E82-10206
FC-L1-04096
JSC-17138
CR-167485

AgRISTARS

A Joint Program for Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing

Foreign Commodity Production Forecasting

October 1981

1980 U.S. CORN AND SOYBEANS EXPLORATORY EXPERIMENT: FINAL REPORT

J. T. Malin and J. G. Carnes

(E82-10206) AGRISTARS: FOREIGN COMMODITY PRODUCTION FORECASTING. THE 1980 US CORN AND SOYBEANS EXPLORATORY EXPERIMENT Final Report (Lockheed Engineering and Management) 183 p HC A09/MF A01 CSCL 02C G3/43
N82-23579
Unclas
00206

Lockheed Engineering and Management Services Company, Inc.
1830 NASA Road 1, Houston, Texas 77058

Lyndon B. Johnson Space Center
Houston, Texas 77058

1. Report No.
JSC-17138, FC-L1-04096

2. Government Accession No.

3. Recipient's Catalog No

4. Title and Subtitle
1980 U.S. Corn and Soybeans Exploratory Experiment: Final Report

5. Report Date
October 1981

6 Performing Organization Code

7. Author(s)
J. T. Malin and J. G. Carnes

8. Performing Organization Report No
LEMSCO-16573

9. Performing Organization Name and Address
Lockheed Engineering and Management Services Company, Inc.
1830 NASA Road 1
Houston, Texas 77058

10. Work Unit No.

11. Contract or Grant No.
NAS 9-15800

12. Sponsoring Agency Name and Address
National Aeronautics and Space Administration
Lyndon B. Johnson Space Center
Houston, Texas 77058 Technical Monitor: J. L. Dragg

13. Type of Report and Period Covered
Technical Report

14. Sponsoring Agency Code

15. Supplementary Notes
The Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing is a joint program of the U.S. Department of Agriculture, the National Aeronautics and Space Administration, the National Oceanic and Atmospheric Administration (U.S. Department of Commerce), the Agency for International Development (U.S. Department of State), and the U.S. Department of the Interior.

16. Abstract
The U.S. Corn and Soybeans Exploratory Experiment, which was completed during FY80, consisted of evaluations of two technology components of a production forecasting system: classification procedures (crop labeling and proportion estimation at the level of a sampling unit) and sampling and aggregation procedures. The results from the labeling evaluations indicate that the corn and soybeans labeling procedure works very well in the U.S. Corn Belt with full-season (after tasseling) Landsat data. The procedure should be readily adaptable to corn and soybeans labeling required for subsequent exploratory experiments or pilot tests. The machine classification procedures evaluated in this experiment were not effective in improving the proportion estimates. The corn proportions produced by the machine procedures had a large bias when the "bias" correction was not performed. This bias was caused by the manner in which the machine procedures handled spectrally impure pixels. Alternatives to the machine processing techniques used in this experiment should be investigated to see if more effective techniques can be found. The simulation test indicated that the weighted aggregation procedure performed quite well. Although further work can be done to improve both the simulation tests and the aggregation procedure, the results of this test show that the procedure should serve as a useful baseline procedure in future exploratory experiments and pilot tests.

17. Key Words (Suggested by Author(s))
Aggregation
AgRISTARS
Analyst labeling
Corn and soybeans
FCPF
Proportion estimation
Remote sensing

18. Distribution Statement

19. Security Classif. (of this report)
Unclassified

20. Security Classif. (of this page)
Unclassified

21. No. of Pages
139

22. Price*

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

PRECEDING PAGE BLANK NOT FILMED

FC-L1-04096
JSC-17138

1980 U.S. CORN AND SOYBEANS EXPLORATORY EXPERIMENT:

FINAL REPORT

Job Order 72-415

This report describes the U.S. Corn/Soybean Exploratory Experiment of the Foreign Commodity Production Forecasting project of the AgRISTARS program.

PREPARED BY

J. T. Malin and J. G. Carnes

APPROVED BY

LOCKHEED-EMSCO

B. L. Carroll, Manager
Crop Applications Department

NASA

J. L. Dragg, Project Manager
Foreign Commodity Production
Forecasting Project (FY80)

LOCKHEED ENGINEERING AND MANAGEMENT SERVICES COMPANY, INC.

Under Contract NAS 9-15800

For

Earth Observations Division

Space and Life Sciences Directorate

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

October 1981

LEMSCO-16573

PREFACE

The Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing is a multiyear program of research, development, evaluation, and application of aerospace remote sensing for agricultural resources, which began in fiscal year 1980. This program is a cooperative effort of the U.S. Department of Agriculture, the National Aeronautics and Space Administration, the National Oceanic and Atmospheric Administration (U.S. Department of Commerce), the Agency for International Development (U.S. Department of State), and the U.S. Department of the Interior.

The work which is the subject of this document was performed by the Earth Resources Applications Division, Space and Life Sciences Directorate, Lyndon B. Johnson Space Center, National Aeronautics and Space Administration and Lockheed Engineering and Management Services Company, Inc. The tasks performed by Lockheed Engineering and Management Services Company, Inc., were accomplished under Contract NAS 9-15800.

The following personnel assisted in compiling this report, in carrying out the tests reported here, or in providing technical inputs and consultation. These include H. O. Hartley, T. H. Hughes, and R. L. Sielken of Texas A&M University; Project Manager J. L. Dragg (FY 1980), Experiments Manager R. O. Hill, R. M. Bizzell, A. H. Feiveson, C. R. Hallum, and L. C. Wade of the National Aeronautics and Space Administration, Lyndon B. Johnson Space Center; and L. M. Abotteen, J. E. Baird, C. L. Dailey, S. A. Davidson, and J. H. Smith of Lockheed Engineering and Management Services Company, Inc.

PRECEDING PAGE BLANK NOT FILMED

CONTENTS

TABLES

FIGURES

1. INTRODUCTION

During the first year (fiscal year 1980) of the Foreign Commodity Production Forecasting (FCPF) project of the Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing (AgRISTARS) program, two exploratory experiments were performed to develop and evaluate techniques. This report describes the U.S. Corn and Soybeans Exploratory Experiment. The other experiment, the U.S./Canada Wheat and Barley Exploratory Experiment, is described in the 1980 U.S. Wheat and Barley Exploratory Experiment Final Report (ref. 1).

The overall purpose of the FCPF project is to develop and test procedures for using aerospace remote sensing technology to provide more objective, timely, and reliable crop production forecasting in foreign areas. To develop technology for use in foreign areas, the FCPF project builds upon existing remote sensing technology and extends this technology to additional crops and regions (ref. 2).

2. SUMMARY

2.1 PURPOSE AND SCOPE

The overall purpose of the U.S. Corn and Soybeans Exploratory Experiment was to develop objective, timely, and reliable technology for production forecasting of corn and soybeans, and to conduct exploratory testing of this technology using data from the U.S. Corn Belt. The technology was made up of two sets of procedures. One set, the classification procedures, was designed to separate corn and soybeans and provide proportion estimates at the level of a sampling unit (5- by 6-nautical-mile segment). The other set was designed to optimally allocate samples simultaneously for multiple crops and to make regional-level crop area and production estimates that make optimum use of available segment proportion estimates. These sets of procedures were to be evaluated for use as components of a baseline technology for adaptation to corn and soybeans production forecasting in foreign regions. The experiment plan for these evaluations was developed in 1979 during the transition year before AgRISTARS (ref. 3).

2.2 TECHNOLOGY DESCRIPTIONS

2.2.1 CLASSIFICATION PROCEDURES

An analyst/computer-based technology has been developed for estimating the proportion of small grains and wheat area in 5- by 6-nautical-mile sample segments. The U.S. Corn and Soybeans Exploratory Experiment was the first attempt to extend segment-level proportion estimation techniques to other crops. The segment-level proportion estimates were obtained by labeling selected pixels from the segment as training for a maximum likelihood classifier. In one version of the procedure, the results from the classification were corrected for bias by using an independent set of labeled pixels. Pixel labeling was done using an objective procedure based on labeling techniques developed during previous experiments. This marks the first time an objective procedure was used to label pixels instead of relying entirely on the experience and insight of highly trained analysts to obtain pixel labels.

2.2.2 SAMPLING AND AGGREGATION PROCEDURES

The multicrop optimum allocation procedure determines optimum sample sizes in strata for simultaneous estimates of one, two, or three crop categories. It minimizes overall sample size while maintaining sample coefficients of variation (C.V.'s) below specified levels for each crop.

The optimal aggregation procedure uses a weighting and strata grouping scheme that is designed to make optimum use of available segment proportion estimates in combination with historical crop statistics. This procedure combines strata and differentially weights current proportion estimates and historical ratios to take account of stratum sample sizes and within-stratum variances. It is designed to make stable large-area aggregated estimates even when there are high rates of data loss and sizable proportion estimation variances.

2.3 TEST DESCRIPTIONS AND RESULTS

2.3.1 CLASSIFICATION PROCEDURES VERIFICATION TEST (CPVT)

The two objectives of this test were to (1) determine the accuracy of the newly developed objective labeling procedure and recommend improvements and (2) determine the effectiveness of the maximum likelihood classification procedure in producing corn and soybean proportion estimates. In this test, 1978 full-season Landsat data from 25 segments distributed across the U.S. Corn Belt were processed. Evaluations were performed by comparing the labeling and classification results to digitized ground-truth crop inventories for the segments.

Labeling accuracy was best on spectrally pure (Type I) dots and good on spectrally mixed (Type II) dots. This labeling accuracy is comparable to the accuracies previously achieved for small grains. Some unclear labeling instructions were discovered. When these were clarified in a later test, even better labeling accuracies were achieved. The results indicate that the corn and soybeans labeling procedure performs very well in the U.S. Corn Belt with full-season data. This procedure should be readily adaptable for subsequent experimentation and testing.

Proportion estimates produced by the machine clustering and classification procedure were no better than estimates made directly using Type II dots as a random sample. Use of the procedure resulted in underestimation of corn by an average of 4 percent and underestimation of soybeans by 6 percent. Alternatives to the machine processing techniques used in this experiment should be investigated to determine whether more effective techniques can be found.

2.3.2 SIMULATED AGGREGATION TEST (SAT)

The primary objective of this test was to evaluate the sampling and aggregation components of the production estimation system. This test was a simulation test on an optimum multicrop allocation of 204 segments in the corn belt. Proportion estimation variances and National Oceanic and Atmospheric Administration (NOAA) yield model variances were taken into account in the allocation. Proportion estimation variances were estimated from processing 88 segments using the corn and soybeans estimation procedure. One hundred simulation runs were performed in which simulated segment estimates were randomly designated as lost at each of five loss rates, and aggregated estimates of acreage and production were made. The distributions of aggregated estimates were compared against actual acreage and production as reported by the USDA.

The simulation tests showed that the allocation procedure was producing estimates with CV's in good agreement with the expected value of 5 percent. The tests of the aggregation procedure demonstrated that the procedure introduced no bias into the aggregated area and production estimates for acquisition rates as low as 10 percent. The increase in CV's resulting from reduced acquisition rates were reasonably small. Estimates of CV's produced by the procedure correspond closely to the actual CV's of the simulated sample. The procedures should serve as a useful baseline component for large-area estimation of acreage and production in future experiments.

3. CLASSIFICATION PROCEDURES VERIFICATION TEST DESCRIPTION

3.1 OBJECTIVES AND SCOPE

The two objectives of this test were (1) to determine the accuracy of the newly developed objective labeling procedure and recommend improvements for use in the SAT, and (2) to determine the accuracy of the proportion estimation procedure. This test involved carrying out the procedures on a sample of test segments for which comparison ground-truth data were collected.

