright rejection; no data were collected for two sets; and several others were eliminated for

other reasons. A total of 26 sets were selected for analysis, and, upon detailed examination, several of these were found to have cloud-cover problems. Thus, roughly 70 percent of the data sets were eliminated because of excessive cloud cover. (See table III.)

A majority of the selected data were of good quality. A few problems which affected data analysis procedures and/or results were observed.

1. Occasional erratic data were present throughout individual scan lines or portions of lines.
2. Differences existed among the mean values obtained from the six detector channels that comprised each spectral band as averaged over a large sample of the data.
3. Differences were observed in the variances from the detector channels over the same data sample.

3.5.2 Geometric Correction and Registration

The digital form of the ERTS-1 data [computer-compatible tape (CCT)] contains several geometric distortions, including scale differential, altitude and attitude variations, Earth rotation skew, orbit velocity change, scan-time skew, nonlinear scan sweep, scan-angle error, and frame rotation. The scale and skew errors were the most significant, with rotation to north-orientation highly desirable. A two-step process was developed by LARS to geometrically correct ERTS-1 data over small areas. It was applied to all data for CITARS (ref. 3). The output form used for CITARS is such that when the data are printed on an 8-line-per-inch, 10-column-per-inch computer line printer the resulting scale is approximately

1:24,000 and the image is north-oriented (ref. 2, vol. V). Comparisons made using topographic maps indicated a scale error of approximately 1 to 2 percent. Having geometrically correct and scaled data greatly facilitated the task of locating section and field coordinates.

Registration of multiple images of the same scene was accomplished through use of the LARS image registration system (ref. 4). A measure of registration error was obtained from the checkpoint residuals of the least squares polynomials used in the image correlation. This statistic averaged less than 0.5 of an image sample (pixel), and in practice additional checkpoints were located whenever the root-mean-square (rms) error for either lines or columns exceeded 0.5.

3.5.3 Determination of Section and Field Coordinates

Determining section and field coordinates was a major preparatory task to classifying the ERTS-1 data. First, a manual method for locating fields displayed the ERTS-1 data as single-band, gray-scale, line printer maps (ref. 2, vol. V). After manually locating all fields and sections in the ERTS-1 data, the precision was determined to be inadequate to meet the maximum error requirement of one pixel. Therefore, a previously developed, computer-assisted method of transforming map coordinates to ERTS-1 data was employed by ERIM to locate section corners and define coordinates for sections (ref. 5). Final standard errors of estimate for control points were less than 0.5 and typically between 0.2 and 0.4 of an ERTS-1 pixel; that is, 15 to 30 meters (49.5 to 99.0 feet) on the ground.

3.6 ADP TECHNIQUES FOR MSS DATA PROCESSING

The basic ADP techniques were grouped into three divisions: (1) standard processing (SP) techniques, (2) preprocessing (PSP) techniques, and (3) multitemporal processing (MSP) techniques. The term "standard" refers to an ADP technique for classifying single-pass data which have not been radiometrically preprocessed.

Each of these ADP techniques consists of a computer-implemented software system and a method or procedure by which an analyst can convert multispectral data into ground-cover, class-identification information on a pixel-by-pixel basis.

The CIP of ADP techniques is sensitive to the manner in which the classifier is trained, the types of MSS data input (such as preprocessed and multitemporal), and the spectral bands used for recognition. At the beginning of CITARS, most of the procedures used generalized analysis algorithms and required decisions on the part of the analyst which could significantly affect the CIP. To permit quantitative evaluation and meaningful comparison of techniques, subjective processes had to be held to a minimum; therefore, only well-defined and repeatable procedures were followed for CITARS. Each of the ADP techniques was documented in detail, and the documented procedures were followed rigidly (ref. 2, vol. I, pp. 33-39).

3.6.1 LARS ADP Techniques

The analysis techniques used by LARS utilized the LARSYS 3 multispectral data analysis system. Its theoretical basis and details of the algorithm implementation are described by Swain (ref. 6) and Phillips (ref. 7). The analysis procedure was described in detail by Davis and Swain (ref. 8) and NASA/JSC (ref. 2, vols. I and VI). The procedures were designed to provide repeatable results, inasmuch as variation caused by analysts is minimized. The analysis procedures are described briefly in the following subsections.

3.6.1.1 Class definition and refinement.- Four major classes – corn, soybeans, wheat (for selected missions), and all other ground covers – were defined. These major classes were divided into subclasses where spectral variability within a class was so great as to result in multimodal probability distributions for that class. Subclasses were isolated by clustering quarter-section field centers. All four ERTS-1 bands were used for clustering. A systematic method which minimized the total number of subclasses and avoided multimodal subclass distributions was used for interpreting information on the separability of subclasses (ref. 8).

3.6.1.2 Classification.- Each data set was analyzed using two versions of the maximum likelihood classification algorithm. Gaussian probability density functions were assumed for both procedures. The first classification method, LARS/SP1, was the maximum likelihood classification rule assuming equal prior probabilities for all classes and subclasses. This rule has been in common usage for remote-sensing data analysis for some time.

The second method, LARS/SP2, used class weights proportional to the class prior probabilities. This approach is more nearly optimal in that the Bayesian error criterion (minimum expected error) is preferred. Class weights may be based on any reasonably reliable source of information. In CITARS the class weights were computed from county acreage estimates made by the USDA the previous year. Class weights were divided among the subclasses in proportion to the number of points in each subclass as determined by the clustering procedure.

3.6.1.3 Display and tabulation of results.- The results of the classification were displayed using a discriminant threshold of 0.1 percent. This low threshold eliminated only the data points which were very much different from the major class characterizations. Thresholded points were counted in the category "other." A computer program generated results tabulations for training fields, test fields, and test sections in both printed form and on punched cards.

3.6.2 ERIM ADP Techniques

The digital data processing and analysis procedures defined by ERIM for use in the CITARS study reflected concern for the calculational efficiency of recognition processors, the need for extending recognition signatures from training areas to other geographic locations and/or observation conditions, and the CITARS requirement for minimizing the need for analyst judgment (ref. 2, vol. VII). The following subsections give a brief summary of the procedures.

3.6.2.1 Training.- The training of the processor (that is, the establishment of class signatures for recognition) was a crucial step in MSS data processing. Although multimodal signatures frequently have been employed, the use of one signature per major class was selected for CITARS processing because of simplicity, processing efficiency, and the fact that a combination of individual field signatures could result in a single signature that encompasses more of the variability of the class than is represented by a multimodal signature. An objective, reproducible procedure based on a chi-squared test was devised to reject anomalous "outlier" fields before the formation of a combined signature, in order to develop signatures representative of healthy crops at a reasonable stage of maturity for the time of season. Signatures for classes other than the major ones were included only if they were found to be confused with the major crops on preliminary recognition runs over training data.

3.6.2.2 Recognition without preprocessing.- Two types of decision algorithms were used: a linear rule, ERIM/SP1; and a more conventional quadratic (Gaussian maximum likelihood) rule, ERIM/SP2. The linear decision rule was chosen because: it requires substantially less computer time for recognition calculations; it has been used successfully in many applications at ERIM; and it has been found to provide comparable recognition accuracy in previous tests (ref. 9). Use of the quadratic rule permitted another comprehensive comparison of the two rules. Both rules apply a threshold test (0.001 probability of rejection) based on a quadratic calculation for the signature of the prevailing class; points failing the test would be classified as being other than the major crops considered.

3.6.2.3 Recognition with signature extension preprocessing.- It was recognized that changes in atmospheric and other local conditions could cause changes in the signal levels received by the MSS for different areas and at different times. The region of signature applicability could be extended beyond the region used for training by employing signature-extension preprocessing techniques (ref. 10). Nonlocal recognition denotes recognition performed on segments other than those from which signatures were extracted. Nonlocal recognition was carried out once before and once after preprocessing corrections for signature extension had been applied for both linear (ERIM/PSP1) and quadratic decision (ERIM/PSP2) rules. The preprocessing method used on the CITARS project was a mean-level-adjustment (MLA) procedure derived from an average over diverse ground covers within the local segment for signature extraction and a comparable average within the nonlocal segment to be classified.

3.6.2.4 Summarizing results.- The results obtained with each procedure were summarized in a standardized form for subsequent analyses of variance. Separate summaries were made for field-center pixels and for entire sections.

3.6.3 EOD ADP Techniques

The EOD evaluated two techniques: one for single-pass data, EOD/SP1 (ref. 2, vol. VIII, parts 1-6), and another for multitemporal MSS data, EOD/MSP1 (ref. 2, vol. VIII, part 7).

For single-pass data the EOD utilized the Iterative Self-Organizing Clustering System (ISOCLS, ref. 11) implemented at JSC and a Gaussian maximum likelihood classifier

to generate the class and subclass statistics. The training fields for corn, soybeans, and wheat were submitted to independent runs using the ISOCLS routine to generate class and, if necessary, subclass statistics (for example, corn 1, corn 2, corn 3). The training fields for class "other" were then submitted to the same clustering scheme to generate class and subclass statistics for all other ground cover. The training fields, test fields, and test sections were classified with the Gaussian maximum likelihood classification algorithm using the statistics previously generated from the clustering process.

Multitemporal data were constructed by combining the data from two or more ERTS-1 passes over a site to form additional features; for example, with ERTS-1 passes in periods I and II, the four features from pass 1 were combined with the four from pass 2 to produce an eight-channel observation vector. Once the multitemporal observations were formed, 11 fields were deleted because of clouds, and processing was performed using the standard EOD procedure (EOD/SP1).

3.7 STATISTICAL ANALYSIS OF CROP IDENTIFICATION PERFORMANCE

The basic questions proposed in the objectives were answered by a series of analyses of variance and blocked-rank tests. The CIP of the ADP techniques was characterized in two different ways: field centers (commonly called test fields) and whole areas (entire sections). The analyses for field centers considered the elements of the performance matrix, e_{ij} , the estimated probability of classifying a

nonboundary (field-center) pixel from class j as class i. For whole areas, the analyses examined the differences between the estimated proportion of class i and its true proportion. The analyses were performed for both local and nonlocal recognition data sets. In this section the analyses of variance are discussed in detail, followed by a brief discussion of the nonparametric or blocked-rank test.

3.7.1 Analyses of Variance

The analyses of variance fall into two main categories: overall analyses and specific, or section-by-section, analyses. The analyses are further divided into analyses concerning: local recognition of corn, soybeans, and "other"; nonlocal recognition of corn, soybeans, and "other"; and multitemporal recognition.

Overall analyses of variance were run for the purpose of comparing procedures over all the data sets for local and nonlocal field centers and whole areas. The experimental unit was a combination of each data set and procedure; that is, results were aggregated over all sections within a data set (ref. 2, vol. IX, appendix A).

In order to compare procedures for specific counties or times, or to compare counties, times, types of nonlocal recognition, and so forth, the size of the experimental unit had to be reduced; thus, a section was chosen as the basic unit. Appropriate interactions between sections and other factors were then used as estimates of error in the analysis-of-variance F-tests (ref. 2, vol. I, p. 42).

In each analysis of variance, as many sections as possible were used. Sometimes sections were removed for one or more of the following reasons.

1. Cloud cover or bad data lines prevented accurate proportion estimation.
2. The ADP processing results were not available.
3. Photointerpreted proportions were not reliable.
4. Maintenance of a balanced design was desirable.

The sections used for a given segment were consistent within an analysis but were not necessarily the same for all analyses.

To evaluate the classification accuracy on the field-center data, the estimated performance matrix was computed for each section in a segment (specific analyses) and for all sections of a segment together (overall analyses). The average of the diagonal elements of the matrix is the average conditional class accuracy.

To apply the analysis of variance in comparing classification accuracy, a single measure of classification performance is needed. One measure of error is the sum of the off-diagonal elements of the performance matrix; that is, the total errors of both commission and omission. One of the major assumptions of analysis of variance is that the variance of the dependent variable in a particular treatment combination is independent of the mean of that combination. Since the elements of the estimated performance matrix can be

considered to be binomially distributed, it can be shown (ref. 12) that the transformation matrix,

$$h_{ij} = \frac{2}{\pi} \text{ arc sin} \sqrt{e_{ij}} \tag{1}$$

stabilizes the variance of the h_{ij}; hence, the sum of the off-diagonal elements of the transformed performance matrix is more suitable for analysis of variance than the corresponding sum of the e_{ij}. The transformation is monotonic so that low/high values of h_{ij} correspond to low/high values of e_{ij}. Furthermore, the use of h_{ij} tends to prevent extreme results on a few sections from dominating a treatment mean. The dependent variable used in the analyses of variance which compare classification accuracy was the sum of the off-diagonal elements of the transformed performance matrix. An average interclass error of 10 percent in the three-class case is an average conditional class accuracy of 80 percent.

In the case of whole areas, the proportion estimation accuracy q was measured by examining differences between the photointerpreted (true) and computer-estimated proportions. This simple difference or bias describes performance for individual crops, whereas the rms error is indicative of overall performance.

$$q^2 = \sum_{i=1}^{K} \frac{\left(\hat{P}_i - P_i\right)^2}{K} \tag{2}$$

For the measure of area estimation accuracy given in equation (2), K is the number of classes, $\hat{P}_i$ is the estimated proportion of crop i, and P_i is the photointerpreted proportion of crop i. These measures were calculated for all analyzed sections of a segment and also for their aggregate.

Along with the true proportions, the bias of the proportion estimate obtained by counting pixels classified as a particular crop must be considered, where the bias depends on the matrix of conditional probabilities of classifying a pixel as one crop when it is of another crop (or mixture of crops) and, as well, on the true proportions present. For this reason, the rms error might be questioned as a reliable measure of accuracy for a procedure, inasmuch as the true proportions and the matrix of conditional probabilities for a particular procedure could be such that the bias is very large or, conversely, almost zero, thus making the procedure appear very inaccurate or very accurate.

It is true, however, that the bias tends to decrease as the accuracy of the classifier increases. Also, on a section-by-section basis, the true proportions vary considerably; if a procedure does well on most or all sections in a segment, one cannot attribute the result to classification errors canceling one another. Instead, it must be concluded that the procedure is in fact accurate.

For this reason, computing the mean square errors on a section-by-section basis and averaging them over a data set should be a reliable indicator of performance.

In the actual specific analyses of variance, it became necessary to transform the mean square error in each section because the variance of the q^2 values were approximately proportionate to their mean. To reduce the effect of this relationship, the following transformation was chosen (ref. 13).

$$y = \log(100Kq^2 + 0.2) \tag{3}$$

The lowest possible value of y is -1.609, representing complete agreement between the computer-estimated and the photointerpreted proportions.

Within three classes, a y-value of 1.0 corresponded to an absolute error of about 0.09 in each class; a y-value of 3.0 represented very poor estimation – an error of about 0.25 in each class.

3.7.2 Nonparametric Tests

The relative ranks of the procedures for each data set were used to test for an overall significant difference between procedures. To do this, a form of blocked-rank test (ref. 14) was utilized.

