(E78-10021) EVALUATION OF SIGNATURE EXTENSION ALGORITHMS Interim Technical Report, 15 May 1976 - 31 Aug. 1977 (Environmental Research Inst. of Michigan) 76 p HC A05/MF A01 CSCL 05B G3/43

N78-12498

Unclas 00021

NASA CR-
ERIM 122700-29-T
7.8-1 0 0 2 1
NASA CR-
151537

Interim Technical Report

EVALUATION OF SIGNATURE EXTENSION ALGORITHMS

Alex P. Pentland
Infrared and Optics Division

September 1977

Prepared for
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Johnson Space Center
Houston, Texas 77058
Contract No. NAS9-14988

ENVIRONMENTAL
RESEARCH INSTITUTE OF MICHIGAN
FORMERLY WILLOW RUN LABORATORIES THE UNIVERSITY OF MICHIGAN
BOX 618 • 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.

Disclaimers. This document 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 document 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 document.

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 document 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.	2. Government Accession No.	3 Recipient's Catalog No.
122700-29-T		

4. Title and Subtitle	5. Report Date
Evaluation of Signature Extension Algorithms	September 1977
	6. Performing Organization Code
7. Author(s)	8 Performing Organization Report No.
Alex P. Pentland	122700-29-T
9. Performing Organization Name and Address	10. Work Unit No.
Environmental Research Institute of Michigan Infrared and Optics Division P.O. Box 8618. Ann Arbor, Michigan 48107	11. Contract or Grant No. NAS9-14988
12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered May 15, 1976 through August 31, 1977
National Aeronautics and Space Administration Johnson Space Center Houston, Texas 77058	14. Sponsoring Agency Code

15. Supplementary Notes

This work was performed for the Earth Observations Division of the NASA Johnson Space Center. Mr. I. Dale Browne was the Technical Monitor.

16. Abstract

The primary aim of this effort which was part of Task 3 of NASA Contract NAS9-14988 was to test and evaluate signature extension techniques which are candidates for inclusion in future large area crop inventory systems.

Four types of signature extension techniques were examined: haze correction algorithms, data stratification procedures, training sample selection strategies for multisegment training, and crop development classifiers.

Algorithms tested which correct for haze effects in Landsat data were CROP-A [1] and XSTAR [2], both developed by Task 2 of this contract. Two data stratification procedures were evaluated, one by the University of California, Berkeley (UCB) [4], and one produced by Johnson Space Center (JSC) personnel [5]. The training sample selection strategy examined was Procedure B [3] developed by Task 1 of this contract. The crop development classifiers evaluated include the Delta Classifier [6], several "green indicator" classifiers, and one developed by this task.

The haze correction algorithm XSTAR, the data stratifications of both UCB and JSC and the training sample selection strategy Procedure B all showed promise for inclusion in future large area crop inventory systems.

17. Key Words (Suggested by Author(s))		18. Distribution Statement
Remote Sensing Large Area Crop Inventory Multispectral Processing Signature Extension	Haze Correction Stratification Multisegment Training Time Track Classification	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 + 82	

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

JSC Form 1424 (Rev Nov 75) NASA — JSC

PREFACE

This report describes part of a comprehensive and continuing program of research in multispectral remote sensing of the environment from aircraft and satellites and the supporting effort of ground-based researchers in recording, coordinating, and analyzing the data gathered by these means. The basic objective of this program is to improve the utility of remote sensing as a tool for providing decision makers with timely and economical information from large geographical areas.

The feasibility of using remote sensing techniques to detect and discriminate between objects or conditions at or near the surface of the earth has been demonstrated. Applications in agriculture, urban planning, water quality control, forest management, and other areas have been developed. The thrust of this program is directed toward the development and improvement of advanced remote sensing systems and includes assisting in data collection, processing and analysis, and ground truth verification.

The specific focus of the work reported herein was the testing, analysis and evaluation of several types of signature extension algorithms. Four types of signature extension related techniques were examined: haze correction algorithms, data stratification procedures, training sample selection strategies for multisegment training, and crop development classifiers.

The research covered in this report was performed under NASA Contract NAS9-14988. The program was carried out in ERIM's Infrared and Optics Division which is directed by R. R. Legault, an Institute Vice-President. The Project Director was Q. A. Holmes, Head of the Information Systems and Analysis Department; and R. F. Nalepka, Head of the Multispectral Analysis Section (MAS) was the Principal Investigator. The Institute number for this report is 122700-29-T.

The authors wish to acknowledge the administrative and technical guidance provided by Mr. R. R. Legault, Dr. Q. A. Holmes, and Mr. R. F. Nalepka, and the great amount of technical help given by Mr. John Stinson and Mr. Robert Beswick. We especially wish to acknowledge Ms. Darlene Dickerson, Mrs. Elizabeth Hugg, Mrs. Jody Watters, and Mrs. Martha Warren who steadfastly typed and prepared this report and earlier materials.

CONTENTS

FIGURES

TABLES

viii

PAGE INTENTIONALLY BLANK

TABLES (Cont'd)

1
SUMMARY

Several algorithms and procedures which are candidates for inclusion in a large area crop inventory system were evaluated. These algorithms and procedures may be 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 which were 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 wheat (Kansas) and spring wheat (North Dakota) areas, and appears to offer great promise to large area crop inventory systems.

The training sample selection strategy available for testing was Procedure B [3]. Although this algorithm was not extensively and completely tested, due to the algorithm becoming available only recently, first results also show promise for future large area crop inventory systems.

In the third category, stratifications of the data, two distinct stratifications were available for testing; a stratification of the data produced by UCB [4] and one produced by JSC [5]. These stratifications yielded a significant increase in classification accuracy, however it appears that both could be considerably improved. These stratifications should be further tested using a multisegment training strategy in order to more clearly establish their performance.

In the final category, green development-trajectory classifiers, several contenders were tested. Four unitemporal green development classifiers were evaluated, with and without haze correction, the

Delta Classifier [6] was examined, and a crop development classifier was tested which was developed as a result of signature modeling efforts under this task. Results obtained using such classifiers are promising, but additional more extensive testing is recommended using a more substantial data base covering several growing seasons.

2
INTRODUCTION

Large area crop inventories using Landsat data have shown some considerable success to date. However the cost of processing is still very high, primarily because each sample segment must be individually processed by an Analyst Interpreter (AI). Signature extension, the ability to infer the signature of a crop based on signatures from selected segments and features which can be automatically extracted from the segments, would significantly lower processing cost by reducing the amount of AI-data interaction required.

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 accomplish accurate signature extension. It is the goal of this report to provide some of the necessary information about the effectiveness of these approaches in order to allow the development of a more effective large area crop inventory system.

This report covers four types of signature extension techniques and procedures:

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

It should be borne in mind that algorithms from several (or all) of the above categories will likely be incorporated into any successful signature extension system.

Section 3 of this report deals with haze correction algorithms, of which two examples have been tested: CROP-A [1] and XSTAR [2].

Section 4 reports on tests of a training sample selection strategy called Procedure B [3].

Section 5 covers evaluations of two stratifications of the data: one by UCB [4] and one by JSC [5]. Optimal stratifications of the data are also investigated.

Section 6 reports on tests of several green development and trajectory classifiers, including the Delta Classifier [6] and a green development classifier produced as a byproduct of a signature modeling effort under this task.

The final section, number 7, is a discussion of the ramifications of the results reported in the previous sections as regards the future of signature extension and large area crop inventories. In addition, recommendations for future activities are included in this section.

3
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 Turner model of the atmosphere [7,8] to measure and correct for the effects of haze.

3.1 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 an affine transformation as indicated by CROP-A. The classification results using the untransformed signatures may then be compared to the results using CROP-A transformed signatures, and some conclusions drawn.

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 which embodies our current ideas of what the next generation of large area crop inventory systems may look like. This test bench is described fully in Appendix III.

The PROCAMS test bench consists of five subsystems: preprocessing, data compression, training, signature transformation, and classification. The preprocessing subsystem screens the data for clouds, cloud shadow, water and bad data points, and then optionally applies corrective algorithms for removing haze or sun angle effects. The compression subsystem employs either the field mean approach described briefly above, or randomly samples the data when proportion estimation results are desired. The training subsystem employs ERIM's clustering algorithm [9] to obtain signatures. The signature transformation subsystem is really only for CROP-A, all other signature extension techniques tested are incorporated in either the preprocessing or the classification subsystems. The final subsystem which carries out the classification employs a sum-of-likelihoods classifier which is similar to the one employed in LACIE CAMS.

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 present 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. Figures 1, 2, and 3 show what can happen when the materials in both sites are not the same. All three figures show

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

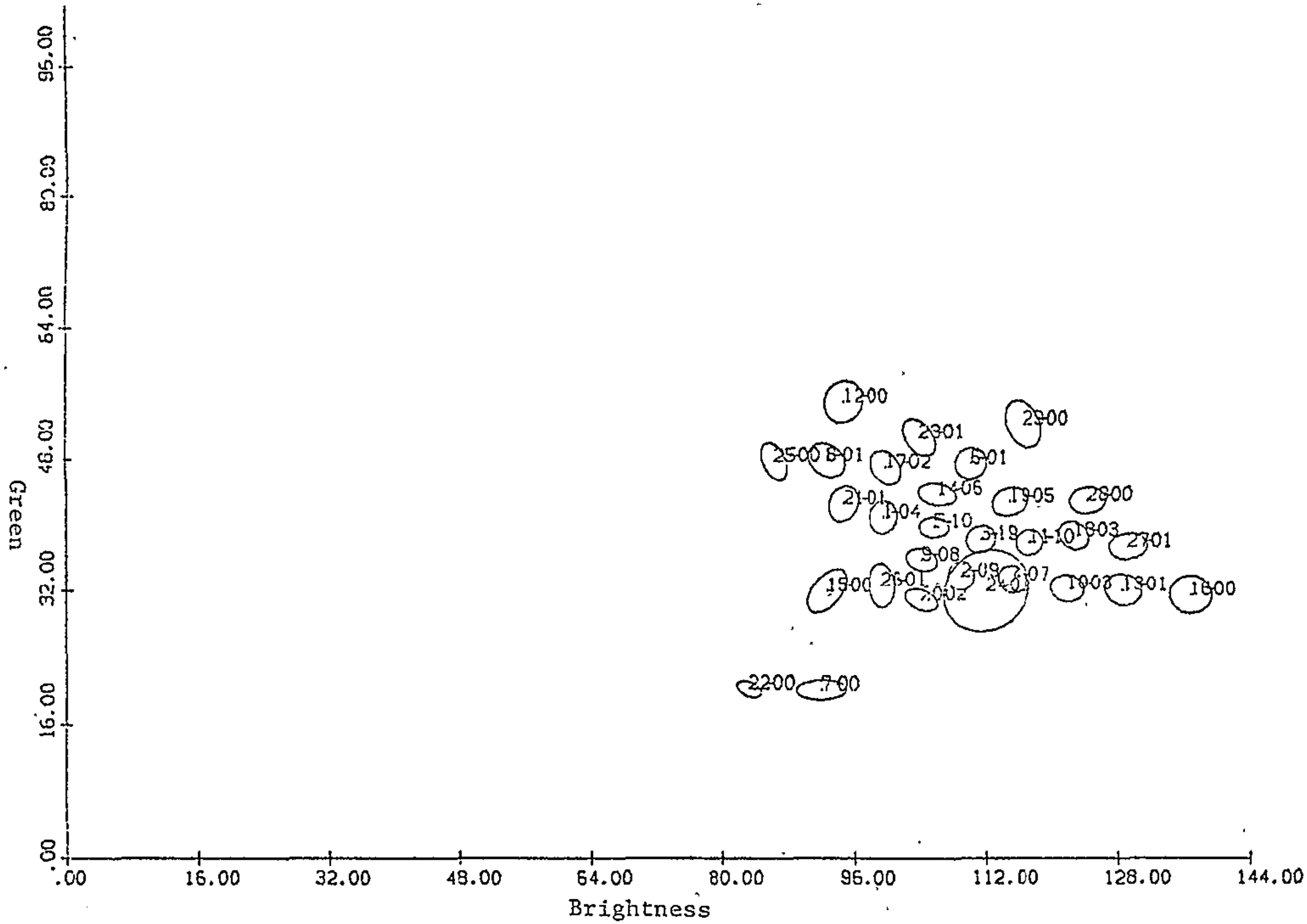


FIGURE 1. KEARNY SRS UNSUPERVISED
(27 MAY 74)