3.2 METHOD

3.2.1 PROCEDURE DESCRIPTION

The procedure used to process the segments for this test is shown in figure 1. Using Landsat and ancillary data, an objective labeling procedure was used to label two sets of pixels from each segment. The major steps in the labeling procedure are shown in figure 2. The procedure is set up to provide increasingly more detailed labeling information at each step in the procedure. The first step consists of a decision tree labeling logic which is used to separate the pixels into cropland and noncropland. The pixels labeled cropland in the first step are separated into summer crops and "other crops" in the second step. This step also uses a decision tree labeling logic. The third step uses a greenness/brightness scatter plot for the separation acquisition to separate the summer crop pixels into corn and soybeans. Labeling methodology is described in a report by C. L. Dailey and K. M. Abotteen (ref. 5), which is included in this document as appendix B.

The first set of analyst-labeled pixels (called Type I dots) is used as training for a clustering algorithm which grouped all of the pixels in the segment into clusters on the basis of their spectral values. Each of the resulting clusters is labeled as corn, soybeans, or "other" using the labeled Type I dot closest to the mean of the cluster. On the basis of the means and variances for each cluster, a maximum likelihood classification of every pixel in the segment is performed. Using the second set of analyst labeled dots (called Type 2 dots) as a random sample of the segment, the proportion based on the

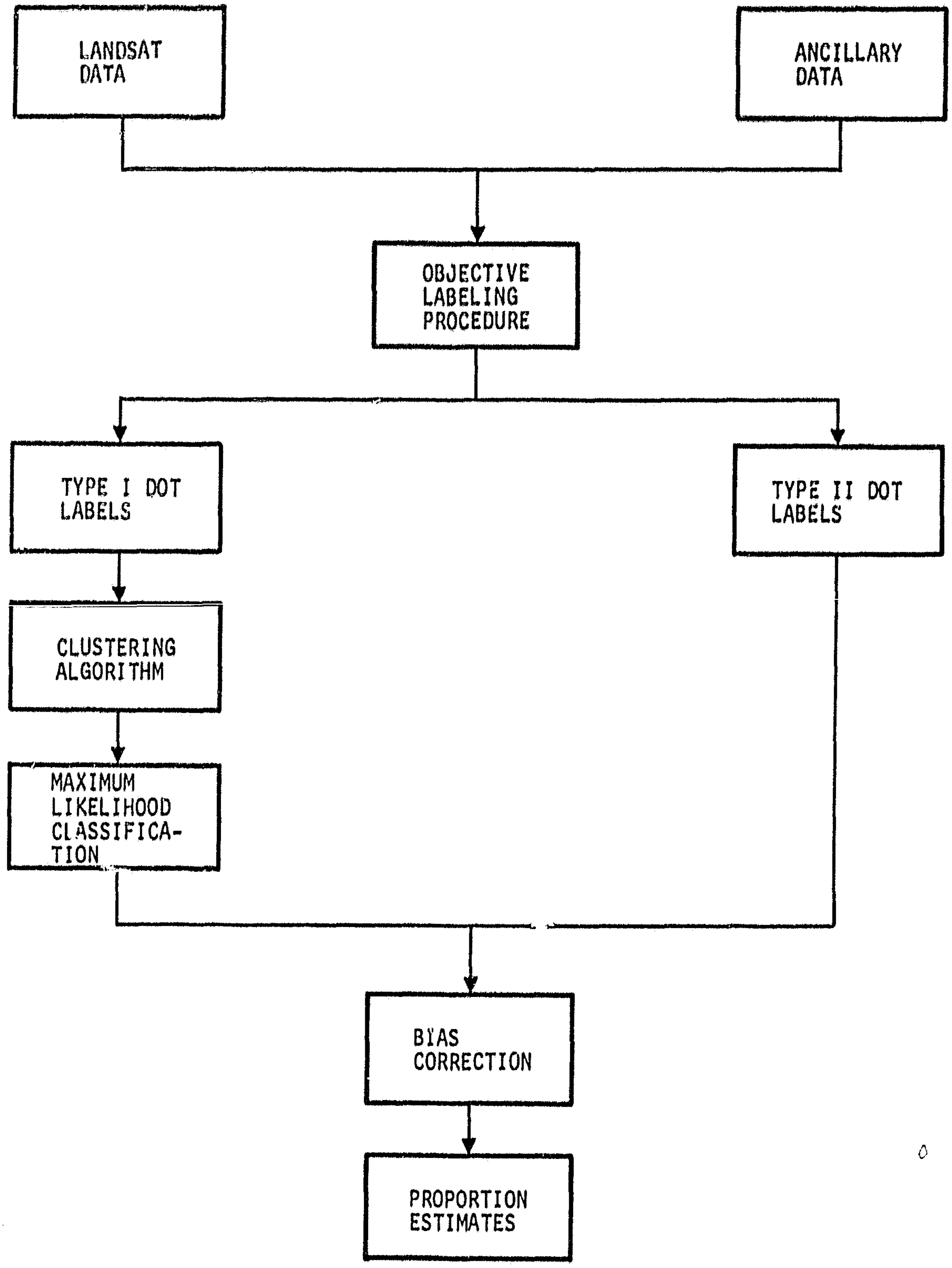


Figure 1.- Diagram showing procedure for processing segment for the Classification Procedures Verification test.

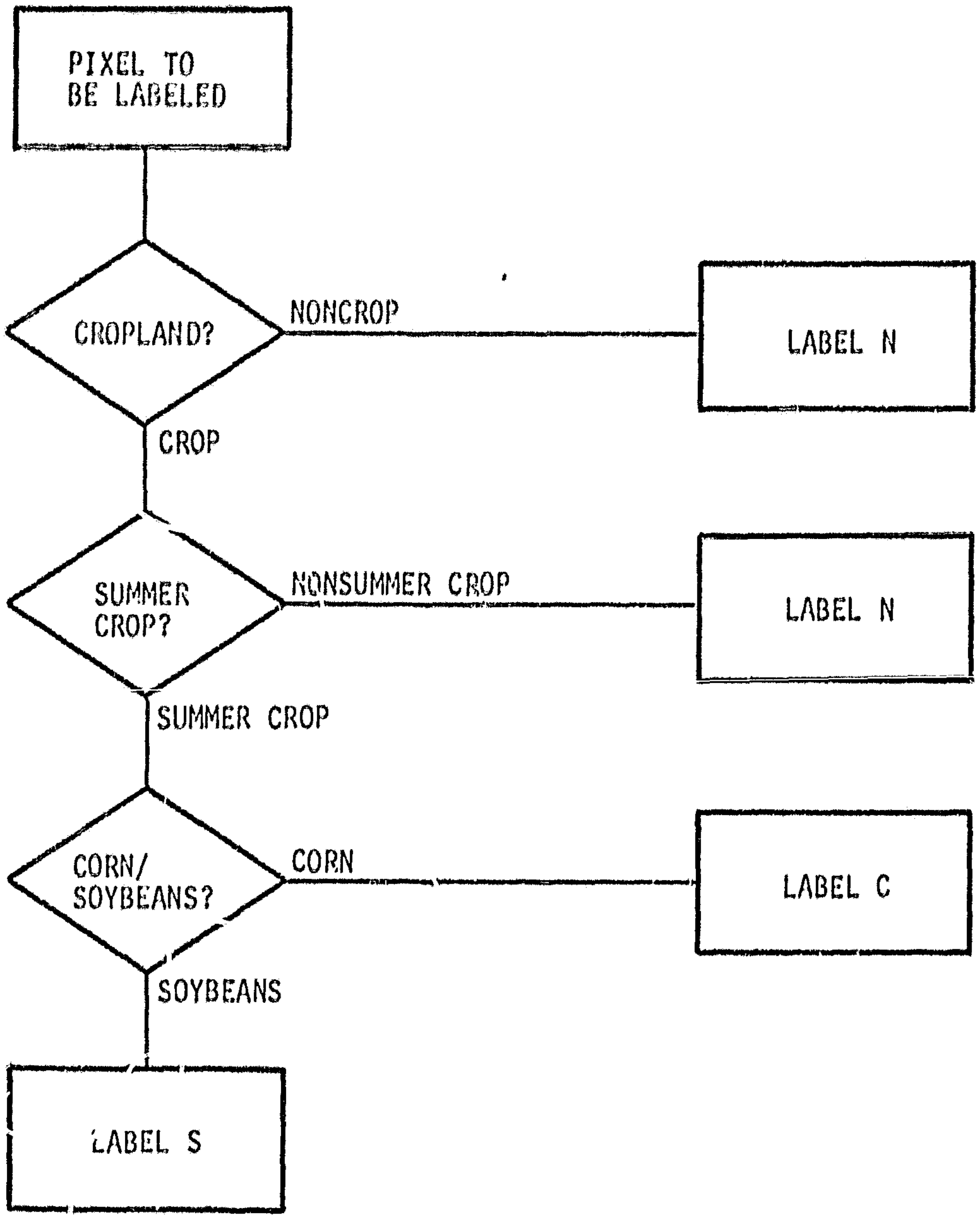


Figure 2.- Diagram showing the major steps in the labeling procedure for the Classification Procedures Verification test.

classification is corrected for any "bias" introduced by the classification process.

3.2.2 DESIGN AND DATA SET

The CPVT consisted of labeling and proportion estimation on 25 segments from four agrophysical units (APU's) in the U.S. Corn Belt using Landsat data from the 1978 crop year. The locations of the segments used in the CPVT are shown in figure 3.

The segments in the CPVT were processed independently by three groups of analysts. Each segment was processed by at least two of the groups. The test followed a rigid experiment design so that analysis of variance techniques could be used to determine if the quality of the labeling and proportion estimation results were dependent on the group doing the labeling or on the APU in which the segment was located (ref. 6). All of the evaluations were performed by comparing the labeling and classification results to the digitized ground-truth crop inventories.

3.3 RESULTS AND EVALUATION

In the CPVT, statistical tests were performed to determine if there was a significant difference in the quality of the labeling and proportion estimation results due to the group performing the processing or the region in which the segment was located. The measures of quality used were dot labeling accuracy, percentage of correct classification, and proportion estimation error. A regional difference was observed for the dot labeling accuracy for soybeans. The labeling of soybeans was significantly less accurate in a predominantly corn-producing region than in the regions where soybeans were more prevalent. A group effect was found in the dot labeling accuracy for corn. One group produced significantly more accurate dot labeling for corn. Investigation showed that the difference was due to a difference in the way the group placed the separation line on the scatter plots for corn and soybeans.