In this test, the null hypothesis H_0 is that for each data set the ranks are randomly assigned. The test is

performed by computing the $(m - 1)$-by-1 vector $\bar{R}$ which contains the average rank for each procedure[1] and then calculating

$$q = (\bar{R} - R_0)'K^{-1}(\bar{R} - R_0) \tag{4}$$

where m is the number of procedures, and R_0 and K are the mean vector and covariance matrix, respectively, for $\bar{R}$ under H_0. (It can be shown that R_0 and K are simple known functions of m and the number of data sets.) If H_0 is true, then q should have approximately a chi-square distribution with $m - 1$ degrees of freedom.

---

[1]One procedure must be left out so that K is non-singular; however, the value of q does not depend on which procedure is left out.

TABLE I.— ERTS-1 COVERAGE SCHEDULE FOR TEST SEGMENTS

ERTS-1 cycle	Month	Period	Date of overflight along track				
			L	M	N	O	P
18	June	I	8	9	10	11	12
19	June	II	26	27	28	29	30
20	July	III	14	15	16	17	18
21	August	IV	1	2	3	4	5
22	August	V	19	20	21	22	23
23	September	VI	6	7	8	9	10
24	September	VII	24	25	26	27	28
25	October	VIII	12	13	14	15	16

Counties covered:

L/M	M/N	N/O	O/P
Huntington and Shelby Counties, Indiana	White County, Indiana	Livingston and Fayette Counties, Illinois	Lee County, Illinois

TABLE II.— COMPARISON OF CROP IDENTIFICATIONS MADE BY ASCS AND BY PHOTOINTERPRETATION

Cover type	Statistic	ASCS total	Photointerpreted		
			Total	Correct	Commission error
Corn	Number of fields	50	51	46	5
	Percentage	100.0	102.0	92.0	10.0
	Number of acres	1,181	1,197	1,165	32
	Percentage	100.0	101.3	98.6	2.7
Soybeans	Number of fields	65	66	61	5
	Percentage	100.0	101.5	93.8	7.7
	Number of acres	1,550	1,540	1,523	16
	Percentage	100.0	99.3	98.3	1.1
Other	Number of fields	108	106	99	7
	Percentage	100.0	98.1	91.7	6.5
	Number of acres	879	874	838	36
	Percentage	100.0	99.4	95.3	4.1

TABLE III.– ERTS-1 MSS ACQUISITION RESULTS

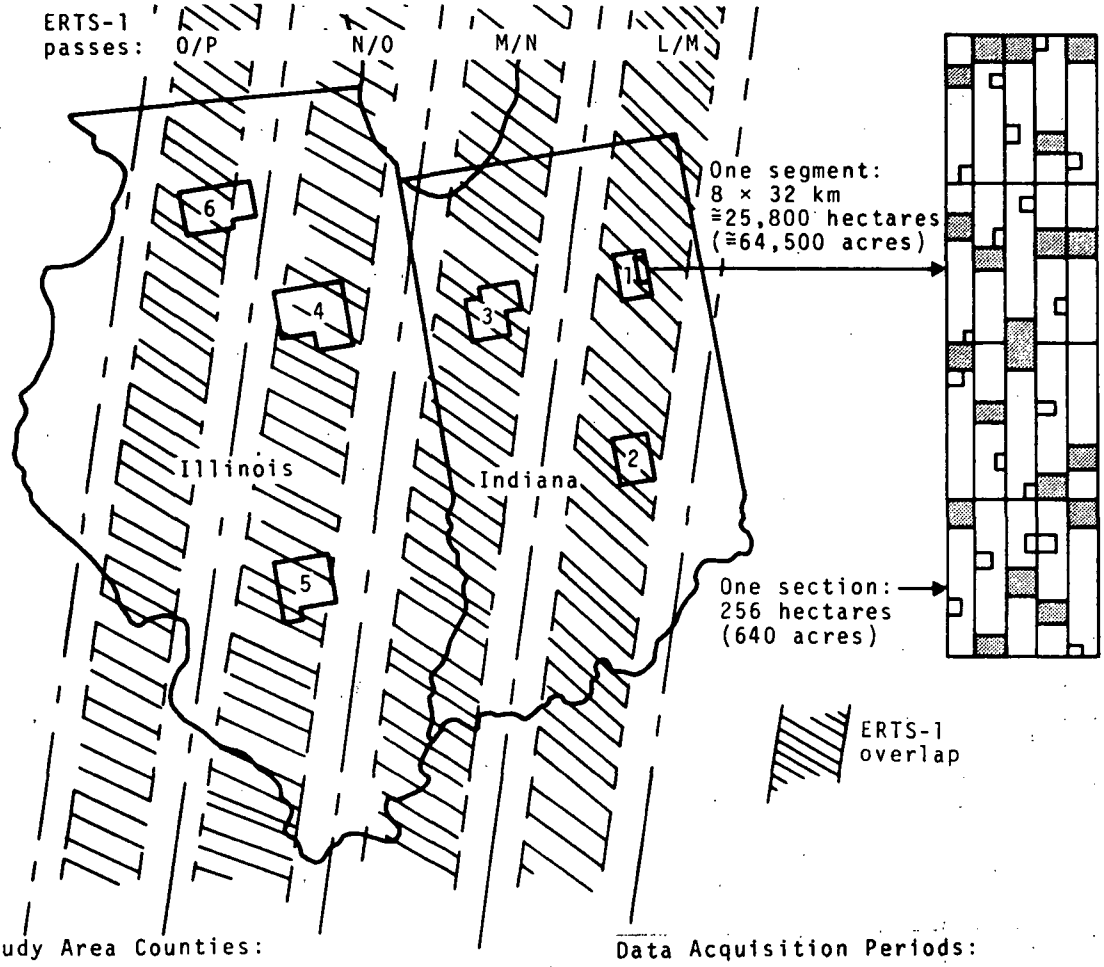
Segment	Pass	Period						
		I 6/8-12/73	II 6/26-30/73	III 7/14-18/73	IV 8/1-5/73	V 8/19-23/73	VI 9/6-10/73	VII 9/24-28/73
Huntington	1	1320-15534	F	S	F	F	S	1428-15520
	2	1321-15593	F	[a]1357-15590	F	S	S	S
Shelby	1	1320-15541	F	F	F	F	F	1428-15523
	2	F	F	P	F	P	1411-15581	S
White	1	1321-15593	F	S	S	S	1411-15581	S
	2	1322-16051	F	S	F	1394-16042	F	S
Livingston	1	1322-16051	F	1358-16045	1378-16043	S	F	F
	2	P	[b]1341-16104	P	P	F	F	P
Fayette	1	[a]1322-16054	S	1358-16051	S	1394-16044	F	F
	2	[b]1323-16112	[a]1341-16111	1359-16105	S	S	F	F
Lee	1	F	S	1359-16100	S	F	F	S
	2	F	S	[c]1360-16155	1378-16153	F	S	F

Code: F = >80 percent clouds on a frame basis; S = >80 percent clouds on a segment basis; P = incomplete coverage.

[a]The data set contains some bad data lines.

[b]The ERTS-1 band 4 data are not usable in this set.

[c]This data set contains some isolated clouds over individual fields.



Study Area Counties:

Indiana	Illinois
1. Huntington	4. Livingston
2. Shelby	5. Fayette
3. White	6. Lee

Data Acquisition Periods:

I – 6/08-12/73	V – 8/19-23/73
II – 6/26-30/73	VI – 9/06-10/73
III – 7/14-18/73	VII – 9/24-28/73
IV – 8/01-05/73	

Ground Truth:

ASCS – 20 quarter sections (white) each ERTS-1 pass

Photointerpretation – 20 sections (black) each ERTS-1 pass

Figure 1.– Technology assessment data set, June through September 1973.

4.0 RESULTS AND DISCUSSION

Performance data obtained with the several recognition processing procedures were studied to determine their respective abilities to recognize the major crops of the CITARS sites. Analysis of variance was used to investigate the effects of several experimental factors on recognition performance. Comparisons were made, both for local and nonlocal recognition. (See ref. 2, vol. IX, appendixes A and B, for tabulations of data performance and summaries of the analyses of variance.)

Many of the performance measures developed for the analyses of variance used in these comparisons have lower values than are potentially attainable from the CITARS data. For example, times of year and situations when recognition was poor were sometimes averaged with the best performances in establishing overall performance measures for an ADP procedure. Also, as discussed earlier, repeatability and removal of analyst judgment from the ADP procedures were emphasized, which may have reduced some performance levels.

The experimental factors of interest included effects of segment characteristics, observation conditions, effects of crop maturity or time of year, and comparison of performances for field centers and entire sections. The majority of the analyses were made with corn and soybeans as the major crops. Wheat was present in early June but in such small amounts that results with it as the major crop are not considered reliable and are not presented here.

4.1 COMPARISON OF ADP PROCEDURES FOR LOCAL RECOGNITION

The majority of analyses compared one standard procedure from each organization. Additional analyses were made to compare alternative procedures: use of prior probability information by LARS, linear-versus-quadratic decision rules by ERIM, and the use of multitemporal data by EOD.

4.1.1 Comparison of Standard Procedures

One of the major CITARS objectives was to determine if significant differences existed in local recognition performance among the three standard procedures: LARS/SP1, ERIM/SP1, and EOD/SP1. The results are summarized in tables IV, V, VI, and VII.

The overall analysis of variance on segment aggregates of data for all time periods showed significant differences (alpha level 0.05 unless otherwise specified) between the procedures, as well as different performance rankings for field-center and whole-area recognition (ref. 2, vol. IX, pp. 19-34). The ERIM/SP1 performed significantly better than either LARS/SP1 or EOD/SP1 on the field-center data. On the other hand, LARS/SP1 performed significantly better than ERIM/SP1 on proportion estimation for whole areas. Differences between EOD/SP1 and the other two procedures were not significant for proportion estimation.

Significant differences between the procedures were found also in the specific analyses of variance that were conducted. In seven of ten specific analyses, differences in field-center classification accuracy were significant, with

ERIM/SP1 ranking first in six of these seven and EOD/SP1 ranking third in five. For proportion estimation in whole areas, eight of ten specific analyses indicated significant differences between procedures; LARS/SP1 was most consistent, ranking first in one analysis and second in all others, while EOD/SP1 and ERIM/SP1 alternated between first and third with EOD/SP1 ranking first four times.

One overall measure of proportion estimation performance is the rms error of proportions in the aggregation of pixels in each data set. The LARS/SP1 had the smallest average rms error (0.095) over the 15 data sets, followed by EOD/SP1 with 0.108 and ERIM/SP1 with the largest at 0.150, as shown in table VI. These tabulations indicate the best performance on proportion estimates when averaged over all data sets was by LARS/SP1, and the worst was by ERIM/SP1.

As can be seen from tables V and VII, all procedures overestimated corn and soybeans proportions and underestimated the proportion of "other." This overestimation of major crops was consistent, even on a section-by-section basis within each segment. Certain parameters might be adjusted within these procedures to reduce that bias (see section 4.5), or a bias correction scheme could be applied to the classification result. Thus, since the performance numbers include both bias and variance, the procedures should not be judged too harshly against the whole-area comparisons.

Table IV suggests a possible source for these consistent overestimates of corn and soybeans proportions. In this

instance, field-center pixels for "other" have the lowest average rates of correct classification. This suggests that other ground covers were more frequently misclassified as corn or soybeans than corn or soybeans were misclassified as "other," thus leading to consistent overestimates of corn and soybeans proportions. The poorer recognition results for the class "other" may have resulted from the fact that the proportion of training samples available for "other" was low in relation to corn and soybeans; hence, a poorer representation of variability for the class "other" by the training statistics resulted (table VIII). In addition, proportion estimates obtained by classified pixel counts were biased and the magnitude of the bias is dependent upon both the actual proportion of the crop and the matrix of conditional probabilities of classification.

The procedure which obtained the highest average rate of correct classification of field-center pixels — in this instance ERIM/SP1 — did not, on the average, achieve the best proportion estimation as determined from counts of classified pixels. The ERIM/SP1 did rank first in the specific analysis which compared proportion estimation performances at the best time of year for field-center recognition (late August; see section 4.4).

The ERIM/SP1 procedure differed from the other two procedures in two major ways: It used a different decision rule, and it used a different training procedure to establish recognition signatures. Since ERIM/SP2 used a quadratic decision rule (like EOD/SP1 and LARS/SP1) but still performed more like ERIM/SP1 (see section 4.3), the training

procedures were examined and compared for possible explanations of the opposite performance rankings of the procedures on field-center and whole-area data. Proportion-dependent bias and other considerations are discussed in section 4.6.

During training, EOD and LARS employed clustering to establish multimodal major crop signatures, whereas ERIM established a single-recognition signature for each major crop and additional signatures only for selected subclasses "other." The selection of these other subclasses was based on field-center statistics and recognition of training data. In one instance, as many as eight signatures for "other" were used, while in two instances only one was used. An analysis of results showed a relatively high negative correlation between the number of other signatures used and overall rms error in proportion estimation (fig. 2). That is, the fewer the signatures used, the greater the rms error tended to be. Correspondingly, the overall proportion of "other" tended to be more underestimated with fewer signatures. Another useful test would be to use LARS/SP1 and/or EOD/SP1 signatures with the ERIM/SP2 linear rule to permit a direct comparison of results for different training procedures with the linear decision boundaries.

4.1.2 Use of Prior Probability Information

The LARS/SP1 procedure used a maximum likelihood Gaussian classifier which assumed that the frequency of occurrence of each class was the same for all classes. The Lars/SP2 procedure used prior probability information in the form of class weights and was designed to maximize overall correct classification. Theoretically, the use of the correct values

for the frequency of occurrence of each class would maximize overall performance for field-center pixels.

The overall results of the equal and unequal prior probability procedures were compared statistically. The result indicated that the use of historical data as a basis for prior probabilities did not affect whole-area or field-center performance significantly for either local or non-local recognition. However, in interpreting this result, it must be remembered that LARS/SP2 was an attempt to maximize overall performance, and the results for field centers have been measured by average classification accuracy. In addition, the quality of the prior probabilities used must be examined.

The unequal prior probabilities were based on the 1972 crop acreage estimates made by the Statistical Reporting Service, USDA, for each county. While it was anticipated that the probabilities derived from these figures would not be the true probabilities for 1973, no major change was expected.

The USDA figures were available only on a county basis, whereas CITARS examined only an 8- by 32-kilometer segment of each county. Furthermore, performance was examined on only 20 of the 100 sections in the segment. The crop proportions varied significantly from section to section; therefore, crop proportions based on county estimates may not apply.

It was concluded that while prior probability information in the form of class weights should be used when available

(because such usage has a sound theoretical basis) it may not in practice provide much, if any, improvement in performance. Further tests are recommended to determine the sensitivity of the classifier to class weights.

4.1.3 Linear-Versus-Quadratic Decision Rule

The ERIM employed both the linear and quadratic decision rules on the CITARS test data. No evidence was found in the results to indicate that the quadratic procedure was better than the linear one; if anything, the linear classification procedure gave slightly better results at approximately one-third the computational cost for the implementations used. These results agree with previous comparisons by Crane and Richardson (ref. 9).

More specifically, for local recognition of crop proportions in whole areas, the linear procedure (ERIM/SP1) had an average rms error of 0.150 over 15 local data sets, compared with 0.187 for the quadratic procedure (ERIM/SP2). Furthermore, the linear procedure had the lower rms error in 11 of the 15 data sets.