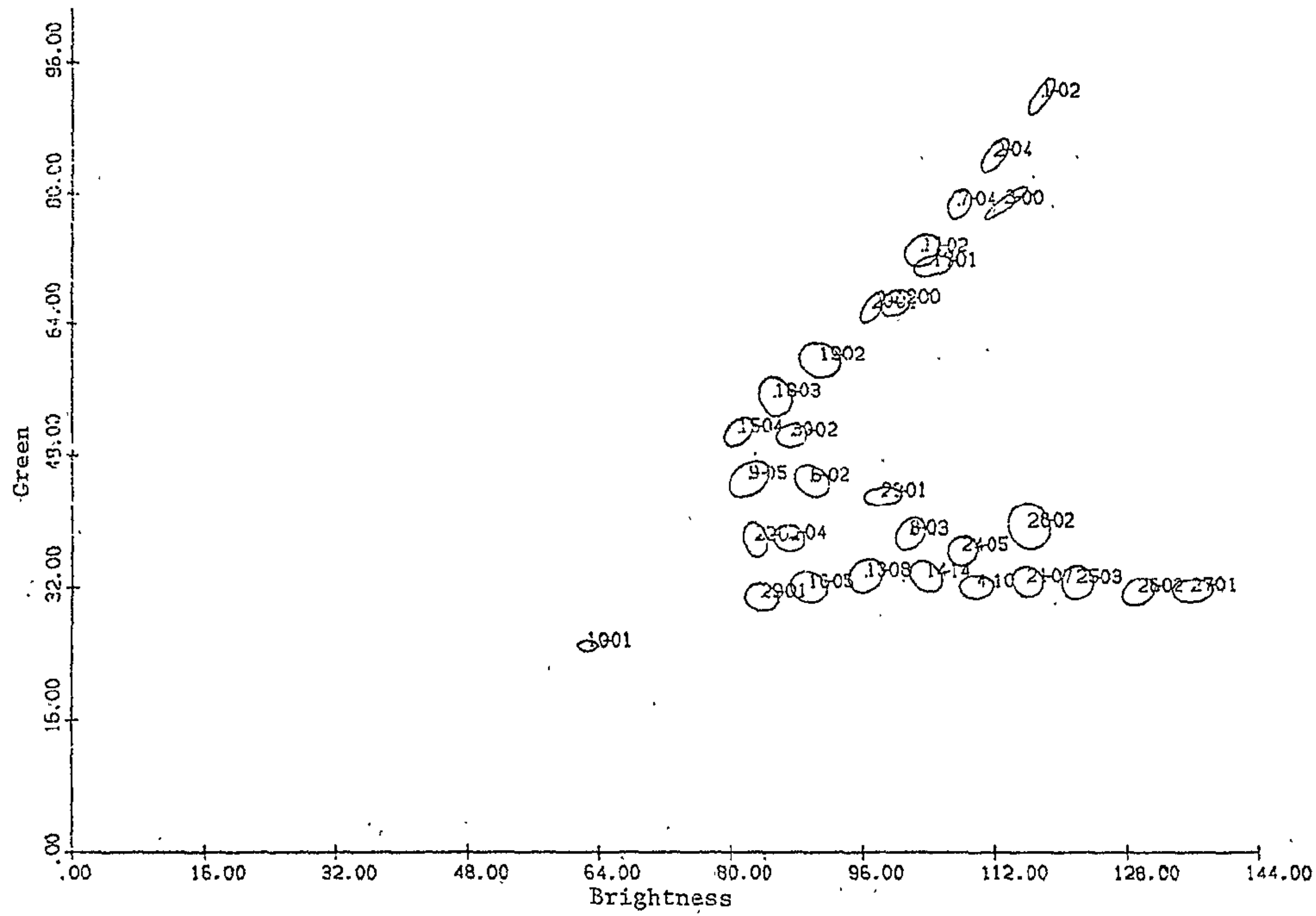


FIGURE 2. FINNEY ITS SUPERVISED
(26 MAY 74)

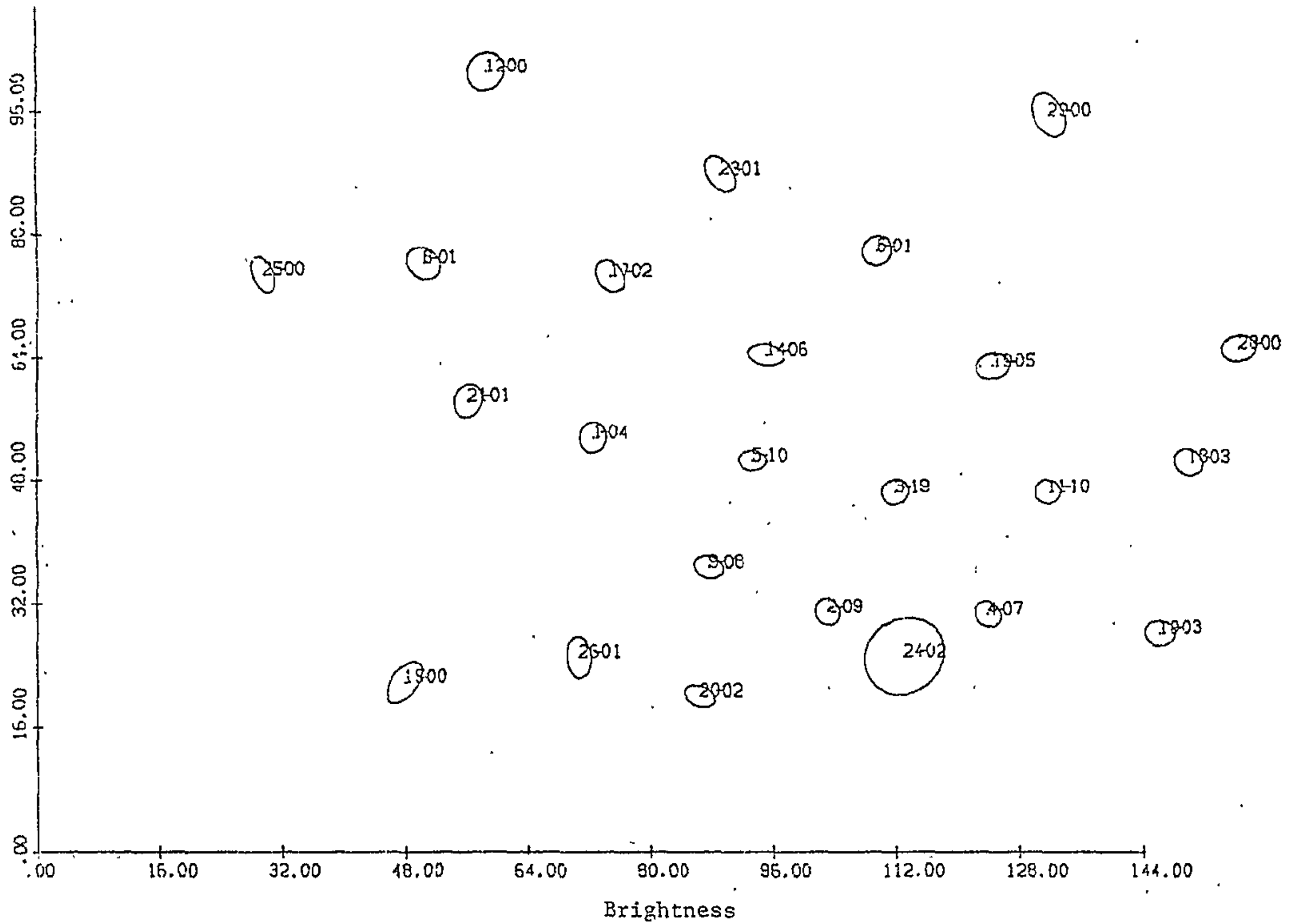


FIGURE 3, KEARNY SRS UNSUPERVISED
(27 MAY 74 - FINNEY CROP-A)

cluster plots in Tasselled Cap transformed space [10]. Figure 1 shows the clusters from the recognition site in Kearny County, Kansas. Figure 2 shows clusters from the training site in Finney County, Kansas. Note that Finney County contains quite a bit of extremely green material, the result of extensive irrigation. Kearny County contains almost none of this material. Figure 3 shows the Kearny clusters transformed by CROP-A to match the Finney cluster distribution. The result is clearly in error. In order to avoid errors of this type, cluster matching algorithms must be employed only on scenes with the same materials. Although stratification on this basis is conceptually possible, the practical problems involved have not yet been solved.

3.2 EVALUATION OF XSTAR

XSTAR is a haze correction algorithm which employs a model of haze effects derived from the ERIM atmospheric model [7]. Briefly, the XSTAR uses shifts of the data distribution in the Tasselled Cap yellow direction to measure the amount of haze present, and then corrects for the effects of this haze using its 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. In the experiments to evaluate XSTAR all possible test site-recognition site pairs were used.

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 lables. Because the accuracy of the AI crop type labels was in doubt for the North Dakota segments, the accuracy of these labels was checked for two of the sites using ground truth in the form of high altitude photography. The AI accuracy was 94% for one of the sites, and 97.5% for the other, with all errors being ones involving small numbers of pixels. The effect of these errors was minimized by a clustering algorithm which eliminates clusters with less than one pixel for each channel in the data (in this case, 16 pixels).

The PROCAMS system was used as the test bench in both experiments, with the preprocessing subsystem being updated to use the program SCREEN [11] which replaces the program BADLIN and CLOUD.

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.

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.

