"Made available under NASA sponsorship in the interest of early and wide dissemination of Earth Resources Survey Program information and without liability for any use made thereof."

NASA CR-
ERIM 122700-33-F
78-10016
NASA CR-
151552

Final Report

AN EVALUATION OF THE SIGNATURE EXTENSION APPROACH TO LARGE AREA CROP INVENTORIES UTILIZING SPACE IMAGE DATA

RICHARD C. CICONE, JOHN L. STINSON,
and RONALD J. BALON

NOVEMBER 1977

Principal Investigator
Richard F. Nalepka

(E78-10016) AN EVALUATION OF THE SIGNATURE EXTENSION APPROACH TO LARGE AREA CROP INVENTORIES UTILIZING SPACE IMAGE DATA Final Technical Report, 15 May 1976 - 14 Nov. 1977 (Environmental Research Inst. of
N78-12496
HC A06/MF A01
Unclas
G3/43 00016

Prepared for
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Johnson Space Center
Earth Observations Division
Houston, Texas 77058
Contract No. NAS9-14988, Task 3
Technical Monitor: I. Dale Browne/SF3

ENVIRONMENTAL
RESEARCH INSTITUTE OF MICHIGAN
FORMERLY WILLOW RUN LABORATORIES THE UNIVERSITY OF MICHIGAN
BOX 8618 • ANN ARBOR • MICHIGAN 48107

NOTICE

Sponsorship. The work reported herein was conducted by the Environmental Research Institute of Michigan under Contract NAS9-14988 for the National Aeronautics & Space Administration, Johnson Space Center, Houston, Texas 77058. I. Dale Browne was Technical Monitor for NASA. Contracts and grants to the Institute for the support of sponsored research are administered through the Office of Contracts Administration.

Other Final Reports. This is one of 6 final reports on this contract.

Disclaimers. This report was prepared as an account of Government sponsored work. Neither the United States, nor the National Aeronautics & Space Administration (NASA), nor any person acting on behalf of NASA:

(A) Makes any warranty expressed or implied, with respect to the accuracy, completeness, or usefulness of the information, apparatus, method, or process disclosed in this report may not infringe privately owned rights; or

(B) Assumes any liabilities with respect to the use of, or for damages resulting from the use of any information, apparatus, method, or process disclosed in this report.

As used above, "person acting on behalf of NASA" includes any employee or contractor of NASA, or employee of such contractor, to the extent that such employee or contractor of NASA or employee of such contractor prepares, disseminates, or provides access to any information pursuant to his employment or contract with NASA, or his employment with such contractor.

Availability Notice. Request for copies of this report should be referred to:

National Aeronautics & Space Administration
Scientific & Technical Information Facility
P. O. Box 33
College Park, Maryland 20740

Final Disposition. After this document has served its purpose, it may be destroyed. Please do not return it to the Environmental Research Institute of Michigan.

1. Report No. NASA CR- ERIM 122700-33-F	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle An Evaluation of the Signature Extension Approach to Large Area Crop Inventories Utilizing Space Image Data		5. Report Date November 1977 6. Performing Organization Code
7. Author(s) R. C. Cicone, J. L. Stinson, R. J. Balon		8. Performing Organization Report No. 122700-33-F
9. Performing Organization Name and Address Environmental Research Institute of Michigan Infrared & Optics Division P.O. Box 8618 Ann Arbor, Michigan 48107		10. Work Unit No. Task 3 11. Contract or Grant No. NAS9-14988
12. Sponsoring Agency Name and Address National Aeronautics & Space Administration Johnson Space Center Houston, Texas 77058		13. Type of Report and Period Covered Final Technical Report 15 May 76 - 14 Nov 77 14. Sponsoring Agency Code

15. Supplementary Notes

The work was performed for the Earth Observations Division.
I. Dale Browne (SF3) was the Technical Monitor.
Mr. Richard F. Nalepka was ERIM's Principal Investigator.

16. Abstract

The objective of this project is to test and evaluate algorithms and procedures which embody the signature extension approach to large area crop inventories utilizing space image data. The study was carried out in two phases.

Phase I investigated signature extension techniques which are candidates for inclusion in an overall signature extension procedure. Four types of signature extension techniques or signature extension related procedures were examined: haze correction algorithms, data stratification procedures, training sample selection strategies for multisegment training, and crop development classifiers. Phase I results reported herein are supplemented by a more detailed interim technical report, Evaluation of Signature Extension Algorithms, ERIM Report No. 122700-29-T.

Phase II addressed the evaluation of overall signature extension procedures. Three tasks were carried out: (1) a specification of the multisegment signature extension experiment design, (2) preparatory stages in the execution of the experiment, and (3) an analysis of LACIE analyst interpreter labeling errors. Phase II has been carried out on the basis that continued test and evaluation of the signature extension approach will be pursued during the next contract year.

17. Key Words (Suggested by Author(s))		18. Distribution Statement
Landsat Signature Extension Stratification LACIE Multisegment Haze Correction	Test and Evaluation Analyst-Interpreter Remote Sensing	Initial distribution is listed at the end of this document.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price*
Unclassified	Unclassified	x + 128	

*For sale by the National Technical Information Service, Springfield, Virginia 22161

JSC Form 1424 (Rev Nov 75)

NASA

PREFACE

This document reports processing and analysis efforts on one task of a comprehensive and continuing program of research in multispectral remote sensing of the environment. The research is being carried out for NASA's Lyndon B. Johnson Space Center, Houston, Texas, by the Environmental Research Institute of Michigan (ERIM). The basic objective of this program is to develop remote sensing as a practical tool for obtaining extensive environmental information quickly and economically.

The specific focus of the work reported herein was on the test and evaluation of the signature extension approach to large area crop inventories. This final report is complemented by an interim technical report ERIM 122700-29-T entitled, "Evaluation of Signature Extension Algorithms", by Alex P. Pentland.

The research covered in this report was performed under Contract NAS9-14988 during the period 15 May 1976 to 14 November 1977. Mr. I. Dale Browne (SF3) served as the NASA Contract Technical Monitor, and Mr. M. C. Trichel (SF3) was NASA Task Monitor. At ERIM, the work was performed within the Infrared and Optics Division, headed by Richard R. Legault, Vice-President of ERIM, in the Information Systems and Analysis Department, headed by Dr. Quentin A. Holmes. Mr. Richard F. Nalepka, head of the Multispectral Analysis Section, served as Principal Investigator, Mr. Richard Cicone and Mr. Alex Pentland shared responsibilities as Task Leader.

The authors wish to acknowledge the assistance of other ERIM staff members who have participated in the development of techniques in the LACIE agricultural context examined herein. Mr. Richard Kauth and Dr. Wyman Richardson contributed to the design of the multisegment signature extension experiment reported herein. Mr. Robert Beswick provided able support. Ms. Darlene Dickerson, Mrs. Elizabeth Hugg and Ms. Martha Warren provided efficient and accurate typing support throughout the contract period and for this report.

CONTENTS

IV
PAGE INTENTIONALLY BLANK

PRECEDING PAGE BLANK NOT FILMED

CONTENTS (Cont'd)

FIGURES

FIGURES (Cont'd)

TABLES

TABLES (Cont'd)

1
SUMMARY

The overall objective of the research reported herein was to initiate an evaluation of the signature extension approach to large area crop inventories utilizing space image data. The Large Area Crop Inventory Experiment (LACIE) is an attempt to establish the feasibility of inventorying the production of wheat on a world-wide basis by utilizing Landsat data. A basic 5x6-mile sampling region or segment is employed and wheat production statistics are aggregated over estimates made within each segment. The current estimation technique employed is called Procedure 1. This technique extracts training data from each segment, applying the resultant measured statistics in classifying the segment. This local training and classification procedure requires that each segment be manipulated by an intervening Analyst Interpreter (AI). Multisegment training and classification techniques attempt to reduce the need for AI intervention. This is carried out by extracting training statistics from a subset of segments and employing the statistics or signatures to other segments, hence the term signature extension.

The activity was carried out in two phases. First, several algorithms and procedures which were candidates for inclusion in a large area crop inventory system were separately evaluated. Second, preparation was made to conduct an extensive signature extension systems evaluation incorporating those candidate algorithms and procedures which showed promise for crop inventories in a multisegment environment, and an analysis was carried out to investigate the Analyst Interpreter stage in crop inventory.

The algorithms and procedures evaluated in the first phase of this program are divided into four distinct types:

1. Haze correction algorithms
2. Training sample selection strategies

3. Data stratification procedures
4. Permanently trained green development-trajectory classifiers.

The algorithms tested which fall into category one, haze correction algorithms, are CROP-A [1] and XSTAR [2]. The XSTAR algorithm has been extensively tested in both winter and spring wheat areas and offers substantial benefit to large area crop inventory systems.

The training sample selection strategy available for testing was a preliminary version of Procedure B [3]. First results show its promise for future large area crop inventory systems.

In the third category, stratifications of the data, two were available for testing: a static stratification defined by UCB [4], and one defined by JSC [5]. Employment of these stratifications yielded an increase in classification accuracies. It appears that these stratifications should be further tested using a multisegment training strategy in order to clearly establish their contribution to improved performance in this environment.

In the final category, green development-trajectory classifiers, several algorithms were tested. Four unitemporal green development classifiers, with and without haze correction, the Delta Classifier [6], and a crop development classifier were tested. Results obtained are promising, but additional testing is recommended using a more substantial data base covering several growing seasons.

The second phase of the program revolved about three basic concerns:

1. The definition and advanced design of an experiment to examine the overall signature extension approach
2. Preparatory phases required to conduct such an experiment
3. Analysis of the nature of analyst interpreter errors and the sensitivity of the signature extension approach to analyst interpreter errors.

The design of the multisegment signature extension experiment required a definition of five basic components of an experiment including: (1) the definition of the systems under test, (2) definition of performance measures, (3) definition of the measurement procedures, (4) specification of factors, parameters and levels desired, and (5) specification of data sets. The systems to be evaluated incorporate the static stratifications defined by UCB and JSC, Procedure B defined by ERIM, data preprocessing filters including haze correction defined by ERIM, and Multisegment Procedure 1. The particular performance measure of most interest will be the measure of variation in wheat proportion estimate as a function of training gain. The results of the multisegment signature extension approach are to be compared to standard LACIE Phase III local classification results.

Preparatory phases carried out to expedite the execution of this experiment have included both data base specification and software development. A preclassification technique was developed to facilitate the evaluation of classification performance where training parameters, like the number of training segments, would be varied to establish the variation in performance.

The specification of a data base for testing led to an analysis of the nature of Analyst Interpreter (AI) errors detected in the labeling of wheat and non-wheat for training purposes. The AI's basic tool is a false color image product generated from Landsat digital data using a Production Film Converter (PFC) that maps Landsat bands 4, 5 and 7 into blue, red and green colors. The product currently in use is called Product 1. It was determined that classification performance in a multisegment environment is sensitive to AI labeling errors. Analysis of the image product indicated significant differences in the color of wheat from one segment to another at the same stage in the crop calendar. This is attributed to the technique employed in the generation of the image product as well as to the effect of other ancillary parameters such as land use, haze and sun angle conditions.

2
INTRODUCTION

The Large Area Crop Inventory Experiment (LACIE) is an attempt to establish the feasibility of inventorying the production of wheat on a world-wide basis through the use of Landsat space image data. The experiment can be structured into four basic components: (1) an overall geographical stratification of the regions of interest, (2) a sampling strategy within strata utilizing five by six mile segments as the basic sampling unit, (3) an estimation system for wheat production within a strata, and (4) an aggregation of results. The techniques employed have shown success to date. However, the cost of the third component, the within strata estimation system, is high, primarily because each sample segment must be individually processed by an Analyst Interpreter (AI). Multisegment signature extension, the ability to infer the signature of a crop in many segments from a selected subset of segments and features, would significantly lower processing cost by reducing the amount of AI data interaction required. In addition, the stratified selection of data samples for training purposes may provide more robust signatures resulting in improved performance.

Many different approaches have been proposed to solve part or all of what is referred to as 'the signature extension problem' -- finding a technique or, more likely, a collection of techniques (a procedure) to succeed at the accurate inventory of crops over a large area through signature extension. It is the objective of this report to (1) initiate an evaluation of the overall signature extension-acreage estimation approach, and (2) perform an evaluation of the components of that approach.

The activity carried out to address these objectives was conducted in two phases.

The goal of the first phase was to provide some of the necessary information concerning the effectiveness of candidate techniques and procedures, and to identify technical needs in order to construct the overall signature extension procedures for extensive evaluation. Four signature extension techniques and related procedures were evaluated: (1) haze correction algorithms, (2) training sample selection strategies, (3) data stratification procedures, and (4) green development-trajectory classifiers.

The goals of the second phase of activity were twofold. First, the evaluation of multisegment signature extension procedures was begun through a specification of the experiment design and an initiation of preparatory phases required to conduct such an experiment. Secondly, an analysis of the cause and effect of Analyst Interpreter labeling errors was initiated. One specific concern was the sensitivity of signature extension classification results to AI labeling errors.

Section 3 of this report deals with Phase I of this project. Section 3.1 reports tests of two haze correction algorithms tested: CROP-A [1] and XSTAR [2]. Section 3.2 reports on tests of a preliminary version of a training sample selection strategy called Procedure B [3]. Section 3.3 covers evaluations of two stratifications of data: one by UCB [4] and one by JSC [5]. Section 3.4 reports on tests of several green development and trajectory classifiers, including the Delta Classifier [6] and a green development classifier. Section 3.5 is a discussion of the ramifications of the Phase I project results.

Section 4 of this report deals with Phase II of this project. Section 4.2 introduces the multisegment experiment design. Section 4.3 describes the preparatory phases of this experiment. Section 4.4 describes the Analyst Interpreter labelling error analysis carried out. Section 4.5 summarizes the observations, conclusions and recommendations derived during Phase II of this project.

3
PHASE I: EVALUATION OF SIGNATURE EXTENSION TECHNIQUES

The overall goal of this task is to evaluate the multisegment signature extension approach to large area crop inventories. Signature extension pertains to the ability to infer the signature of a crop in a group of segments based on signatures from a selected subset of segments. One motivation for this approach to crop inventory is that it would lower processing cost by reducing the amount of Analyst Interpreter/data interaction required. A second motivation was born out of research on specific signature extension techniques. The signature of a particular crop, that is, its statistical characteristics as a function of spectral, temporal and ancillary conditions, may be better understood and more accurately estimated in a multisegment environment. The goal of Phase I of this project is to study certain signature extension techniques that appear to have promise and to recommend whether the development of an accurate large crop inventory system using signature extension techniques is a feasible goal.

3.1 APPROACH

Four types of signature extension techniques or related procedures are examined:

1. Haze correction algorithms
2. Training sample selection strategies
3. Data stratification procedures
4. Green development-trajectory classifiers.

These techniques were evaluated using a compressed data base of LACIE blind sites as is described in Appendix I. That data base is known as the Fields Data Base and consists of the mean values for each field designated by an Analyst Interpreter during the LACIE operation.

3.2 HAZE CORRECTION ALGORITHMS

Two examples of haze correction algorithms were tested by this task. The first, CROP-A [1], is a cluster-matching algorithm. The other algorithm tested, XSTAR [2], employs a simplification of the ERIM radiative transfer model [7,8] to measure and correct for the effects of haze.

3.2.1 EVALUATION OF CROP-A

The cluster-matching algorithm CROP-A was tested over ten sample segments in Kansas using acquisitions from early and late May 1974 (see Appendix I.1 for a more complete description of the data set).

The form of the evaluation experiment was to perform unitemporal, matching-biophase signature extension between these sample segments, first applying signatures from one segment directly to other segments with no transformation of the mean or covariance of the signatures, and then to repeat these extensions after transforming the mean and covariance of the signatures using CROP-A transformation.

Classification results were obtained for each segment by classifying mean vectors computed from several wheat and non-wheat fields in the segment, instead of classifying every pixel. This permitted a great many classifications to be run relatively economically. That field mean classification results are strongly indicative of pixel-by-pixel classification results are shown in a study reported in Appendix II.

The performance measure used in the comparison between untransformed signature extension and CROP-A transformed signature extension was the average accuracy of the field mean classification. This average accuracy is the average of the percent of wheat field means correctly classified and the percent of non-wheat field means correctly classified.

The CROP-A experiment was carried out on a test bench known as PROCAMS. PROCAMS (PROtotype CAMS) is a system of programs developed at ERIM and is described fully in Appendix III.

The major results of the CROP-A evaluation experiment are seen in Table 1. Briefly, the classification results using CROP-A transformed signatures were not as good as the classification results using untransformed signatures.

The primary difficulty with CROP-A seems to be that it makes the assumption that the same materials are presented in both training and recognition scenes in order to make training cluster-recognition cluster pairings. This assumption is quite often not true, and can account for very large errors.

TABLE 1. COMPARISON OF FIELD MEAN CLASSIFICATION RESULTS USING LOCAL, UNTRANSFORMED AND CROP-A TRANSFORMED SIGNATURES

CLASSIFICATION USING:	NUMBER OF CASES	AVERAGE ACCURACY (%)	STANDARD DEVIATION OF AVERAGE ACCURACY (%)
Local Signatures	10 (Early May)	90.7	8.2
	10 (Late May)	87.5	10.4
CROP-A Transformed Signatures	12 (Early May)	78.3	15.0
	31 (Late May)	67.8	19.0
Untransformed Signatures	12 (Early May)	85.0	9.1
	31 (Late May)	72.9	15.5

3.2.2 EVALUATION OF XSTAR

XSTAR is a haze correction algorithm which employs a model of haze effects derived from the ERIM radiative transfer model [7]. Briefly, the XSTAR uses shifts of the data distribution in a linear combination of Landsat channels known as the yellow direction in the Tasselled Cap transformation [9] to measure the amount of haze present, and then corrects for the effects of this haze using the haze model [8]. In all tests of XSTAR, a simple cosine correction was also used to correct for sun angle effects.

The standard used to evaluate XSTAR was similar to that used for CROP-A, namely, compare classification results for untransformed signature extension and for signature extension where all data sets have first been corrected to a standard haze condition using XSTAR.

Two different experiments were conducted to evaluate XSTAR. The first was conducted using 1975-76 multitemporal (first and second biowindows*) data over 23 sample segments in Kansas for a total of 506 extensions. The second experiment was conducted using 1975-76 multitemporal (first, second and third biowindows) data over 18 sample segments in North Dakota (306 possible extensions), where the crop of interest is spring wheat. Appendices I.3 and I.4 contain a full description of these data sets.

In the Kansas experiments the performance measures used were the field mean classification accuracy and the proportion estimation accuracy. In the North Dakota experiment the true spring wheat proportions were unavailable, and so only the field mean classification accuracy was used. The LACIE fields data base as of day 315 provided the field definitions and crop type labels.

While both the field mean classification and proportion estimation results were fairly good when using XSTAR it was noted that the XSTAR corrected results were no better than the untransformed results. This was initially quite puzzling, because examination of cluster plots both before and after XSTAR correction showed that XSTAR was doing an adequate job of correction for haze and other effects.

---

*Currently, the term biowindows (or alternatively biophases) refers to a division of the crop calendar into four parts. Each division is related to important phases in the growth pattern of wheat. Biowindow 1 refers to the pre-emergent to the emergent stage. For winter wheat this would be the period from planting about September (about Julian date 285) through winter dormancy. Biowindow 2 refers to the wheat greening up period to the point of heading. Biowindow 3 is associated with post-heading and the senescent stages. The final biowindow refers to the harvesting stage in the growth cycle of wheat.

The explanation for these results is found in the method of classification used: our method of classification was to use a sum-of-likelihoods classifier with no rejection threshold. It was this lack of a rejection threshold which caused untransformed signature extension to yield results comparable to the results obtained when using XSTAR. According to the haze model used by XSTAR, the principal effect of haze is to shift the data distribution along the brightness axis of the Tasselled Cap transformed data space. It happens, however, that the principal direction of discriminability between wheat and non-wheat is orthogonal to this, parallel to the green direction of the transformed space. Thus, the decision boundary formed by the sum-of-likelihoods classifier is essentially parallel to the brightness axis. As the amount of haze in a scene varies the data distribution moves along this plane but does not cross it; thus, without thresholding, the decision boundary formed from a training site in a high haze condition was still reasonably effective in a test site with a low haze condition and vice versa.

The fact that not thresholding acts as a haze correction technique is true only because the primary direction of discriminability between wheat and non-wheat is orthogonal to the primary direction of haze shift. With crops other than wheat, this haze compensation effect will not continue to hold true. Further, it can be seen that using a threshold introduces a large bias, and significantly increases the RMS error in proportion estimation.

In the multisegment training tests on 74 winter wheat data sets over 39 Kansas segments (see Section 4) every proportion estimate using a classification threshold was less accurate than the corresponding estimate without a threshold. Examination of this result showed that in every case as the classification threshold was made smaller, the accuracy of the proportion estimates increased. A more thorough discussion of these results may be found in the interim technical memorandum "Evaluation of Signature Extension Algorithms" [10].

It is hypothesized that this increase in accuracy is due to picking up additional types of wheat which were not represented in the training segment.

Because of the effects which occur when no classification threshold is used, the North Dakota experiment was also run with and without a classification threshold.

Table 2 shows the average classification accuracy for thresholded and unthresholded classifications on XSTAR-corrected and uncorrected data. The performance of unthresholded classification on XSTAR corrected data is statistically no different than the unthresholded performance on uncorrected data, but when a classification threshold is used the performance on uncorrected data drops sufficiently to make the performance on XSTAR corrected data significantly* better than the performance on uncorrected data. The conclusion that may be reached from this is that the XSTAR correction is in fact aligning the data distributions from different sample segments, but that the unthresholded classification is unimproved because the classifier decision boundary is parallel to the principal direction of haze shift, as explained above.

TABLE 2. PERFORMANCE OF CLASSIFICATION ON XSTAR CORRECTED AND UNCORRECTED SPRING WHEAT DATA (Average of 318 Signature Extensions)

	Average Field Mean Classification Accuracy	
	Thresholded Classification**	Unthresholded Classification
XSTAR Corrected	60.10%	60.35%
Uncorrected	57.17%	61.65%

**0.001 Rejection Threshold

*The significance level of 0.01 is used throughout this report.

An analysis of the factors which were important in determining the difference between performance on XSTAR corrected and on uncorrected data indicated that the number of time periods involved in the classification was the only significant factor, although the haze level was also a significant factor at the 0.1 level. As more data acquisitions are added to the classification the chance of an acquisition with differing haze levels between the training and test sites increases, and so the uncorrected accuracy remains the same or drops in spite of the additional information in the classification, while the XSTAR corrected accuracy increases.

The conclusion to be reached from these results is that XSTAR performs a haze correction function which increases the accuracy of field mean classification and proportion estimation as compared to untransformed signature extension using a sum-of-likelihoods classifier with a rejection threshold.

3.3 TRAINING SAMPLE SELECTION STRATEGIES

Another activity pursued under this contract by another task was developing and demonstrating a training and classification technique called Procedure B [3]. This technique incorporates a training sample selection strategy together with an unconventional classification technique. In order to separate the effects of the training procedure from the effects of the classification procedure, and in order to evaluate the effect of this training sample selection strategy on a LACIE-like system, early in the contract period the PROCAMS test bench was modified to incorporate the training sample selection strategy of a preliminary version of Procedure B.

The following is a description of the resulting classification procedure, referred to as Multisegment CAMS. First, apply the training sample selection strategy of Procedure B to a large collection of LACIE sample segments. This selection strategy selects a number of sample segments as training segments. These XSTAR-corrected training sample segments are then clustered as if they were simply one large, contiguous portion of the data. The set of clusters generated (signatures) are then applied directly to all of the (XSTAR corrected) sample segments within the original large data set, using the normal maximum likelihood classifier.

In the original Procedure B demonstration, six LACIE sample segments were chosen to serve as training for all of the Kansas sample segments. In all of the following experiments, these same six segments were used for training both Procedure B and Multisegment CAMS classification. Local classification, used as a comparison, uses signatures extracted on a segment by segment basis from the Fields Data Base (see Appendix I.4 for a complete description of the data base*). Multisegment CAMS and the local classification were run without a classification threshold on the maximum likelihood classifier.

A comparison of proportion estimation accuracy for Procedure B, Multisegment CAMS, and the 75-76 LACIE procedure of local training and classification was carried out over 28 sample segments. None of the differences in proportion estimation accuracy or bias were statistically significant, due to the relatively large variance in the proportion estimates.

A comparison using 74 Kansas data sets was carried out between Multisegment CAMS and local training and classification. Again the differences in proportion estimation accuracy (variance) were not statistically significant, but now with the larger sample size Multisegment CAMS revealed a statistically significant bias.

These results did not include a bias correction procedure such as is being incorporated into LACIE. When considering an environment

*The Fields Data Base consists of a number of fields, extracted from LACIE Blind Sites, that have been designated and labeled by an Analyst Interpreter. This labeling was carried out late in the year (Julian Date 315) which enabled the AI to use all available Landsat imagery showing crop development throughout the year.

where it is anticipated that a bias correction procedure such as Procedure 1 will be used, the training gain advantage enjoyed by a method such as Multisegment CAMS is largely nullified by the need for an AI to process every sample segment anyway, for bias correction purposes. If, however, the bias of a procedure were a relatively consistent function of the true proportion (or ancillary variables), then the AI would need to process only enough sample segments to allow for the estimation of the bias correction function.

Such is the case with Multisegment CAMS. Because the same set of signatures is used for all sample segments, much of the bias is predictable. This is not true for local training and classification methods. In the 74 data sets over Kansas, bias which was a function of the true proportion of wheat accounted for only 5% of the error in the local training and classification procedure, as compared to 30% of the error in the Multisegment CAMS procedure.

Thus a linear bias correction rule trained over only the six original training segments and then applied to the proportion estimates for all of the data sets considerably improves the accuracy of Multisegment CAMS, while the accuracy of local training and classification is affected relatively little.

The difference in proportion estimation accuracy (variance) between Multisegment CAMS (as bias corrected) and local training and classification (corrected or uncorrected) is statistically significant at the 5% level. Neither of the biases are statistically significant.

The above results indicate that a Procedure 1/CAMS system, modified to incorporate the Multisegment CAMS training and bias corrected procedures, might enjoy a large training gain advantage, together with increased accuracy, as compared with the 75-76 LACIE procedures. It is also possible that a Procedure 1/Multisegment CAMS system would be more consistently accurate (in addition to being less expensive to run) than a Procedure 1/local CAMS system if the AI's turn out to have a large or randomly varying bias because of the consistent estimable bias of Multisegment CAMS.

ORIGINAL PAGE IS
OF POOR QUALITY

3.4 DATA STRATIFICATION

Data stratification is the grouping of segments on the basis of similarity in segment physical features which affect the performance of signature extension. The primary difficulty in stratifying the data is that it is not known which features of a segment (which we will hereafter refer to as ancillary variables) affect the performance of signature extension.

For this reason the emphasis of the task in this area was twofold. First, examine existing stratifications of the data and determine their relationship to signature extension performance. Second, use the actual performance of signature extensions to determine what factors are most important in determining signature extension performance.