The labeling accuracies for the CPVT are shown in table 1. The labeling accuracy is comparable to the small-grains labeling accuracies previously

ORIGINAL PAGE IS
OF POOR QUALITY

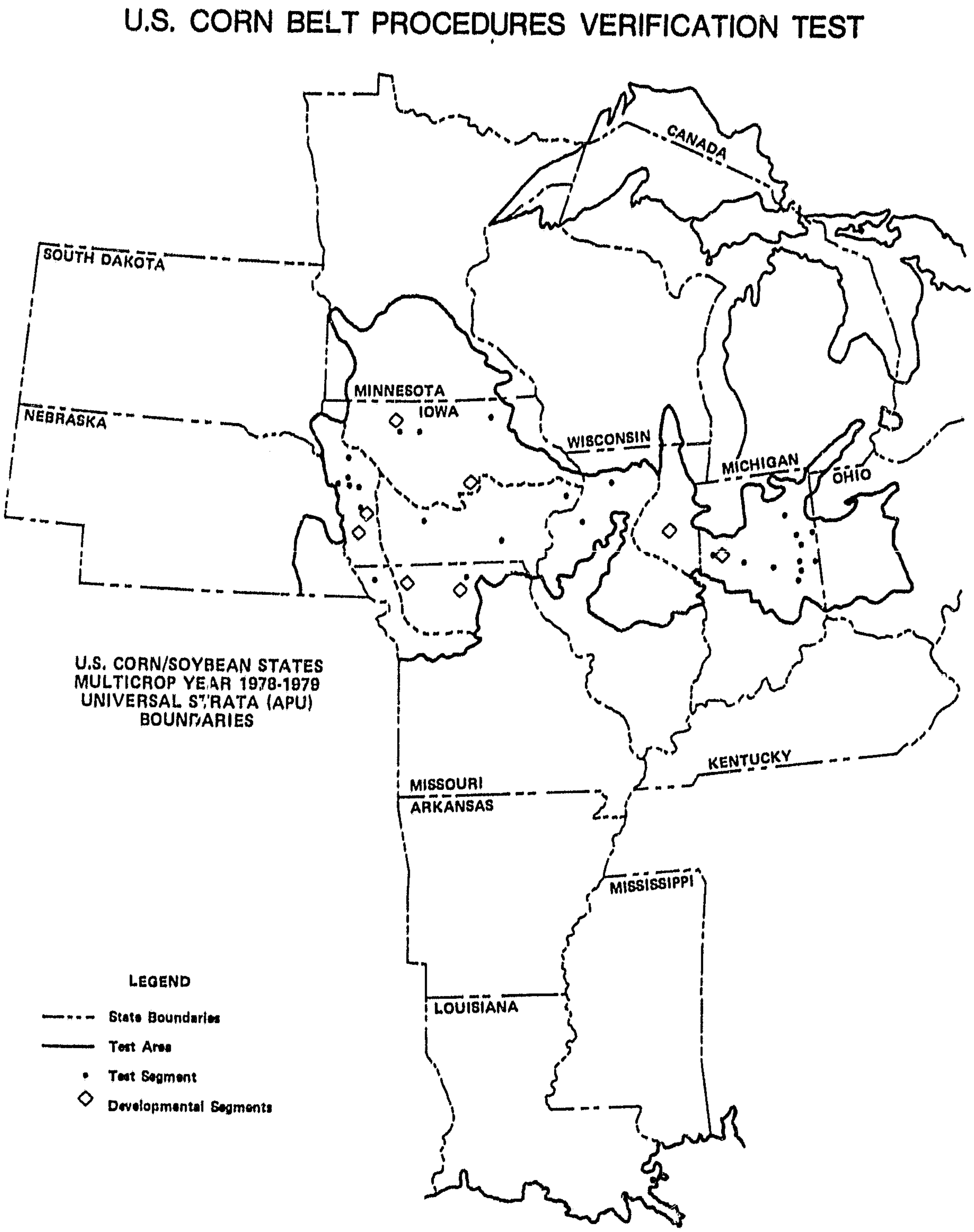


Figure 3.- Map showing locations of the segments used in the Classification Procedures Verification test.

achieved during the Large Area Crop Inventory Experiment (LACIE). The labeling for Type I dots was better than for Type II dots. This difference results from the fact that the Type I dots are required to be spectrally pure, while the Type II dots can be spectrally mixed. It is, therefore, natural to expect better labeling accuracy on dots which are representative of a particular crop, rather than a mixture of signatures from more than one crop.

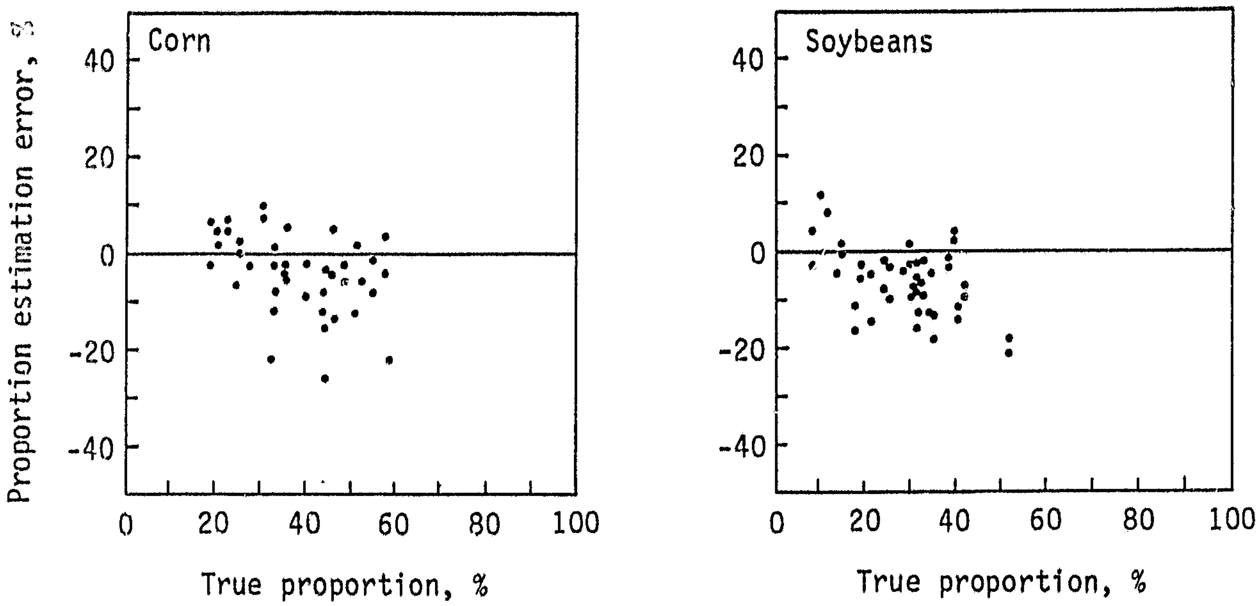
The proportion estimation errors as a function of the true proportion are shown for both corn and soybeans in figure 4. The average proportion of corn in the segments was 38 percent. The machine processing procedure underestimated the corn proportion by an average of 4 percent. The average proportion of soybeans was 28 percent. The procedure underestimated the soybeans proportion by 6 percent. All of the bias and half of the variability in the proportion estimation errors were the result of dot labeling errors. The proportion estimates produced by the procedure were not any better than estimates obtained by using the Type II dots as a random sample. Therefore, the machine processing (i.e., clustering and classification) did not improve the results.

Since the labeling and classification accuracies were much better for spectrally pure pixels than for mixed pixels, a study was made on the segments in this test to determine if accurate proportion estimates could be obtained from classification information for spectrally pure pixels. In order to perform the study, analysts assigned each of the pure pixels with its ground-truth label, and a proportion estimate was made using only these pixels. Figure 5 shows the proportion estimation errors for two criteria for pixel purity. Pixels which meet the "one-half pixel" purity criterion are at least one-half pixel from the field boundaries. Pixels which meet the "one pixel" criterion are at least one pixel from the field boundaries. The results indicate that proportion estimates based only on pure pixels can be biased and have a great deal of variability. In the data set used in this test, the corn estimates showed a positive bias.

This test is described in detail in a report by J. G. Carnes and J. E. Baird (ref. 4), which is included in this document as appendix A.

TABLE 1.- SUMMARY OF DOT LABELING RESULTS FOR THE CLASSIFICATION PROCEDURES VERIFICATION TEST

Ground-truth category	Percent correctly labeled	
	Type 1 dots	Type 2 dots
Corn	83	73
Soybeans	79	64
Other	93	86
All categories	86	75



Mean error = -4%
Standard deviation = 8%
RMSE = 9%

(a) Corn — proportion estimate.

Mean error = -6%
Standard deviation = 7%
RMSE = 9%

(b) Soybeans — proportion estimate.

Figure 4.- Proportion estimation errors as a function of the true proportion for both corn and soybeans.

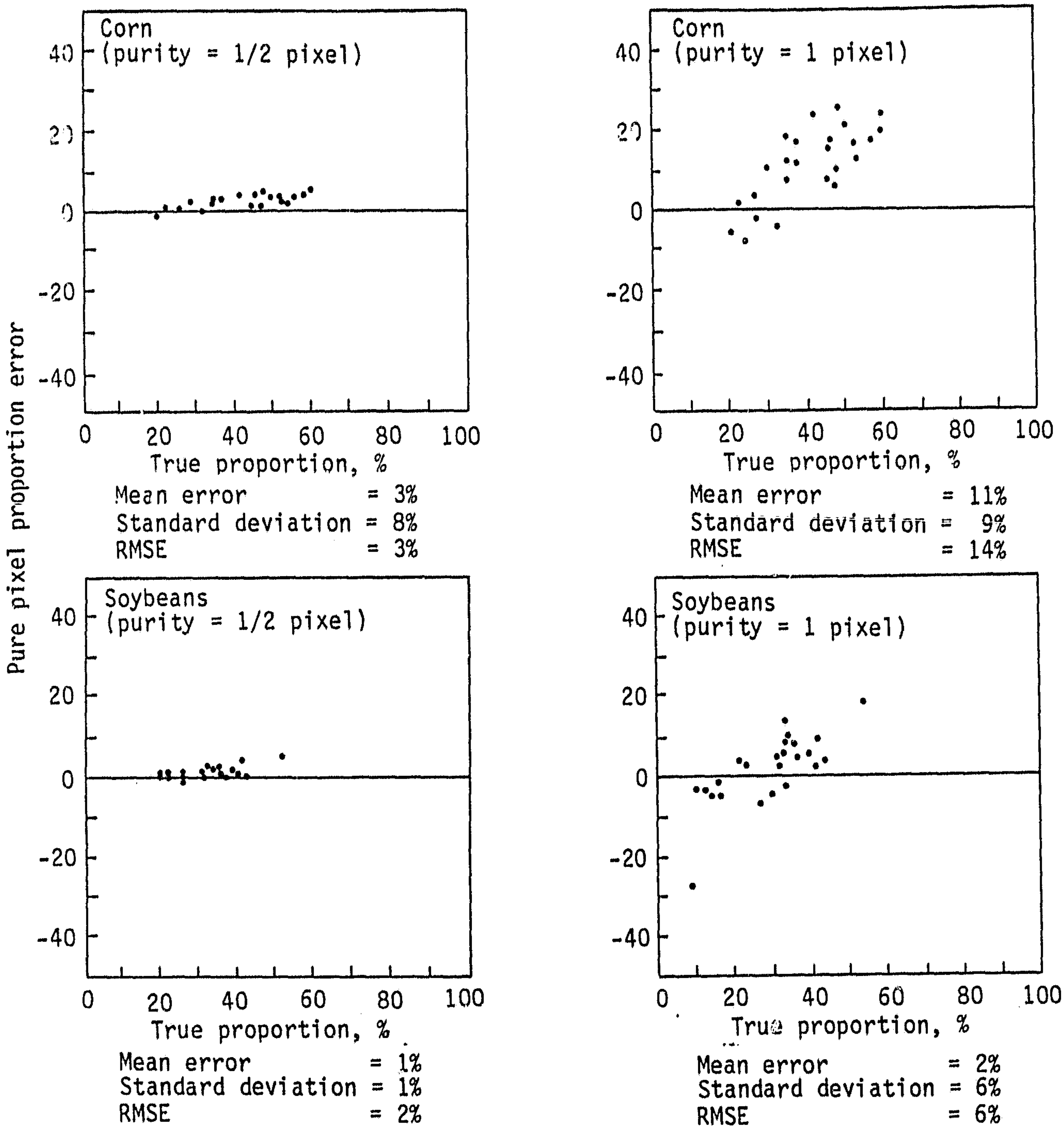


Figure 5.- Proportion errors when only pure pixels are used to determine the proportions.