The physical explanation for the success of not thresholding as a signature extension technique is shown in Figure 4. According to the

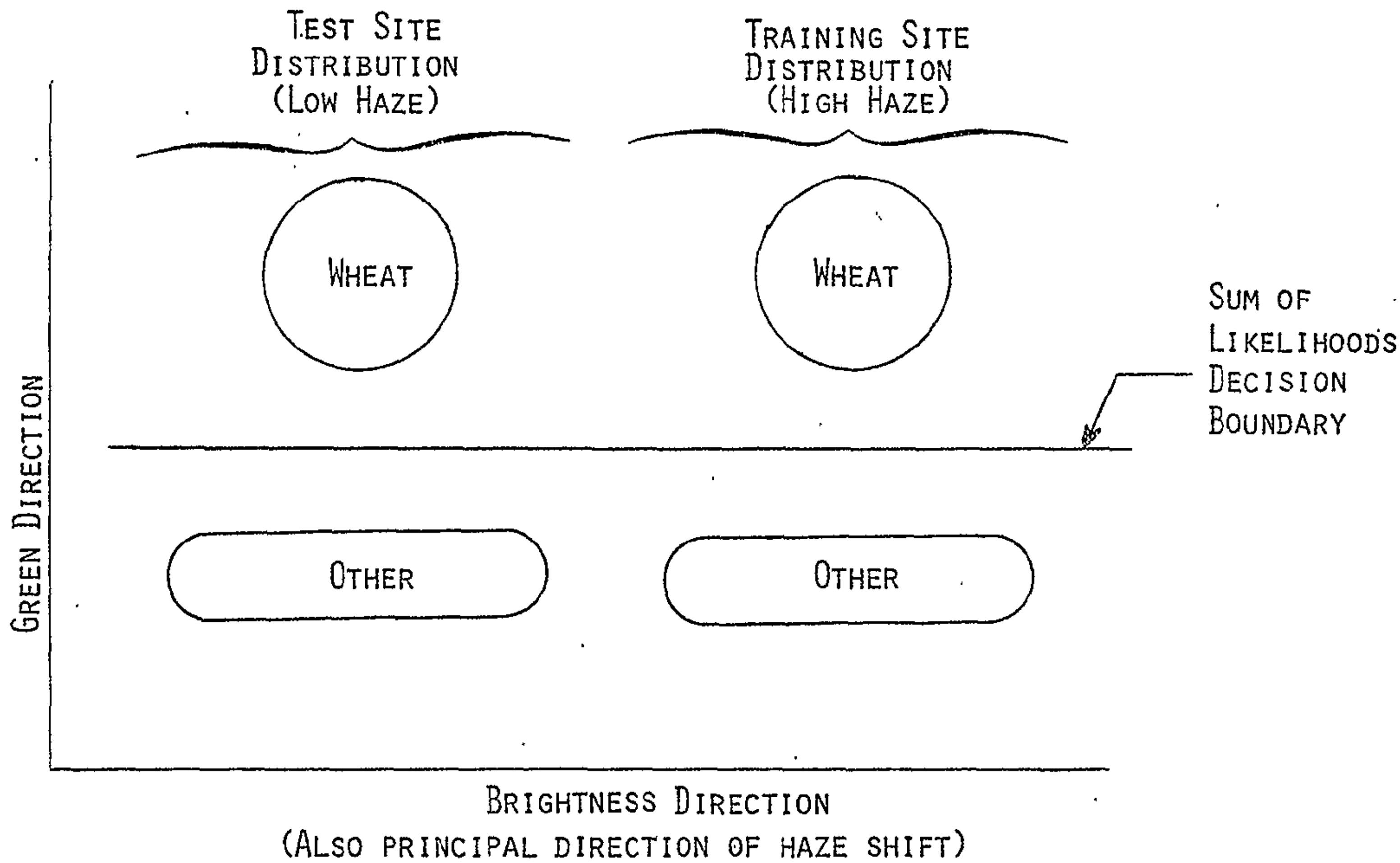


FIGURE 4. ILLUSTRATION OF THE EFFECT OF NOT THRESHOLDING IN THE PRESENCE OF HAZE

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 appears that there is an even more important effect arising from not using an alien rejection threshold in classification. Local training and classification proportion estimates both with and without a rejection threshold (one which would theoretically reject 0.1% of the data) were obtained using the 1975-76 Landsat data over 23 segments in Kansas. The results of these classifications are shown in Table 2. 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. Table 3 shows a typical result comparing proportion estimation accuracy with and without a threshold. The difference is statistically significant.

TABLE 2. EFFECTS OF THRESHOLDING ON PROPORTION ESTIMATION OVER 23 SEGMENTS IN KANSAS

	Threshold = 0.1%		No Threshold	
	Estimated Proportion of Wheat	RMS Error	Estimated Proportion of Wheat	RMS Error
Local Training	16.6%	11.79%	23.7%	10.86%
True Proportion of Wheat	23.0%		23.0%	

TABLE 3. CLASSIFICATION THRESHOLDS AND PROPORTION ESTIMATION ACCURACY

Local Training and Classification Using:	Estimated Proportion of Wheat (True = 23.7%)	RMS Error for Proportion Estimation
Rejection Threshold = 0.1%	9.4%	19.10%
No Rejection Threshold	23.6%	15.19%

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. Care must be used in applying this result, however, because the data used in these tests was previously screened to remove water, clouds, cloud shadows, and bad data.

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 4 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

TABLE 4. PERFORMANCE OF CLASSIFICATIONS 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

classification is unimproved because the classifier decision boundary is parallel to the principal direction of haze shift, as explained above.

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

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

a significant factor at the 0.1 level. Table 5 shows the effect of the number of time periods used on thresholded classifications using XSTAR corrected and uncorrected data. As more passes are added to the classification the chance of a pass 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 significantly 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.

TABLE 5. EFFECT OF NUMBER OF PASSES USED IN CLASSIFICATION
(Average over 318 extensions in North Dakota)

	Average Field Mean Classification Accuracy	
Number of Passes Used	Thresholded Classification on XSTAR Corrected Data	Thresholded Classification on Uncorrected Data
2	60.4%	56.9%
3	60.4%	54.0%
4	67.6%	55.2%

4
TRAINING SAMPLE SELECTION STRATEGIES

During this year, Task 1 of this contract developed and demonstrated a training and classification technique called Procedure B. 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, the PROCAMS test bench was modified to incorporate the training sample selection strategy 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 involves screening the segments for bad data, and applying the XSTAR correction to them. This training sample 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. This produces a set of clusters which are supposed to contain all of the variability of the original large data set after XSTAR correction. These 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. The training for the local classification used as a comparison comes from the Day 315 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.

Table 6 shows a comparison of accuracy in proportion estimation for Procedure B, Multisegment CAMS and the 75-76 LACIE procedure of local training and classification over the 28 sample segments on which Procedure B has been used. None of the differences in proportion estimation accuracy or bias are statistically significant, due to the relatively large variance in the proportion estimates.

TABLE 6. COMPARISON OF PROPORTION ESTIMATION ACCURACY USING PROCEDURE B, MULTISEGMENT CAMS AND LOCAL CLASSIFICATION ON 28 DATA SETS OVER KANSAS (6 Training Sites)

	Estimation Proportion Wheat (True = 20.4%)	RMS Error for Proportion Estimates
Procedure B	23.5%	9.93%
Multisegment CAMS	16.6%	12.67%
Local Training/ Classification	20.9%	10.69%

Table 7 shows a comparison of accuracy in proportion estimation between Multisegment CAMS and local training and classification over all 74 data sets in Kansas. Again, the differences in proportion estimation accuracy (variance) are not statistically significant, but now with the larger sample size Multisegment CAMS reveals a statistically significant bias.

TABLE 7. COMPARISON OF PROPORTION ESTIMATION ACCURACY USING MULTISEGMENT CAMS, LOCAL CLASSIFICATION ON 74 DATA SETS OVER KANSAS (6 Training Sites)

	Estimated Proportion Wheat (True = 23.7%)	RMS Error for Proportion Estimates
Multisegment CAMS	18.5%	15.05%
Local Training/ Classification	23.6%	15.19%

The results shown in Tables 6 and 7 do not include a bias correction procedure such as is being incorporated into LACIE. When considering an environment 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, where the number and relative spectral positioning of the signatures changes from segment to segment. 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, as shown in Table 8.

TABLE 8. BIAS CORRECTION RULES (Developed on the 6 Training Segments)

	Corrected Proportion Estimate of Wheat-74 Segments (True = 23.7%)	RMS Error of Corrected Proportion Estimate - 74 Segments
Multisegment CAMS	22.9	11.44%
Local Training/ Classification	20.8	14.12%

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 much cheaper 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.

5
DATA STRATIFICATION

Data stratification is the grouping of segments on the basis of similarity in segment features which affect the performance of signature extension. This idea has always been an attractive one, primarily because a good data stratification would allow a great reduction in the amount of training required to achieve a desired level of performance. 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, or how important these features might be.

For this reason the emphasis of this task in this area was two-fold. 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.

5.1 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 (see Appendix I.1 for a complete description of the data set). 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, although the test sites were from too small a region to be really definite. Figures 5 and 6 show envelopes drawn by hand around the clusters from each of several sites. Note that all of the envelopes for the ten sites are similar, which suggests that there is very little difference in haze level or soil color between them.

These two figures also illustrate an ad hoc attempt at stratification of the sites into two groups. This stratification has a statistically significant effect on classification accuracy -- but not the effect of dividing the data into two groups within which there is a high accuracy of signature extension classification. This stratification separates the sites into those with good classification results (Figure 5) and those with poor classification results (Figure 6). The sites with poor classification results -- Morton, Grant, South Stevens and North Stevens -- are all from the southwestern corner of Kansas, which suggests that some effect such as a local drought may be responsible for their poor performance.