3.4.1 EXAMINATION OF AVAILABLE DATA STRATIFICATION

Two data stratifications were available for testing. The first of these was developed by the University of California, Berkeley (UCB) [4], and the second was developed by Johnson Space Center (JSC) personnel [5].

The UCB stratification was first examined in conjunction with the CROP-A evaluation, using unitemporal Landsat data, collected in May 1974 over 10 segments in Kansas. The UCB stratification was broken down into three levels of coarseness: the original UCB stratification, a coarser version of the original stratification, and an even coarser version which ignored soil type differences.

The performance of within-strata signature extensions was then compared to the performance of across-strata extensions, for each of the three coarseness-levels of the UCB stratification, and for both CROP-A transformed and untransformed signature extensions. The result was that there was no statistically significant difference between within-strata and across-strata signature extension performance, regardless of whether CROP-A transformed or untransformed signatures were used. This seemed to indicate that the stratification was too fine, and that a much coarser stratification would probably suffice.

The UCB and JSC stratifications were later examined much more carefully during the evaluation of XSTAR on 1975-76 multitemporal Landsat data collected over 23 sample segments in Kansas (see Appendix I.3 for a complete description of the data). The form of the evaluation experiment was to first perform all signature extensions possible among the 23 segments (a total of 506 extensions) first using untransformed signature extension, and then using XSTAR-corrected signature extension. The field mean performance of each of these extensions was then tabulated, and the field mean performance of the within-strata extensions was compared to the field mean performance of the across-strata extensions.

The original UCB stratification is composed of four parts: a very fine soil stratification, a stratification based on land use and irrigation in the segments, a stratification into three groups based on a ten-year average of degree days for the segments, and a stratification into four groups based on a ten-year average of the amount of precipitation in a segment. These four parts of the stratification are then combined (via a Cartesian cross-product of the three) to produce what is referred to as the UCB data stratification. The soil stratification resulted in a partitioning of our 23 data segments into 23 partitions. As a result signature extension analysis could not be carried out. Our analysis was therefore restricted to three parts.

Each of the three component parts of this stratification were then examined in combination and separately as well.

The difference between the within-strata accuracy and the across-strata accuracy in classification of field means was not found to be statistically significant when the land use/irrigation portion of the UCB stratification was used to stratify the data.

Stratifying using either the degree day portion or the precipitation portion of the UCB strata produced a difference between within-strata accuracy and the across-strata accuracy which was significant at the 0.05 level. Within-strata accuracy was 72.8% for degree days strata and 82.4% for precipitation. Across-strata accuracy was 67.3% and 66.2% respectively.

The greatest difference between within-strata and across-strata accuracy was found when the degree day and the precipitation portions of the UCB stratification were both used to stratify the data into a total of twelve groups. Within-strata accuracy was 86.5% and across-strata 66.6%. This difference was significant at the 0.001 level.

An observation made from this analysis is that since precipitation and degree days are related to crop development, the primary effect of the successful portions of the UCB data stratification is to insure a similar degree of crop development in both the training and test segments.

The analysis of the JSC data stratification was somewhat different. Because none of the components of the stratification were available to us, no analysis of the components could be conducted. JSC strata defines "groups" and "subgroups". Three levels of generalization of the JSC stratification were analyzed at a "group" level. First, the performance of the "suggested" training segment-test segment extensions were analyzed. Second, the performance of extensions from any segment designated as a training segment to any segment designated as a test segment (both within the same strata) was examined. Third, the performance of extensions between any segments within the same strata was evaluated. In all three cases the accuracy of the extensions under examination were compared to the average across-strata signature extension accuracy. The "subgroups" defined in the JSC data stratification were ignored in these evaluations, since none of these subgroups had more than one of our testing segments in them.

Analysis of the first level of generalization, i.e., the "suggested" extensions, could not be effectively carried out since it was found that there were only two examples of such extensions within our data set, hence no significant results could be obtained.

Fourteen out of the 506 possible extensions were between designated training and designated test segments in the same strata, the second level of generalization. The field mean accuracy of these fourteen was not much different than the average field mean accuracy, and what difference there was was not statistically significant.

The third level of generalization of the JSC stratification examined, where all extensions within the same strata were compared to the across-strata extensions, had a different result. The average of the field mean accuracies of the within-strata extensions was found to be significantly higher than the average across-strata accuracy (70.5% vs. 62.6%).

Whereas the JSC stratifications yielded less substantial improvement in the field mean accuracy than the UCB stratification, the important issue realized is that partitioning of segments does yield improved performance in field mean accuracy and therefore potentially useful in a multisegment environment wherein proportion estimates are of interest. In addition, the UCB strata analysis indicated that physical variables associated with crop calendar afforded the best results. This underlines the importance of accurate crop calendar information. It is our judgement that a similar analysis of JSC component variables would yield the same observation.

3.4.2 RELATIONSHIP OF ANCILLARY INFORMATION TO SIGNATURE EXTENSION PERFORMANCE

For each signature extension technique there is a unique best stratification of the data which matches the assumptions on which the development of the technique was based.

Thus, logically, one would need to choose a signature extension algorithm and then choose a data stratification to match that particular algorithm. The simplest method to obtain the data stratification for a particular algorithm is to use the actual performance of the algorithm on various test-training pairs to determine what test segment-training segment differences affect classification performance. This is what was done for both XSTAR corrected signature extension and for untransformed signature extension.

The technique used to investigate the relationship between various ancillary variables and the performance of signature extension between those segments is a fairly straightforward one.

First, train separately on every site in the test set and then extend each of these sets of training statistics to every other site in the test set.

Secondly, pair the performance figures obtained from each of the signature extensions with a list of ancillary variables which describe the extension.

Third, use this list of ancillary variables to describe or characterize the successful extensions.

This characterization of the successful signature extensions can then be used to derive the "best" stratification for the particular signature extension algorithm used in the first step. This is done by using the characterization of the successful extensions (possibly a linear equation in the ancillary variables) to predict which extensions are most likely to be successful. These pairs of extensions with the best predicted performance are then said to be within the same strata, and thus the stratification is complete.

This process was carried out first using 1975-76 Landsat data over 23 segments in Kansas (see Appendix I.3 for a complete description of this data set), and later using 1975-76 Landsat data over 18 segments in North Dakota (see Appendix I.4 for a complete description of this data set). The list of ancillary variables used in performing this analysis is shown in Table 3.

Using the Kansas data set, the experiment was first carried out using untransformed signature extension, as a control case. The characterization of the successful signature extensions was accomplished

TABLE 3. LIST OF ANCILLARY VARIABLES

I. GENERAL:

Degree Days (10 Year Average)

Land Use (% Agriculture)

Precipitation (10 Year Average)

Latitude

Longitude

Elevation

II. PASS SPECIFIC (Calculated for Each Pass):

Sun Angle

View Angle

Julian Date

Crop Calendar (Robertson Scale) [4]

Difference Between Sites in Mean of Soils Area in Landsat Space

Difference Between Sites in Mean of Green Development Area in Landsat Space

Haze Diagnostic Calculated by XSTAR from Yellow Shift of Data

Difference Between Sites in Additive Factor Calculated by XSTAR

Difference Between Sites in Multiplicative Factor Calculated by XSTAR

Haze Value Calculated by XSTAR from Yellow Shift of Data using a stepwise linear regression technique. The results of this stepwise linear regression are given in Table 4 below.

TABLE 4. RESULTS OF STEPWISE LINEAR REGRESSION OF UNTRANSFORMED SIGNATURE EXTENSION RESULTS VS ANCILLARY INFORMATION

Important Factors	Cumulative Standard Error	Cumulative R^2
DIFFERENCE BETWEEN TRAINING AND TEST SITE OF:		
Mean of Soils Region in Landsat Space, Biowindow 1	14.50	0.124
Longitude	14.27	0.153
View Angle, Biowindow 1	14.14	0.170
XSTAR Additive Factor, Biowindow 2	14.05	0.183
Crop Calendar, Biowindow 2	13.98	0.192
Sun Angle, Biowindow 2	13.82	0.212

The final regression equation incorporating all of these factors was used to predict performance of untransformed signature extension between various pairs of sites. The predicted performance can be used to generate a stratification which meets training gain or performance criteria specified by the user. When the desired training gain was 1.2, four out of the 23 sites were classified by signature extension rather than local training, a savings of 20% in training cost. Using this 1.2 training gain stratification the proportion estimation bias in this 23 segment sample is not statistically significant.

This experiment was then repeated using XSTAR, in place of untransformed signature extension. Table 5 shows the results of the stepwise linear regression of XSTAR's results versus the ancillary information.

TABLE 5. RESULTS OF STEPWISE LINEAR REGRESSION OF XSTAR CORRECTED SIGNATURE EXTENSION RESULTS VS ANCILLARY INFORMATION

Important Factors	Cumulative Standard Error	Cumulative R^2
DIFFERENCE BETWEEN TRAINING AND TEST SITE OF:		
Mean of Green Development Region in Landsat Space, Biowindow 1	15.461	0.080
Longitude	15.176	0.116
Crop Calendar, Biowindow 2	15.031	0.134
Latitude	14.937	0.146
Sun Angle, Biowindow 2	14.853	0.158

This regression was used to define stratification of the data as was done with the regression equation obtained for the untransformed signature extension case. Proportion estimation results for XSTAR corrected signature extension using the 1.2 training gain stratification again, does not have a statistically significant bias.

When the above experiments were repeated using 1975-76 Landsat data over 18 North Dakota segments, the resultant regression equations accounted for so small a portion of the total variance in field mean accuracy it was useless in determining a stratification of the data. The conclusion to be drawn from this result is that all of the eighteen North Dakota sites were within the same stratum, as far as could be discerned using our list of ancillary data.

3.4.3 THE UTILITY OF STRATIFICATIONS OF THE DATA

Section 3.4.1 illustrated that static data stratifications based on similarities between segments in average degree days and average precipitation yield a considerable improvement in field mean classification accuracy. Section 3.4.2 showed that other, often pass-specific ancillary variables could be useful in a data stratification, and that such stratifications could be used to significantly lower the operating cost of a large area crop inventory system.

It appears, therefore, that the stratification work done by UCB and JSC should be extended to include dynamic or pass-specific ancillary variables. These data stratifications should also be evaluated in a multisegment training environment.

3.5 GREEN INDICATOR AND CROP DEVELOPMENT CLASSIFIERS

The general approach taken by signature extension classification techniques has been to use some aspect of the wheat growth pattern as viewed by Landsat as a criterion for classification. Classifiers based on a green indicator calculate a "green number" from the Landsat data, and claim that during some period of time only wheat pixels will display green numbers within a certain range. Crop development classifiers are more sophisticated; they employ a model of what wheat looks like to Landsat as a function of time of year to classify wheat from non-wheat.

3.5.1 TESTS OF SEVERAL CLASSIFIERS

The performance of several green indicator classifiers was investigated using 1975-76 sample segment data over 23 Kansas blind sites (see Appendix I.3 for a more complete description of this data set). The formulas for the green indicators tested are shown in Table 6.

For each of these green development indicators a decision threshold was trained over all of the field means in all of the test sites, and the field mean classification accuracy was noted. This procedure was applied to the first biowindow and second biowindow passes separately, and then repeated using XSTAR haze corrected data. Table 7 summarizes these results for Biowindows 1 and 2.

TABLE 6. GREEN DEVELOPMENT INDICATORS AND THEIR FORMULAS

Name	Formula*
G	CH 1 - CH 4 + 96
TVI	$\sqrt{(CH\ 4 - CH\ 2)/(CH\ 4 + CH\ 2) + 0.5}$
Ratio 7/5	CH 4/CH 2
Tasselled Cap Green	(CH1 x -0.28972) + (CH2 x -0.56199) + (CH3 x 0.599153) + (CH4 x 0.49070)

TABLE 7. PERFORMANCE OF GREEN DEVELOPMENT INDICATORS

	Average Field Mean Accuracy (percent):			
	Untransformed Data		XSTAR Corrected Data	
Indicator	Bio 1	Bio 2	Bio 1	Bio 2
G	70	82	72	84
TVI	77	81	76	81
Ratio	76	81	75	82
Tasselled Cap Green	76	80	72	80

*CH1 through CH4 correspond to Landsat Bands 4 through 7.

These field mean classification accuracies imply that the green development indicators hold considerable promise as proportion estimators. Results of pixel-by-pixel proportion estimation over 23 segments using the G indicator in Biowindow 2, and the TVI indicator in Biowindow 1 displayed a very large bias of about 10-16%. Further, the variance of the error in proportion estimation for these indicators was very large. This seemed to show that a more sophisticated approach was required than the "if it's that green then, it must be wheat" model employed by these green indicator classifiers.

The Delta Classifier does use a more sophisticated model of wheat development. Accordingly, we used this technique to classify each of the 23 test sites, comparing the field mean classification accuracy of the Delta Classifier to ancillary information via a regression. It was concluded that such a classifier must include ancillary variables in the decision rule, so that the stage of crop development can be more accurately known.

3.5.2 CROP DEVELOPMENT INVESTIGATIONS

An investigation into the properties of wheat development and discriminability was initiated with the purpose of determining what information was necessary to construct an accurate crop development classifier. The first step of this investigation was to determine what information was needed to discriminate wheat from non-wheat. Two questions were asked. First, what combinations of passes over a site are needed during the growing season? And second, is Landsat data two dimensional?, (i.e., do the first two channels of the Tasselled Cap transform, brightness and greenstuff, contain by far the majority of the information needed for spectral discrimination)?

To investigate each of these questions, 322 signature extensions were carried out using five acquisition dates from the 1973-74 data over 12 Kansas sites.

The data set contained passes from five dates: 20 October, 20 April, 9 May, 27 May and 12 June. All combinations were tested for performance both locally and in signature extension. The best single date was 20 April, with 9 May and 27 May trailing in accuracy by 5 and 10% respectively. The combination of 20 October and 20 April proved to be the best combination of passes with no other combination approaching this accuracy.

Investigating the information distribution in the Tasselled Cap transform it was confirmed that most of the information needed to distinguish wheat from non-wheat is contained within the first two components of this transform, namely brightness and greenstuff. It was shown that the classification accuracy using these two channels was only about 3% less than the accuracy using all four Landsat channels.

The results of this investigation guided us in the next step of the investigation, which was the development of a statistical model of wheat development. The data base used for this modeling effort consisted of field means and ancillary information about those fields, drawn from 74 multitemporal data sets over 39 Kansas ITS and blind sites. Appendix I.4 gives a complete description of the sites and the ancillary information used.

This empirical modeling has resulted in a pair of models which predict the green and brightness development of a wheat pixel during the second biowindow based on a statistical regression on the first biowindow Landsat signal with ancillary data.

The green development model incorporates the following ancillary information (listed in order of importance):

- Number of days into the growing season when data was acquired
- Amount of greenness displayed by green development arm of the Tasselled Cap
- Crop calendar
- 10-year average of degree days

The brightness model incorporates these ancillary variables (again, in order of importance):

- Average brightness of scene
- Brightness displayed by green development arm of Tasselled Cap
- Greenness displayed by green development arm of Tasselled Cap
- Sun angle

These two models were combined in a Development Model Classifier, in the same manner as the Delta Classifier incorporates a crop development model. The decision boundary of this classifier was then trained on the second biowindow of all 74 Kansas data sets, which resulted in an average field mean classification accuracy of 78.1%. When the normal maximum likelihood classifier was trained on all 74 data sets the resulting accuracy was 75.4%, showing that inclusion of the ancillary information into the decision rule via the two models improved field mean classification accuracy.

In order to determine the stability of these models, the coefficients of the models were redetermined using 81 fields from 12 randomly selected data sets. The coefficients of the models developed on only 12 data sets were quite similar to the coefficients of the model developed using all 74 data sets.

As a further test of similarity, the new models were incorporated into a Development Model Classifier and the coefficients of the classifier were then trained over these same 12 data sets; thus the classifier was constructed using information from only 81 fields in 12 data sets. This classifier was then used to classify all 74 data sets, resulting in an average accuracy of 76.5%. Table 8 shows how the accuracies of several other classifiers compare to this accuracy.

The results of this modeling appear encouraging enough to warrant further testing and development in the future. Of particular interest would be a model which was applicable throughout the crop year. Such a model could provide an ideal AI key, as well as the basis for a classifier.

TABLE 8. COMPARISON OF SEVERAL CLASSIFIERS

Classifier	Number of Landsat Acquisitions Used	Field Mean Classification Accuracy (Average Over 74 Data Sets)
Development Model Classifier (trained on 12 data sets)	2 (Biowindows 1,2)	76.5%
Maximum Likelihood (trained on all 74 data sets)	1 (Biowindow 2)	75.4%
Delta Classifier	3 (Biowindows 1,2 or 3,4)	70.1%
Multisegment CAMS	4	74.0%

3.6 PHASE I: CONCLUSIONS AND RECOMMENDATIONS

The development of an accurate large area crop inventory system using signature extension techniques is a feasible goal. As we understand it now such a system would employ haze and sun angle corrected data in a multisegment training and classification scheme which would be applied within some stratification of the data. Support for this view of signature extension is contained in the following discussion of conclusions about each of the four types of signature extension algorithms tested.

Two examples of haze correction algorithms were tested: CROP-A [1] and XSTAR [2].

CROP-A was tested in a unitemporal mode on data collected in 1973-74 over ten sample segments in Kansas. Because of the uniformly low level of haze present in these segments, no conclusion could be reached about CROP-A's ability to compensate for haze. It was noted, however, that in some cases CROP-A made serious errors which actually degraded classification performance. For this reason CROP-A was deemed unsuitable for general application in large area crop inventories, and was dropped from further consideration.

The haze correction algorithm XSTAR was tested in a multitemporal mode on 1975-76 LACIE sample segment data over 23 blind sites in Kansas and 18 sample segments in North Dakota, providing a wide range of haze levels and other conditions for evaluation of the algorithm. It was found that this algorithm substantially improved signature extension classification accuracy when a sum-of-likelihoods classifier was used with an alien rejection threshold. Further, the accuracy of classification using the XSTAR haze correction was substantially the same regardless of haze level or differences between the test and training sites.

An interesting and useful observation made during the tests was that when no alien rejection threshold was used in the sum-of-likelihoods classifier, untransformed signature extension achieved the same level of classification accuracy as XSTAR haze corrected signature extension. The explanation for this not totally expected result is that the wheat/non-wheat decision boundary is typically nearly parallel to the principal direction of shifts in the data due to haze. Thus classification accuracy is often little affected by haze level differences between test and training sites given that no alien rejection threshold is used in the classifier, that the only class of interest is wheat and that the appropriate acquisitions are available.

The training sample selection strategy available for testing at this time was a preliminary version of Procedure B [3]. This training sample selection strategy was used to select six sample segments as training for all Kansas sample segments, a training gain of almost 12 to 1 (12 recognition sites for each training site). Multitemporal proportion estimation results obtained by using the six selected sample segments as training for classification of 74 multitemporal data sets were extremely encouraging, and in fact were not statistically different from multitemporal local training and classification proportion estimation results (i.e., using all 74 data sets for training).

One of the major findings of the above study was that nearly all of the bias in the proportion estimates of the multisegment training and classification procedure resulted from the particular configuration of the signature set used for classification, rather than from peculiarities of the recognition sample segments. This meant that the proportion estimation bias could be accurately corrected simply by estimating the bias on the original six training segments. The bias corrected proportion estimates of the multisegment training and classification procedure were extremely accurate and had a low variance when compared to local training and classification. This finding may have important ramifications for reducing the cost and increasing the accuracy of bias correction procedures.

The third category of techniques and procedures examined was stratification of the data. Two stratifications of the data were available, one carried out by the University of California, Berkeley [4] and another derived at JSC [5]. These stratifications were evaluated by comparing the performance of within-strata and across-strata signature extensions, both before and after XSTAR haze correction, using multitemporal sample segment data. Both of these stratifications significantly and substantially improved signature extension classification performance.

The primary beneficial effect of these stratifications seemed to be that they matched together segments with the same stage of crop development. It was shown that these stratifications could be improved by incorporating certain dynamic or pass-specific ancillary information about the segments into the stratification procedure. These data stratifications require further evaluation in conjunction with a multisegment training and classification system.

The fourth category of signature extension techniques examined was that of green indicator and crop development trajectory classifiers. It was found that such classifiers can be made robust enough to be applicable to a broad range of sample segments, and probably without needing to be retrained each year. However these classifiers also displayed an unacceptably high variance in proportion estimation accuracy, due to the existence of a fairly large number of sample segments with unusual development patterns.

It appears that in order to make such classifiers sufficiently accurate for current day needs they will need to be modified to incorporate sufficient ancillary information (such as a crop calendar) into the decision rule to account for sample segments with atypical development patterns. The crop development modeling undertaken by this task has been a first step towards solving this problem.

A recommendation of this task is that a further evaluation experiment be carried out which closely examines the potential of the multisegment training and classification approach to signature extension. Such an evaluation should also include an examination of the usefulness of haze correction and data stratification techniques in a multisegment environment.

4
PHASE II: EVALUATION OF MULTISEGMENT SIGNATURE EXTENSION PROCEDURES

Phase I of this task addressed the evaluation of signature extension techniques. The goals of the second phase of activity were two-fold. Of first concern is the evaluation of multisegment signature extension procedures. That is, an analysis of the effectiveness of systems that incorporate those techniques evaluated in Phase I. The second concern of this phase of activity relates to an analysis of the Analyst Interpreter's role in a multisegment signature extension environment. Phase II has been carried out with the expectation of continued test and evaluation of the signature extension approach through the next contract year. Three specific activities were carried out:

1. The definition and advanced design of an experiment to examine the overall signature extension approach
2. Preparatory phases to conduct such an experiment
3. Analysis of the nature of analyst interpreter errors and the sensitivity of the signature extension approach to analyst interpreter errors.

4.1 BACKGROUND

The LACIE Phase III operation employs a classification and mensuration strategy called Procedure 1 [11]. Procedure 1 provides an environment wherein a large number of domestic or foreign 5x6 mile segments are classified using local training procedures. Crop proportion estimates for wheat are computed and bias corrected. Training is accomplished by clustering all pixels within a segment. The clustering algorithm is seeded by a subset of labeled dots derived from 209 points that occur at the nodes of a 10x10 pixel grid superimposed on the LACIE segment

The clusters are named wheat or non-wheat by their association to another subset of the 209 points that have been labeled by an analyst interpreter using false color photo image interpretation techniques. These clusters are then used to classify every pixel in the segment from which they were derived using a sum of likelihood quadratic classifier. Proportion estimates are derived for wheat and non-wheat and bias corrected by multiplying the estimates using a performance matrix derived from a third subset of the 209 dots. The procedure is labor intensive in that each segment must be processed by an intervening analyst interpreter. Proportion estimates are, in addition, sensitive to AI labeling errors.

The multisegment signature extension environment is one wherein an attempt would be made in reducing the need for local training. That is, to process certain segments automatically without an intervening analyst interpreter. A certain subset of segments would be designated training sites. Training data would be derived from these segments and used in classification throughout. Hence, specific segments can be more intensely photointerpreted for training, hopefully with a resultant reduction in labeling error.

The multisegment signature extension approach, however, poses a twofold requirement: an appropriate training segment selection approach, and a bias correction approach employing non-local performance expectations. Any operational system addressing the multisegment signature extension approach to large area crop inventories is operating under the one basic constraint that the smallest sampling unit is a 5x6 mile LACIE segment.

Research in signature extension has been based on selecting a minimal set of training segments within a given area stratification. This requires that a given area to be mensurated must first be stratified into partitions of relatively homogeneous class characteristics. A multisegment signature extension test and evaluation experiment must examine proposed partitions for signature extension as well as classification and mensuration procedures within the context of these partitions. Hence the overall objectives of this investigation will be to evaluate current UCB [4] and JSC [5] signature extension stratifications to determine if these products:

1. Increase the efficiency of the multisegment training selection technique termed Procedure B, and
2. Provide an efficient means for sampling to be used for classification and mensuration employing a Procedure 1 operation extended into a multisegment environment.

4.2 ADVANCED MULTISEGMENT SIGNATURE EXTENSION EXPERIMENT DESIGN

4.2.1 APPROACH AND DESIGN SUMMARY

The design of a multisegment signature extension experiment requires a specification of five basic components of an experiment. These components include:

1. The systems under test
2. The performance measures
3. The measurement procedures
4. The parameters, factors, and levels desired
5. The data sets.

Each of these components are described in the following sections provided to more specifically detail this experiment. An overview of the experiment is provided in the following.

The overall signature extension approach to large area crop inventories operates within the basic constraint that 5x6 mile Landsat data segments are the basic sampling unit in estimating the proportion of crops within a region of interest. The experiment to be conducted will evaluate three procedures designed to function in a multisegment environment. Each of these three procedures will be evaluated in light of specified static stratifications of data.

The three procedures to be evaluated shall be termed 'Multisegment Procedure 1', 'Procedure B' and 'Modified Procedure B'. The third procedure is a hybrid of the first two incorporating the training strategy of Procedure B and the estimation strategy of Procedure 1.

The static data stratification to be examined includes: 1) a local strategy wherein each segment is its own stratum; this is equivalent to the current Procedure 1 strategy; 2) fixed boundary strategy as defined by UCB and JSC; and 3) arbitrary strategy wherein all available segments are in one stratum. Hence we will examine strategies that, for m segments, define either m strata, or one stratum, or some number, n, between these extremes. The first strategy can be thought of as a 'Baseline' strategy since it currently is LACIE operational.

Each specified multisegment crop inventory procedure will be evaluated in light of each of the three categories of data stratification. The fixed boundary stratification strategy will, in addition, evaluate three approaches to training and classification for each procedure: 1) within strata training, within strata or local classification, 2) within strata training, across strata or global classification, and 3) within strata training, weighted global or across strata classification. Figure 1 flowcharts the experiment as described to this point.

In addition to the evaluation of the specific procedures in their overall performance with respect to ground truth and the current LACIE approach, the sensitivity of each procedure as a function of a number of parameters will be examined to some extent. Of particular interest is the behavior of these approaches in light of certain data preprocessing algorithms, specifically haze and sun angle external effects corrections and data compressions using the greenness and brightness channels of the Tasselled Cap transformation and/or BLOB spatial/spectral clustering. Another very important measure of each system is performance as a function of training gain. Other procedure-specific parameters will be analyzed as described in Section 4.2.5.