4. SIMULATED AGGREGATION TEST DESCRIPTION

4.1 OBJECTIVES AND SCOPE

This test was accomplished in two studies. The first study involved proportion estimation of corn belt segments to provide estimates of variability of segment proportion estimates and to evaluate the classification procedures as they were modified following the CPVT. This study is described in a report by S. A. Davidson (ref. 7), which is included in this document as appendix C.

The second study was the simulation study that used the proportion estimation variances derived in the first study. The objectives of the simulation study were to (1) verify that the optimum multicrop sample allocation procedure provided correct sample allocations among the strata, (2) validate the new aggregation and variance estimation logic, and (3) determine the robustness of the procedure under random nonresponse. This study is described in a report by J.H. Smith (ref. 8), which is included in this document as appendix D.

4.2 METHOD

4.2.1 PROCEDURE DESCRIPTIONS

The labeling procedure used in the SAT was essentially the same as that used in the CPVT. The changes made as a result of the CPVT were mainly improvements in the clarity of the procedure. The proportion estimation procedure was modified from the procedure used in the CPVT. On the basis of a study performed by the Supporting Research project of the AgRISTARS program (ref. 9), the objective of providing estimates of variability of segment proportions and resource considerations, the decision was made not to perform the bias correction on the initial proportion estimates in the SAT. Therefore, the proportion estimation procedure involved labeling of the Type I dots, classification of the segment, and proportion estimation by enumeration of pixels in the class of interest.

The multicrop allocation procedure tested in the second part of the SAT formulates the allocation problem in terms of nonlinear programming. The sample size is minimized using a Lagrangian Multiplier technique, subject to the constraints that the sample C.V.'s for each crop not exceed a given value (ref. 10).

The aggregation procedure tested in the second part of the SAT is shown in figure 6. It consists of a technique for using historical data to compensate for the loss of data in a particular stratum (ref. 11). The technique involves a weighting procedure which places more reliance on historical data as the classification results become less reliable because of data loss or errors in the classification results.

4.2.2 DESIGN AND DATA SET

The 88 segments in the SAT were each processed once. Twenty-three of the segments had been processed in the CPVT. These were processed in the SAT, but by a different analyst group. Thirty-five additional segments with ground-truth inventories were processed and used in the evaluations. For 30 segments no ground-truth data were available. The locations of the segments used in this test are shown in figure 7. Evaluations of the labeling and proportion estimation accuracies were performed using the segments for which ground-truth information was available.

The simulation test of the aggregation procedure was performed by setting up an allocation of 204 simulated segments in 12 strata in the states of Illinois, Indiana, and Iowa. Historical data were used to determine the mean crop proportions within strata. The distribution of segment proportions was determined from the historical variability and from the empirical variances observed in the classification results. State-level historical data were used to determine mean yields, and the distribution of yield estimates was determined using NOAA yield model variance.

A Monte Carlo simulation was performed in which segments were randomly designated as "lost". For each loss rate, 100 simulations were performed to obtain aggregated estimates of production.

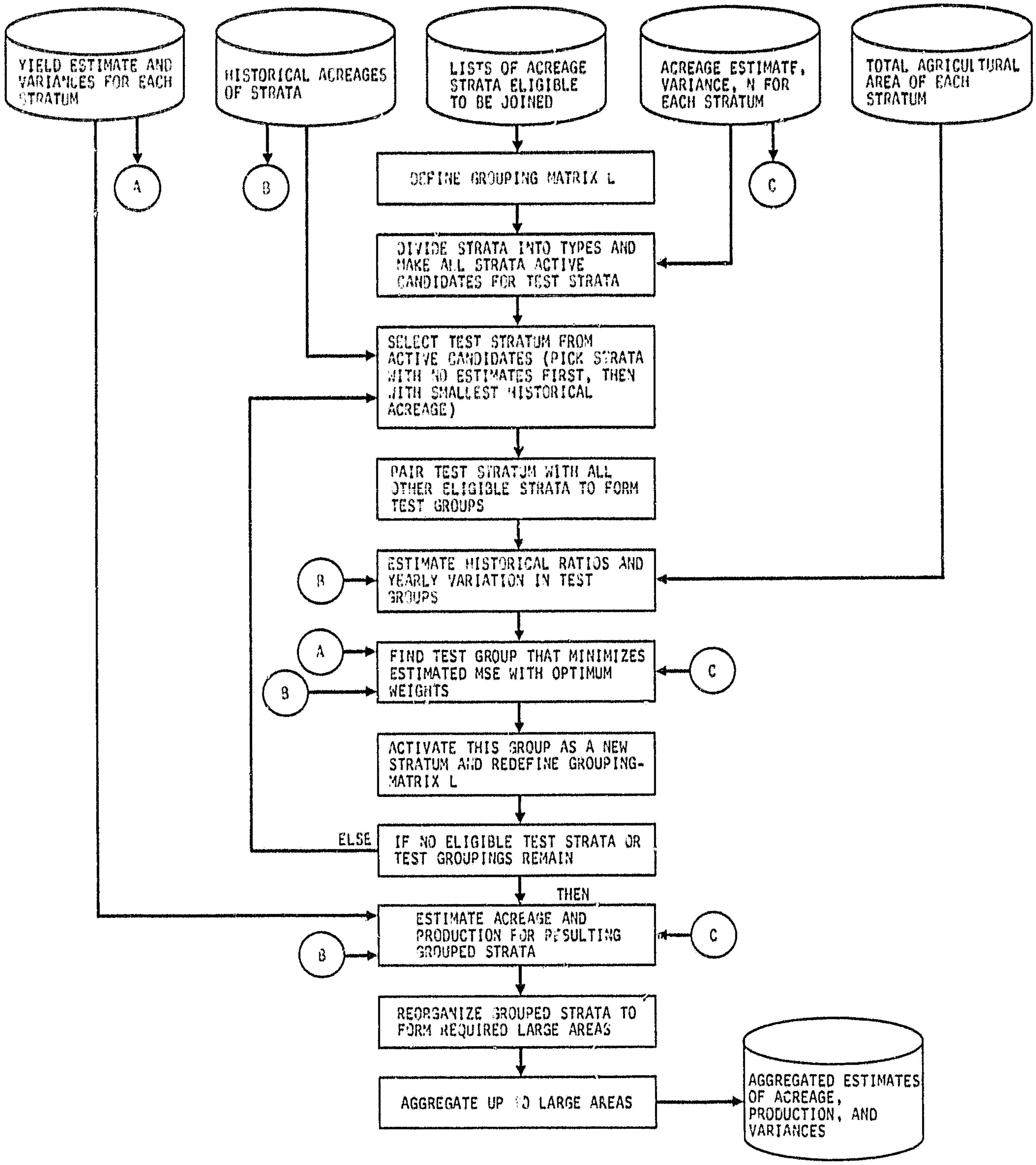


Figure 6.- Grouped Optimal Aggregation Technique.

SIMULATED AGGREGATION TEST

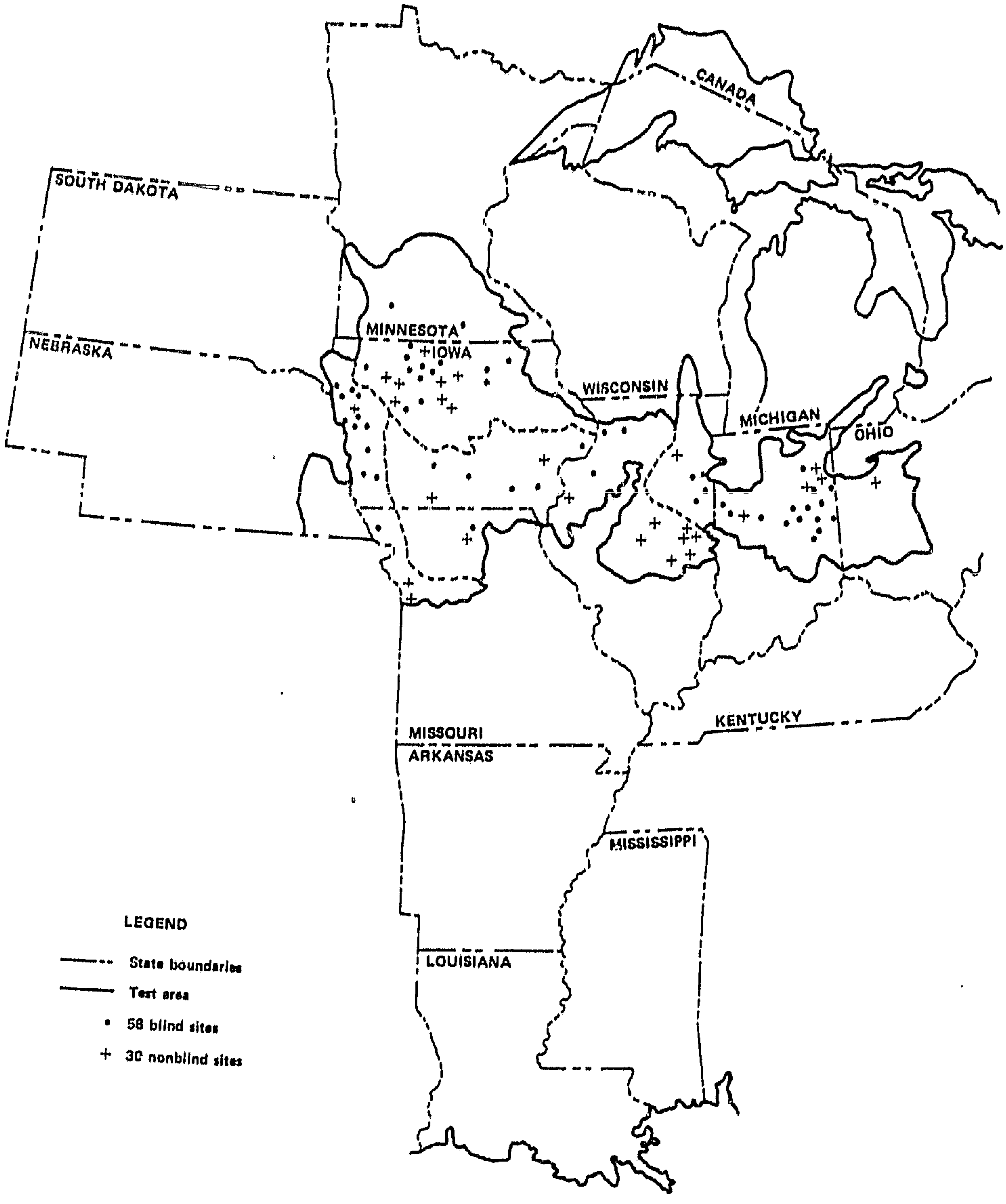


Figure 7.- Map showing locations of the segments used in the Simulated Aggregation test.

4.3 RESULTS AND EVALUATION

In the SAT, the labeling accuracy was better than the accuracy in the CPVT. Table 2 shows a comparison of the labeling accuracies in the two tests. The improvement in the labeling accuracy for the second test was due to changes in the labeling procedure recommended on the basis of the first test and to an improved procedure for selecting acquisitions.