Average classification accuracy in field centers for local recognition was 0.639 for the linear procedure, compared with 0.606 for the quadratic procedure. Again, the linear procedure had the better performance in 11 of the 15 sets.

A quadratic decision rule is theoretically optimal for minimizing the overall probability of misclassification for Gaussian (normal) distributions with known parameters

and the assumption of equal prior probabilities. Therefore, it might be expected that the quadratic rule would outperform a linear decision rule. Possible reasons for the equal or better performance of the linear procedure are set out in the following discussion.

Since identical signatures were used throughout for the two decision procedures, there was no confounding effect because of different training procedures. Thus, the only reasons for performance differences are the different shapes of decision surfaces defined by the two rules and their relationships to the test data. For equal covariance matrices, the two rules would form identical decision surfaces. For situations with different covariance matrices, the linear decision rule that was utilized (ERIM/SP1) adjusted its decision surfaces to minimize the average probability of misclassification between each pair of classes, utilizing all covariance information. However, the disparity between linear and quadratic decision surfaces increases as the dispersion patterns defined by the covariance matrices become increasingly different in shape, orientation, and/or size. For example, figure 3 is a hypothetical example which illustrates a difference in space assigned to the less dispersed class A by the two decision lines based on the indicated training data.

Both the procedure used to establish signatures and the relative amounts of major crop and other training data tended to give more dispersed signatures for the major crops, especially during the early and middle parts of the growing season. The effect of such a tendency would be to have less of the decision space assigned to class "other"

by quadratic decision surfaces; for example, consider class B to be major and class A to be "other" in figure 3. In test data (both field centers and whole areas), the quadratic rule always, with but one exception, estimated less "other" and more major classes than did the linear rule for the same signatures, with the linear rule having a smaller magnitude of bias for "other."

In field-center training data, the quadratic rule slightly outperformed the linear rule in 12 of 15 cases (ref. 2, vol. VII), the opposite of what occurred with test data and what would be expected for normal distributions with known parameters. The fact that quadratic performance on training data was the better of the two reduces the likelihood that several other possible explanations were responsible for the equal or better linear performance on test data. The first of these other possible explanations is that the selection of other signatures was based on linear calculations with training data. Second, the training data may not have been normal. Finally, there can be a portion of the signal space in which the decision of the linear rule depends on the order in which signatures are considered (in any such overlap region, those considered last, which happened to be signatures for class "other" in CITARS, are favored); yet, in one test where the order of signatures was reversed, the results changed only slightly, and the new linear results still were better than the quadratic.

The next consideration is that of how well the training data represented the test data and the consequences of any differences between them on recognition using the two rules. The single signature for each major crop usually was based

on pixels from 10 or more individual fields so that they tended to represent much or most of the variability present in major crop signals. In contrast, only a few samples usually were available for each other class present; and, frequently, other classes in the test data were not represented in the training data. Because the quadratic decision surfaces in these cases tended to be closer bound to the other signatures, they consequently were more sensitive to differences between the statistics of other training and test data and to the presence of data not represented by the signatures used, including mixture pixels. The hypothetical example of figure 3 illustrates these effects also.

In addition to recognition performance, the relative costs of implementation and processing are operational performance considerations. Consideration of these latter items was not an objective of CITARS; however, computational efficiency was a major factor in ERIM's choice of the linear rule (a factor of three faster than the quadratic rule) for its principal procedure. The relative speeds of any two computational procedures are implementation dependent; that is, speed depends on both the machine used and the way the calculations are carried out. The two ERIM procedures are reasonably well balanced in their programming sophistication and flexibility. Other procedures have been defined, such as table look-up procedures (ref. 15), which are more efficient for performing quadratic calculations and may be only slightly less expensive than the comparable implementation of a linear rule. However, such procedures have storage limitations which restrict the number of channels that can be used practically. For instance, it might be difficult to utilize them for processing multitemporal data with more than four channels being used for recognition.

Another type of implementation is with special-purpose processors such as the Multivariate Interactive Digital Analysis System (MIDAS, ref. 12), to perform calculations at much higher rates than are possible on general-purpose digital computers.

To summarize, the linear decision rule performed as well as, or better than, the quadratic decision rule for the signature sets utilized. The relative dispersion volumes of major and other signatures tended to cause greater underestimation of the class "other" by the quadratic rule. Comparison of the two types of rules on CITARS signature sets obtained by another procedure, which might yield more equal dispersion volumes and hence more similar decision boundaries, would be a useful addition to these results.

4.1.4 Multitemporal Analysis

The physical phenomena associated with MSS data acquired on more than one date (multitemporal data) indicate that temporal differences in the spectral signatures from one groundcover class to another should be many times larger than spectral differences on any single date. In the development of the CITARS task a key question was: Can multitemporal data be used to improve classification performance? Actually, the question has two parts:

1. Does the temporal dimension of multitemporal data provide additional information useful for discriminating crops?
2. Can an ADP procedure be developed to extract the additional information expected from the multitemporal data?

To determine if multitemporal data processing improved the CIP, the CIP for each of several combinations of times for the Fayette County segment was compared to the CIP obtained at any single time (ref. 2, vol. VIII, part 7). Only the Fayette County segment contained sufficient cloud-free data for multitemporal processing. This segment had cloud-free data during periods I, II, III, and V.

Table IX compares the CIP's achieved for the several combinations of multitemporal data with the best single-pass results. Analysis-of-variance tests for significant differences between multitemporal recognition and the three standard procedures (LARS/SP1, ERIM/SP1, and EOD/SP1) showed multitemporal classifications to be significantly better than classifications using the three main procedures. (This included both field-center recognition and crop proportion estimation.) The results for the combination of time periods I and II were significantly better than for period II data alone. When periods II and III were combined, the results were significantly better than for period III data alone; and when data for periods I, II, III, and V were combined, the results were significantly better than for period V alone.

In summary, the CITARS task has shown that:

1. A procedure can be implemented to utilize multitemporal data (in this case a simple modification of the standard single-pass procedure EOD/SP1).
2. The overall CIP achieved by the use of multitemporal data was superior to the single-pass results for the cases considered.

Additional classifications of different areas, crops, and times will be required to test the adequacy of the ADP procedure and to determine how much CIP can be improved using multitemporal data. Finally, the benefits of improved CIP should be weighed against the cost of registering multitemporal sets of ERTS-1 data and the increased classification costs when using additional features.

4.2 NONLOCAL RECOGNITION

In nonlocal recognition, the signatures used to classify pixels in a segment were generated either from a different segment or from a different satellite pass than the data being classified. Nonlocal recognition performance proved to be substantially poorer than local performance for both field centers and proportion estimates of whole areas. Preprocessing with an MLA procedure was found to improve average nonlocal recognition performance.

4.2.1 Comparison of Local and Nonlocal Results Using Nonpreprocessing ADP Procedures

Twenty nonlocal recognition cases were analyzed for the three main ADP procedures. The average field-center CIP obtained with nonlocal signatures was only 78 percent of that obtained with local signatures, as shown in table X and figure 4. A degradation also was observed for whole areas, where the nonlocal average rms error of crop proportion estimates was 23 percent greater than that obtained locally.

A blocked-rank test did not indicate any significant difference between the main procedures for either nonlocal whole areas or nonlocal field centers. However, differences between procedures were considerable for some particular analyses of variance. Although not consistently, EOD/SP1 tended to perform the best for whole areas with ERIM/SP1 tending to be the worst. On the other hand, ERIM/SP1 usually was best for field-center analyses. The averages reflected in table X show these tendencies.

Table X also shows comparisons of alternative ADP procedures employed by ERIM and LARS. No significant differences were evident for nonlocal performances. Using identical signatures, the linear decision rule (ERIM/SP1) had slightly better average nonlocal performances for field centers and whole areas than did the quadratic decision rule (ERIM/SP2). It ranked ahead of the quadratic in 12 of 20 whole-area cases and 13 of 20 field-center cases.

Average nonlocal performance, with *a priori* crop proportion information based on the previous year's harvest in each county (LARS/SP2), was slightly worse than that based on equal prior probabilities (LARS/SP1). The latter ranked higher in 11 of 20 whole-area cases and 8 of 20 field-center cases. (Also, see sections 4.2.2 and 4.2.3.)

4.2.2 Comparison of Results Obtained With and Without Preprocessing

Signatures from one segment and time may not accurately recognize data from another segment or time because of the following factors.

1. Differences in crop characteristics between the segments or between the sample used for training and the test segment data
2. Differences in observation (scan-angle and/or atmospheric haze content)
3. Differences in sensor characteristics

Discussions of crop and atmospheric factors in the CITARS data sets are presented in section 4.3. Their implications for interpretation of nonlocal results are discussed in section 4.2.3.

One straightforward preprocessing procedure, MLA, was used in two CITARS procedures, ERIM/PSP1 and ERIM/PSP2. Preprocessing increased average, nonlocal, field-center classification by 11 percent (table X) of the accuracy without it. It also decreased the average rms error of proportion estimates by 5 percent. Based on overall results for the 20 cases, the average nonlocal performance ranking of the preprocessing procedure ERIM/PSP1 was higher than any of the three main procedures.

The overall analysis of variance for whole areas did not indicate a significant difference between ERIM/SP1 and ERIM/PSP1; however, ERIM/PSP1 (the preprocessing technique) exhibited better performance in 13 of 20 cases. In specific analyses, MLA exhibited a significant but not consistent effect; in four analyses, preprocessing was significantly better, in two significantly worse, and in most not significantly different. In five of the latter analyses, it was observed that for those cases in which MLA preprocessing performed better the other procedures performed worse, and vice versa.

The preprocessing method (ERIM/PSP1) produced better overall results for field centers in 17 of 20 nonlocal cases, a significant result at the 0.001 alpha level. In three specific analyses, the addition of preprocessing made a significant improvement, and it never significantly degraded field-center performance in other analyses.

Nonlocal performances of linear and quadratic decision rules with preprocessing were compared with the same situation for local recognition, and the results were the same. The average performance of the linear procedure was slightly better for both field centers and whole areas (see table X). Also, the linear procedure with MLA outranked the corresponding quadratic procedure in 16 of 20 cases for whole areas (a significant result at the 0.01 alpha level) and in 13 of 20 field-center cases. No specific analyses of variance were made comparing these two procedures.

4.2.3 Interpretation of Nonlocal Recognition Results

The nonlocal recognition results were analyzed for trends or patterns relating to the combinations of training and recognition segments and other factors which could have caused differences in signals and signatures and led to the unsatisfactory results obtained.

A fairly strong dependence on the segment used for training was found for whole areas. In specific analyses for the July 14 through 18 time period, often it was found that significantly better performance was achieved when the signatures applied were from the same segment on another day rather than from a different segment. This indicates

that the variability in the test fields was better represented by training fields in the same segment than by those in other segments. Signatures from two or more other segments produced inconsistent results on test data for a given segment. Also, it was found that reversing the direction (that is, exchanging the roles of training and test segments) could make a significant difference in nonlocal recognition performance. This further indicates variability in the characteristics of training data and points to the importance of having both an adequate sample for training and appropriate training procedures.

Another source of differences between data sets was in the atmospheric conditions which existed at the times of data collection. Substantial differences in haze level (measured by optical depth) and in the path radiance and atmospheric transmittance, which are dependent on haze level and on other parameters, were noted for nonlocal recognition. Sensor scan angle is another parameter which should not be discounted (refs. 16 and 17). As discussed further in section 4.3.2, a negative correlation existed between nonlocal field-center recognition performance with the three main ADP procedures and the difference in haze level between training and recognition segments. In other words, an increase in the difference resulted in a decrease in performance.

The MLA preprocessing procedure counteracted the atmosphere-dependent signal changes to a degree and reduced the amount of correlation between differences in haze level and recognition performance. The MLA procedure also took into account other conditions which could cause the average signals from each pair of segments to differ. To the extent

that the averaged areas had the same types and proportions of ground covers present, the procedure should have performed at its best. Some of the interactions observed between preprocessing and nonpreprocessing procedures were caused by signature adjustments which were greater than differences between local signatures. Although both additive and multiplicative changes are present in signals, MLA can affect only the additive correction.

The ERIM carried out an analysis supplementary to CITARS by applying a preprocessing signature-extension procedure (ref. 2, vol. VII) for Multiplicative and Additive Signature Correction (MASC, ref. 18). The MASC procedure employs an analysis of unsupervised clusters in each pair of segments to develop a signature transformation. This analysis is not exactly comparable to other CITARS results (ref. 2, vol. IX, pp. 31-34), as a different form of the data was utilized. (For instance, results for ASCS-visited and photointerpreted fields in the nonlocal areas were combined.) The results are presented here to illustrate a potential technique for solving the signature extension problem. The time period chosen was August 21, when maximum field-center classification accuracy was obtained. Average field-center accuracy was about 80 percent using ERIM/SP1 for local recognition in the Fayette and White County segments. Without preprocessing, nonlocal recognition accuracy between the two decreased to about 40 percent in each direction. The use of MLA preprocessing (ERIM/PSP1) increased average accuracy to 74 percent using Fayette signatures in the White segment but only to 46 percent using White signatures in the Fayette segment. When the MASC preprocessing transformation

was applied to the White signatures, field-center accuracy increased to 80 percent, the accuracy achieved with local signatures.

The results of the CITARS effort have manifested some of the problems associated with nonlocal recognition; that is, differences in scene condition, atmospheric state, and sensor configuration can produce signal changes and seriously degrade recognition performance in nonlocal areas. These problems must be overcome for truly operational, remote-survey operations. It is possible that a simple preprocessing procedure potentially could improve nonlocal recognition performance and that use of more sophisticated signature extension techniques may produce further improvements. It is noteworthy that this procedure is potentially capable of compensating for effects such as haze but not necessarily for on-the-ground differences such as the effects of various stages of crop maturity.

4.3 SEGMENT EFFECTS ON CROP IDENTIFICATION PERFORMANCE

One of the CITARS objectives was to quantify and evaluate segment or location effects on CIP. Segment effects include both site characteristics and atmospheric effects. Significant differences existed among the segments in the CIP as measured by both classification accuracy of field-center pixels and crop proportion estimates for whole areas. These results are summarized by time period in tables XI and XII. The segment effects, however, are difficult to isolate since they are confounded with time period and crop maturity; that is, data were not available for every segment for each time period. It is not possible with available data to describe quantitatively the effects of location

on CIP without making further assumptions about these effects. For example, simple averaging over time periods could not give a meaningful figure for comparing segments since the segments are not observed at the same time periods. It is possible, however, to obtain estimates of expected segment and time responses if one is willing to assume a noninteractive model for the expected response; that is,

$$E(y_{ij}) = \alpha_i + \beta_j \tag{4}$$

where $E(y_{ij})$ is the expected response (for example, CIP), α_i is the ith segment effect, and β_j is the *jth* time effect. Under this model the expected response for the *ith* segment (averaged over time) and the *jth* time (averaged over segments) can be estimated from the segment/time data in tables XI and XII. (See ref. 2, vol. IX, appendix C, for further details.) The results are shown in the margins designated "expected segment response" and "expected time response." Although little more can be done to separate the effects of location from those of time period quantitatively, some of the major characteristics of the sites can be described qualitatively and associated with the CIP.