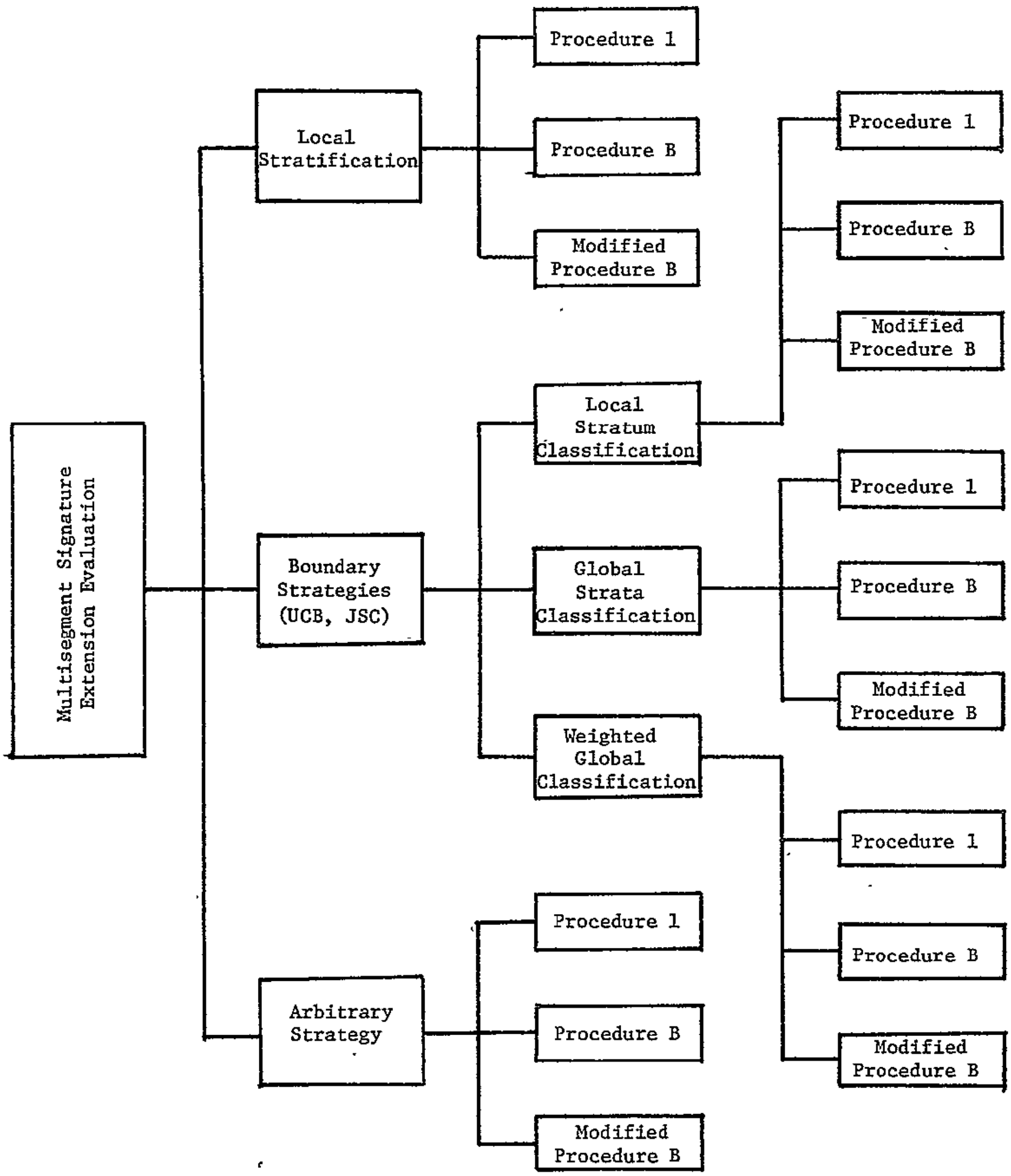

FIGURE 1. FLOW DIAGRAM OF MULTISEGMENT SIGNATURE EXTENSION PROCEDURE EVALUATION

Performance measures of interest include not only performance accuracy but also strata wide performance based on the distribution of performances from individual segments. Data to be employed will consist of LACIE blind sites in the Great Plains as described in Section 4.2.6.

4.2.2 SYSTEMS UNDER TEST

A multisegment signature extension classification and mensuration system employing space image data is comprised of four basic components:

1. Data preprocessing requirements
2. A training strategy
3. A proportion estimation strategy
4. Post classification bias correction strategy.

The training strategy involves both the training sample selection strategy and signature determination. Keep in mind that the sampling strategy requires the selection of training pixels or fields constrained to specific 5x6 mile segments within a given stratification of data. Signature determination is the process of establishing information representative of the classes of data or specific features of interest within strata. Various classes of signature determination strategies are available. One prominent strategy applies to statistical modeling of classes. This strategy assumes that the data is Gaussian or Normally distributed. Another strategy may employ analytic and empirical signature modeling. We shall restrict our analysis to statistical strategies.

The systems to be considered in this test and evaluation of multisegment signature extension procedures are illustrated in Table 9.

TABLE 9. PRINCIPAL PROCEDURAL STRATEGIES FOR TEST AND EVALUATION

DATA PREPROCESSING	TRAINING SELECTION	 SIGNATURE DETERMINATION	PROPORTION ESTIMATION	POST BIAS CORRECTION
SUN ANGLE CORRECTION	UCB STRATA - Random Selection - Procedure B	PROCEDURE B	WITHIN STRATA - Procedure B - Sum of Likelihoods • all pixels • blobs • 209 dots	PERFORMANCE MATRIX CORRECTION (Procedure 1)
HAZE CORRECTION	JSC STRATA - Random Selection - Procedure B	CLUSTERING - Procedure 1 (209 dots) - Field Means - Blobs	ACROSS STRATA - Procedure B - Sum of Likelihoods • all pixels • blobs • 209 dots	REGRESSION VS. ESTIMATE (ERIM)
DATA COMPRESSION - BLOB - Tasselled Cap - Manually-Defined Fields	ARBITRARY STRATA - Random Selection - Procedure B			REGRESSION VS. ANCILLARY DATA (TAMU)

Numerous components are specified. An operational procedure employs a subset of these components and may traverse different paths. For example, one approach may (1) employ haze corrected pixel data, (2a) Procedure B training selection strategy within UCB stratification, (2b) determine signatures by clustering pixels, (3) employ sum of likelihoods classification, and (4) bias correct as in Procedure 1. It is not feasible to examine all possible paths through this array of procedural components. In addition, many systems with potential in a multisegment environment are not herein specified. For example, the proportion estimation strategies specified may rely on multitemporal acquisitions of data. Numerous multitemporal classifiers have been proposed. Further testing of these, however, is required outside of the multisegment framework. The systems proposed herein are those that have been in our opinion tested adequately to warrant further examination in the multisegment environment.

The performance of these procedures must be evaluated not only with respect to one another, but also with respect to a base line system. That system will be the standard Procedure 1 employed in a local or single-segment environment.

The principal procedural strategies indicated in Figure 1 operate within a partitioning framework. These strategies primarily include Procedure B and a version of Procedure 1 adapted to the multisegment environment. A composite system wherein a Procedure B training segment selection strategy is employed and a Procedure 1-like estimation strategy is used in conjunction with the training strategy is another conceivable processing strategy to be tested. The next two sections are presented to provide information with regard to Procedure B and Procedure 1 training strategies in a multisegment/partitioning environment.

4.2.2.1 Generalized Procedure 1 Training Strategy

We have noted that the multisegment signature extension approach poses a training segment selection problem. Resultant classification is sensitive to variational differences between training and test segments. Procedure 1 employs a single segment or local approach to training and classification to eliminate those differences. Extending Procedure 1 to a multisegment environment requires partitioning segments into 'like' groupings. The designation of these static stratifications using physical variables such as soil type and precipitation is an attempt to associate segments in a manner that would minimize the spectral differences between like classes in segments belonging to the same strata. These strata can be used in two ways:

1. For Training Selection Purposes: To insure that all spectral classes are represented in choosing segments from every strata to be used across all segments in classification.

2. For Classification Purposes: Segments would be classified using training data determined within their strata only.

In either case the Procedure 1 training strategy must be carried out in a multisegment environment. The following is a generalization of the signature extraction strategy to which Procedure 1 can be easily adapted.

Consider n strata and m segments where $n \leq m$. Segment s_{ij} is the j^{th} segment of the i^{th} strata S_i. Let the signature set for segment s_{ij} be $SIG(s_{ij})$. Let the training data for stratum S_i be $T(S_i)$. Call the Procedure 1 clustering function $\prod$, then

$$SIG(s_{ij}) \equiv \prod_{k=1}^{n} \omega_k T(S_k) \qquad (1)$$

where ω_i is a weight for each stratum.

If

$$\begin{array}{ll} k = i & \omega_k = 1 \\ k \neq i & \omega_k = 0 \end{array} \tag{2}$$

then the strata are being used for classification purposes, i.e., the segment is classified using signatures computed within the stratum of which it is a member.

If

$$\omega_i = \omega_j \quad \text{for all } i,j \tag{3}$$

then the strata are being used for training purposes only, i.e., a segment is classified using all signatures, but insuring that each stratum is represented by training data.

The value of introducing this notation is twofold. First of all, the same signature extraction strategy currently employed locally in Procedure 1 can be employed in multisegment signature extraction. Procedure 1 is simply the case where each segment is its own stratum and ω_i is defined as in (2). Secondly in computing SIG(s_{ij}) (the signature set to be applied to segment s_{ij}) the training data from stratum S_i, ($s_{ij} \subset S_i$), may be weighted more than training data from other strata. This recognizes that important information for any one segment appears in every stratum, however, it is more likely that training data within the same strata would be more significant.

4.2.2.2 Character of the Procedure B Training Selection Process

The training segment selection strategy that would be employed in adapting Procedure 1 to a multisegment environment would likely be carried out through random selection of a number of segments to satisfy a training gain requirement. The accuracy and variance in the estimate as a function of training gain is an important factor to be measured in this experiment.

For a given training gain, one's confidence that a particular random selection of training segments is adequate would be closely related to the measured variance in the estimate given a different collection of training segments satisfying the gain requirement. Procedure B is an attempt to provide a systematic technique in training segment selection that would insure that the segments selected for training are adequate at a level of confidence higher than random selection. This approach is based on the same philosophy as static stratifications of regions by use of physical variables. That philosophy being that there are natural groupings of data, and sampling should be carried out to insure representation of these natural groupings. Whereas static stratifications base groupings on physical variables, Procedure B groups data within strata dynamically as a function of measured spectral variables. These groupings are dynamic in the sense that as additional spectral information is added, for example additional temporal acquisitions, then the spectral strata 'boundaries' may shift. Sampling is carried out to insure representation within each natural spectral grouping. The efficiency of this automatic segment selection approach in comparison to the random segment selection approach is of interest.

4.2.3 PERFORMANCE MEASURES

Evaluation of the multisegment signature extension procedures under test will be characterized by a set of performance measures. These can describe performance within a segment, within a stratification of data and across all strata. Performance measures can be descriptive or analytic.

4.2.3.1 Descriptive Performance Measures

Descriptive performance measures characterize a procedure in reference to the baseline system, in this case the LACIE Phase III Procedure 1. The three performance measures to be considered include:

1. The differences in classification error
2. The differences in wheat proportion error
3. An estimate of the overall training gain

These performance measures provide a basis for comparison between Procedure 1 and signature extension procedures employing partitioning.

4.2.3.2 Analytic Performance Measures

Analytic measures characterize the performance of a particular signature extension approach in reference to the ground truth. A primary objective of error analysis is to estimate and describe the distribution of errors over many data sets. An understanding of this distribution provides insight to the functioning of the system under test and may provide post-classification corrective measures. Analytic measures to be considered include:

1. Bias in Proportion Estimate: The displacement of the mean of the predicted wheat proportion over a set of segments or strata from the true proportion.

2. Correlation in Proportion Estimate: The degree of correlation between predicted wheat proportion over a set of segments or strata to the true proportion.

3. Mean Square Error in Proportion Estimate: The sum of the square of the distance of each estimate from the true proportion; this is a measure of the accuracy of the estimate without bias correction.

4. Variance in the Proportion Estimate: This measure is identical to the mean square error except employed after bias correction.

5. R^2: This measure is the square of the correlation coefficient; R^2 can be thought of as the percent of variation about a regression line that can be accounted for by the dependent variable in the regression equation.

Figure 2 is a display of six hypothetical test results. Each illustrates the effectiveness of various analytic performance measures in describing the results. The ground truth proportion estimate is plotted for a set of segments versus the predicted estimates. The 45° line indicates the correct estimate.

Figure 2(a) illustrates a test result that is unbiased, highly correlated to the truth and with low variance in the estimate. Figure 2(b) diagrams a biased result that is correlated with a high R^2 about the dashed regression line. Figures 2(c) and 2(d) are both uncorrelated results, however Figure 2(c) is not biased and with greater variance than Figure 2(d). Whereas the variance of Figure 2(d) is lesser, the mean square error could be greater. Figure 2(e) illustrates a biased result that is highly correlated to the truth with a very low variance. This result could be bias corrected by simply shifting it toward the 45° line. Figure 2(b) could be similarly corrected, but would result in a higher variance in error. However, a multiplicative and additive correction would result in an equivalently low variance estimate. Figure 2(f) is somewhat similar to Figure 2(c). Both results are unbiased, and both have high variance in the estimates. However, whereas the results shown in Figure 2(c) are not well correlated to the truth, Figure 2(f) is negatively correlated. This information may give added insight in the analysis of the systems under test.

4.2.4 MEASUREMENT PROCEDURES

Section 4.2.2 indicated that an evaluation will be carried out for three procedures: multisegment Procedure 1, Procedure B, and a modified Procedure B. Each of these procedures will in turn be evaluated

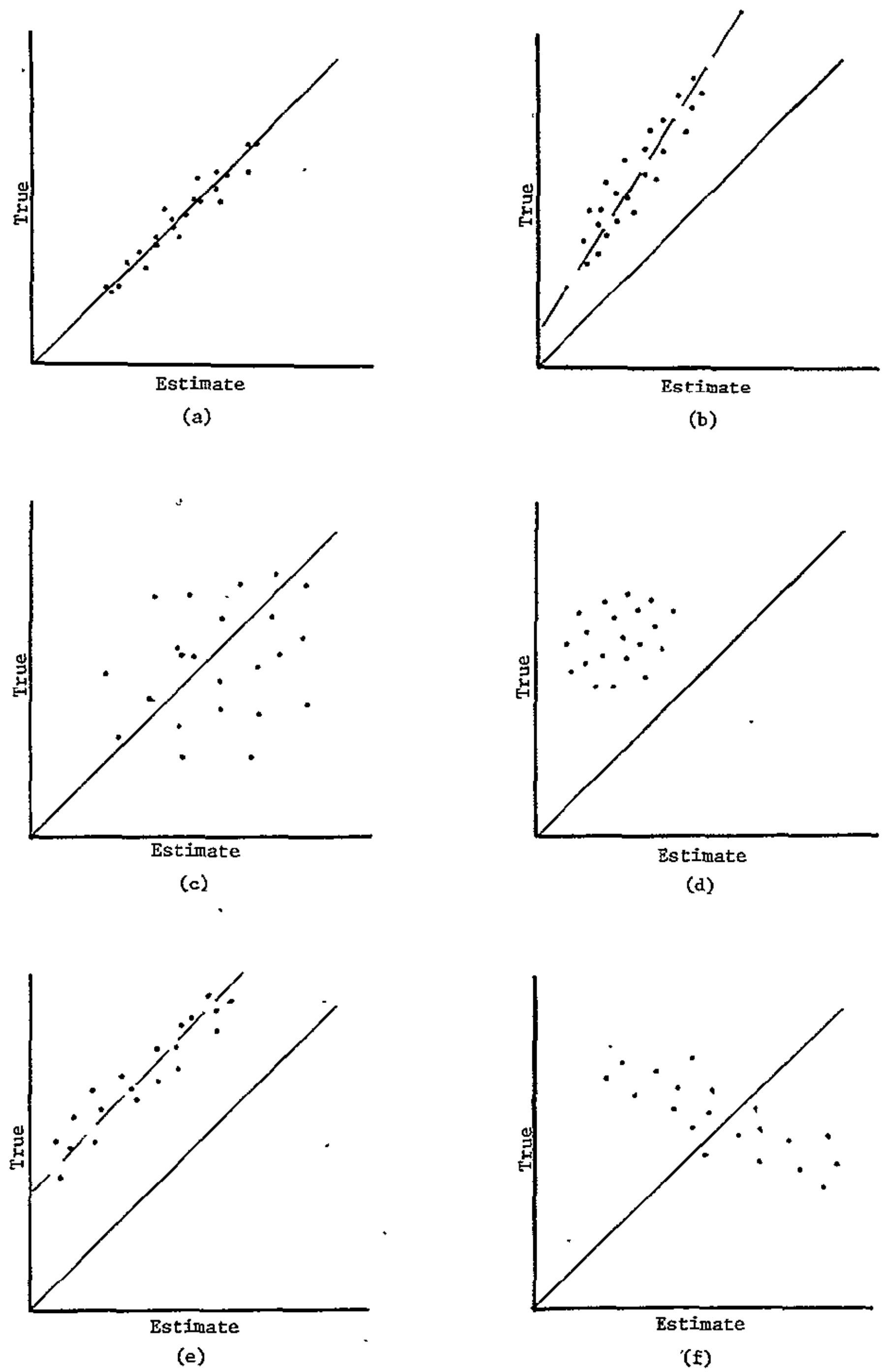

FIGURE 2. SIX HYPOTHETICAL MULTISEGMENT SIGNATURE EXTENSION TEST RESULTS

in light of four physical stratifications of data: local, UCB, JSC, and arbitrary partitions.

Any evaluation of the inherent value of static stratification in a multisegment environment will require that the measures of performance discussed in Section 4.2.3 are statistically significant. As a result a large number of classifications must be performed for a large number of segments with procedural parameters varied at each classification (see Section 4.2.5). This demands judicious selection of the data base (see Section 4.2.6) and a classification strategy that minimizes cost.

The Procedure B classification strategy is described in Reference [12]. The sum of likelihoods classification strategy is summarized in the following. Appendix IV contains a more detailed specification of this strategy.

The parameters varying most rapidly in the proposed evaluation are training parameters, for example, the number of training segments employed. Ordinarily this would require the determination of a set of signatures and computation of proportion estimates for each set of training parameters. A procedure has been devised and termed 'preclassification' which delays the need for setting training segment selection parameters until after signature determination and after classifying the data set.

The preclassification procedure to be employed in the test and evaluation of signature extension procedures is as follows:

1. Select the set of segments potentially available for training.

2. Determine signatures from each training segment independently from others.

3. Employ the following classification procedure:

   a. Classify each segment using the signatures from each other segment, determining a wheat and non-wheat likelihood (i.e., for m training segments, each segment is classified m times).

   b. Select the subset of segments to be used for training.

   c. Sum likelihoods from each training segment and determine wheat proportion estimate.

   d. For testing purposes, repeat (b) and (c) for each variation in the training segment selection process.

Proportion estimation can be carried out for a variety of training segment sets, simply by summing likelihoods corresponding to the appropriate training segments. Clustering and likelihood calculation, the two most complex operations computationally, do not have to be recomputed for each different set of training data. Appendix IV describes how this preclassification procedure is logically equivalent to a more standard approach.

4.2.5 PARAMETERS, FACTORS AND LEVELS

A number of conditions in the evaluation of specific multisegment signature extension procedures will be varied. This is carried out in order to examine the sensitivity of the procedures to various parameters. The underlying objective here is to understand not only that a specific approach is or is not successful, but to understand why as well.

Parameters of particular interest in this evaluation are listed and briefly described in the following.

1. Number of Training Segments: It is critical to evaluate the performance of an approach as a function of training gain, that is, the ratio of the total number of segments processed

to the number of segments used for training. The training gain is a measure of the system's efficiency. Hence, the number of segments used for training must be varied. Not only will the number of training segments be varied, but the specific ones employed for a specific training gain will be as well. This is required in a Procedure 1 context in order to measure the variance in the estimate as a function of the random training segment selection strategy. Concerns associated with experiment cost effectiveness resulting from this requirement have been addressed in Section 4.2.4 and in Appendix IV.

2. Preprocessing: Phase I of this project evaluated certain data preprocessing strategies and concluded that they may be of considerable value in a multisegment environment. The benefits of haze and sun angle external effects corrections and data compression in using the Tasselled Cap transformation and blobbing need to be evaluated in a multisegment signature extension environment.

3. Training Weights as a Function of Strata: Every segment to be classified may be so classified using training data from within the local strata in which it belongs as well as from other strata. Appendix IV discusses a weighting that will vary from segment to segment associating a level of confidence in the training data drawn from different strata as applied to a specific segment. Three sets of weights will be evaluated. The first associates a full confidence in training data from the local strata and a zero confidence level in all other training data. In effect physical stratification of the data is used not only for training but also for classification. A second weighting may employ an equal level of confidence

in training data independent of the strata from which it is drawn. A third weighting may employ a higher level of confidence in local stratum training data and a lesser level of confidence in other data. The third approach suggests that physical stratifications of the data are not truly static boundaries, but rather confidence thresholds. Hence a confidence weighting as a function of some distance measure may be appropriate. The nature of that distance measure is still to be investigated.

4. The Number of BLOB-Clusters: This parameter pertains to Procedure B. A blob-cluster, or B-cluster, is the spectral stratification of the data described in Section 4.2.2.2. It is a matter of investigation to analyze the sensitivity of Procedure B to the number of spectral strata employed.

5. The Random Draw of BLOBS for B-Cluster Labeling: The estimation mechanism in Procedure B requires that each B-cluster or spectral stratification be estimated by a technique wherein a random draw of BLOBS within the B-cluster are labeled and aggregated. This approach may be employed as well for the AI labeling of fields for Multisegment Procedure 1 training purposes as an alternative to dot labeling. The system's sensitivity to the number of the blobs using this approach is of concern.

4.2.6 DATA SETS

In an effort to attain statistically significant results, the data base for this experiment will contain a large number of LACIE blind site segments. However, in order to keep processing costs within reason, four compressions of the data will be considered: (1) the augmented AI Fields Data Base, (2) BLOB compression, (3) 209 dot samples, and (4) ground truth Fields Data Base.

Augmented Fields Data Base

The augmented fields data base is described in Appendix I. This represents a set of segments for which an analyst interpreter designated and labeled specific fields for training. Section 4.3 describes a process carried out to augment the data base with additional fields. This data base is drawn from Kansas and North Dakota representing both winter and spring wheat. Due to its availability, initial testing of measurement procedures and signature extraction should be carried out using this data base

BLOB Compression

BLOB is a spectral-spatial clustering technique that groups data into field-like shapes. It is of interest to us to analyze this data preprocessing technique to determine how accurately actual field shapes are estimated and more importantly, to measure the accuracy of crop proportion estimates based on BLOB classification. This technique is of particular interest in that it forms the basic unit of data in Procedure B.

209 Dot Samples

Upon overlaying a 10x10 pixel grid to a LACIE segment, 209 pixels are represented at the nodes of the grid. Currently in LACIE Phase III operations these '209 dots' are used in various stages including labeling of samples, cluster seeding, cluster labeling and bias correction. The 209 dots for our purposes represent a reasonable random sampling of the segment to be used for proportion estimation of wheat and non-wheat.

Ground Truth Data Base

A task is currently underway wherein a number of LACIE blind sites in the Great Plains are being processed to incorporate ground truth, stratification and ancillary information. These data are expected to be available within a six month period. As they become available, it is our intention to phase out the use of the augmented fields data base and replace it with these data statistically summarized on a field by field basis.

4.3 FIELDS DATA BASE PREPARATION AND AUGMENTATION

One of the important efforts in preparation for test and evaluation of multisegment signature extension procedures is the development of an adequate data base. The proper selection and labeling of training fields within each test site is an essential part of the development of this data base. The Fields Data Base, used for test and evaluation of signature extension algorithms of Phase I of this project will be used initially for the extraction of signatures and testing of multi-segment signature extension procedures. To insure that the AI Fields Data Base properly represents each segment, the following procedure was carried out using LACIE Blind Site 1975-76, Day 315 Fields Data Base. This data included 38 Kansas and 18 North Dakota test sites (see Appendix I.4 for a complete description of the data base).

1. Compare AI field designations with large scale annotated ground truth high altitude photos and correct any AI labeling errors.
2. Determine the degree to which AI field selection simulates random field selection on a segment by segment basis.
3. Augment the fields data base to insure a simulated random selection process.

4.3.1 LOCATING AI FIELD DESIGNATION ERRORS

The AI designations ("wheat" or "other") of defined fields were checked against ground truth labels on aerial photographs of the scenes involved. This was done for 32 1975-1976 LACIE blind sites in Kansas and 16 in North Dakota. For each segment three accuracy measures were computed. They were defined as follows:

1. $$\text{TOTAL ERROR} = \frac{\text{total no. of mis-labeled fields}}{\text{total no. of defined fields}}$$

2. $\text{MISSED WHEAT} = \dfrac{\text{no. of fields labeled other when actually wheat}}{\text{no. of defined fields actually wheat}}$

3. $\text{FALSE WHEAT} = \dfrac{\text{no. of fields labeled wheat when actually other}}{\text{no. of defined fields actually other}}$

A summary of the accuracy figures appears in Table 10. Analyst-Interpreter accuracy on the North Dakota segments was not as good as on the Kansas segments. This may be attributable to the presence of the confusion crop barley in North Dakota and to the practice of strip cropping. Two of the segments in Kansas, No. 1164 (68.4% false wheat) and No. 1860 (54.5% missed wheat) were found to have anomolously large error figures. The number of field designations changed per segment ranged from 0 to 12, averaging about 3.3 corrections per segment. An average segment contains about 30 fields.

TABLE 10. SUMMARY OF AI ACCURACY MEASURES

	North Dakota		Kansas	
Error	Ave. Error	Std. Dev.	Ave. Error	Std. Dev.
Total	17.2%	6.7%	11.4%	8.1%
Missed Wheat	26.7%	14.7%	20.0%	10.5%
False Wheat	6.1%	5.5%	3.3%	7.4%
$\frac{\text{Missed Wheat}}{\text{False Wheat}}$ Ratio	4.4		6.4	

One observes the AI makes far fewer mistakes of labeling other crops as wheat than the reverse mistake of labeling wheat as other. The ratio MISSED WHEAT/FALSE WHEAT is 4.4 in North Dakota and 6.4 in Kansas. This indicates the presence of a source of variation in the appearance of wheat which is misleading the AI. An unknown source of variation is not likely to make a crop other than wheat look like wheat.

The AI looks at the development of a crop at key points in time, the "biophases" of wheat, and the pattern of development is central in the decision process. It is unlikely that small, random variations in the appearance of fields would cause a non-wheat crop to be shifted into this pattern. It is more likely to shift a wheat field beyond the thresholds of the wheat pattern as the AI conceives it. A statistical investigation exploring AI error for these segments is reported in a following section of this report.

4.3.2 SIMULATING A RANDOM TRAINING SELECTION

As has been described earlier, the Fields Data Base was selected to conduct the test and evaluation of signature extension algorithms in order to provide a compression of the data. This would both be representative of the individual segments and result in a cost effective analysis. Initially it was acceptable to assume that the Analyst Interpreter could accurately represent the segments through field selection. That is, the AI designated fields were representative of the segments in the sense that the variability in the data was accounted for. It became a concern, however, that introducing human interaction would bias representative selection. That is to say, the Analyst Interpreter was not properly simulating a random training field selection process. A random field selection process would insure, in a statistical sense, that the variability in each scene was properly sampled. This concern led us to establish a procedure, termed CHECK, whose function is to establish how closely AI field designations simulate random field selection. The following CHECK procedure was devised:

1. Histogram the multitemporal segment of data:
   - Tasselled Cap brightness and green channels
   - three bins per channel selected to separate observed modes
2. Histogram AI designated training pixels using the same bins.