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

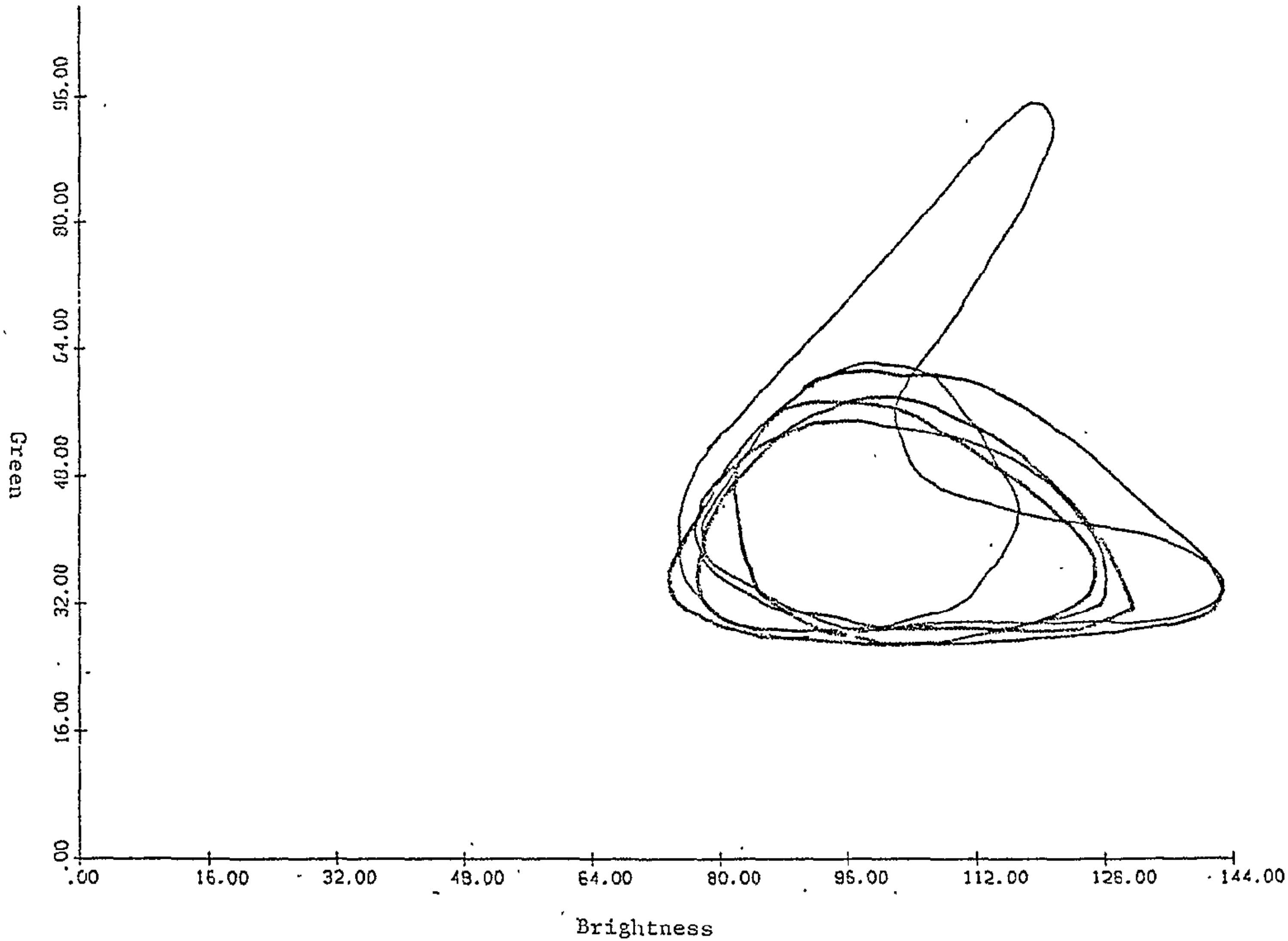


FIGURE 5. DATA ENVELOPES FOR SEVERAL SITES WITH GOOD CLASSIFICATION RESULTS

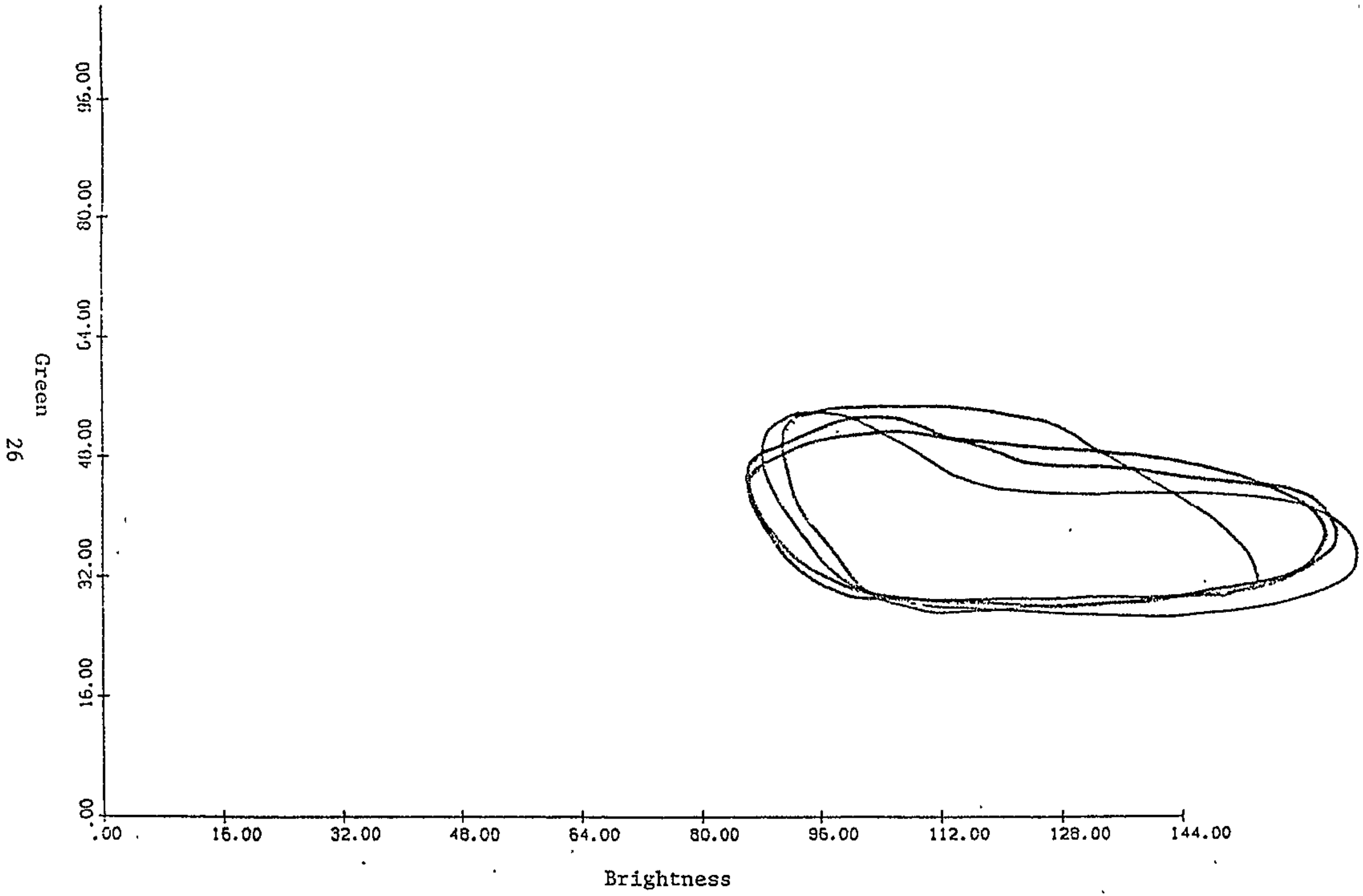


FIGURE 6. DATA ENVELOPES FOR SEVERAL SITES WITH POOR CLASSIFICATION RESULTS

were 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 UCB stratification is composed of three parts: a very fine 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 three 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.

It was found that of the 506 extensions we had full information about the UCB stratification for only 169 extensions, and only four of these were within-strata extensions. As a result, even though these four extensions had an average field mean accuracy of about 80%, as compared to 70% overall average field mean accuracy, the difference was not statistically significant.

Each of the three component parts of this stratification were then examined separately in a similar fashion. Table 9 shows the result of these examinations.

The difference between the within-strata accuracy and the across-strata accuracy was not found to be statistically significant when the land use/irrigation portion of the UCB stratification was used to stratify the data. In fact the within-strata accuracy was slightly lower than the across-strata accuracy.

Stratifying using either the degree day portion of 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.

The greatest difference between within-strata and across-strata accuracy was found when the degree day and the precipitation portions

TABLE 9. FIELD MEAN ACCURACY ANALYSIS OF PORTIONS OF THE UCB DATA STRATIFICATION

Portion(s) of the UCB Stratification Used	# Extensions Within-Strata	Average Accuracy Within-Strata (%)		# Extensions Across-Strata	Average Accuracy Across-Strata (%)	
		XSTAR	Untransformed		XSTAR	Untransformed
Land Use and Irrigation	12	67.2	68.2	157	70.4	69.4
Degree Days (10 year average)	74	72.8	72.8	95	67.3	66.6
Precipitation (10 year average)	41	82.4	80.1	128	66.2	65.9
Degree Days and Precipitation Together	26	86.5	84.5	143	66.6	66.6

of the UCB stratification were both used to stratify the data into a total of twelve groups. This difference was significant at the 0.001 level.

The conclusion reached from this analysis is that 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. However, three levels of generalization of the JSC stratification were analyzed. 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, of course, 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. It should be noted that the "sub-groups" defined in the JSC data stratification were ignored in these evaluations, because none of these subgroups had more than one of our testing segments in them.

When the suggested signature extensions were examined it was found that there were only two examples of such extensions within our data set, so 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 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. Table 10 shows the results obtained. The differences are significant at the 0.005 level.

TABLE 10. FIELD MEAN ACCURACY ANALYSIS OF JSC DATA STRATIFICATION

	XSTAR Corrected Signature Extension	Untransformed Signature Extension
Extensions Within-Strata (46 cases)	70.5%	69.0%
Extensions Across-Strata (444 cases)	62.6%	62.0%

5.2 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. This best stratification is usually different from the best stratification for any other algorithm.

For instance, CROP-A needs to have a stratification which provides it was test segment-training segment pairs with the same crops present in both segments. XSTAR needs no such restriction, but currently requires that the haze level within each segment be fairly uniform.

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 of the difference in various ancillary variables (segment features) between test segment and training segment to 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. This involves $n^2 - n$ signature extensions and classifications, where n is the number of sites in the test set.

Secondly, pair the performance figures obtained from each of these $n^2 - n$ signature extensions with a list of ancillary variables which describe the extension -- for instance, difference between the two sites in degree days, precipitation, sun angle, and so forth.

Third, use this list of ancillary variables to characterize the successful extensions -- for instance, one might perform a multiple linear regression between the ancillary variables and the signature extension performance figure.

Lastly, this characterization of the successful signature extensions can 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 11.

TABLE 11. LIST OF ANCILLARY VARIABLES

I. GENERAL:

Degree Days (10 Year Average)	Longitude
Land Use (% Agriculture)	Elevation
Precipitation (10 Year Average)	
Latitude	

II. PASS SPECIFIC (Calculated for Each Pass):

Sun Angle

View Angle

Julian Date

Crop Calendar (Robertson Scale)

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 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 using a stepwise linear regression technique which adds variables one at a time to the regression equation, starting with the most significant and continuing until none of the remaining variables have an effect on the regression equation which is significant at the 0.05 level. The results of this stepwise linear regression are given in Table 12 below.