4.3.1 Site Characteristics

Site characteristics which might affect CIP include soil type, field size, cropping practices, crop calendar, and weather.

Soil type *per se* probably does not materially affect CIP. (Note: Soil color could affect crop signatures and subsequent CIP, especially early in the growing season.)

However, it does affect several other factors which in turn can influence CIP. For instance, soil type has a major influence on the uniformity of an area. Similar soils occurring over large areas are generally associated with large fields, fewer crops, and uniform crop growth. In the Corn Belt, corn and soybeans are the predominant cover types in areas having uniform, productive soils. Of the CITARS test sites, Livingston County had the fewest different soil types, the largest field sizes, and most uniform fields. Classifications of Livingston County were among the best in CITARS.

At the other extreme are areas having very heterogeneous and diverse soils. These areas are characterized by small fields, a diverse set of cover types (such as small grains, forages, and woods, in addition to corn and soybeans), and less uniform growing conditions. (For example, a difference in soil-moisture-holding capacity can significantly affect crop growth.) These characteristics describe the Huntington County segment. The other four segments fall somewhere between Livingston and Huntington in most of these site characteristics.

The correlation of field size to CIP is shown in figure 5 and below.

	Quarter sections	Full sections
Overall rms error	-0.675[a]	-0.633[b]
Average rms error over sections	-.519[b]	-.536
Average conditional accuracy	.112	.089

[a]Denotes significance at 0.05 alpha level.

[b]Denotes significance at 0.001 alpha level.

Field size did not significantly affect the recognition of field-center pixels (r = 0.089), but it was negatively correlated with proportion estimates for whole areas (-0.633 and -0.536 for overall and average over sections, respectively). This indicates that the proportions of corn, soybeans, and "other" were estimated more accurately in segments with larger fields than those with smaller fields. This effect is attributed to:

1. The percentage of mixture pixels (that is, pixels falling on field boundaries and containing two or more cover types) is smaller in areas having large fields.
2. Large fields generally are associated with uniform soils and crop growth.
3. The proportion of class "other" was higher in areas with small fields, and indications were that proportion estimation accuracy is negatively correlated with the amount of class "other" present (see section 4.6.1).

The problem of mixture pixels would be reduced by increasing the spatial resolution of the scanner system, since there is no real assurance that mixture pixels are classified in the same proportions as pure pixels or as the cover types that occur in the area. This is one possible source of bias in the proportion estimates.

The decreased performance associated with more heterogeneous areas with smaller fields is to be expected, and there is no easy solution for this problem. It is clear, however, that additional information is required to train the classifiers so that the existing classes will be represented accurately.

4.3.2 Atmospheric Effects

A study was conducted to investigate the effect of atmospheric haze level on CITARS classification accuracy. The effect on local classification was investigated by plotting the classification accuracy obtained at the various sites as a function of the optical depth for the 0.5-micrometer band. The effect on nonlocal classification was investigated by plotting the classification accuracy as a function of the difference between the haze levels at training and at test sites. The values obtained for the optical depth at 0.5 micrometer for the various CITARS passes are shown in table XIII.

Figure 6 shows plots of the accuracy of local classification as a function of optical depth. The correlation of the data in figure 6 is -0.602. Since this figure contains points corresponding to a number of sites and a number of different passes, considerable scatter was expected to be in the data. Up to an optical depth of 0.4 micrometer there is little indication of dependence on haze level. If the points for larger haze levels are included, a weak negative correlation seems to exist between optical depth and local classification accuracy. (The correlation coefficient = -0.60 with significance at an alpha level of 0.05.) Previous theoretical calculations (ref. 13) and simulations (ref. 14) have shown that uniform haze level would have little or no effect on local classification. However, real haze is never perfectly uniform; therefore, some deteriorating effect on classification accuracy would be expected. The results shown in figure 6 must be interpreted with caution, as other possible contributing factors (such as site effects and Sun angle) have not been considered here.

Figure 7 shows the classification accuracies obtained for nonlocal classification as a function of the difference in haze level between the training and the test sites, with and without MLA preprocessing (table XIII). The correlation coefficient for ERIM/SP1 and ERIM/SP2 (without preprocessing) was -0.769 (significant at an alpha level of 0.001) compared to -0.280 (not significant) for ERIM/PSP1 and ERIM/PSP2 (with preprocessing). These results indicate that differences in haze levels between training and classification sites can adversely affect CIP but that some of these effects can be removed or adjusted by a preprocessing procedure such as MLA.

4.4 EFFECT OF CROP MATURITY ON CROP IDENTIFICATION PERFORMANCE

It is well known that crop maturity stage affects the remote identification of crops. Most previous studies have been limited to one or two dates near the optimum time for discriminating the crops of interest. One of the CITARS objectives was to determine the effect of crop maturity on CIP. The available results assessing the effect of time period are summarized in tables XI and XII for recognition of field-center pixels and proportion estimates for whole areas, respectively. Again, the analysis is severely limited by missing segments. Certain trends may be observed, however, in these expected values when computed using a non-interactive prediction model (see section 4.3 for a discussion of the model). The field-center CIP's increased from periods II to V and then decreased significantly in periods VI and VII. Proportion estimation errors, on the other hand, were approximately the same for all time periods (except for period III, when they were greater).

The peak CIP associated with the late August period is attributed to two factors.

1. Prior to this time soybeans had not reached their fullest vegetative growth and ground cover, thereby increasing the probability of confusion with other cover types having less ground cover and leaf area.
2. By late August all cornfields have tasseled.

Prior to late August, then, corn and soybeans are still growing and developing, and there is more variability among fields of both corn and soybeans. For example, in early August some soybean fields may have almost complete ground cover while others have only partial cover. Similarly, not all fields of corn tassel at the same time; however, by late August differences in growth and development equalize. The rapid decrease in CIP in September is attributed to the onset of senescence. Again, the variability among fields increases during this time; and the amount of ground cover, particularly for soybeans, decreases. By late September most soybean fields will have lost all their leaves, causing soil to have a major influence on the spectral response.

The peak of expected proportion estimation error in mid-July corresponded with the greater variability among corn and soybean fields at that time. It was noted, however, that variability in performance among procedures at any given time was much greater for proportion estimation than for field-center classification.

One further note on the effect of crop maturity should be made with regard to nonlocal training and recognition: Crop maturity differences must be taken into account when

an attempt is made to transfer training statistics from one area to another some distance away. If the crops in the two areas are not in nearly the same stage of maturity, poor results can be expected. In particular, maturity differences are most likely to exist in the north-south direction, since planting and growth of crops are highly correlated with latitude.

4.5 EFFECT OF DATA PREPARATION ON CROP IDENTIFICATION PERFORMANCE

4.5.1 Effects of Multitemporal Registration

To facilitate the classification of multitemporal ERTS-1 data without having to locate section and field coordinates in each segment/date combination of data, the satellite passes over each segment were registered as part of the data preparation phase (ref. 2, vol. V). An experiment was performed to determine if registration had any effect on CIP and, if so, the magnitude of the effect.

In the experiment, CIP's obtained with registered and nonregistered forms of ERTS-1 data were compared. Both forms of data were geometrically corrected. The coordinates of sections and fields used for the registered data were the same as those used in the CITARS classifications. The coordinates from approximately the same fields were located in the nonregistered data by manually placing the photo-overlays over the ERTS-1 imagery. A one-to-one correspondence of fields in both data sets was not used because doing so would have eliminated fields which were required for training. About 80 percent of the fields were common to both data sets. The same procedure was used to select

pixels from fields; that is, one "guard" pixel was selected between a field boundary and any selected pixel. The same classification procedures (that is, LARS/SP1 and LARS/SP2) were applied to both the registered and nonregistered data sets for all five segment/date combinations. Recognition performances for fields and proportion estimates for sections were tabulated, and an analysis of variance was performed to determine if any significant differences existed between the registered and nonregistered data.

The results of the comparison of overall classification accuracy for field-center pixels for the two forms of multitemporal data are summarized as follows.

Segment	Period	Correct field-center classification, percent	
		Nonregistered (a)	Registered (a)
Fayette	II	42.1	51.7
Fayette	III-1	72.5	52.8
Huntington	III	64.7	45.3
Livingston	IV	71.6	66.8
White	V	77.2	75.7

[a]The figures depict the mean of the LARS/SP1 and LARS/SP2 procedures.

The analysis of variance indicated that no significant difference in classification accuracy existed between the registered and nonregistered data. However, in two cases, Fayette, period III-1, and Huntington, period III, higher performance was obtained with the nonregistered data. This may be attributed to having different samples of training and test fields for the two data sets. Other CITARS results

show that there were significant differences in recognition for different selections of fields from the same segment (see section 4.5.2). Problems with registration, if any, would be expected to appear first in segments with small field sizes, such as those in Fayette and Huntington Counties. The size of the sample of fields is considered the most likely cause of differences, although in one case (Fayette, period II) higher performance was obtained with the registered data.

4.5.2 Effects of Training Set Selection

One of the objectives of CITARS was to examine the effect on CIP when the training set selection was varied. Originally two training sets, each containing 10 quarter sections, were to have been available for comparison. However, as training fields were selected, it became obvious that 10 quarter sections would not provide an adequate training sample; thus, two sets were combined to provide a 20-quarter-section training set.

To vary the training set for this experiment, 10 pilot sections and 10 test sections were used to train the classifier. The CIP for each of these training sets was compared to the CIP for the 20-quarter-section training set. Since the 10-section samples were twice as large as the 20-quarter-section sets, it was possible to estimate the effect of training set size as well as sample selection.

Ten data sets were classified using the 10 pilot sections as the basis for training the classifier (pilot as train) and then classified again using the 10 test sections as the basis for training (test as train). The analysis

procedures were the same as for other classifications of ERTS-1 data performed by LARS (that is, LARS/SP1 and LARS/SP2).

The pilot-as-train classifications were compared to the regular CITARS classifications (train as train) by examining the overall CIP of field-center pixels from the 10 pilot sections (pilot as test). This method of comparison was used to avoid biasing the CIP by testing samples which were used to train the classifier. Analysis of variance was performed on both overall classification accuracy and proportion estimates.

The results of the various combinations of training and test samples are summarized in table XIV. Using proportion estimates as the dependent variable in the analysis of variance, training and test fields yielded significantly different results. Since different sampling procedures were followed in their selection, training and test sections could be from two different populations. Analysis of the classification accuracies showed that the test and pilot results were significantly different. This result is attributed to variations in sampling, either the size of the sample or bias in selection. Since random sampling was used to divide the sections into test and pilot sections, the differences are attributed to normal variations in sampling a population.

Table XIV shows the number of pixels in each training set. In only four cases the number of pixels in the test or pilot fields was approximately twice as great as in the training fields; thus, the effect of training set size could

not be fully evaluated. However, in those four cases the presence of more training pixels did not cause significant improvement in CIP.

The results of this experiment indicate:

1. Significant differences in CIP can be obtained with different samples of training fields.
2. Training set size alone probably was not the primary factor that limited the accuracy of the CITARS classifications.

4.6 CLASSIFICATION ACCURACY AND PROPORTION ESTIMATION

4.6.1 Comparison of Field Center Versus Whole Area

When the relative performances of the main procedures on both field-center and whole-area data used for the analyses of variance are surveyed, it becomes noticeable that ERIM/SP1 tends to be the best procedure with respect to field-center classification accuracy but the worst on proportion estimation of whole sections.

In order to seek explanations of this apparent anomaly, further studies were made on four data sets [LE(6), LI(5), WH(11), and HU(13)] exhibiting the property that ERIM/SP1 was the best procedure on field-center accuracy but the worst on whole-section proportion estimation [see table XV(a) and (b)].

Four possible explanations for this behavior are:

1. Only the test sections were used in field-center analyses, whereas all cloud-free test and pilot sections were included in the whole-area comparisons.
2. The proportion estimation procedure based on counting classified pixels is biased; hence, the bias may cause ERIM/SP1 to have a larger mean-square error on proportion estimation even though it has a smaller average probability of misclassification.
3. A procedure which classifies pure pixels accurately may not necessarily give good proportion estimates when mixture or boundary pixels are classified.
4. The proportions of class "other" in the whole areas were usually larger than for field centers, which made ERIM/SP1 appear to be worse for whole areas.

It is unlikely that factor 1 is the answer since, first, the test pilot sections were selected at random; second, the loss of a section because of clouds should be independent of the theoretical performance of a procedure on that section; and, third, analyses of variance involving the four data sets showed the means for procedures to be significantly different. (That is, even taking the random variation from section to section into account under the hypothesis that the procedures were equivalent on the average, it was unlikely that the observed procedure means could have been diverse as they were.) Furthermore, the blocking efficiency on the White and Huntington data sets was so poor that at least as effective a test could have been made by choosing sections at random for each procedure instead of comparing procedures for the same sections.

If factor 2 were true, then ERIM/SP1 should have been worst on proportion estimation using field-center pixels. This was not the case, however. Table XV(c) shows the performance of the three procedures on field-center proportion estimation using the same dependent variable as for whole areas.

A comparison of tables XV(a) and XV(c) shows that except for the case of HU(13), those procedures that were best for field-center average probability of misclassification were also best on field-center proportion estimation, and vice versa. For this reason, it is suggested that factor 3 (mixture pixels) and not bias is largely responsible for discrepancies between proportion estimation and classification accuracy results. Also, studies disclosed that factor 4 (actual proportion of class "other") was a probable contributing cause of the observed performances.

4.6.2 Bias in Proportion Estimation

It is well known that if pure pixels are being classified and classified pixels are being counted to obtain an estimate of crop proportions, the resulting estimate e is biased; that is,

$$E(\hat{e}) \neq \alpha \tag{5}$$

where

$\hat{e}$ = the estimated proportion vector

α = the true proportion vector

E( ) = the expectation operator

In fact, it is easily shown that $E(\hat{e}) = P\alpha$ where $P = (p_{ij})$ is the matrix of conditional probabilities of classifying a pixel from class j as class i.

Attempts were made to reduce biases in CITARS proportion estimates by computing a corrected or inverted proportion estimator $\hat{\alpha}$, which was the solution to the problem:

Minimize $(\hat{e} - P\hat{\alpha})^T(\hat{e} - P\hat{\alpha})$ with respect to $\hat{\alpha}$
subject to the constraints $\Sigma\hat{\alpha}_i = 1$ and $\hat{\alpha}_i \geq 0$.

The matrix P was obtained from results of field-center classification of the test sections.

The whole-area proportion estimates from four data sets were corrected for bias on a section-by-section basis, and new values of the analysis-of-variance dependent variable V were computed. Table XVI shows mean values of V to be consistently higher (less accuracy) than they were in table XV-B; that is, the known-to-be-biased method of pixel counting gave much better results than the corrected estimator.

Two possible explanations are presented immediately:

1. A confusion matrix for whole-area (that is, mixture) pixels is vastly different from one for pure pixels.
2. The section-to-section variation in crop signatures is so great that one confusion matrix cannot be applied to all sections.