3. Establish a criterion based on Step 1 as to which bins were significant.

4. Compare two histograms to determine whether all significant bins were represented by AI training pixels.

5. Use a histogram map (similar to cluster map) to select additional training to insure that each significant bin is represented by training pixels.

Keep in mind that the purpose of carrying out this procedure was to insure that the AI training field selection process was not biased in simulating a random training field selection process. Random training selection statistically insures that important clusters of data would be represented in proportion to their density. For example, should ten percent of a scene fall into a particular spectral class, random sampling of the scene would insure that, on the average, ten percent of the samples would fall into that spectral class. The histogram approach was used since important clusters of data would tend to fall into the same bins. By histogramming the data into bins, the AI field designation could be augmented by selecting samples from larger bins that were missed by the AI.

Using data from two acquisition dates, four channels, there were 81 possible bins or classes in which a pixel could fall. To decide which bins were most important to examine, the data was grouped according to size. The first group consisted of all bins containing more than 5% of the data, the second more than 1% of the data, the third and fourth groups were cut off at the 0.5% and 0.1% levels. Figure 3 shows a plot of bin size vs. average percent of the test site included in each group. Only 25% of the data fall in bins containing over 5% of the pixels, but 83% of the pixels are contained in the 1% level group. Figure 4 is a plot of bin size vs. the number of bins within a group. The number of bins per group ranges from three to 67.

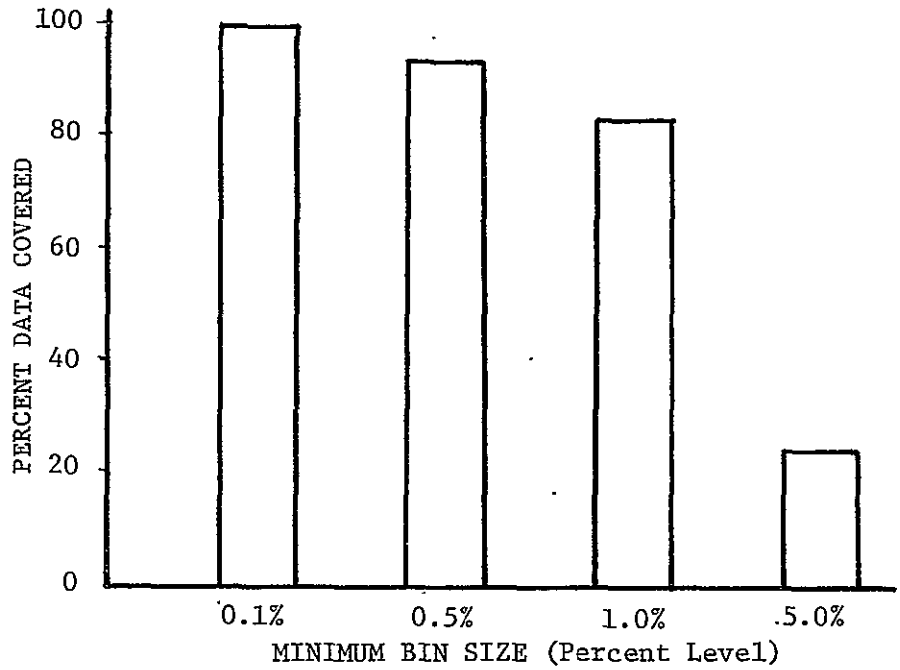

FIGURE 3. DATA COVERAGE VS. MINIMUM BIN SIZE

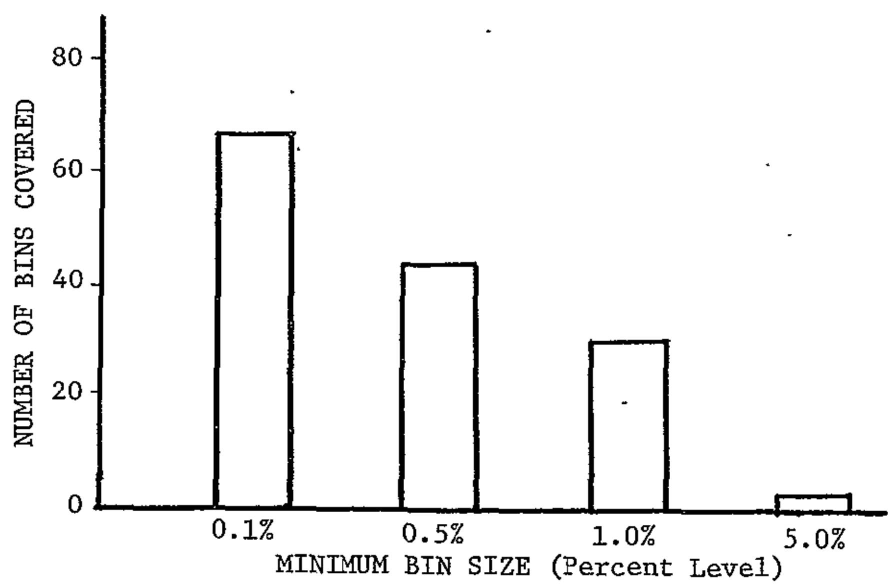

FIGURE 4. NUMBER OF BINS VS. MINIMUM BIN SIZE

The 1% level was found to be the most optimum group to work with when using two dates, containing 83% of the data in approximately 31 bins.

Several observations were made when comparing the training histograms to the segment histograms on a bin by bin basis for 13 Kansas segments.

In general if the bin contained n% of the segment data then it contained (n $\pm$ 2.5)% of the AI designated training data. Cases where this was not true usually involved the larger bins containing greater than 7% of the data. In these cases if the bin contained n% of the segment then it might contain (n $\pm$ n/2)% of the training. Thus larger bins were generally represented by AI designations. However bins containing less than 2.5% of the total data may be completely missed by AI training. This introduces a non-random character to the training data. This type of missed training was found in 7 of the 13 test sites. There was an average of 2.5 bins per segment not found in the AI designated training sets, with as many as 11 bins not represented by training in some segments.

Using the histogram maps (Figure 5) and ground truth photos new fields were determined to complete the training set. On the example histogram map one can see definite field structure. The blank areas symbolize data in bins with less than 1% of the data. These areas are usually field boundaries and represent a mixture of vegetation types. The field-like structure of the histogram indicates that important bins that were not sampled by the AI are actually fields. Hence a better simulation of random training selection could be achieved by augmenting the Fields Data Base with fields representing important bins that were not represented by the AI fields. This was carried out for all of the segments in the test data base. Overall there were 23 new polygons added to the first 13 training segments examined, with as many as nine added to a single site.

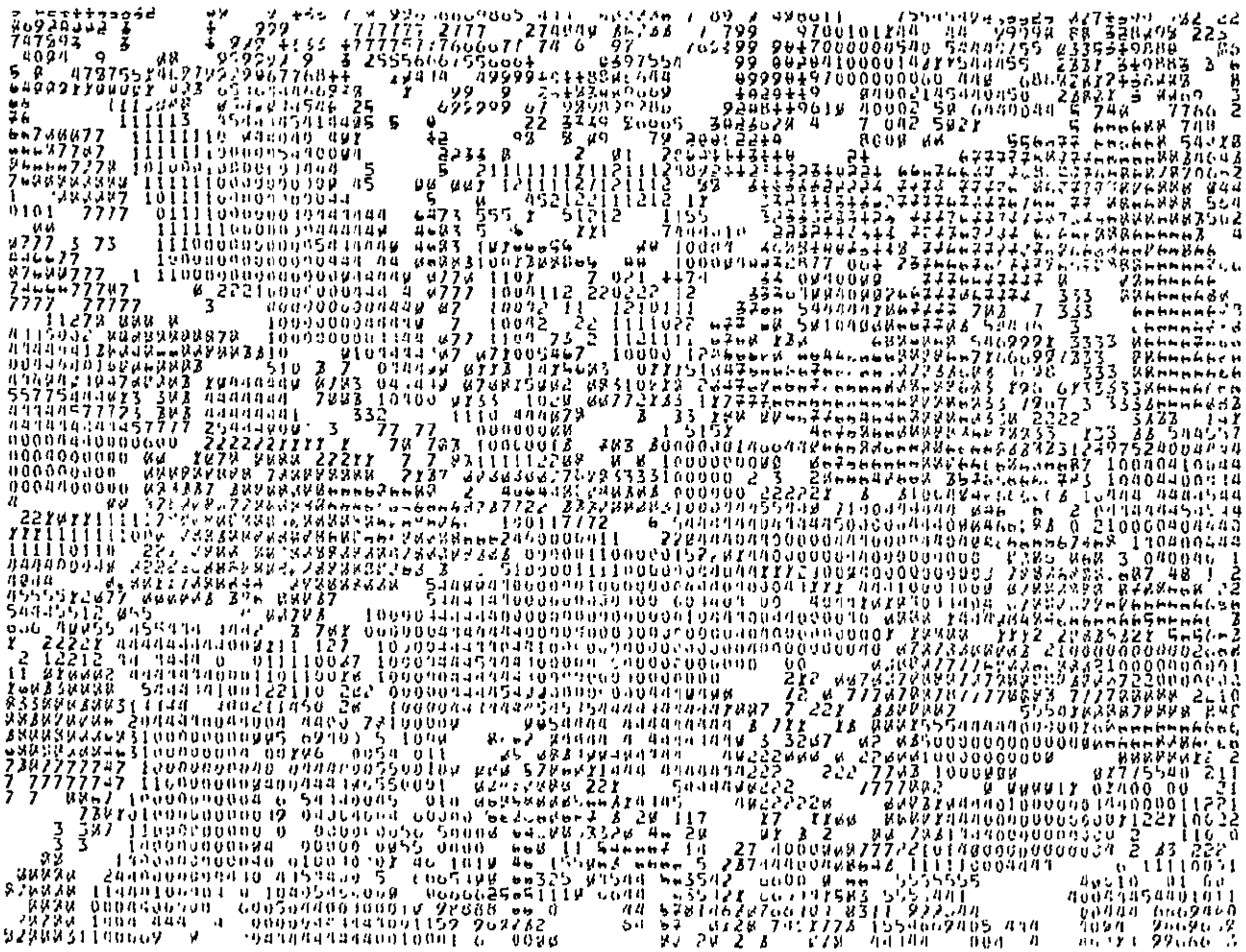

FIGURE 5. CHECK HISTOGRAM MAP OF PORTION OF SEGMENT 1154

4.3.3 FURTHER ANALYSIS USING CHECK

CHECK provides a framework within which any training selection procedure may be examined to establish bias or non-random characteristics. It can also be used to examine the characteristics of the data as a function of the temporal dimensionality of the data. It is well known statistically that an increase in the dimensionality of the data provides not only the potential for more information, but also the need for more, or at least more accurate, training sample selection in order to describe the information content of various classes of data.

CHECK was used to examine the effectiveness of two training procedures as a function of additional multitemporal data acquisitions. The two procedures include the AI training field designation, and sampling based on the selection of every tenth pixel in every tenth line of data. (The second procedure is not exactly equivalent to the training procedure employed in the LACIE Phase II Procedure I system.) The purpose of this exercise was to establish how a fixed sampling of data behaves as new information is added.

The CHECK procedure was carried out for data sets containing two, then three and four multitemporal data acquisitions. The data was histogrammed into three levels in the Tasselled Cap brightness and green channels for each set of acquisitions. For two biophases, there were a possible 81 bins of data (3^4 or three levels for each of four channels of data). Three biophases provided a potential for 3^6 or 729 bins, and four biophases a potential for 6561 bins. Histograms were examined for bins containing 0.1, 0.5, 1.0 and 5.0 percent of the total number of pixels per segment. The 0.1% level was the only level wherein 80% or more of the data in each segment was represented for each set of acquisitions.

A number of observations can be made in examining these histograms. Comparing the 209 point histograms to the segment histograms on a bin by bin basis for two biophases one finds a closer-to-randomly-selected training set than represented by the AI selected fields. If a bin contained n% of the segment data then the bin contained about n ± 1.5% of the 209 point training set, regardless of bin size. This is no surprise since the 209 points were selected by arbitrarily superimposing a grid on the data set.

Upon extending the CHECK procedure to three and four acquisition dates, employing these fixed sampling criteria leads to interesting results. Figure 6 illustrates three methods: wall-to-wall ground truth represented by the total number of bins, AI labeling, and use of the 209 point grid. The 0.1% curves are presented since this covers the majority of data points in all three acquisition cases. Notice that as the number of time periods increases, increasing the dimensionality of the data, the amount of training required also increases.

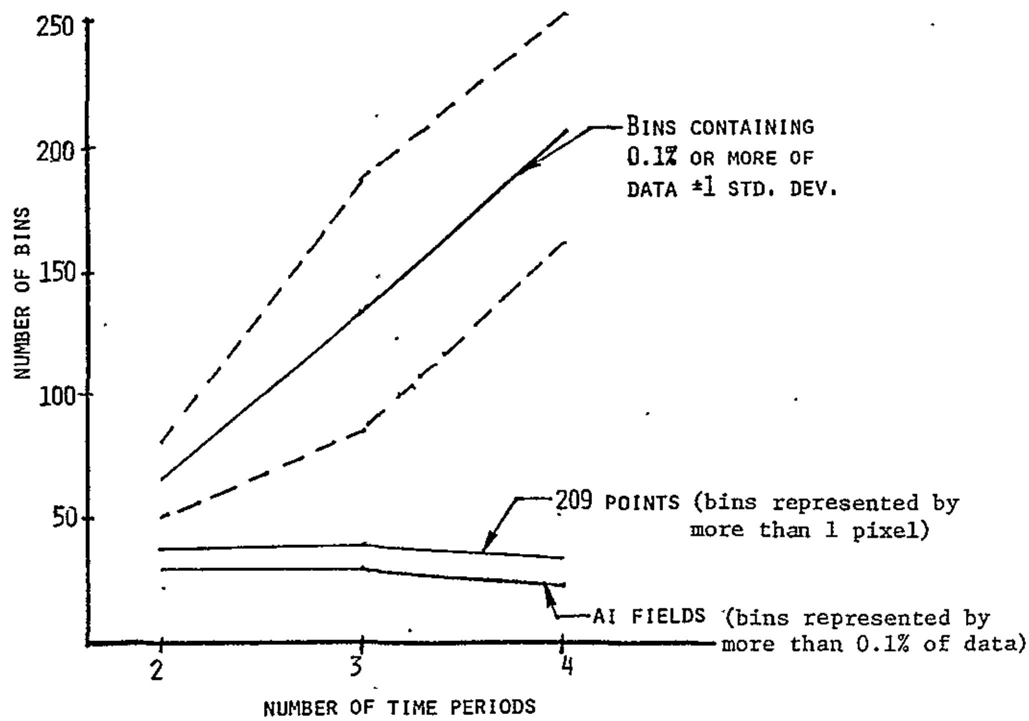

FIGURE 6. NUMBER OF BINS CONTAINING 0.1% OR MORE OF DATA COVERED BY TRAINING AS DETERMINED THROUGH CHECK

The 209 point or AI labeling method does not adequately represent the data in and of themselves. Additional information must be provided. This is precisely what the LACIE Procedure 1 training selection approach attempts to address by augmenting the training selection based on 209 points with an associated clustering algorithm. The alternative approach would be to sample employing wall-to-wall ground truth. Whereas wall-to-wall ground truth may not be a feasible approach, we plan to investigate the use of 209 points in cluster labeling as in Procedure 1 as well as a field seeking algorithm like BLOB in conjunction with the CHECK procedure as a technique to determine representative training fields.

4.4 ANALYST INTERPRETER LABELING ERROR ANALYSIS

Section 4.3 described activity that related to correcting Analyst Interpreter labeling errors in a number of 1975-76 LACIE Blind Site segments in Kansas and North Dakota that currently comprise the test data base described in Appendix I. This was accomplished by comparing the crop labels of AI designated fields to ground truth annotated high altitude photography. An analysis of the nature of these labeling errors was of interest for several reasons.

The Analyst Interpreter functions in a multisegment/multitemporal environment. The labeling of wheat and non-wheat is carried on a segment at a time, utilizing several false color Landsat images representing various biophases in the wheat crop calendar. The AI currently is provided with false color imagery generated by a Production Film Converter employing a specific color coding technique [13]. These images are termed Product 1's. In addition to these images, other aids are provided to assist the AI in understanding the local scene characteristics that may affect the apparent colors of wheat and non-wheat. However, multisegment signature extension is carried out by the AI each and every time the AI labels wheat or non-wheat using the non-segment specific, or global, information accumulated by experience.

The error analysis carried out attempts to quantitatively address questions pertaining to the influence of the technique employed in the generation of Product 1's upon the AI's ability to correctly label wheat and upon subsequent classifications based on inaccurate signatures derived from mislabelled samples. Two specific concerns warranted our analysis of the Product 1. First of all, the product is generated using segment specific, not global, parameters, and secondly, external effects, like haze and sun angle, are not accounted for.

4.4.1 APPROACH

The analysis of the nature of Analyst Interpreter labeling errors was carried out in six stages:

1. Comparison of AI designations with ground truth labels and measurement of error rates. Section 4.3 described the AI error found to be present in 46, 1975-76 LACIE blind sites and the error statistics generated for each segment.

2. A brief consideration of the effect AI labeling errors have on accuracy of proportion estimation. Described in Section 4.4.2 below.

3. A search for correlation between extent of labeling error and various segment specific ancillary variables. Described in Section 4.4.3 below.

4. Development of a data base with field means of Landsat data for three biophase acquisitions per segment and a technique for display of the data in color space. Described in Section 4.4.4 below.

5. Diagnostic work relating color error with various acquisition and segment specific variables. Intended approach shown in Section 4.4.5.

6. Exploration of possible improvements in generation of false color imagery. Plans are indicated in Section 4.4.6.

Stages one through four have been completed at the time of this writing. Work on stages five and six is in progress.

4.4.2 EFFECT OF LABELING ERRORS ON PROPORTION ESTIMATION

Our consideration of the influence on proportion estimation of mislabeled training fields was not intended to be definitive. We wished to obtain a general idea, based on the data already at our disposal, of the variance in proportion estimation which might be attributed to mislabeling. As one indicator we considered segments with missed wheat error but no false wheat error. We plotted missed wheat error versus the fraction of wheat in scene that was detected, i.e., the ratio of the proportion estimate in local classification mode, to the ground truth proportion of wheat in the scene. Figure 7 reveals a tendency for detected proportion of wheat to fall off quickly with missed wheat error. It suggests that for error greater than 24% about 60% wheat detection may be expected. The missed wheat error statistic is only a crude measure of the amount of misinformation given to the classifier, which probably accounts for much of the scatter in Figure 7. Even so the missed wheat variable accounts for about 40% of variance in the detected proportion of wheat.

4.4.3 CORRELATION OF LABELING ERRORS WITH ANCILLARY VARIABLES

Analyst-Interpreter accuracy measures were regressed against the following set of segment specific variables:

1. Ground truth percentage of wheat in the segment.
2. Long term average for growing season of Degree-Day sum.
3. Long term average for growing season of Precipitation
4. Elevation.
5. Latitude
6. Longitude

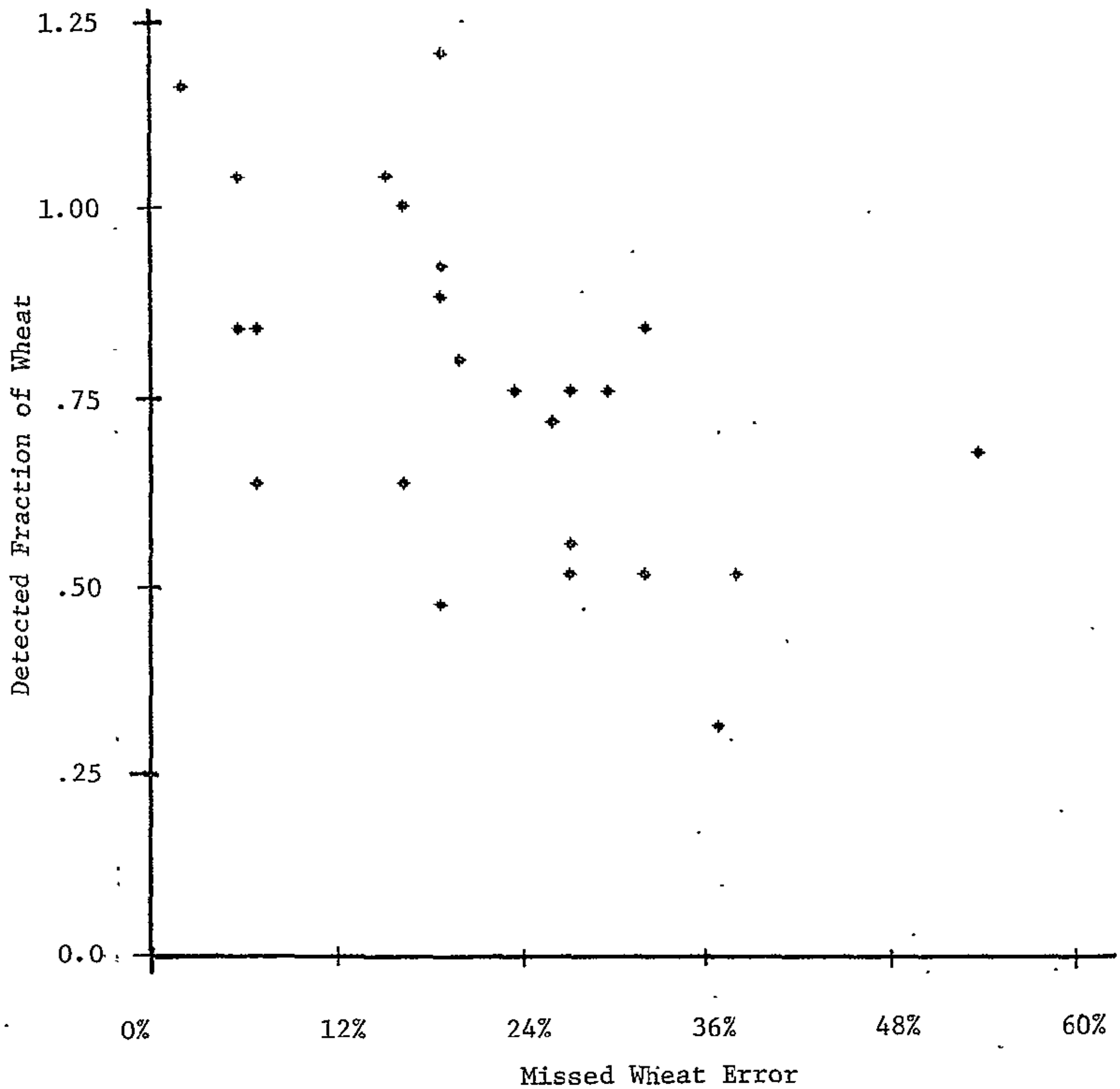

FIGURE 7. FRACTION OF WHEAT IN SCENE DETECTED VERSUS MISSED WHEAT ERROR

In a somewhat unexpected result, we found AI accuracy not to be correlated with percentage of wheat in a segment. Figure 8 demonstrates the independence of missed wheat error from percentage of wheat in a segment. A study conducted by Coberly, Tubbs and Odell [13], indicated the Product 1 might be susceptible to color distortion in scenes with very little wheat and scenes dominated by wheat. This concern stemmed from the fact that bias and scale values used in generating the Product 1 are computed on the basis of variability in the contents of a scene. Logically, the amount of wheat in a scene is an important factor in how homogeneous the scene will appear. In the study cited, wheat and non-wheat signatures were used to generate artificial scene statistics, assuming different proportions of wheat, and these statistics were used to compute corresponding bias and scale values. These values indicated color distortion in scenes with little wheat (a lot of variability) and scenes largely composed of wheat (little variability). The fact that AI error rates are not a function of proportion of wheat in a scene makes us suspect that the study cited was too simplistic in its assumptions. Proportion of wheat in a scene may be one factor in color error but in real life it is one among many. The conclusion of the study, that Product 1 is susceptible to distortion, is still valid. However, the range of factors involved and the significance of color shifts produced, have yet to be explored.

The other variables tested also proved uncorrelated with the single exception of latitude. Latitude was found correlated to AI total error with $r = -.60$ at a significance level below 0.001. As Figure 9 shows this is not a tight correlation but it appears to be real.

We interpret this to mean there exists a factor which

1. Characteristically varies with geographic latitude of a segment and
2. Is capable of influencing AI accuracy in a fairly strong manner.

