JSC 09390

CROP IDENTIFICATION TECHNOLOGY ASSESSMENT FOR REMOTE SENSING (CITARS)

VOLUME VII

DATA PROCESSING AT THE ENVIRONMENTAL RESEARCH INSTITUTE OF MICHIGAN, ANN ARBOR

National Aeronautics and Space Administration

LYNDON B. JOHNSON SPACE CENTER

Houston, Texas

June 1975

JSC 09390

CROP IDENTIFICATION TECHNOLOGY ASSESSMENT FOR REMOTE SENSING (CITARS)

VOLUME VII

DATA PROCESSING AT THE ENVIRONMENTAL RESEARCH INSTITUTE OF MICHIGAN, ANN ARBOR

PREPARED BY

William A. Malila
W. A. Malila

Daniel Rice
D. P. Rice

Richard C. Cicone
R. C. Cicone

APPROVED BY

Robert M. Bizzell
Robert M. Bizzell, Project Manager

Andrew E. Potter
Andrew E. Potter, Chief
Research, Test, and Evaluation Branch

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
LYNDON B. JOHNSON SPACE CENTER
HOUSTON, TEXAS

June 1975

TECHNICAL REPORT INDEX/ABSTRACT

1. TITLE AND SUBTITLE OF DOCUMENT
Crop Identification Technology Assessment for Remote Sensing (CITARS), Volume VII: Data Processing at the Environmental Research Institute of Michigan, Ann Arbor

2. JSC NO.
JSC-09390

3. CONTRACTOR/ORGANIZATION NAME
Environmental Research Institute of Michigan

4. CONTRACT OR GRANT NO.
NAS9-14123

5. CONTRACTOR/ORIGINATOR DOCUMENT NO.
109600-12-F

6. PUBLICATION DATE (THIS ISSUE)
June 1975

7. SECURITY CLASSIFICATION
Unclassified

8. OPR (OFFICE OF PRIMARY RESPONSIBILITY)
Andrew E. Potter

9. LIMITATIONS
GOVERNMENT HAS UNLIMITED RIGHTS [X] YES [ ] NO
IF NO, STATE LIMITATIONS AND AUTHORITY

10. AUTHOR(S)
W. A. Malila
D. P. Rice
R. C. Cicone

11. DOCUMENT CONTRACT REFERENCES
WORK BREAKDOWN STRUCTURE NO.
CONTRACT EXHIBIT NO.
DRL NO. AND REVISION
DRL LINE ITEM NO.

12. HARDWARE CONFIGURATION
SYSTEM
SUBSYSTEM
MAJOR EQUIPMENT GROUP

13. ABSTRACT

This document describes the recognition processing and results at the Environmental Research Institute of Michigan for the Crop Identification Technology Assessment for Remote Sensing. The objectives were to define recognition processing procedures eliminating analyst judgment, determine the abilities of the procedures to recognize major crops, and investigate the effects of several factors on recognition processing.

Prescribed data processing procedures were used. The procedures used linear and quadratic decision rules, including a preprocessing transformation for signature extension to nonlocal recognition segments. The analyses of the prescribed output results and supplementary processing are described.

The results of the Crop Identification Technology Assessment for Remote Sensing will be applied extensively in the Large Area Crop Inventory Experiment.

14. SUBJECT TERMS

Data processing	Multispectral band scanners	
Earth Resources Technology Satellites	Recognition	
Farm crops	Remote sensors	

Page Intentionally Left Blank

PREFACE

Volume VII reports the processing and analysis by the Environmental Research Institute of Michigan for the Crop Identification Technology Assessment for Remote Sensing. The institute participated in most aspects of the assessment from planning to production and analysis of results.

This report presents both the recognition processing results obtained and analysis of the results at the Environmental Research Institute of Michigan, P.O. Box 618, Ann Arbor, Michigan 48107. The report also includes a summary of recognition procedures employed at the institute and descriptions of other participation of the institute in the Crop Identification Technology Assessment for Remote Sensing.

Pages 1 through 124 and A-1 through A-6 are the text of the report as prepared by the Environmental Research Institute of Michigan. Only minor changes were made to match the format and style of the other volumes of this series.

For convenience, the authors frequently used nonmetric units of measure used by the Agricultural Stabilization and Conservation Service of the U.S. Department of Agriculture.

Page Intentionally Left Blank

GLOSSARY

Acre — unit of measure equaling 4,046 meters^2.

ANOVA — analysis of variance.

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

CIP — crop identification performance, the quantitative assessment of crop inventories in specified areas using remote sensing, photointerpretation, and automatic data processing.

CITARS — Crop Identification Technology Assessment for Remote Sensing.

Diff — difference.

EOD — Earth Observations Division of the Lyndon B. Johnson Space Center, National Aeronautics and Space Administration, Houston, Texas.

ERIM — Environmental Research Institute of Michigan, Ann Arbor.

ERIM-PSP1 — nonlocal recognition at ERIM with the linear decision rule and preprocessing.

ERIM-PSP4 — nonlocal recognition at ERIM with the quadratic decision rule and preprocessing.

ERIM-SP1 — nonlocal recognition at ERIM with the linear decision rule without preprocessing.

ERIM-SP2 — local recognition at ERIM with the quadratic decision rule.

ERTS-1 — first Earth Resources Technology Satellite, which orbits the Earth 14 times daily in a circular, Sun-synchronous, near-polar orbit at a 915-kilometer altitude. The satellite views the same Earth scene every 18 days. The ERTS-1 was renamed LANDSAT-1 in January 1975.

FAY — Fayette County, Illinois, segment. It is sometimes also abbreviated Fay in the tables in this volume.

Field — spatial sample of digital data of a known ground feature selected by a CITARS researcher.

G.T. — ground truth proportion in a segment.

Ground truth — ground observations by the ASCS of selected 0.4-kilometer (0.25-mile) sections in each of the six selected counties in Indiana and Illinois.

HUN — Huntington County, Indiana, segment.

Inch — unit of measure equaling 2.54 centimeters.

JSC — Lyndon B. Johnson Space Center, National Aeronautics and Space Administration, Houston, Texas.

LARS — Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana.

LIV — Livingston County, Indiana, segment.

Local recognition — the classification of ERTS-1 CITARS data according to crops using statistics from the same data set as the data classified.

MASC — Multiplicative and Additive Signature Correction algorithm for signature extension.

MIDAS — Multivariate Interactive Digital Analysis System.

Mile — unit of measure equaling 1.609 kilometers.

MLA — mean level adjustment, a technique for signature extension.

MSS — multispectral scanner.

NA — no adjustment.

Nonlocal recognition — the classification of ERTS-1 CITARS data according to crops using statistics from another data set from a different segment in the same period as the data classified.

Other — the recognition class of CITARS data which includes all ground features except the major crops, corn and soybeans, for all periods except the first. For the first period, wheat is the only major crop.

PAST — pasture. It is sometimes also abbreviated PASTUR in the tables in this volume.

PI — photointerpretation.

Pixel — picture element, one instantaneous field of view recorded by the ERTS-1 MSS. One ERTS-1 pixel covers about 4,400 meters^2 (1.09 acres). One frame has about 7.36×10^6 pixels, each described by four radiance values.

Quarter section — one-quarter of a section selected for ASCS ground truth.

RMS — root mean square.

RMS Dev. — RMS deviation.

Section — 2.6-kilometer^2 (1-mile^2) township and range section in one of the six selected county segments in Indiana and Illinois.

Segment — 256-kilometer^2 (100-mile^2) area measuring 8 by 32 kilometers (5 by 20 miles) selected in each of the six CITARS counties in Indiana and Illinois.

SHE — Shelby County, Indiana, segment.

Signature — color, tone, brightness, texture, and pattern of a field or crop as it appears on remotely sensed data.

Signature extension — the transformation of recognition signatures obtained from one segment for use in recognition on another segment to minimize differences in data caused by atmospheric or other observational differences between segments.

SR&T — supporting research and technology.

Threshold — boundary in spectral space beyond which a pixel has such a low probability of inclusion in a given class that the pixel is excluded from the class.

Test — type of CITARS data used to evaluate CIP.

TR — trailer.

Training — type of CITARS data from which the spectral characteristics are computed for use in supervised multispectral classification of ERTS-1 data. Training field statistics form the input to maximum likelihood computations for establishing decision boundaries to discriminate between test samples.

UT — untransformed.

WDS — woods.

WHI — White County, Indiana, segment.

Page Intentionally Left Blank

CONTENTS

Section | Page

Page Intentionally Left Blank

TABLES

FIGURES

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

This report describes the recognition processing effort that was carried out at ERIM on CITARS data and presents and discusses the overall results obtained. CITARS was a joint research task for crop identification technology assessment for remote sensing, in which ERIM was one of three organizations performing recognition processing on ERTS-1 multispectral scanner data collected over Illinois and Indiana during 1973. Analysis of variance techniques are being employed to analyze the more detailed results generated at ERIM, along with similar results generated by the other two organizations, and will be documented in a joint final volume of CITARS reports.

The CITARS objectives were to define early in 1973 recognition processing procedures for ERTS data that eliminated analyst judgment as much as possible, determine their respective abilities to recognize the major crops of the area, and to investigate the effects of several experiment factors on recognition performance. Among these factors were time of year, geographic location, field-center versus full-section (whole area) recognition, recognition in areas other than those used for training, signature extension techniques for improving non-local recognition, and linear versus quadratic decision rules.

The processing of CITARS ERTS data was carried out with the pre-defined procedures. The results were good in some instances and less satisfactory in others. Upon completion of the prescribed processing some supplementary analyses were performed using processing techniques that had been developed subsequent to the definition of the CITARS procedures and were based on greater experience with ERTS data. Intensive analyses of the data base for one of the six CITARS segments also were performed. These supplementary studies indicated that the results reported for the standard CITARS processing mode do not measure up to the full potential for recognition processing on the CITARS data sets.

One of the major problems and delays in CITARS data processing was encountered before recognition processing was begun. It was the problem of accurately locating oneself in the ERTS data and defining field-center pixels for training and testing. Field-center pixels were required so that the problems introduced by pixels along boundaries which contained signals from two or more different materials could be studied separately from field-center recognition of "pure" pixels of the various crop and background covers. The problem was compounded by the use of spatial registration techniques to transfer field coordinates from date to date for each segment. Although not a prime responsibility of ERIM, we eventually employed our computer-assisted procedures to locate section corners in the ERTS data and check the accuracy of coordinates determined manually elsewhere.

The major crops of prime interest were corn and soybeans. Performance in recognizing them varied throughout the growing season as the crops matured. Our best single time for recognizing corn, soybeans, and other field-center pixels was late August when an 80% correct local recognition rate was achieved with ERIM procedures; this value was increased several percent by corrections made to the data base after detailed comparisons with ground truth data in a supplementary analysis effort. This conclusion regarding optimum time, and others that follow, must be qualified because suitable cloud-free ERTS-1 data were not obtained for more than two ERTS cycles for any one of the six segments analyzed during the main part of the growing season. Consequently, variability in results between segments is confounded with time factors. In a supplementary analysis effort, multitemporal data for the Fayette segment were shown to improve field-center recognition accuracies above single-time levels.

Crop proportion estimates in full-section data generally were biassed in favor of the major crops. The presence of mixtures of two or more ground covers in individual resolution elements increased such errors. For example in August, mixtures of trees and other covers were frequently mis-recognized as corn. Indications were that the problem was worst for segments with the smallest average field sizes. Another factor in the bias was our fixing of parameters, early in the processing effort, primarily on the basis of field-center analyses. Better balanced parameters should be established in the future. Supplementary processing efforts also were made in one segment for full-section recognition. With multitemporal data, improved proportion estimates with lower variance were obtained. On a single-time (August) segment, the application of a new nine-point mixtures estimation algorithm produced substantial improvement (minimal bias and variance) over the standard CITARS result.

Non-local recognition with unadjusted signatures produced a substantial average reduction in recognition performance from levels attained with local signatures. Signature extension preprocessing by mean level adjustment improved recognition performance on the average in both field centers and full sections. In one instance where the mean-level-adjustment procedure only partially reduced the severe degradation from local recognition performance, a supplementary analysis effort with a more sophisticated signature extension algorithm, also developed at ERIM, was successful in matching local recognition performance. Further development and testing of signature extension techniques is recommended.

The use of both a linear decision rule and the more conventional quadratic decision rule by one organization with identical signatures and test data provided a good opportunity to compare results. Processing costs with the ERIM "best linear" rule were about one-third those of the quadratic rule for CITARS processing. In performance, it tended to be slightly better than the quadratic rule, on the average, but the variability in performance probably is too great to say that the difference is significant. Nevertheless, the equal or better performance at a third the cost is a distinct advantage for the linear decision rule.

There was substantial variability in the amounts and types of pixels available for training in the various segments. In general, too few "other" pixels were available. Corn and soybean pixels usually were abundant, but the greater-than-expected variability in soybean maturities was not always adequately represented in the training data. Also, in some instances, our single signature for each major crop might better have been multimodal for soybeans. Our procedure for establishing "other" signatures should be revised because it is too dependent on ground truth identifications and the specific mix of other-crop pixels that happen to be available from training fields.

Since indications from the processing results were that training data often did not adequately represent the test data, we conducted a supplementary analysis of one data set. In one instance, we used a different set of fields for training, i.e., fields from half of the former test sections. In others, we trained on all field-center pixels to estimate the maximum performance achievable with this data set. In the latter instances, over 90% correct field-center recognition was obtained for single-time data and 94% correct with seven-channel multitemporal data from four time periods. We recommend that studies be made to establish procedures and criteria for determining how much and what type training data are required for establishing signatures that are representative of test data.

We carried out an intensive comparison of the cover-type designations of field center pixels in the Fayette 21 August segment with ground truth data and found a number of discrepancies which, when corrected, improved field-center recognition performance by several percentage points. Similarly in July for the same segment, we found that quite a few "soybean" fields were recognized by the "other" subclass "bare soil". In fact, a number of soybean fields were immature and recently planted in mid July. Field-center recognition accuracy for soybeans increased by 10% when these bare-soil recognitions were considered to be correct soybean recognitions. Another of our concerns was that certain ground-cover categories that are inherently mixtures of two or more ground cover types be eliminated from the field-center analyses and be considered only in the full-section analyses which address the mixture and boundary problems.

Wheat was the major crop of interest in early June. Only two segments had as much as 7% or 8% wheat planted. In these two, there were insufficient training and test pixels available to make a reliable assessment of wheat recognition capability. Furthermore, the validity of some of the few wheat field identifications transmitted to the data analysts is in question.

A general comment about the training procedures used is that they were made "analyst-independent" at the expense of optimality, i.e., without first having sufficient time and effort to develop optimum procedures for ERTS data. Nevertheless, the consistent use of prescribed procedures was of benefit to the analysis of results, even if they are not optimum.

Regarding the procedures used to obtain ground truth, the randomization of choice of ground areas for periodic ASCS visits was made independent of field size. Consequently, many fields for which extensive ground truth was available were too small to extract ERTS field-center pixels for training. We recommend that field size be a factor in the choice of fields for ground visits in the future.

2
INTRODUCTION

CITARS denotes a joint research task for crop identification technology assessment for remote sensing [1,2]. Participants were the Earth Observations Division (EOD) of the Johnson Space Center (JSC), the Environmental Research Institute of Michigan (ERIM), the Laboratory for the Applications of Remote Sensing of Purdue University (LARS), and the Agricultural Stabilization and Conservation Service (ASCS) of the U. S. Department of Agriculture.

2.1 OBJECTIVES OF CITARS

The major objective of CITARS was to quantify the crop identification performance achievable by multispectral recognition processing techniques operating on ERTS-1 data. Techniques developed and/or implemented at EOD, ERIM, and LARS, were applied in parallel to data sets collected throughout the 1973 growing season (June-September) from segments in six counties in Indiana and Illinois. The major crops of interest were corn and soybeans for all but the early June time period. For early June, the major crop of interest was wheat, but it was found in appreciable, though still small ($\leq$8%), proportions in only two of the six segments.

Among the questions the CITARS task was designed to answer are the following:

(a) How do corn, soybeans, and wheat identifications vary with time during the growing season?

(b) Does crop identification performance (CIP) vary among different geographic locations (which may have different soils, weather, management practices, crop distributions, and field sizes)?

(c) Is there a difference between recognition performance in field centers and in full sections which include boundary elements?

(d) Can statistics acquired from one time or location be used to adequately identify crops at other locations and/or times?

(e) Can a signature extension technique be used to improve CIP in non-local areas above those obtained in (d)?

(f) How much variation in CIP is observed among different data analysis techniques? For example, are there differences in performance and cost between linear and quadratic decision rules?

2.2 DESCRIPTION OF THE DATA

ERTS-1 multispectral scanner (MSS) data were obtained on each 18-day cycle of satellite passes over the six 5x20-mile segments. The segments were situated in overlap zones between passes so the coverage potentially was available on two successive days of each cycle for each segment. The data collection window spanned five days in each cycle because of the East-West dispersion of the segments. Two pairs of segments were situated so both members of a pair were covered on the same day.

Cloudy conditions over the test segments eliminated many of the data sets potentially available for processing and analysis [3]. A total of 18 segments were recognition processed at ERIM using local training data and 24 using non-local training data. "Local" training data are data from the same segment and same day as the test data and the recognition operation is termed "local recognition". When the training data are from a different segment or a different day than the test data, a "non-local recognition" combination exists. The 24 non-local recognitions were selected by the CITARS participants from the greater number of available combinations, in order to hopefully satisfy desired analyses of variance without undue expenditures of resources.

Ground "truth" in the form of field observations was collected by ASCS every 18 days, coincident with the ERTS passes, in 20 quarter-sections in each segment. ERTS data from these quarter sections were used in training for recognition processing.

Aerial photography also was collected over each segment several times throughout the growing season. Photointerpretation was carried out at EOD to provide crop identification information in 20 test sections in each segment. These photo-identified fields were used in evaluating recognition processing performance.

2.3 WORK PRIOR TO RECOGNITION PROCESSING

There was a considerable amount of work required in preparation for the recognition processing. After designing the experiment and planning the data collection, in early 1973, procedures were pre-defined for use in processing at each of the three organizations. Since all three organizations were to use a common data base for training and testing, there were certain tasks that were assigned to a single organization to conserve resources. For example, EOD provided photomaps, photointerpretation, ground-truth overlays, and collated and disseminated ASCS ground observations. LARS was responsible for editing and preparing the ERTS-1 data for analysis and spatially registering data from the successive passes over each segment; they also were assigned responsibility for locating and specifying coordinates of field-center pixels for training and testing. ERIM performed some ERTS data quality analyses.

2.3.1 DEFINITION OF PROCESSING PROCEDURES

Major requirements of CITARS procedures for processing ERTS-1 MSS data were that they be made "analyst independent" as much as possible and that the pre-defined procedures be adhered to for all prescribed recognition processing. The timing of these requirements precluded the development of optimal procedures for several reasons. The judgments of an experienced analyst were an integral part of our normal procedures, as was the common approach in the remote sensing community. ERTS-1 had been launched less than a year earlier and we at ERIM had not had at that time a great deal of experience in processing ERTS data, especially over extended regions. There are some differences in the problems associated with the two modes of data collection. Finally, the time available for specifying the procedures was short.

Although the procedures developed should not be considered optimal, their standardization removed one possible source of variation from analyses of the results. Furthermore, there was a substantial benefit to ERIM researchers in having to consider the problem of removing analyst-dependent aspects from processing procedures. This experience helped the continued development of existing and new processing techniques during the time CITARS processing and analysis was taking place with the pre-defined procedures.

2.3.2 SPECIFICATION OF FIELD-CENTER PIXELS

A major problem and source of delay encountered in preparations for CITARS recognition processing was that of identifying and specifying and verifying field-center pixels for training and testing. Field-center pixels were desired so that boundary pixels which contain mixtures of two or more materials would not be used for training or in the field-center phase of the evaluation of recognition performance; the use of spatial registration by a nearest neighbor algorithm for transferring coordinates from one date to the next forced the use of more stringent criteria for field-center pixels than might otherwise have been employed. This task was carried out at LARS but eventually had to be augmented by computer-assisted procedures, developed and applied at ERIM, which substantially reduced human errors.

Two computer-assisted procedures were used. First, special digital line-printer maps were made on which crop-type symbols were printed in all locations where field-center pixels were indicated on field coordinate cards; some of these maps even were color coded to distinguish between training and test pixels. When used in conjunction with the ground truth overlays which were of approximately the same scale as the maps, one could quite readily find obvious errors in the coordinate cards. They also were very useful in manual checking of other coordinates which were correct or less obviously in error.

Consistency in the placement of ground-truth overlays was found to be a major problem with the purely manual procedure that was being employed to specify pixels in the test sections and training quarter-sections which were scattered throughout the 100 sq. mi. segments. Therefore, ERIM assumed responsibility for locating section corners and specifying the coordinates of test sections and training quarter sections in all segments. This was accomplished through the use of computer-assisted procedures which had been developed under other ERTS investigations at ERIM [4].

A map transformation from Earth coordinates on a rectified aerial photograph to ERTS data coordinates was calculated for each segment using roughly 30 control points for each calculation. The control points were located visually in the rotated and geometrically corrected ERTS data and by coordinate digitization on the photograph. A map transformation then was computed by the method of least squares; ERTS coordinates of the few control points with large residuals (>1 pixel) were checked and modified or deleted, as appropriate, and the transformation was recomputed. Next, the transformation was applied to all section corners of interest (whose locations on the photograph had been digitized at the same time as the control points) to find their fractional line and column coordinates in the ERTS data. Final standard errors of estimate (for control points) were less than 0.5 and typically between 0.2 and 0.4 ERTS pixels, i.e., 15 to 30 meters on the ground. The RMS error in digitizing the location of the individual points was on the order of three meters on the ground (errors of roughly 0.005 inch or less on a photograph at a scale of 1:24,000).

These section corner coordinates (calculated in fractional ERTS line and column coordinates) then were used in the manual location of field boundaries of individual fields within the sections. A major advantage of the procedure was that it preserved the relative positions of all points considered with an accuracy that could not be matched manually. Another feature of the ERIM procedure was utilized to generate ERTS data coordinates for each outlined section . All pixels whose centers fell inside lines connecting the vertices (again, located by fractional coordinates) were automatically included on coordinate definition cards.

This computer-assisted procedure could readily have been applied to locate coordinates of individual test and training fields as well. With it one also could have expeditiously obtained coordinates in different passes without their spatial registration, which then would have been required only for multitemporal analyses.

2.3.3 DATA QUALITY TESTS AT ERIM

Previous experiments had shown that some ERTS-1 MSS data suffered from occasional noise or other degradations that usually affected complete or substantial portions of scan lines across the image. Often these irregularities occurred at scan-line intervals that were multiples of six, corresponding to one of the six individual detectors used for each spectral band of the MSS.

To look for detector*-related irregularities, we computed histograms and sample statistics for many of the unrotated data sets in groups of scan lines corresponding to each of the six detectors per ERTS band; we also utilized similar statistics computed by LARS for other of the data sets. Tests were

*The irregularities usually did not originate with the actual detector element, but rather with some element of the subsequent signal processing chain.

made of the variability of detector means and detector variances within bands; plots of the ratio of standard deviation to mean were made for individual detectors and results are summarized in Refs. 2 and 5.

Bad lines when detected were examined on line-printer maps of rotated data and compared with the ground-truth overlays to determine which, if any, test or training fields were affected by each bad line. A tabulation of such fields was prepared and forwarded to LARS for use in editing the fields included in the common data base for processing and analysis.

2.4 REMAINDER OF THIS REPORT

This report presents the overall recognition processing results obtained at ERIM. Analyses of these results also are included. Further, some supplementary analyses carried out using techniques and procedures different from those defined for use on CITARS are presented.

Recognition results are presented for two groups of test pixels. The first group was composed of field-center pixels only and, appropriately, formed the basis for "field-center" analyses. The second group, for "full-section" analyses, was composed of all pixels within the 1-sq-mi test sections -- thus, this group includes boundary and other mixture pixels, farmsteads, roads, and, in some cases, urban areas.

The remainder of this report presents a summary of ERIM CITARS procedures, presentations and discussions of training results and test results, and supplementary analyses. Additional insights will be provided by results of the analysis of variance effort which is being conducted on more detailed recognition results (i.e., results on a section-by-section basis) than are reported here and will also compare results from the three organizations. These analyses will be documented in a joint report that will conclude CITARS documentation.

3
SUMMARY OF ERIM CITARS PROCEDURES

A stated goal of the CITARS task was to assess the crop identification capability of that current remote-sensor-data processing technology which could be documented in an unambiguous way so as to eliminate the need for judgment on the part of the data analyst. The techniques assessed in this program were defined prior to the start of data processing and do not include certain advanced techniques which are in various stages of development at ERIM.

The major emphasis of research at ERIM has been on those problems which in our opinion are key to the development of operational remote-sensor survey systems for large areas. These key problems include (1) the throughput rate of recognition processors, (2) the need for extending signatures from training areas to other geographic locations and to other observation conditions, and (3) the misclassifications caused by the relatively large size of the spatial resolution element of data from satellite sensors.

The procedures used on the CITARS project reflect the above concerns. For example, the linear decision rule used [6] reduces the amount of general-purpose digital computer time required for recognition, compared to the more conventional quadratic rule, and has shown comparable accuracies in previous tests.* The use of both decision rules in CITARS, with common sets of signatures and test data, provided another opportunity to compare the rules. The

---

*Other ways of increasing throughput, such as special-purpose computers, also are being explored at ERIM under other contracts, e.g., the development of a hybrid special-purpose/general-purpose digital image processor, MIDAS (Multivariate Interactive Digital Analysis System) [7].

training procedure used produces a minimum number of signatures, which also reduces general-purpose digital computer time.

Preprocessing for signature extension is an important part of the outlined CITARS tasks performed at ERIM, but only the most straight-forward of the several different techniques that have been developed and are under investigation at ERIM was specified for use on the project. The analysis of recognition accuracies over large areas that include field boundaries and non-agricultural materials has been recognized as an important problem; it is being addressed by the ERIM technique for estimating proportions of unresolved objects which was not part of the CITARS processing carried out.

The ERIM procedures for processing CITARS data are outlined below, except for the procedures used for testing data quality; results of data quality tests are included in other reports [2 and 5]. The CITARS Task Design Plan [1] explains all procedures in detail, along with their rationale. Those steps which have been revised or more completely specified since that document was published are explained in detail by Appendix I.

The parenthesized numbers following the steps below refer to the numbering system for steps used by the Task Design Plan document.

3.1 DATA PREPARATION AND TRAINING

Steps 1-11 were performed for all ERIM procedures.

(1) Reformat the ERTS data which had been rotated, geometrically corrected, and/or registered by LARS. (1.1)

(2) Reformat and verify the field-definition coordinates prepared by LARS. (1.3.1)

(3) Extract a signature for each ASCS (training) field, using program STAT on the designated field-center pixels. (1.4.1, 1.5.2*)

(4) Combine signatures of each given crop type to form a single overall signature for each crop type. Use program SIGCOM, which employs an iterative procedure to reject outlier fields from the combination. [The resulting major-crop signatures (corn and soybeans, or wheat) are used in Step (5); the others are saved for possible use at step 9.] (1.4.2, 1.5.3*)

(5) Use program CLASFY, with major-class signatures only, to assign pixels of the data set (according to ERIM's "best linear" rule) to the classes without using a threshold. The output tape will contain recognition results and the likelihood function exponents for each pixel. (1.4.3.1)

(6) Use program HIST to form a histogram for each major crop of the likelihood function exponents of its training pixels that are correctly recognized. (1.4.3.2)

(7) If necessary, expand the covariance matrix of each major crop so as to insure that the 0.001 rejection threshold will be likely to accept at least 99% of the pixels of that major crop (in the absence of competition from other crops). This step is meant to be some protection against possible undesirable effects caused by actual statistics being non-Gaussian. [This step was never found necessary on any data set according to our procedure.] (1.4.3.3*)

---

*Refer to Appendix I for current full description of this step.

(8) Use program CLASFY (linear rule), with only the major-class signature(s) (after step 7), to preliminarily recognize those ASCS (training) field centers not designated as major crops. If step 7 is not required, the CLASFY run of step 5 is used in place of this step. A rejection threshold of 0.001 is applied. Any crop type which has more than 10% of its pixels (and more than 2 pixels) classified as any major crop is considered as a significant "other" crop type. The others are not used. (1.5.1)

(9) Use program POMPOM to compute the probabilities of misrecognition (based on the best linear rule applied to the available signatures) which are needed below to decide whether or not "other" class signatures need to be subdivided into two or more subclasses.