To test assumption 2 independent of 1, the field-center proportion estimates as obtained from pilot sections were corrected using the aggregated P-matrix obtained from test sections. The result should be valid if the conditional

probabilities of misclassification are assumed to be about the same for all sections. An examination of table XVII shows the assumption is probably false; that is, assumption 2 is a likely explanation for the poor results obtained with a corrected proportion estimator (compare with table XV-C). No data are available for EOD, since field-center pilot sections were not processed.

In order to reduce the section-to-section variation, the correction procedure was applied to the whole-area estimates aggregated over all sections in each data set. Poor results indicated that assumption 1 is also true — that a confusion matrix estimated from classifying pure pixels cannot be effectively used to unbias a whole-area estimate.

Table XVIII shows the rms error before and after correction where one aggregated estimate is made for each data set. Note that in almost every case, the uncorrected proportion estimate was more accurate.

TABLE IV.– MEAN FIELD-CENTER CLASSIFICATION ACCURACY FOR THREE STANDARD PROCEDURES AVERAGED OVER 15 DATA SETS

Class	Procedure		
	LARS/SP1	ERIM/SP1	EOD/SP1
Corn	0.66	0.70	0.62
Soybeans	.59	.68	.61
Other	.50	.53	.46
Rms error	.58	.64	.57

TABLE V.– OVERALL BIAS AND RMS ERROR IN PROPORTION ESTIMATION FOR THREE STANDARD PROCEDURES AVERAGED OVER 15 DATA SETS

Class	Procedure		
	LARS/SP1	ERIM/SP1	EOD/SP1
Corn	0.063	0.064	0.025
Soybeans	.033	.059	.081
Other	-.096	-.124	-.106
Rms error	.095	.150	.108

TABLE VI.— AVERAGE CONDITIONAL CLASSIFICATION ACCURACY AND RANKINGS OF THREE STANDARD PROCEDURES

Data set (a)	Procedure					
	LARS/SP1		ERIM/SP1		EOD/SP1	
	Accuracy	Rank	Accuracy	Rank	Accuracy	Rank
HU(6)	0.607	3	0.670	2	0.688	1
HU(13)	.484	2	.555	1	.425	3
SH(12)	.502	3	.536	2	.546	1
SH(13)	.384	3	.551	1	.492	2
WH(10)	.742	2	.797	1	.607	3
WH(11)	.609	1	.581	3	.590	2
LI(5)	.588	2	.695	1	.579	3
LI(7)	.700	1	.694	2	.611	3
FA(4)	.544	3	.668	1	.572	2
FA(5)	.538	2	.654	1	.500	3
FA(6)	.620	2	.670	1	.605	3
FA(9)	.797	2	.809	1	.747	3
LE(5)	.539	3	.566	1	.545	2
LE(6)	.559	1	.547	3	.556	2
LE(8)	.551	2	.597	1	.440	3
Mean over all data sets	0.584	2.1	0.639	1.5	0.567	2.4

[a]Data sets: HU is Huntington, SH is Shelby, WH is White, LI is Livingston, FA is Fayette, and LE is Lee County. Number in parentheses = [(period number - 1) × 2] + pass number . Example: HU(6) is the Huntington County, Indiana, segment for period III, pass 2; [(3 - 1) × 2] + 2 = 6 .

TABLE VII.— RMS ERRORS[a] IN PROPORTION ESTIMATION AND RANKINGS OF THREE STANDARD PROCEDURES

Data set	Procedure					
	LARS/SP1		ERIM/SP1		EOD/SP1	
	Error	Rank	Error	Rank	Error	Rank
HU(6)	0.330	3	0.202	2	0.192	1
HU(13)	.131	1	.252	3	.134	2
SH(12)	.027	2	.040	3	.027	1
SH(13)	.151	3	.116	2	.096	1
WH(10)	.065	1	.100	3	.083	2
WH(11)	.057	1	.178	3	.070	2
LI(5)	.004	1	.091	3	.022	2
LI(7)	.013	1	.080	3	.028	2
FA(4)	.115	2	.106	1	.133	3
FA(5)	.144	1	.161	2	.178	3
FA(6)	.154	1	.188	2	.191	3
FA(9)	.158	2	.141	1	.204	3
LE(5)	.020	1	.232	3	.066	2
LE(6)	.025	1	.239	3	.142	2
LE(8)	.034	1	.118	3	.051	2
Mean over all data sets	0.095	1.5	0.150	2.5	0.108	2.1

[a]Overall data set estimates.

TABLE VIII.— SUMMARY OF PLOT AND PIXEL COUNTS

[Ref. 2, vol. VII, table I]

TEST

NUMBERS OF PLOTS AND PIXELS

CROP	FAYETTE 6/10	FAYETTE 6/11	FAYETTE 6/29	FAYETTE 7/16	FAYETTE 7/17	FAYETTE 8/21	SHELBY 6/8	SHELBY 9/7	SHELBY 9/24	HUNTINGTON 7/15	HUNTINGTON 9/24	WHITE 8/21	WHITE 9/7	LIVINGSTON 7/16	LIVINGSTON 8/3	LEE 7/17	LEE 7/18	LEE 8/5
CORN	34 286	34 286	31 271	34 286	34 286	34 286	56 638	56 638	56 638	28 157	28 157	74 628	74 628	49 498	48 466	47 448	49 454	49 454
SOY	46 358	46 358	44 340	46 358	46 358	46 358	39 237	39 237	39 233	35 189	35 189	61 525	61 525	64 819	59 772	48 573	47 569	47 569
WHEAT*	8 65	8 65	8 65	8 65	8 65	8 65	4 36	4 36	4 36			4 26	4 26					
TREE	12 113	11 105	11 110	12 104	12 113	12 113	2 10	2 10	2 10	5 57	5 57	1 4	1 4			4 25	4 25	4 25
PASTURE	6 23	6 23	4 17	6 23	6 23	6 23	3 6	3 6	3 6	5 16	5 16	9 103	9 103	4 10	4 10	10 61	10 61	10 61
GRAIN	12 49	12 49	9 39	12 49	12 49	12 49	3 15	3 15	3 15	1 4	1 4			3 13	3 13	6 16	5 14	5 14
CITY	4 103	2 8	4 103	4 103	4 103	4 103	3 53	3 53	3 53			3 111	3 111	3 388	3 164	4 33	4 33	4 33
HAY	1 9	1 9	1 9	1 9	1 9	1 9						7 73	7 73					
WDS/PAST	2 12	2 12	2 12	2 12	2 12	2 12	1 25	1 25	1 25	4 52	4 52			1 19	1 19			
WATER										4 79	4 79							
WEED												1 4	1 4					
OTHER							15 55	15 55	15 55	13 612	13 612	4 48	4 48	1 4	1 4	7 49	7 49	7 49
OATS										1 4	1 4							
QUARRY												1 25	1 25					
TR PARK														1 5	1 5			

TRAINING

CROP	FAYETTE 6/10	FAYETTE 6/11	FAYETTE 6/29	FAYETTE 7/16	FAYETTE 7/17	FAYETTE 8/21	SHELBY 6/8	SHELBY 9/7	SHELBY 9/24	HUNTINGTON 7/15	HUNTINGTON 9/24	WHITE 8/21	WHITE 9/7	LIVINGSTON 7/16	LIVINGSTON 8/3	LEE 7/17	LEE 7/18	LEE 8/5
CORN	9 70	9 64	9 68	9 70	9 70	9 70	24 160	24 160	24 160	6 62	6 62	24 474	24 474	11 173	11 173	14 128	15 131	15 131
SOY	7 56	7 56	15 115	17 131	18 139	21 192	7 35	11 55	11 55	15 159	15 159	19 218	19 218	20 248	20 248	11 110	11 110	11 110
WHEAT	9 48	9 48	3 24	2 18	2 18		4 26	1 2	1 2									
TREE	2 104	2 104	2 90	2 104	2 104	2 104	7 20	4 33	5 20	6 59	6 59	3 64	2 65	1 22	1 22	2 225	2 225	2 225
BARE	3 45	3 45	2 12	3 45	3 45	2 12	4 20									1 3	1 3	1 3
BRUSH	1 3	1 3	1 3	1 3	1 3													
CLOVER	1 4	1 4	2 5	2 5	2 5	3 9	1 8	4 31	4 31									
STUBBLE				1 6	1 6			1 4	1 4									
WATER	3 58	3 58	3 58	3 58	3 58	3 58				2 36	2 36							
PASTURE	13 95	12 89								2 5	2 5	2 30	2 30	1 3	1 3	1 9	2 20	2 20
FESCUE							1 2	1 2	1 2									
GRASS							1 4	1 4	1 4									
OATS							1 3							1 9	1 9	2 21	2 21	2 21
WEED			3 45		1 12	1 6												
OTHER														1 42	1 42	1 125	1 125	1 125
QUARRY												1 10	1 19	2 47	2 47			
HAY																1 2	1 2	1 2

*We have serious doubts about the validity of 27 of the 65 test wheat pixels in Fayette; See Sec. 5.2.

**Upper number of each pair denotes number of plots; lower number is pixel count.

TABLE IX.– COMPARISON OF AVERAGE CONDITIONAL CLASSIFICATION ACCURACY AND RMS ERRORS IN PROPORTION ESTIMATES FOR SINGLE "BEST" DATES AND MULTITEMPORAL

Period	Average conditional accuracy		The rms errors in proportion estimates	
	Single "best" date (a)	Multi-temporal	Single "best" date (a)	Multi-temporal
I and II	0.67	0.68	0.070	0.054
II and III	.67	.85	.144	.083
III and V	.81	.85	.141	.094
I, II, III, and V	.81	.89	.141	.069

TABLE X.– COMPARISON OF LOCAL AND NONLOCAL RECOGNITION PERFORMANCES

Comparison	Procedure	Preprocessing	Average classification accuracy over all cases (a)		Average rms error in proportion estimation over all cases	
			Local (b)	Nonlocal	Local (b)	Nonlocal
Major procedures	LARS/SP1	Without	0.586	0.453	0.098	0.157
	ERIM/SP1	Without	.644	.490	.156	.182
	EOD/SP1	Without	.572	.461	.112	.135
	Average		.601	.468	.122	.158
Linear vs quadratic decision rule	ERIM/SP1 (linear)	Without	.644	.490	.156	.182
	ERIM/SP2 (quadratic)	Without	.612	.486	.197	.206
	Average		.628	.488	.177	.194
Use of *a priori* probabilities	LARS/SP1 (equal weights)	Without	.586	.453	.098	.157
	LARS/SP2 (unequal weights)	Without	.594	.445	.129	.177
	Average		.590	.449	.114	.167
Use of MLA preprocessing	ERIM/PSP1 (linear)	With	.644	.556	.156	.157
	ERIM/PSP2 (quadratic)	With	.612	.543	.197	.210
	Average		.628	.550	.177	.184
	ERIM/SP1 (linear)	Without	.644	.490	.156	.182
	ERIM/SP2 (quadratic)	Without	.612	.486	.197	.206
	Average		.628	.488	.177	.194

[a]Twenty nonlocal recognition cases were analyzed for the three classes of corn, soybeans, and "other."

[b]A weighted average was used for local recognition to account for segments that were processed with more than one set of nonlocal recognition signatures.

TABLE XI.– AVERAGE CONDITIONAL CLASSIFICATION ACCURACY AS A FUNCTION OF SEGMENT AND DATE

Data set	Period and date						Expected segment response (a)
	II 6/26 to 6/30/73	III 7/14 to 7/18/73	IV 8/01 to 8/05/73	V 8/19 to 8/23/73	VI 9/06 to 9/10/73	VII 9/24 to 9/28/73	
HU(6) HU(13)		0.65				0.49	0.62
SH(12) SH(13)					0.51	.47	.57
WH(10) WH(11)				0.74	.59		.61
LI(5) LI(7)		.61	0.68				.62
FA(4) FA(5) FA(6) FA(9)	0.59	.58 .64		.78			.61
LE(5) LE(6) LE(8)		.55 .55	.55				.54
Expected time response[a]	0.57	0.60	0.63	0.75	0.56	0.48	–

[a]Using noninteractive model (ref. 2, vol. IX, appendix C).

TABLE XII.– RMS ERRORS FOR PROPORTION ESTIMATES AS A FUNCTION OF SEGMENT AND PERIOD

Data set	Period and date						Expected segment response (a)
	II 6/26 to 6/30/73	III 7/14 to 7/18/73	IV 8/01 to 8/05/73	V 8/19 to 8/23/73	VI 9/06 to 9/10/73	VII 9/24 to 9/28/73	
HU(6) HU(13)		0.268				0.195	0.212
SH(12) SH(13)					0.063	.102	.081
WH(10) WH(11)				0.073	.117		.97
LI(5) LI(7)		.059	0.052				.42
FA(4) FA(5) FA(6) FA(9)	0.130	.177 .191		.170			.156
LE(5) LE(6) LE(8)		.157 .156	.076				.108
Expected time response[a]	0.090	0.154	0.105	0.111	0.117	0.118	–

[a]Using noninteractive model (ref. 2, vol. IX, appendix C).

TABLE XIII.— OPTICAL DEPTH MEASUREMENTS MADE AT THE TIME OF ERTS-1 OVERPASSES

Local recognition		Nonlocal recognition		
Data set	Optical depth (haze)		Recognition (a)	Difference in optical depth
HU(6)	0.3	1	FA(5) → FA(6)	0.04
HU(13)	1.1	2	FA(6) → FA(5)	-.04
SH(12)	.18	3	LE(5) → LE(6)	—
SH(13)	.99	4	LE(6) → LE(5)	—
SH(10)	.11	5	HU(6) → LI(5)	-.13
WH(11)	.12	6	HU(6) → LE(6)	-.14
LI(5)	.17	7	LE(6) → LI(5)	-.27
LI(7)	.28	8	LE(6) → HU(6)	-.14
FA(4)	No data	9	LI(7) → LE(8)	.06
FA(5)	.35	10	LE(8) → LI(7)	.06
FA(6)	.39	11	LI(5) → FA(5)	.18
FA(9)	.27	12	FA(5) → LI(5)	.18
LE(5)	No data	13	WH(11) → SH(12)	.06
LE(6)	.44	14	SH(12) → WH(11)	-.06
LE(8)	.34	15	SH(13) → HU(13)	.11
		16	HU(13) → SH(13)	-.11
		17	FA(6) → HU(6)	-.09
		18	HU(6) → FA(6)	.09
		23	WH(10) → FA(9)	.16
		24	FA(9) → WH(10)	-.16

[a]The recognitions as shown are read:
Training segment pass → recognition segment pass.