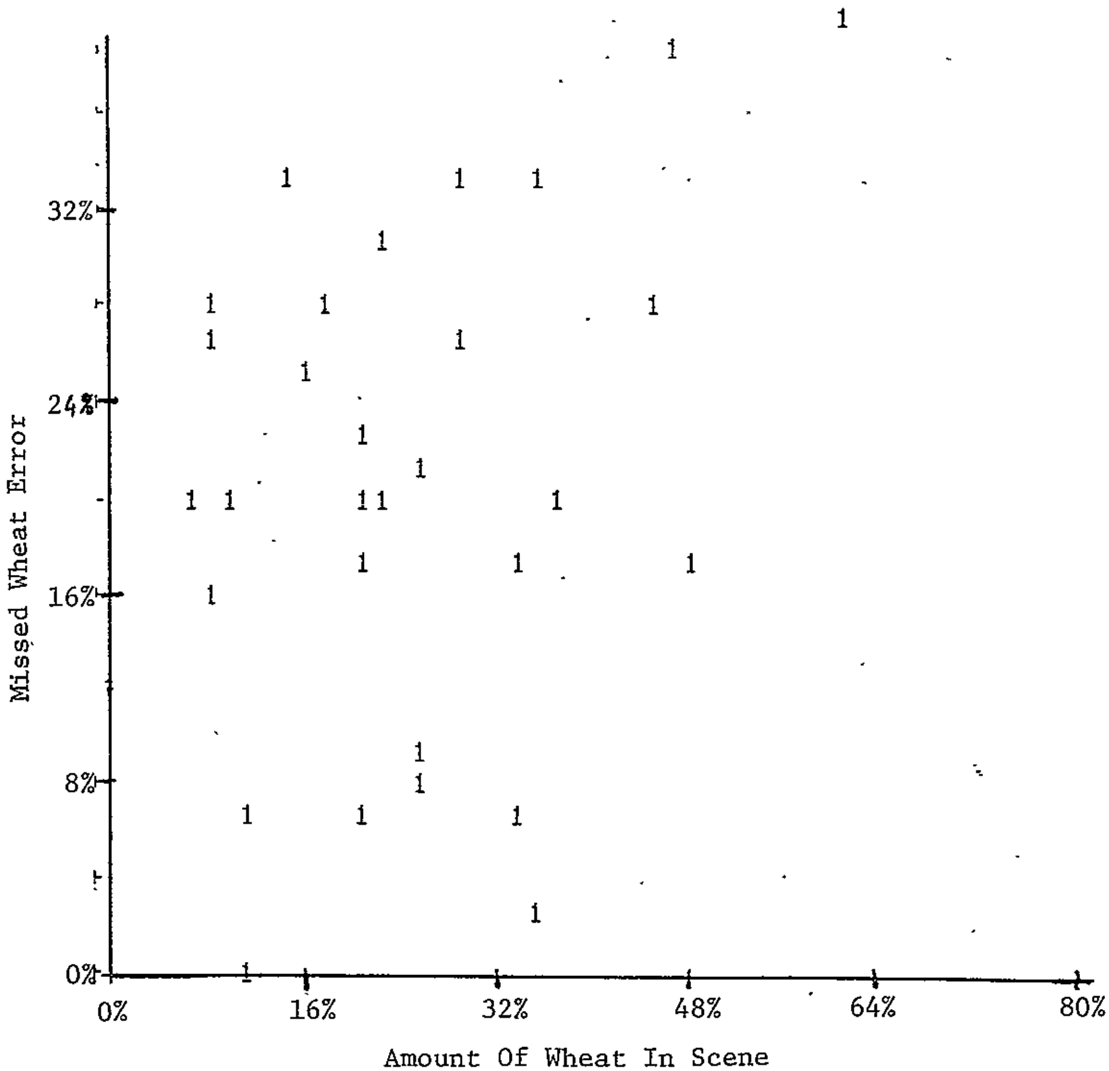

FIGURE 8. MISSED WHEAT ERROR VERSUS PERCENTAGE OF WHEAT IN SCENE

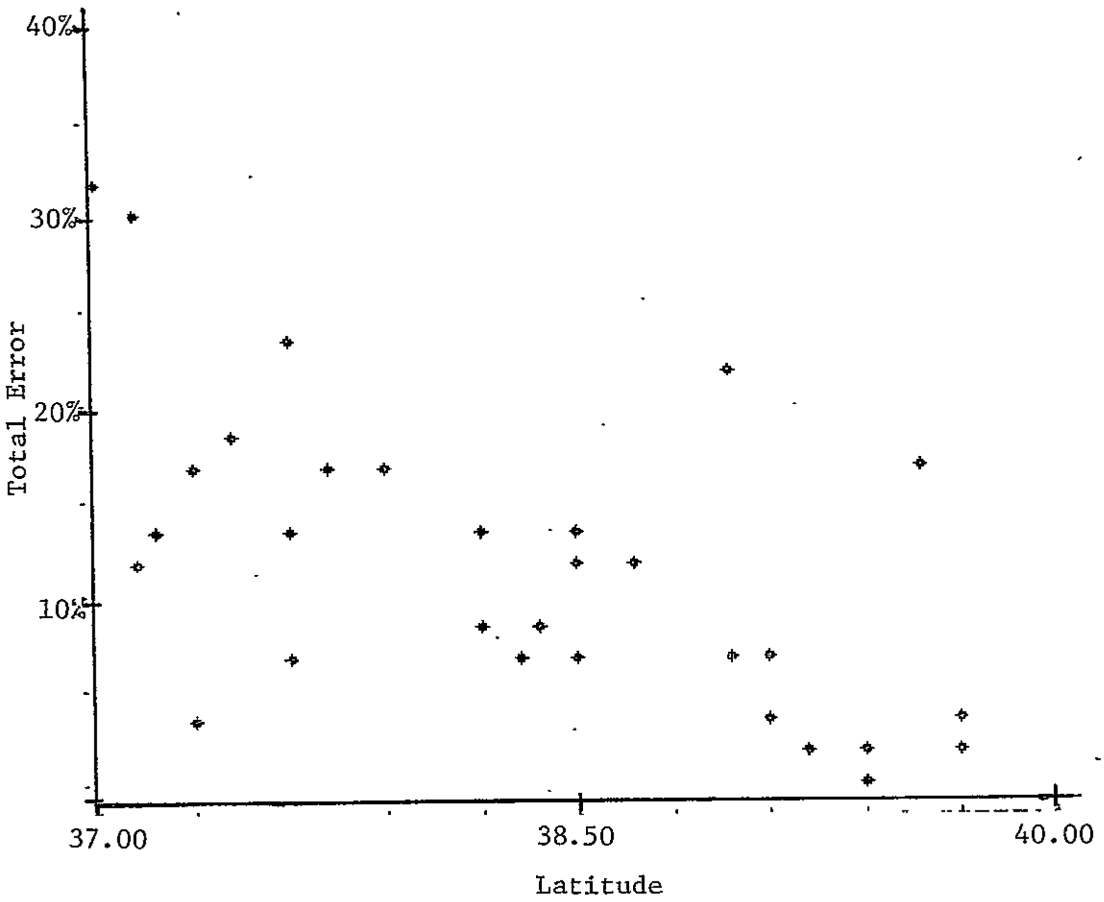

FIGURE 9. SEGMENT LATITUDE VERSUS AI TOTAL ERROR FIGURE

The adjusted crop calendar provided for the AI is a critical factor in the labeling process. The crop calendar also varies characteristically with latitude because of climatic changes. Our first suspicion in this matter is, therefore, that unrecognized inaccuracies in crop calendar adjustment procedure exist which are tied to latitude.

4.4.4 DISPLAY OF DATA IN COLOR SPACE

Effort was directed toward obtaining a display of field mean data in color space, i.e., a chromaticity diagram. The idea of this is to have a graphical portrayal of the distribution of colors of fields as they appear on the Product 1 false-color imagery. Distance in color space is an indicator of distinguishability between colors to the human eye. It was felt a display of the fields in color space would be a direct, insightful tool for addressing the labeling problem. Implementing the technique required three steps.

1. A data base was established containing the following information for each segment (see Appendix V).
   a. The mean value in each of Landsat bands four through seven for each defined field in the scene.
   b. The ground truth designation of each field (wheat or non-wheat).
   c. The AI label for each defined field.
   d. The bias and scale factors used to transform the Landsat data before production of the Product 1 imagery.
2. For each acquisition in the data base an affine transformation was applied to the field mean data of the Landsat channels, exactly as if the data were being prepared for input to the blue, green and red color guns of the PFC, <u>viz</u>:

$$B = A_1X_1 + B_1$$

$$G = A_2X_2 + B_2$$

$$R = A_4X_4 + B_4$$

Here A_i and B_i $(i = 1,2,4)$ are the scale and bias factors for an acquisition as computed by current procedures [13]. After transformation any values of R, G, or B falling outside the range 0-256 (the color intensity range of the PFC) are terminated to the appropriate end point.

This data can be displayed on a two-dimensional chromaticity diagram after a normalization of the variables:

$$r = R/T$$
$$g = G/T$$
$$T = R+G+B \quad \text{(see Figure 11)}$$

It is not necessary to plot b (b = B/T) because of the restraint r+g+b = 1.

3. The (T, r, g) color space is not uniform because one cannot say there is a unique relationship, valid everywhere on the (r,g) graph, between distance and distinguishability of colors. There are transformations with which one can approximate a uniform color scale (UCS). The CIE 1960 UCS diagram is an example. It is defined as a projective transform of the CIE 1931 (x,y)-chromaticity diagram (Figure 12). To map our (r,g,b) space to the standard (x,y,z) chromaticity space the following relations were employed [15]:

$$x = \frac{0.49000r + 0.31000g + 0.20000b}{0.66697r + 1.13240g + 1.20063b},$$

$$y = \frac{0.17697r + 0.81240g + 0.01063b}{0.66697r + 1.13240g + 1.20063b},$$

$$z = \frac{0.00000r + 0.01000g + 0.99000b}{0.66697r + 1.13240g + 1.20063b}.$$

This transformation must be considered approximate in our case because the colors of the PFC are not exactly the standard (R,G,B) primaries. We proceeded on the belief this would allow an improvement in uniformity of the diagram if not optimum uniformity. The CIE UCS mapping is given by the following equations [14]:

$$u = \frac{4x}{-2x + 12y + 3}$$

$$v = \frac{6y}{-2x + 12y + 3}$$

Figure 10 displays the color ranges of the CIE X-Y chromaticity diagram. Figure 11 shows ellipses which represent statistical variation of chromaticity matches. The length of the axes of the ellipses represent the distance in color space required to make two colors just distinguishable to the eye. Observe that this distance is much smaller in the blue area of the (x,y) diagram than in the green. Obviously this space is not uniform. After transformation to (U,V) space (Figure 12) the ellipses are more or less comparable throughout the diagram, indicating improved uniformity. Figures 13 and 14 show a Biowindow 2 LACIE segment in (r,g) space and in (U,V) space.

4.4.5 FACTORS AFFECTING QUALITY OF THE PRODUCT 1

Our approach to investigating the labeling problem has two basic hypotheses behind it:

1. The current method of generating Product 1's introduces color errors which adversely affect the Analyst-Interpreters' ability to correctly label wheat and non-wheat in some instances.

2. An array of factors affect the quality of Product 1's and these factors must be recognized before the production of any standard Landsat film product.

Statement 1 refers to color error. We understand this term along the following lines. Three criteria of film quality are proposed by Toyo Kaneko [16]. These include color level resolution, brightness, and color distortion. The first two are closely related and important for training field selection and delineation. The color distortion criterion is important for training field labeling [17]. Color distortion is the

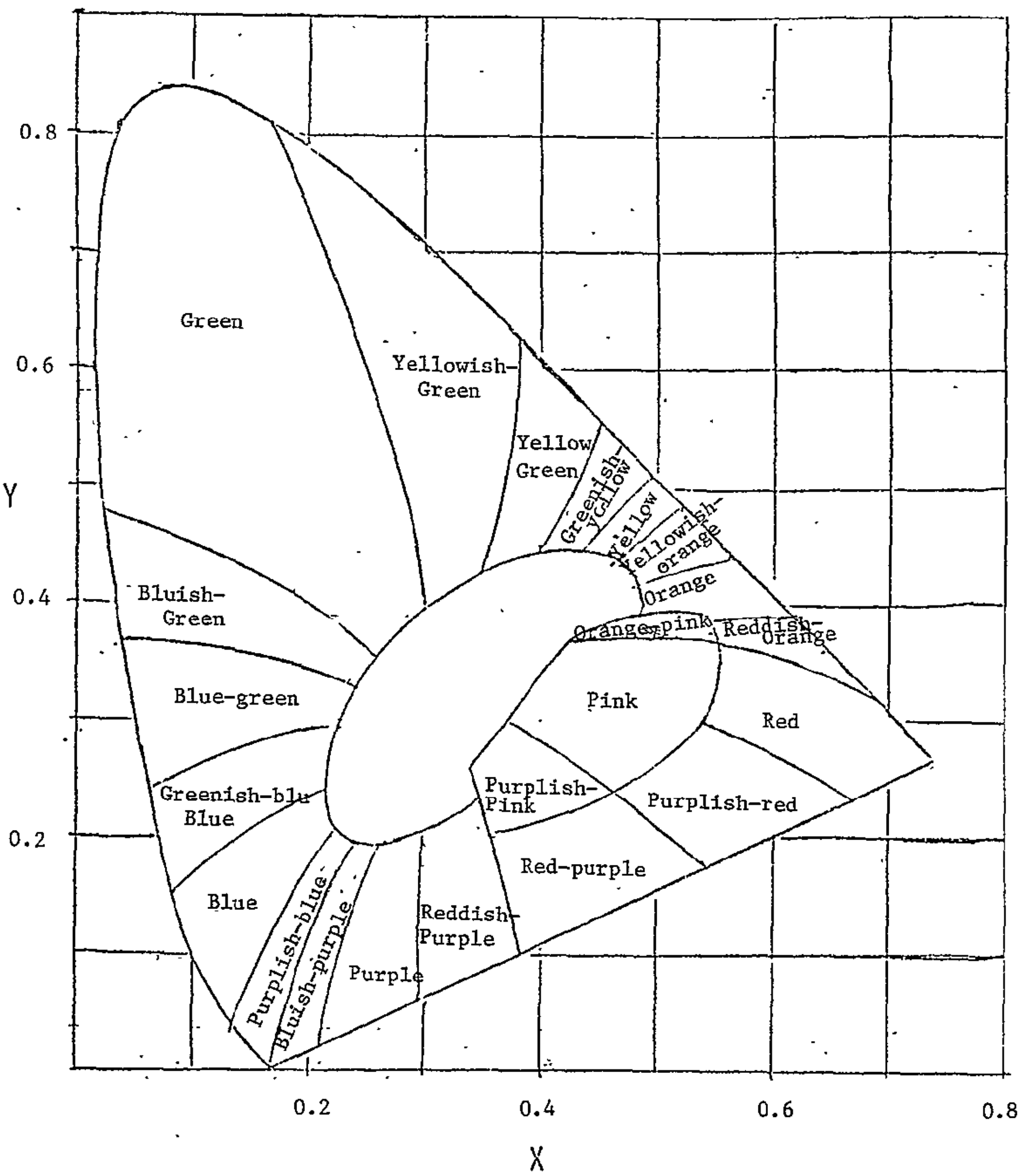

FIGURE 10: COLOR RANGES OF THE CIE X-Y CHROMATICITY DIAGRAM
(After Coberly)

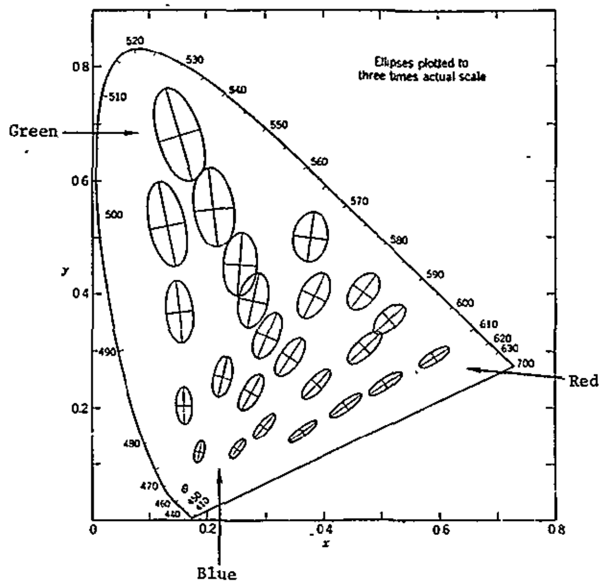

FIGURE 11. CIE 1931 (x,y)-CHROMATICITY DIAGRAM SHOWING STATISTICAL VARIATION OF CHROMATICITY MATCHES (After Stiles, 1946)

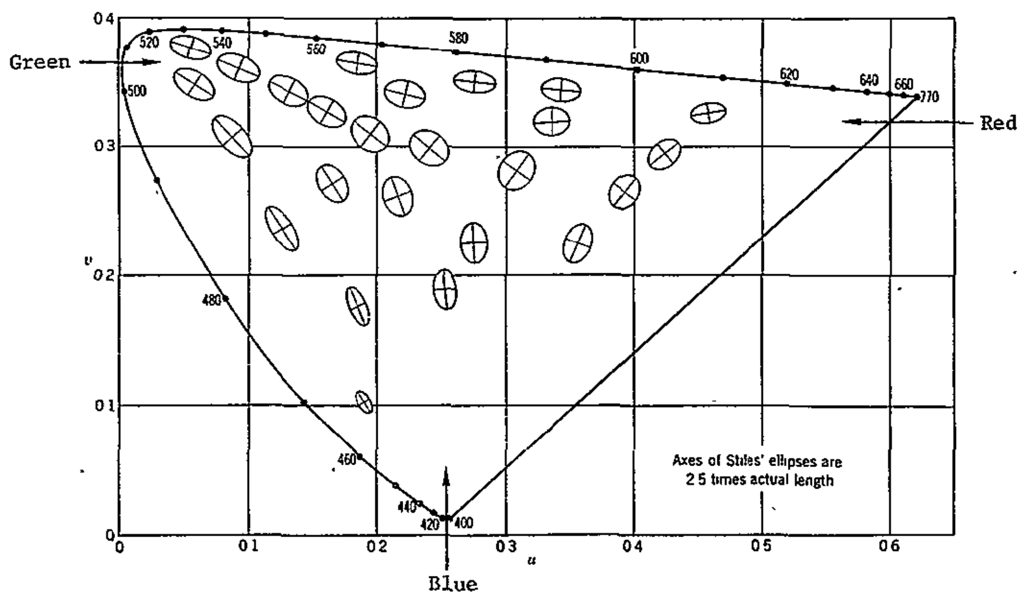

FIGURE 12. CIE 1960 UCS DIAGRAM WITH STILES ELLIPSES OF FIGURE 11

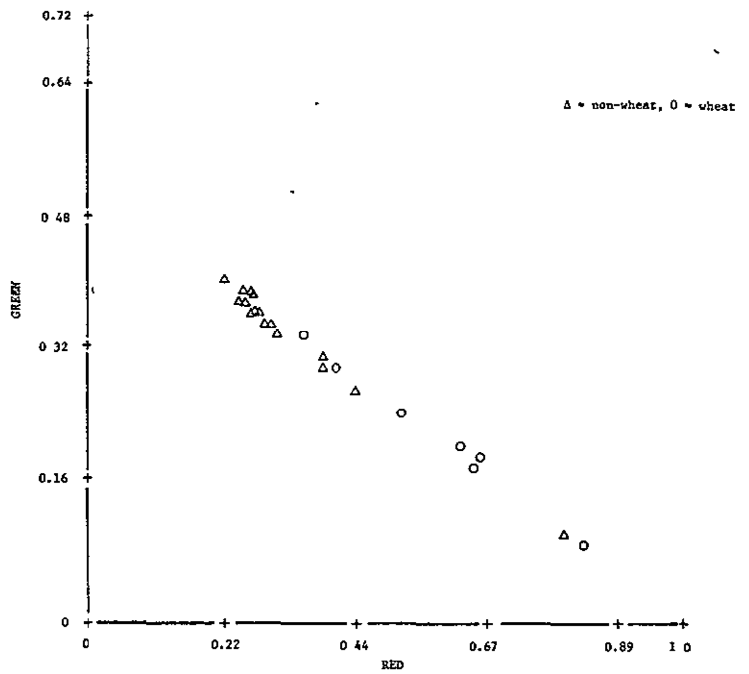

FIGURE 13. (r,g) CHROMATICITY PLOT OF FIELD MEANS FOR SEGMENT 1041, JULIAN DATE 127

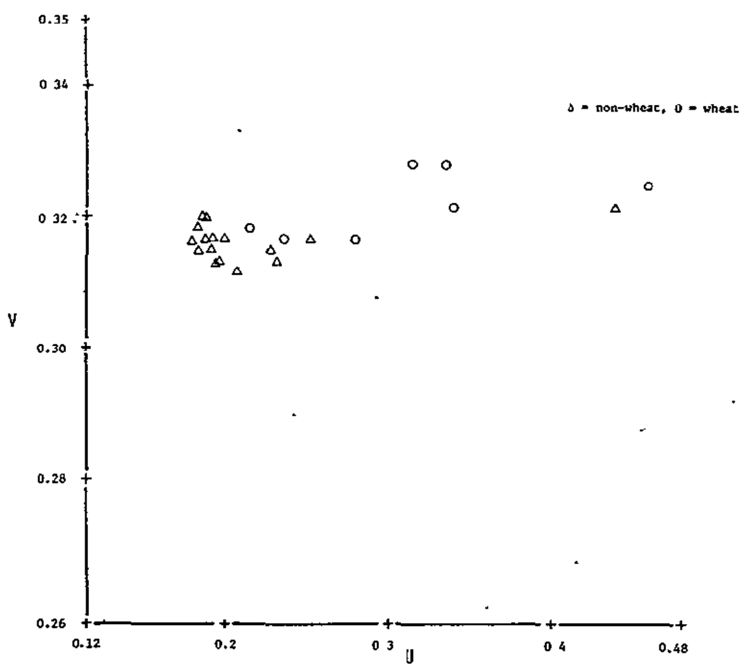

FIGURE 14. (U,V) CHROMATICITY DIAGRAM OF DATA SHOWN IN FIGURE 13

most important criterion in dot labeling. We conceive of color distortion as a change in hue, saturation, and brightness, i.e., colors of a given pixel, from time to time within a given segment. It should be thought of as a change in color from segment to segment of pixels with like reflectance. Color error is, therefore, defined for our purposes as a distortion of color from one segment or time period to another of two objects having the same reflectance. We are implying that the goal of any false color image display is to map objects of the same reflectance into the same color, regardless of place or time of acquisition, and make important differences between objects appear visible to the human eye.

To make our work more direct and quantitative we intend that color error be given analytic measures. For example one might consider the distance in (U,V) color space of the average color of wheat in a scene from some defined reference point as a measure of color error. With a measure of color distortion in hand we will be in a position to address the question of what factors cause color shifting in Product 1 imagery and determine their relative significance. Among the variables we will want to include in this analysis are the following:

a. haze level
b. sun angle
c. soil color
d. crop calendar
e. proportion of wheat in the scene
f. color composition of wheat and non-wheat
g. amount of clouds, water in scene.

Most of these variables are acquisition specific, i.e., are different for each Landsat pass over a particular segment. It is understood that the AI need not have been considering any particular acquisition in his work. We are not looking for correlations between acquisition specific variables and AI error rates; we endeavor to understand the Product 1 in ways which allow it to be generally improved. A reduction in labeling error may then be anticipated.

4.4.6 EXPLORATION OF POSSIBLE IMPROVEMENTS

The technique of display described in Section 4.4.4 gives us a special vantage point from which to explore ramifications of suggested improvements in production of false color imagery. Some suggestions will arise out of the diagnostic work described in Section 4.4.5. Other possibilities which will be evaluated include the following:

1. Correction of data for haze level and sun angle before production of imagery.
2. Use of a different technique for computing bias and scale factors:
   a. Hocutt method
   b. Kaneko method
   c. Krauss method
   d. New methods as our understanding suggests them.
3. Application of the Tasselled Cap transformation to the data prior to generation of imagery. The brightness, greenness and yellow dimensions of the data to be used as inputs to the green, red and blue guns of the PFC after scaling by one or another technique.

4.4.7 DISCUSSION

As a background to the discussion we present some (U,V) chromaticity diagrams of acquisitions available in our data set. In these figures wheat fields are designated by circles and non-wheat fields by triangles. A blackened-in circle or triangle indicates the AI mislabeled the field. Figures 15(a) and (b) show acquisitions of two segments in the second biophase. Figures 16(a) through (d) show biophases one and two for two segments. Figure 17 shows a complete 3 biophase history for one segment.