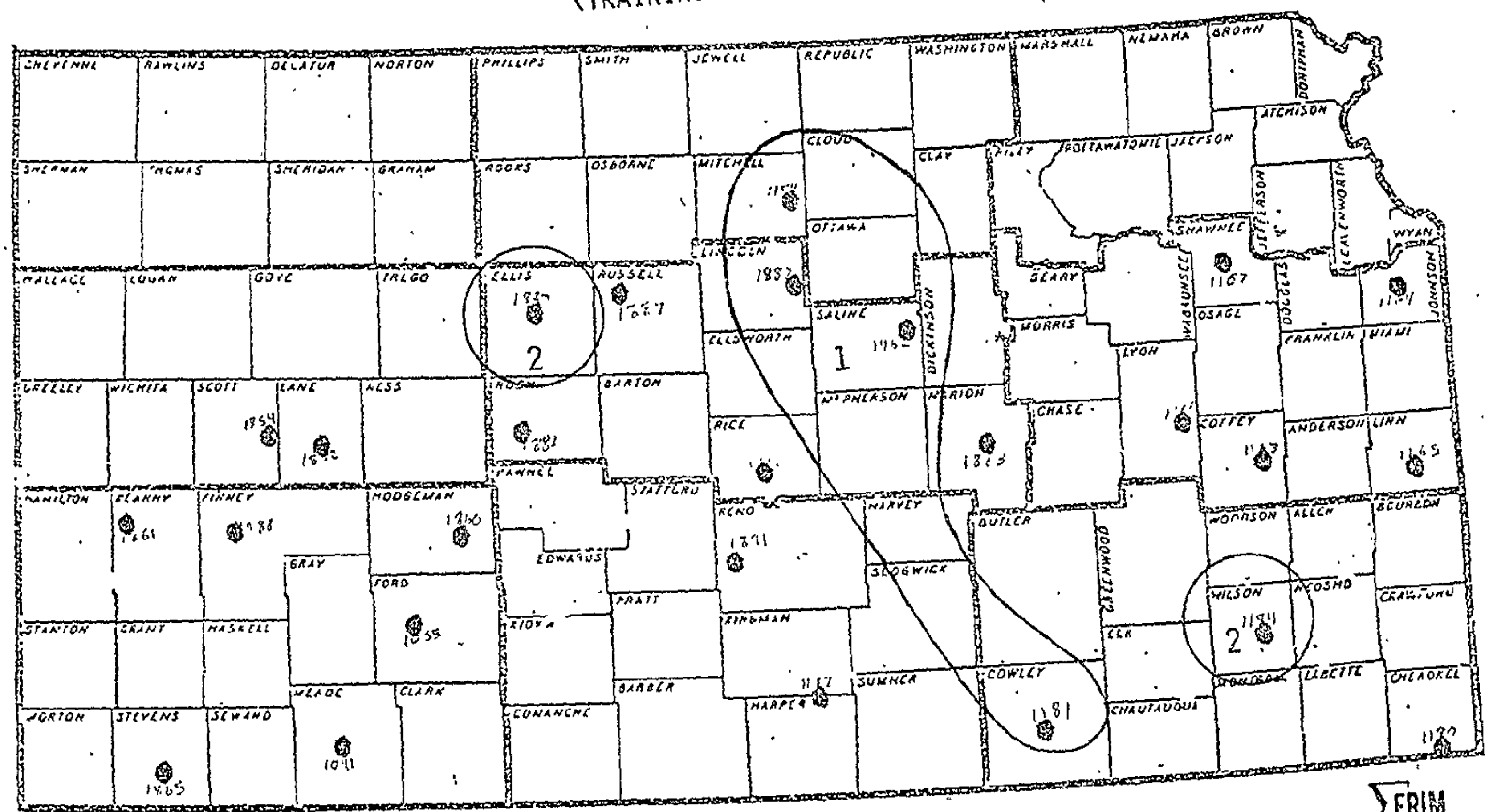
TABLE 12. RESULTS OF STEPWISE LINEAR REGRESSION OF UNTRANSFORMED SIGNATURE EXTENSION RESULTS VS ANCILLARY INFORMATION

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. Figure 7 shows the stratification obtained when the desired training gain is 1.2 (i.e., four out of the 23 sites are classified by signature extension rather than local training, a savings of 20% in training cost). Figure 8 shows average field mean

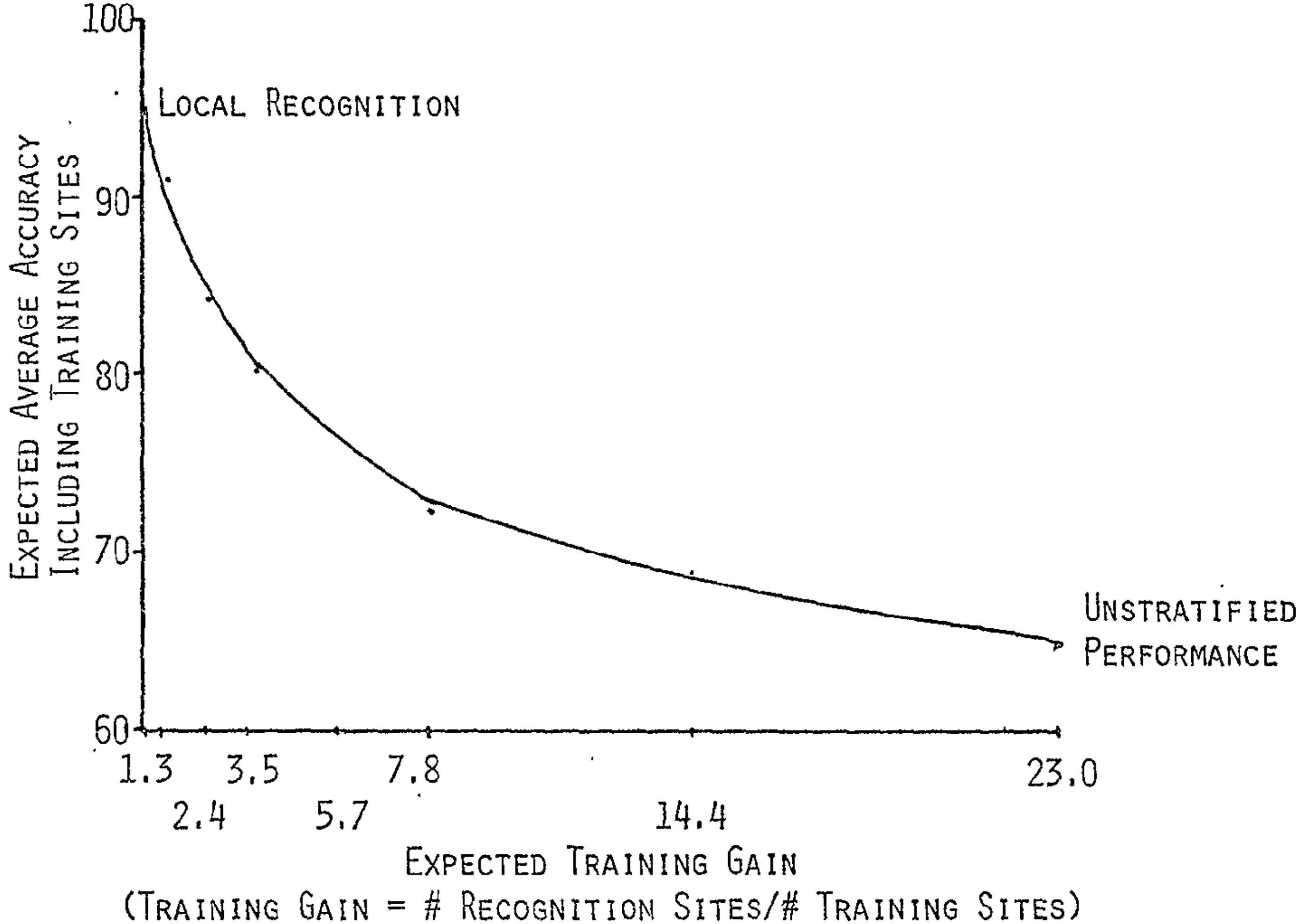
FIGURE 7
UNTRANSFORMED STRATA
KANSAS

(TRAINING GAIN = 1.2)



ERIM

FIGURE 8
UNTRANSFORMED PERFORMANCE ON
COMPRESSED DATA



classification accuracy over the 23 sites as a function of training gain. Table 13 shows pixel-by-pixel proportion estimation results using the 1.2 training gain stratification shown in Figure 7. The proportion estimation bias in this 23 segment sample is not statistically significant.

TABLE 13. UNTRANSFORMED SIGNATURE EXTENSION PROPORTION ESTIMATION RESULTS OVER 23 SITES IN KANSAS

	Estimated Proportion of Wheat	RMS Errors	Standard Deviation of Error
Local Training	23.7%	10.86%	11.12%
Untransformed Signature Extension (Training Gain of 1.2)	25.1%	11.40%	11.52%
True Proportions of Wheat	23.0%		

This experiment was then repeated using XSTAR, in place of untransformed signature extension. Table 14 shows the results of the stepwise linear regression of XSTAR's results versus the ancillary information.

TABLE 14. 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 equation was used to define stratification of the data as was done with the regression equation obtained for the untransformed signature extension case. Figure 9 shows the stratification obtained when the desired training gain is 1.2. Figure 10 shows the relationship of average field mean classification accuracy over the 23 sites as a function of training gain. Table 15 shows pixel-by-pixel proportion estimation results for XSTAR corrected signature extension using the 1.2 training gain stratification shown in Figure 9. Again, this proportion estimation result does not have a statistically significant bias.

TABLE 15. XSTAR CORRECTED SIGNATURE EXTENSION PROPORTION ESTIMATION RESULTS OVER 23 SITES IN KANSAS

	Estimated Proportion of Wheat	RMS Error	Standard Deviation of Error
Local Training	23.7%	10.86%	11.12%
XSTAR Corrected Signature Extension (Training Gain of 1.2)	23.8%	13.19%	13.46%
True Proportions of Wheat	23.0%		

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 (less than 5%) of the total variance in field mean accuracy as to be useless in determining a stratification of the data. The conclusion to be drawn from this 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.