TABLE XIV.— COMPARISON OF DIFFERENT SAMPLES OF TRAINING AND TEST FIELDS

[Data correctly classified for field-center accuracy, percent]

Segment/ period	Source of training data						
	Training fields			Pilot fields		Test fields	
	Training	Pilot	Test	Training	Test	Training	Pilot
HU III	92.3	28.4	80.1	89.7	78.7	87.1	72.7
LI III	78.1	58.8	60.6	81.4	76.1	75.2	71.7
FA III-2	77.8	52.9	63.7	86.8	69.7	89.8	73.7
LE III-2	80.2	53.2	61.7	58.8	64.3	79.7	54.8
LE IV	75.5	62.4	49.9	71.0	57.0	75.9	59.2
WH V	87.9	75.8	74.3	88.3	80.7	84.1	67.0
FA V	90.5	79.7	79.5	84.4	86.3	90.5	85.2
SH VI	77.1	48.0	51.8	76.4	49.2	76.9	58.0
HU VII	81.2	40.9	68.2	87.1	66.8	78.2	60.5
SH VII	73.5	52.9	43.8	64.7	51.6	71.6	61.3
Mean over all segments	81.4	55.3	63.4	78.9	68.0	80.9	66.4

TABLE XV.— COMPARISON OF FIELD-CENTER CLASSIFICATION ACCURACIES AND WHOLE-AREA CROP PROPORTION ESTIMATES

Data set	LARS/SP1	ERIM/SP1	EOD/SP1
(a) Mean values of dependent variable for field-center classification error			
LE(6)	1.341	0.999	1.149
LI(5)	1.142	1.012	1.206
WH(11)	1.111	1.035	1.238
HU(13)	1.139	1.113	1.369
(b) Mean values of dependent variable for whole-area proportion estimation			
LE(6)	1.125	2.212	1.659
LI(5)	0.596	1.314	.462
WH(11)	1.538	2.172	1.074
HU(13)	1.351	2.823	1.644
(c) Mean values of dependent variable for field-center proportion estimation			
LE(6)	2.281	1.777	2.057
LI(5)	2.024	1.722	2.123
WH(11)	1.711	1.493	2.400
HU(13)	1.154	2.080	2.799

TABLE XVI.– MEAN VALUES OF V FOR CORRECTED WHOLE-AREA PROPORTION ESTIMATES

Data set	LARS	ERIM	EOD
LE(6)	2.201	2.680	2.234
LI(5)	2.182	2.168	2.152
WH(11)	2.013	2.262	2.035
HU(13)	2.421	3.052	3.128

TABLE XVII.– MEAN VALUES OF V FOR CORRECTED WHOLE-AREA PROPORTION ESTIMATES – FIELD-CENTER PILOT DATA

Data set	LARS	ERIM
LE(6)	3.116	3.060
LI(5)	2.152	1.097
WH(11)	2.458	2.148
HU(13)	3.863	2.516

TABLE XVIII.– COMPARISON OF RMS ERROR FOR UNCORRECTED AND CORRECTED OVERALL-SEGMENT-PROPORTION ESTIMATES

Data set	LARS		ERIM		EOD	
	Uncorrected	Corrected	Uncorrected	Corrected	Uncorrected	Corrected
LE(6)	0.077	0.137	0.187	0.220	0.121	0.302
LI(5)	.021	.066	.073	.135	.045	.029
WH(11)	.057	.171	.178	.178	.070	.074
HU(17)	.130	.266	.252	.230	.133	.229

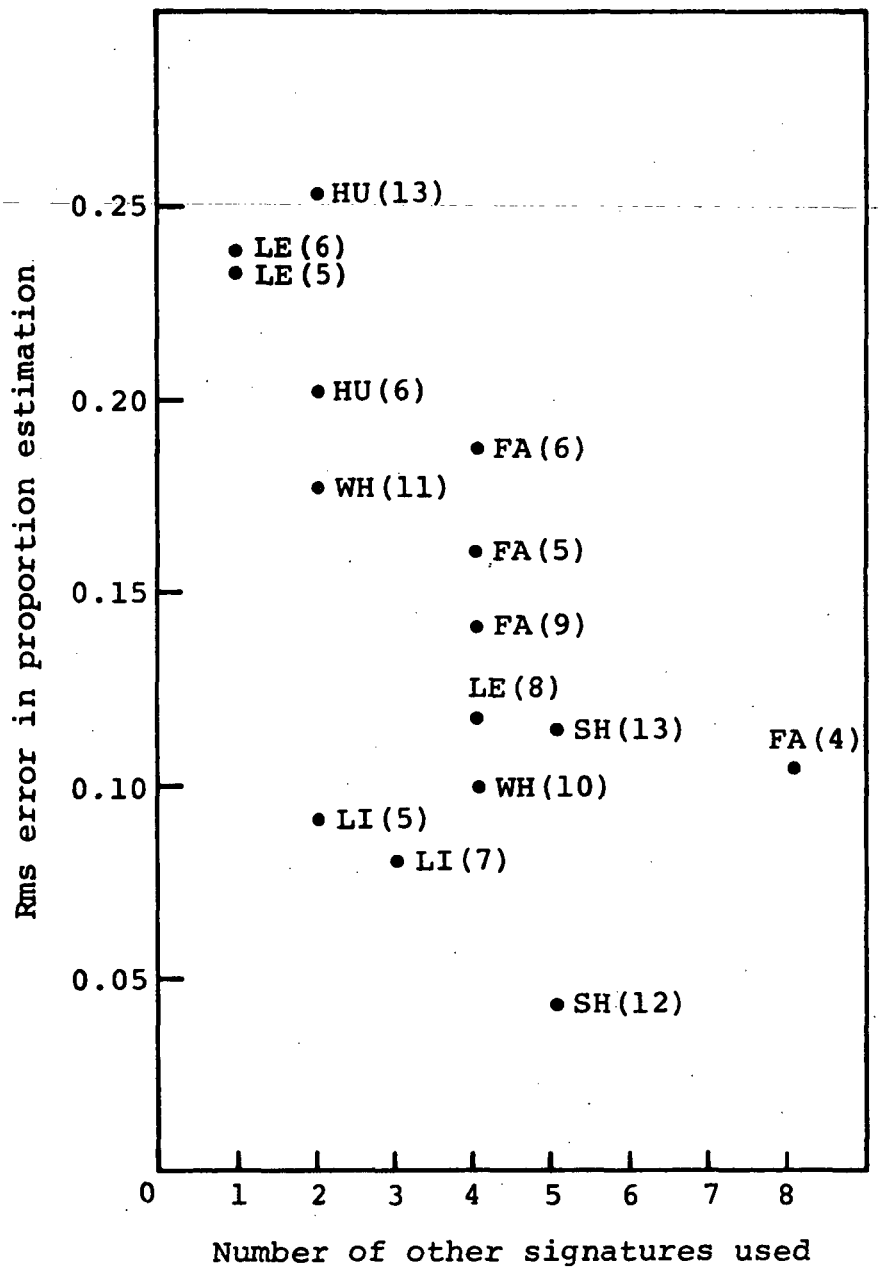


Figure 2.— Effect of number of signatures used on full-section proportion estimation by ERIM/SP1 on CITARS data.

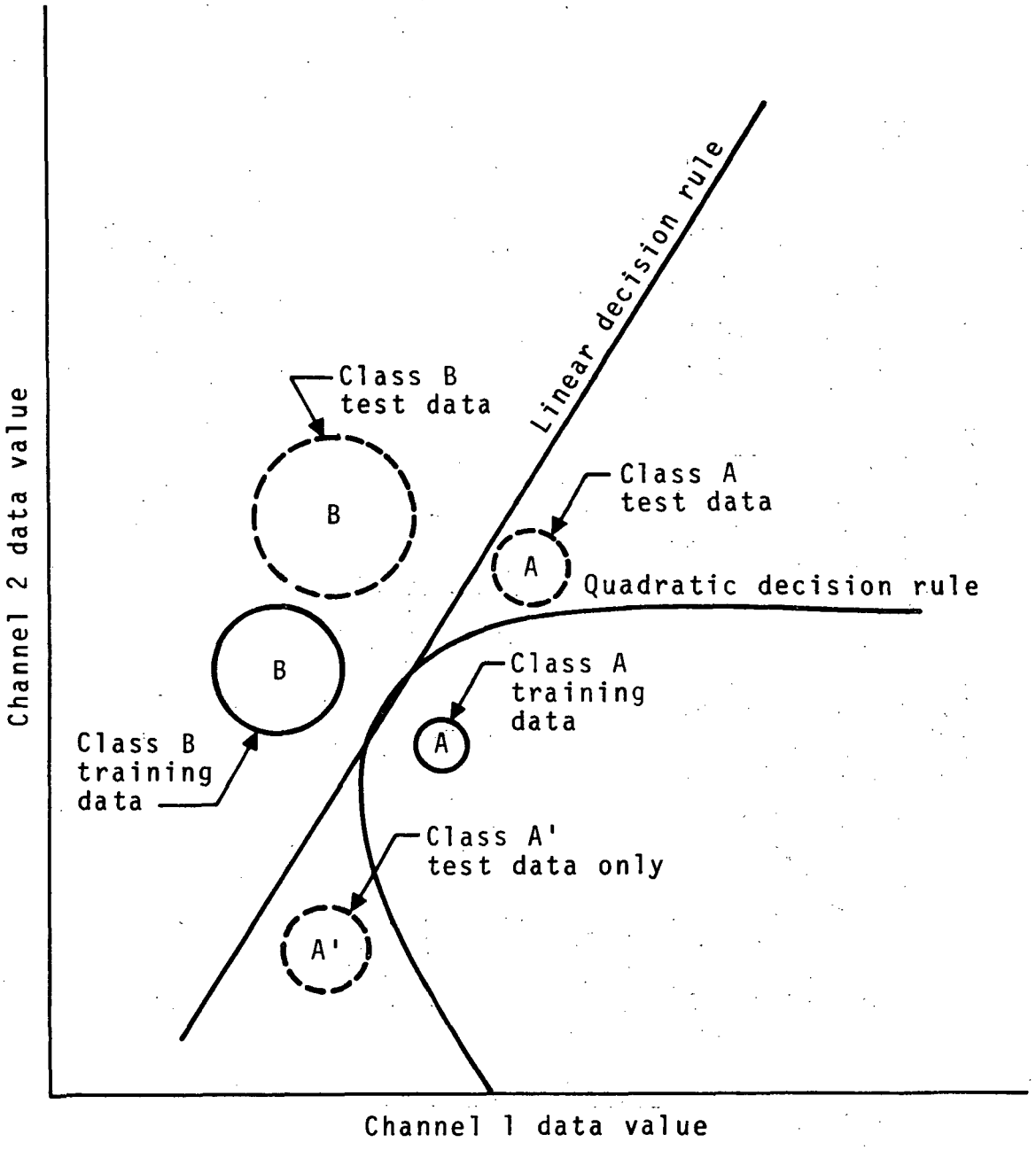


Figure 3.— Illustration of a situation where a linear decision rule would outperform a quadratic decision rule on test data.

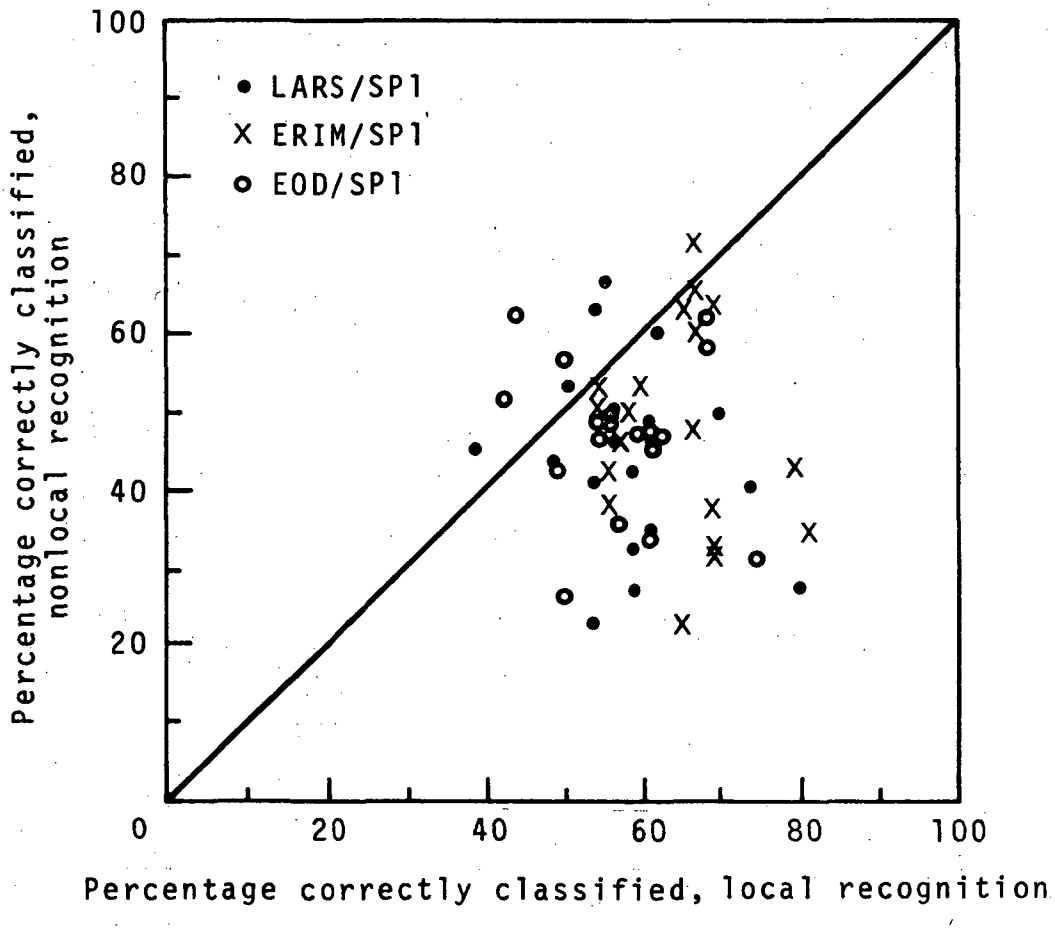


Figure 4.– Comparison of local and nonlocal recognition performance on field-center pixels.