If the probability of misrecognizing a major crop as a given significant other crop is greater by 0.02 for the combined other crop signature (from step 4) than for any individual-field signature of the same other crop type (from step 3), the combination is judged to have introduced excessive extra misrecognition. In this case, the fields of this (other) type are separated (split) into two (or more) groups, each forming a separate combined signature by the method of step 4, so as to minimize the amount of extra misrecognition introduced by the combinations. (1.5.3*)

(10) If the probability of misrecognizing any major crop signature as the other crop signature (after the splitting of step 9) in question is too large (greater than 0.25 or, if the number of other-crop pixels is less than 8, greater than 0.15), then the other crop is not used. (*)

(11) The other signatures remaining after steps 8, 9, and 10, and the major-crop signatures after step 7, are collected to form the final signature set for local recognition.

---

*[Same comment as previous page]

3.2 LOCAL RECOGNITION, LINEAR DECISION RULE (ERIM-SP1)

(12) Run program CLASFY to recognize the data set using the final signatures from step 11.

(13) Run program TALLY on the output from step 12 to gather recognition statistics for each field (field-center pixels only) in the data set and for each section and quarter-section. The statistics extracted are the number of pixels recognized as belonging to each particular signature class and the number of pixels whose exponents are less than the theoretical χ^2 for a 0.001 probability of false rejection.

(14) Run program TOTAL on the TALLY output for the fields. This will combine the statistics for the class "other" and sum the results over each test section individually, over all test sections, and over the training regions. The results are punched on cards according to the standard CITARS format for recognition results.

(15) Run program TOTAL on the TALLY output for the sections and quarter-sections. This will determine and tabulate the number of pixels recognized as each major crop or as "other" in each section, in all sections combined, and in all quarter-sections combined. The results are punched according to the standard CITARS format for recognition results.

3.3 LOCAL RECOGNITION, QUADRATIC DECISION RULE (ERIM-SP2)

(16-19) The steps are identical to steps 12-15 except that QRULE is used in step 12 to implement a quadratic decision rule rather than the linear decision rule of CLASFY.

3.4 NON-LOCAL RECOGNITION, LINEAR DECISION RULE, WITHOUT PREPROCESSING (ERIM-SP1)

(20-23) The steps are identical to steps 12-15 except for the following:

(a) The signature set used (step 12) is from a data set different from the one being processed.

(b) The rejection threshold is 0.0001 (step 13).

3.5 NON-LOCAL RECOGNITION, QUADRATIC DECISION RULE, WITHOUT PREPROCESSING (ERIM-SP2)

(24-27) The steps are identical to steps 20-23 except that QRULE (rather than CLASFY) is used for recognition.

3.6 SIGNATURE EXTENSION PREPROCESSING OF SIGNATURES FOR NON-LOCAL RECOGNITION

Steps 28-29 are performed to obtain preprocessed final signatures for non-local recognition.

(28) Use program STAT to compute the channel mean values for all pixels in training quarter-sections not affected by bad lines or clouds. Do this for the data set supplying the signatures as well as the one to be processed. For each channel, compute the difference between the respective mean values.

(29) For each final signature used in step 20, subtract the mean-value difference in each channel (from step 28) from the corresponding signature mean value. This forms the preprocessed signature set.

3.7 NON-LOCAL RECOGNITION, LINEAR DECISION RULE, WITH PREPROCESSING (ERIM-PSP1)

(30-33) The steps are identical to steps 20-23, except that the preprocessed signature set is used (from step 29).

3.8 NON-LOCAL RECOGNITION, QUADRATIC DECISION RULE, WITH PREPROCESSING (ERIM-PSP4)

(34-37) The steps are identical to steps 20-23, except that the preprocessed signature set is used (from step 29) and that QRULE (rather than CLASFY) is used for recognition.

TABLE 1. SUMMARY OF PLOT AND PIXEL COUNTS

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
TEST																		
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																		
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.

Page intentionally left blank

4
THE TRAINING EFFORT

Discussion of the ERIM training effort centers about three major topics, each presented separately: major-crop signatures, other-crop signatures, and non-local recognition signatures.

4.1 MAJOR-CROP SIGNATURES

Signatures were established for the major crop(s) by using the previously described procedures on training data (in the ASCS-visited quarter sections). The major crops were corn and soybeans for all but the two early June segments for which wheat was of interest.

There are substantial differences in the amount of training data that was available for the various data sets. Table 1 contains the number of individual fields (plots) and total number of pixels available for training in each data set by crop type. For the major crops, corn training data ranged from 6 to 24 fields with 62 to 474 total pixels, while soybeans data ranged from 7 to 21 fields and 35 to 248 pixels. In early June, wheat as a major crop had only 4 and 9 fields with 26 and 48 pixels, respectively, for training in Shelby and Fayette counties.

There also was a disparity in the number of other crops and ground covers represented in the training data. Only three other ground covers were available in two segments, with a maximum of eight in one data set. In numerous instances, however, only a few pixels in one field or plot was available for one of these "other" ground covers.

The numbers of plots and pixels in the test data also are presented in Table 1. A comparison of these with the corresponding numbers for training data shows that there are several instances in which an "other" ground cover is presented in either training or test data but not in both. On the other hand, corn and soybeans are well represented, with the number of test pixels ranging from 157 to 819 in from 28 to 74 fields.

When statistics from individual fields of a major crop were combined and tested for similarity, frequently a field (or two) was rejected by the test and its statistics then were not used in forming the final signature for recognition. Table 2 summarizes the numbers of fields and pixels available for training and used for training, as well as listing those fields that were rejected at this step of the procedure. One or more fields was rejected in the formation of 24 of 33 major crop signatures; however, the number rejected represents only 6% of the total available for training.

The training operations are more easily understood and visualized through an examination of ellipse plots of pairs of data channels (See Fig. 1). Plots were generated for most of the data sets analyzed, and selected plots of field statistics are presented in Fig. 2, parts (A) through (Z). (The sixth column in Table 2 references the part of Fig. 2 associated with each data set.) Each ellipse here represents the signature of an individual field; in other plots each may represent the combined signature of a class.

The mean of the signature is at the center of the ellipse and the shape of the ellipse is determined by both the correlation between the two channels and the relative magnitudes of their variances; for perfect correlation, the ellipse would collapse to a straight line skew to the coordinate axes while, for zero correlation, the axes of the ellipse would be aligned to the coordinate axes. The size of the ellipse is a function of both the variances and the χ^2 level chosen for the display. The χ^2 level is a measure of squared distance from the mean in covariance units, and the ellipses presented in this report are for a constant distance corresponding to $\chi^2 = 1$. For two degrees of freedom, approximately 40% of the points from a normal distribution would lie within the ellipse. However, the other two channels of data used in recognition are not represented on a given ellipse plot, and one should use care in generalizing from two to four dimensions. We have chosen, in most cases, to plot values in channel 4 (ERTS Band 7) vs. channel 2 (ERTS Band 5). These give a quite complete characterization of the signature separability because of the high degree of correlation found between channels 1 and 2 and between channels 3 and 4 (See Fig. 1).

The corn field signatures of Shelby County on 24 September 1973 typify a data set on which the algorithm used to test and combine several field signatures into a single-mode major crop signature works well. Field 19-45 was rejected by the algorithm and was not used in calculating the final combined corn signature. Upon examining the associated ellipse plot (Fig. 2(A)), one finds that field 19-45 (Ellipse #14) is indeed visually distinct from the central cluster of corn field signals as displayed in the plot of channel 4 versus channel 2. An examination of ground truth information indicated that field 19-45 was mislabelled as corn, when in actuality it was alfalfa.

The soybean signature of Lee County on 17 July 1973 (Fig. 2(F)) displays another situation. The rejection threshold used in the CITARS processing was a χ^2 distance of 13.277, corresponding to a 0.01 probability of false rejection under the assumption of normality and four degrees of freedom. Any signature measured to be greater than this distance from the intermediate combined mean and covariance of field signatures was not used in the final calculation of the combined signature. Field 86-30 (Ellipse 32) was rejected, being a χ^2 distance of 14.59 from the overall signature. After examining the ellipses by eye, an analyst would most likely concur with this decision. Field 72-68 (Ellipse 30) was at a χ^2 distance of 12.42 and thereby acceptable according to the algorithm. Yet there may be some doubt in the mind of an analyst concerning this decision because Ellipse 30 (Field 72-68) is visually distinct from the central cluster of ellipses, though less obviously an outlier than Ellipse 32 (Field 86-03) for these two channels. This raises the question as to how appropriate the 13.277 χ^2 threshold may be. Empirical examination of a sampling of CITARS data early in the processing effort aided in the establishment of the 13.277 χ^2 rejection level. This level most often reflected the decision an experienced analyst would make under the condition that only a single-mode signature was acceptable. Situations may arise where this threshold is not adequate yet, since an objective was to identify an automatic procedure, a standard threshold was a prerequisite for CITARS and was used throughout.

The soybean signatures of Livingston County on 16 July 1973 (Fig. 2(N)) offer yet another facet of the procedure. Here, field 89-03 (Ellipse 21) with a χ^2 distance of 13.4715 was rejected from the single-mode representation of soybeans. An analyst may have instead chosen to use two signatures and reject Ellipse 20 as being an outlier from either of what apparently are two clusters of ellipses. Instead, the field of Ellipse 20 was accepted because it was close enough to the combined ellipse of the two clusters.

The appearance of two distinct clusters of signatures in the 16 July Livingston data (Fig. 2(N)) is an indication of the variability of the soybean crop at this time of year. A number of fields were planted or replanted several weeks later than the others and had very low ground cover percentages in mid-July. These fields were spectrally more like bare soil than soybeans. This phenomenon was evident in other counties as well -- for example, Fayette County which is discussed next.

The soybean training field signatures for Fayette, 16 July, are represented in the ellipse plots of Fig. 2(V), channels 4 vs 2. A glance at this figure again reveals the great variability in soybeans at this time of year. Here, the late-planted fields are to the lower right of the centroid of ellipses. Ellipses 20 (Field 22-45) and 23 (Field 64-63) were observed to have 0 to 5% ground cover and the soybean plants were only 3 to 4 inches tall. These fields were more like bare soil than soybeans. Ellipse 22 (Field 55-45) also had 4-inch plants, but its ground cover was listed as 5 to 20%. The majority of the fields had plants 10 to 14 inches tall with ground covers of 20 to 50%. The most mature fields of soybeans were Ellipses 16 (Field 77-08) and 14 (Field 35-13). Both had 80 to 100%

ground cover, were 20 or more inches in height, and were blooming. The algorithm for combining field signatures rejected only the mature field 77-08 (Ellipse 16) in producing the soybean recognition signature. This recognition signature consequently had a large dispersion volume and soybean crop proportions were overestimated in test data, as discussed in the next section.

The soybean fields became much more similar to each other in late August after the late-planted fields had time to catch up with the others. Yet there still was more variability evident for soybeans than for corn. The greater uniformity of corn field signatures was evident at all times of the growing season.

Wheat field signatures in early June were different in the two segments analyzed. Only four training fields were present in the Shelby 8 June segment and they were very similar to each other, producing a compact recognition signature (See Fig. 2(Y)). The nine training fields in the Fayette 10 and 11 June segments exhibited greater variability. One field, 5-48 (Ellipse 11, Fig. 2(R)), was rejected by the field signature combination algorithm. A check of ground truth information showed it to be the only training field still in the "boot" stage of maturity -- an all-green stage as opposed to the yellowing and senescing stages of the other fields. The field represented by Ellipse 16 was accepted, even though it appears to be less mature than the others, so the resulting Fayette wheat signature still was less compact than the Shelby signature.

Having established preliminary major class signatures, a preliminary recognition run was made on major-crop training data to determine whether or not any major-crop signature variances and covariances should be expanded to improve major-crop recognition. It was not found to be necessary to scale any of the major-crop signatures for CITARS processing according to the procedures described in Sec. 3.

As a point of information, the preliminary recognition results obtained with only major-crop signatures are summarized in Table 3 for field-center pixels in nine data sets. These performance numbers represent an upper limit on major-crop recognition for CITARS because the introduction of "other"-class signatures could only detract from the values shown. Not surprisingly, the results for the "other" category are low since there were no other signatures to compete with the major-crop signatures. However, a correlation was noted between extremely low "other" class values here and overestimation of major crop proportions in the final recognition results.

4.2 OTHER-CLASS SIGNATURES

Once the final major-crop signature(s) had been determined, the next task was the selection of appropriate 'other'-class signatures. The procedure defined for CITARS is one that utilizes the ground cover categories found in the training data and empirically determines which are significant by preliminary recognition on training data. The results of the four steps used in selecting other signatures are summarized in Table 4.

In the first step of determining which of the other classes gave significant false alarms to the major classes, anywhere from zero to seven were found, with 1, 2, 1, 4, 1, 4, 4, and 1 segments having 0,...,7 significant other classes, respectively. Whenever a distribution was sufficiently 'far' from the major-crop distribution(s), no false alarms were found. A common example of such a distribution is the water distribution. Figure 3(A) is a plot of combined signature ellipses of all classes for Fayette 16 July 1973 before determination of the final set of signatures. Water (Ellipse 4) is visually far from the other combined crop signatures. Water is so spectrally distinct from corn, soybeans, and wheat that the probability a major-crop field-center pixel will be called water (or vice versa) is very slight. The advantage in deleting such outlying distribution signatures is solely that computing costs are reduced; costs of the more expensive quadratic rule are cut proportionately more than those of the linear rule by a decrease in the number of signatures used. The final results would be no different with or without the signature, because a rejection test is applied to all pixels in recognition and water pixels are far enough from the major crops to be left unclassified, which for the purposes of CITARS is equivalent to being classed as 'other'. The case for closer distributions is not as clear cut. There may be instances where no false alarms are found in training data but where some test data would be close enough to give false alarms that might have been avoided had another signature been used for recognition. This latter type of situation, however, is more a case of non-representative training data than of a fault in the training procedure.

Early June data are especially interesting re the determination of other-class signatures. As seen in Table 4, only one other class (weeds) was significant for Fayette 10 and 11 June, while there was no significant other class for Shelby 8 June. We earlier discussed the compactness of the Shelby wheat signature. This is evident in the plot of combined signatures for all classes in the Shelby 8 June training data (Fig. 3(B)). Although the separation of the means of wheat (Ellipse 9) and oats (Ellipse 7) and fescue (Ellipse 8) are relatively small, the compactness (small size of covariance and, therefore, the ellipse) of wheat results in very low probabilities of misrecognizing fescue or oats as wheat. The converse is also true because of the compactness of the oats and fescue signatures, a compactness resulting largely from the small number of fields and pixels available to establish these signatures. The combined signature for Fayette 11 June wheat, e.g., Ellipse 1 in Fig. 3(C), is less compact than that for Shelby. As discussed earlier in regard to Fig. 2(R), the greater number of wheat fields in Fayette training data contained a greater variety of wheat-field conditions than Shelby and consequently yielded a signature more representative of the field-center test data. Weeds, the only significant other class here, are represented by Ellipse 10, Fig. 3(C).

All pixels from each significant 'other' class were combined to form a single combined signature which then was used for final recognition, unless it was modified in succeeding steps of the procedure. In just over 20% of the cases, splitting (or non-combining) of a specific other class signature was warranted (Step 2). By way of illustrating splitting, consider the situation where two fields that comprise an 'other' class straddle a major-crop signature. When they are combined to form a single signature, a greater percentage of major-crop pixels would be assigned to this 'other' class than would be the case if the two individual signatures were used separately -- thus, splitting would be indicated. Decisions regarding splitting were based on probability of misclassification calculations made using program POMPOM on training signatures, both individual and combined.

Next, (Step 3) a rejection test was applied and just over 40% of the whole and split 'other' signatures were rejected because their presence would yield too high a probability of missing major-crop field-center pixels, again based on probability of misclassification calculations for training signatures. The ellipse plots of Fig. 3(A) can be used to illustrate the problem. Consider the soybean signature (Ellipse 2) for this Fayette 16 July data set. One notices immediately that distributions 8 and 10 (respectively, weeds and stubble) are displayed very near the mean of soybeans and lie wholly within the soybean ellipse. Since these ellipses describe only the two-dimensional situations, we turned to theoretical calculations for quantitative estimates of the various probabilities of misclassification. The probability of misclassifying soybeans as weeds was 35% and as stubble was 20%. Following the procedure described in Sec. 3, both weed and stubble signatures were rejected because these percentages were deemed to be too high (the small number of pixels in the stubble signature also was part of the consideration).

The final step in signature generation was to insure that all variance-covariance matrices were non-singular. The value 0.1 was added to each diagonal term of each singular or ill-conditioned matrix. Fourteen signatures required this procedure (See Table 4). The usual cause of singularity and ill-conditioning was too few data points for use in computing the matrix.

4.3 NON-LOCAL RECOGNITION SIGNATURES

The establishment of signatures for recognition processing on one data set based on training statistics from another was the next consideration. Two approaches were taken. First, training signatures from one data set were applied directly to another without any change. Second, signature-extension preprocessing by an adjustment of mean values was performed on the training signatures of one data set before they were applied to the other data set.

With regard to the direct application of signatures from one data set onto another, one would expect optimum results when spectral characteristics between the two data sets were identical, that is to say, when all factors of variability (atmospheric conditions, stages of crop maturity, etc.) were negligible. These factors were not negligible in the CITARS data sets as can be seen in Fig. 4. Fig. 4 displays differences in signal values of final signatures on a channel-by-channel basis for 13 pairs of data sets. A dot represents each crop listed at the top of the graph. Each pair of data sets represents a non-local recognition case that was studied. The cases, data sets, and crop codes are defined in Table 5. If signature mean values were identical between data sets, each dot would have an ordinate value of zero. Only few points lie on this line and there is substantial variation among crop types for individual cases. Some form of preprocessing to correct for the observed differences is suggested, because one would expect large differences to correlate with poor non-local recognition performance.

The second approach used for non-local recognition processing attempted to correct for the observed differences. It assumes that differences between signatures of the same crop in two different data sets can be estimated by differences in the overall average levels of signals in the two data sets. The quarter section areas were used to estimate average levels for each data set, and differences between these data set average values were calculated. The "X"'s on Fig. 4 mark these differences in each channel. These differences in averages should closely correspond to crop differences for the method to work best. Eighty percent of the crop differences are on the same side of the zero ordinate as the corresponding difference between average values. But in only few cases is the adjustment exact for any given crop and, as noted earlier, there is substantial variation between crops.

We should consider the reasons for the observed differences between segment averages and crop means. Differences between any two data sets in ground cover types and proportions in quarter sections can cause differences between average levels. If crops appearing in one data set also appear in the second and in like proportions, then calculation of average level differences should be accurate; however, if any crop appears in large proportion in one and not the other, then the calculation of the average difference would be biased.

To provide a second type of display of signal differences for four non-local cases, we have included Fig. 5 which presents plots of the calculated correction and the comparative mean values of individual major crops. A crop lying on the correction line would, after preprocessing, have the same mean value as the corresponding crop signature of the associated data set. Aside from differences attributable to atmospheric, scan angle, and illumination changes, there can be differences in the makeup (and reflectances) of individual ground covers. For example, crops might be at different stages of maturity, cover different amounts of soil, or have different soil colors in the two sets of fields available for training. Such differences could cause the calculated differences in means of an individual crop to depart from the difference in segment average values. In a situation where the same segment is observed on two successive days, as in Fig. 5(A) for Fayette 16 and 17 July, such differences would be minimized and here the patterns of individual crops match the adjustment lines quite well. On the other hand, as can be seen in Fig. 5(B), for two different segments in the same time period, there appears to be a substantial difference between the makeup of covers (e.g., soybeans and trees) in the two segments. If the signatures shown truly represent the test (as well as training) data, it is clear that the mean-level-adjustment procedure is not optimal in this instance. In general, one would expect that a correction line whose slope is not restricted to 45^{o} on plots like those of Fig. 5 would perform better. The multiplicative and additive signature correction (MASC) algorithm recently developed at ERIM has such generality, and tests made with it on CITARS data are described and discussed in Sec. 6.

4.4 SUMMARY AND CONCLUSIONS

In determining signatures from training data, a first concern is the amount of training data available. Generally a much greater number of corn and soybean pixels were available in comparison to wheat and 'other' crop pixels. In some cases the number of 'other' crop pixels was inadequate, not even enough to prevent a calculated singular covariance matrix.

While determining major crop signatures for local recognition procedures, 6% of the separate major crop signatures were rejected in calculating the combined major crop signatures. Reasons for rejection varied from incorrect labelling to a high variability of crop characteristics at the particular time of year. There was sufficient evidence to warrant consideration of the use of multi-mode signatures as opposed to single-mode major crop signatures in some instances.

Choosing 'other' crop signatures for final recognition purposes resulted in from zero to eight 'other' class signatures being selected for each data set. The major criterion in the selection of an 'other' class signature as a final signature was the signature's relative 'nearness' to the major class signatures. If the signature was so 'far' away so as to be rendered unnecessary or so 'close' to major class signatures so as to cause too much misrecognition of major class pixels, the signature was rejected. Other problems such as a major class signature 'straddled' by an 'other' class signature, or singular covariance matrices were also adjusted to maximize recognition accuracy.

Two sets of training signatures were used in non-local recognition processing. Signatures from the training data set were used first with no adjustment. To correct for differences between data sets, a second set of preprocessed signatures was determined. These signatures were adjusted in mean level as indicated in the overall quarter-section average differences between segments in non-local data sets. Generally, preprocessed signatures better approximated the actual target data set signature mean values, although differences between crops were pronounced in some non-local data sets.

TABLE 2. SUMMARY OF FIELDS REJECTED FROM MAJOR-CLASS SIGNATURES

SEGMENT	CROP	# FIELDS USED*	# PIXELS USED*	FIELD REJECTED (# PIXELS)	FIGURE 2** PART:	ELLIPSE #
HUN 15 Jul	Corn	6/6	58/58	none	-	---
	Soy	4/15	148/151	84-71(3)	P	22
HUN 24 Sept	Corn	6/6	61/61	none	-	---
	Soy	15/15	136/136	none	-	---
SHE 8 Jun	Wheat	4/4	26/26	none	Y	---
SHE 7 Sept	Corn	24/24	155/155	none	-	---
	Soy	11/11	51/51	none	-	---
SHE 24 Sept	Corn	23/24	152/153	19-45(1)	A	14
	Soy	10/11	51/54	34-08(3)	B	42
WHI 21 Aug	Corn	23/24	449/463	37-44(14)	I	16
	Soy	16/19	202/213	25-43(5)	J	45
				85-41(6)		44
				85-50(2)		35
WHI 7 Sep	Corn	23/24	411/435	46-45(24)	K	18
	Soy	17/19	195/209	25-41(8)	L	33
				25-43(6)		35
LIV 16 Jul	Corn	10/11	127/160	67-03(33)	M	33
	Soy	19/20	218/224	89-03(6)	N	21
LIV 3 Aug	Corn	10/11	135/162	67-03(27)	O	34
	Soy	20/20	231/231	none	-	---
FAY 10 Jun	Wheat	8/9	43/47	5-48(4)	Q	10
FAY 11 Jun	Wheat	8/9	33/36	5-48(3)	R	11
FAY 29 Jun	Corn	9/9	66/66	none	-	---
	Soy	14/15	101/107	77-08(6)	S	31
FAY 16 Jul	Corn	5/9	57/69	95-30(12)	U	36
	Soy	16/17	123/127	77-08(4)	V	16
FAY 17 Jul	Corn	8/9	57/65	64-70(8)	W	16
	Soy	17/18	130/136	77-08(6)	X	25
FAY 21 Aug	Corn	9/9	66/66	none	Z	---
	Soy	19/21	166/177	22-45(8)	T	79
				64-63(3)		76
LEE 17 Jul	Corn	13/14	115/117	92-67(2)	E	22
	Soy	10/11	102/108	86-03(6)	F	32
LEE 18 Jul	Corn	14/15	125/127	92-67(2)	C	23
	Soy	10/11	94/100	86-03(6)	D	33
LEE 5 Aug	Corn	12/15	101/125	31-66(15)	G	16
				20-23(7)		11
				92-67(2)		23
	Soy	10/11	98/104	86-03(6)	H	33

*The column entries are (number used)/(number available).

**The last two columns indicate which part of Fig. 2 and which specific ellipse in that part represents the field in question.