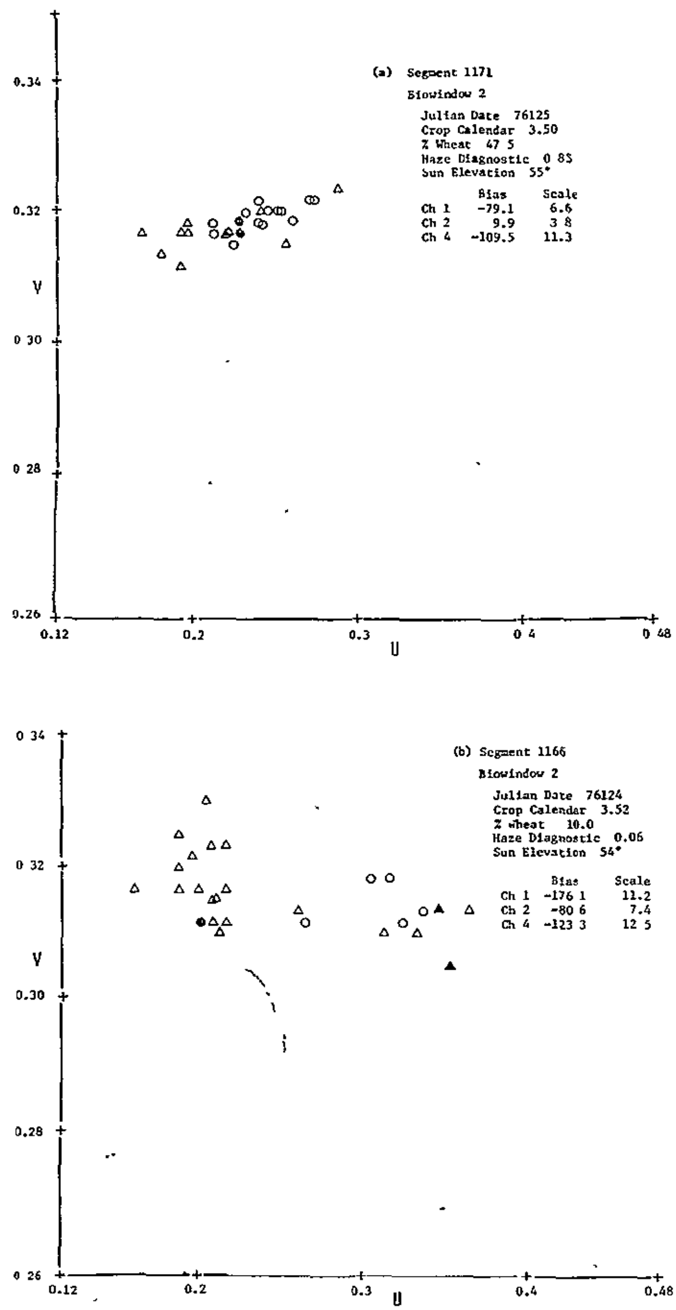

FIGURE 15. TWO SEGMENTS IN SECOND BIOWINDOW

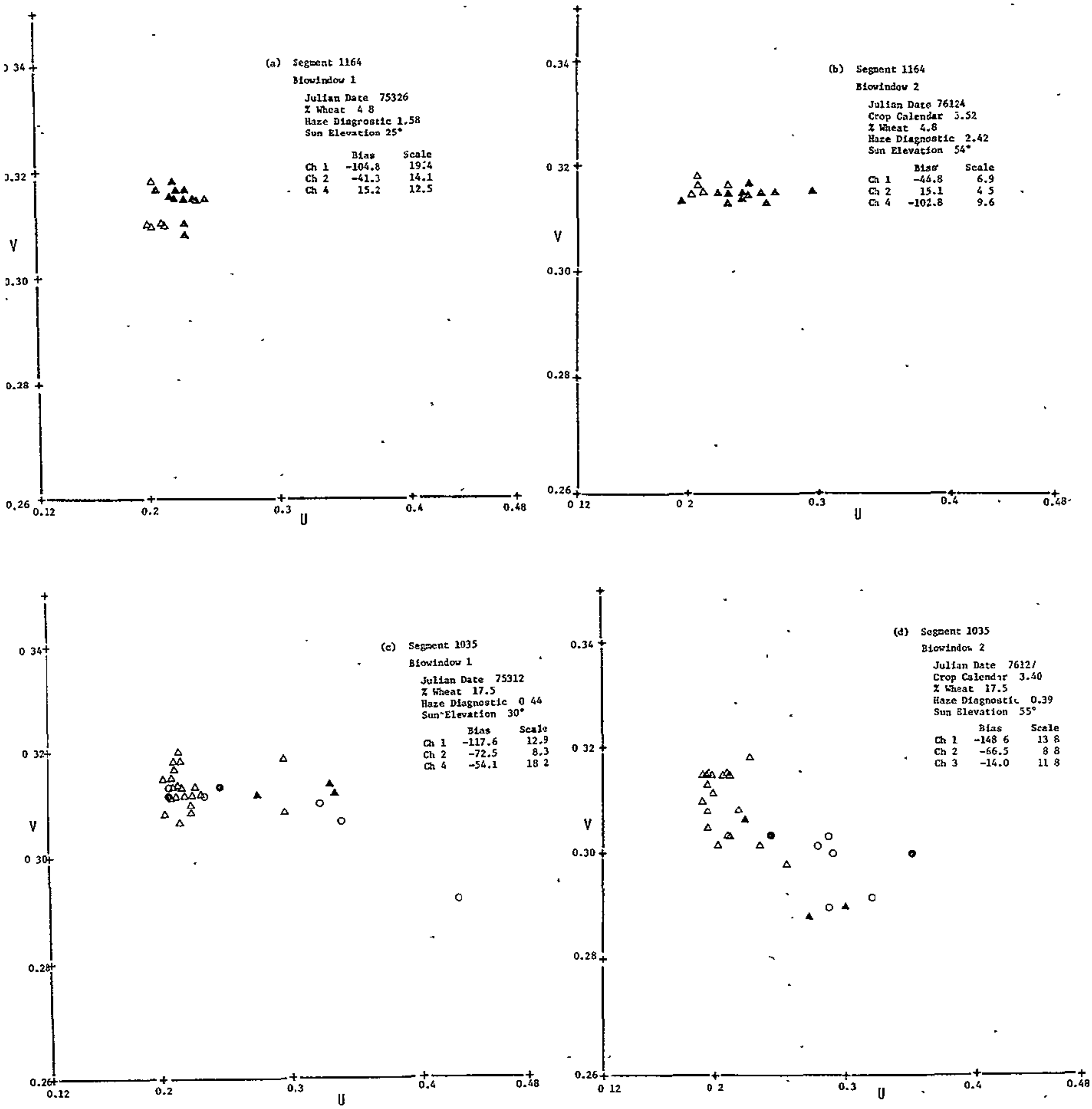

FIGURE 16. BIOWINDOWS 1 AND 2 FOR TWO SEGMENTS

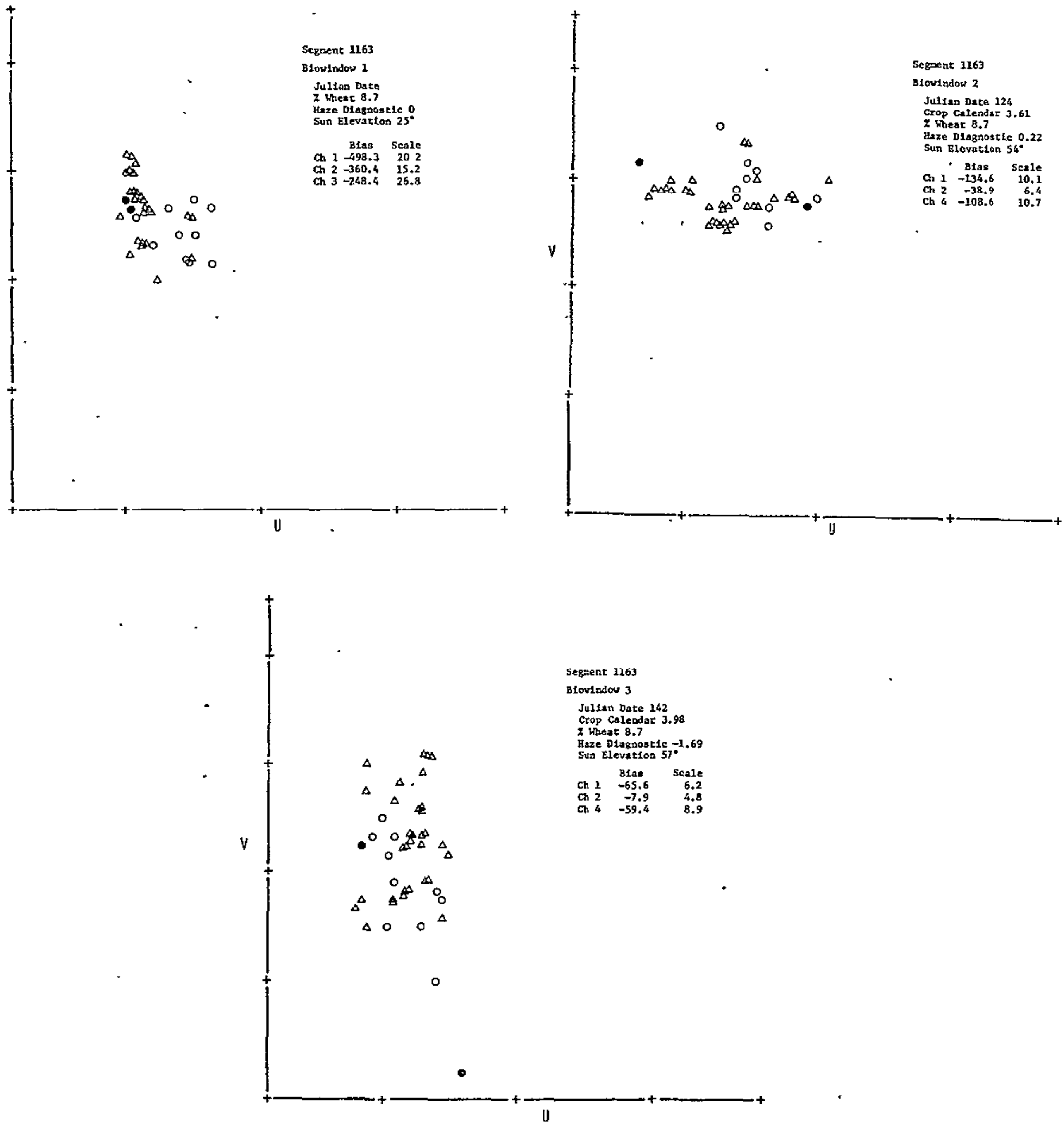

FIGURE 17. THREE BIOWINDOW ACQUISITION HISTORY FOR ONE SEGMENT

Looking at what we have thus far we can point to some disturbing things about Product 1 imagery. Figures 15(a), 15(b) and 16(b) show color space distribution of fields in three segments. Note how different the distribution of wheat color is between these segments, despite the fact the acquisitions were within one day of each other and the crop calendars are virtually identical. We have hypothesized this marked alteration in the wheat color signature from segment to segment is the result of using freely varying bias and scale values for scaling of data and not taking account of haze and illumination (sun angle) effects. The Analyst-Interpreter must interpret imagery using ancillary information, crop calendar estimates, historical agricultural statistics, and ground truth information. This is necessary to allow the AI to adjust the recognition of wheat to each segment and each acquisition. Because of the artificial variability of the Product 1 image, the presence of wheat and its approximate stage of development can never be addressed from the Product 1 image alone.

Consider the interpretation problem of Segment 1164. The color distribution of fields in this segment are shown in Figures 16(a) and 16(b) for acquisitions in biophases 1 and 2. Of the acquisitions made in 1975-76 on 1164, Julian date 124 stands out as the one to potentially distinguish wheat and non-wheat. There were no other acquisitions in the second or third biophases. This acquisition was at the same crop calendar point as the acquisitions of Figure 15. If one adopts the color signature of wheat displayed in Figure 15(a), (i.e., if one overlays the chromaticity diagram of 1164 on 1171) it appears 1164 contains mainly wheat. If one adopts the color signature of 1166, Figure 15(b), it appears 1164 contains little or no wheat.

The AI assigned the label of wheat to 70% of the fields in Segment 1164. In fact, there were no wheat fields among the fields defined on 1164.

Segment 1164 is not a special case of color distortion. It does not have extreme bias and scale values associated with it and could not be flagged by looking at these values. In this case the AI failed to find the proper boundary for interpreting the color of crops in this scene. This is an example of complete miscuing on the crop color signatures for a particular segment. This error is possible because of the artificial variability of the Product 1 which makes it necessary to tailor recognition of wheat to each segment and each acquisition on that segment. This raises for us the concern that even when this tailoring is basically successful the fit may be too unnecessarily tight or too loose. This lies in the realm of the individual AI's interpretation. It is a difficult tailoring task to perform on scant information about qualities of Product 1 imagery. We know the interpretation of false-color imagery can produce completely accurate labeling of fields on some segments. It is our conjecture that a portion of the 21% average missed wheat error and 11% average false wheat error are due to difficulties in interpretation introduced by color signature variability in Product 1 imagery.

A linear discriminant function was trained over all segments and three time periods, to see how well a universal wheat signature could be applied to individual segments. The result of applying the best linear universal discriminant to individual segments was essentially random classification. To illustrate the reason for this we have computed linear discriminant boundaries between wheat and non-wheat on a local, segment by segment basis, for 5 segments with virtually the same crop calendar at acquisition. Figure 18 shows how much these boundaries shift between segments.

The technique of labeling fields by interpretation of false color imagery with shifting color signatures requires two things: 1) substantial local information, ancillary data and ground truth comparison, and 2) self restraint on the part of the interpreter not to apply earlier

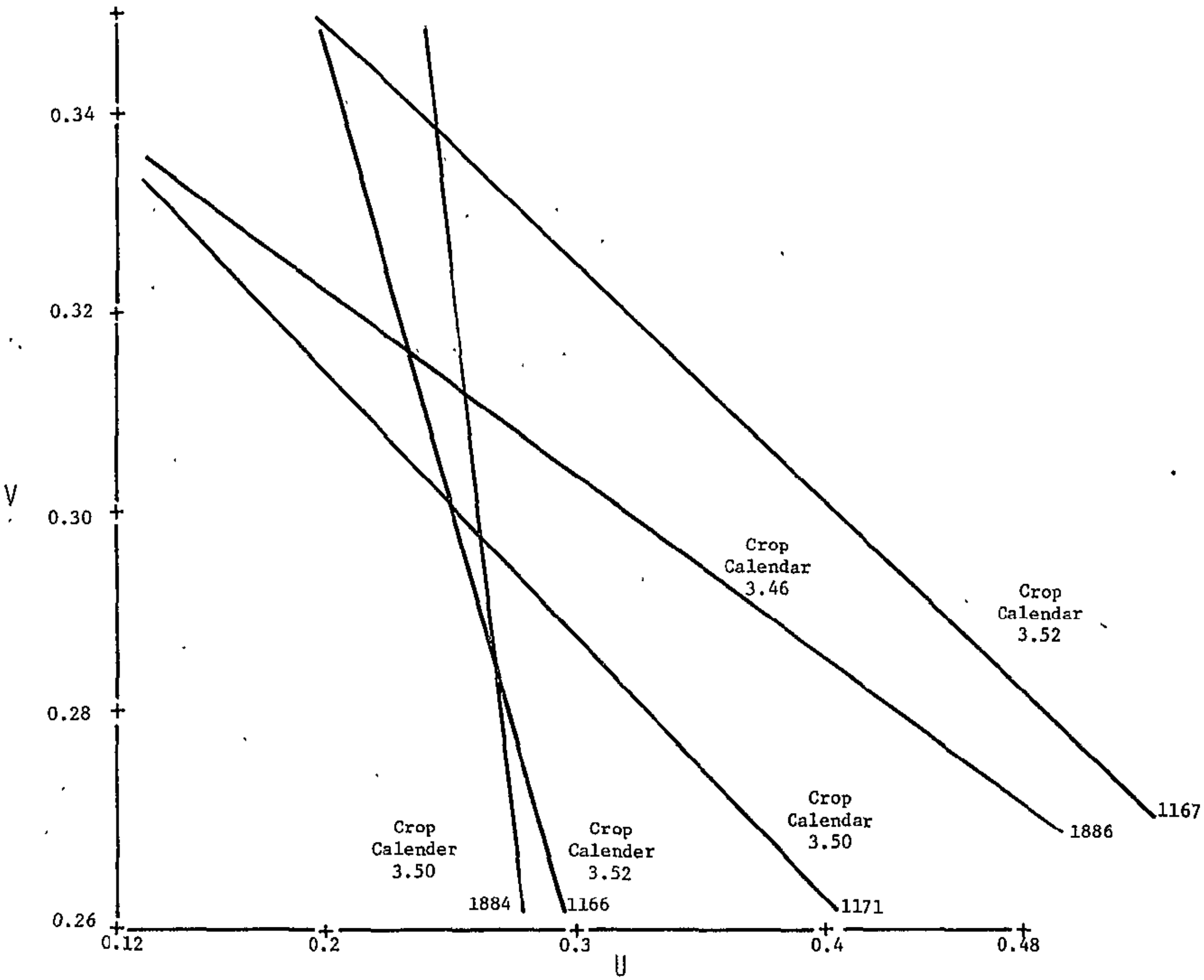

FIGURE 18. LINEAR DISCRIMINANT BOUNDARY BETWEEN WHEAT AND NON-WHEAT FOR FIVE SEGMENTS IN BIOPHASE TWO. (Crop Calendar Range: 3.46-3.52)

training where it might not be valid. We believe this necessity for restraint may contribute to inaccuracy anyway in the form of missed wheat errors.

As an alternative to the above, we propose investigation be directed toward establishing a way of producing imagery with stable discrimination boundaries. We believe the techniques discussed in this section provide the proper tools and we feel the explorations envisaged ought to be carried out.

In the data set we are currently working with we have field means data for 51 acquisitions. These acquisitions are spread among 32 1975-76 Kansas segments and three time periods. The segment numbers along with date, crop calendar, and error statistics are listed in Appendix V. We feel the extent of this data set is only marginal for the analyses we would like to perform. We would hope to have a new, larger set of acquisitions made available to us at a future point in time. This would allow us to be more definitive about qualitative conclusions and would make quantitative analysis feasible.

4.5 PHASE II: CONCLUSIONS AND RECOMMENDATIONS

Phase II of this project has concentrated on a twofold purpose: (1) the specification of an experiment design for the test and evaluation of overall signature extension procedures for large area crop inventory, and (2) an analysis of Analyst Interpreter wheat labeling errors.

Phase I documented that the development of accurate large area crop inventory systems using signature extension techniques is a feasible goal. The evaluation of three such techniques has been specified in the experiment design. These include a multisegment adaptation of Procedure 1, currently employed in LACIE as a local or single segment procedure, Procedure B, developed at ERIM, and a modified version of Procedure B, incorporating the training selection strategy of Procedure B and the classification strategy of Procedure 1.

In addition to the evaluation of these three overall procedures, a number of procedural parameters will be varied to determine the effect on classification results. These parameters include the number of segments used in training, and the incorporation of various data preprocessing techniques, specifically sun angle, haze effect corrections, and data compression strategies.

A most important aspect in the analysis of these multisegment signature extension techniques is their performance as a function of the use of static stratifications of the data. Three sets of stratifications will be employed including: (1) physical stratifications of the data based on ancillary variables as defined by UCB and JSC, (2) an arbitrary stratification wherein all segments are grouped into one stratum, and (3) a 'baseline' stratification wherein each segment is its own stratum, equivalently local or single segment training and classification.

Preparatory stages in the execution of the experiment to evaluate these overall multisegment signature extension procedures included the development of a data set for purposes of initial evaluation. This data set was drawn from the Fields Data Base. One step in its preparation includes the correction of Analyst-Interpreter labeling errors. The ensuing analysis of these labeling errors revealed that classification performance in a multisegment environment was sensitive to AI labeling errors.

In an attempt to understand the nature of these errors in order to provide recommendations as to improved labeling techniques, it was determined that the current procedure used in production of the Landsat Product 1 false color imagery has certain undesirable characteristics. Specifically, the color of wheat differed substantially from segment to segment at the same stages in the crop calendar.

It is recommended that the data base used in the analysis of AI errors be expanded to incorporate additional acquisitions for existing segments as well as additional segments in order to establish a data base that can be analyzed adequately to establish statistical significance. In addition, the technique employed in the analysis of the Product 1 imagery is a most useful approach to the analysis of other false color image products. That technique employs a mapping of field means into color space coordinates transformed into a space wherein Euclidean distance is more closely correlated to the human eye's ability to discriminate colors. Hence analysis of an AI's ability to discriminate wheat from non-wheat can be carried out statistically. A comparison of various image production techniques in this fashion would be of great value. It was also observed that the presence of haze or clouds in a scene may adversely affect image products. Techniques to reduce haze effects and screen clouds should be incorporated into the image production process.

APPENDIX I
DATA PREPARATION

The preparation of an adequate data base for the evaluation of signature extension algorithms was one of the major activities of this task. This activity had two separate phases. First, 1973-74 data was prepared to allow us to begin our first testing immediately. Later when 1975-76 LACIE sample segment data was received, together with the fields data base, activities were begun to prepare a large, comprehensive data base which included ancillary information about the sample segment and the specific passes in the data set.

Because the preparation of data was an ongoing activity, this appendix has been organized to reflect the state of the data base used for testing at the end of each of four periods covered by this report. Thus experiments conducted during the third quarter will refer to Section I.3 of this appendix for a complete description of their data.

I.1 FIRST PERIOD

The Landsat data used during the first period consists of ten 1973-74 LACIE sample segments over Kansas, mainly in the Southwest Crop Reporting District as shown in Figure I-1. Two of the sample segments are Intensive Study Sites (ITS) with wall-to-wall ground truth as determined by ground teams, and the remaining 8 sample segments are Statistical Reporting Service (SRS) sites with field labeling determined by NASA/JSC analysts based upon examination of the imagery itself. Imagery from several Landsat passes over each of these sites is available, and these images have been registered to each other. Table I-1 shows the sample segments, how the ground truth was obtained, and the dates of imagery collection used in the tests reported here.

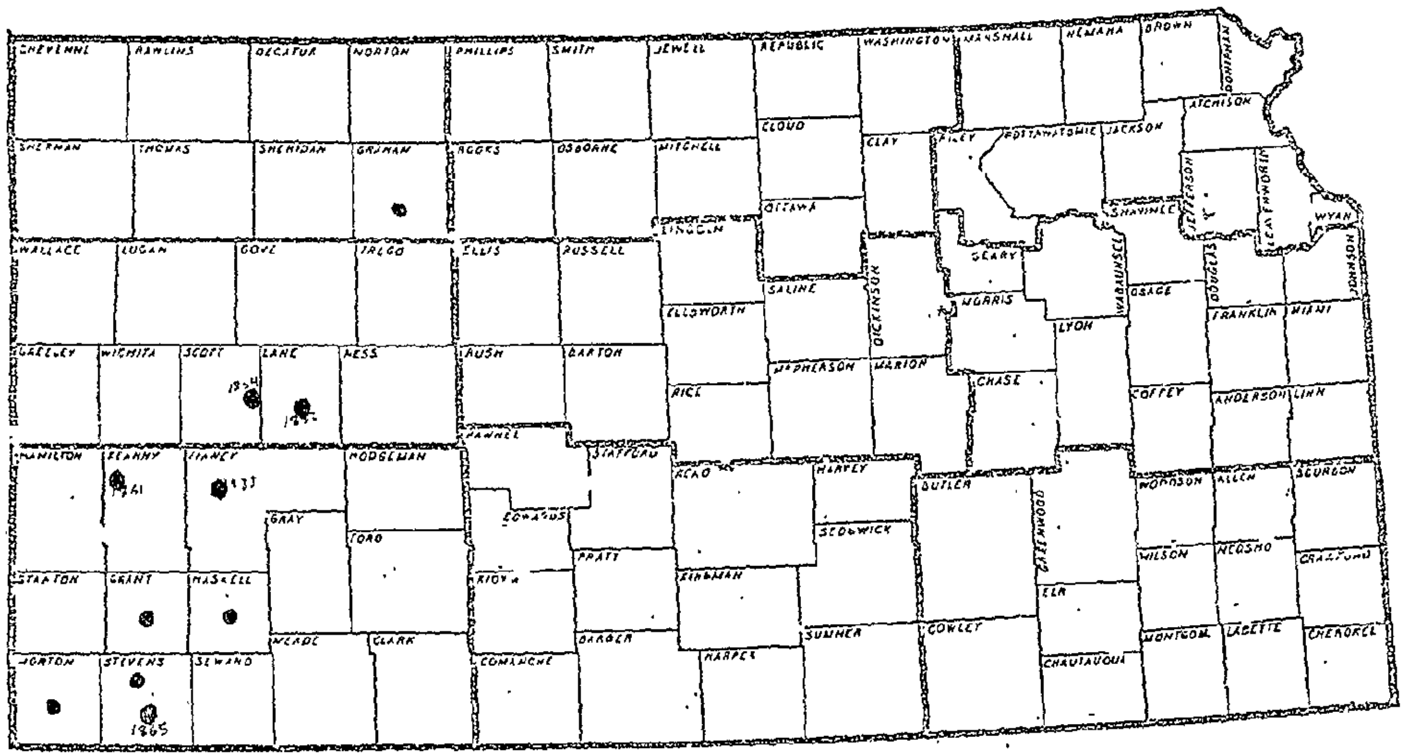

FIGURE I-1. TEST SITES IN KANSAS, 73-74 DATA

TABLE I-1. FIRST PERIOD DATA BASE

Site Name	Sample Segment No.	Ground Truth	Acquisition Dates Used
Morton	1042	ITS	5/8, 5/26
Finney	1034	ITS	5/8, 5/26
Graham	1018	SRS	5/8, 5/26
Lane	1026	SRS	5/8, 5/26
Scott	1029	SRS	5/8, 5/26
Grant	1036	SRS	5/9, 5/27
Kearny	1040	SRS	5/9, 5/27
Haskell	1065	SRS	5/9, 5/27
N. Stevens	1045	SRS	5/9, 5/27
S. Stevens	1045	SRS	5/9, 5/27

I.2 SECOND PERIOD

During the second period, 1973-74 multitemporal LACIE sample segments over 12 sites in Kansas were prepared. Figure I-2 shows their spatial distribution (two of the sites are in Stevens County). Four of these sample segments -- over Ellis, Saline, Morton, and Finney -- are Intensive Test Sites with wall-to-wall ground truth as determined by ground teams, while the remaining eight sample segments are SRS sites with field labeling determined by NASA/JSC analysts based upon examination of the imagery itself. Data from several Landsat passes over each of these sites is available, and has been registered to each other. Table I-2 shows the sample segments, and the dates of imagery collection used in the tests reported here.

ORIGINAL PAGE IS
OF POOR QUALITY

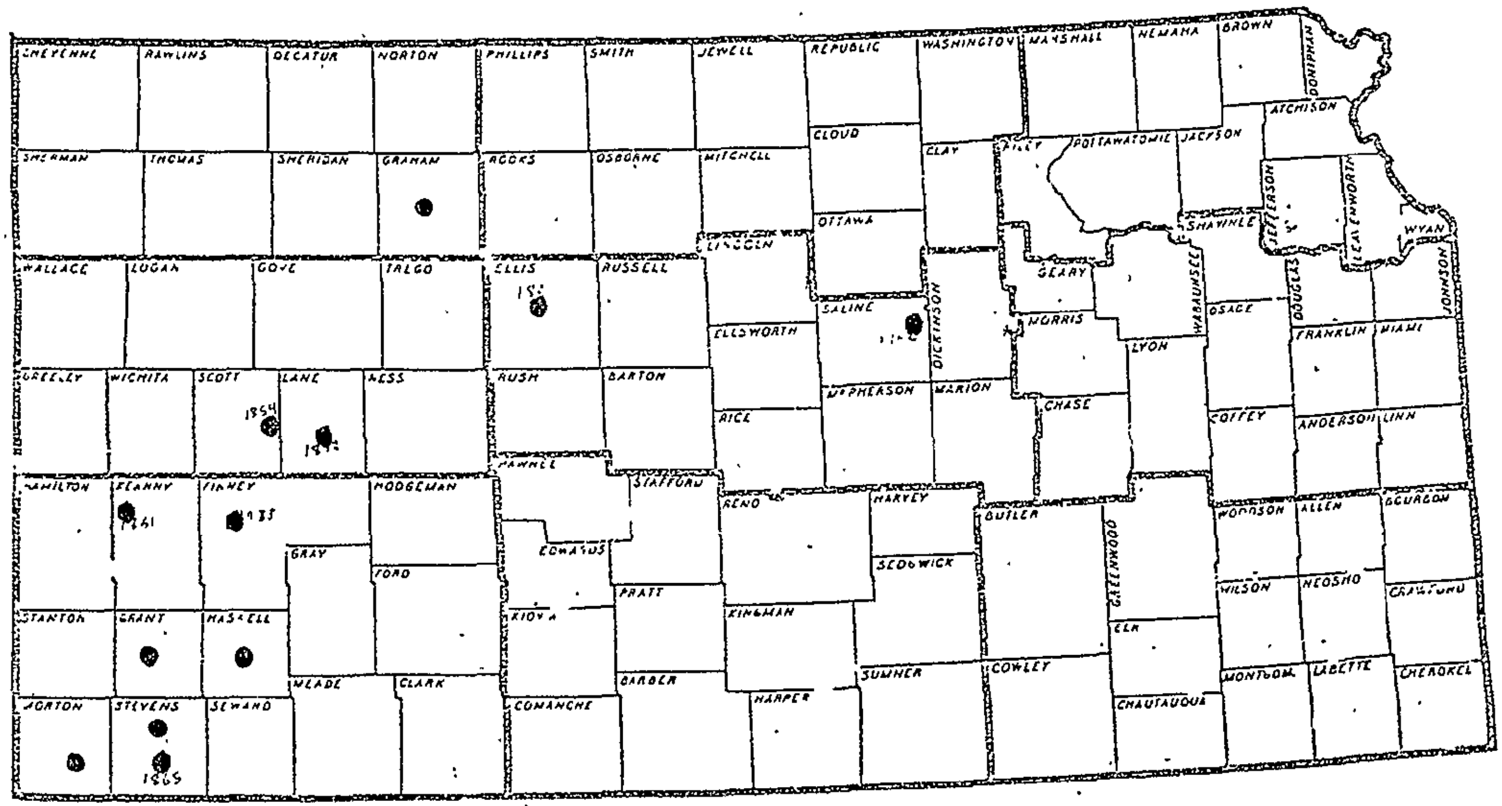

FIGURE I-2. TEST SITES IN KANSAS, 73-74 DATA

TABLE I-2. 1973-74 MULTITEMPORAL LACIE SAMPLE SEGMENTS

Site Name	Sample Segment No.	
Morton	1042	10/23/73, 5/9/74, 5/27/74, 6/7/74
Finney	1034	10/23/73, 4/20/74, 5/8/74, 5/26/74
Saline	1114	10/20/73, 4/18/74
Ellis	1106	10/21/73, 5/26/74, 6/12/74
Graham	1018	10/4/73, 4/20/74, 5/26/74
Lane	1026	10/4/73, 4/20/74, 5/26/74
Scott	1029	10/4/73, 4/20/74, 5/26/74
Grant	1036	10/23/73, 5/9/74, 5/27/74
Kearny	1040	10/23/73, 5/9/74, 5/27/74
Haskell	1065	10/23/73, 5/9/74, 5/27/74
N. Stevens	1045	10/23/73, 5/27/74, 6/14/74
S. Stevens	1045	10/23/73, 5/27/74, 6/14/74

I.3 THIRD PERIOD

After receipt in December 1976 of a large data set consisting of the 75-76 LACIE sample segments over the U.S., together with the Fields Data Base as of Day 315, the following data base was prepared.