FIGURE 9. XSTAR STRATA
KANSAS
(TRAINING GAIN = 1.2)

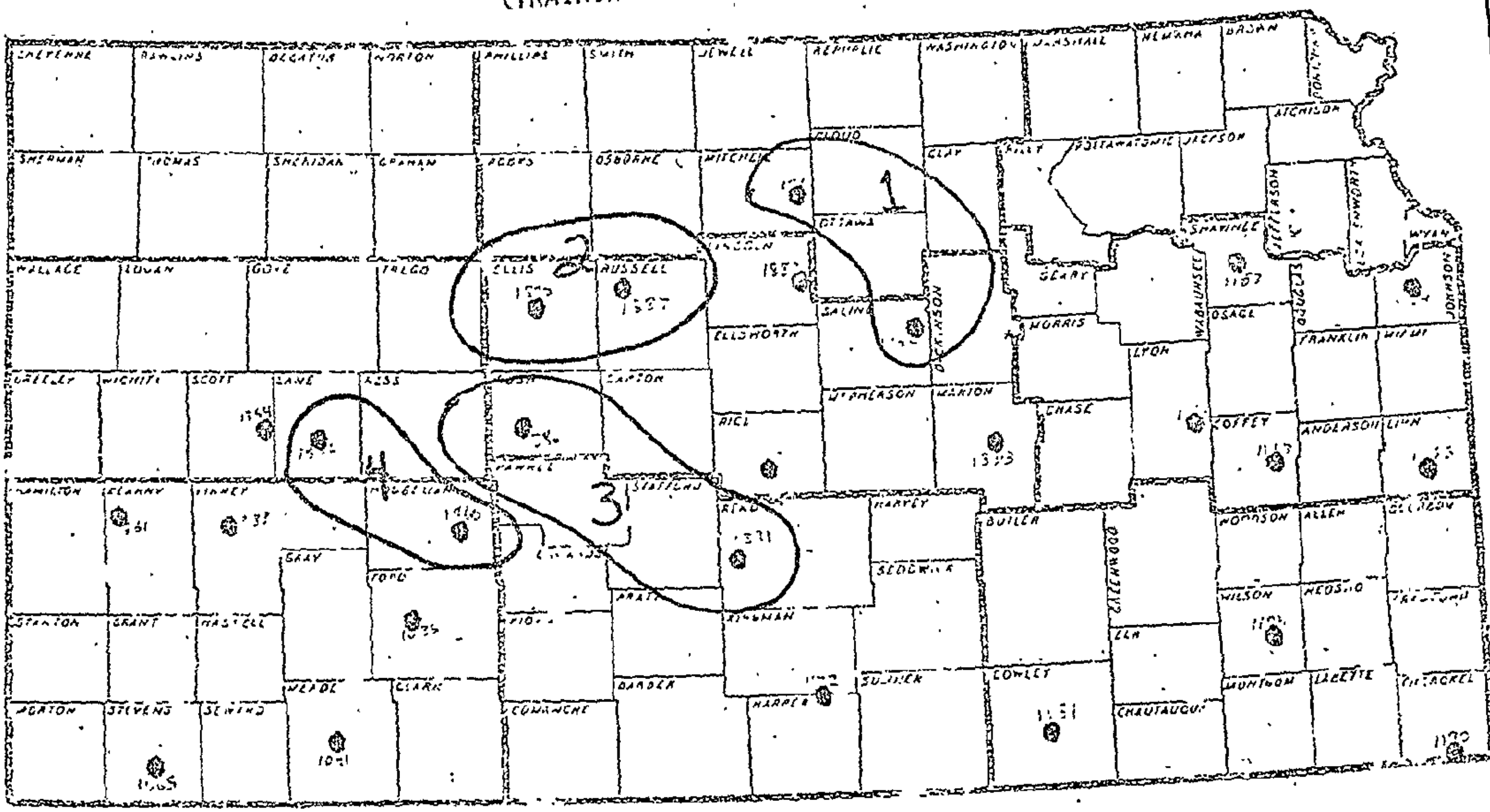
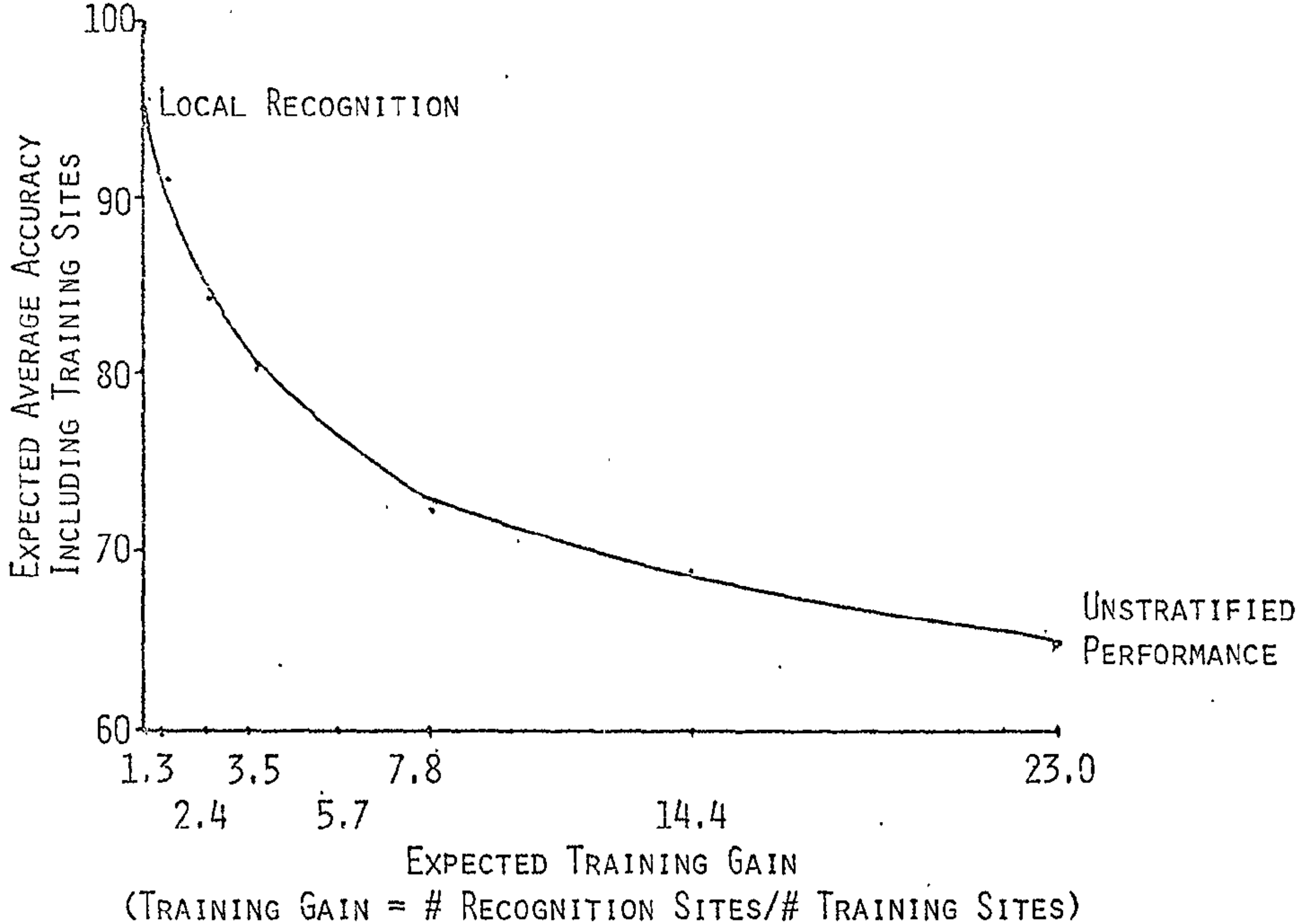


FIGURE 10
XSTAR PERFORMANCE ON
COMPRESSED DATA



5.3 THE UTILITY OF STRATIFICATIONS OF THE DATA

Section 5.1 showed 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 5.2 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.

6
GREEN INDICATOR AND CROP DEVELOPMENT CLASSIFIERS

Any classification technique which employs a decision rule which has been trained in one place or time and can be used to classify in a different place or time is accomplishing signature extension. The general approach taken by these 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. Thus during the relevant time period, any pixel with a green number within this range is to be called wheat. 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, so that any pixel whose Landsat signal values are sufficiently close to what the model predicts is called wheat. The Delta classifier is an example of such a classifier.

6.1 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 16.

TABLE 16 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*-0.28972 + CH2*-0.56199 + CH3*0.599153 + CH4*0.49070

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. The field mean accuracies obtained in this fashion are an upper bound on the performance of these green development indicators as a classification procedure. Table 17 summarizes these results for Biowindow 1, and Table 18 summarizes the results for Biowindow 2.

TABLE 17 PERFORMANCE OF GREEN DEVELOPMENT INDICATORS
BIOWINDOW 1

Indicator	Average Field Mean Accuracy: Untransformed Data	Average Field Mean Accuracy: XSTAR Corrected Data
G	70	72
TVI	77	76
Ratio 7/5	76	75
Tasselled Cap Green	76	72

TABLE 18 PERFORMANCE OF GREEN DEVELOPMENT INDICATORS BIOWINDOW 2

Indicator	Average Field Mean Accuracy: Untransformed Data	Average Field Mean Accuracy: XSTAR Corrected Data
G	82.4	83.9
TVI	81.2	81.3
Ratio 7/5	81.2	82.2
Tasselled Cap Green	80.3	79.9

These field mean classification accuracies seemed to indicate that the green development indicators hold considerable promise as proportion estimators. Results of pixel-by-pixel proportion estimation over the 23 segments using the G indicator in Biowindow 2, and the TVI indicator in Biowindow 1 are given in Table 19.

TABLE 19 PROPORTION ESTIMATION RESULTS OF GREEN DEVELOPMENT INDICATORS OVER 23 SITES IN KANSAS

Indicator	Estimated Proportion of Wheat
TVI, Biowindow 1	39.8%
G, Biowindow 2	33.9%
True Proportion of Wheat	23.0%

As can be seen from table 19 the green indicators, even when optimally trained on field means, displayed a very large bias. Further, the variance of the error in proportion estimation for these indicators was very large. This seemed to indicate that a more sophisticated approach was required than the "if its 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. It requires good data from three different biowindows in order to make a discrimination between wheat and non-wheat. Accordingly, we used the Delta Classifier to classify each of the 23 test sites, and obtained an average field mean classification accuracy of 71%. It should be pointed out, however, that while one pass was available in each of the four biowindows, these passes were not selected with an eye to optimizing the Delta Classifier's performance. This not-terribly-high field mean accuracy led us to investigate the reasons for these problems. Comparing the field mean classification accuracy of the Delta Classifier to ancillary information via a regression, it was discovered that the following four factors significantly affected the performance of the Delta Classifier in Kansas:

Degree Days (10 Year Average)
Precipitation (10 Year Average)
Longitude
XSTAR's Haze Coefficient Gamma,
in Biowindow 3

It was concluded that in order to be successful, such a classifier must include ancillary information (such as a crop calendar) in the decision rule, so that the stage of crop development can be more accurately known.

6.2 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? And second, is Landsat data two dimensional?, (i.e., do the first two channels of the Tasselled Cap transform, brightness and greenstuff, contain all the discriminability information?)

To investigate each of these ideas, 322 signature extensions were carried out using 1973-74 data over 12 Kansas sites (see Appendix 1.2 for a more complete description of the data set) this included all possible extensions with matching biophases. The results of the investigation into these two questions are briefly summarized below:

1. Best Dates for Classification. The data set contained passes from five dates: 20 October, 20 April, 9 May, 27 May, and 12 June. All combinations of these dates were tested for performance both locally and in signature extension. The best single date was found to be the 20 April date, with the average accuracies of the 9 May and 27 May dates trailing by 5 and 10% respectively. There was a tie for the best combination of passes: any combination of passes containing both the 20 October and 20 April dates performed about equally, and no other combination of two passes approached the accuracy of this October-April combination.

2. Information Distribution in the Tasselled Cap Transform. Each of the 322 extensions were also performed using only the first two components of the Tasselled Cap Transform -- Brightness and Greenstuff. It was found that average accuracy using only these two channels was about 3% less than the accuracy using all four Landsat channels; for multitemporal extensions, average accuracy decreased by about 3% for each time period added beyond the first time period as compared to untransformed accuracy. This trend did not hold whenever the April pass was one of the passes used in the extension; in these cases there was no significant decrease in accuracy. It is hypothesized that most of the information needed to

distinguish wheat from non-wheat can be obtained from the green development seen by Landsat at any fixed point in a crop calendar, and that the green development information is contained within the first two components of the Tasselled Cap transform.