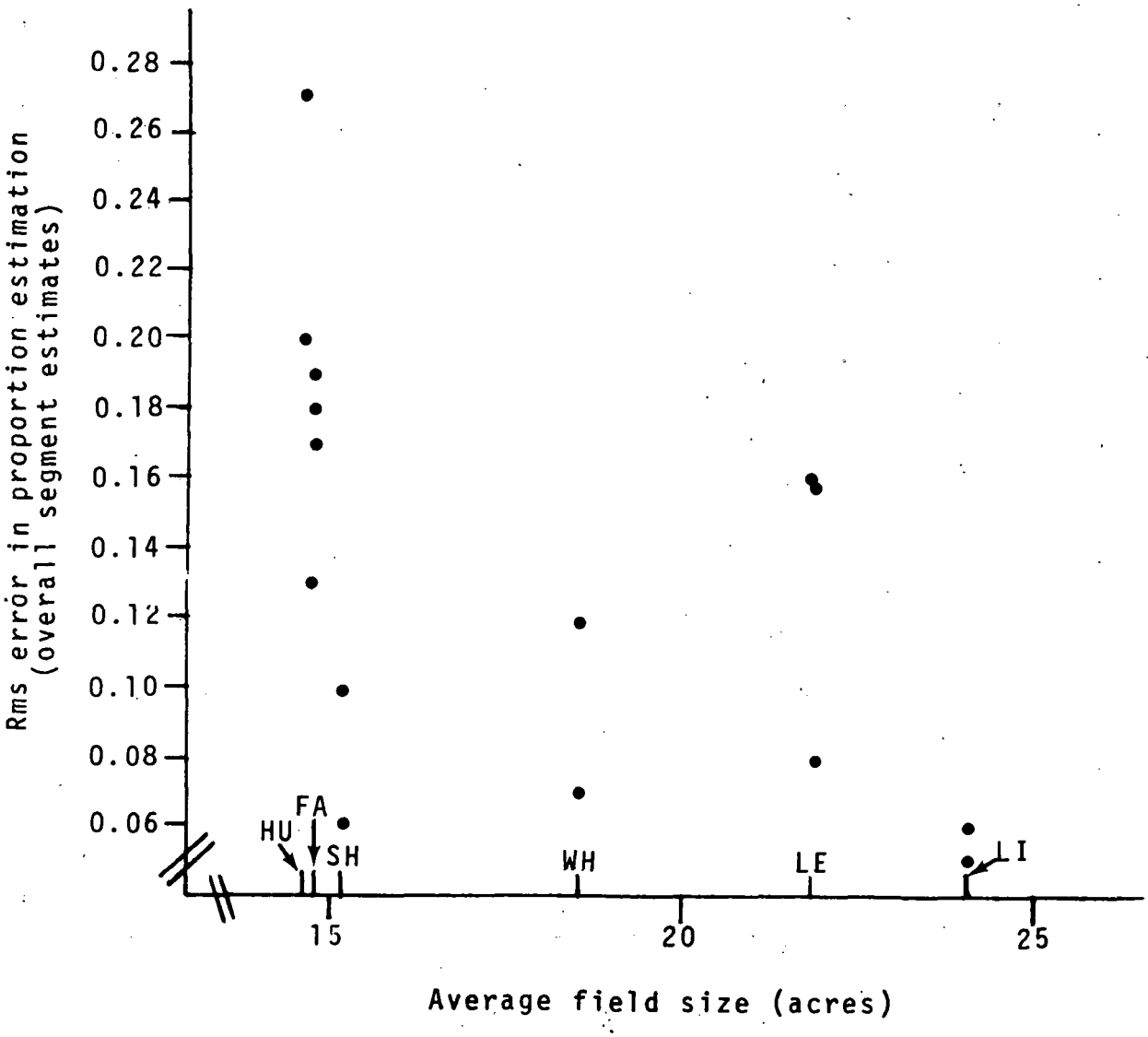


Figure 5.– Correlation of field size and crop identification performance.

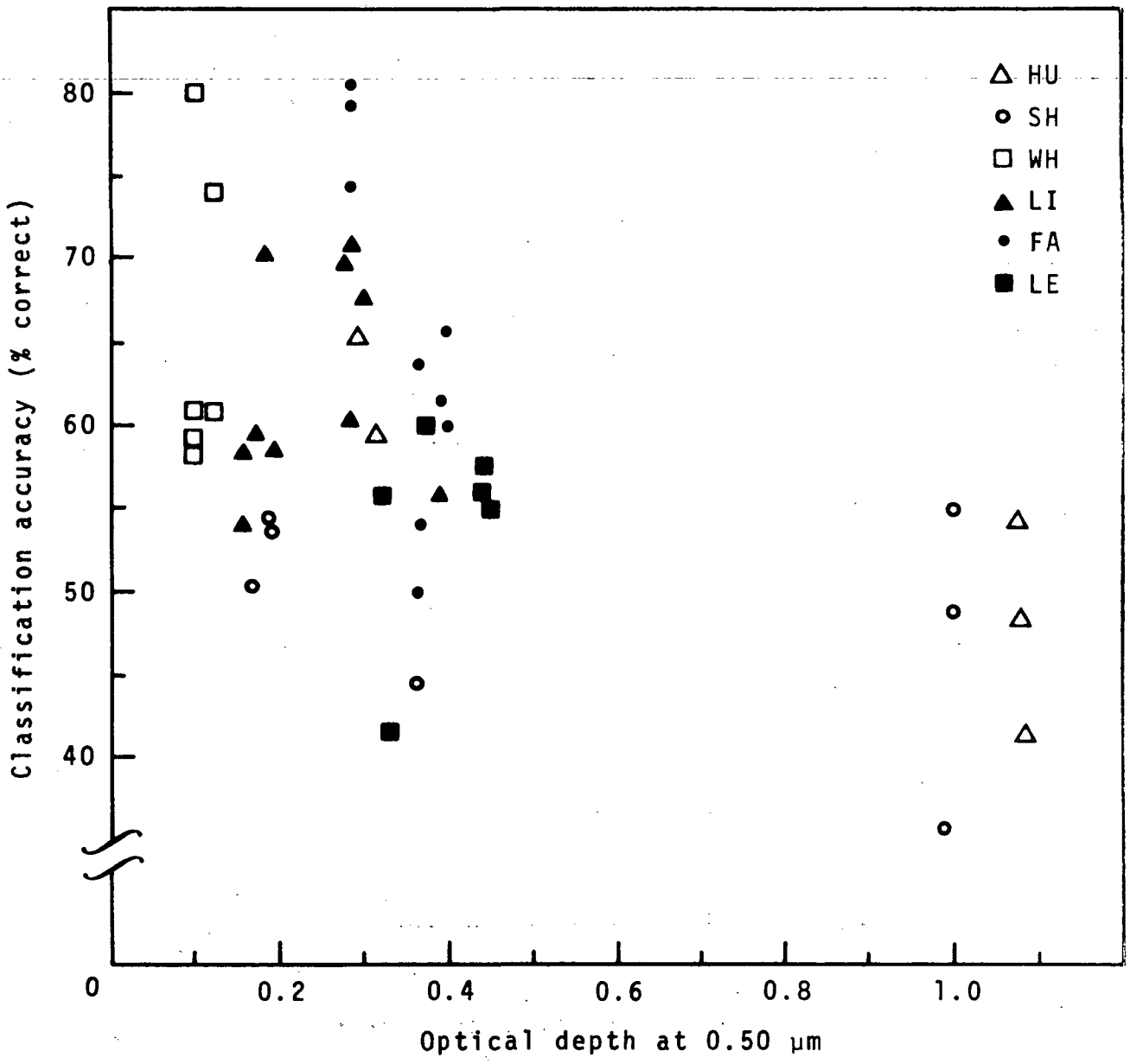
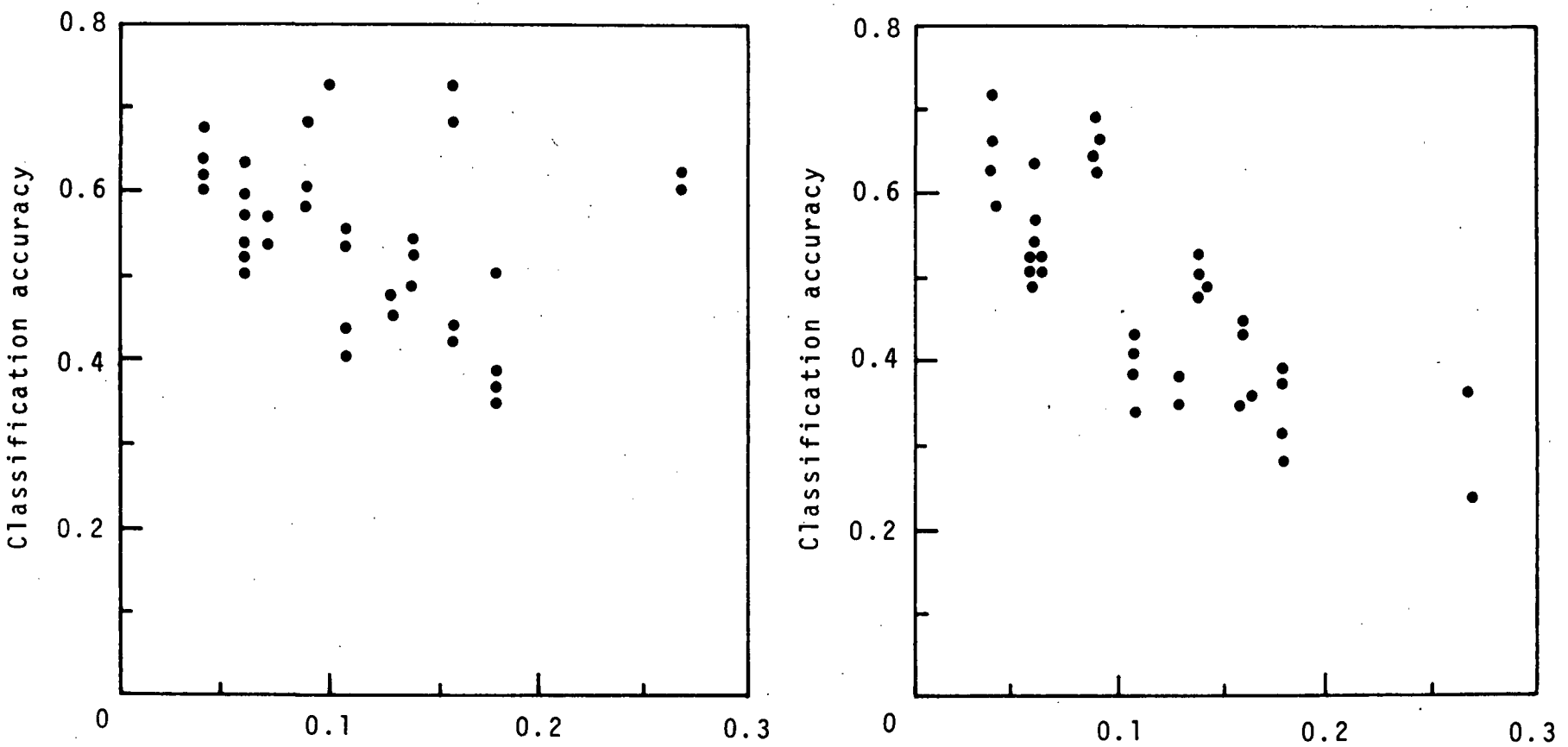


Figure 6.— Graph showing relation of optical depth to local classification accuracy for main procedures.



Difference in optical depth between local and nonlocal segments

(a) With MLA preprocessing. (b) Without preprocessing.

Figure 7.— Comparison of nonlocal classification accuracy as a function of the difference in optical depth between training and classification segments.

5.0 SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

The overall objective of the CITARS was to quantify the CIP resulting from the remote identification of corn, soybeans, and wheat using ADP techniques developed at EOD, ERIM, and LARS. The ADP techniques were evaluated for local and nonlocal recognition. Specific objectives (section 2.0) included performance comparisons to determine if and how CIP's varied (1) with time during the growing season, (2) among different geographic locations, and (3) among the different data analysis techniques. Additional objectives were to determine (4) whether nonlocal signal statistics could be used successfully for crop identification, (5) if the use of radiometric preprocessing could extend training statistics and improve nonlocal performance, and (6) whether the use of multitemporal data could increase the CIP.

To accomplish the objectives, five major tasks had to be completed.

1. Acquisition and preparation of an ERTS-1 data set with sufficient ancillary data to support the experimental objectives and design
2. Computer-aided processing of this data set with selected classification algorithms and procedures
3. Quantification of the CIP's in a manner which would permit quantitative evaluation of the ability of these procedures to satisfy the requirements of agricultural applications
4. Statistical analysis to evaluate quantitatively the impact of major factors known to affect CIP

5. Interpretation of these results (a) to ascertain the underlying physical factors responsible for the results, (b) to draw inferences as to the status of the technology as it relates to agricultural applications, and (c) to make recommendations as to where the technology must be strengthened.

This section briefly summarizes the results of the several key technical tasks which had to be performed to accomplish the scientific objectives. The results are summarized for the six major CITARS objectives, and conclusions and recommendations are presented.

5.1 KEY TECHNICAL ACCOMPLISHMENTS

As discussed in section 3.0, an ERTS-1 data set with supporting ancillary data was acquired and prepared. Except for completeness of satellite coverage (two-thirds of the ERTS-1 scenes were unacceptable because of excessive cloud cover) and insufficient amounts of wheat in some sites, the data set met the requirements of the CITARS experimental design. Assembly of the data set included:

1. Acquisition of crop identification and other agronomic ground-truth data by the ASCS
2. Acquisition and interpretation of color IR aerial photography to extend the field identification data acquired by ASCS to additional sections
3. Registration and geometric correction of multitemporal ERTS-1 MSS data for the test segments
4. Location of field and section coordinates in the ERTS-1 data.

In addition, repeatable, analyst-independent ADP procedures had to be defined and documented, and measures of CIP had to be determined.

Periodic crop observations of fields used to train the classifiers were made by the ASCS throughout the growing season. Photointerpretation of multidate aerial photography was successfully used to increase the size of the data base. The photointerpreted data were used to evaluate ERTS-1 data classification accuracy in field centers. In addition, crop area proportion measurements were made from the aerial photography and used to evaluate proportion estimates derived from the pixel-by-pixel classifications of the ERTS-1 data.

Multiple ERTS-1 passes were registered with an average error of less than one-half pixel. This made multitemporal classifications of the data possible and eliminated the need to locate field and section coordinates in each ERTS-1 scene.

The need to maximize the number of pure pixels selected from the relatively small fields in several of the segments made selection of field coordinates more difficult than expected. Manual methods were found to be inadequate, and a computer-aided method of transforming digitized photomap coordinates to ERTS-1 line and column coordinates was used. The latter method is recommended for use in future projects requiring precise definition of ERTS-1 data coordinates.

A key task prior to the start of ERTS-1 data classifications was to define and document data analysis procedures which were repeatable, easily followed, and yet incorporated

the judgment and skill of experienced analysts. Although it was recognized that restricting analyst decisions might reduce the CIP, variability caused by analyst judgment had to be minimized in order to make meaningful comparisons of results. Limited tests were made using the LARS ADP procedures. These tests indicated that, for the CITARS data set, results produced by the ADP procedures used were comparable to those obtainable using procedures with considerably more analyst interaction.

An important accomplishment of CITARS was the use of quantitative measures of CIP and statistical evaluation of results. The statistical evaluation consisted of analyses of variance and blocked-rank tests for comparisons involving factors such as ADP procedure, location, acquisition date, and use of preprocessing. Two variables, average conditional classification accuracy of pure field-center pixels and the rms error of proportion estimates for entire sections, were used as measures of CIP. Section-to-section variability was used in analyses of variance to determine if differences among the factors were significant. The analyses of variance revealed several significant differences; however, the power of many of the tests was limited because of missing data, the amount of variability present, and the failure of the dependent variable to adequately describe performance of a section independently of the composition of that section. Continued use and development of these tools for remote sensing experiments are recommended.

5.2 RESULTS AND DISCUSSION

The statistical analyses provided a key to the quantitative assessment of remote sensing technology for crop

identification in field centers and for crop area estimation. Previous results have been confirmed in some instances, whereas in others unanticipated results have led to reconsiderations and new insights into certain aspects of the technology. The remainder of this section summarizes the major results and conclusions from the CITARS experiments.