The Landsat data used consisted of 75-76 Landsat data over 21 Blind Sites and two Intensive Test Sites (ITS) in Kansas. These 23 sites represented all of the Blind Sites and ITS sites in Kansas with cloud-free passes in early Biowindow one, and in Biowindow two. Only these two passes were used in any of the experiments described in this report, although a pass from each of the remaining biowindows was also prepared. These four passes were merged to form multitemporal data sets, and then screened to eliminate areas covered by cloud, cloud shadow or water in any of the four biowindows.

Signatures were computed for each of these 23 sites, and a data tape consisting of field means was also produced. The Fields Data Base as of Day 315 was used in these steps.

The final step in data preparation was to prepare a list of ancillary information for each of the sites. The types of ancillary information and the range of each ancillary variable appears below in Table I-3. Figure I-3 shows the distribution of these sites in Kansas.

I.4 FOURTH PERIOD

The fourth period data base consisted primarily of 74 data sets over 38 sample segments in Kansas (35 blind sites and 3 intensive test sites) and 18 data sets over 18 sample segments in North Dakota. Each of the data sets consists of four acquisitions of 75-76 LACIE sample segment data, one from each crop development biowindow whenever possible. Only the first two biowindows of the Kansas data and the first three biowindows of the North Dakota data were ever used. Along with the Landsat data is ancillary data pertaining to the sample segment, and to the various Landsat acquisitions used in the data set.

The fields data base as of Day 315 was used to provide the field designations which were used in lieu of ground truth in our evaluations. Tables I-4 and I-5 show the ranges of important ancillary variables for the winter wheat and spring wheat data, respectively. The ancillary variable called "crop calendar" is the Robertson crop calendar, and the variable "gamma" is the haze factor calculated by XSTAR [2]. The haze levels represented in these data sets span a fairly broad range.

2-D

TABLE I-3. ANCILLARY VARIABLES AND THEIR RANGE

Ancillary Variable	Range
GENERAL:	
Degree Days (10 Year Average)	2060 - 2470
Land Use (% Agriculture)	10% - 100%
Precipitation (10 Year Average)	7.2" - 12.9"
Latitude	37.1^{o} - 39.2^{o}
Longitude	94.9^{o} - 101.5^{o}
Elevation	900' - 3350'
PASS SPECIFIC (Calculated for Each Pass):	
Sun Angle	56^{o} - 67^{o}; 35^{o} - 46^{o}
View Angle	-5.5^{o} - 4.5^{o}; -6.0^{o} - 4.0^{o}
Julian Date	294 - 349; 87 - 127
Crop Calendar (Robertson Scale)	0 - 0; 2.76 - 3.66
CALCULATED FROM DATA:	
Difference Between Sites in Mean of Soils Area in Landsat Space	0 - 37.73; 0 - 48.65
Difference Between Sites in Mean of Green Development Area in Landsat Space	0 - 35.77; 0 - 60.72
Haze Diagnostic Calculated by XSTAR from Yellow Shift of Data	-1.36 - 0.86; -4.26 - 0.73
Difference Between Sites in Additive Factor Calculated by XSTAR	0 - 19.06; 0 - 17.04
Difference Between Sites in Multiplicative Factor Calculated by XSTAR	0 - 0.14; 0 - 0.42
Haze Value Calculated by XSTAR from Yellow Shift of Data	-0.06 - 0.03; -0.22 - 0.03

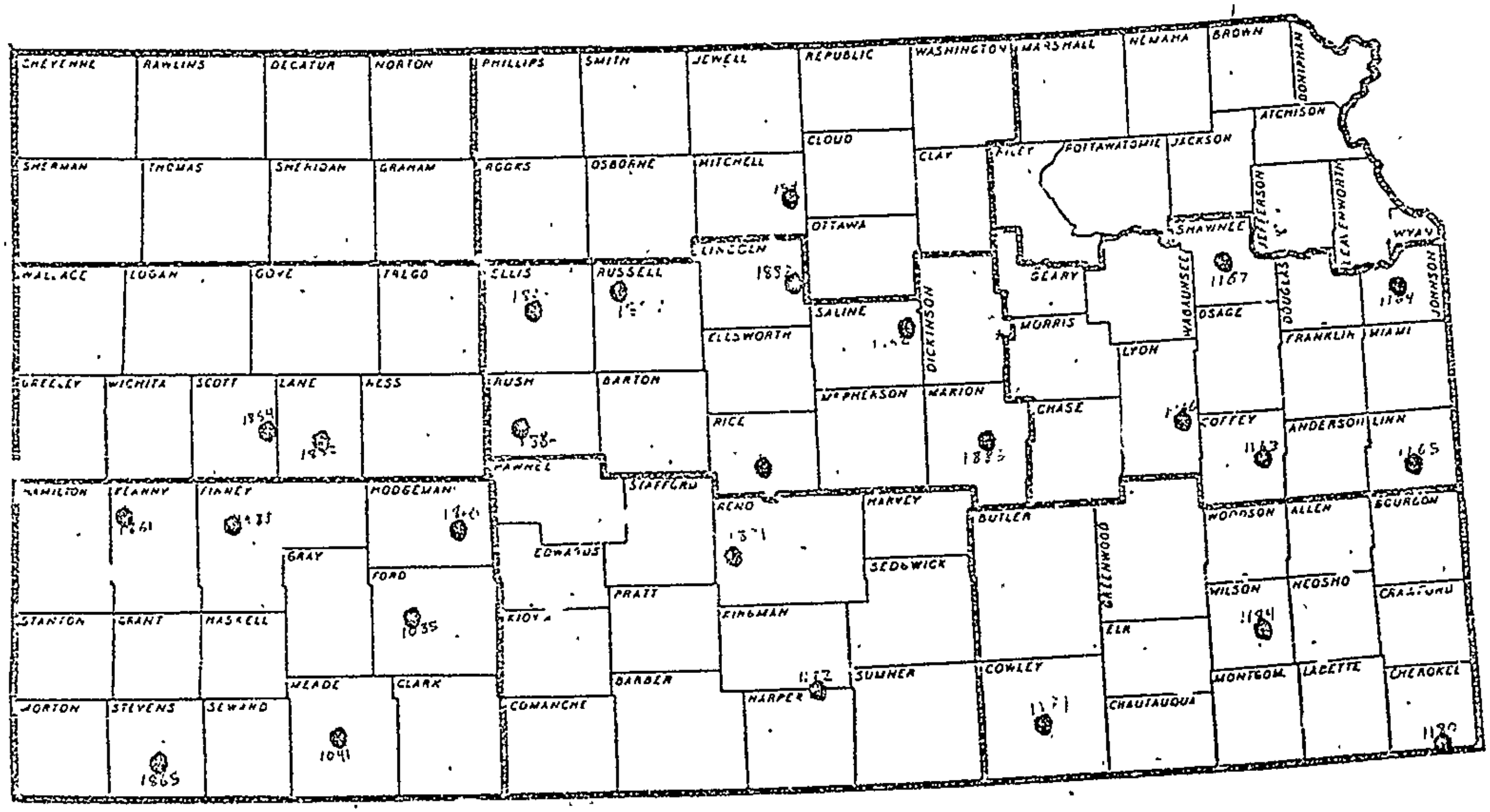

FIGURE I-3. TEST SITES IN KANSAS, 75-76 DATA

TABLE I-4. RANGE OF ANCILLARY DATA
Winter Wheat (Kansas) Data

DEGREE DAYS	1910 - 2525	ELEVATION	900′ - 3350′
PRECIPITATION (INCHES)	1 - 15	LATITUDE	37.0° - 39.7°
% AGRICULTURE	5 - 100	LONGITUDE	94.8° - 101.5°

BIOWINDOW 1

JULIAN DATE 291-90 CROP CALENDAR 0 - 3.3 SUN ANGLE 46° - 68° GAMMA -.08 - .23

BIOWINDOW 2

JULIAN DATE 90-138 CROP CALENDAR 3.0 - 3.6 SUN ANGLE 35° - 46° GAMMA -.5 - .19

BIOWINDOW 3

JULIAN DATE 135-163 CROP CALENDAR 3.3 - 4.8 SUN ANGLE 31° - 36° GAMMA -.22 - .19

BIOWINDOW 4

JULIAN DATE 163-200 CROP CALENDAR 4.5 - 6.0 SUN ANGLE 31° - 34° GAMMA -.25 - .17

TABLE I-5. RANGE OF ANCILLARY DATA

Spring Wheat (North Dakota) Data

DEGREE DAYS	2360 - 2520	ELEVATION	950′ - 2600′
PRECIPITATION (INCHES)	7.8 - 9.2	LATITUDE	46.2^0 - 48.8^0
% AGRICULTURE	5 - 100	LONGITUDE	96.7^0 - 103.8^0

<u>TIME PERIOD 1</u>			
JULIAN DATE 127-131	SUN ANGLE	33^0 - 39^0	GAMMA -.11 - .12
<u>TIME PERIOD 2</u>			
JULIAN DATE 144-150	SUN ANGLE	33^0 - 39^0	GAMMA -.5 - .1
<u>TIME PERIOD 3</u>			
JULIAN DATE 164-186	SUN ANGLE	33^0 - 39^0	GAMMA -.41 - .14
<u>TIME PERIOD 4</u>			
JULIAN DATE 198-204	SUN ANGLE	33^0 - 39^0	GAMMA -.01 - .18

APPENDIX II

CLASSIFICATION ACCURACY USING COMPRESSED DATA

COMPRESS is an optional data compression procedure within PROCAMS. The object of data compression is to greatly reduce the processing time required to run portions of PROCAMS and therefore reduce the cost of processing the data. COMPRESS computes a mean value for the pixels contained within each training field.

This data compression normally is performed after the preprocessing and training stages of PROCAMS and before classification.

However, before we begin to conduct extensive experiments on compressed data, we would like to know whether or not it is valid to draw inferences about results for normal uncompressed data from results obtained using compressed data.

To answer this question we examined two different types of classification: local classification and signature extension results using untransformed signatures from another site. Both compressed and uncompressed data were used for each type of classification. Nine LACIE sample segments from 1973-74 Landsat data over Kansas were used for this test. Most of the sample segments are from the Southwest Crop Reporting District of Kansas, all are from western Kansas.

Table II-1 shows local classification accuracy for Morton and Finney Counties, early in May and late in May. A comparison of average classification accuracy on compressed and uncompressed data is given. The difference between average classification accuracy using compressed and uncompressed data is 1.2%. The standard deviation of the difference in classification accuracy using the compressed and uncompressed data is 2.78%.

TABLE II-1. LOCAL CLASSIFICATION ACCURACY (Compressed vs Uncompressed Data)

Site		Classification Accuracy (%) Compressed	Classification Accuracy (%) Uncompressed
Morton	Early May	96	91
Finney	Early May	97	98
Morton	Late May	92	90
Finney	Late May	97	98
	Average:	95.5	94.3

Table II-2 shows signature extension results using untransformed signatures from remote sites. The classification accuracy is given for compressed and uncompressed data for each of twenty cases. Six of the signature extensions are from the early May data and fourteen from the late May data. The average of the difference in the classification accuracy between compressed and uncompressed data is 7.9%. The standard deviation of the difference between classification accuracies is 6.89%. The correlation coefficient between the compressed and uncompressed data is 0.856. This correlation is significant at the 0.0005 level.

These results would tend to support the belief that inferences can be drawn about the overall performance of various algorithms on normal uncompressed data from the results of tests of these algorithms on compressed data.

TABLE II-2. UNTRANSFORMED SIGNATURE EXTENSION RESULTS COMPARING COMPRESSED AND UNCOMPRESSED DATA

			Accuracy (%)	
Site From	Site To	Time Period	Not Compressed	Compressed
Morton	Finney	Early May	91	93
Morton	Grant	Early May	60	85
Morton	Haskell	Early May	78	88
Finney	Morton	Early May	76	80
Finney	Grant	Early May	71	90
Finney	Haskell	Early May	100	99
Morton	Finney	Late May	54	50
Morton	Graham	Late May	61	72
Morton	Grant	Late May	69	75
Morton	Haskell	Late May	77	86
Morton	N. Stevens	Late May	82	87
Morton	S. Stevens	Late May	57	66
Finney	Morton	Late May	53	55
Finney	Graham	Late May	64	75
Finney	Lane	Late May	85	84
Finney	Scott	Late May	87	97
Finney	Grant	Late May	54	75
Finney	Haskell	Late May	64	79
Finney	N. Stevens	Late May	55	61
Finney	S. Stevens	Late May	50	49
		Average:	69.4	77.3

ORIGINAL PAGE IS OF POOR QUALITY

APPENDIX III

DESCRIPTION OF THE PROCAMS TEST BENCH

A signature extension algorithm cannot stand alone; it requires data quality control programs, signature extraction techniques, a classifier and other related procedures and processes to form a complete classification system. For the testing of signature extension algorithms, the classification system PROCAMS was used as the test bench into which various techniques were incorporated for evaluation. PROCAMS, whose development was begun by ERIM during the FY76 contract period, was designed to be a state-of-the-art test bench for a wide range of data processing algorithms, including signature extension algorithms.

The PROCAMS system consists of several modules which can be grouped into five general subsystems: preprocessing, data compression, training, signature transformation, and classification. A brief description of the five subsystems of PROCAMS follows, together with a flow chart (Figure III-1).

The preprocessing portion of PROCAMS consists of set-up programs, data quality algorithms, and, optionally, a haze correction technique. Originally there were two routines which performed the function of preparing the data for PROCAMS. These are PRECAMS, a subroutine to set up the header record with information needed for subsequent processing, and SUBTIME, a subroutine which selects the spatial and temporal subset of the data which is to be processed and modifies the header information accordingly. Data quality algorithms include subroutine BADLINE, which detects and flags bad data lines using a data channel which is appended for just this purpose, and subroutine CLOUD which identifies and similarly records pixels which correspond to clouds, cloud shadow, and water. These four programs were later replaced by one program called SCREEN [18]. The final (aid optional) stage of the preprocessing is haze correction.

98
PAGE INTENTIONALLY BLANK

PRECEDING PAGE BLANK NOT FILMED

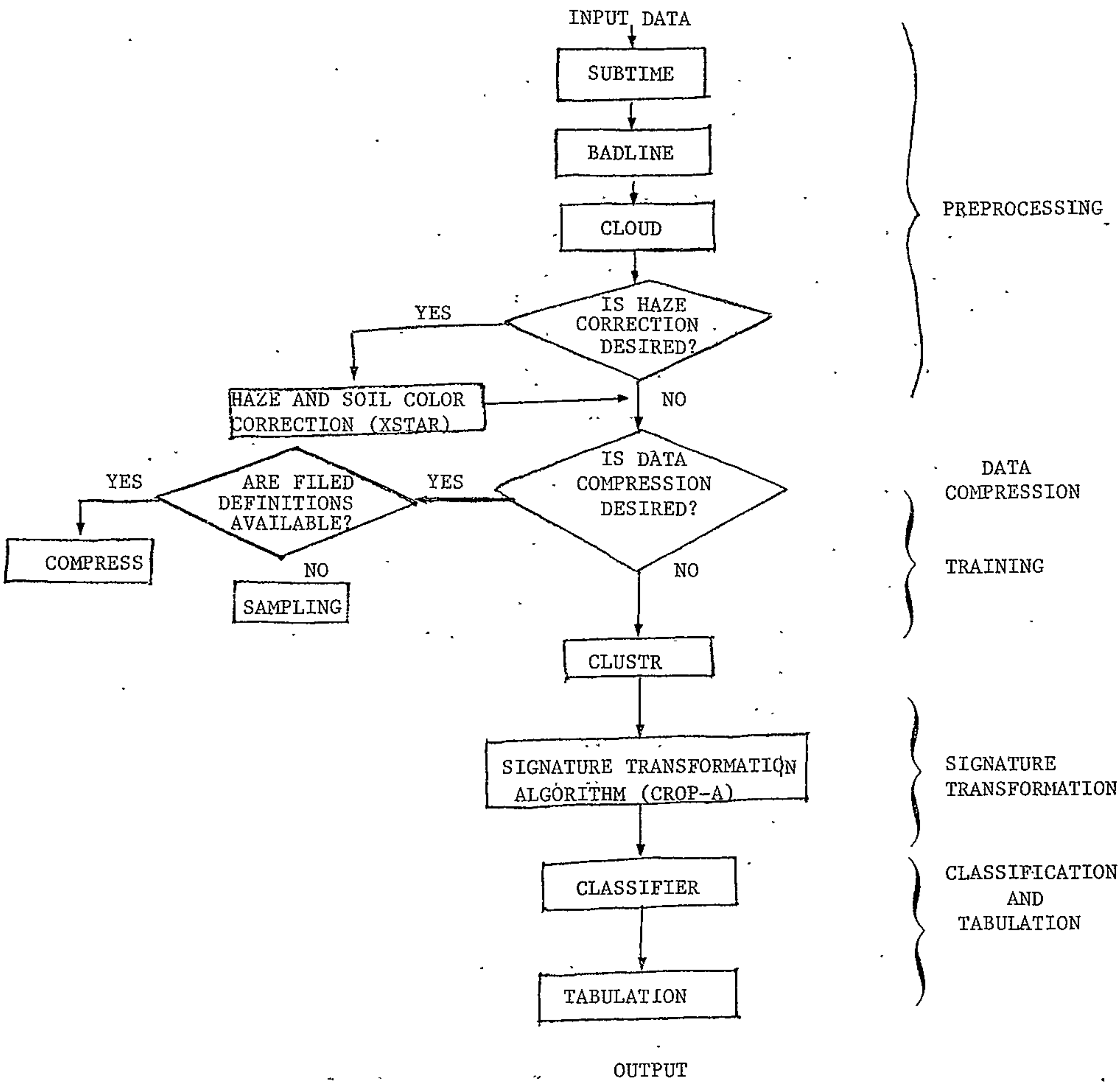

FIGURE III-1. FLOW CHART OF THE PROCAMS SYSTEM

Data compression is an optional step in PROCAMS which is used to lower processing costs when several passes through the data are anticipated. Two types of data compression were used in PROCAMS. The first data compression technique computes the average signal values over each field to produce a mean value or "average pixel". This subroutine, called COMPRESS, yields data compression ratios of up to 100 to 1. This technique is applicable only when fields have been defined.

When proportion estimation results are desired, the data may be sampled randomly to achieve an effective data compression.

The third step of PROCAMS (training) is implemented in ERIM's clustering algorithm CLUSTR.

The fourth subsystem in PROCAMS (signature transformation) is signature extension, a role which is filled by the cluster matching routine CROP-A developed by ERIM.

The final portion of PROCAMS consists of the classification and tabulation programs. PROCAMS uses a sum-of-likelihoods decision rule for its classifier, similar to the one used in the LACIE classification and mensuration subsystem. Properly trained, this classifier has been shown to perform nearly as well as any classifier yet designed.

APPENDIX IV

TWO APPROACHES TO MULTISEGMENT PROCEDURE 1

This appendix addresses the classification technique to be used in evaluating static stratification in a multisegment environment. We have termed this approach 'preclassification'.

Overall Objective

Develop an experiment design which will efficiently and effectively evaluate static stratification of space image data in a multisegment signature extension environment for the purpose of large area crop inventory.

Environment and Training Selection

The current LACIE Procedure 1 provides an environment wherein a large number of segments are classified using local training procedures and crop proportion estimates computed by pixel count.

The multisegment signature extension environment is one wherein an attempt would be made in reducing the need for local training. A certain subset of segments would be designated training sites. Clusters would be computed from these segments, labeled according to their association to training dots, and used in classification throughout. Hence, specific segments can be more intensely photointerpreted for training, hopefully with a resultant reduction of labeling error.

The multisegment signature extension approach, however, poses a training segment selection problem. The resultant classification is sensitive to variational differences between training and test segments. The designation of static stratifications of segments using variables such as soil type and precipitation is an attempt to associate segments in a manner that would minimize the spectral differences between like classes in segments belonging to the same strata. These stratifications can then be used in one of two ways:

1. For training selection purposes: To insure that all spectral classes are represented in choosing training segments from every strata to be used across all segments in classification.

2. For classification purposes: Segments would be classified using signature clusters determined within their stratum only.

These two applications of static stratification in multisegment signature extension can be generalized.

Consider n strata and m segments where $n \leq m$.

Segment s_{ij} is the j^{th} segment of the i^{th} stratum S_i.

The signature set for segment s_{ij} is $SIG(s_{ij})$.

The training data for stratum S_i is $T(S_i)$.

Call the clustering function $\prod$, then

$$SIG(s_{ij}) \equiv \prod_{k=1}^{n} \omega_k T(S_k)$$

where ω_i is a weight for each stratum.

If for $k = i \quad \omega_k = 1$

$k \neq i \quad \omega_k = 0$

then Case 2 above is implied, i.e., the segment is classified using signatures computed only within its own stratum.

If

$$\omega_i = \omega_j \quad \text{for all } i,j$$

then Case 1 is implied, i.e., a segment is classified using all signatures, but insuring that each stratum is represented.

The value of introducing this notation lies in that the weights ω_k can vary anywhere between the two cases. For example, it may be useful to use stratification for training and in computing $SIG(s_{ij})$, weighting the training data from stratum i ($T(S_i)$ more than for other strata. This recognizes that important information for any one segment appears in every stratum, however, it is more likely that training within the same stratum would be more significant.

Terminology

For purposes of further discussion, reference will be made to three partitions of data: (1) within segment, all pixels from a 5x6 mile LACIE segment; (2) within stratum, all segments within a defined stratification of segments; (3) within universe, all segments.

Problem

Any evaluation of the inherent value of static stratification in a multisegment environment will require measures of performance that are statistically significant. These measures may include: (1) within segment classification accuracy, (2) within stratum classification accuracy, and (3) within universe classification accuracy. Each of these measures may be determined as a function of training gain. As a result, a large number of classifications must be performed for a large number of segments, varying the training data at each classification. The cost of such an experiment could be prohibitive. What legitimate training and classification algorithm should be employed to maximize testing efficiency? In other words, what logical extension of Procedure 1 into a multisegment environment will be required to evaluate static stratification?

Two Approaches to Multisegment Procedure 1

The following pages document two approaches to extending Procedure 1 into a multisegment environment. The second approach is called preclassification and is described to be logically equivalent to the first approach. The first approach is a straightforward extension of Procedure 1. Before getting into the details of each, consider the following graphics in order to group the salient aspects of each approach.

The first approach combines the training data first, extracts signatures from the combined training data set, then estimates proportions for wheat and non-wheat. Preclassification differs in that information from the training segments is not combined until after likelihoods are calculated. The particular advantage of this approach for test and evaluation purposes lies in the fact that training segment selection does not have to be carried out first. The details of these two approaches are described in the following sections.

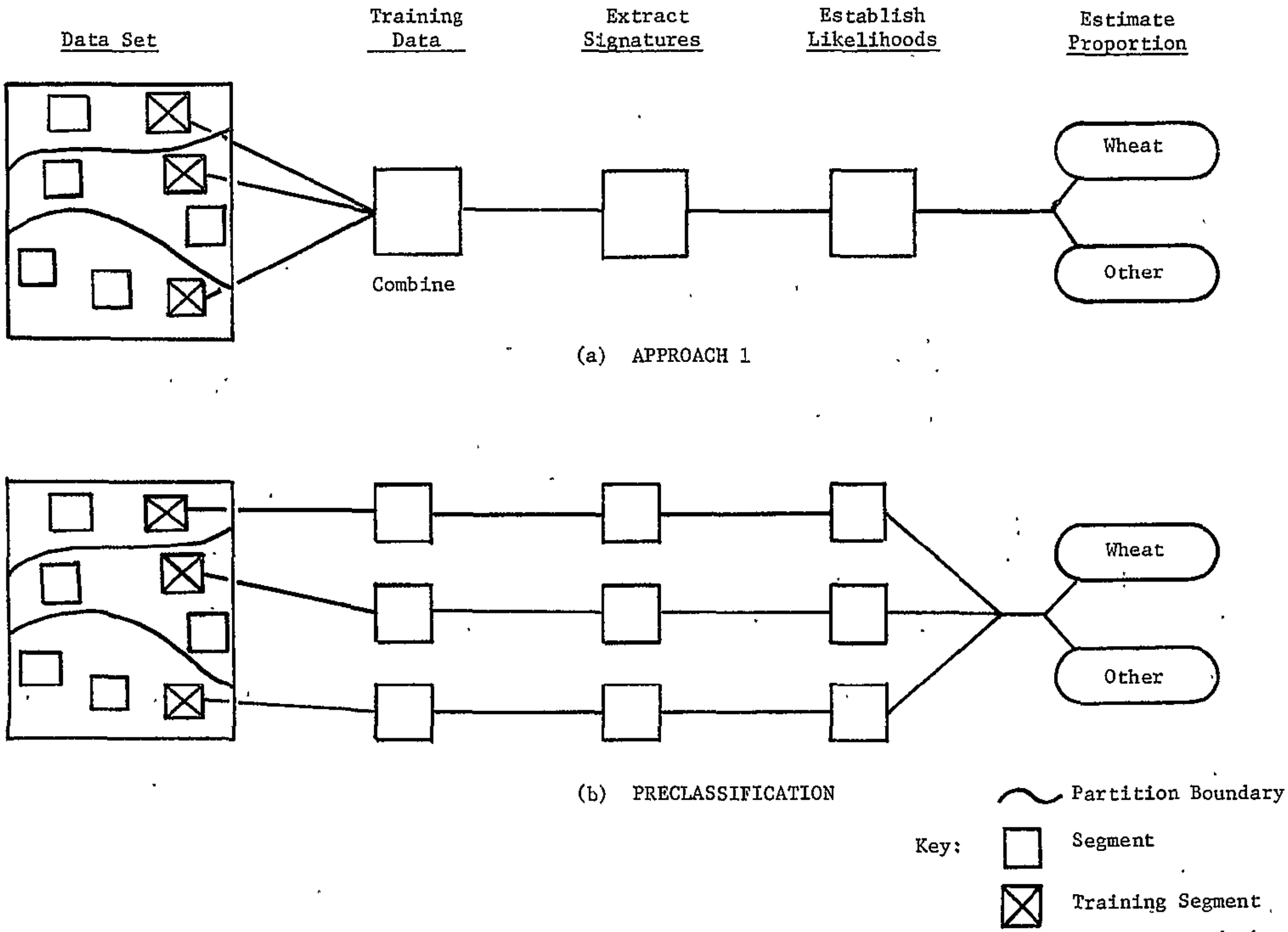

FIGURE IV-1. ILLUSTRATIONS OF TWO APPROACHES IN EXTENDING PROCEDURE 1 TO A MULTISEGMENT ENVIRONMENT

APPROACH 1

Consider the following approach:

1. Select training segments from each strata
2. Merge training segment data together
3. Cluster the training pixels into subclasses
4. Calculate proportion estimates using sum of likelihoods.

First note that in an evaluation experiment, using this approach would result in clustering and classification of data each time training parameters are changed.

However, the procedure is a straightforward extension of Procedure 1. Important decisions must be made along the way.