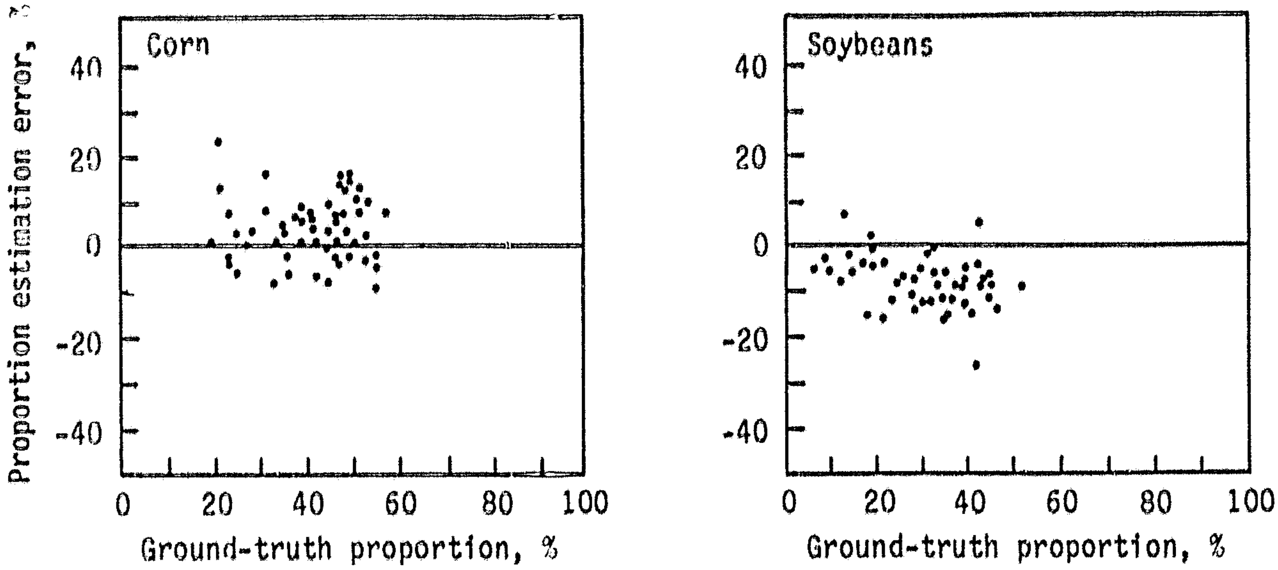
The proportion estimation results for the SAT are shown in figure 8. The results for soybeans proportion estimation were comparable to those obtained in the CPVT. The average soybeans proportion in the segments was 30 percent. The procedure underestimated the soybeans proportion by an average of 8 percent. For corn, the average proportion was 41 percent. In the SAT, the procedure overestimated the corn proportion by 5 percent, while in the CPVT, the proportions were underestimated by 4 percent. The change in bias between the two tests is due to the fact that a bias correction was not performed in the SAT. The classification procedure was trained using only spectrally pure pixels. When only pure pixels are used in training, a classification is produced which is representative of the pure areas of the segment, rather than of the entire segment. As the pure pixel studies showed, this will produce a positive bias in the classification results.

The simulation tests of the sampling and aggregation procedures were set up to provide large area production estimates with a CV of 5 percent for both corn and soybeans at a 100 percent acquisition rate. The aggregation procedure was tested to determine if the CV estimates computed by the procedure were correct, if any bias was introduced into the aggregated estimates because of nonresponse, and if the CV's at reduced response rates were reasonable.

The simulation tests showed that the allocation procedure was producing estimates with CV's in good agreement with the expected value of 5 percent (CV = 4.7 percent for corn and CV = 5.2 percent for soybeans). The tests of the weighted aggregation procedure demonstrated that the procedure introduced no bias into the aggregated area and production estimates for acquisition rates

TABLE 2.- COMPARISON OF LABELING ACCURACY FOR CPVT AND SAT TESTS

Ground-truth categories	Percent correctly labeled	
	CPVT (Type I dots)	SAT
Corn	86	93
Soybeans	79	88
Other	93	96
All categories	86	92



Mean error = 5%
Standard deviation = 7%
RMSE = 8%

Mean error = -8%
Standard deviation = 6%
RMSE = 9%

Figure 8.- Summary of results for the Simulated Aggregation test.

as low as 10 percent. Figure 9 shows the CV's resulting from reduced acquisition rates for area and for production. These variances are reasonable, and the average CV estimates produced by the procedure correspond closely to the CV's of the simulated sample.

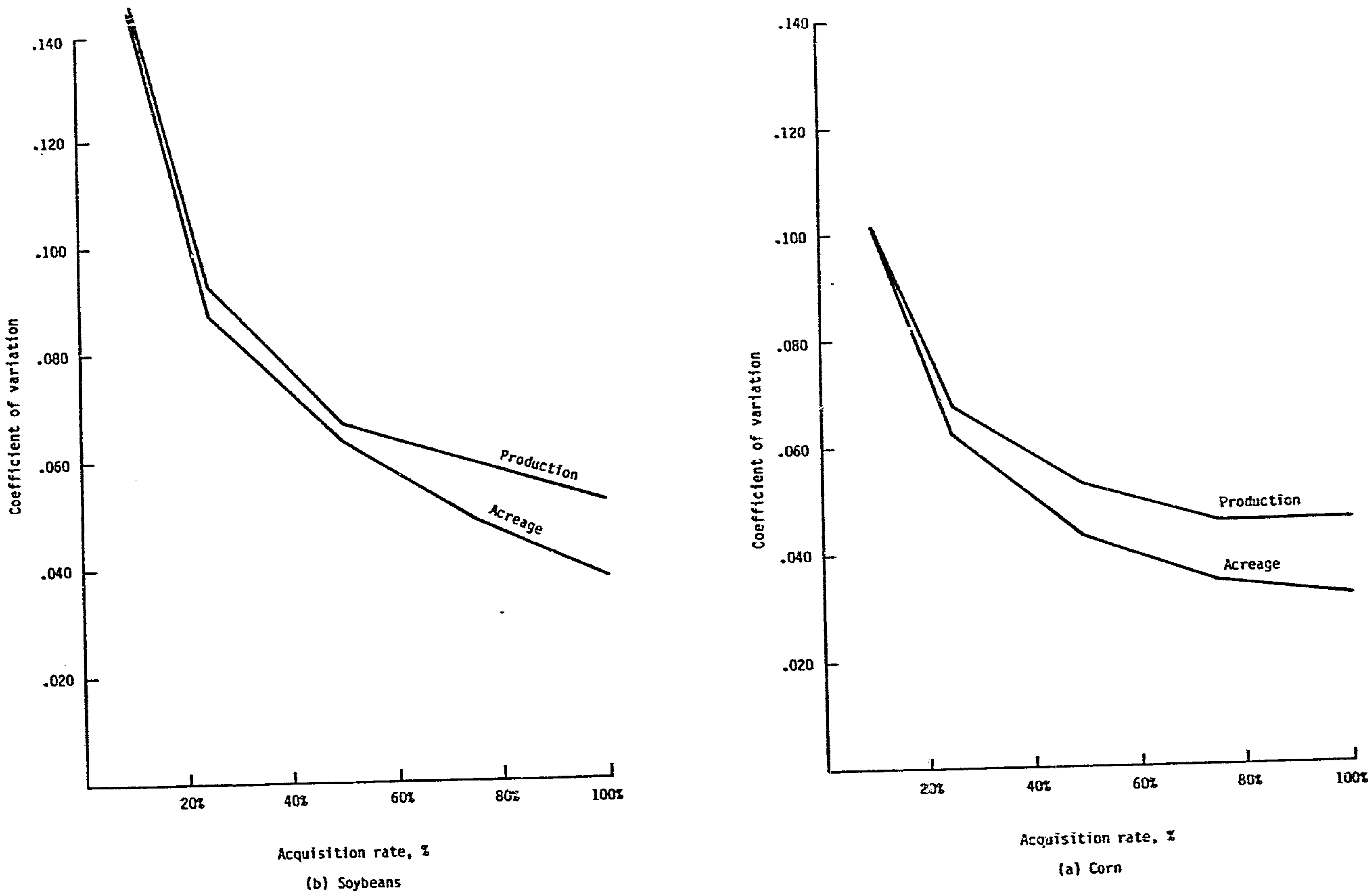


Figure 9.- Coefficients of variation resulting from reduced acquisition rates for acreage and production for three states.

5. CONCLUSIONS AND RECOMMENDATIONS

The results from the labeling evaluations indicate that the corn/soybeans labeling procedure performs very well in the U.S. Corn Belt with full-season (after tasseling) Landsat data. The procedure should be readily adaptable to corn/soybeans labeling required for subsequent exploratory experiments or pilot tests.

The machine classification procedures evaluated in this experiment were not effective in improving the proportion estimates. The corn proportions produced by the machine procedures had a large bias when the "bias" correction was not performed. This bias was caused by the manner in which the machine procedures handled spectrally impure pixels. Alternatives to the machine processing techniques used in this experiment should be investigated to see if more effective techniques can be found.

The simulation test indicated that the weighted aggregation procedure performed quite well. Although further work can be done to improve both the simulation tests and the aggregation procedure, the results of this test show that the procedure should serve as a useful baseline procedure in future exploratory experiments and pilot tests.

6. REFERENCES

1. 1980 U.S. Wheat and Barley Exploratory Experiment: Final Report. (To be published).

2. Technical Program Plan - AgRISTARS. January 15, 1980.

3. LACIE Transition Project Accuracy Assessment Fiscal 1979 Interim Plan. LACIE-00634, JSC-13770, September 1979.

4. Carnes, J. G.; and Baird, J. E.: Evaluation of Results of U.S. Corn and Soybeans Exploratory Experiment - Classification Procedures Verification Test. LEMSCO-14386, JSC-16339, September 1980.

5. Dailey, C. L.; and Abotteen, K. M.: Corn/Soybeans Decision Logic Development and Testing. LEMSCO-14811, JSC-16380, October 1980.

6. LACIE Transition Project Accuracy Assessment Fiscal Year 79 Interim Plan, LEC-13753, JSC-13770, August 1979.

7. Davidson, S. A.: Segment-Level Evaluation of the Simulated Aggregation Test: U.S. Corn and Soybean Exploratory Experiment. LEMSCO-15116, JSC-16820, September 1980.

8. Smith, J. H.: Test of Grouped Optimal Aggregation Technique. (To be published).

9. Havens, K. A.; and Abotteen, K. M.: Evaluations of Three-Category Classification. LEC-13498, JSC-16012, August 1979.

10. Malin, J. T.; Smith, J. H.; Cheffin, R. E.; Norwood, D. F.; McLean, T. K.: Sampling and Aggregation Software Systems Modules Description. LEMSCO-16221, JSC-17136, August 1981.