The results of this investigation guided is in the next step of the investigation, which was the development of a fairly sophisticated model of wheat development as seen by Landsat, as a function of both time and ancillary information relating to crop development, haze level, illumination of the site and so forth. 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 throughout the second biowindow.

The green development model, which has a correlation with observed signals of 0.907 and a residual error of three counts, incorporates the following ancillary information (listed in order of importance):

- Number of days into 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, which has a correlation with observed signal values of 0.80 and a residual error of 6.7 counts, 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 incorporated into 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 only 75.4%, showing that inclusion of the ancillary information into the decision rule via the two models had significantly improved classification accuracy.

Such models (or classifiers) are useful only if they are stable in the sense that if they are constructed or trained on only a small portion of the data they still yield approximately the same results as if all the data were used. If they are stable in this sense, then they derive their accuracy from underlying physical processes and may well be applicable (with perhaps small changes) to other places and other years. At the worst, if they are stable then they can be accurately trained anew each year using only a small number of sample segments, and at a correspondingly small cost.

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 20 shows how the accuracies of several other classifiers compare to this accuracy.

TABLE 20. 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%

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.

7
CONCLUSIONS AND RECOMMENDATIONS

The overall conclusion of this report is that 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 CROP-A made serious errors which actually degraded classification performance (as compared to simply applying signatures from one segment directly to a different segment, called untransformed signature extension) whenever the types of materials found in the training and test sites were substantially different. For this reason CROP-A was deemed to be 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 the XSTAR haze correction was substantially the same regardless of haze level or differences between the test and training sites.

An interesting discovery 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. Two factors were responsible for this unexpected result. First, 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 and that the only class of interest is wheat. The second factor in this result is noise introduced by errors in the testing procedure which may have had the effect of degrading the classification accuracy of XSTAR corrected signature extension. Two sources of noise were discovered in the testing procedure. The major source of noise came from the AI field designations and crop labels that were used in computing performance figures. Later analysis disclosed that the AI had approximately a 7% crop labeling error rate in Kansas and a 14% crop labeling error rate in North Dakota. Another source of noise was programming error which resulted in truncating the haze diagnostic vector to integer values. This truncation is not considered to be a serious source of noise.

The training sample selection strategy available for testing at this time was 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 over 38 Kansas blind and ITS sites were extremely encouraging, and in fact were not statistically different from multitemporal local training and classification proportion estimation results.

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 accomplished 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 collected over 23 blind sites in Kansas. 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, such as the Delta Classifier. Several such classification schemes were examined using the 74 multitemporal data sets collected over 38 Kansas blind and ITS sites. 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 made a first step towards solving this problem.

The 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.

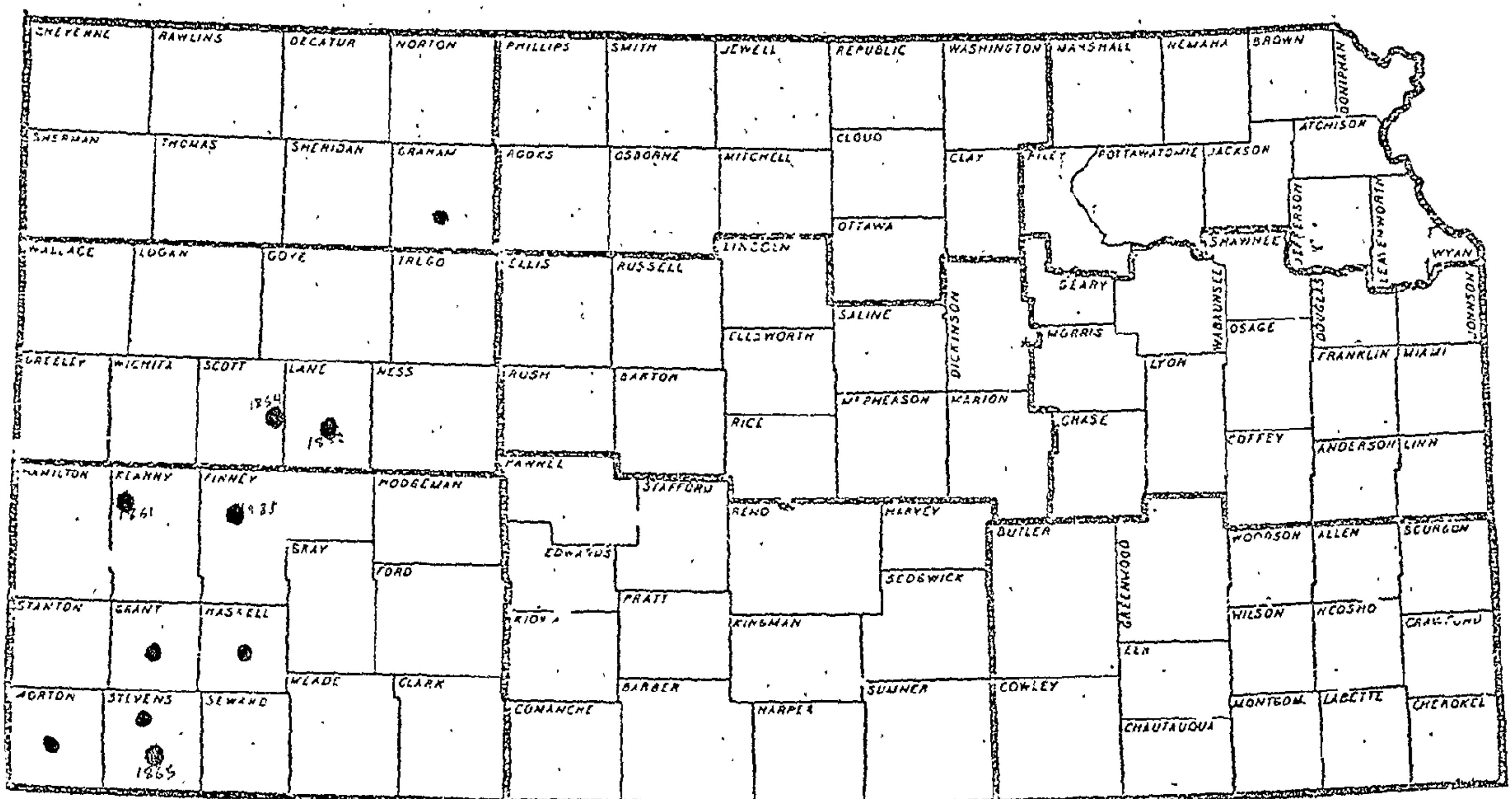
APPENDIX I
DATA PREPARATION

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.



ERIM

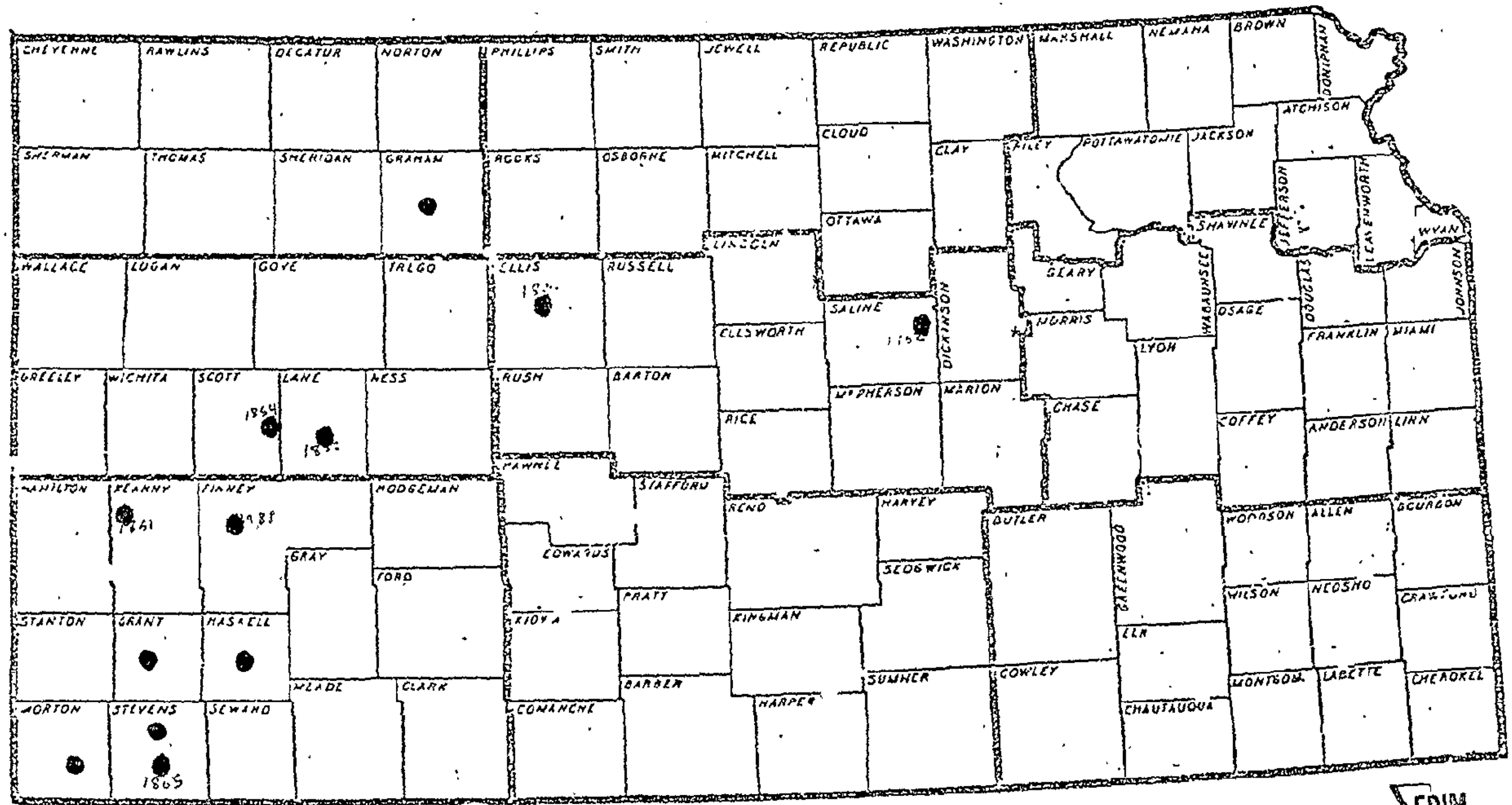
FIGURE I-1. TEST SITES IN KANSAS, 73-74 DATA

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.