5.2.1 ADP Procedures

Major differences were found in the results for the three principal ADP procedures. The ERIM/SP1 was consistently better in field-center recognition than the LARS/SP1 and EOD/SP1 procedures. However, for whole-area proportion estimation, LARS/SP1 was the most consistent and had the lowest average rms error. Possible reasons for this are (1) the method of training and (2) the decision rule used, both of which factors will be discussed below in more detail. It also was found that all three principal procedures consistently overestimated the proportion of major crops in the segments. Possible reasons for this occurrence are (1) bias in the proportion estimation method and (2) the presence of pixels containing two or more cover classes. These results indicate that field-center recognition of pure pixels, which commonly has been used to evaluate CIP, is not a reliable indicator of proportion estimates for whole areas.

The three main procedures tested differed in two ways. Both LARS/SP1 and EOD/SP1 used a clustering procedure to define training statistics (usually several classes for each major crop) and employed a quadratic decision rule. The ERIM/SP1, on the other hand, formed a single signature

for each major crop and used a variable number of signatures for "other" and a linear decision rule which was optimized on a class-pairwise basis. The differences in performance among the three procedures were attributed to the method of training rather than the decision rule used, since similar results (high ranking for field-center recognition and low ranking for whole-area proportion estimates) were obtained for ERIM/SP2, a quadratic decision rule classifier which used the same signatures as ERIM/SP1. The disparity in rankings was minimized or nonexistent for late August when inherent crop variations were the least and greater numbers of other signatures were selected by ERIM. Thus, it was concluded the major reason for the differences in rankings for the two types of performance is the use by ERIM/SP1 and ERIM/SP2 of single signatures to represent classes having considerable variation.

It can be shown that proportion estimates based on aggregated pixel classifications, as in CITARS, are inherently biased because the expected performance depends on the true proportions present, as well as on the performance matrix of the classifier for individual pixels. It was observed that the rms error in proportion estimation averaged over all procedures was positively correlated with the percentage of "other" in the test sections and negatively correlated with average field size. The latter result is a strong indication that field boundary pixels containing two or more cover classes were a major source of the biased proportion estimates for whole areas. The evidence is that conventional processing techniques using training based on pure field-center pixels cannot be relied on to produce unbiased proportion estimates for whole areas containing mixture pixels.

Another unexpected CITARS result is the lack of improvement of LARS/SP2 (nonequal, major-class, prior probabilities) over LARS/SP1 (equal prior probabilities). Theoretically, apart from boundary pixels, the Bayesian classifier should produce its minimum error rate when correct values for the frequency of occurrence of each spectral class are utilized as parameters in the classification rule. The LARS/SP1 procedure assumed the likelihood that each spectral class would occur equally in the scene. The LARS/SP2 included a procedure for estimating the prior probabilities based on existing agricultural statistics (see section 4.1.2). For CITARS classifications, the LARS/SP2 procedure utilizing unequal prior probabilities did not produce an improvement over LARS/SP1, which assumed them to be equal. This is attributed in part to the fact that the agricultural statistics used were at the county level only (differing by as much as 20 percent from the true proportions) and the test sections were subsets of the county and not randomly located within it. Boundary pixels are another possible cause. The authors do not believe that use of prior probability information in the form of class weights should be discouraged solely on the basis of the CITARS analysis, since it does not constitute a definitive test. Instead, further tests are recommended to determine the sensitivity of the maximum likelihood classifier to class weights.

In other experiments, LARS showed that significant differences in CIP can be obtained with different training sets and that training set size alone does not determine the adequacy of a training set. These results and those discussed earlier point out the dependence of CIP on the development of training statistics.

5.2.2 Nonlocal Recognition and Preprocessing

Comparisons of local and nonlocal performance indicated that average classification accuracy of field-center performance for nonlocal recognition was reduced by 22 percent of that obtained locally. For whole-area proportion estimates, the average rms error of nonlocal classifications was 23 percent greater than for local classifications. Haze level differences between the training and recognition segments were found to be quite well correlated ($r = -0.77$) with degradations in nonlocal classification performance. Other factors, each of which could affect the representativeness of signatures, were regional differences in soil type, agricultural practices, crop maturity, scene composition, training set selection, and MSS scan angle. The results clearly indicate the problems in extending training statistics over space and/or time.

Preprocessing with a relatively simple MLA algorithm (ERIM/PSP1) produced a slight but statistically significant improvement over the three principal procedures in nonlocal field-center recognition; however, no significant improvement in proportion estimates for whole areas was evident. Preprocessing substantially reduced the correlation between field-center performances and differences in haze levels. The inconsistent results with MLA processing are attributed in part to differences in scene composition, for which the technique does not account. Additional research is required to improve on the signature adjustment technique tested and to account better for spectral variability caused by scene composition.

5.2.3 Multitemporal Data Analysis

One segment with several clear ERTS-1 overpasses was analyzed. The use of this multitemporal data (EOD/MSP1) resulted in significant increases in CIP compared to single date classifications. While substantial improvements in performance were obtained for the segment analyzed using basically the same data analysis procedures as for single-date data, new analysis procedures should be researched and developed, taking into account the increased complexity of multitemporal scenes. The use of multitemporal data requires a more complex data processing system (registration, increased data base size, and more complex data analysis procedures), but the increased performance may well justify the added complexity.

5.2.4 Effect of Site and Crop Characteristics

Significant differences in CIP existed among the six test segments. Field size was found to be correlated with proportion estimation performance but not with field-center recognition. The correlation of field size with the accuracy of crop proportion estimates is attributed primarily to the related decrease in the percentage of pixels containing mixtures of crops as field size increased. In addition, it was observed that large fields tended to be more uniform, and areas having larger fields had relatively fewer fields of class "other." Both of these factors contributed to improved performance.

The crop calendar was also an important factor influencing CIP. Field-center performance increased to a peak in late

August when maturity differences among fields of a particular crop (corn or soybeans) were least and the amount of ground cover the greatest. Performance decreased rapidly in September as the crops senesced. Whole-area proportion estimation accuracy was about the same for all time periods except mid-July, when it increased considerably because of greater differences in crop maturity and ground cover.

5.2.5 Relation of Crop and Sensor Characteristics

Two key factors influencing CIP with remote sensor data are (1) the nature of the spectral variation among and within the classes to be identified and (2) the capability of the sensor to measure the spectral variation. An understanding of the relationship of these factors may help explain the levels of CIP obtained in CITARS. In several instances it was found that accurate identification of corn, soybeans, and "other" was not possible even when all the fields analyzed were used to train the classifier. This may have been caused by a lack of differences in the spectral characteristics of the three classes or by the inability of the ERTS-1 MSS to resolve and precisely measure the differences present. The latter is suspected to account for at least a part of the problem. Crop classifications made during the 1971 Corn Blight Watch Experiment [19] using MSS data with more spectral bands, narrower bands, and greater sensitivity and dynamic range showed that these same cover types could be more accurately identified. Additional comparisons of ERTS-1 and aircraft-acquired MSS or other high-spectral-resolution data, such as those available from the current LACIE field measurements project, will be needed to verify this point.

5.3 CONCLUSIONS

The CITARS has provided a quantitative assessment of the 1973-era technology for remote identification of major agricultural crops. The use of quantitative measures of classification performance and statistical evaluations of the results have been important parts of the technology assessment. The major conclusions from the CITARS experiments are:

1. The CIP's for corn and soybeans varied throughout the growing season, with field-center accuracy being maximized in late August.
2. The probability of correct classification of field-center pixels was not well correlated with, and thus is not a reliable indicator of, proportion estimation performance.
3. Proportion estimation accuracy was strongly correlated with both field size and the proportions of major crops in the segment, but field-center classification accuracy was not. Boundary pixels containing two or more cover types were recognized as major contributors to the bias in proportion estimates.
4. The manner in which ground cover classes were selected and used to train the classifier strongly influenced the amount of bias in proportion estimates.
5. Both the probability of correct classification and proportion estimation accuracy were decreased when training statistics for a different location or date were used.
6. An MLA algorithm for first-order adjustments to training statistics used for nonlocal classifications increased the probability of correct classification of field-center pixels but did not improve proportion estimates for whole areas.

7. The use of multitemporal data improved both proportion estimation accuracy and probability of correct classification.

In addition, it has been shown that relatively automatic data analysis procedures can be defined; these procedures can produce repeatable results, are suited for processing relatively large volumes of data, and incorporate (to a large degree) the judgment and expertise of experienced analysts.

5.4 RECOMMENDATIONS

The CITARS provides valuable direction for future research and development of remote sensing technology and guidelines for the design of operational crop production survey systems utilizing remote sensing technology. Recommendations from CITARS include:

1. Continued use and development of quantitative measures of CIP and statistical evaluation of classification results
2. Continued development of improved methods for training classifiers
3. Research and development of methods to improve the accuracy of crop proportion estimates for whole areas
4. Further tests to determine the sensitivity of maximum likelihood classifiers to the use of prior probability information and of linear classifiers to different signature sets
5. Additional research, development, and testing of two complementary approaches to nonlocal recognition: (a) more

sophisticated preprocessing algorithms and (b) stratification of areas based on their similarity with respect to agricultural factors

6. Development of data analysis procedures which account for the increased complexity of multitemporal data and reap the benefits of the potentially greater information content afforded by multitemporal data

7. Additional comparisons of ERTS-1 and other multispectral data sources to determine the adequacy of the ERTS-1 MSS in terms of the number, width, and placement of its spectral bands, the signal-to-noise ratio, and its sensitivity, dynamic range, and spatial resolution.

6.0 REFERENCES

1. Hall, F. G.; Bauer, M. E.; and Malila, W. A.: First Results from the Crop Identification Technology Assessment for Remote Sensing (CITARS). Proc. Ninth Int. Symp. on Remote Sensing of Environment (Ann Arbor, Mich.), 1974, pp. 1171-1191.

2. NASA/JSC: Crop Identification Technology Assessment for Remote Sensing (CITARS). 1975.

   Vol. I Task Design Plan. JSC-09384.

   Vol. II Ground-Truth Field Observations of the Agricultural Stabilization and Conservation Service. JSC-09385:

   Part 1 Huntington County, Indiana
   Part 2 Shelby County, Indiana
   Part 3 White County, Indiana
   Part 4 Livingston County, Illinois
   Part 5 Fayette County, Illinois
   Part 6 Lee County, Illinois

   Vol. III The First Earth Resources Technology Satellite (ERTS-1) Multispectral Scanner Data Acquisition. JSC-09386.

   Vol. IV Image Analysis. JSC-09387:

   Part 1 Crop Identification Extension
   Part 2 Evaluation (to be published)

   Vol. V The First Earth Resources Technology Satellite (ERTS-1) Data Preparation. JSC-09388.

   Vol. VI Data Processing at the Laboratory for Applications of Remote Sensing, Purdue University, Lafayette, Indiana. JSC-09389.

   Vol. VII Data Processing at the Environmental Research Institute of Michigan, Ann Arbor. JSC-09390.

Vol. VIII Data Processing at the Earth Observations Division, Lyndon B. Johnson Space Center. JSC-09391:

Part 1 Huntington County, Indiana
Part 2 Shelby County, Indiana
Part 3 White County, Indiana
Part 4 Livingston County, Illinois
Part 5 Fayette County, Illinois
Part 6 Lee County, Illinois
Part 7 Multitemporal Analysis of Fayette County, Illinois

Vol. IX Statistical Analysis of Results. JSC-09392.

Vol. X Interpretation of Results. JSC-09393.

3. Anuta, P. E.: Geometric Correction of ERTS-1 Digital Multispectral Scanner Data. LARS Information Note 103073, Purdue University (W. Lafayette, Ind.), 1973.

4. Anuta, P. E.: Spatial Registration of Multispectral and Multitemporal Digital Imagery Using Fast Fourier Transform Techniques. IEEE Trans. on Geosci. Electron., 8, 1970, pp. 353-368.

5. Malila, W. A.; Hieber, R. H.; and McCleer, A. P.: Correlation of ERTS MSS Data and Earth Coordinate Systems. Proc. Conf. on Machine Processing of Remotely Sensed Data, Purdue University (W. Lafayette, Ind.), 1973, pp. 2b, 1-14.

6. Swain, P. H.: Pattern Recognition: A Basis for Remote Sensing Data Analysis. LARS Information Note 111572, Purdue University (W. Lafayette, Ind.), 1972.

7. Phillips, T. L., ed.: LARSYS Version 3 User's Manual. LARS, Purdue University (W. Lafayette, Ind.), 1973.

8. Davis, B. J.; and Swain, P. H.: An Automated and Repeatable Data Analysis Procedure for Remote Sensing Applications. Proc. Ninth Int. Symp. on Remote Sensing of Environment. 1974, pp. 771-774.

9. Crane, R. B.; and Richardson, W.: Rapid Processing of Multispectral Scanner Data Using Linear Techniques. Conf. on Earth Resources Observation and Information Analysis System (Tullahoma, Tenn.). Report 31650-107-S, Mar. 1972.

10. Kriegler, F. W.; Malila, W. A.; Nalepka, R. F.; and Richardson, W.: Preprocessing Transformations and Their Effects on Multispectral Recognition. Proc. Sixth Int. Symp. on Remote Sensing of Environment (Ann Arbor, Mich.), 1969.

11. Minter, R. T.: Computer Program Documentation ISOCLS Iterative Self-Organizing Clustering Program, Program C094. Prepared by Lockheed Electronics Company, Inc., under contract NAS 9-12200, 1972.

12. Kriegler, F.; Marshall, R.; Lampert, S.; Gordon, M.; Connell, C.; and Kistler, R.: Multivariate Interactive Digital Analysis System (MIDAS): A New Fast Multispectral Recognition System. Proc. Conf. on Machine Processing of Remotely Sensed Data, Purdue University (W. Lafayette, Ind.), Oct. 1973, pp. 4B, 51-64.

13. Snedecor, G. W.; and Cochran, W. G.: Statistical Methods. Sixth ed. The Iowa State University Press (Ames), 1967.

14. Puri, M. L.; and Sen, P. K.: Nonparametric Methods in Multivariate Analysis. John Wiley & Sons (N.Y.), 1971.

15. Eppler, W. G.: An Improved Version of the Table Look-Up Algorithm for Pattern Recognition. Proc. Ninth Int. Symp. on Remote Sensing of Environment, Environmental Research Institute of Michigan (Ann Arbor), Apr. 1974, pp. 793-812.

16. Kan, E. P.: Effects of Atmospheric Correction on Classification. LEC report TM642-620 (Houston), 1972.

17. Malila, W. A.; and Nalepka, R. F.: Advanced Processing and Information Extraction Techniques Applied to ERTS-1 MSS Data. Proc. Third ERTS Symp. NASA SP-351, 1973, pp. 1743-1772.

18. Henderson, R. G.: Signature Extension Using the MASC Algorithm. Proc. Symp. on Machine Processing of Remotely Sensed Data, Purdue University (W. Lafayette, Ind.), June 1975.

19. NASA/JSC: Corn Blight Watch Experiment. Final Report, vol. III, Experiment Results, 1974.

20. Baker, K.: A User's Guide to the University of Houston Linear Feature Selection Program (Program 0010). NASA working document, 1973.[1]

21. Potter, J.; and Shelton, M.: Effect of Atmospheric Haze and Sun Angle on Automatic Classification of ERTS-1 Data. Proc. Ninth Int. Symp. on Remote Sensing of Environment, 1974, pp. 865-874.[1]

[1]This reference was not called out in text but is included for the purpose of providing a complete list of all CITARS research material.