TABLE 3. PRELIMINARY TEST RESULTS FOR MAJOR-CROP SIGNATURES ONLY (ERIM-ERTS-SP1)

		PERCENT CORRECT OF POINTS IN EACH CLASS			
SEGMENT	DATE	CORN	SOYBEANS	WHEAT	OTHER (Not Classified)
HUN	15 JUL	70.7	87.3		22.9
LIV	16 JUL	65.5	85.6		38.7
FAY	16 JUL	95.1	94.4		0.8
FAY	17 JUL	97.2	92.2		1.3
LEE	17 JUL	78.0	88.9		5.5
LEE	18 JUL	77.5	88.0		6.6
LIV	3 AUG	77.7	54.4		49.3
FAY	21 AUG	92.7	88.3		16.0
SHE	8 JUN			52.8	

TABLE 4. SELECTION OF "OTHER"–CLASS SIGNATURES

PASS	ALGORITHM	ITEMS
1	ACCEPTANCE	M,S,D
2	SPLITTING	AP,NSP
3	REJECTION	A,R
4	SINGULARITY	SC,NSC
F	FINAL	M,O,D

KEY:

M	a major crop
S	a 'significant' crop
SP	splitting required
NSP	no splitting
A	accepted
R	rejected
SC	singular covariance
NSC	non singular
O	'other' class signature
D	delete

HUN 15 JUL	PASS 1	2	3	4	F
CORN	M	–	–	–	M
SOY	M	–	–	–	M
TREE	S	NSP	A	NSC	O
PASTUR	S	NSP	A	NSC	O
WATER	D	–	–	–	D
GRASS	S	NSP	R	–	D

HUN 24 SEP	PASS 1	2	3	4	F
CORN	M	–	–	–	M
SOY	M	–	–	–	M
TREE	S	NSP	A	NSC	O
PASTUR 1	S	SP	R	–	D
2			A	SC	O
WATER	D	–	–	–	D
GRASS	S	NSP	R	–	D

SHELBY 8 JUN	PASS 1	2	3	4	F
WHEAT	M	–	–	–	M
SOY	D	–	–	–	D
TREE	D	–	–	–	D
CORN	D	–	–	–	D
OATS	D	–	–	–	D
BARE	D	–	–	–	D
CLOVER	D	–	–	–	D
FESCUE	D	–	–	–	D
GRASS	D	–	–	–	D

SHELBY 7 SEPT	PASS 1	2	3	4	F
CORN	M	–	–	–	M
SOY	M	–	–	–	M
FESCUE	S	NSP	R	–	D
GRASS	S	NSP	R	–	D
WHEAT	S	NSP	A	SC	O
CLOVER 1	S	SP	A	NSC	O
2			A	NSC	O
STUBBLE	S	NSP	A	SC	O
TREE	S	NSP	A	NSC	O

SHELBY 24 SEPT	PASS 1	2	3	4	F
CORN	M	–	–	–	M
SOY	M	–	–	–	M
FESCUE	S	NSP	A	SC	O
GRASS	S	NSP	A	SC	O
WHEAT	S	NSP	A	SC	O
CLOVER	S	SP	A	NSC	O
			R	–	D
STUBBLE	S	NSP	R	–	D
TREE	S	NSP	A	NSC	O

TABLE 4 (Continued)

WHITE 21 AUG	PASS 1	2	3	4	F
CORN	M	-	-	-	M
SOY	M	-	-	-	M
PASTUR 1	S	SP	A	NSC	O
2			A	NSC	O
QUARRY	D	-	-	-	D
WOODS 1	S	SP	A	NSC	O
2			A	NSC	O

WHITE 7 SEPT	PASS 1	2	3	4	F
CORN	M	-	-	-	M
SOY	M	-	-	-	M
PASTUR 1	S	SP	R	-	D
2			R	-	D
QUARRY	S	NSP	A	NSC	O
WOODS 1	S	SP	A	NSC	O
2			R	-	D

LIV 16 JUL	PASS 1	2	3	4	F
CORN	M	-	-	-	M
SOY	M	-	-	-	M
PASTUR	S	NSP	A	SC	O
TREE	S	NSP	A	NSC	O
OATS	S	NSP	R	-	D
QUARRY	D	-	-	-	D
OTHER	D	-	-	-	D

LIV 3 AUG	PASS 1	2	3	4	F
CORN	M	-	-	-	M
SOY	M	-	-	-	M
PASTUR	S	NSP	A	SC	O
TREE	S	NSP	R	-	D
OATS	S	NSP	A	NSC	O
QUARRY	D	-	-	-	D
OTHER	S	NSP	A	NSC	O

FAY 10 JUN	PASS 1	2	3	4	F
WHEAT	M	-	-	-	M
CORN	D	-	-	-	D
SOY	D	-	-	-	D
WATER	D	-	-	-	D
TREE	D	-	-	-	D
BARE	D	-	-	-	D
BRUSH	D	-	-	-	D
CLOVER	D	-	-	-	D
WEEDS	S	NSP	A	NSC	O

FAY 11 JUN	PASS 1	2	3	4	F
WHEAT	M	-	-	-	M
CORN	D	-	-	-	D
SOY	D	-	-	-	D
WATER	D	-	-	-	D
TREE	D	-	-	-	D
BARE	D	-	-	-	D
BRUSH	D	-	-	-	D
CLOVER	D	-	-	-	D
WEEDS	S	NSP	A	NSC	O

FAY 16 JUL	PASS 1	2	3	4	F
CORN	M	-	-	-	M
SOY	M	-	-	-	M
TREE	S	NSP	A	NSC	O
WHEAT 1	S	SP	R	-	D
2			A	NSC	O
BRUSH	S	NSP	R	-	D
CLOVER	S	NSP	A	SC	O
BARE	S	NSP	A	NSC	O
STUBBLE	S	NSP	R	-	D
WEEDS 1	S	SP	R	-	D
2		R	R	-	D
WATER	D	-	-	-	D

FAY 17 JUL	PASS 1	2	3	4	F
CORN	M	-	-	-	M
SOY	M	-	-	-	M
TREE	S	NSP	A	NSC	O
WHEAT 1	S	SP	R		D
2			A	NSC	O
BRUSH	S	NSP	R	-	D
CLOVER	S	NSP	A	SC	O
BARE	S	NSP	A	NSC	O
STUBBLE	S	NSP	R	-	D
WEEDS	S	NSP	R	-	D
WATER	D	-	-	-	D

TABLE 4 (Concluded)

FAY 21 AUG	PASS 1	2	3	4	F
CORN	M	-	-	-	M
SOY	M	-	-	-	M
TREE	S	NSP	A	NSC	O
WATER	D	-	-	-	D
BARE	S	NSP	A	NSC	O
BRUSH	S	NSP	R	-	D
CLOVER	S	NSP	A	NSC	O
WEED	S	NSP	A	SC	O

FAY 29 JUN	PASS 1	2	3	4	F
CORN	M	-	-	-	M
SOY	M	-	-	-	M
BARE 1	S	SP	A	NSC	O
2			A	NSC	O
BRUSH	S	NSP	A	SC	O
CLOVER	S	NSP	A	NSC	O
WATER	D	-	-	-	D
WEEDS	S	NSP	A	NSC	O
WHEAT 1	S	SP	A	SC	O
2			R	-	D
3			A	NSC	O
TREE	S	NSP	A	NSC	O

LEE 17 JUL	PASS 1	2	3	4	F
CORN	M	-	-	-	M
SOY	M	-	-	-	M
TREE	S	NSP	A	NSC	O
OATS 1	S	SP	R	-	D
2			R	-	D
BARE	S	NSP	R	-	D
HAY	S	NSP	R	-	D
OTHER	D	-	-	-	D
PASTUR	S	NSP	R	-	D

LEE 18 JUL	PASS 1	2	3	4	F
CORN	M	-	-	-	M
SOY	M	-	-	-	M
TREE	S	NSP	A	NSC	O
OATS 1	S	SP	R	-	D
2			R	-	D
BARE	S	NSP	R	-	D
HAY	S	NSP	R	-	D
PASTUR 1	S	SP	R	-	D
2			R	-	D
OTHER	D	-	-	-	D

LEE 5 AUG	PASS 1	2	3	4	F
CORN	M	-	-	-	M
SOY	M	-	-	-	M
TREE 1	S	SP	R	-	D
2			A	NSC	O
OATS 1	S	SP	R	-	D
2			A	NSC	O
BARE	S	NSP	R	-	D
HAY	S	NSP	A	SC	O
PASTUR	S	NSP	A	NSC	O
OTHER	D	-	-	-	D

TABLE 5. CODES USED TO DESCRIBE CITARS DATA SETS

Segments:

1 - Huntington
2 - Shelby
3 - White
4 - Livingston
5 - Fayette
6 - Lee

Times:

1 - June 8-12 pass 1
2 - June 8-12 pass 2
3 - June 26-30 pass 1
4 - June 26,30 pass 2
5 - July 14-18 pass 1
6 - July 17-18 pass 2
7 - Aug 1-5 pass 1
8 - Aug 1-5 pass 2
9 - Aug 19-23 pass 1
10 - Aug 19-23 pass 2
11 - Sep 6-10 pass 1
12 - Sep 6-10 pass 2
13 - Sep 24-28 pass 1

Nonlocal Recognition:

ANOVA CODE	SIGNATURES FROM		TO	ERIM CODE
1 -	5(5)	→	5(6)	DE
2 -	5(6)	→	5(5)	ED
3 -	6(5)	→	6(6)	IJ
4 -	6(6)	→	6(5)	JI
5 -	1(6)	→	4(5)	GF
6 -	1(6)	→	6(6)	GJ
7 -	6(6)	→	4(5)	JF
8 -	6(6)	→	1(6)	JG
9 -	6(8)	→	4(7)	OT
10 -	4(7)	→	6(8)	TO
11 -	4(5)	→	5(5)	FD
12 -	5(5)	→	4(5)	DF
13 -	3(11)	→	2(12)	QP
14 -	2(12)	→	3(11)	PQ
15 -	2(13)	→	1(13)	SR
16 -	1(13)	→	2(13)	RS
17 -	5(6)	→	1(6)	EG
18 -	1(6)	→	5(6)	GE
19 -	2(1)	→	5(1)	CB
20 -	5(1)	→	2(1)	BC
21 -	5(1)	→	5(2)	BA
22 -	5(2)	→	5(1)	AB
23 -	3(10)	→	5(9)	LN
24 -	5(9)	→	3(10)	NL

Time
Segment

TABLE 5 (Continued)

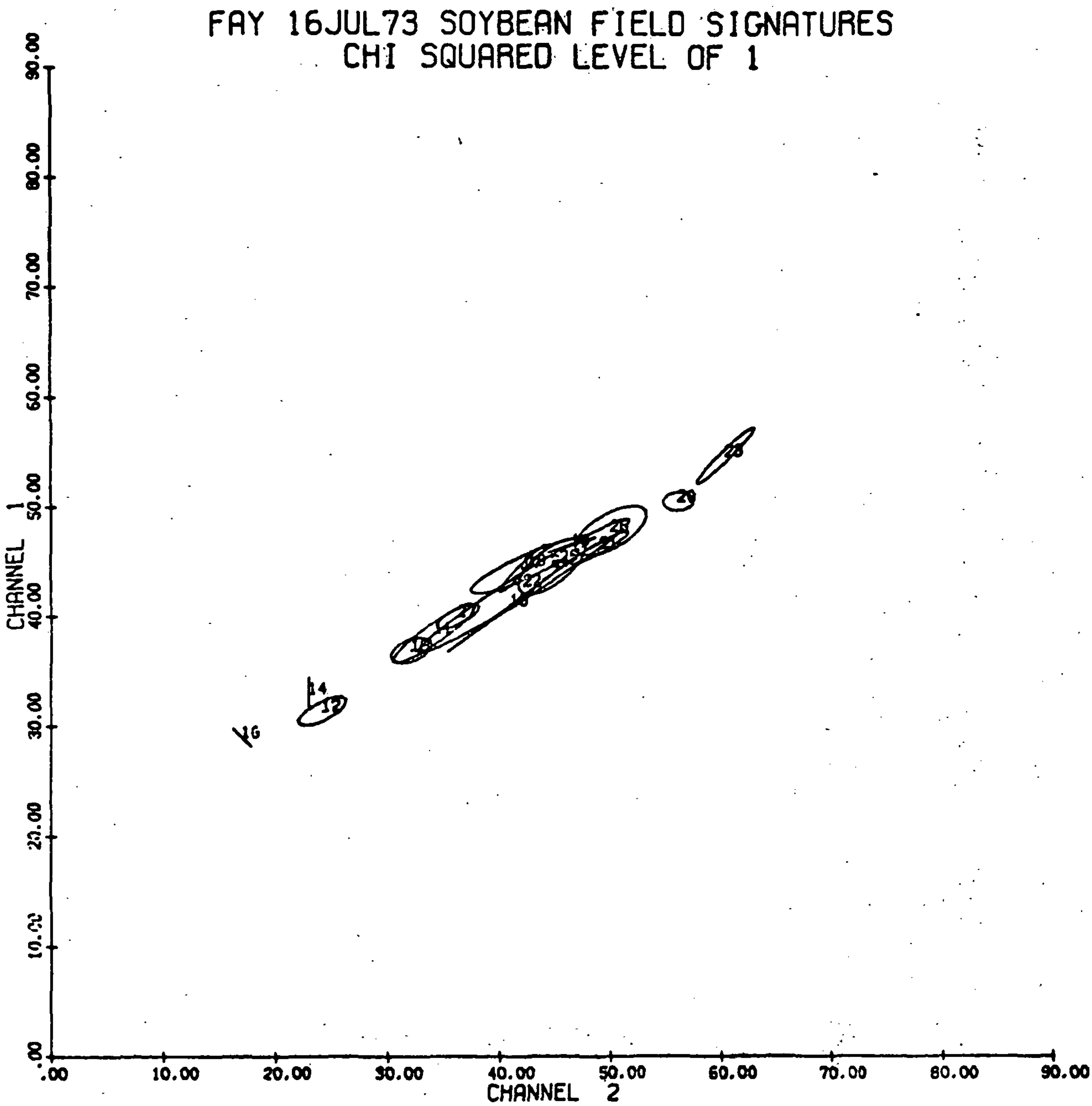
ERIM Code	ANOVA Code	Segment	Date
A	5(2)	FAY	11 JUN
B	5(1)	FAY	10 JUN
C	2(1)	SHE	8 JUN
D	5(5)	FAY	16 JUL
E	5(6)	FAY	17 JUL
F	4(5)	LIV	16 JUL
G	1(6)	HUN	15 JUL
I	6(5)	LEE	17 JUL
J	6(6)	LEE	18 JUL
L	3(10)	WHI	21 AUG
N	5(9)	FAY	21 AUG
O	4(7)	LIV	3 AUG
P	2(12)	SHE	7 SEP
Q	3(11)	WHI	7 SEP
R	1(13)	HUN	24 SEP
S	2(13)	SHE	24 SEP
T	6(8)	LEE	5 AUG
Y	5(3)	FAY	29 JUN

Time

Segment

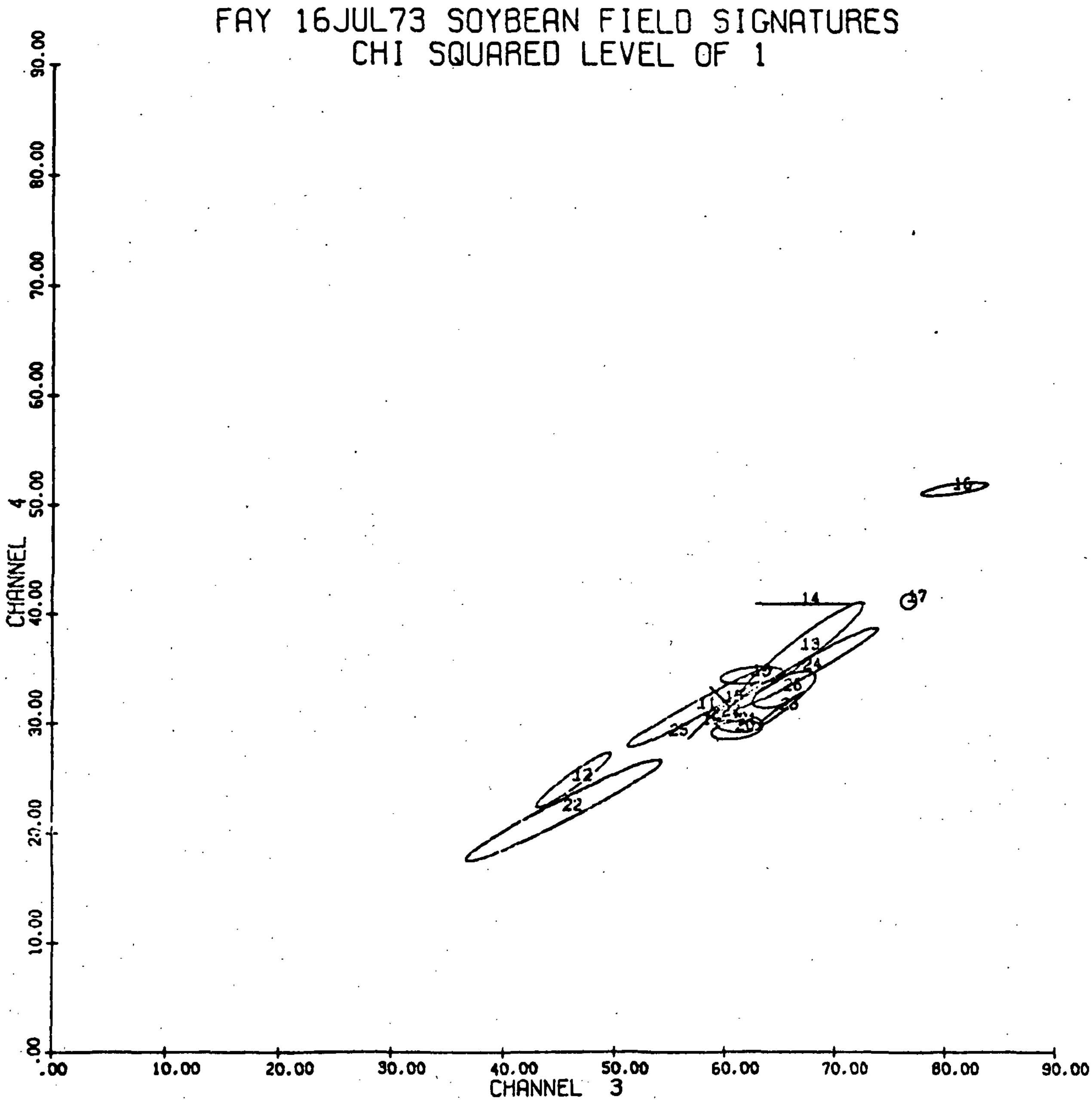
ERIM CROP CODES (Partial List):

B - Bare Soil
C - Corn
E - Weeds
L - Clover
P - Pasture
S - Soybeans
T - Trees
W - Wheat



(A) CHANNELS 1 AND 2.

FIGURE 1. SAMPLE ELLIPSE PLOTS OF SEPARATE SOYBEAN FIELD SIGNATURES.



(B) CHANNELS 3 AND 4.

FIGURE 1. SAMPLE ELLIPSE PLOTS OF SEPARATE SOYBEAN FIELD SIGNATURES.

FIGURE 2. UNIT CONTOUR ELLIPSE PLOTS OF SEPARATE MAJOR-CROP FIELD SIGNATURES
(CHANNEL 4 VS. CHANNEL 2 UNLESS OTHERWISE SPECIFIED)

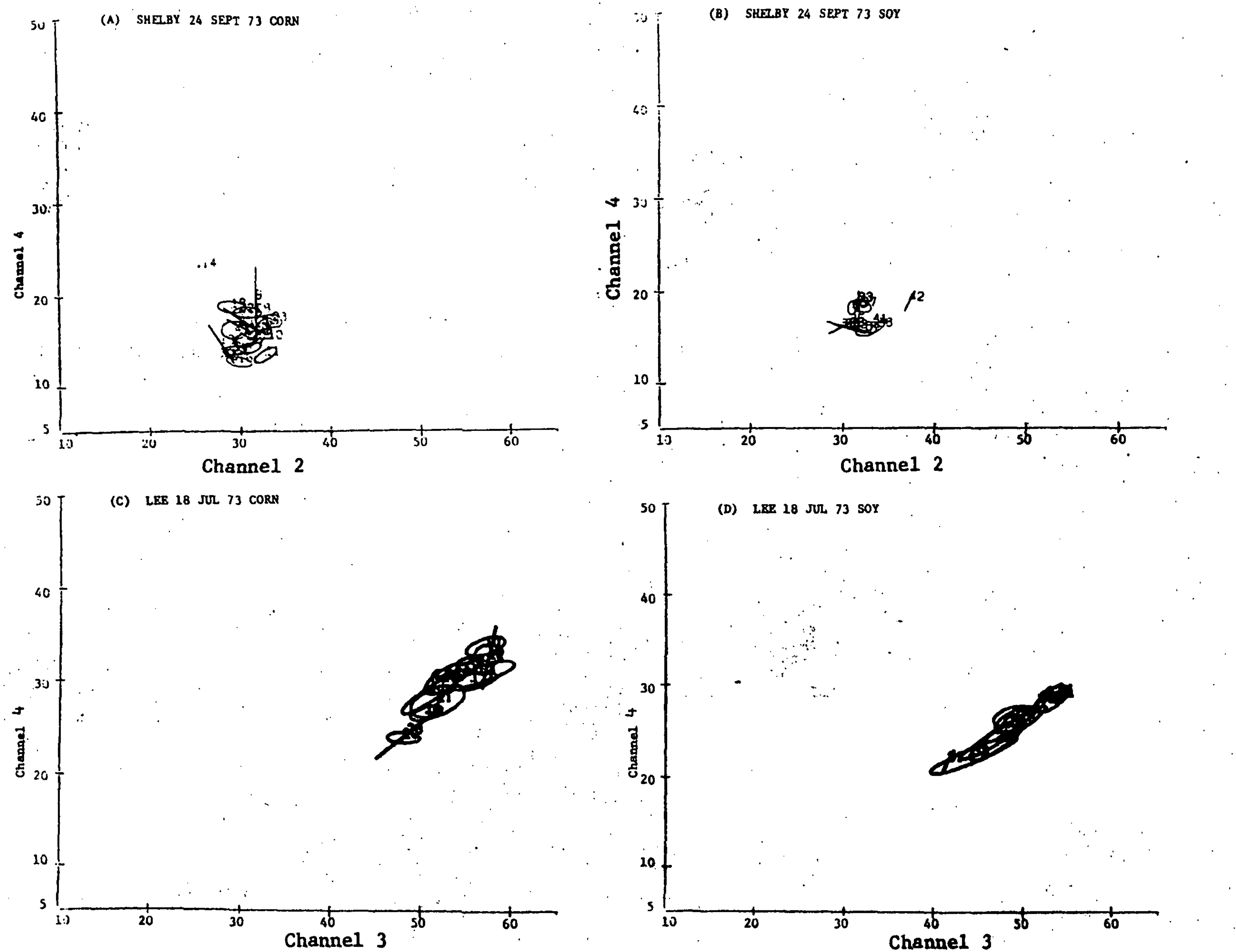


FIGURE 2 (Continued)

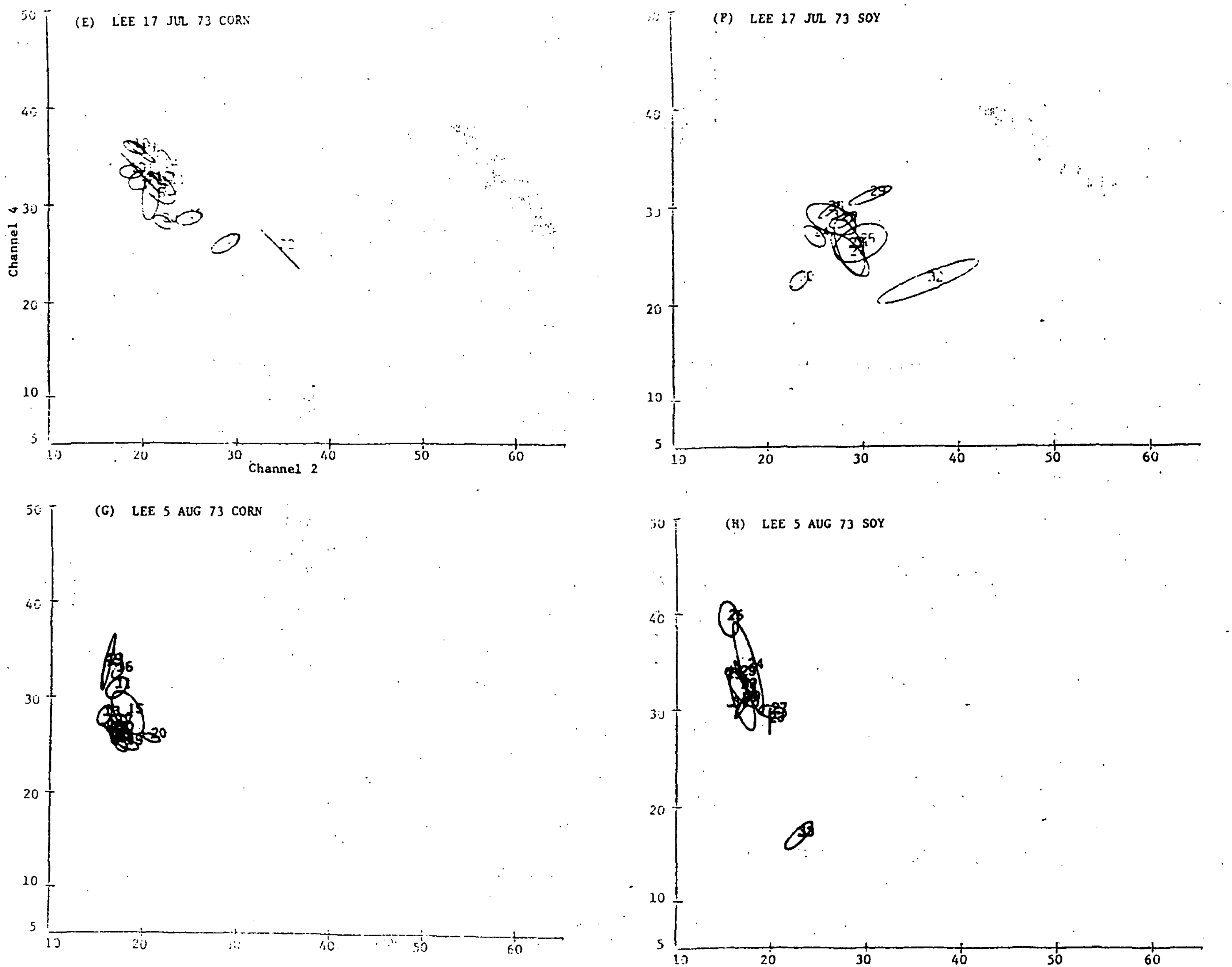


FIGURE 2 (Continued)

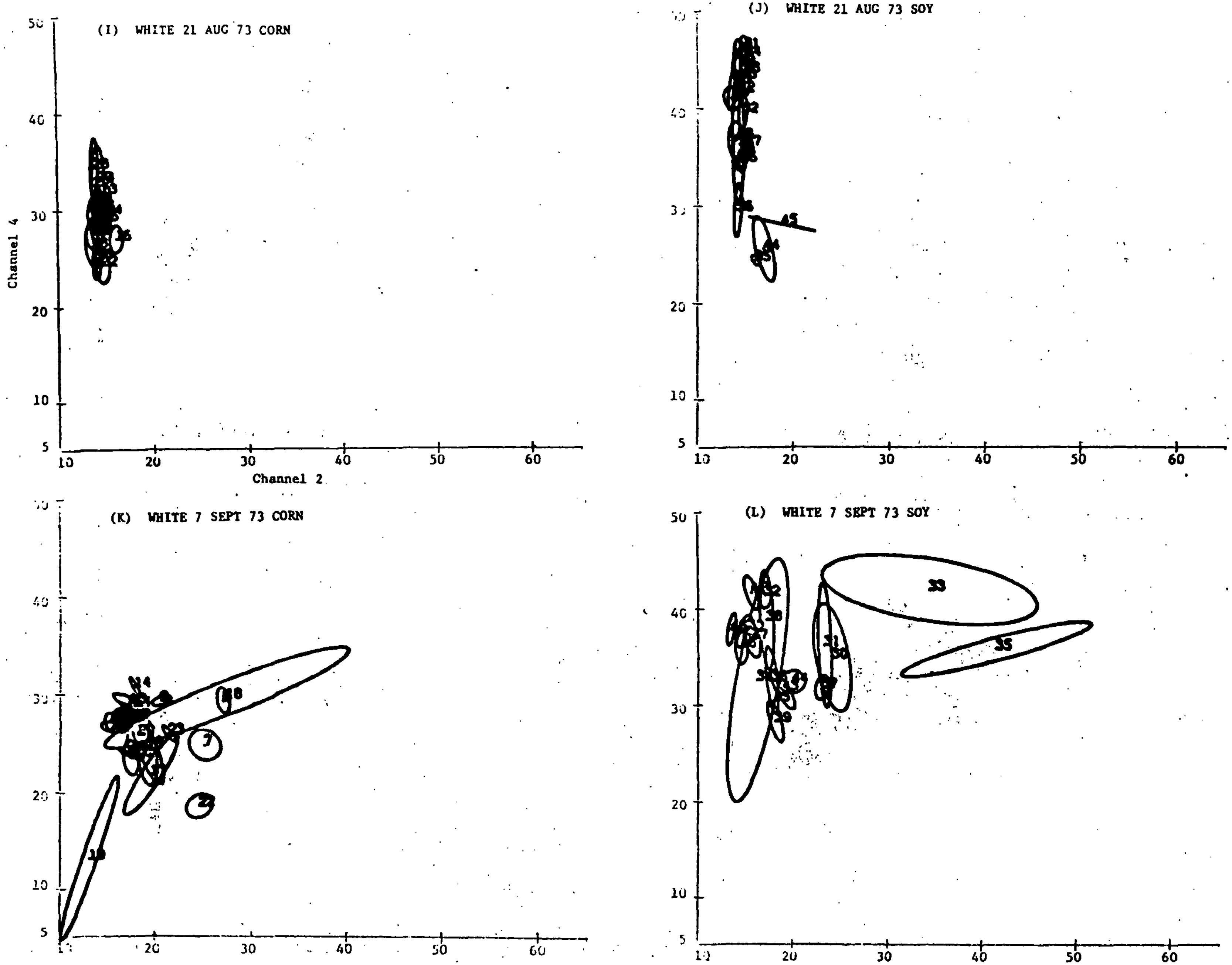


FIGURE 2 (Continued)