1. Weighting Training Segments Due to Random Selection Process

First of all, the selection of training segments must be carried out in a manner that would simulate the random selection of training fields. On an average the number of randomly selected fields would be in like proportions from stratum to stratum as a function of the total number of fields in each stratum. For example, suppose the universe of data is comprised of two strata, each with ten fields. If six of those fields were to be selected at random from the twenty, one would expect each stratum to be represented by three. To simulate this, training segments should be drawn from each strata in like proportions. Suppose, however, that two strata contained 8 and 6 segments respectively. If the training gain desired was 3.0, i.e., one-third of each strata required for training, the first stratum would require 2.7 segments, the second 2 segments. Since the selection of 2.7 segments is not possible, one may round and select 3 segments. In order to reflect this adjustment affecting the random character of the selection, weights need to be assigned to the training data as follows:

For segment s_{ij}, the j^{th} segment - the i^{th} stratum, S_i,

Let s'_{ij} be a training segment

Let t_i be the number of segment in the i^{th} stratum and t'_i the number of training segment in the i^{th} stratum.

Each segment s'_{ij} is assigned the weight:

$$\omega'_i = \frac{t_i}{t'_i}$$

Recalling the earlier defined weight ω_i in the definition of a set of training clusters, we can extend its definition to

$$\omega_i = \mu_i \omega'_i$$

where μ is related as mentioned to the classification technique employed. This more fully discussed in what follows:

2. Weighting Training Segments With Respect to Classification Segments

As was mentioned earlier, data stratification could be used for purposes of training only, or for purposes of classification as well. The technique employed is related by a factor μ_i of the weight ω_i assigned to each pixel of training data. If you recall, if $\mu_i = 1$ for all segments in strata i, then classification of segment s_{ij} is determined only by those signature clusters defined from stratum S_i. However, this weight may be adjusted to better represent one's confidence in the training data available in each stratum when applied to an arbitrary segment. This approach implies that the classifier has no confidence in applying signatures derived from data from other stratum. Another approach is to employ equal levles of confidence. An interim approach may be to establish confidence levels empirically. For example, for purposes of our test and evaluation the experiments constructed in FY77 provide within stratum and across strata classification results.

The weight μ_i may be assigned so that segment s_{ij} from stratum S_i would have associated weights μ_k and μ_i.

μ_k = average error in signature extension $S_k \rightarrow S_i$ for all $k \neq i$

μ_i = average error in signature extension $S_i \rightarrow S_i$
(i.e., segments extend within stratum)

μ_k is applied to training data from S_k

μ_i is applied to within stratum training data

Note that these weights would vary from segment to segment. Since clusters are computed before classification, each strata would require a different set of signature clusters rendering this approach impractical for test and evaluation purposes and making it clumsy for an operational system.

3. Weighting Clusters in Sum of Likelihoods

Training pixels within training segments can be selected using a technique that attempts to insure representativeness, much as the CAMS AI training selection approach, or selected randomly, as in the Procedure 1 209-point technique. The former requires that each derived cluster be weighted equally in classification when computing sum of likelihood. That is, pixel $\overline{x}$ is wheat if

for m wheat clusters and n non-wheat clusters with likelihoods p_W and p_N respectively

$$\frac{1}{n} \sum_{i=1}^{n} p_{iW} > \frac{1}{m} \sum_{j=1}^{m} p_{jN}$$

However random selection of training pixels requires that:

$\overline{x}$ is wheat if

$$\sum_{i=1}^{n} n_i p_{iW} > \sum_{i=1}^{m} n_j p_{jN}$$

where n_k is the number of pixels in cluster k. That is to say, clusters are not equally weighted but in proportion to the number of samples they represent.

APPROACH 2 (Preclassification)

Consider the following alternative approach:

1. Select training segments from each stratum
2. Cluster the training pixels into subclasses independently from each training segment
3. Employ the following classification procedure:
   i. Classify each segment independently using the clusters from each other segment, determining a wheat and non-wheat likelihood (i.e., for m training segments, each segment is classified m times).
   ii. Sum likelihoods from each training segment to determine wheat proportion estimate.

This approach offers two advantages for the test and evaluation of multi-segment signature extension.

First of all, determination of likelihoods can be performed before training segments are selected. Clusters can be computed for every segment and applied in classifying every other segment. Proportion estimation can be carried out for a variety of different training segments, simply by summing the computed likelihoods corresponding to the training segments. Clustering and likelihood calculation does not have to be recomputed for each different set of training data.

More graphically, consider the following situation: given 5 training segments each pixel $\overline{x}$ would have a vector associated with it as follows:

$$(\overline{x}, \overline{p}_W, \overline{p}_N)$$

where:

$\overline{x}$ is the n channel mean vector

$\overline{p}_W$ are wheat likelihoods corresponding to each of 5 training segments

$\overline{p}_N$ are non-wheat likelihoods corresponding to each of 5 training segments

Using Segments 2 and 3 as training $\overline{x}$ would be wheat if

$$p_{W2} + p_{W3} > p_{N2} + p_{N3}$$

Using Segments 1 and 5 as training $\overline{x}$ would be wheat if

$$p_{W4} + p_{W5} > p_{N4} + p_{N5}$$

A second advantage is that the weighting factor μ can be applied at classification, i.e., tally time to reflect a stratum training confidence level. For example, if Segments 2 and 3 in the above example were representatives of strata i and j, then a pixel $\overline{x}$ from stratum S_i would be wheat if:

$$\mu_i p_{W2} + \mu_j p_{W3} > \mu_i p_{N2} + \mu_j p_{N3}$$

What needs to be established is whether this technique appropriately simulates the first approach. The essential difference is that in the first approach clusters are determined for all training pixels at once, rather than separate sets of clusters for each training segment. A subclass appearing in two segments would be represented at tally time by two clusters, whereas only one cluster would appear using Approach 1.

We shall assume that the selection of training is done randomly. Algebraically, the procedure is as follows:*

1. Determine the likelihood that $\overline{x}$ is wheat given each training segment.

Given n training signatures $SIG(s_{ij})$ for the j^{th} segment of the i^{th} stratum

then the likelihood that a pixel $\overline{x}$ belongs to the wheat (or non-wheat) signature sig_i is $p_W(\overline{x}|sig_i)$ or $p_N(\overline{x}|sig_i)$

The sum of likelihoods that $\overline{x}$ is wheat is given by $p_W(\overline{x}|SIG(s_{ij}))$ where:

$$p_W(\overline{x}|SIG(s_{ij})) = \sum_{k=1}^{n} n_k p_W(\overline{x}|sig_k)$$

where n_k is the number of training pixels in sig_k

*Shown for wheat, similarly for non-wheat.

The total number of training wheat pixels in $SIG(s_{ij})$ is given by

$$\eta_W = \sum_{k=1}^{n} n_k$$

2. <u>Determine $\bar{x}$ is wheat given all training segments.</u>

Let a set of m training segments be represented by $\{s_{ij}\}$.
Let the signatures derived from these training segments be $SIG\{s_{ij}\}$.
Then the likelihood that a pixel $\bar{x}$ is wheat is given by

$$p_W(\bar{x} | SIG\{s_{ij}\})$$

where

$$p_W(\bar{x} | SIG\{s_{ij}\}) = \frac{\sum_{k=1}^{m} \omega_k p_W(\bar{x} | SIG(s_{ij}))}{\sum_{k=1}^{m} \omega_k \eta_{Wk}}$$

where ω_k is the weight earlier defined in Approach 1. $\bar{x}$ is wheat if

$$p_W(\bar{x} | SIG\{s_{ij}\}) > p_N(\bar{x} | SIG\{s_{ij}\})$$

Approach 2 is an appropriate simulation of the Approach 1 under the assumption of random selection of training pixels within a segment. Differences in the training procedures are accounted for by weighting, at classification, each computed cluster subclass by its number of pixel members. Hence, using Approach 2, a subclass appearing in two segments, though represented by two clusters, are weighted in such a way so as to contribute the same likelihoods as the corresponding single cluster that would have been computed using Approach 1.

APPENDIX V

DESCRIPTION OF DATA EMPLOYED IN ANALYST-INTERPRETER LABELING ERROR ANALYSIS

The following tables, V-1 to V-3, list the LACIE Blind Site Acquisitions for three biowindows employed in the Analyst-Interpreter Labeling Error Analysis described in Section 4.4. Pertinent ancillary information is also encoded in these tables as well as summarized in table V-4.

TABLE V-1. BIOWINDOW ONE ACQUISITIONS

Segment Number	Julian Date 1975	Crop* Calendar	Missed Wheat Fraction	False Wheat Fraction
1035	312	0.0	0.28	0.12
1041	312	0.0	0.28	0.12
1154	311	0.0	0.03	0.02
1163	327	0.0	0.18	0.0
1164	326	0.0	0.0	0.70
1165	326	0.0	0.0	0.07
1166	327	0.0	0.16	0.10
1167	327	0.0	0.28	0.0
1171	364	0.0	0.13	0.0
1172	328	0.0	0.28	0.0
1176	364	0.0	0.44	0.0
1179	364	0.0	0.20	0.0
1181	345	0.0	0.08	0.0
1852	295	0.0	0.20	0.05
1854	295	0.0	0.28	0.0
1865	349	0.0	0.20	0.0
1880	311	0.0	0.15	0.0
1882	311	0.0	0.33	0.0
1883	328	0.0	0.0	0.0
1887	311	0.0	0.07	0.0

* 0.0 implies information not available.

TABLE V-2. BIOWINDOW TWO ACQUISITIONS

Segment Number	Julian Date	Crop Calendar	Missed Wheat Fraction	False Wheat Fraction
1020	128	3.17	0.09	0.0
1035	127	3.40	0.28	0.12
1041	127	3.40	0.28	0.12
1154	090	2.76	0.03	0.02
1163	124	3.61	0.18	0.0
1164	124	3.52	0.0	0.70
1165	124	3.61	0.0	0.07
1166	124	3.52	0.16	0.10
1167	124	3.52	0.28	0.0
1171	125	3.50	0.13	0.0
1184	124	3.66	0.23	0.0
1851	127	3.22	0.28	0.06
1861	128	3.30	0.17	0.08
1865	127	3.42	0.20	0.0
1884	125	3.50	0.18	0.0
1886	127	3.46	0.27	0.07
1887	127	3.35	0.07	0.0

TABLE V-3. BIOWINDOW THREE ACQUISITIONS

Segment Number	Julian Date 1976	Crop Calendar	Missed Wheat Fraction	False Wheat Fraction
1019	164	4.60	0.07	0.0
1163	142	3.98	0.18	0.0
1167	142	3.93	0.28	0.0
1169	144	4.00	0.27	0.35
1180	141	4.11	0.24	0.02
1854	154	4.14	0.28	0.0
1857	154	4.10	0.33	0.10
1861	164	4.55	0.17	0.08
1865	136	3.58	0.20	0.0
1880	127	3.34	0.15	0.0
1882	152	4.15	0.33	0.0
1887	135	3.55	0.07	0.0

TABLE V-4. DESCRIPTION OF ANCILLARY DATA

	Variable	N	Minimum	Maximum	Mean	Std Dev
1.	Segment	39	1019.0	1988.0	---	---
2.	Number of Wheat Fields	39	0.0	32.0	12.4	6.32
3.	Number of Other Fields	39	9.0	46.0	21.0	8.42
4.	Number of Missed Wheat Fields	39	0.0	8.0	2.05	1.97
5.	Number of False Wheat Fields	39	0.0	12.0	0.90	2.11
6.	Fraction of Missed Wheat	38	0.0	0.44	0.16795	0.128
7.	Fraction of False Wheat	39	0.0	0.706	0.04912	0.126
8.	Fraction of Total Error	39	0.0	0.706	0.103	0.125
9.	Number of Fields	39	17.0	75.0	33.4	12.7
10.	Julian Date 1	39	294	127	---	---
11.	Julian Date 2	39	311	128	---	---
12.	Julian Date 3	39	364	199	---	---
13.	Degree-days	38	1910.0	2540.0	2245.7	146.38
14.	Crop Calendar 1	39	0.0	3.4	0.49	1.07
15.	Crop Calendar 2	39	0.0	3.66	2.7	1.32
16.	Crop Calendar 3	39	0.0	6.0	4.0	0.90
17.	GAMMA 1	38	- 0.6	0.22	0.03	0.07
18.	GAMMA 2	37	- 0.22	0.20	0.01	0.07
19.	GAMMA 3	38	- 0.26	0.14	- 0.03	0.09
20.	Elevation	39	0.0	3500.0	1882.1	826.11
21.	THETA 1	39	35.0	69.0	---	---
22.	THETA 2	39	35.0	68.0	---	---
23.	THETA 3	39	31.0	68.0	---	---
24.	Precipitation	39	0.0	15.0	7.9	4.40
25.	Land Use	39	0.0	4.0	2.3	1.67
26.	Latitude	39	37.0	39.70	38.3	0.80
27.	Longitude	39	94.8	101.80	98.4	2.06
28.	Haze Diagnostic 1	39	- 1.36	4.61	0.53	1.44
29.	Haze Diagnostic 2	39	- 4.26	3.67	0.21	1.39
30.	Haze Diagnostic 3	39	- 4.45	2.96	- 0.71	1.76

REFERENCES

1. Lambeck, P. F., and D. P. Rice, "Signature Extension Using Transformed Cluster Statistics and Related Techniques," Report No. 109600-70-F, Environmental Research Institute of Michigan, Ann Arbor, Michigan, May 1976.

2. Lambeck, P., W. Malila, and J. Gleason, "Task 2: Signature Extension Preprocessing Algorithm Development," Quarterly Progress Report, 17 May - 14 August 1976, ERIM Report No. 122700-5-L, NASA Contract Number NAS9-14988, Environmental Research Institute of Michigan, Ann Arbor, Michigan, August 1976.

3. Information Systems and Analysis Department, "Partitioning Requirements for Signature Extension," ERIM Technical Memo, 122700-13-R, Environmental Research Institute of Michigan, Ann Arbor, Michigan, January 1977.

4. Colwell, R. N., C. M. Hay, and R. W. Thomas, "Application of Photointerpretive Techniques to Wheat Identification, Signature Extension and Sampling Strategy," NASA Contract Number NAS9-14565, Space Sciences Laboratory, University of California, Berkeley, California, 14 May 1976.

5. "Final Partitioning of Kansas," Lockheed Electronics Company Memo, Reference 642-1935, May 6, 1976.

6. "MRAD Phase I LACIE QA Results," JSC Internal Note, 76-FM-16, September 1, 1976.

7. Turner, Robert E., "Atmospheric Effects in Multispectral Remote Sensor Data," ERIM Report Number 109600-15-F, Environmental Research Institute of Michigan, Ann Arbor, Michigan, May 1975.

8. Malila, W. A., R. B. Crane, and R. E. Turner, "Information Extraction Techniques for Multispectral Scanner Data," Report Number, WRL 31650-74-T, Willow Run Laboratories, Ann Arbor, Michigan, June 1972, Appendix I.

9. Kauth, R. J., and G. Thomas, "The Tasselled Cap--A Graphic Description of the Spectral-Temporal Development of Agricultural Crops as seen by Landsat," Purdue/LARS Symposium on Machine Processing of Remotely Sensed Data, Purdue University, West Lafayette, Indiana, June 1976.

116

PAGE INTENTIONALLY BLANK

10. Pentland, A. P., "Evaluation of Signature Extension Algorithms," ERIM Technical Memo, 122700-29-T, Environmental Research Institute of Michigan, Ann Arbor, Michigan, September 1977.

11. Phase III Detailed Analysis Procedure, NASA/JSC, LACIE-00720, JSC-11693, Houston, Texas, August 1977.

12. Kauth, R. J. and W. Richardson, "Procedure B: A Multisegment Training Selection and Proportion Estimation Procedure for Processing Landsat Agricultural Data," ERIM Report Number 122700-31-F, Environmental Research Institute of Michigan, Ann Arbor, Michigan, November 1977.

13. Coberly, W. A., J. Tubbs, P. L. Odell, "The Influence of False Color Infrared Film on Training Errors in LACIE."

14. Judd, D. B., G. Wyszecki, Color in Business Science and Industry, John Wiley & Sons, New York, 1975.

15. Wyszecki, G., and W. S. Stiles, Color Science, John Wiley & Sons, New York, 1967.

16. Kaneko, T., "Flagging Algorithms for the Current LACIE PFC Products," IBM Memorandum to J. Engvoll, S. Tinkler, NASA, December 3, 1976.

17. Austin, W. W., "PFC Film Products for LACIE Segments," Memorandum to L. M. Flores, NASA, December 6, 1976.

18. Lambeck, P. F., "Signature Extension Preprocessing for Landsat MSS Data," ERIM Report Number 122700-32-F, Environmental Research Institute of Michigan, Ann Arbor, Michigan, November 1977.

Technical and Final Report Distribution List

NASA Contract NAS9-14988

NAME	NUMBER OF COPIES
NASA/Johnson Space Center Earth Observations Division Houston, Texas 77058	
ATTN: Mr. Robert MacDonald/SF	(1)
ATTN: Dr. J. D. Erickson/SF3	(1)
ATTN: Dr. F. Hall/SF2	(1)
ATTN: Mr. J. Murphy/SF2	(1)
ATTN: Dr. R. P. Heydorn/SF3	(1)
ATTN: Mr. M. C. Trichel/SF3	(1)
ATTN: Mr. K. J. Demel/SF3	(1)
ATTN: Dr. F. P. Weber/SF5	(1)
ATTN: Mr. I. D. Browne/SF3	(5)
ATTN: Mr. J. Dragg/SF4	(1)
ATTN: Ms. E. Laity/SF121	(2)
ATTN: Mr. G. E. Graybeal/SF4	(2)
ATTN: Mr. K. Baker/SF4	(1)
NASA/Johnson Space Center LACIE Development & Evaluation Department Houston, Texas 77058	
ATTN: Mr. B. L. Carroll/C09	(1)
NASA/Johnson Space Center Mathematical Physics Branch Houston, Texas 77058	
ATTN: Dr. H. G. De Vezin, Jr./FM8	(1)
NASA/Johnson Space Center Earth Resources Program Office Office of the Program Manager Houston, Texas 77058	(1)
NASA/Johnson Space Center Earth Resources Program Office Program Analysis & Planning Office Houston, Texas 77058	
ATTN: Dr. O. Glenn Smith/HD	(1)

NAME	NUMBER OF COPIES
NASA/Johnson Space Center Earth Resources Program Office Systems Analysis and Integration Office Houston, Texas 77058	
ATTN: Mr. Richard A. Moke/HC	(1)
ATTN: Mr. M. Jay Harnage, Jr./HC	(1)
NASA/Johnson Space Center Technical Library Branch Houston, Texas 77058	
ATTN: Ms. Retha Shirkey/JM6	(4)
NASA/Johnson Space Center Technical Support Procurement Houston, Texas 77058	
ATTN: Mr. Ben McGuire/BB6	(1)
Earth Resources Laboratory, GS Mississippi Test Facility Bay St. Louis, Mississippi 39520	
ATTN: Mr. D. W. Mooneyhan	(1)
EROS Data Center U.S. Department of Interior Sioux Falls, South Dakota 57198	
ATTN: Mr. G. Thorley	(1)
Department of Mathematics Texas A&M University College Station, Texas 77843	
ATTN: Dr. Larry Guseman	(1)
NASA/Johnson Space Center Mission Planning & Analysis Division Houston, Texas 77058	
ATTN: Mr. Eugene Davis/FA	(1)

NAME	NUMBER OF COPIES
U.S. Department of Agriculture Agricultural Research Service Washington, D.C. 20242 ATTN: Dr. Robert Miller	(1)
U.S. Department of Agriculture Soil & Water Conservation Research Division P.O. Box 267 Weslaco, Texas 78596 ATTN: Dr. Craig Wiegand	(1)
U.S. Department of Interior USGA National Center Mail Stop 115 Geography Program Reston, Virginia 22092 ATTN: Dr. James R. Anderson	(1)
Director, Remote Sensing Institute South Dakota State University Agriculture Engineering Building Brookings, South Dakota 57006 ATTN: Mr. Victor I. Myers	(1)
U.S. Department of Interior Fish & Wildlife Service Bureau of Sport Fisheries & Wildlife Northern Prairie Wildlife Research Center Jamestown, North Dakota 58401 ATTN: Mr. Harvey K. Nelson	(1)
U.S. Department of Agriculture Forest Service 240 W. Prospect Street Fort Collins, Colorado 80521 ATTN: Dr. Richard Driscoll	(1)

NAME	NUMBER OF COPIES
U.S. Department of Interior Geological Survey 801 19th Street, N.W. Washington, D.C. 20242 ATTN: Mr. M. Deutsch	(1)
U.S. Geological Survey 801 19th Street, N.W., Room 1030 Washington, D.C. 20242 ATTN: Dr. Jules D. Friedman	(1)
U.S. Department of Interior Geological Survey Federal Center Denver, Colorado 80225 ATTN: Dr. Harry W. Smedes	(1)
U.S. Department of Interior Geological Survey Water Resources Division 901 S. Miami Ave. Miami, Florida 33130 ATTN: Mr. Aaron L. Higer	(1)
University of California School of Forestry Berkeley, California 94720 ATTN: Dr. Robert Colwell	(1)

NAME	NUMBER OF COPIES
Environmental Remote Sensing Applications Laboratory Oregon State University Corvallis, Oregon 97331 ATTN: Dr. Barry J. Schrumpf	(1)
U.S. Department of Interior Director, EROS Program Washington, D.C. 20242 ATTN: Mr. J. M. Denoyer	(1)
U.S. Department of Interior Geological Survey GSA Building, Room 5213 Washington, D.C. 20242 ATTN: Mr. W. A. Fischer	(1)
NASA Wallops Wallops Station, Virginia 23337 ATTN: Mr. James Bettle	(1)
Purdue University Purdue Industrial Research Park 1200 Potter Drive West Lafayette, Indiana 47906	
ATTN: Dr. David Landgrebe	(1)
ATTN: Mr. Terry Phillips	(1)
ATTN: Dr. Marvin Bauer	(1)
ATTN: Dr. Philip Swain	(1)
U.S. Department of Interior EROS Office Washington, D.C. 20242 ATTN: Dr. Raymond W. Fary	(1)

NAME	NUMBER OF COPIES
U.S. Department of Interior Geological Survey 801 19th Street, N.W. Washington, D.C. 20242 ATTN: Mr. Charles Withington	(1)
U.S. Department of Interior EROS Office Washington, D.C. 20242 ATTN: Mr. William Hemphill	(1)
Chief of Technical Support Western Environmental Research Laboratories Environmental Protection Agency P.O. Box 15027 Las Vegas, Nevada 89114 ATTN: Mr. Leslie Dunn	(1)
NASA/Langley Research Mail Stop 470 Hampton, Virginia 23365 ATTN: Mr. William Howle	(1)
U.S. Geological Survey Branch of Regional Geophysics Denver Federal Center, Building 25 Denver, Colorado 80225 ATTN: Mr. Kenneth Watson	(1)
NAVOCEANO, Code 7001 Bay St. Louis, MS 39520 ATTN: Mr. J. W. Sherman, III	(1)

NAME	NUMBER OF COPIES
U.S. Department of Agriculture Administrator Agricultural Stabilization and Conservation Service Washington, D.C. ATTN: Mr. Kenneth Frick	(1)
Pacific Southwest Forest & Range Experiment Station U.S. Forest Service P.O. Box 245 Berkeley, California 94701 ATTN: Mr. R. C. Heller	(1)
University of Texas at Dallas Box 688 Richardson, Texas 75080 ATTN: Dr. Patrick L. Odell	(1)
Department of Mathematics University of Houston Houston, Texas 77004 ATTN: Dr. Henry Decell	(1)
Institute for Computer Services and Applications Rice University Houston, Texas 77001 ATTN: Dr. M. Stuart Lynn	(1)
U.S. National Park Service Western Regional Office 450 Golden Gate Avenue San Francisco, California 94102 ATTN: Mr. M. Kolipinski	(1)

NAME	NUMBER OF COPIES
U.S. Department of Agriculture Statistical Reporting Service Room 4833, South Bldg. Washington, D.C. 20250 ATTN: G. F. Hart/W. H. Wigton	(2)
U.S. Department of Agriculture Statistical Reporting Service Washington, D.C. 20250 ATTN: Mr. H. L. Trelogan, Administrator	(1)
Ames Research Center National Aeronautics & Space Administration Moffett Field, California 94035 ATTN: Dr. D. M. Deerwester	(1)
Goddard Space Flight Center National Aeronautics & Space Administration Greenbelt, Maryland 20771 ATTN: Mr. W. Alford, 563 ATTN: Dr. J. Barker, 923	(1) (1)
Lewis Research Center National Aeronautics & Space Administration 21000 Brookpark Road Cleveland, Ohio 44135 ATTN: Dr. Herman Mark	(1)
John F. Kennedy Space Center National Aeronautics & Space Administration Kennedy Space Center, Florida 32899 ATTN: Mr. J. P. Claybourne/AA-STA	(1)
NASA/Langley Mail Stop 214 Hampton, Virginia 23665 ATTN: Mr. James L. Raper	(1)

NAME	NUMBER OF COPIES
Texas A&M University Institute of Statistics College Station, Texas 77843 ATTN: Dr. H. O. Hartley	(1)
Texas Tech University Department of Mathematics P.O. Box 4319 Lubbock, Texas 79404 ATTN: Dr. T. Boullion	(1)
University of Tulsa Math-Sciences Department 600 South College Tulsa, Oklahoma 74104 ATTN: Dr. William A. Coberly	(1)
S&D - DIR Marshall Space Flight Center Huntsville, Alabama 35812 ATTN: Mr. Cecil Messer	(1)
Code 168-427 Jet Propulsion Laboratory 4800 Oak Grove Drive Pasadena, California 91103 ATTN: Mr. Fred Billingsley	(1)
NASA Headquarters Washington, D.C. 20546	
ATTN: Mr. W. Stoney/ER	(1)
ATTN: Mr. Leonard Jaffe/ER	(1)
ATTN: Mr. M. Molloy/ERR	(1)
ATTN: Mr. James R. Morrison/ERR	(1)
ATTN: Ms. Ruth Whitman/ERR	(1)

NAME	NUMBER OF COPIES
Mr. James D. Nichols Space Sciences Laboratory, Room 260 University of California Berkeley, California 94720	(1)
Texas A&M University Remote Sensing Center College Station, Texas 77843 ATTN: Mr. J. C. Harlan	(1)
U.S. Department of Agriculture 12th & Independence, SW Room 3745-S Washington, D.C. 20250 ATTN: Mr. Clark Ison LACIE Project Office (FAS)	(1)
University of Arkansas Mathematics Department Fayetteville, Arkansas 72704 ATTN: Dr. Jack D. Tubbs	(1)
U.S. Department of Agriculture Foreign Agricultural Service Washington, D.C. 20250 ATTN: Dr. Howard L. Hill	(1)
University of California Remote Sensing Laboratory 129 Mulford Hall Berkeley, California 94720 ATTN: Ms. Claire M. Hay	(1)
IBM 1100 NASA Road One Houston, Texas 77058 ATTN: Mr. R. E. Oliver/Code 56	(1)