ERIM

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

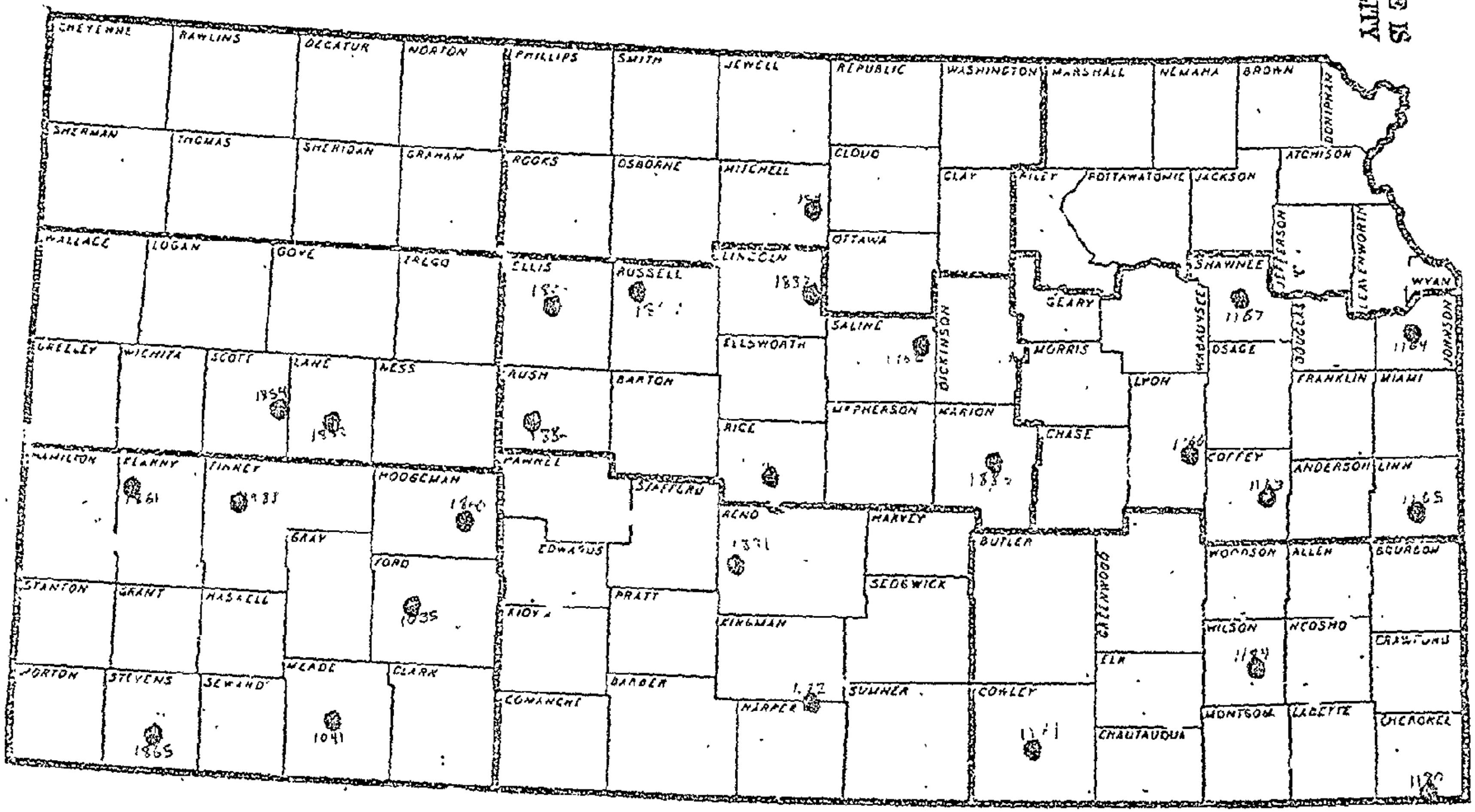
The fourth period data base consisted primarily of 74 data sets over 38 sample segments in Kansas (35 blind sites and 3 intensive test sites) and 18 data sets over 18 sample segments in North Dakota. Each of the data sets consists of four acquisitions of 75-76 LACIE sample segment data, one from each 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. Limited comparisons of the Kansas field designations with actual ground truth showed no discrepancies. North Dakota (spring wheat) field designations were then compared with ground truth over two of the sample segments. The analyst interpreters were found to have accuracies of 94% and 97.5% over these two sample segments.

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.

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



ERIM

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°

<u>Biowindow 1</u>

Julian Date 291-90 Crop Calendar 0 - 3.3 Sun Angle 46° - 68° Gamma -.08 - .23

<u>Biowindow 2</u>

Julian Date 90-138 Crop Calendar 3.0 - 3.6 Sun Angle 35° - 46° Gamma -.5 - .19

<u>Biowindow 3</u>

Julian Date 135-163 Crop Calendar 3.3 - 4.8 Sun Angle 31° - 36° Gamma -.22 - .19

<u>Biowindow 4</u>

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

TIME PERIOD 1

JULIAN DATE 127-131 SUN ANGLE 33^0 - 39^0 GAMMA -.11 - .12

TIME PERIOD 2

JULIAN DATE 144-150 SUN ANGLE 33^0 - 39^0 GAMMA -.5 - .1

TIME PERIOD 3

JULIAN DATE 164-186 SUN ANGLE 33^0 - 39^0 GAMMA -.41 - .14

TIME PERIOD 4

JULIAN DATE 198-204 SUN ANGLE 33^0 - 39^0 GAMMA -.01 - .18

APPENDIX II
CLASSIFICATION ACCURACY USING COMPRESSED DATA
(John Stinson)

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	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

Site From	Site To	Time Period	Accuracy (%) Not Compressed	Accuracy (%) 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

APPENDIX III
DESCRIPTION OF THE 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 [11]. The final (aid optional) stage of the preprocessing is haze correction.

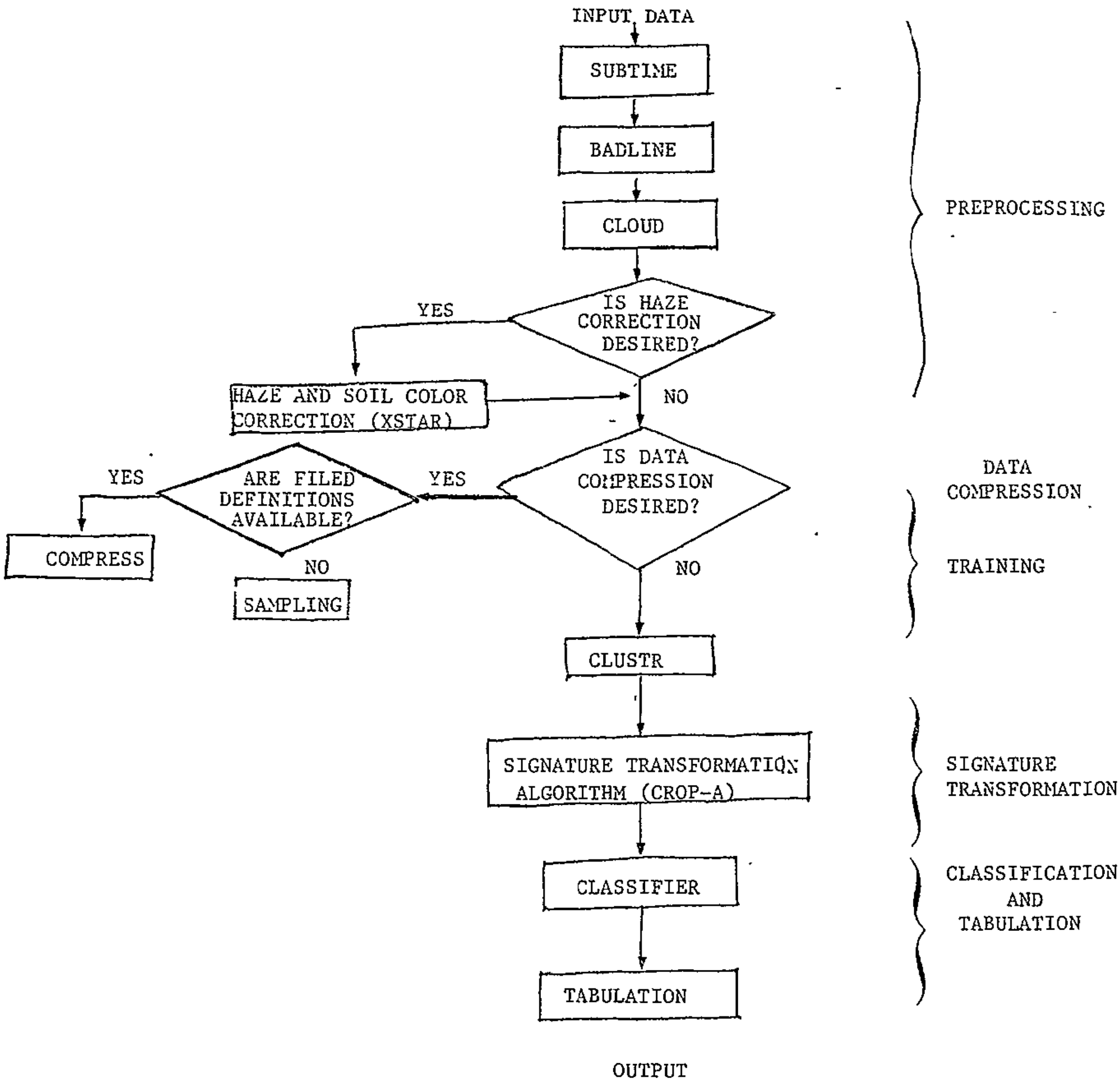


FIGURE III-1. FLOW CHART OF THE PROCAMS SYSTEM

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 [9].

The fourth subsystem in PROCAMS (signature transformation) is signature extension, a rôle 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 [12].

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 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 No. 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", Report No. 109600-15-F, Environmental Research Institute of Michigan, May 1975.

8. Malila, W. A., R. B. Crane, and R. E. Turner, "Information Extraction Techniques for Multispectral Scanner Data", Report No. WRL 31650-74-T, Willow Run Laboratories, Ann Arbor, Michigan, June, 1972, Appendix I.

9. Horwitz, H. M., J. T. Lewis, and A. P. Pentland, "Estimating Proportions of Objects From Multispectral Scanner Data", Report No. 109600-13-F, Environmental Research Institute of Michigan, Ann Arbor, Michigan, May 1975.

REFERENCES (Cont'd)

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

11. Lambeck, P. F., Internal Memorandum IS-PFL-1272 to R. F. Nalepka, "Implementation of the XSTAR Haze Correction Algorithm and Associated Preprocessing Steps for Landsat Data", Environmental Research Institute of Michigan, Ann Arbor, Michigan, 18 March 1977 (NASA/JSC Memorandum SF3/77-332).

12. Richardson, W., and A. P. Pentland, "Evaluation of Algorithms for Estimating Wheat Acreage From Multispectral Scanner Data", Report No. 109600-69-F, Environmental Research Institute of Michigan, Ann Arbor, Michigan, May 1976.

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 Water Resources Division 500 Zack Street Tampa, Florida 33602 ATTN: Mr. A. E. Coker	(1)
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 Naval Research Laboratory Washington, D.C. 20390 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)