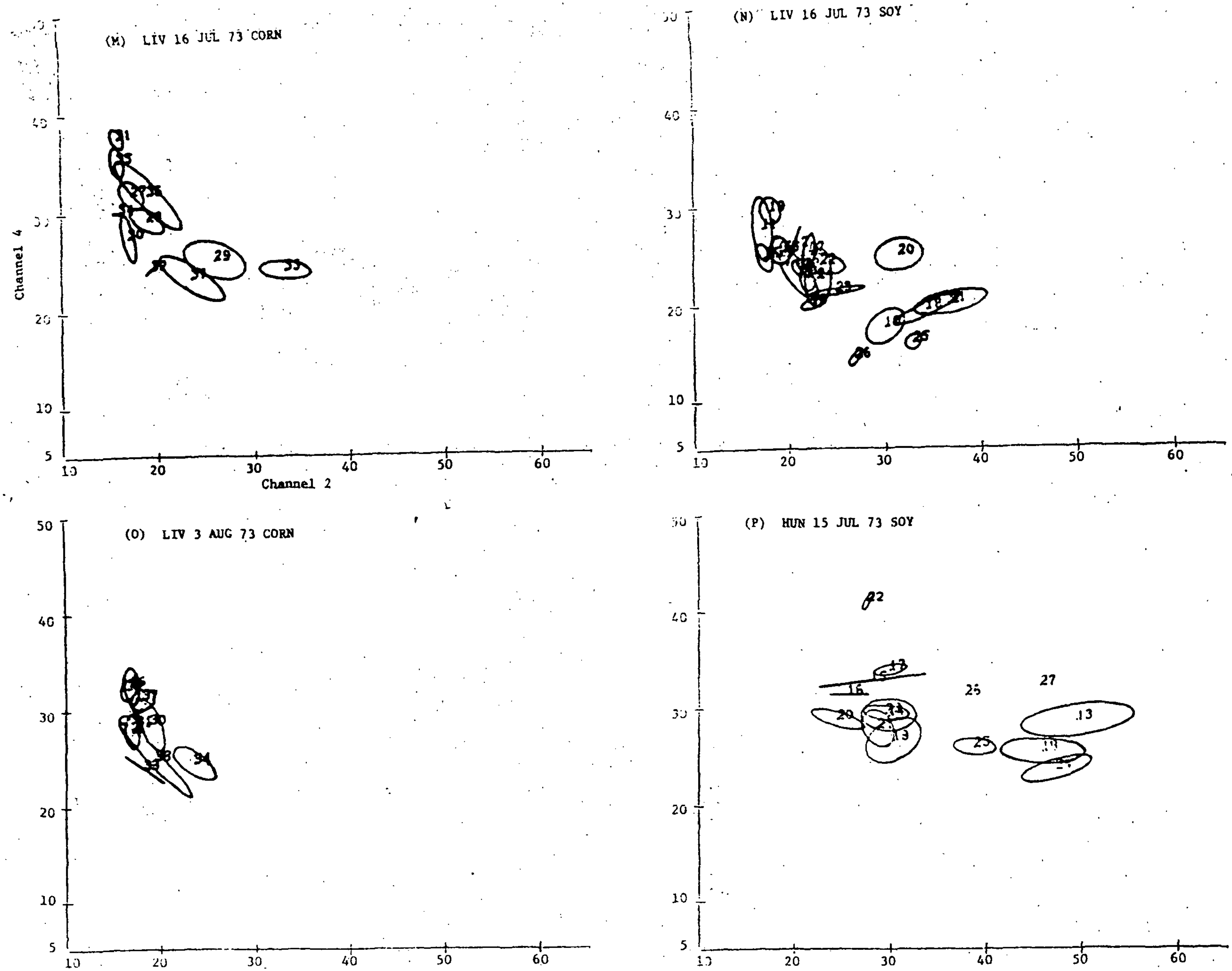


FIGURE 2 (Continued)

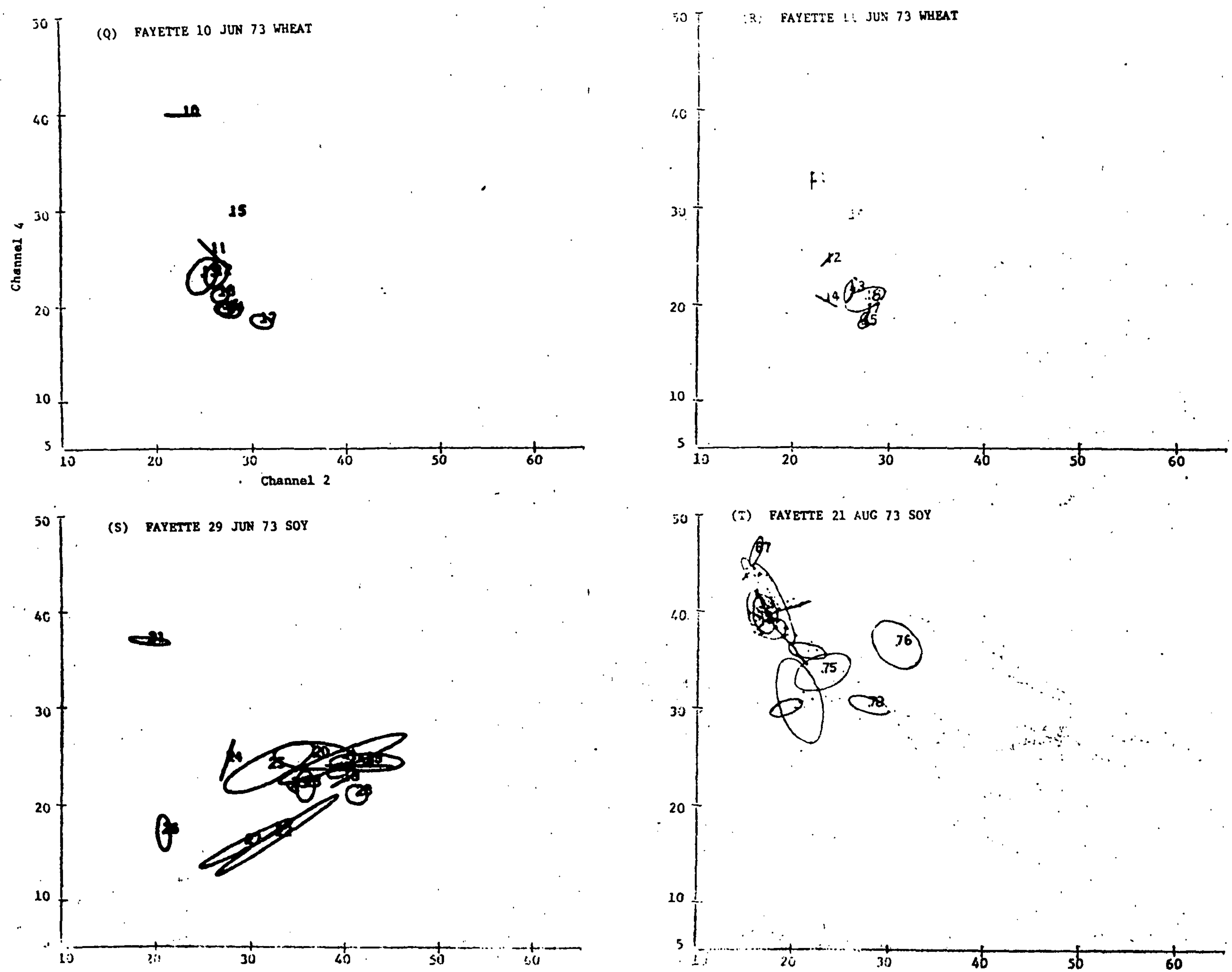


FIGURE 2 (Continued)

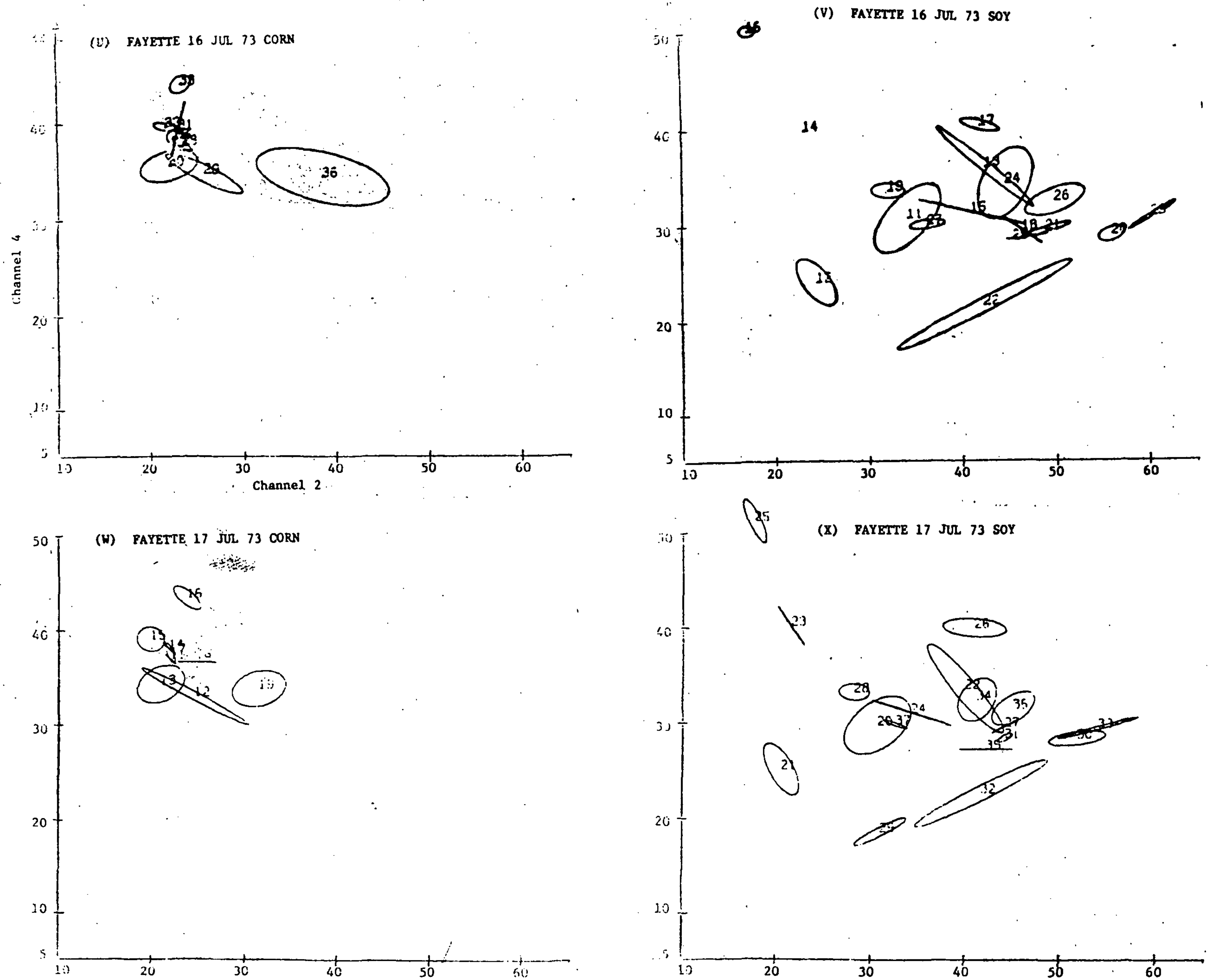
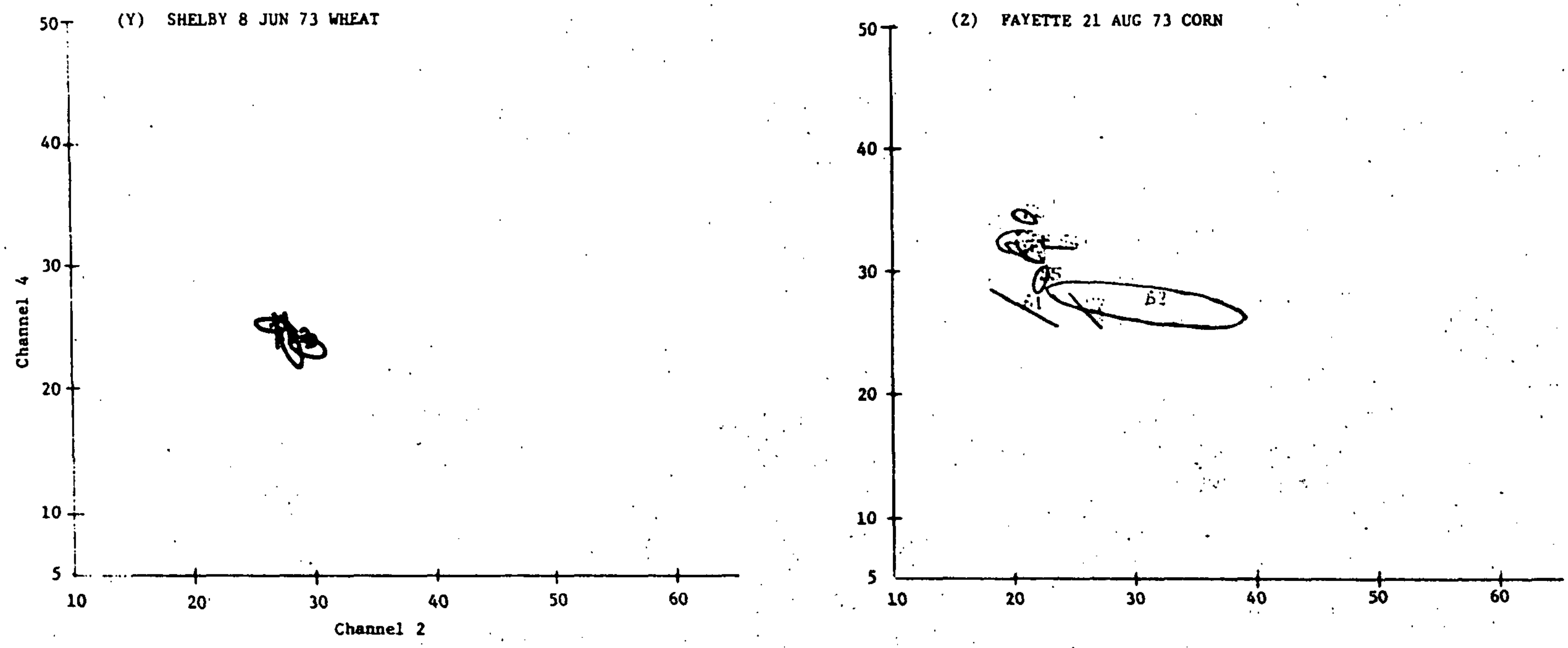
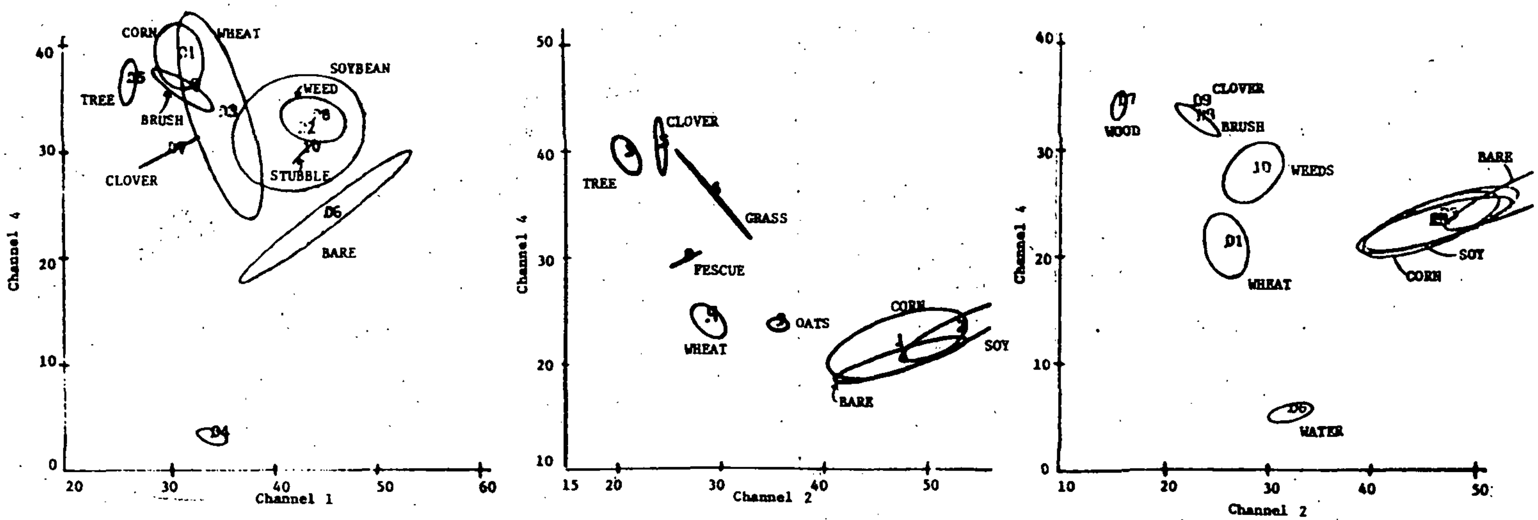


FIGURE 2 (Continued)





(A)

Fayette 16 July 1973

(B)

Shelby 8 June 1973

(C)

Fayette 11 June 1973

FIGURE 3: UNIT CONTOUR ELLIPSES FOR MAJOR FIELD SIGNATURES

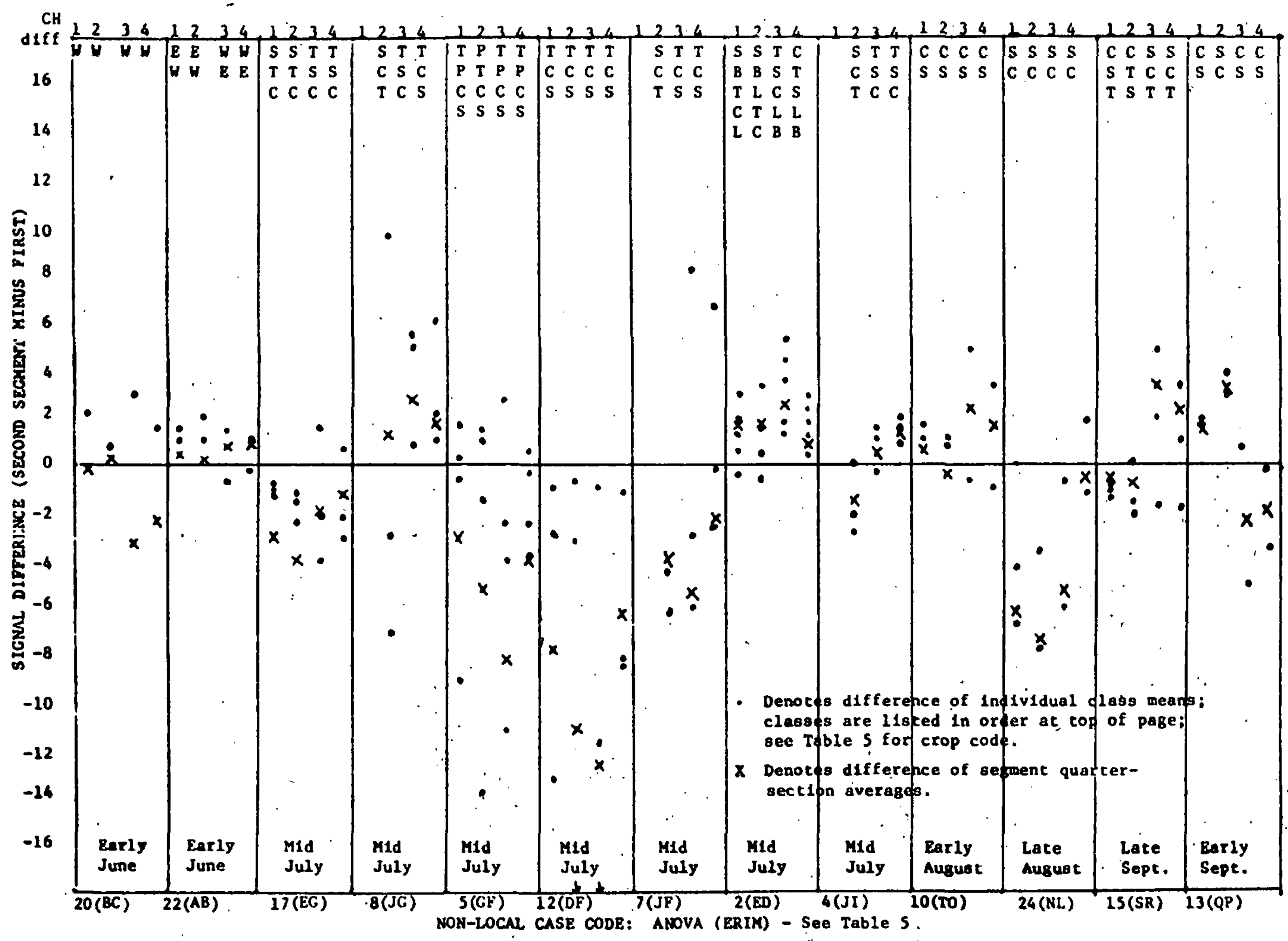


FIGURE 4. SIGNAL DIFFERENCES BETWEEN SEGMENTS FOR THE AVERAGE SCENE (MLA) AND INDIVIDUAL SCENE CLASSES.

FIGURE 5: COMPARISON OF MAJOR-CROP SIGNALS FOR NON-LOCAL RECOGNITION PROCESSING (Continued)

(A) NON-LOCAL CASE 1: 5(5) → 5(6)

(B) NON-LOCAL CASE 5: 1(6) → 4(5)

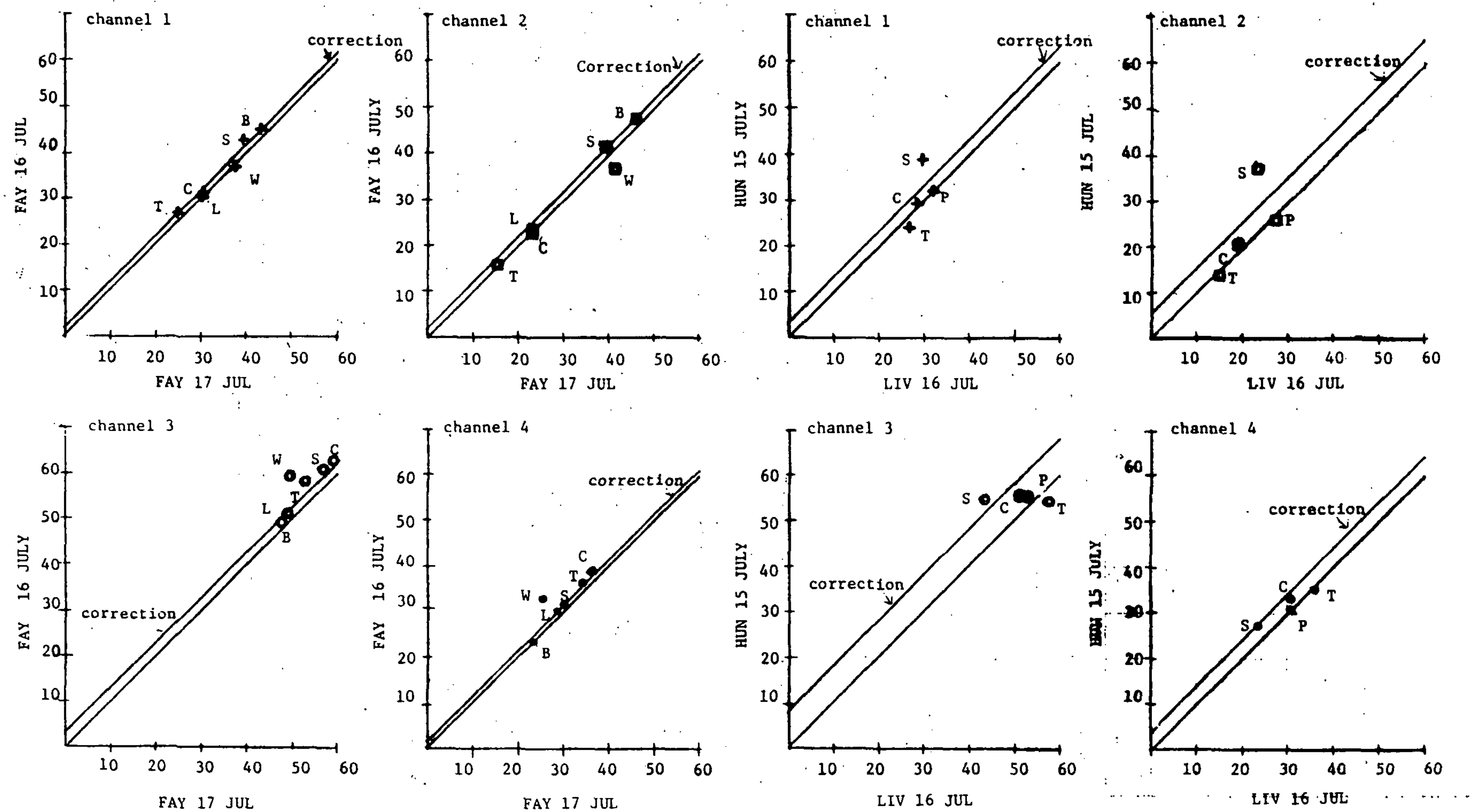
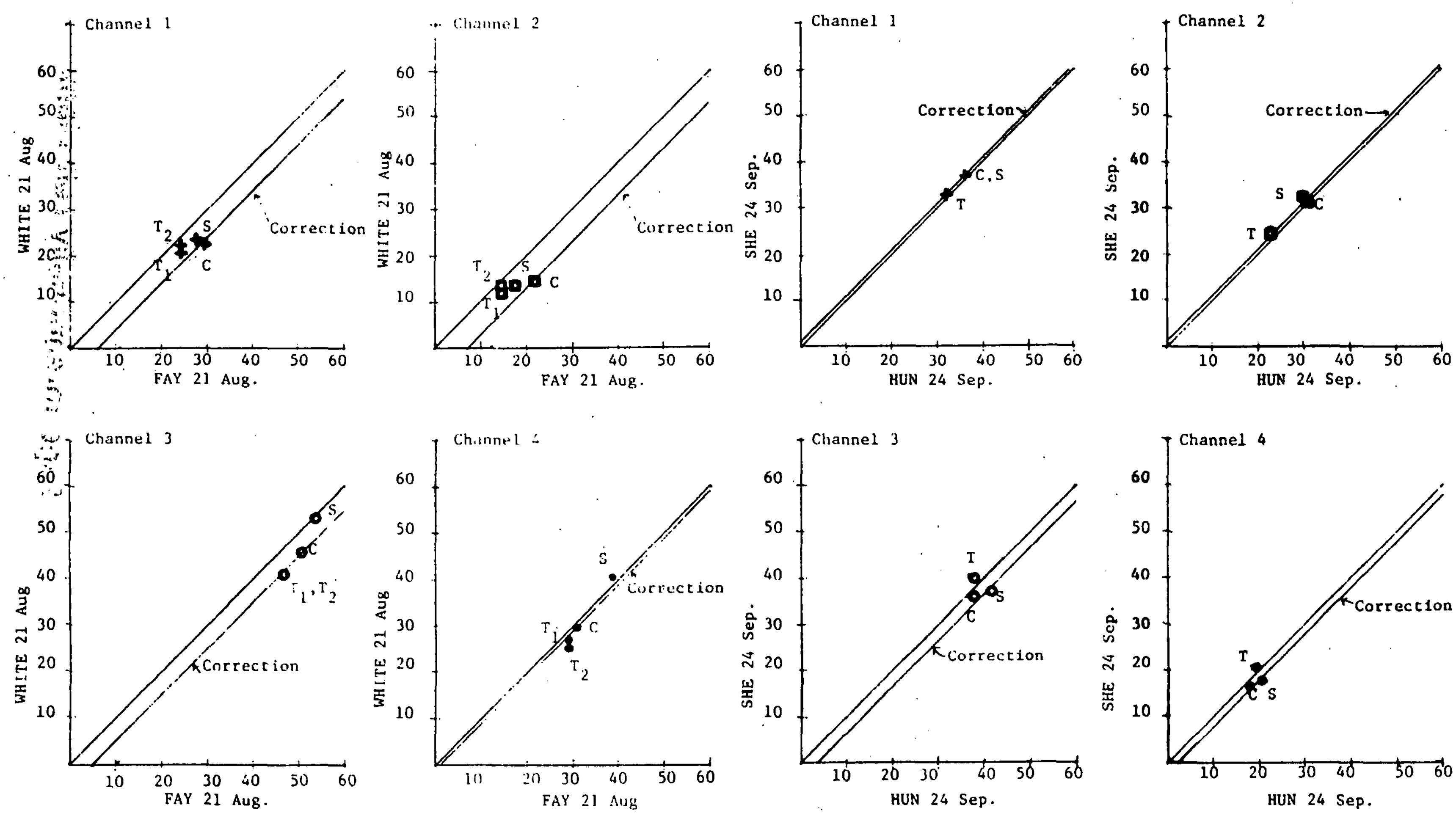


FIGURE 5: COMPARISON OF MAJOR-CROP SIGNALS FOR NON-LOCAL RECOGNITION PROCESSING (Concluded)

(C) NON-LOCAL CASE 23: 2(10) → 5(9)

(D) NON-LOCAL CASE 15: 2(13) → 1(13)



Page Intentionally Left Blank

5
RECOGNITION PROCESSING RESULTS

Once final signatures were established (See Sec. 4), final recognition results were generated systematically by using these signatures to recognize test data. All four channels of ERTS data were used to process every data set except the Lee 18 July segment and non-local recognition cases involving Lee 18 July signatures; for these exceptions, Channel 1 (ERTS Band 4) was omitted because of ERTS data quality problems. Results then were compiled with post-recognition analysis programs. Results cards were produced and forwarded to EOD for use in an analysis of variance (ANOVA) effort. Also, a variety of summary tables and graphs were generated and analyzed to produce the observations and conclusions that are presented in this section. Upon receipt of the ANOVA results, additional conclusions and/or modifications will be generated and reported in the joint final report for CITARS.