11. Feiveson, A. H.: Weighted Ratio Estimation for AgRISTARS. JSC-16861, February 1981.

APPENDIX A

EVALUATION OF RESULTS OF U.S. CORN AND SOYBEANS EXPLORATORY EXPERIMENT - CLASSIFICATION PROCEDURES VERIFICATION TEST

AgRISTARS

FC-LO-00423
JSC-16339

A Joint Program for Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing

Foreign Commodity Production Forecasting

September 1980

EVALUATION OF RESULTS OF U.S. CORN AND SOYBEANS EXPLORATORY EXPERIMENT - CLASSIFICATION PROCEDURES VERIFICATION TEST

J. G. Carnes and J. E. Baird

Lockheed Engineering and Management Services Company, Inc.
1830 NASA Road 1, Houston, Texas 77058

Lyndon B. Johnson Space Center
Houston, Texas 77058

PRECEDING PAGE BLANK NOT FILMED

1. Report No. JSC-16339; FC-L0-00423

2. Government Accession No.

3. Recipient's Catalog No.

4. Title and Subtitle
Evaluation of Results of U.S. Corn and Soybeans Exploratory Experiment – Classification Procedures Verification Test

5. Report Date
September 1980

6. Performing Organization Code

7. Author(s)
J. G. Carnes and J. E. Baird

8. Performing Organization Report No.
LEMSCO-14386

9. Performing Organization Name and Address
Lockheed Engineering and Management Services Company, Inc.
1830 NASA Road 1
Houston, Texas 77058

10. Work Unit No.

11. Contract or Grant No.
NAS 9-15800

12. Sponsoring Agency Name and Address
National Aeronautics and Space Administration
Lyndon B. Johnson Space Center
Houston, Texas 77058 Tech. Monitor: R. O. Hill

13. Type of Report and Period Covered
Technical Report

14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

The U.S. Corn and Soybeans Exploratory Experiment – Classification Procedures Verification Test was undertaken to evaluate the classification procedure utilized in making crop proportion estimates for corn and soybeans using remotely sensed data. The procedure was derived during the Transition Year of the Large Area Crop Inventory Experiment in the Earth Observations Division at the NASA Lyndon B. Johnson Space Center. Analysis of variance techniques were applied to classifications performed by 3 groups of analysts who processed 25 segments selected from 4 agrophysical units (APU's). Group and APU effects were evaluated to determine factors which affected the quality of the classifications. The classification results were evaluated to determine the effectiveness of the procedure in producing corn and soybeans proportion estimates.

17. Key Words (Suggested by Author(s))
Classification procedure evaluation, clustering, dot labeling procedure, machine classification, multicrop proportion estimates

18. Distribution Statement

19. Security Classif. (of this report) Unclassified

20. Security Classif. (of this page) Unclassified

21. No. of Pages 39

22. Price*

PRECEDING PAGE BLANK NOT FILMED

FC-LO-00423
JSC-16339

TECHNICAL REPORT

EVALUATION OF RESULTS FROM THE U.S. CORN AND SOYBEANS EXPLORATORY EXPERIMENT — CLASSIFICATION PROCEDURES VERIFICATION TEST

Job Order 74-402

This report describes Accuracy Assessment activities of the Foreign Commodity Production Forecasting project of the AgRISTARS program.

PREPARED BY

J. G. Carnes and J. E. Baird

APPROVED BY

Mike Pore

M. D. Pore, Supervisor
Accuracy Assessment Section

B L Carroll

B. L. Carroll, Manager
Commodity Forecasting Department

LOCKHEED ENGINEERING AND MANAGEMENT SERVICES COMPANY, INC.

Under Contract NAS 9-15800

For

Earth Observations Division
Space and Life Sciences Directorate
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
LYNDON B. JOHNSON SPACE CENTER
HOUSTON, TEXAS

September 1980

LEMSCO-14386

PREFACE

The investigation which is the subject of this document was undertaken in support of the Foreign Commodity Production Forecasting project of the Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing program. Under Contract NAS 9-15800, scientists of Lockheed Engineering and Management Services Company, Inc., evaluated the results which are reported for the Earth Observations Division, Space and Life Sciences Directorate, of the National Aeronautics and Space Administration, Lyndon B. Johnson Space Center.

CONTENTS

TABLES

FIGURES

PRECEDING PAGE BLANK NOT FILMED

1. INTRODUCTION

The purpose of the U.S. Corn and Soybeans Exploratory Experiment — Classification Procedures Verification Test was to evaluate the performance of the adapted Large Area Crop Inventory Experiment (LACIE) Transition Year (TY) classification procedure for corn and soybeans. See reference 1 for a discussion of the procedure used in this test. In this test, 25 segments selected from four agrophysical units (APU's) were processed by three groups of analysts. Analysis of variance techniques were used to determine the factors which were important to the quality of the classifications performed. The factors evaluated were group effects and APU effects. The classification results were evaluated to determine the effectiveness of the procedure in producing corn and soybeans proportion estimates.

2. FACTORS AFFECTING THE QUALITY OF THE CLASSIFICATIONS

The segments used in this test were from APU's 14, 24, 25, and 28 located in Missouri, Iowa, Illinois, and Indiana. Because APU 24 had a small number of segments and APU's 14 and 24 were reasonably similar, APU's 14 and 24 were merged and designated APU 14 for evaluation purposes.

Three groups of analysts processed the segments. Group I processed 19 of the segments, whereas groups II and III each processed 18 segments. The allocation of the segments among the groups and APU's is shown in table 1. The linear model and related assumptions used in the analyses of variance are described in reference 2.

The following measures of classification quality were used in the analyses of variance:

a. Proportion estimation error

b. Percentage of picture elements (pixels) correctly classified

c. Reduction in the expected proportion estimate variance if a bias correction were applied to the classification results

d. Analyst dot labeling accuracy

The factors were tested for their effects in the following order: first, interaction between groups and APU's; second, group effects; and, third, APU effects. If a significant result was obtained at one stage, it was impossible to test for significant results at a later stage.

Table 2 shows the average proportion estimation error and average absolute proportion error for corn and soybeans by group and by APU. Significant differences are indicated by numbers in parentheses following the values. No significant effects were found in the results for corn. For soybeans, a significant difference in the proportion errors was found between groups II and III. The absolute proportion error was significantly different for APU 14.

TABLE 1.- DISTRIBUTION OF SEGMENTS BY GROUP AND BY APU

[Parentheses indicate processed data which were not used in the analyses of variance]

APU	Segment number	Group I	Group II	Group III
14	135	X		X
	202	(X)	(X)	(X)
	864		X	X
	865	X		X
	877	X	X	
	880		X	X
	881	X	X	(X)
	882	(X)	(X)	(X)
25	107	X	X	
	141	X		X
	144		X	X
	205	X	X	
	800	(X)	(X)	
	807		X	X
	809	X		X
28	123	X	X	
	127	(X)	(X)	(X)
	133		X	X
	832	X		X
	837	(X)	(X)	(X)
	842	X	X	
	843	(X)	(X)	
	852	X		X
	853	(X)		(X)
	860		X	X

TABLE 2.- PROPORTION ESTIMATION ERRORS

[Significant differences are indicated by number in parentheses following the values]

	Corn		Soybeans	
	Average error, %	Average absolute error, %	Average error, %	Average absolute error, %
Group I	-6.3	7.4	-6.6	7.4
Group II	-3.1	8.1	-9.0(1)	9.0
Group III	-4.8	7.1	-4.0(1)	7.0
APU 14	-5.8	7.4	-2.3	4.5(2)(3)
APU 25	-3.6	5.9	-7.3	9.0(2)
APU 28	-4.8	9.3	-9.9	9.9(3)
Overall	-4.7	7.5	-6.5	7.8

The results for the percentage of pixels correctly classified are shown in table 3. An interaction between groups and APU's for the percentage of correct classification (PCC) for class "other" made it impossible to determine group and APU effects for the PCC for "other." The only significant result was a group effect for the PCC for corn, where the group III result was significantly different from the group I and II results.

The results of reductions in variance are shown in table 4. In analyzing the results for corn, a significant interaction between groups and APU's made it impossible to test for group and APU effects individually. There were no significant effects for soybeans.

Tables 5 and 6 show the dot labeling accuracy for type 1 and type 2 dots. There were group effects for the type 1 dot labeling accuracy for corn and for the overall category. In both cases, group III was significantly different from groups I and II. A significant APU effect was sho [illegible] the labeling accuracy for class "other" in both the type 1 and type 2 [illegible] In both cases, APU 14 was significantly different from APU's 25 and 28.

In summary, the observed group effects involved dot labeling accuracy and PCC for corn. In both cases, group III was consistently less accurate than groups I and II. Since all three groups were given the same training and were to follow the same procedures, it would appear that there was some misunderstanding of the procedure for corn by group III.

The observed APU effects involved dot labeling accuracy and proportion estimation error for soybeans. In both cases, APU 14 had less accurate results than APU's 25 and 28. It appears that dot labeling for soybeans is more difficult in APU 14. It is interesting to note that, although the dot labeling for type 1 dots showed a significant difference, the PCC for the classifications based on these dots did not show a significant difference.

TABLE 3.- PERCENTAGE OF PIXELS CORRECTLY CLASSIFIED

[Significant differences are indicated by number in parentheses following the values]

	Corn PCC	Soybeans PCC	"Other" PCC	Overall PCC
Group I	73.2(1)	64.2	72.1	72.6
Group II	75.6(2)	52.5	68.7	70.8
Group III	62.6(1)(2)	53.9	75.6	68.4
APU 14	77.8	59.9	67.1	72.4
APU 25	69.9	49.8	72.2	70.3
APU 28	63.6	60.9	77.1	69.2
Overall	70.4	56.9	72.1	70.6

TABLE 4.- PERCENTAGE OF REDUCTION IN VARIANCE EXPECTED IF BIAS CORRECTION IS PERFORMED ON CLASSIFICATION RESULTS

	Corn	Soybeans
Group I	61.0	53.2
Group II	62.8	59.3
Group III	61.6	59.5
APU 14	58.9	55.4
APU 25	62.2	59.2
APU 28	64.3	57.4
Overall	61.8	57.3

TABLE 5.- TYPE 1 DOT LABELING ACCURACY

[PCL = percentage of dots correctly labeled; significant differences are indicated by the number in parentheses following the values]

	Corn PCL	Soybeans PCL	"Other" PCL	Overall PCL
Group I	88.3(1)	79.9	89.5	86.8(3)
Group II	89.2(2)	76.2	88.3	86.8(4)
Group III	67.0(1)(2)	66.1	85.8	77.8(3)(4)
APU 14	83.5	83.3	76.9(5)(6)	83.5
APU 25	85.9	65.1	89.6(5)	82.7
APU 28	75.1	73.8	97.1(6)	85.1
Overall	81.5	74.1	87.9	83.8

TABLE 6.- TYPE 2 DOT LABELING ACCURACY

[PCL = percentage of dots correctly labeled; significant differences are indicated by the number in parentheses following the values]

	Corn PCL	Soybeans PCL	"Other" PCL	Overall PCL
Group I	66.9	70.4	85.9	74.9
Group II	70.5	60.6	86.5	74.3
Group III	64.5	61.1	80.7	70.9
APU 14	70.8	72.8	76.6(1)(2)	73.6
APU 25	70.7	61.8	89.3(1)	76.3
APU 28	60.5	57.5	87.2(2)	70.3
Overall	67.3	64.0	84.4	73.4

3. CLASSIFICATION PROCEDURE EVALUATION

In order to determine the effectiveness of the classification procedure in producing proportion estimates, the various stages in the classification procedure must be investigated. One way of doing this is to calculate proportion estimates based only on the information available at a particular stage. By comparing the accuracy at the different stages, one can determine which steps are necessary and which steps are not.

The classification procedure consists of the following steps:

a. Two sets of dots are labeled as corn, soybeans, or "other" by the analyst.

b. Using one set of analyst-labeled (type 1) dots as seed pixels, all pixels in the segment are grouped into clusters on the basis of their spectral values.

c. Each of the clusters is labeled as corn, soybeans, or "other" by the analyst-labeled type 1 dot closest to the mean of the cluster.

d. On the basis of the means and variances for each cluster, every pixel in the segment is classified as corn, soybeans, or "other."

e. Using the second set of analyst-labeled (type 2) dots as a random sample of the segment, the proportions based on the classification are corrected for any bias introduced by the classification process.

Proportion estimates can be calculated at the following four stages in the classification procedure:

a. At the dot labeling stage, the type 2 dots can be aggregated on the basis of their labels to determine a proportion.

b. At the clustering stage, a proportion can be determined by aggregating the pixels in a cluster on the basis of the label assigned to the cluster.

c. At the classification stage, a proportion can be determined by aggregating the pixels on the basis of the labels assigned by the classifier.