5.1 SUMMARY OF RECOGNITION RESULTS

Recognition results obtained for all prescribed data sets with the various ERIM procedures are summarized in Table 6. The majority of processing and analysis efforts was directed toward the recognition of corn and soybeans. Only for three early June data sets (last page of Table 6) was wheat the crop of interest.

Overall performance in corn and soybean recognition was best in late August, averaging 80% correct for field centers in the two available segments.

Lower accuracies were achieved both earlier and later in the growing season. Full-section results did not so clearly depend on the time of season, but some of the better results obtained were in late August. There was substantial variation between segments, but the usual tendency was to over-estimate major-crop proportions. There does not appear to be a high correspondence between overall recognition performance in field centers and in full sections.

In late August, field-center corn and soybeans were about equally well detected. Soybean signatures exhibited a much greater variability due to a wide range of planting dates and consequent variability in maturity in early and mid season.

The speed and computational advantage of the ERIM "best linear" decision rule did not result in a degradation of its performance in relation to the more conventional quadratic decision rule. In fact, this linear rule tended to give slightly better performance than the quadratic rule, with the same signatures and test data and for about 1/3 the general-purpose digital computer cost.

Non-local recognition processing with unadjusted signatures usually produced a degradation in recognition performance from the levels achieved with local signatures, i.e., by an average 12% decrease in field-center correct recognition and more than a 20% increase in RMS deviation from true crop proportions for full-section recognition. On the average, signature extension preprocessing by mean level adjustment improved non-local recognition performance both for field centers (+6%) and full sections. For the late August data, a dramatic improvement of +34% was obtained for non-local recognition (with linear rule) in White by adjusting Fayette signatures. A corresponding improvement of only +8% was obtained in Fayette by adjusting White signatures. Differences in the dispersion volumes of the two sets of signatures were noted.

The evaluation of wheat recognition performance was severely hampered by a lack of training and test data for wheat in the CITARS test segments. Furthermore, special procedures had to be used to obtain ground truth for test wheat fields and the validity of wheat fields designated for testing is placed in doubt.

We investigated some of the reasons for lower-than-hoped-for performance on the CITARS data and noted some ways in which performance measures might have been increased by as much as ten to twenty percent. These are discussed in Sec. 5.3 and Sec. 6. Sec. 6 also presents improved results obtained with procedures more advanced than those defined for, and used in, the standardized CITARS data processing.

5.2 DETAILED DISCUSSION AND PERFORMANCE COMPARISONS

The CITARS test plan was designed to answer a number of specific questions regarding crop recognition performance. In this section, we discuss the recognition results in detail in the context of these key questions. Because the many factors in the test design are interrelated, it is next to impossible to discuss them independently, so forward and backward referencing and some repetition are necessary in the sections that follow.

5.2.1 BEST PERFORMANCE AND BEST TIME FOR CORN AND SOYBEAN FIELD-CENTER RECOGNITION

Linear-rule field-center local recognition results from Table 6 are presented graphically in Fig. 6. One is immediately struck by the great variability in detection performance throughout the growing season for the three classes -- corn, soybeans, and other. Concentrating first on the dashed line, which represents the average correct percentage of pixels in all classes, we find that overall accuracies ranged from as low as 51% correct in late Sept. to a high of 80% correct in late August; the average accuracy was 64% correct. Table 7 presents a tabular listing of the overall recognition accuracies. It is risky to draw definitive conclusions regarding the best time periods because of the lack of continuity in the data; that is, we did not have useable data from each segment in each time period so time of year differences are confounded with between-segment differences.

The two late-August segments achieve the same best overall performance level in different ways. In Fayette County, the major crops are recognized with 86% to 87% accuracy but "other" recognition is lower at 70%. The opposite is true in White County where "other" field-center pixels are recognized with 90% accuracy while corn and soybeans have 72% and 78% correct recognition. The dispersion volumes of the Fayette signatures are 9 and 30 times greater than White soybean and corn signatures, respectively. As will be discussed later, the full-section proportions of corn were much over-estimated in Fayette and slightly underestimated in White. Soybean proportions were much more accurately estimated in both data sets, being more accurate in Fayette.

Local recognition of field centers by the quadratic decision rule followed the same overall trend as the linear rule but fared slightly poorer, with a range of 38% to 80% and an average of 59% correct.

Fig. 6 also presents performance curves (% detection) for each of the three classes. We see that for several data sets one of the classes has a better performance than was achieved in late August; however, the other two classes are correspondingly lower, causing the overall performance to be lower than the best in late August. We can postulate reasons for the better performance in late August. Perhaps the major reason is the fact that soybeans were planted at times differing by as much as six weeks and exhibited much more variability in percent ground cover during the early and middle parts of the growing season. This variability is evident in the field signature ellipse plots of Fig. 2 and also in Fig. 7 and Table 8 which were generated under another task of this SR&T contract [8]. Also, the dispersion volumes of the soybean signatures tended to be larger in mid July than later. Corn fields had reached their full height and cover and had tasseled by late August or earlier and probably did not begin to senesce significantly until later. Differences in soil color also are minimized in the latter part of the growing season when ground cover reaches its maximum values. However, it was noticed that confusion between corn and soybeans was as closely related to segment as to time of year.

5.2.2 BEST PERFORMANCE AND BEST TIME FOR CORN AND SOYBEAN FULL-SECTION RECOGNITION

Next, we consider full-section recognition in more detail. Full-section recognition yielded a proportion of each area assigned to each crop category. Each pixel was assigned in full to a crop category, even though it may have been on the boundary between fields. As a measure of overall performance, we computed the RMS deviation of recognized proportions from the true proportions:

$$\text{RMS Deviation} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (p_i - \hat{p}_i)^2} \times 100\%$$

where

p_i = true proportion of crop i (from ground truth) in the test areas within the segment,

$\hat{p}_i$ = proportion of total area recognized as crop i,

and

n = no. of crop categories (3 or 2).

This measure varies from 0 (perfect) to 100 (for the two-crop case) and 82 (for the three-crop case).

For local linear-rule recognition performance on full sections, this measure varied from 24% to a respectable 2.8% on data sets for which corn and soybeans were the major crops (See Table 9). With the quadratic rule, the range was somewhat worse, 30% to 4%. The best linear-rule performance was for the Shelby 7 September data set, while White 21 August was second best and Fayette 21 August was seventh best. From the data, there is not a time trend in full-section performance that is as clear as that for field centers, although more consistency is apparent in August and early September.

Full-section performances on the individual classes -- corn, soybeans, and other -- are summarized in Table 10. It presents the ratio of recognized to true crop proportions for each local recognition case. Most of the time, one or both major-crop proportions are overestimated.

We examined areas in the Fayette 21 August data set where full-section results were poor, and made the following two observations. First, urban areas, which contain mixtures of all sorts of ground cover types, were substantially misrecognized as corn. Second, wooded areas, especially on the boundaries or in sparse woods, tended to be misrecognized as corn. See additional discussion in Sec. 6.6.

For corn, the most accurate proportion estimates occurred in mid July and August segments -- Livingston 16 July and 3 August, and Lee 17 and 18 July. However, the worst results, all overestimates, also occurred in mid July (Huntington 15 July and Fayette 16 and 17 July). In contrast, the earlier field-center accuracies for corn were higher for the Huntington and Fayette data sets than for the Livingston and Lee sets.

A possible explanation is that Lee and Livingston are characterized by fields of larger average size (See Table 11) than Huntington and Fayette. Therefore, a greater fraction of the area in the latter segments consists of mixture pixels. Also, the corn signature mean was more centrally located in ERTS signal-space than other signatures. It was surrounded on one flank by trees, another by soybeans, and a third flank by other agricultural types. The soybean signature had most ground-cover types on a single flank. Mixtures of signal values from separate flanks of corn are likely to be misrecognized as corn, but it is more difficult to encounter such a mixture that looks like soybeans. Since Huntington and Fayette test sections are known to have more tree areas than Livingston and Lee, even more mixture pixels were misrecognized as corn.

Beyond the above comments, it is difficult to say more about best times of year, since no single data set was available at many periods through the growing season.

5.2.3 PERFORMANCE FOR WHEAT RECOGNITION

Three early-June data sets were processed with wheat as the only major crop. Overall, 86% to 96% of the test field-center pixels were correctly assigned to the proper class. Performance on wheat pixels was not nearly so good, ranging from 23% to 53% correct on the specified wheat test pixels. The overall results are so high because the class "other" was accurately recognized and only about 7% of the test area was wheat, according to the ground truth. Furthermore, we consider the wheat field-center detection values to be incorrect and misleading (lower than actual) because of errors and uncertainties in the ground truth information for wheat test fields, as well as because of insufficient wheat training and test pixels.

As noted earlier, there were only four test wheat fields with 36 field-center pixels in Shelby County and eight test fields with 65 pixels in Fayette County. Of the 65 in Fayette, 37 were in one large field labeled as Field 29-29. We believe that only 10 of these 37 are valid wheat field-center pixels, so we recomputed field-center recognition performance on that basis and reported the revised results as well as those based on the prescribed field coordinates. The area in question clearly consists of three separate fields in post-June data sets. The photointerpreter lists it as wheat stubble flanked by two soybean fields. Analysis of the June ERTS data, both image and digital values, show that the side fields are distinct from the center in June as well as later. This places the validity of the entire set of test wheat fields in doubt. It is our understanding that the test wheat fields were obtained after the fact from farmers' recollections and were noted on aerial photographs. This was necessary because the only available aerial photography was collected after harvest and photointerpretation of wheat proved to be inaccurate. It could be that field boundaries were not accurately marked in the field or were transmitted inaccurately.

The 27 pixels in question comprise over 40% of the wheat test pixels in Fayette. When they were omitted from the two Fayette local field-center analyses, wheat recognition increased by 9% and 20%, while overall performance increased by 2%.

Ground truth proportions for Shelby are not given in Table 6 because we were told that acreages for some of the sections were in doubt and have not yet learned which they are. The recognized proportions listed are for all 20 test sections. We expect the proportion of wheat in Shelby to be close to that in Fayette, in which case the recognized proportion would be quite accurate. On the other hand, the recognized Fayette wheat proportion is double the ground-truth value. The reasons for the over recognition in Fayette probably are related to the signature characteristics discussed earlier in Sec. 4.1. One characteristic was that there was a greater variability in the maturity of the Fayette training data which produced a signature with a greater dispersion volume than the compact pattern in Shelby.

The greatest source of false wheat detections in Fayette field centers was the urban class (27-29 of 49 pixels), followed by soybeans, hay, and corn in lesser amounts. It would appear that mixture pixels prevalent in the urban areas are a major problem that could be avoided by proper stratification to eliminate urban areas from processing.

5.2.4 FULL-SECTION VS. FIELD-CENTER PERFORMANCE

The test design included both field-center and full-section processing so that the possibly degrading effects of pixels that represent mixtures of two or more ground covers could be separated from the simpler question of how well relatively "pure" samples of the major crops could be recognized and distinguished from their backgrounds. Most field-center pixels represented single classes of ground cover, but mixtures were present and caused problems in the woods/pasture category and the urban category, as discussed further in Sec. 6.6.

The specification of field-center pixels included a requirement for a minimum one-pixel border between each field-center pixel and the field boundary. This requirement was especially needed since nearest-neighbor operations were used to rotate and scale the data and to place later passes in spatial registration with the first (reference) data set.

In Sec. 5.2.2, a measure for evaluating full-section performance was defined. This measure was included in Table 6 and compiled more conveniently in Table 9 for corn/soybean data sets.

To help determine whether or not there was correlation between performances in field centers and full sections, we prepared Fig. 8 for the linear decision rule. There appears to be little or no correlation between good performance on field-center recognition and good performance on full-section recognition. Local recognition, marked by dots, shows no important pattern and non-local recognition (without adjustment), marked by X's, show a slight negative correlation, if any. The best data set for full-section proportion estimation (Shelby 7 Sept., coded P on Fig. 8) had lower than average field-center performance. It must be that compensating errors were made in the data so that the recognized proportions matched the true proportions. The analysis of variance effort will give a measure of the variance of these estimates and one would expect larger variances for the data sets where field-center accuracy is low. Another point that can be made is that the use of the same Shelby 7 Sept. signatures with the quadratic decision rule produced substantially poorer estimates of the crop proportions.

5.2.5 NON-LOCAL RECOGNITION PERFORMANCE, WITHOUT AND WITH PREPROCESSING

The capability to use signatures from one county or state and use them to obtain accurate recognition in another is desirable for large area survey operations. The CITARS data sets provided an opportunity to test this capability both without and with signature-extension preprocessing.

The tables presented earlier in this section have contained results obtained with non-local signatures as well as local ones. For example, the field-center results in Table 7 show that the application of unadjusted non-local signatures reduced overall correct recognition by an average of 12% for the linear decision rule, although six of twenty cases showed some improvement in recognition. The use of simple mean-level-adjustment preprocessing for signature extension caused, on the average, a 6% improvement in non-local recognition, although there were a few cases where a drop was noted (in only two cases did the drop exceed 5%, however). When the quadratic rule was used, the same pattern appeared, although local recognition accuracy was lower to start with, dropped only 8%, and the use of preprocessing more consistently improved non-local recognition performance.

For the few instances in which the same segment was viewed on successive days, local performances were comparable between days. However, there were differences in non-local performances with exchanged signatures.

The most dramatic changes in non-local performance occurred between the Fayette and White 21 August data sets, where the best local recognition results were obtained for field centers. Non-local results without adjustment were about half of the local results. With mean-level-adjusted Fayette signatures, field-center recognition in White rose by 34% to 74% correct. Only an 8% increase was obtained with adjusted White signatures on Fayette data.

Comparisons similar to those above can be made for the full-section performance measures presented in Table 9. Average non-local performance with the linear decision rule is degraded somewhat from that obtained locally (an increase of four percentage points in the RMS deviation).

However, the use of mean-level-adjustment preprocessing brought the average performance back to the local level. As with the field-center cases, there were some exceptions to these trends. When the quadratic rule was used, performance was worse than with the linear rule and there was no clear-cut difference between local and non-local recognition, either with or without preprocessing.

For the wheat data, there was little difference between the overall performance between the two Fayette data sets for local, non-local without adjustment, and non-local with adjusted signatures. Wheat field-center recognition improved when 11 June signatures were applied to 10 June data. Shelby overall performance degraded with non-local signatures from Fayette 10 June, and preprocessing with mean-level adjustment did not help. Extension from Fayette 11 June to Shelby 8 June was not prescribed, although 11 June was the registration reference data set for Fayette County.

5.2.6 PERFORMANCE DIFFERENCES BETWEEN SEGMENTS

A number of factors combine to cause variability between segments, including mix of crop types present, field sizes, differences (in soils, climate, farming practices, etc.) between counties, differential changes during the growing season, and differences between pixels used for training and those used for test.

Prior presentations of field-center recognition results (e.g., Fig. 6 and Table 6) have shown that the relative performances among the crops were by no means consistent between segments in the same time period. Sometimes corn showed best performance, sometimes soybeans, and sometimes "other".

The variability is present in all recognition results -- linear rule and quadratic, field center and full section, local and non-local. Only in July were data available for more than two segments for comparative analysis, except for a few second-day coverages of segments.

We have noted earlier that a pair of data sets for the same county on successive days invariably resulted in similar final signatures and similar local recognition results, even though small differences existed. Other pairs suffered the variability described above. For example, the pairs Fayette 16 and 17 July and Lee 17 and 18 July had overall local field-center recognition results (Table 7) within a few percent of each other, but performance for Huntington 15 July was appreciably lower and Livingston 16 July higher. Had this not been the case, one might have assumed that the training procedure and recognition performance might be too sensitive to small differences or random factors between such pairs of data sets. Rather there seem to be significant differences between data sets for different counties, causing the training procedure to respond quite differently. The variety of non-local recognition results, even after preprocessing, is another indication of differences between the segments.

In general the major crops had better percent correct recognition than other crops in all segments, although there are a few exceptions in July and August. One can see from Table 6 that full-section recognition consistently under-recognized pixels of the "other" class. Thus the ERIM procedures erred on the side of false alarms of major crops in "other" fields rather than false alarms of "other" in major crop fields. Furthermore, the use of quadratic decision rule rather than the linear rule accentuated this tendency even more.

Substantial differences exist between segments in the numbers of "other" signatures that were available for consideration, selected, and rejected. Higher performance for "other" occurred when a higher percentage of the other classes were accepted, but overall performance in field centers does not appear to be well correlated with performance on other-class pixels.

Wide variations in the relative signature dispersion volume (non-compactness of the signature, as measured by the determinant of the covariance matrix) were noted between segments and time periods. The unusually wide soybean signature in Fayette 16 and 17 July may be in part responsible for the large overestimate of the soybean proportion in full-section recognition. Also, within a given time period, the percent of other field centers correctly identified was quite consistently related to the compactness of the major-crop signatures. The most important thing, obviously, is where the test pixels happen to fall with respect to the signatures (both large and small).

5.2.7 LINEAR VS. QUADRATIC DECISION RULE

The CITARS processing results indicate that recognition by ERIM's "best linear" decision rule is slightly better than the more expensive and yet more widely used quadratic decision rule. This is shown for local and non-local processing and for full-section and field-center processing, by comparing the overall performance in each category averaged over all data sets.

The data sets probably exhibit too much variance of overall performance to support the hypothesis that the linear rule is significantly better than the quadratic rule (under CITARS conditions) with a high degree of confidence. However, the point is that even if they have the same performance, the 2/3 cost reduction of the linear rule is significant and is a distinct advantage of the linear rule. This result confirms other tests of the two rules conducted at ERIM.

One more difference between the two decision rules is that, for the CITARS processing, the quadratic rule over-assigned pixels to major crop classes and under-assigned pixels to the "other" class to a greater extent than did the linear rule. That is, the quadratic rule gave fewer other-class false alarms in major crop fields, but the linear rule gave fewer major-class false alarms in other fields.

The cost advantage of the linear rule accrues from the fact that a quadratic calculation to determine whether or not to threshold the pixel and assign it to the null class is made only for the "winning" signature, whereas the quadratic rule requires the calculation for every signature. Therefore, the more signatures used, the greater is the cost advantage of the linear rule. An average of four signatures was used in CITARS processing.

5.3 CONCLUSIONS AND RECOMMENDATIONS

Crop recognition performance varied throughout the growing season as crops matured. The best single time for recognizing corn, soybeans, and other field centers was late August when an 80% correct recognition rate was achieved on the prescribed test data.

Crop proportion estimates in full-section data were biassed in favor of the major crops. The presence of mixtures of two or more ground covers in individual pixels increased such errors, for example, mixtures of trees and other covers along woodlots and in urban areas frequently were recognized as corn; the elimination of urban and other non-agricultural areas from processing is recommended, wherever possible. Also, the parameters in our procedures were fixed early in the processing, primarily on the basis of field-center analyses; further development of procedures with a better balance of emphasis should reduce the observed bias.

The ERIM "best linear" decision rule was found to have a distinct cost advantage ($\sim$ 1/3 the cost) over the more conventional quadratic rule with equal or better recognition performance on the CITARS data.

Non-local recognition with unadjusted signatures produced a substantial average reduction in recognition performance from levels attained with local signatures. Signature extension preprocessing by mean level adjustment improved recognition performance on the average in both field centers and full sections. We recommend that other, more sophisticated signature extension techniques be evaluated, for example, the MASC procedure which is discussed in Sec. 6.

There was substantial variability in the amounts and types of pixels available for training in the various segments. In general, too few "other" pixels were available; the quantity of corn and soybean pixels usually was sufficient, but they often did not adequately represent the variability of the test data. The variability in soybean field maturities was greater than we had anticipated, and in some instances, our single signature for each major crop might better have been made multi-modal for soybeans. Our procedure for establishing "other" signatures should be revised because it is too dependent on ground truth identifications and the specific mix of other-crop pixels that happen to be available from training fields.

There were insufficient training and test pixels for wheat to make a reliable assessment of wheat recognition capability. Furthermore, the validity of some field identifications as transmitted to analysts is in question.

This raises another point regarding the procedures used to obtain ground truth. The randomization of choice of ground areas for periodic ASCS visits was made independent of field size. Consequently, many fields for which extensive ground truth is available were too small to extract ERTS field-center pixels for training. We recommend that field size be a factor in the choice of fields for ground visits in the future. Furthermore, studies should be made to establish criteria for determining how much training data is required for establishing signatures that are representative of the test data.

The specification of field-center pixels for analysis proved to be a more difficult task than many people expected. Although addressed in other

CITARS documentation, we note here that computer-assisted procedures, like the one developed and applied at ERIM on the CITARS data, are highly desirable.

Finally, supplementary analyses were performed on a limited set of CITARS data to better understand the sources of error and to explore the use of more advanced processing techniques. Results with the other techniques are presented in Sec. 6. We found a few corrections that should be made in the data base which resulted in the improvement of standard field center results by up to 20 percentage points for the sets analyzed. While discussed in detail in Sec. 6, they also are summarized below.

First, in retrospect we have discovered errors in crop identification in all data sets where we looked for them. A detailed examination of Fayette 21 August uncovered sufficient errors to raise overall field-center performance from 80% to 85% correct. Other segments with poorer performance may experience an even larger benefit from similar checking.

Second, in July and early August, test fields labeled "soybeans" varied in maturity on a continuum from mature to bare soil, and there were some test fields labeled bare soil. There is no way for ERTS to distinguish bare soil that was labeled soybeans from any other kind of bare soil, and recognition suffered errors whichever category was assigned to bare soil pixels. For Fayette 17 July for example, when we hand-tallied all bare pixels as belonging to category soybeans, recognition accuracy rose from 64% to 76%. When bare pixels were assigned to category other, recognition accuracy was 74%. In either case, the soybean signature was still large due to the wide variation of maturities.

Third, certain target categories are inherently mixtures of ground cover types, for example cities and woods-pasture. Others form problematic mixtures on their boundaries with agricultural covers, for example corn recognitions occur on forest boundaries. Deletion of urban areas from Fayette 17 July test data improved field-center recognition results to above 80% (including the improvement outlined in the above paragraph), for a total improvement of 16 to 20 percentage points.

There are other situations which have not had time and resources to investigate further. For instance, the use of non-specific "other" training and test pixels, i.e., pixels labeled 'other' instead of trees, pasture, etc. These areas quite possibly could include many mixture pixels. A major concern is the fact that there are 612 of such 'other' pixels in the Huntington test data with only 157 corn, 189 soybean, and 212 additional "other" pixels, and they have a strong impact on the field-center performances for the segment. Another item for exploration would be the marked change in White corn and soybean field signature patterns between 21 August and 7 September (Figs. 2, parts I-L).

In summary, the incompleteness of the ERTS coverage throughout the season and the variability in the data limit the number of definitive answers that can be given to the questions posed in the CITARS objectives. Perhaps most clear is that there is no accuracy penalty associated with use of the faster linear decision rule. The degradation in performance with unadjusted non-local signatures is quite clear. Mean level adjustment was shown to help somewhat on the average, and other signature extension methods should be investigated.

	EXT	TYPE	NO OTHER SIGS	TEST LINEAR FIELD CENTER C	S	O	AVG. OVER PTS.	FULL SECTION C	S	O	RMS DEV.	TEST QUADRATIC FIELD CENTER C	S	O	AVG. OVER PTS.	FULL SECTION C	S	O	RMS DEV.	TRAINING LINEAR FIELD CENTER C	S	O	AVG. OVER PTS.	TRAINING QUADRATIC FIELD CENTER C	S	O	AVG. OVER PTS.
FAY		G.T						19.7	29.3	51.0						19.7	29.3	51.0									
29 JUN	YY		8	69.0	61.8	69.6	66.7	33.5	17.6	48.9	10.5	82.3	68.2	40.8	62.1	48.7	25.8	25.5	22.4	77.9	45.2	92.0	76.9	82.4	65.2	90.3	82.1
(Y)																											
FAY		G.T.						19.7	29.3	51.0						19.7	29.3	51.0									
16 JUL	DD		4	93.4	65.4	37.5	63.2	29.3	41.5	29.2	15.5	94.8	66.8	23.3	59.0	31.1	50.6	18.3	23.5	80.0	74.8	96.7	87.5	81.4	77.9	95.0	87.7
(D)	02ED	NA	4	96.9	64.2	27.4	60.2	83.4	41.0	20.6	41.3	93.4	69.6	14.0	56.2	41.7	47.0	11.3	28.1	80.0	71.0	67.7	70.6	81.4	77.1	62.5	69.7
	02ED	MLA		97.3	62.8	45.8	66.6	35.5	35.9	28.6	16.3	95.5	67.6	18.6	57.8	40.3	44.5	15.2	25.4	81.4	67.9	89.2	81.9	84.3	74.0	76.5	77.0
	11FD	NA		16.1	2.0	66.6	29.3	18.1	3.4	78.5	21.8	78.0	9.2	27.4	35.3	33.3	21.3	45.5	9.6	27.1	10.7	77.7	50.4	65.7	19.8	74.1	57.1
	11FD	MLA		24.8	41.3	43.3	37.4	14.9	36.0	49.1	4.8	85.3	55.9	9.3	47.4	27.2	58.4	14.4	27.3	5.7	32.8	84.1	57.1	65.7	57.3	39.4	48.7
FAY		G.T.						19.7	29.3	51.0						19.7	29.3	51.0									
17 JUL	EE		4	96.2	64.2	40.6	64.5	36.6	38.5	24.9	18.7	96.5	71.8	25.1	61.6	40.1	47.4	12.4	27.3	90.0	62.6	90.8	82.2	88.6	67.6	88.4	82.2
(E)	01DE	NA	4	92.7	65.9	57.2	70.2	28.4	33.2	38.4	9.1	95.1	70.4	32.6	63.5	29.7	49.5	20.8	21.8	81.4	51.1	91.6	77.8	80.0	72.7	84.9	80.4
	01DE	MLA		90.6	73.5	31.0	62.7	31.2	42.6	26.2	17.6	90.6	77.7	17.9	59.3	30.4	54.6	15.0	26.1	72.9	66.9	77.3	73.5	74.3	77.0	70.9	73.3
	18GE	NA	2	94.4	80.4	33.2	67.0	37.1	43.3	19.6	22.2	96.2	81.6	18.4	62.5	40.4	50.3	9.3	29.5	81.4	72.7	61.8	68.0	80.0	75.5	59.8	67.6
	18GE	MLA		82.2	84.4	52.9	72.2	36.3	36.4	27.3	17.2	82.2	88.3	32.4	66.0	36.7	48.0	15.3	25.3	65.7	64.7	73.3	69.6	64.3	71.9	52.2	60.0
FAY		G.T.						19.7	29.3	51.0						19.7	29.3	51.0									
21 AUG	NN		4	87.4	85.5	69.8	80.3	36.8	28.6	34.6	13.7	87.8	87.4	61.5	78.0	40.2	30.6	29.2	17.3	90.0	75.0	98.9	87.4	88.6	78.1	97.9	88.0
(N)	23LN	NA	4	0.3	15.6	89.6	38.5	2.6	2.2	95.2	31.5	1.7	18.7	86.1	38.7	3.6	3.4	92.9	29.9	0.0	10.1	89.6	33.3	0.0	12.0	55.2	28.4
	23LN	MLA		23.8	19.8	88.2	46.1	15.8	10.2	74.0	17.4	26.2	23.2	83.7	46.3	18.1	13.2	68.8	13.9	34.3	15.1	100.0	54.0	40.0	18.8	99.0	55.9

Key: (1) Field-center results are in % correct recognition; full-section results are in % of total area.

(2) Crops: C = Corn, S = Soybeans, W = Wheat, O = Other

(3) G. T.: Ground truth proportion in segment.

(4) A letter (found beneath name) is associated with each data set. The double letter code in the column marked EXT (extension), e.g., AB, implies the signatures of data set A extended to data set B. If the letter is duplicated, e.g., AA, this implies local recognition. The associated number refers to the non-local task design number code (See Table 5).

(5) MLA: mean level adjustment

(6) NA: no adjustment

(7) RMS Dev.: See Table 9.

TABLE 6. SUMMARY OF ERIM CITARS PROCESSING RESULTS WITH LINEAR AND QUADRATIC DECISION RULES

TABLE 6
(Continued)

	EXT	TYPE	NO. OTHER SIGS	TEST LINEAR FIELD CENTER C	S	O	AVG. OVER PTS.	TEST LINEAR FULL SECTION C	S	O	RMS DEV.	TEST QUADRATIC FIELD CENTER C	S	O	AVG. OVER PTS.	TEST QUADRATIC FULL SECTION C	S	O	RMS DEV.	TRAINING LINEAR FIELD CENTER C	S	O	AVG. OVER PTS	TRAINING QUADRATIC FIELD CENTER C	S	O	AVG. OVER PTS.
HUN		G.T						18.7	22.3	59.0						18.7	22.3	59.0									
15 JUL	GG		2	68.8	83.1	49.2	57.3	32.8	36.4	30.8	19.9	68.8	87.3	34.5	47.6	40.1	42.6	17.3	29.5	77.4	86.2	91.3	86.2	87.1	92.5	91.3	91.1
(G)	08 JG	NA	1	62.4	29.1	52.5	50.1	27.0	31.0	42.0	12.0	72.0	32.8	45.1	46.8	31.7	34.2	34.1	17.6	51.6	27.7	93.3	53.2	72.6	27.0	84.6	54.2
	08JG	MLA		51.6	36.5	59.3	54.6	18.4	30.5	51.0	6.6	68.8	37.0	53.4	52.8	23.9	32.2	43.9	10.8	16.1	27.0	93.3	46.2	41.9	27.7	90.4	50.5
	17EG	NA	4	76.4	59.3	62.3	63.7	28.8	34.9	36.3	16.1	84.1	78.8	28.2	43.8	33.0	54.1	12.9	33.4	61.3	38.4	87.5	58.5	64.5	57.2	59.6	59.4
	17IG	MLA		75.2	74.1	34.0	46.0	36.2	48.4	15.4	31.0	77.7	77.8	17.5	35.3	39.2	53.8	7.0	37.0	80.6	43.4	52.9	53.5	80.6	56.0	28.8	53.0
HUN		G.T						18.7	22.3	59.0						18.7	22.3	59.0									
24 SEP	RR		2	79.6	38.6	48.4	51.0	45.7	27.0	27.3	24.2	77.1	39.7	30.3	38.1	40.2	43.7	16.1	30.3	91.9	59.1	86.5	74.2	95.2	73.0	85.6	81.2
(R)	15SR	NA	5	17.8	59.8	50.7	47.8	11.3	27.6	61.1	5.4	22.3	48.7	50.7	46.6	17.5	27.0	55.5	3.5	14.5	18.2	88.5	40.0	24.2	22.6	88.5	44.0
	15SR	MLA		63.7	63.5	39.0	46.2	28.0	29.5	42.6	11.7	69.4	56.6	38.2	45.4	33.0	26.9	40.2	13.9	41.9	27.0	91.3	50.5	58.1	26.4	86.5	51.7
SHE		G.T						37.8	24.4	37.8						37.8	24.4	37.8									
7 SEP	PP		5	60.2	55.3	46.0	56.5	41.5	27.4	31.1	4.7	75.9	49.4	35.0	62.4	56.4	21.7	21.9	14.2	57.3	88.2	85.5	70.1	76.2	90.2	81.6	80.1
(P)	13QP	NA	2	97.6	33.3	25.0	70.6	75.2	11.7	13.1	26.9	98.9	33.8	23.4	71.2	79.1	12.0	8.9	30.0	91.5	60.8	26.3	69.1	92.7	66.7	2.6	64.6
	13QP	MLA		89.5	48.9	31.5	70.3	62.0	25.8	12.2	20.4	91.7	48.5	10.3	67.9	68.5	26.0	5.5	25.7	80.5	74.5	39.5	68.7	84.1	74.5	5.3	61.9
SHE		G.T						37.8	24.4	37.8						37.8	24.4	37.8									
24 SEP	SS		5	49.4	59.5	56.5	52.9	24.4	39.0	36.6	11.5	48.4	55.3	45.0	49.3	31.4	35.7	32.9	8.0	35.6	85.5	84.1	56.5	42.5	80.0	81.0	58.6
(S)	16RS	NA	2	93.7	5.5	19.0	60.4	75.8	14.7	9.5	27.9	68.3	22.3	10.5	47.5	62.8	34.0	3.2	25.3	90.6	9.1	27.0	60.1	80.0	21.8	11.1	
	16 RS	MLA		78.7	29.1	24.0	57.6	44.3	41.9	13.9	17.5	16.6	41.6	10.0	50.6	39.0	54.8	6.1	25.4	57.5	43.6	38.1	50.4	56.3	58.2	11.1	46.4
WHI		G.T.						36.4	31.0	32.6						36.4	31.0	32.6									
21 AUG	LL		4	71.7	77.5	90.4	78.4	30.3	25.0	44.8	8.6	75.0	78.5	88.3	79.6	30.7	26.6	42.7	7.2	71.7	86.2	99.0	79.3	77.5	76.7	99.1	83.3
(L)	24NL	NA	4	11.5	57.5	62.2	40.0	15.8	20.6	63.6	22.3	9.9	73.5	50.3	41.8	20.1	26.3	53.5	15.5	1.9	64.7	74.0	28.5	1.7	77.1	62.5	30.3
	24 NL	MLA		76.0	83.8	58.9	74.3	41.1	35.9	22.9	6.8	76.4	86.7	42.9	71.4	44.8	39.1	16.1	11.7	81.4	87.2	57.7	79.9	80.0	88.1	49.0	78.1
WHI		G.T						36.4	31.0	32.6						36.4	31.0	32.6									
7 SEP	QQ		2	82.0	73.3	18.8	63.0	54.1	31.4	14.5	14.6	81.8	74.1	18.5	63.1	55.2	32.5	12.2	16.0	86.5	83.9	39.5	79.2	86.3	83.9	37.7	78.8
(Q)	14PQ	NA		32.8	71.8	23.6	43.7	25.3	34.3	40.5	8.1	49.5	67.2	17.3	47.3	43.9	28.3	27.8	5.4	37.7	73.5	48.2	47.9	51.0	63.5	35.1	51.9
	14PQ	MLA		44.1	70.5	4.8	43.1	30.7	31.9	37.4	4.3	71.0	65.5	3.6	52.0	51.7	27.8	20.4	11.4	60.3	70.6	9.6	56.0	76.8	65.1	6.1	63.6

Table 6 (Continued)

	EXT	TYPE	NO OTHER SIGS	TEST LINEAR FIELD CENTER C	S	O	AVG. OVER PTS.	FULL SECTION C	S	O	RMS DEV.	TEST QUADRATIC FIELD CENTER C	S	O	AVG. OVER PTS.	FULL SECTION C	S	O	RMS DEV.	TRAINING LINEAR FIELD CENTER C	S	O	AVG. OVER PTS.	TRAINING QUADRATIC FIELD CENTER C	S	O	AVG. OVER PTS.
LIV		G.T.						32.5	37.7	29.8						32.5	37.7	29.8									
16 JUL	FF		2	55.0	85.0	68.6	72.4	30.8	49.6	19.6	9.1	59.4	83.9	29.4	63.3	36.4	54.5	9.1	15.6	49.1	84.3	54.5	66.4	55.5	84.3	51.2	67.6
(F)	12DF	NA	4	24.9	33.3	39.2	32.4	15.2	31.9	52.9	17.0	21.1	72.9	16.2	44.0	13.2	70.6	16.2	23.4	11.0	27.0	68.2	31.2	10.4	82.3	51.2	52.4
	12DF	MLA		13.9	73.1	21.0	43.3	3.7	65.0	31.3	22.9	18.1	79.4	15.3	46.0	5.3	72.2	22.6	25.7	11.0	70.6	54.5	48.0	17.9	78.2	46.3	51.8
	05GF	NA	2	73.1	21.9	20.5	36.0	38.0	26.0	36.0	8.3	73.3	24.8	6.6	34.0	59.1	32.4	8.6	19.9	50.3	23.0	57.7	39.5	54.3	37.1	46.3	44.7
	05GF	MLA		64.7	70.0	9.6	53.4	25.3	53.3	21.4	11.0	67.3	62.9	5.9	49.9	33.8	56.4	9.8	15.8	44.5	56.5	23.6	45.2	44.5	54.8	23.6	44.5
	07JF	NA	1	5.4	33.9	32.6	25.5	11.8	33.2	55.0	19.0	35.5	42.6	29.4	37.3	20.3	37.6	42.1	10.0	4.0	25.4	65.8	27.7	34.1	34.7	57.7	39.7
	07JF	MLA		62.9	68.4	53.3	63.0	29.9	50.2	19.9	9.3	65.3	67.9	52.8	63.4	30.1	51.9	18.0	10.7	43.9	48.8	43.9	46.1	48.0	47.6	43.1	46.7
LIV		G.T.						32.5	37.7	29.8						32.5	37.7	29.8									
3 AUG	OO		3	77.0	43.3	87.9	60.7	34.2	25.2	40.6	9.6	85.4	56.3	58.1	65.9	38.2	35.7	26.1	4.1	59.5	55.2	98.0	64.9	71.1	67.7	96.0	74.3
(O)	10TO	NA	4	16.5	56.9	89.3	48.7	6.1	41.8	52.2	20.1	20.2	62.0	88.8	52.6	8.1	44.9	47.1	17.8	10.4	71.8	86.1	55.5	12.7	74.6	81.3	56.4
	10TO	MLA		24.7	48.1	90.2	46.8	8.8	30.5	60.8	22.9	27.9	53.2	90.2	50.6	10.5	33.8	55.7	19.7	19.7	58.1	87.8	52.6	20.8	65.3	83.7	55.3
LEE		G.T.						33.4	28.5	38.1						33.4	28.5	38.1									
17 JUL	II		1	63.4	89.0	17.4	68.5	35.4	55.0	9.6	22.5	59.2	90.8	15.2	67.5	34.9	57.6	7.6	24.4	61.7	91.8	79.0	77.7	65.6	90.0	78.4	77.8
(I)	04JI	NA	1	37.9	75.0	24.2	53.4	31.2	42.8	25.9	10.9	44.5	79.8	23.1	57.9	32.6	45.1	22.3	13.2	42.2	82.7	83.4	74.8	59.4	80.9	77.1	74.2
	04JI	MLA		49.1	86.3	19.2	62.2	28.0	55.5	16.5	20.2	49.3	87.3	18.7	62.7	28.1	57.7	14.2	22.0	56.3	91.8	80.8	77.7	55.5	91.8	78.4	76.1
LEE		G.T.						33.4	28.5	38.1						33.4	28.5	38.1									
18 JUL	JJ		1	59.7	89.5	14.8	67.0	30.2	59.4	10.4	24.0	62.1	89.8	15.4	68.1	31.3	59.2	9.5	24.3	64.9	88.2	75.8	75.7	67.2	87.3	75.5	75.8
(J)	031J	NA	1	70.5	98.4	5.5	73.9	25.9	71.3	2.7	32.4	73.8	97.0	1.1	73.8	29.1	69.7	1.1	32.1	67.9	92.7	63.6	69.5	70.2	92.7	54.3	64.2
	031J	MLA		73.1	89.8	11.0	71.6	38.3	54.4	7.3	23.4	70.5	90.9	9.9	71.0	40.0	55.4	4.6	25.1	77.1	86.4	70.7	74.7	80.2	82.7	71.2	75.0
	06GJ	NA	2	49.1	78.0	32.4	60.2	14.1	46.1	39.8	15.1	73.6	60.8	13.2	58.4	40.0	43.3	16.7	15.5	48.9	79.1	47.0	52.9	65.6	75.5	23.5	41.1
	06GJ	MLA		55.7	86.1	28.6	66.0	16.3	52.6	31.1	17.5	68.5	78.9	15.4	65.4	28.4	53.2	18.5	18.4	52.7	81.8	38.4	48.8	61.8	83.6	22.2	47.0
LEE		G.T.						33.4	28.5	38.1	0					33.4	28.5	38.1									
5 AUG	TT		4	41.2	59.2	78.6	55.4	18.8	28.5	52.7	11.9	45.6	62.2	72.5	57.5	23.3	31.6	45.1	7.3	52.7	63.6	84.1	74.1	59.5	70.9	81.1	74.9
(T)	090T	NA	3	96.3	50.4	46.2	67.1	55.6	23.0	21.3	16.4	85.0	47.3	28.6	58.7	58.1	28.8	13.1	20.3	90.1	54.5	37.9	51.5	89.3	48.2	30.8	45.8
	050T	MLA		93.8	56.8	45.6	69.0	51.6	27.4	21.0	14.4	93.2	56.4	29.1	66.1	55.8	32.0	12.2	19.9	87.0	63.6	35.4	50.9	88.5	55.5	31.1	47.1

TABLE 6
(Continued)

	EXT TYPE	NO OTHER SIGS	TEST LINEAR FIELD CENTER W	TEST LINEAR FIELD CENTER O	TEST LINEAR FIELD CENTER AVG OVER PTS	TEST LINEAR FULL SECTION W	TEST LINEAR FULL SECTION O		TEST QUADRATIC FIELD CENTER W	TEST QUADRATIC FIELD CENTER O	TEST QUADRATIC FIELD CENTER AVG OVER PTS	TEST QUADRATIC FULL SECTION W	TEST QUADRATIC FULL SECTION O	TEST QUADRATIC FULL SECTION O	TRAINING LINEAR FIELD CENTER W	TRAINING LINEAR FIELD CENTER O	TRAINING LINEAR AVG. OVER PTS.	TRAINING QUADRATIC FIELD CENTER W	TRAINING QUADRATIC FIELD CENTER O	TRAINING QUADRATIC AVG. OVER PTS.
FAY	G.T					7.6 7.2	92.4 92.8	20 sec 26 sec				7.6 7.2	92.4 92.8	20 sec 26 sec						
11 JUN	AA	1	47.7	90.2	87.2	15.6	84.4		49.2	89.2	86.3	16.8	83.2		81.3	97.6	96.0	81.3	97.6	96.0
(A)	21 BA NA	1	43.1	87.2	84.0	19.1	80.9	26 sec	46.2	86.5	83.6	20.7	29.3		38.3	97.4	96.0	83.3	97.4	96.0
	21 BA MLA		40.0	88.0	84.6	16.5	83.5	26 sec	46.2	87.2	84.3	19.2	80.8		79.2	98.1	96.2	83.3	97.9	96.4
Revised FIELD 29-29	AA	1	63.2	90.2	89.1				65.8	89.2	88.2									
FAY	G.T					7.6 7.2	92.4 92.8	20 sec 26 sec				7.6 7.2	92.4 92.8	20 sec 26 sec						
10 JUN	BB	1	23.1	91.8	87.4	12.6	87.4		26.2	90.9	86.7	15.2	84.8		87.5	97.5	96.5	89.6	97.5	96.7
(B)	22 AB NA	1	40.0	90.8	87.5	12.2	87.8	26 sec	35.4	89.9	86.4	12.9	87.1		83.3	94.9	93.8	83.3	94.9	93.8
	22 AB MLA		46.2	89.0	86.2	15.2	84.8		43.1	87.9	85.1	16.4	83.6		87.5	94.3	93.6	87.5	94.5	93.8
	19 CB NA	0	41.5	85.6	87.0	12.3	87.7		41.5	87.5	88.6	12.3	87.7		41.7	91.9	87.4	41.7	95.6	90.3
	22 CB MLA		29.2	89.8	86.0	16.3	83.7		29.2	89.8	86.0	16.4	83.6		20.8	94.0	86.7	20.8	94.0	86.7
	15 CB																			
REVISED FIELD 29-29	BB	1	31.6	91.8	89.5				36.8	90.9	88.8									
SHE	G.T																			
8 JUN	CC	0	52.8	97.5	96.0	7.9	92.1		52.8	97.5	96.0	7.9	92.1		100	99.2	99.3	100	92.0	99.3
(C)	20 BC NA	1	77.8	88.7	88.4	11.1	88.9		83.3	88.1	88.0	14.1	85.9		42.3	98.0	92.8	42.3	97.6	92.4
	20 BC MLA		52.8	90.0	88.7	7.5	92.5		55.6	89.4	88.2	9.9	90.1		15.4	99.2	91.4	15.4	98.0	90.3

TABLE 7. COMPARISON OF OVERALL LOCAL AND NON-LOCAL FIELD-CENTER RECOGNITION ACCURACIES ON DATA SETS FOR WHICH CORN AND SOYBEANS ARE DESIGNATED AS MAJOR CROPS.

		PERCENTAGE OF FIELD-CENTER PIXELS ASSIGNED TO CORRECT CLASS					
		LINEAR RULE			QUADRATIC RULE		
PROCESSED DATA SET	NON-LOCAL TRAINING DATA SET	LOCAL	NON-LOCAL	NON-LOCAL WITH PREPROC.*	LOCAL	NON-LOCAL	NON-LOCAL WITH PREPROC.*
LIV 16 JUL	FAY 16 Jul	72.4	32.4	43.3	63.3	44.0	46.0
"	HUN 15 Jul	"	36.0	53.4	"	34.0	49.9
"	LEE 18 Jul	"	25.5	63.0	"	37.3	63.4
LIV 3 Aug	LEE 5 Aug	60.7	48.7	46.8	65.9	52.6	50.6
LEE 17 Jul	LEE 18 Jul	68.5	53.4	62.2	67.5	57.9	62.7
LEE 18 Jul	LEE 17 Jul	67.0	73.9	71.6	68.1	73.8	71.0
"	HUN 15 Jul	"	60.2	66.0	"	58.4	65.4
LEE 5 Aug	LIV 3 Aug	55.4	67.1	69.0	57.5	58.7	66.1
FAY 16 Jul	FAY 17 Jul	63.2	60.2	66.6	59.0	56.2	57.8
"	LIV 16 Jul	"	29.3	37.4	"	35.3	47.4
FAY 17 Jul	FAY 16 Jul	64.5	70.2	62.7	61.6	63.5	59.3
"	HUN 15 Jul	"	67.0	72.2	"	62.5	66.0
FAY 21 Aug	WHI 21 Aug	80.3	38.5	46.1	78.0	38.7	46.3
HUN 15 Jul	LEE 18 Jul	57.3	50.1	54.6	47.6	46.8	52.8
"	FAY 17 Jul	"	63.7	46.0	"	43.8	35.3
HUN 24 Sep	SHE 24 Sep	51.0	47.8	46.2	38.1	46.6	45.4
SHE 7 Sep	WHI 7 Sep	56.2	70.6	70.3	43.3	71.2	67.9
SHE 24 Sep	HUN 24 Sep	52.9	60.4	57.6	49.3	47.5	50.6
WHI 21 Aug	FAY 21 Aug	78.4	40.0	74.3	79.6	41.8	71.4
WHI 7 Sep	SHE 7 Sep	63.0	43.7	43.1	63.1	47.3	52.0
Average		64.4	51.9	57.6	58.7	50.9	56.4

* Preprocessing by mean-level adjustment

TABLE 8. PHYSICAL AND BIOLOGICAL CHARACTERISTICS OF SELECTED SOYBEAN FIELDS, FAYETTE COUNTY, ILLINOIS, 1973

Characteristic	Field / Date	VALUE FOR INDICATED FIELD: 35-13	44-12	55-43	64-63	69-41	69-49	88-66
Height (inches)	Jun 10/11*	3	2	3	0	0	0	0
	Jun 29/30*	10	6	8	1	6	6	4
	Jul 16/17*	24	10	20	4	12	12	12
	Aug 3/4	30	30	30	12	30	30	26
	Aug 21/22*	36	36	40	22	36	38	38
	Sept 8/9	36	36	40	26	36	38	38
	Sept 26/27	0	36	40	28	36	38	38
Ground Cover (%)	Jun 10/11	0-5	0-5	0-5	Bare	0-5	0-5	Bare
	Jun 29/30	5-20	5-20	0-5	0-5	5-20	5-20	0-5
	Jul 16/17	80-100	20-50	20-50	0-5	20-50	20-50	20-50
	Aug 3/4	80-100	80-100	80-100	5-20	50-80	50-80	80-100
	Aug 21/22	80-100	80-100	80-100		80-100	80-100	90-100
	Sept 8/9	50-80	80-100	80-100	80-100	80-100	80-100	80-100
	Sept 26/27	0	80-100	80-100	80-100	80-100	80-100	80-100
Stage of Maturity (See Below)	Jun 10/11	1	1	1	0	0	0	0
	Jun 29/30	1	1	1	1	1	1	1
	Jul 16/17	2	1	1	1	1	1	1
	Aug 3/4	2	3	2	1	3	3	3
	Aug 21/22	4	3	4	2	3	3	3
	Sept 8/9	5	5	5	3	5	5	5
	Sept 26/27	21	6	6	5	5	6	6

*ERTS-1 coverage obtained

Stage of Maturity Key:

1 = Pre-bloom
2 = Blooming
3 = Early Pod Set
4 = Late Pod Set
5 = Turning yellow, leaves dropping
6 = Mature
21 = Harvested

TABLE 9. COMPARISON OF OVERALL LOCAL AND NON-LOCAL FULL-SECTION RECOGNITION PERFORMANCE ON DATA SETS FOR WHICH CORN AND SOYBEANS ARE DESIGNATED AS MAJOR CROPS

PROCESSED DATA SET	NON-LOCAL TRAINING DATA SET	RMS DEVIATION+ FROM TRUE CROP PROPORTIONS (PERCENT) — LINEAR RULE: LOCAL	LINEAR RULE: NON-LOCAL	LINEAR RULE: NON-LOCAL WITH PREPROC.*	QUADRATIC RULE: LOCAL	QUADRATIC RULE: NON-LOCAL	QUADRATIC RULE: NON-LOCAL WITH PREPROC.*
LIV 16 Jul	FAY 16 Jul	9.1	17.0	22.9	15.6	23.4	25.7
"	HUN 15 Jul	"	8.3	11.0	"	19.9	15.8
"	LEE 18 Jul	"	19.0	9.3	"	10.0	10.7
LIV 3 Aug	LEE 5 Aug	9.6	20.1	22.9	4.1	17.8	19.7
LEE 17 Jul	LEE 18 Jul	22.5	10.9	20.2	24.4	13.2	22.0
LEE 18 Jul	LEE 17 Jul	24.0	32.4	23.4	24.3	32.1	25.1
"	HUN 15 Jul	"	15.1	17.5	"	15.5	18.4
LEE 5 Aug	LIV 3 Aug	11.9	16.4	14.4	7.3	20.3	19.9
FAY 16 Jul	FAY 17 Jul	15.5	41.3	16.3	23.5	28.1	25.4
"	LIV 16 Jul	"	21.8	4.8	"	9.6	27.3
FAY 17 Jul	FAY 16 Jul	18.7	9.1	17.6	27.3	21.8	26.1
"	HUN 15 Jul	"	22.2	17.2	"	29.5	25.3
FAY 21 Aug	WHI 21 Aug	13.7	31.5	17.4	17.3	29.9	13.9
HUN 15 Jul	LEE 18 Jul	19.9	12.0	6.6	29.5	17.6	10.8
"	FAY 17 Jul	"	16.1	31.0	"	33.4	37.0
HUN 24 Sep	SHE 24 Sep	24.2	5.4	11.7	30.3	3.5	13.9
SHE 7 Sep	WHI 7 Sep	2.8	26.9	20.4	14.4	30.0	25.7
SHE 24 Sep	HUN 24 Sep	11.5	27.9	17.5	8.0	25.3	25.4
WHI 21 Aug	FAY 21 Aug	8.6	23.3	6.8	7.2	15.5	11.7
WHI 7 Sep	SHE 7 Sep	14.6	8.1	4.3	16.0	5.4	11.4
Average		15.5**	19.2	15.7	19.2**	20.1	20.6

\+ Note that small values are best; values were computed as: $\text{RMS Deviation} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(p_i - \hat{p}_i)^2} \times 100\%$

where p_i = true proportion of crop i in data set (from ground truth)

$\hat{p}_i$ = proportion of area recognized as crop i

n = number of crop classes

* Preprocessing by mean-level adjustment.

** Weighted by number of non-local data sets present.

TABLE 10. RATIO OF RECOGNIZED TO TRUE CROP PROPORTIONS FOR LINEAR AND QUADRATIC RULES IN THOSE SEGMENTS FOR WHICH CORN AND SOYBEANS ARE MAJOR CROPS. (LOCAL PROCESSING PROCEDURES)

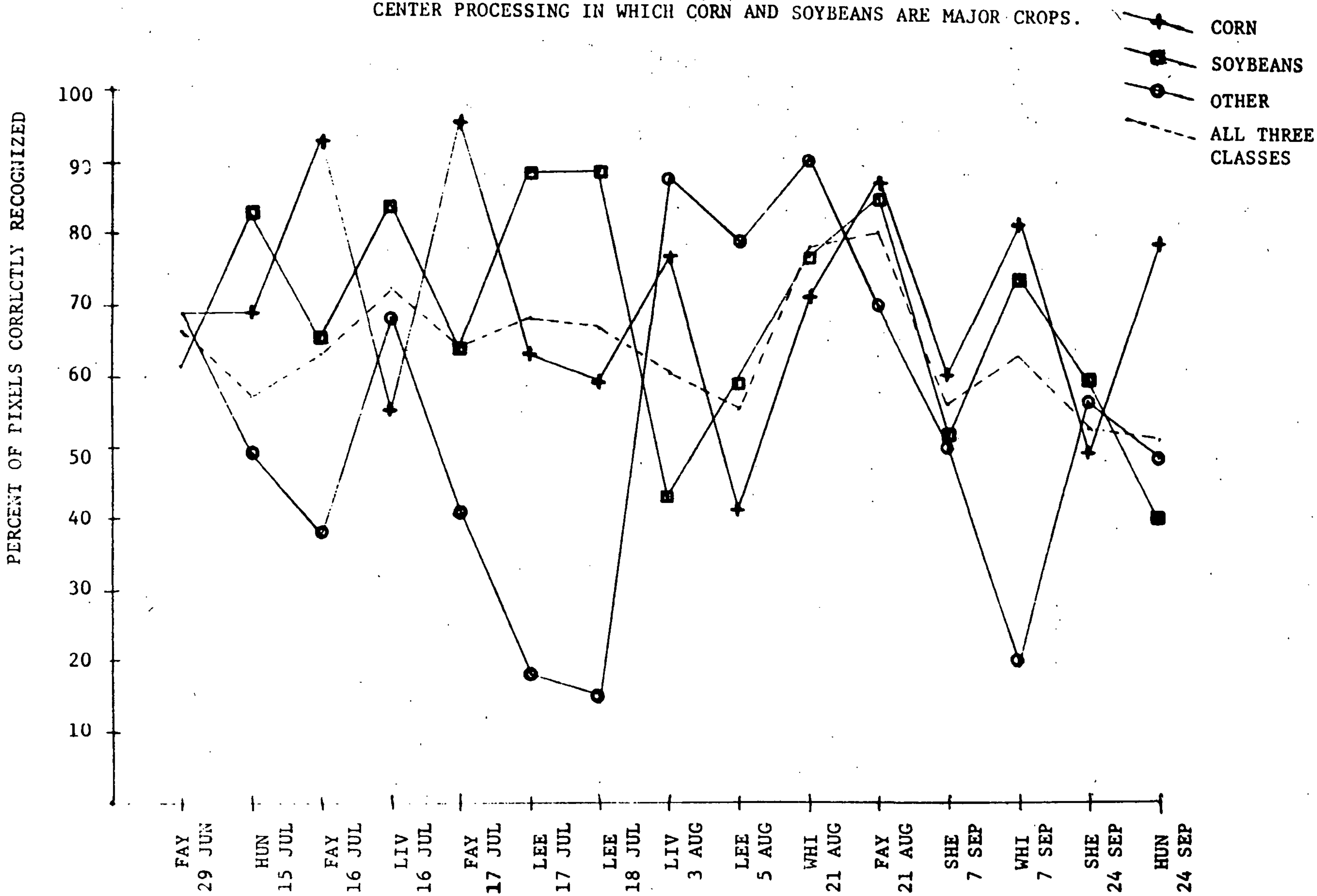
	LINEAR RULE			QUADRATIC RULE		
	CORN	SOY	OTHER	CORN	SOY	OTHER
HUN 15 Jul	1.75	1.63	.52	2.14	1.91	.29
HUN 24 Sep	2.44	1.21	.46	2.15	1.96	.27
SHE 7 Sep	1.10	.98	.92	.86	.41	1.52
SHE 24 Sep	.65	1.60	.97	.83	1.46	.87
WHI 21 Aug	.83	.81	1.37	.84	.86	1.31
WHI 7 Sep	1.49	1.01	.44	1.52	1.05	.37
LIV 16 Jul	.95	1.32	.66	1.12	1.45	.31
LIV 3 Aug	1.05	.67	1.36	1.18	.95	.88
FAY 29 Jun	1.70	.60	.96	2.47	.88	.50
FAY 16 Jul	1.49	1.42	.57	1.58	1.73	.36
FAY 17 Jul	1.86	1.31	.49	2.04	1.62	.24
FAY 21 Aug	1.87	.98	.68	2.04	1.04	.57
LEE 17 Jul	1.06	1.93	.25	1.04	2.02	.20
LEE 18 Jul	.90	2.08	.27	.94	2.08	.25
LEE 5 Aug	.56	1.00	1.38	.70	1.11	1.18

TABLE 11. SUMMARY OF FIELD SIZES IN ASCS-IDENTIFIED QUARTER SECTIONS

COUNTY		CORN	SOY	WHEAT	OTHER	TOTAL
	ACRES	1498	813	36	620	3550
Lee	NO. FIELDS	42	31	2	34	160
	AVG. SIZE	35.6	26.2	18.0	18.2	22.1
	ACRES	1239	1073	39	569	2969
Livingston	NO. FIELDS	33	27	2	33	87
	AVG. SIZE	37.5	39.7	19.5	17.2	34.1
	ACRES	733	287	416	1358	3193
Fayette	NO. FIELDS	37	11	26	92	217
	AVG. SIZE	19.8	26.0	16.0	14.7	14.7
	ACRES	1836	510	38	954	3753
White	NO. FIELDS	42	13	2	41	146
	AVG. SIZE	43.7	39.2	19.0	23.3	25.7
	ACRES	1888	540	323	753	3648
Shelby	NO. FIELDS	71	24	15	61	189
	AVG. SIZE	26.5	22.5	21.5	12.3	19.3
	ACRES	831	618	63	986	2756
Huntington	NO. FIELDS	39	25	6	54	148
	AVG. SIZE	21.2	24.7	10.4	18.3	18.6

From Ref. 2.