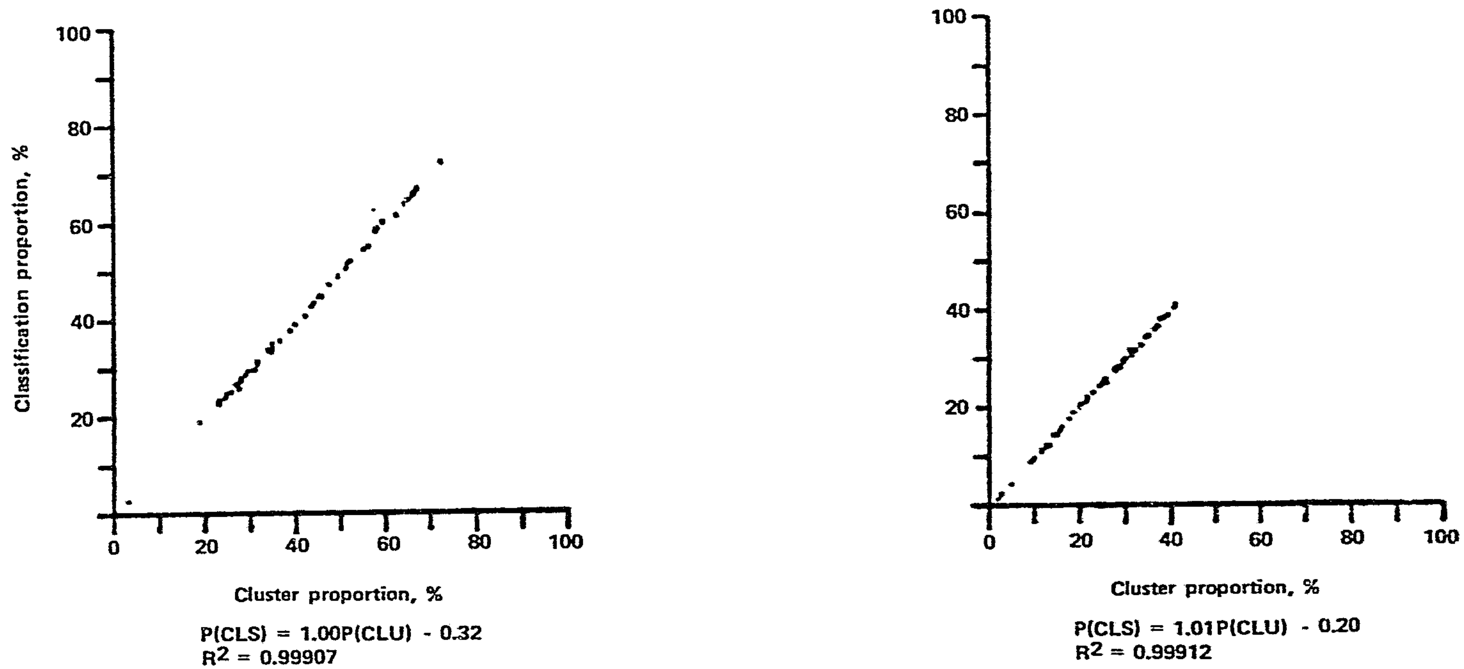
d. At the bias-correction stage, the final estimate produced by the procedure can be used.

The set of classifications used in this evaluation is listed in table 1. For the purposes of evaluating the classification process, five of the classifications were not used: 882 and 127 by group I; 881 by group II; 837 and 860 by group III. Eliminating these classifications resulted in each segment being represented twice by two different groups. Groups I and II were represented 17 times each, whereas group III was represented 16 times.

Although it is possible to determine a proportion at the clustering stage, clustering proportions are not presented. The cluster-based proportions are not included because the cluster and classification proportions are essentially identical. Figure 1 shows the classification proportions P(CLS) as a function of the cluster proportions P(CLU) for the segments involved in this evaluation. The linear regressions shown in the figure indicate an almost perfect correlation between the two proportion estimates ($R^2 = 0.99907$). Therefore, proportion estimates are calculated for the type 2 dots, classification, and bias-correction stages.

Figure 2 shows the errors in the proportion estimates as a function of the true proportion. The mean error, standard deviation, and mean square error for each estimator are presented in table 7 (page 3-7). The mean error is a measure of the bias in the estimator. The standard deviation is a measure of the estimator's variability. The mean square error is an indication of the overall performance of the estimator.

The mean error for corn was negative at the dot labeling and bias-correction stages and positive at the machine classification stage. The mean square errors were nearly the same at the dot labeling and bias-correction stages. This indicates that the machine processing did not improve the proportion estimate. The type 2 dots produced as good an estimate by themselves as when they were used to establish a bias-correction factor for the machine classification.

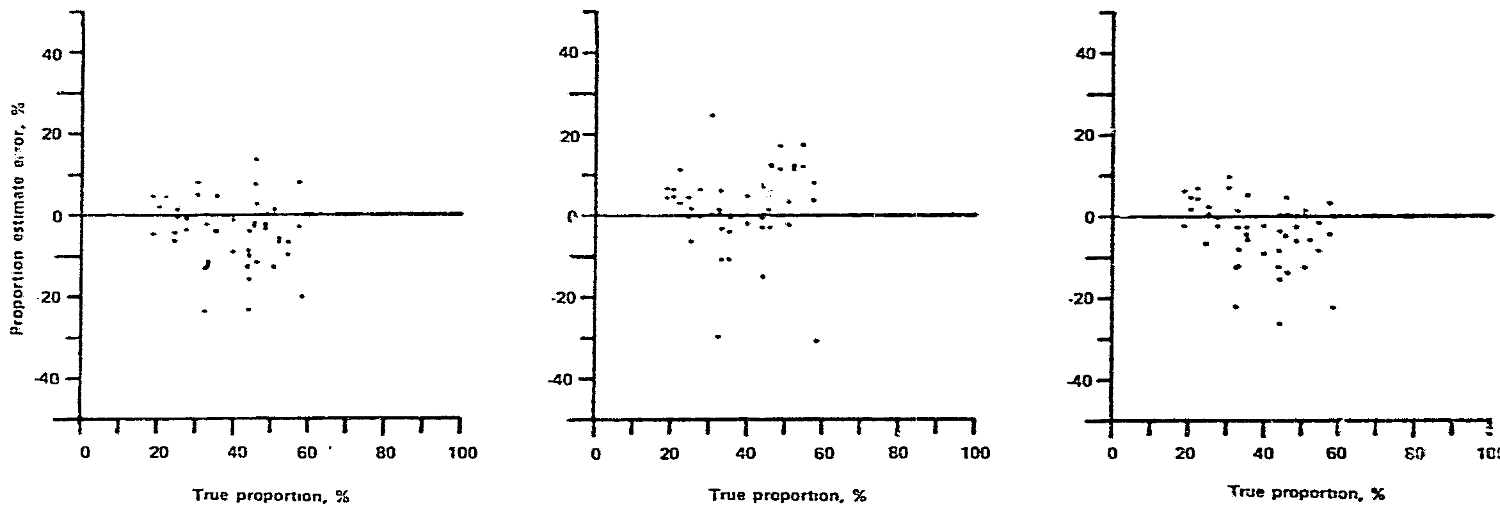


(a) Corn.

(b) Soybeans.

Figure 1.- Comparison between classification proportions and cluster proportions.

ORIGINAL PAGE IS
OF POOR QUALITY

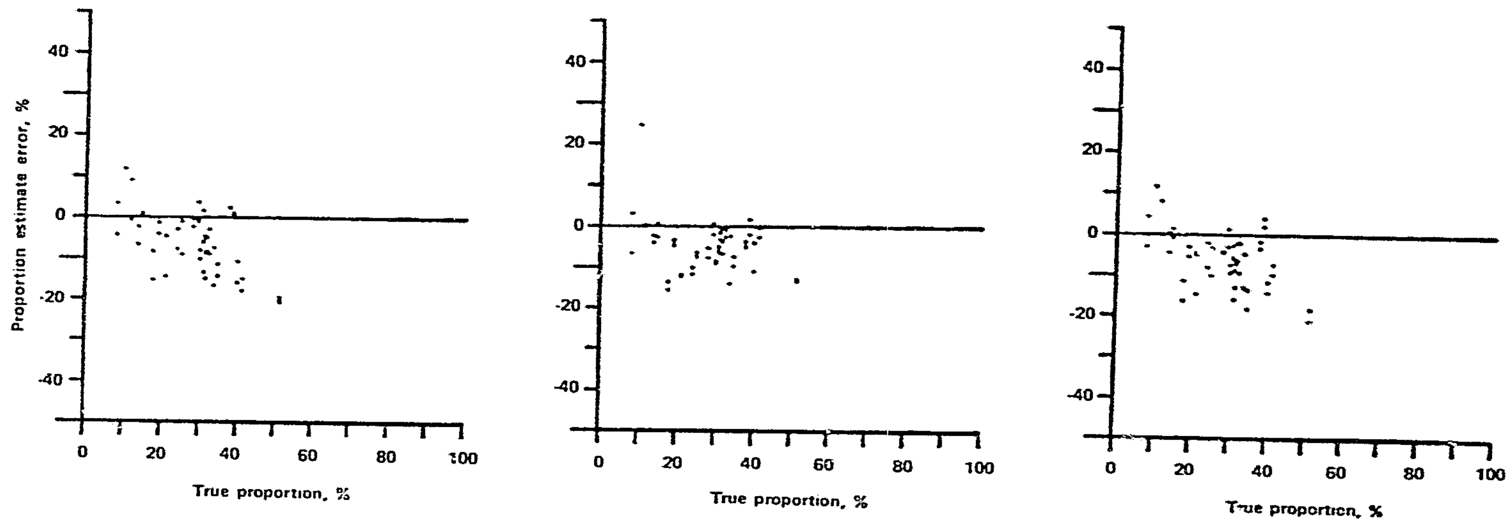


(a) Corn — dot estimate.

(b) Corn — classification estimate.

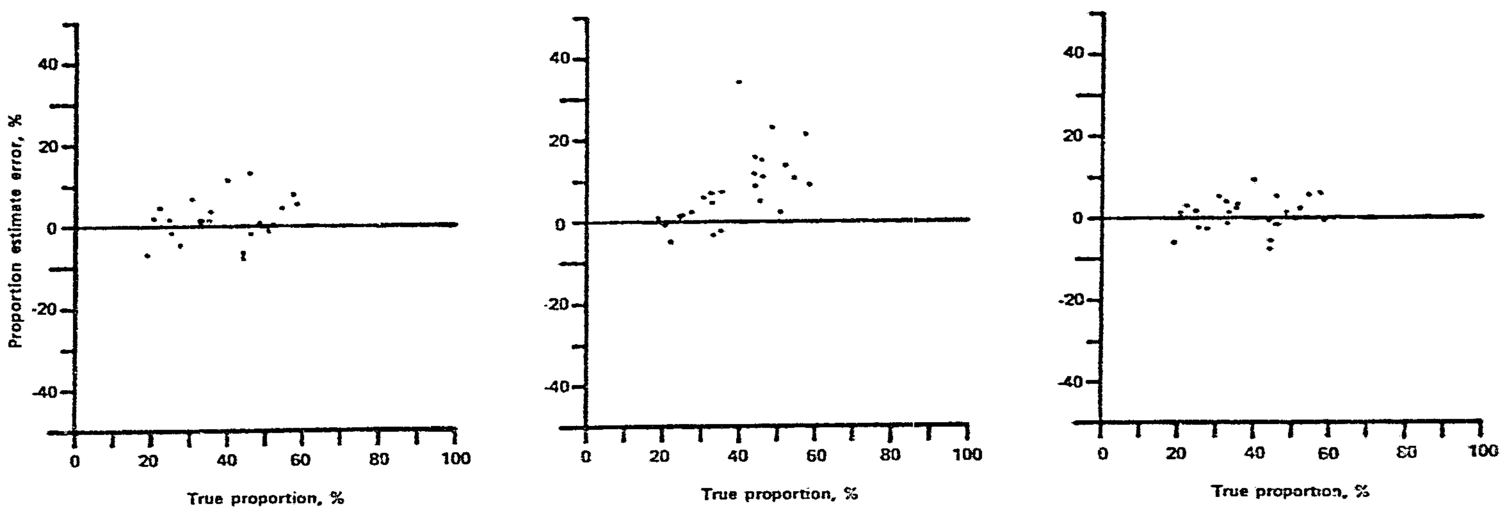
(c) Corn — bias-corrected estimate.