FIGURE 6. LINEAR RULE RECOGNITION RESULTS FOR ALL LOCAL FIELD-CENTER PROCESSING IN WHICH CORN AND SOYBEANS ARE MAJOR CROPS.



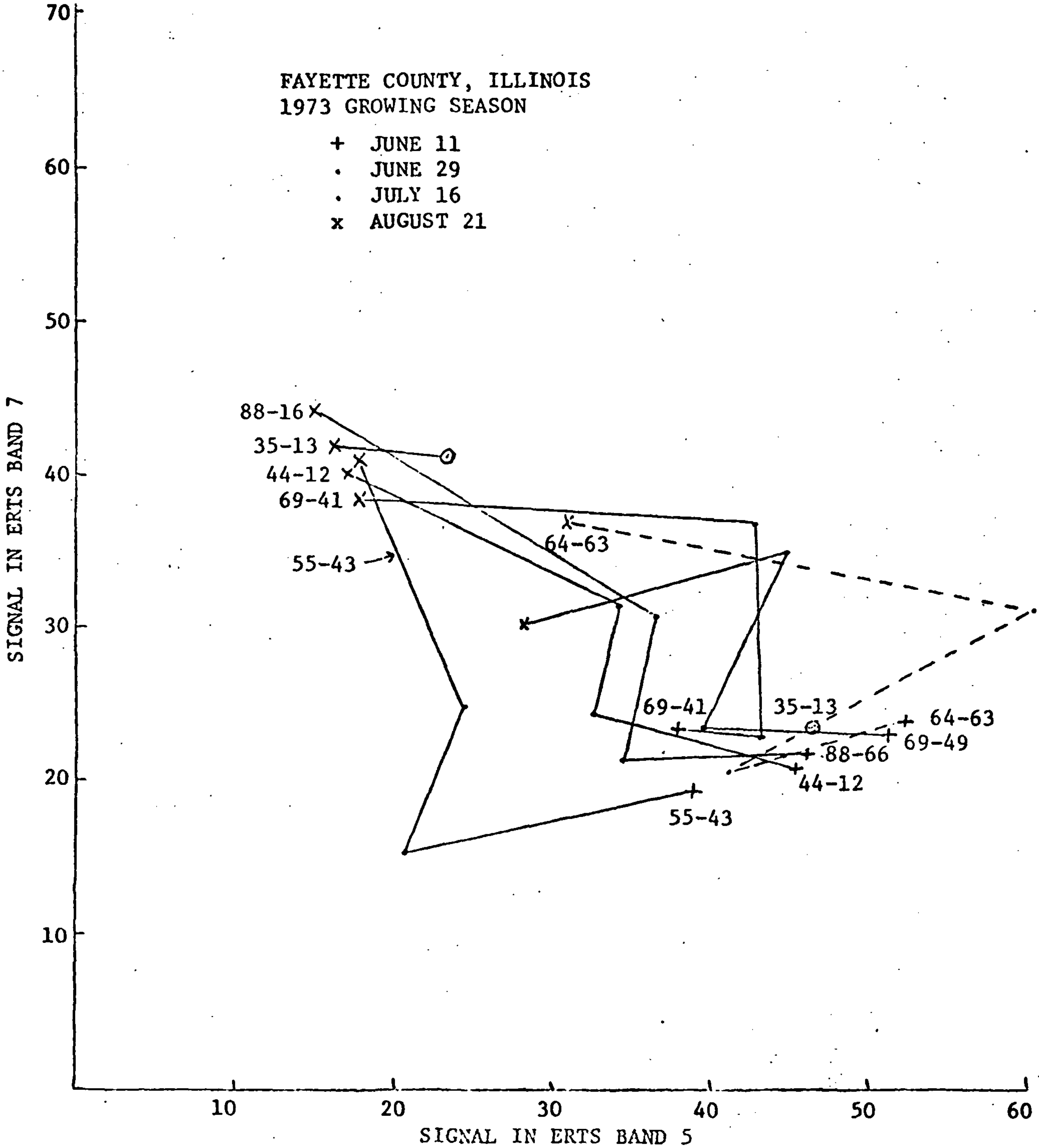


FIGURE 7. TEMPORAL PATHS OF ERTS SIGNATURE MEANS FOR SOYBEAN FIELDS

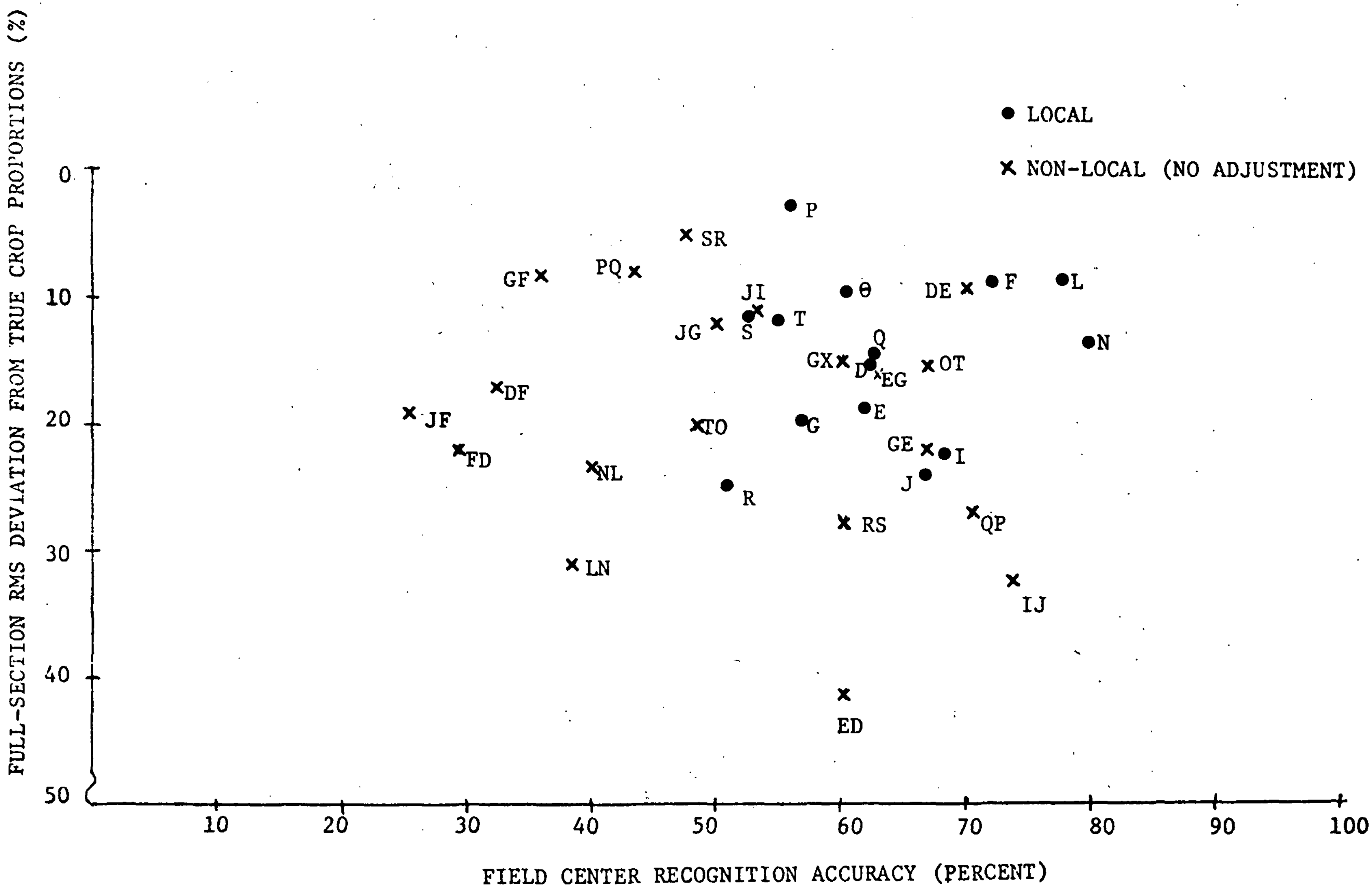


FIGURE 8. COMPARISON OF FIELD-CENTER AND FULL-SECTION RECOGNITION RESULTS WITH LINEAR DECISION RULE

6
SUPPLEMENTARY ANALYSIS AND PROCESSING

Supplementary analysis and processing steps were carried out at ERIM on a subset of the CITARS data to gain better understandings both of the reasons for some lower than expected recognition performances obtained with the standardized procedures utilized for CITARS and of the maximum performances that might be expected. The latter aspect was explored through the use of both additional data for training and developmental processing techniques that are more advanced than those defined for use on CITARS. Major parts of the supplementary effort were carried out under supporting Research and Technology tasks at ERIM other than the CITARS task.

The effort was directed primarily at the Fayette segment, with emphasis on the 21 August pass. For this time period, standard recognition performance for corn, soybeans, and other was the best, averaging 80% correct on the CITARS-defined field centers with our linear decision rule and CITARS procedures. This number does not fully reflect either the field-center accuracy achieved or that which potentially is achievable with this data set, as summarized below and discussed later in this section.

Increases up to 85% correct were computed after errors in the crop identifications assigned to some pixels were corrected and pixels that represented mixtures of two or more ground cover types were eliminated from the field-center data set.

A major problem affecting recognition performance appears to have been inadequate or insufficient data for training. In order to estimate the full potential for recognition on this 21 August data set, we trained and tested on all available field-center pixels, both ASCS-ground-truthed (CITARS training) and photointerpreted (CITARS test). Recognition results up to 90% correct were obtained with various approaches for this single time period.

Multi-temporal procedures also were applied to the corrected CITARS-defined Fayette field-center data. Seven channels from four time periods were used to obtain 85% correct with ASCS training fields only, 87% correct with two additional soybean signatures obtained from six test fields, and 94% correct when all field-center pixels were used for establishing the signatures.

Signature extension between the Fayette 21 August segment and the White 21 August segment with the CITARS mean-level-adjustment procedure had results that differed substantially, depending on which segment's signatures were adjusted and applied to the other segment. In both cases, the percentage correct for non-local field-center recognition with unadjusted signatures was about half that obtained with local signatures. Although mean-level adjustment of Fayette signatures produced results comparable to those with local signatures in White, mean-level adjustment of White signatures produced only a small improvement over the non-adjusted White signatures in Fayette. The application of a signature extension procedure newly developed at ERIM produced a result in Fayette with White signatures that equalled the local-signature result. The procedure is called MASC for Multiplicative and Additive Signature Correction.

The supplementary analyses did not fully consider the question of estimating the proportions of crops in the full sections, because of time and resource limitations; however, some effort was expended. The standard CITARS procedures did a good job of estimating the proportion of soybeans in the aggregate of the 20 test sections. However, the estimated corn proportion was nearly double the true proportion, with the proportion of "other" being correspondingly lower. An analysis was made of the ground covers in sections where corn was substantially over-estimated. These sections were found preponderantly to contain much trees, brush, and/or urban areas. The problem appears to be that pixels which represent a mixture of trees and another ground cover are frequently misrecognized as corn. The use of multitemporal data gave more accurate proportion estimates with lower RMS deviation.

A very recent development of ERIM processing techniques investigations, namely the nine-point mixutres algorithm, was tested on the Fayette 21 August test sections. This algorithm is a combination of one of our newer mixtures estimation (pixel proportion estimation) algorithms and the ERIM multi-element decision rule. The application of this technique improved the corn and "other" proportion estimates substantially, while retaining the good accuracy of the soybean proportion estimate.

Turning to time periods other than August, we examined the field-center recognition results for Fayette, 16 and 17 July. Corn had a high detection percentage in both -- 93% and 96%, respectively. Soybeans were under recognized (64-65%) with many pixels being assigned to the "bare soil" subclass of "other". There actually were late-planted soybean fields that had a bare-soil appearance in mid-July. If either pixels recognized as bare soil are considered to be part of the class "soybeans" or late-planted soybean fields are considered to be "other" at this time period, soybean field-center recognition percentages increase to 85-95% correct and the overall percentages correct increase from 63-64% to 74-76%. Deletion of urban areas produced further increases in overall field-center accuracy to 81-84%.

Finally, wheat recognition in the Shelby 8 June and Fayette 10 and 11 June segments was examined. In Sec. 5, the problems associated with the ground-truth determinations of wheat fields for test are discussed. To eliminate as much as possible any uncertainty about the identity of the pixels being tested, it was decided to use only fields visited by ASCS personnel, i.e., those in the CITARS training quarter sections. Training was performed on the Fayette 11 June ASCS fields and testing was performed on the ASCS fields of Shelby 8 June and Fayette 10 June. The MASC algorithm was used first to transform the 11 June signatures* to the measurement conditions of the other segments. Of the wheat in the two new "test" sets, 88-100% of it was correctly recognized with 92-94% of other pixels correctly recognized. These field-center results are substantial improvements over percentages obtained on the original test data. When results for combined ASCS and test fields were tallied, wheat field-center recognition was 83-93% correct with 84-95% correct for "other" pixels.

*The investigator in this case chose to use different training procedures than were employed for the standard CITARS processing so the resulting signatures sets were not identical.

6.1 CHANGES IN FIELD-CENTER DESIGNATIONS FOR FAYETTE, 21 AUGUST

On the final test field coordinate designations prepared by LARS for Fayette 21 August, we checked each cover-type designation against the available ground truth information. On the basis of this, several fields were found to be in error, and several were deleted from analysis. Table 12 lists the specific changes made and why.

In some cases, the photo-interpreter indicated change of crop type during the growing season, for example, small grain replanted to soybeans, or a wheat field becoming bare soil, and the changes were not reflected on the coordinate cards. The category woods/pasture contained mixtures of both trees and pasture and was not a useful pure cover type. Two fields were deleted because of contradictory ground truth information. Three soybean fields were deleted because they were sparse, very immature or otherwise not consistent with proper stands of soybeans. Urban areas also were deleted since in our judgement they contain mixtures of two or more spectral classes and could be eliminated from processing by stratification.

We deleted mixture classes from field-center analysis because their presence interferes with proper measurement of field-center recognition accuracy and because, in our opinion, the full-section analysis is the proper vehicle to study their effect on recognition.

6.2 RECOGNITION RESULTS USING REVISED FIELD-CENTER DEFINITIONS FOR FAYETTE 21 AUGUST

With the Table 12 revisions to the field-center pixel definitions, a re-calculation of results obtained with the CITARS procedure and signatures shows an overall increase to 82.2% correct with urban areas included and 83.5% correct with them excluded. Performance matrices for these cases are presented in Table 13.

We also compared ASCS field observation records with the cover types assigned to the pixels used for training. A few changes were made in one or two "other" subclasses, resulting in a modification on one "other" class signature when our CITARS procedures were applied to the revised data. With the new signature set, recognition accuracies increased to 83.2% correct with urban areas and 85.0% correct with urban areas excluded (See Table 13).

6.3 BARE SOIL VS. SOYBEANS IN JULY FAYETTE DATA

Since in the CITARS region the planting date of soybeans varied from June thru July, 1973, the maturity level of soybeans varies accordingly. In July, soybean ground cover varied in a "continuum" from very mature to new seedlings to bare soil. The CITARS field designation was soybeans even if the field were newly planted. And unfortunately, no scanner can tell the difference between bare soil that is <u>called</u> soybeans and bare soil that isn't.

To measure the effect of this problem, we turned to mid-July Fayette segments (16 and 17 July). Here corn field-centers were well-recognized -- 93% and 96%. Soybeans, in their state of highly variable maturity, had 64%

and 65% correct recognition. There was a separate "other "signature for bare soil. We first considered bare soil recognitions (approximately 100 pixels) to be part of the class "soybeans", and re-measured recognition performance. Then we considered immature soybeans to be part of the class "other" and remeasured performance. In both cases, the soybean recognition percentages increased to 85-95% correct, and the overall percentages increased from 63-64% to 74-76% correct.

When, in addition, urban-area pixels (See Table 12) were deleted from the analysis, the overall correct percentages increased to 81-84%. This is a total increase of up to 20% in field-center recognition accuracy.

6.4 USE OF VARIED SAMPLES OF DATA FOR TRAINING

Because the results of recognition tests indicated that the ASCS training field data did not adequately represent the test data, we investigated the use of different and/or additional samples of training data for use in local recognition with the linear decision rule on Fayette 21 August data.

First, half of the test data was used to establish corn and soybean signatures, while retaining the other signatures obtained from the ASCS quarter sections. We shall call the two halves of the test data "Test A" and "Test B". Test B data correspond to those labeled "Pilot" in the field coordinate definitions. Test A field-center data were used to establish the major-crop signatures. When these signatures were used in recognition on the pixels not used for training, the overall result increased to 84.3% from the 82.2% achieved with ASCS signatures, as shown in Table 14. When the test set was Test A & Test B, the same as for the ASCS signatures, the overall result was 88.8% correct. With testing on all field-center pixels, the result was 89.0% correct overall.

Then the ERIM clustering algorithm was used to establish 30 signatures. Recognition results with these signatures (also presented in Table 14) show 90.5% correct.

Thus, we see that recognition results are sensitive to the amount of the data used to establish signatures and to the degree to which they represent the test data.

6.5 USE OF MULTITEMPORAL DATA*

CITARS ERTS data from four time periods over the Fayette segment (10 June, 29 June, 17 July, and 21 August 1973) were merged to form a multitemporal data tape. These data previously had been placed in spatial registration according to a nearest-neighbor algorithm by Purdue/LARS as part of the CITARS data preparations. Two other passes also were available but were not used because each was within one day of one of the selected passes; furthermore, clouds were present on 11 June, and the 16 July data set had some data quality problems. The 29 June data set was included despite a number of bad data lines since, otherwise, the late-June time period would not be represented.

The intersection of pixel assignments for the four time periods was used to define a subset of the field-center pixels that had been identified for training in the CITARS data preparations. This subset excluded fields that

---

*The results reported herein were generated by R. Hieber and W. Malila under Task VI of the ERIM SR&T contract [8 and 9].

had data problems at one or more time periods, and was used in training for the reported multitemporal analysis. Four-character labels were assigned to the pixels of each field, one character to represent the ground-truth class at each of the time periods. There was only one label for all the corn pixels, but four different types of labels were needed to represent the variety of soybean planting dates and maturities.

A pixel-by-pixel clustering algorithm [11] was used to establish a set of clusters from the field-center training pixels. One feature of the algorithm is a capability to label pixels and use that labeling in the clustering procedure. Thus, one or more clusters was generated for each of the labeled classes. An iterative procedure was used to combine the many clusters generated on the pass through the data into a smaller number for use in recognition. Another feature of the algorithm is its distance measure which accounts for the variances associated with the clusters as well as mean separations.

Only eight of the 16 channels of multitemporal data were used in the clustering procedure. A selection was made of data channels 2 and 4 (ERTS Bands 5 and 7) from each time period; although arbitrary, this selection was made in part because of the high degree of correlation that has been observed between channels 1 and 2 and between 3 and 4.

Recognition processing was performed with four sets of signatures using the ERIM linear decision rule, and field-center results were tabulated for several different threshold levels, i.e., different levels at which a pixel is rejected as being a member of the "winning" recognition class. The primary interest was in results for the test fields used in the standard CITARS processing, although tabulations were made separately for training fields and combined training and test fields. Only seven of the eight channels clustered were used in recognition; channel 4 from 29 June was omitted because of its numerous bad lines.

A total of 21 clusters was defined for the first set of multitemporal recognition runs -- 12 for soybeans, three for corn, and six for "other". These remained after a combination was made of 33 clusters found at an intermediate stage. The recognition results were analyzed, with particular attention being paid to the distribution of misclassifications. As a result of this analysis, two new soybean clusters were established by applying the clustering procedure to pixels from six soybean test fields that were completely missed in the recognition runs. A second recognition run was made using these additional signatures and results tabulated for field centers.

It was decided to estimate an upper limit on recognition performance for this multitemporal data set by using all available field-center data (both ASCS training and photointerpreted test data) to establish signatures. The clustering program was run with one of three labels assigned to each pixel analyzed -- corn, soybeans, or other. A total of 49 clusters was produced. Two more recognition runs were made -- one with signatures from a subset consisting of the 30 clusters which represented the largest number of pixels and the other with all 49.

Field-center recognition results were obtained first with the 21 clusters, seven multitemporal channels, and the ERIM linear decision rule for five different decision thresholds corresponding approximately to 0.001, 0.0003, 0.0001, 0.00001, and 0^+ probability of falsely rejecting a point from the assumed multivariate normal signature distributions.

Corn recognition was 94.8% for all threshold values. Fifteen of 286 corn pixels were missed on each run. The number of non-corn pixels falsely recognized as corn rose from 24 at 0.001 threshold to 69 for 0^+ threshold.

Soybean recognition ranged from 72.1% to 91.3% correct, depending on threshold. None of the missed soybean pixels were falsely assigned to another recognition class; all were rejected by the threshold test. This fact indicates that the training data were not fully representative of the soybean test fields.

Recognition of the "other" class was 86.1% correct for the 0.001 threshold and decreased monotonically as the rejection threshold approached zero.

The overall percentage of points correctly recognized was largest (84.6%) for the 0.0003 rejection threshold, and results with it are presented in Table 15. This threshold value also was best with the augmented signature set discussed in the next paragraph.

Because of the previously noted failure of the signatures to recognize pixels from a number of soybean fields, the soybean signatures were augmented with two others determined from six fields that were completely missed with the 0.001 threshold (these represent 43 of the 100 soybean pixels missed at that threshold). The pixels from these fields were clustered and two new soybean signatures were established. Table 15 shows an improvement in soybean recognition from 77% to 84% and overall recognition from 85% to 87% when this additional pair of signatures was used with a 0.0003 threshold on the original CITARS field definitions. When the corrected field definitions were used, the overall average rose to 89% correct. Still, over 50 soybean pixels remained unclassified, i.e., rejected by the threshold test.

As a further demonstration of the need for more representative training data, we clustered all training and test field-center pixels to establish the more complete 30 and 49 signature sets for use in recognition. The results of recognition runs with these signatures are presented in Table 16. The overall field-center accuracy increased to 93% and 94% with the two sets of signatures. Soybeans had the greatest increase in values, old values were 77-84% and new ones 92-96%.

Full-section results with multitemporal data are discussed in Section 6.6. It is worth noting that the linear decision rule permitted use of the large number of signatures without prohibitive computer costs. The more signatures that are employed, the greater is its cost advantage over the quadratic decision rule.

6.6 ANALYSIS OF FULL-SECTION RESULTS FOR FAYETTE 21 AUGUST

Full-section recognition results do not depend on the accuracy with which field-center pixels can be located and identified, except for data used in training. Test data here included all pixels within square-mile sections and, thusly, included a variety of boundary and mixture pixels. The evaluation of these results consequently depends on "wall-to-wall" ground truth so that the true proportions of all crops in each test section can be computed for comparison with the proportions recognized in multispectral scanner data.

Full-section recognition results produced with the CITARS procedures for the Fayette 21 August segment (linear decision rule) were examined and compared with photointerpreted ("ground-truth") proportions on a section-by-section basis. The computer-recognized proportions of corn were all greater than the corresponding ground-truth proportions (See Table 17). In 14 of the test sections, the proportion of corn estimated from ERTS data exceeded the ground-truth proportion by 15% of the total section area, and for five of these 14 it was 25% or more. The physical composition of these sections was examined on the photomap and ground-truth overlays. A common constituent was trees, often in spatially extended or mixed patterns, as noted under "Comments" in Table 17. The largest discrepancy was in Section 94 which contains an urban area with stands of trees and tree-lined streets.

In contrast to the substantial over-recognition of corn, soybeans were much more accurately recognized and tended to be under-recognized when discrepancies occurred (See Table 18).

The observed over-recognition of corn proportions and equal or under-recognition of soybean proportions in the Fayette 21 August segment should not be attributed indiscriminately to other data sets. For example recognized proportions in data collected over White County on the same day also were examined on a section-by-section basis. There it was found that the proportions of both corn and soybeans usually were under-estimated. One explanation for this pattern is that, while the overall field-center recognition accuracies were approximately the same in these two segments, the major crop field centers were more accurately recognized than other field centers in the Fayette segment while the opposite was true for the White segment.

Full section results also were tabulated for the 49-signature multitemporal recognition run (See Sec. 6.4) and compared with the above results obtained using the standard CITARS procedures on the Fayette 21 August data. It so happens that soybean proportions were accurately and equally well recognized by both the single-time CITARS processing and the multitemporal processing. However, corn presented a much different situation, because the estimated proportion of corn for the single-time processing was nearly double the true amount and there was a large variance in estimates for individual sections. The corn proportions estimated from the multitemporal data were much more accurate, the overall corn proportion being almost exactly the same as the ground truth proportion and the variance in section estimates reduced to 1/4 of the single-time variance (See Table 19); however, the variance in corn estimates still exceeded that for soybeans.

The RMS deviation measure of full-section recognition performance, introduced in Sec. 5.2.2, shows an excellent performance for the multitemporal data.

6.7 USE OF ERIM PROPORTION ESTIMATION (MIXTURES) ALGORITHMS*

When a pixel represents a ground resolution element with more than one material present in a substantial amount, the pixel cannot be properly recognized by conventional multispectral recognition rules. One effect of such errors could be inaccurate acreage estimates for the crop(s) of interest.

The ERIM proportion estimation algorithms estimate the proportions of constituent materials within individual pixels (or within edited average pixel over a large area) by using the spectral information available in multiple channels of data.

An improved mixtures estimation procedure, LIMMIX, recently was developed by H. Horwitz, J. Lewis, and A. Pentland under another task of the ERIM's SR&T contract.[10] It allows consideration of mixtures of two materials at a time, as well as greater numbers if desired, when the pixel is not definitely assigned to a single material. Another development on yet another SR&T task has been multielement processing by W. Richardson[12]. The results reported below represent a combination of these two techniques as implemented by A. Pentland [11]. The combined procedure referred to as the "nine-point mixtures algorithm", was tested on CITARS data from the Fayette segment for corn and soybeans on 21 August 1973.

---

*The results reported herein were generated by H. Horwitz, A. Pentland, and J. Lewis under Task IV of the ERIM SR&T Contract [10, 11].

The training procedure requires signatures which may be obtained in any of several ways, e.g., from combinations of training field-center pixels (as in CITARS) or from clustering of pixels. In preliminary applications of the procedure the algorithm has been applied to training data for different values of the parameters (η_1^2, η_2^2, η_3^2) of the procedure so optimum parameter values can be established for use on test data.

The algorithm was applied to the Fayette 21 August data. Training was performed on the combination of original ASCS training data plus Test-B (pilot) data because experience had shown that the training data were not fully representative of the field-center test data (as discussed in Section 6.4) and because the selection of quarter sections for ASCS visits was biassed toward very high proportions of agricultural fields, in contrast to the random selection of test sections.

The result of this operation was encouraging for the ASCS and Test-B data, as shown in Table 20. The optimum parameters were $\eta_1^2 = 20$, $\eta_2^2 = 2.5$, and $\eta_3^2 = 2.5$.

Finally, the algorithm with these optimum parameters was applied to Test-A data from the Fayette August 21 segment. As shown in Table 20, the technique was a substantial improvement over conventional recognition for proportion estimation and gave a very satisfactory result in comparison to the original CITARS result.

Furthermore, over the Test-A sections, the RMS error between true corn percentage and the estimated corn percentage was only 3.53. For soybeans, the corresponding figure was 4.33. Compare these results with Tables 17-19 in Section 6.6.

6.8 NON-LOCAL RECOGNITION WITH NEW SIGNATURE EXTENSION ALGORITHM*

The MASC (Multiplicative and Additive Signature Correction) algorithm for signature extension was recently developed at ERIM by Dr. Robert Henderson under Task II of this ERIM SR&T contract [13]. With this algorithm, a signature correction transformation is determined for extending signatures from one site to another. The transformation applies both a multiplicative and an additive correction term to each signature. For example, the MASC correction for transforming signature means of one crop from one site, W, for use in another site, F is

$$\hat{m}_F^{(i)} = a_{F|W}^{(i)} m_W^{(i)} + b_{F|W}^{(i)}$$

where $m_W^{(i)}$ is the mean value for the crop in channel i for the area W,

$a_{F|W}^{(i)}$ and $b_{F|W}^{(i)}$ are the multiplicative and additive correction coefficients, respecitvely, for transforming W signatures to F conditions,

and $\hat{m}_F^{(i)}$ is the adjusted signature mean value in channel i for use in area F

---

*The results reported herein were generated by R. Henderson under Task II of the ERIM SR&T Contract [13].

The factors $\{a^{(i)}_{F|W}\}$ are used also to scale the signature variance-covariance matrices.

While the transformation coefficients $\{a^{(i)}, b^{(i)}\}$ could be determined from radiometric and atmospheric measurements made in the two sites at the time of data collection, such measurements usually are not taken and other factors, such as bidirectional reflectance, can cause differences as well. Alternatively, these coefficients can be based on the results of unsupervised pixel-by-pixel data clustering procedures in the two sites of interest. The clusters are paired between sites and used in conjunction with a linear regression program which computes the coefficients. It is not necessary to identify the ground cover classes associated with the various clusters, because they are paired according to their relative signal values. Furthermore, it is not necessary that the proportions of the cover classes be the same in both sites, although ideally the same cover classes should be present in both sites.

Earlier in Sec 5.2, we noted that non-local recognition performance between the Fayette and White segments on 21 August was very poor, approximately half of the local values. The mean-level-adjustment method of signature extension produced good results in transforming Fayette signatures for use on White data, but only a small increase in recognition was obtained with transformed White signatures on Fayette data. The new MASC signature extension algorithm then was applied to these data to see if it could improve on the latter performance.

When untransformed signatures from White were applied to both training and test in Fayette, some 240 km away, an overall recognition accuracy of only 28% correct was achieved (See Table 21).

Recognition of the major crops was especially poor, only 1.7% for corn and 10.0% for soybeans.

Very much improved results, with an overall average of 80% correct, were obtained for Fayette through the MASC transformation of White signatures. As shown in Table 22, corn and soybean recognition improved to 83% correct and overall accuracy to 80% correct. These results are comparable to those obtained using local signatures on Fayette (Table 6) and exhibit substantially more improvement than did the mean-level adjustment procedure.

The MASC algorithm also was applied to early-June CITARS data where wheat was the major crop of interest. Because of the previously discussed difficulties with ground truth for wheat in test areas, only training field data were used in the tests described below. That is, signatures extracted from training fields in the Fayette 11 June segment were transformed and tested on data from the training fields of Fayette 10 June and Shelby 8 June which were collected under different observation conditions. The MASC transformation again was developed by the analysis of clusters generated in the two segments by an unsupervised clustering algorithm.

Results of the recognition tests made with and without MASC signature transformations are presented in Table 23. It is seen that wheat field-center recognition is improved substantially with equal or only slightly degraded other recognition. Results are presented for combined ASCS and test fields as well as for ASCS fields only.

TABLE 12. CHANGES MADE IN TEST FIELD-CENTER PIXEL DESIGNATIONS FOR THE FAYETTE SEGMENT, 21 AUGUST, DURING SUPPLEMENTARY ANALYSES

1. CHANGES IN FIELD COVER TYPE:

FIELD	FORMER TYPE	CORRECT TYPE	REASON
17-31	Small Grain	Tree	Confusion between field 31 (trees) and 31A (small grain)
29-09	Tree	Soybeans	PI and photomap indicate soybeans
11-15	Small Grain	Soybeans	Field replanted to soybeans
39-18	Small Grain	Bare Soil	PI indicates this change
29-37	Wheat	Bare Soil	PI indicates this change
29-33	Wheat	Bare Soil	PI indicates small grain to bare soil; this is center part of field pixels designated 29-29
58-48	Small Grain	Idle	PI indicates "small grain or idle, probably idle"
16-11	Small Grain	Idle	PI indicates idle
56-12	Wheat	Hay	PI indicates this change

2. DELETIONS FROM FIELD-CENTER ANALYSIS

FIELD	TYPE	REASON
39-17	Woods/Pasture	Mixture of two different cover types
39-35	Woods/Pasture	Mixture of two different cover types
45-04	Trees	Mixture, more pasture than trees
45-18	Trees	Mixture of trees and pasture
45-22	Trees	Mixture of woods, pasture, pond, field
58-04	Trees	Sparse woods with pasture

TABLE 12. (Cont'd)

FIELD	TYPE	REASON
16-03	Pasture	Mixed with trees
33-07	Pasture	Varied area, diverted pasture, crossing of field boundary apparent on photomap
29-29	Wheat	Really covers 3 fields, the 2 side ones being soybeans, the center one bare (see 29-33 above)
17-17	Wheat	The large field called wheat based on ASCS survey shows several different fields on photomap and PI data.
17-43	Soybeans	Incomplete ground cover, mostly bare soil on Sept. 7 according to PI
19-62	Soybeans	Very immature, PI comments point to largely bare soil in Aug.
11-13	Soybeans	Large ditches with soil response in region pixels selected
94-08	Urban	Contains mixture pixels and is clearly non-agricultural
94-14	Urban	Contains mixture pixels and is clearly non-agricultural
94-26	Urban	Contains mixture pixels and is clearly non-agricultural
94-34	Urban	Contains mixture pixels and is clearly non-agricultural

TABLE 13. RECOGNITION RESULTS USING REVISED FIELD CENTER DEFINITIONS, FAYETTE 21 AUGUST SEGMENT, ERIM LINEAR DECISION RULE, CITARS PROCEDURES (0.001 THRESHOLD)

ORIGINAL CITARS RESULTS

	INCLUDING URBAN AREAS					EXCLUDING URBAN AREAS				
			RECOGNIZED AS:					RECOGNIZED AS:		
TRUE CLASS	% PIXELS CORRECT	NO. PIXELS	CORN	SOYBEAN	OTHER	% CORRECT	NO. PIXELS	CORN	SOYBEAN	OTHER
Corn	87.4%	286	250	18	18	87.4%	250	250	18	18
Soybeans	85.5%	358	22	306	30	85.5%	306	22	306	30
Other	69.8%	374	91	22	261	71.6%	271	55	22	194
	80.3%	1018				82.0%	915			

RESULTS WITH REVISIONS OF FIELDS AS GIVEN IN TABLE 12

Corn	87.4%	286	250	18	18	87.4%	286	250	18	18
Soybeans	87.1%	357	23	311	23	87.1%	357	23	311	23
Other	70.3%	269	73	7	189	73.5%	166	37	7	122
	82.2%	912				83.5%	809			

RESULTS WITH TABLE 12 FIELD REVISIONS PLUS CHANGED SIGNATURE SET (NEW SIGNATURE BASED ON CHANGED TRAINING FIELD ID'S)

Corn	87.4%	286	250	18	18	87.4%	286	250	18	18
Soybeans	89.1%	357	24	318	15	89.1%	357	24	318	15
Other	71.1%	269	71	7	191	72.3%	166	39	7	120
	83.2%	912				85.0%	809			

TABLE 14. RECOGNITION RESULTS USING VARIED SAMPLES OF TRAINING DATA, FAYETTE 21 AUGUST

	TRAINING SET	TEST SET	TRUE CLASS	% PIXELS CORRECT
(1)	ASCS Only (CITARS RESULT)	TEST A & TEST B	Corn	87.4%
			Soybeans	87.1%
			Other	70.3%
				82.2%
(2)	TEST-B For Corn & Soybeans; ASCS For Other	TEST A + ASCS For Corn & Soy ; TEST A + TEST B For Other	Corn	79.7%
			Soybeans	82.9%
			Other	87.0%
				84.3%
(3)	TEST-B For Corn & Soybeans; ASCS For Other	TEST A & TEST B	Corn	87.8%
			Soybeans	91.0%
			Other	87.0%
				88.8%
(4)	ASCS Only	TEST A + TEST B + ASCS	Corn	87.9%
			Soybeans	82.9%
			Other	82.1%
				83.9%
(5)	TEST-B For Corn & Soybeans; ASCS For Other	TEST A + TEST B + ASCS	Corn	85.7%
			Soybeans	86.7%
			Other	93.4%
				89.0%
(6)	ASCS + TEST A + TEST B (30 Signatures)	TEST A + TEST B + ASCS	Corn	86.7%
			Soybeans	90.6%
			Other	92.9%
				90.5%

NOTES: (1) Linear decision rule

(2) Test data are field centers modified as in Table 12 with urban areas retained.

TABLE 15. EFFECTS OF TWO ADDED SOYBEAN SIGNATURES AND CORRECTED FIELD DEFINITIONS ON MULTITEMPORAL RECOGNITION, FAYETTEE SEGMENT

	ORIGINAL CITARS FIELD DEFINITIONS		REVISED FIELD DEFINITIONS
	TRAINING-FIELD CLUSTERS ONLY	WITH TWO SUPPLEMENTAL SOYBEAN CLUSTERS	WITH TWO SUPPLEMENTAL SOYBEAN CLUSTERS
TEST FIELD CENTERS, % POINTS CORRECT:			
CORN	94.8	94.8	94.8
SOYBEANS	77.1	84.4	84.9
OTHER	84.0	84.4	86.6
AVG. OVER POINTS	84.6%	87.1%	88.5%
TEST FIELD POINTS MISSED: (REJECTED/MISCLASSIFIED)			
CORN	9/6	14/1	9/6
SOYBEANS	82/0	56/0	54/0
FALSE DETECTIONS TO:			
CORN	28	28	27
SOYBEANS	33	33	10
OTHER	96	70	68
TOTAL	157	131	105

NOTES: (1) 0.0003 rejection threshold

(2) 7 channels

(3) 21 signatures based on training field clusters

(4) The test pixels used are the revised CITARS field-center set with urban pixels retained (See Table 12).

TABLE 16. MULTITEMPORAL RECOGNITION RESULTS FOR TRAINING ON ALL FIELD-CENTER PIXELS

TOTAL NUMBER OF SIGNATURES	TRUE CLASS	% PIXELS CORRECT	TOTAL NO. PIXELS	TEST PIXELS RECOGNIZED AS: CORN	SOYBEAN	OTHER
30	CORN	90.6%	286	259	2	25
	SOYBEANS	92.2%	357	0	329	28
	OTHER	95.2%	269	11	2	256
		92.5%	912			
49	CORN	89.2%	286	255	5	26
	SOYBEANS	96.1%	357	0	343	14
	OTHER	96.3%	269	10	0	259
		94.0%	912			

NOTES: (1) Threshold for 0.0003 probability of false rejection.

(2) The test pixels used are the revised CITARS field-center set with urban pixels retained (See Table 12).

(3) The 30 signatures are a subset of the 49.

TABLE 17 FULL-SECTION CORN RECOGNITION RESULTS FOR THE FAYETTE 21 AUGUST SEGMENT, WITH STANDARD CITARS PROCEDURES

CITARS SECTION	GROUND-TRUTH PROPORTION (%)	RECOGNIZED PROPORTION (%)	DIFFERENCE ≥15%	DIFFERENCE ≥25%	COMMENTS
2	19.8	36.4	✓		Trees, brush, contoured land
11	15.9	27.2			
15	9.3	28.0	✓		Expressway; trees in one corner
16	15.9	37.9	✓		Two roads; many tree-lined field borders
17	7.7	32.5		✓	Trees and Brush
19	11.7	35.4	✓		
20	6.8	33.0		✓	Many thin areas with trees
26	37.9	41.9			
29	18.4	41.4	✓		Many trees in pasture
33	27.4	45.2	✓		Farm Fields only - on exception
34	26.1	33.5			
36	69.3	70.9			
39	11.7	38.8		✓	Trees and brush
41	10.8	28.1	✓		More than half trees; hilly terrain
45	19.9	28.8			
56	54.2	54.8			
58	6.5	28.6	✓		Trees and brush
67	11.4	42.8		✓	Stream pattern lined with trees
80	1.7	16.1	✓		40% Trees with many curved boundaries
94	6.3	40.5			Primarily urban with many trees
Total	19.7	36.8			

RMS Error = 19.98

$$\text{RMS Error} = \sqrt{\frac{1}{N}\sum_{i=1}^{20}(\hat{p}_i - p_i)^2}$$

TABLE 18. FULL-SECTION SOYBEAN RECOGNITION RESULTS FOR THE FAYETTE 21 AUGUST SEGMENT WITH STANDARD CITARS PROCEDURES

CITARS SECTION	GROUND-TRUTH PROPORTION (%)	RECOGNIZED PROPORTION (%)
2	41.7	39.0
11	70.2	60.8
15	33.3	38.5
16	29.1	38.5
17	23.0	13.2
19	23.2	18.2
20	38.2	29.6
26	54.1	54.4
29	21.0	26.1
33	43.4	32.8
34	44.7	43.2
36	9.9	14.0
39	16.1	10.3
41	14.4	5.6
45	37.5	36.6
56	18.4	28.2
58	16.2	11.9
67	35.4	32.4
80	15.5	11.2
94	1.5	1.6
Total	29.3	28.6

RMS Error = 6.39

TABLE 19. COMPARISON OF FULL-SECTION RECOGNITION RESULTS, FAYETTE SEGMENT

	CORN	SOYBEAN	OTHER	RMS DEVIATION*
GROUND TRUTH PROPORTION (%)	19.7	29.3	51.0	
CITARS RESULTS, 21 AUGUST				
Recognized Proportion (%)	36.8	28.6	34.6	13.7
RMS Error Between Sections (%)	19.98	6.38		
MULTI-TEMPORAL RESULTS				
Recognized Proportion (%)	19.6	26.8	53.6	2.1
RMS Error Between Sections (%)	9.51	6.16		

* See Definition on Table 9; this a measure of overall performance for all crops in the segment.

TABLE 20. RESULTS OF NINE-POINT MIXTURES ALGORITHM APPLICATION TO FAYETTE 21 AUGUST DATA

	CORN	SOYBEANS	OTHER	RMS DEVIATION
Proportion of training pixels (ASCS & TEST-B (Pilot))	21.50	34.88	43.62	2.2
Training ground-truth proportion	24.54	33.63	41.78	
Proportion of Test-A Pixels	15.85	31.41	52.74	1.5
Test-A Ground-Truth Proportion	14.16	31.06	54.78	
CITARS Original Result (TEST-A + TEST-B)	36.8	28.6	34.6	13.7
Ground-Truth Proportions (TEST-A+TEST-B)	19.7	29.3	51.0	

TABLE 21. NON-LOCAL RECOGNITION RESULTS USING UNTRANSFORMED SIGNATURES FROM WHITE 21 AUGUST DATA

FAYETTE 21 AUGUST ASCS + TEST DATA

TRUE CLASS	NO. PLOTS	NO. PIXELS	% CORRECT	NO. PIXELS RECOGNIZED AS: CORN	SOYBEANS	OTHER
CORN	43	356	1.7%	6	-	350
SOYBEANS	66	549	10.0%	-	55	494
OTHER	46	461	70.9%	126	8	327
TOTAL	155	1366		132	63	1171
AVG. OVER POINTS			28.4%			

TABLE 22. NON-LOCAL RECOGNITION RESULTS USING MASC-TRANSFORMED SIGNATURES FROM WHITE 21 AUGUST DATA

TRUE CLASS	NO. PLOTS	NO. PIXELS	% CORRECT	NO. PIXELS RECOGNIZED AS: CORN	SOYBEANS	OTHER
CORN	43	356	83.4%	297	33	26
SOYBEANS	66	549	83.2%	28	457	64
OTHER	46	461	72.2%	110	18	333
TOTAL	155	1366		435	508	423
AVG. OVER POINTS			79.6%			

TABLE 23. RECOGNITION RESULTS USING UNTRANSFORMED AND MASC SIGNATURES FOR WHEAT

Data Set Signature Extracted From	Signature Transformation Applied*	Data Set Signatures Applied To	Field-Center Pixel Recognition, Percent Correct	
		ASCS and Test	Wheat	Other
Fay, 11 June	UT	Fay, 10 June	64.0%	89.3%
Fay, 11 June	MASC	Fay, 10 June	93.0%	84.2%
Fay, 11 June	UT	Shelby, 8 June	41.5%	95.9%
Fay, 11 June	MASC	Shelby, 8 June	83.0%	95.0%
		ASCS Only		
Fay, 11 June	UT	Fay, 10 June	72.9%	97.3%
Fay, 11 June	MASC	Fay, 10 June	100%	94.3%
Fay, 11 June	UT	Shelby, 8 June	17.6%	96.9%
Fay, 11 June	MASC	Shelby, 8 June	88.2%	92.0%

*UT = Untransformed

MASC = Multiplicative and Additive Signature Correction

Page Intentionally Left Blank

7

REFERENCES

1. Task Design Plan for the Crop Identification Technology Assessment for Remote Sensing (CITARS) Phase of the Large Area Land Resources and Crop Acreage Inventory Applications System Verification Test Project, NASA Johnson Space Center, Earth Observations Division, Houston, Texas, 1973.

2. Hall, F. G., Bauer, M. E., and Malila, W. A., "First Results From the Crop Identification Technology Assessment for Remote Sensing (CITARS)", Proceedings of Ninth International Symposium on Remote Sensing of Environment, ERIM, Ann Arbor, Michigan, April 15-19, 1974.

3. Bizzell, R. M., CITARS Data Acquisition, Part 1: ERTS-1 MSS, NASA Johnson Space Center, Earth Observations Division, Houston, Texas (circa 1974).

4. Malila, W. A., R. H. Hieber, and A. P. McCleer, "Correlation of ERTS MSS Data and Earth Coordinate Systems", Proceedings of Conference on Machine Processing of Remotely Sensed Data, Purdue University, W. Lafayette, Ind., 16-18 Oct. 1973.

5. Simmons, W. R., D. M. Freeman, and M. E. Bauer, ERTS Data Preparation for CITARS, Technical Report, LARS, Purdue University, West Lafayette, Indiana, (in preparation).

6. Crane, R. B., and W. Richardson, "Rapid Processing of Multispectral Scanner Data Using Linear Techniques", Remote Sensing of Earth Resources, Vol. 1, edited by F. Shahrokhi, University of Tennessee Space Institute, Tullahoma, Tennessee, 1972.

7. Kriegler, F. J., D. Christenson, M. Gordon, R. Kistler, S. Lampert, R. Marshall, and R. McLaughlin, MIDAS, Prototype Multivariate Interactive Digital Analysis System - Phase 1, Volume 1: System Description, Report NASA CR-132463, Environmental Research Institute of Michigan, Ann Arbor, Mich., August 1974.

8. Malila, W. A., and R. H. Hieber, "Task VI, Multitemporal Analysis" Quarterly Progress Report, No. 109600-6-L, NASA Contract NAS9-14123, Environmental Research Institute of Michigan, Ann Arbor, Michigan, Nov. 1974.

9. Malila, W. A., R. H. Hieber, and R. C. Cicone, "Task VI, Multitemporal Analysis", Quarterly Progress Report, No. 109600-10-L, NASA Contract NAS9-14123, Environmental Research Institute of Michigan, Ann Arbor, Mich., Feb. 1975.

10. Horwitz, H., J. Lewis, and A. Pentland, "Task IV, Proportion Estimation", Quarterly Progress Report, No. 109600-3-L, NASA Contract NAS9-14123, Environmental Research Institute of Michigan, Ann Arbor, Mich., Aug. 1974.

11. Horwitz, H., J. Lewis, and A. Pentland, "Task IV, Proportion Estimation", Quarterly Progress Report, No. 109600-10-L, NASA Contract NAS9-14123, Environmental Research Institute of Michigan, Ann Arbor, Mich., Feb. 1975.

12. Richardson, W., A Study of Some Nine-Element Decision Rules, Report NASA-CR ERIM 190100-32-T, Environmental Research Institute of Michigan, Ann Arbor, Mich., July 1974.

13. Henderson, R., "Subtask II.1 of Task II, Signature Extension and Other Special Techniques", Quarterly Progress Report, No. 109600-10-L, NASA Contract NAS9-14123, Environmental Research Institute of Michigan, Ann Arbor, Mich., Feb. 1975.

APPENDIX

COMPLETE DESCRIPTION OF PROCEDURES WHICH DIFFER FROM THOSE GIVEN IN THE TASK DESIGN PLAN

APPENDIX

COMPLETE DESCRIPTION OF PROCEDURES WHICH DIFFER FROM THOSE GIVEN IN THE TASK DESIGN PLAN

The Task Design Plan (Ref. 1) specifies in detail the procedures ERIM was to follow in processing CITARS data. Since that time, some of the procedures have been more fully specified, and a few minor changes were made in others. Those portions of the procedures whose description has changed are given below in full. They use the same numbering system found in Ref. 1, and each section replaces the corresponding numbered section of Ref. 1.

The following is the list of changes made.

Section	Changes
1.4 - 1.4.2.2	Unspecified parameters now given; minor changes in wording.
1.4.3.2 - 1.4.3.3	Method of expanding major-crop covariance matrices given.
1.5.1	Changes in wording.
1.5.2	Change in editing criterion used in extracting training statistics.
1.5.3 - 1.5.4	Complete description of preparation of other-class signatures.

[Sections 1. through 1.3.3: no change.]

1.4. DEFINE MAJOR-CLASS SIGNATURES FOR CLASSIFICATION

1.4.1. EXTRACT STATISTICS FOR FIELDS OF MAJOR CROPS (CORN, SOYBEAN, AND WHEAT). Program STAT will be used to extract signal statistics from the designated field-center pixels of those ASCS ground-truthed corn, soybean, and wheat fields that are selected by NASA as training fields. The standard editing procedure of program STAT will be used to delete from the statistics those pixels which are too dissimilar to the others.

1.4.2. COMBINE, TEST, REJECT, AND RECOMBINE FIELD STATISTICS. The training fields will be analyzed independently for each of the three major classes, and recognition signatures generated. This step will be performed by program SIGCOM.

1.4.2.1. Combine the Field Statistics. All training-field statistics for a given class will be combined by subroutine COMSCL into one interim class signature, with equal weights used for large fields ($\geq$ (20) pixels) and lesser weights used for smaller fields. The weights for fields of less than (20) pixels will be $(N_i/20)^{1/2}$ times the large-field weight, where N_i is the number of pixels in the i-th small field.

1.4.2.2. Test and Reject Individual Field Statistics. The mean vector of each individual field will be tested against the interim combined class signature derived in the above step by evaluating the interim combined quadratic form at the field mean of the individual field and flagging the field as questionable if the value exceeds the χ^2 value of NCHAN, where NCHAN is the number of channels being used.

The signatures from all non-questioned fields will be reprocessed by subroutine COMSCL to produce a new signature for the field elimination test that follows. The weighting for these field signatures will be the same as discussed in Section 1.4.2.1.

Each individual field will be tested against this new combined class signature by evaluating the new combined quadratic form at the mean of the individual field; if the value exceeds the χ^2 value for 0.01 probability of rejection, (which for 4-channel signatures is 13.277) the field will be eliminated from further consideration in training.

[Sections 1.4.2.3 through 1.4.3.1: no change.]

1.4.3.2. Histogram Exponents. One histogram will be made with program STAT of the exponents generated by correct classifications for each of the three classes. For example, the histogram for corn will be for all pixels which are both from those corn training fields used to derive the final corn signature and recognized as corn. The exponent limit necessary to accept (75%) of the pixels will be read off each histogram, giving a separate value for each of the three classes.

1.4.3.3. Test and Scale the Covariance Matrices. The following factor will be computed for each major crop.

$$\text{factor} = \frac{H_{.75}}{\chi^2_{.75}} \times \frac{\chi^2_{.01}}{\chi^2_{.001}} = H_{.75} \times \frac{1}{7.491} \text{ (4 channels)}$$

where

$H_{.75}$ = exponent limit to reject 25% of histogrammed pixels (from Step 1.4.3.2).

$\chi^2_{.75}$ = theoretical chi-square value for 0.25 probability of rejection.

$\chi^2_{.01}$ = theoretical chi-square value for 0.01 probability of rejection.

$\chi^2_{.001}$ = theoretical chi-square value for 0.001 probability of rejection.

If this factor is less than 1.0, no scaling will be done. If this factor exceeds 1.0, the covariance matrix will be expanded, using program ASCALE, by multiplying each element by the factor.

These three signatures, after they possibly undergo scaling, will be used for the major crops in subsequent steps.

The purpose of this is to make it likely that classification using major crop signatures will correctly classify at least 99% of major crops training pixels as major crops, using a χ^2 threshold of 0.001 probability of false rejection. Since there are not sufficient training pixels to define adequately the tail of the histogram (Step 1.4.3.2), so the 3rd quartile point is chosen to represent it, assuming its shape is that of the χ^2 curve with an expanded scale. Expanding the covariance matrix has the effect of scaling the histogram so that more points fall less than the 0.001 rejection threshold used by the classifier. The above formula attempts to align the 0.01 (99% accepted) position of the histogram with the 0.001 chi squared rejection threshold.

1.5. DEFINE "OTHER" CLASS SIGNATURES

1.5.1. IDENTIFY THE SIGNIFICANT "OTHER" CLASSES. The major crop signatures (and no others) will be used in a classification run over the "other" training field data. This run will be evaluated for a classification threshold set for 0.001 probability of false rejection.

Any other-class field with 20 pixels or fewer which has (two) or more pixels, or any larger field with more than (10%) of its pixels, classified as corn, soybeans, and/or wheat, will then be deemed a significant other-class field.

1.5.2. EXTRACT STATISTICS FOR "OTHER"-CLASS FIELDS. Program STAT. will be used to extract signal statistics from the designated field-center pixels of those ASCS ground-truthed fields which were selected by Step 1.5.1 as significant "other"-class fields. The standard editing feature of program STAT will be used. In practice, this step may be done in the same job as Step 1.4.1, omitting signatures from any field that is not a significant "other" class.

1.5.3. COMBINE, TEST, AND RECOMBINE THE FIELD STATISTICS.

1.5.3.1. Combine "Other" Field Statistics. The statistics for the fields in each significant "other" class will be combined into one signature by the procedure of Step 1.4.2.

1.5.3.2. Test and Possibly Regroup Field Statistics. Given a specific significant minor-crop type, we perform the following test for each major crop. The probabilities of misclassification of the major crop (combined signature) as the minor crop (each field separately, and all fields combined) are computed using program POMPOM and are examined. The largest of these values for each field is compared to the value for all fields combined. If the latter exceeds the former by more than 0.02, then the minor-crop fields are regrouped as described below. Otherwise, the signature of all fields combined is selected.

If subdivision is called for, the following is done. If only two fields of the minor crop are present, the two separate field signatures are selected. If more than two fields exist for the significant minor crop, the fields will be divided into two groups, each to be combined separately, in such a way that the largest probability of misclassifying a combination as major crop is smaller than for any other possible grouping. The recombination will be performed according to the procedure of Step 1.4.2. If the resultant combinations still satisfy the above criterion for regrouping, the signatures will be grouped into three or more groups.

1.5.4. HANDLE SIGNATURES WITH SINGULAR COVARIANCE MATRICES. Whenever a signature is formed from NCHAN or fewer pixels, where NCHAN is the number of spectral channels, the covariance matrix is singular and its computer representation is nearly singular and perhaps even not non-negative definite. Whenever an ill-conditioned covariance matrix is encountered, it will be forced to a minimum size by adding 0.1 to each diagonal element of the matrix.

[All subsequent sections: no change.]