Figure 2.- Proportion estimates using analyst labeling as input.



(d) Soybeans — dot estimate. (e) Soybeans — classification estimate. (f) Soybeans — bias-corrected estimate.

Figure 2.- Concluded.

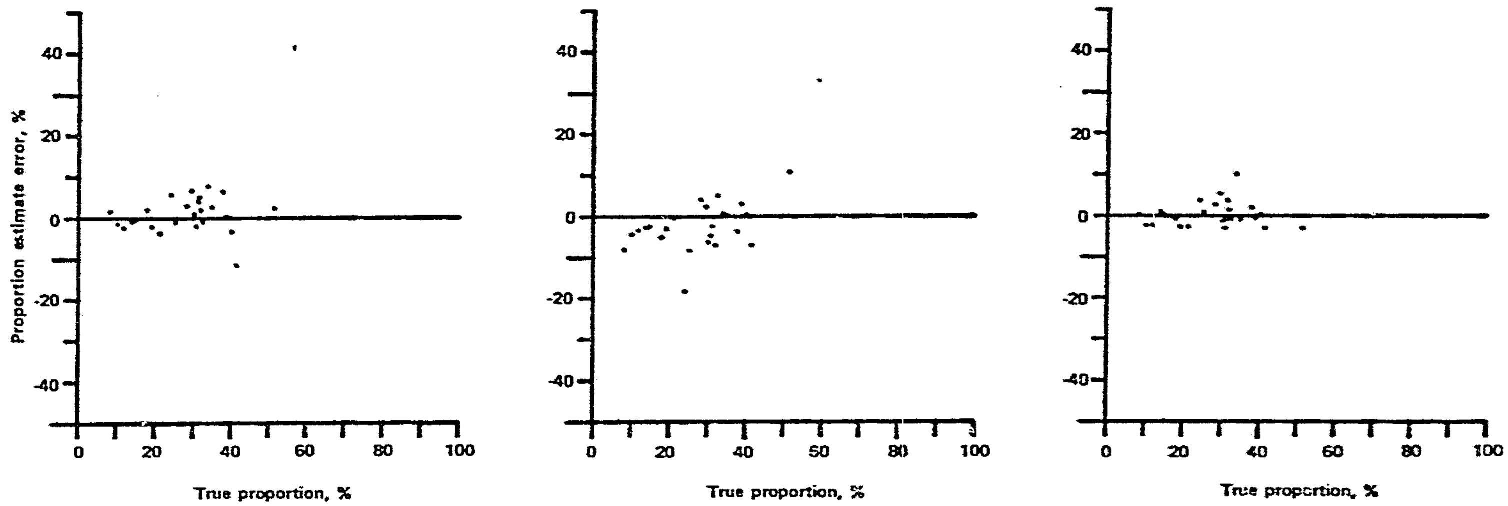


(a) Corn — dot estimate.

(b) Corn — classification estimate.

(c) Corn — bias-corrected estimate.

Figure 3.- Proportion estimates using ground-truth labeling as input.



(d) Soybeans — dot estimate.

(e) Soybeans — classification estimate.

(f) Soybeans — bias-corrected estimate.

Figure 3.- Concluded.

TABLE 9.- U.S. CORN AND SOYBEANS EXPLORATORY EXPERIMENT — CLASSIFICATION ERRORS USING GROUND-TRUTH LABELS AS INPUT

Source of classification	Corn			Soybeans		
	Mean error	Standard deviation	Mean square error	Mean error	Standard deviation	Mean square error
Type 2 dots as random sample	1.55	5.19	28.3	1.00	4.14	17.5
Machine classification	8.21	8.98	144.7	-2.28	5.63	35.6
Bias-corrected machine classification	1.00	4.07	17.0	0.47	3.08	9.3

TABLE 10.- U.S. CORN AND SOYBEANS EXPLORATORY EXPERIMENT — CLASSIFICATION IMPROVEMENT USING GROUND-TRUTH LABELS AS INPUT

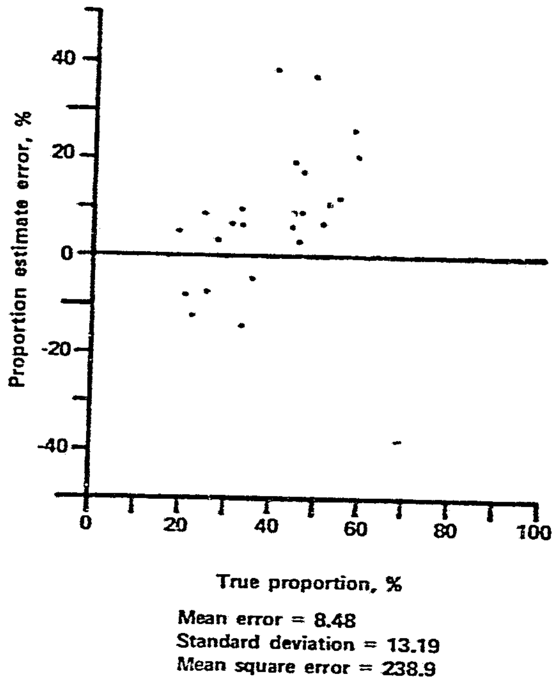
Classification sources compared	Corn		Soybeans	
	Processing improved, %	Mean improvement	Processing improved, %	Mean improvement
Machine classification vs. type 2 dots	20	-5.05	36	-1.26
Bias correction vs. machine classification	76	5.70	76	2.24
Bias correction vs. type 2 dots	60	0.65	64	0.98

improvement is not great enough to warrant the effort involved in performing the machine classification.

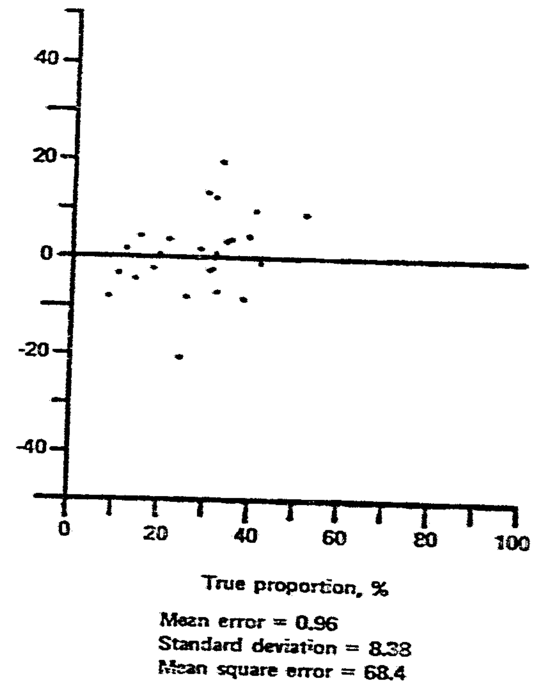
The most interesting feature of the ground-truth-based classification results is the large mean error in the machine classification proportions for corn. The plot in figure 3 shows that the error increases with increased true proportion. In fact, the mean square error of 144.7 (table 9) is larger than the mean square error of 103.8 for the analyst-based machine classification results (table 7). This indicates a serious problem with the procedure, since one would expect the results to improve or remain the same when true labels are substituted for analyst labels.

A possible source for the bias could be that the type 1 dots, used as input for the classification, are not representative of the entire segment. In order to determine if the type 1 dots are representative of the segment as a whole, a proportion estimate can be calculated using the type 1 dots as a random sample of the segment. If the type 1 dots are representative of the segment, the estimate should be unbiased. Figure 4 shows the proportion estimation error for the type 1 dots. As one might expect, the corn estimate has an 8.48-percent positive bias. This is very close to the bias of 8.21 percent in the classification estimate. The type 1 dot estimate shows the same trend as the classification estimate. Therefore, the type 1 dots are not representative of the segment, which is responsible for the bias in the classification results.

The question to consider now is: Why are the type 1 dots a biased sample of the segment? These dots are a set taken from a random grid; thus, the location should not produce a bias. One restriction was placed on the dots: that a dot which falls on a field boundary is not used. In this particular test, type 1 dots were used only if they were more than one-half pixel away from a field boundary. If the proportion is calculated using all of those pixels which meet the purity criterion and this estimate is biased with respect to the true proportion, then the purity restriction on the type 1 dots is the source of the observed bias. Figure 5 shows errors in the proportions based

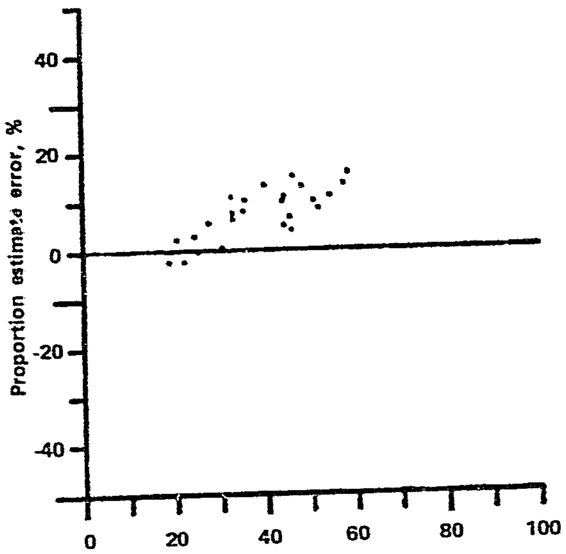


(a) Corn.



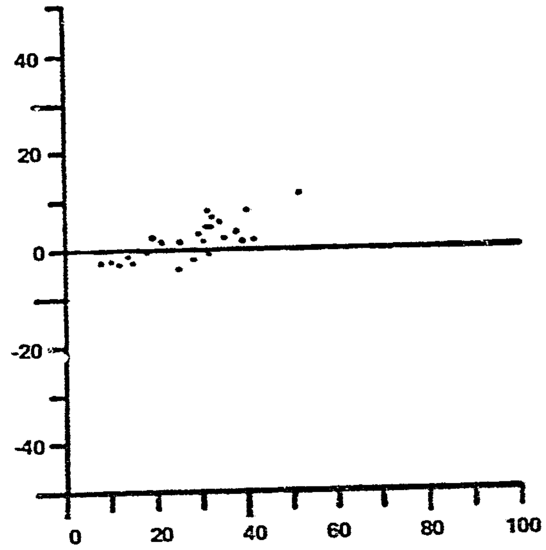
(b) Soybeans.

Figure 4.— Proportion errors based on type 1 dots with ground-truth labels versus true proportions.



Mean error = 7.61
Standard deviation = 5.21
Mean square error = 84.0

(a) Corn.



Mean error = 2.15
Standard deviation = 3.99
Mean square error = 19.9

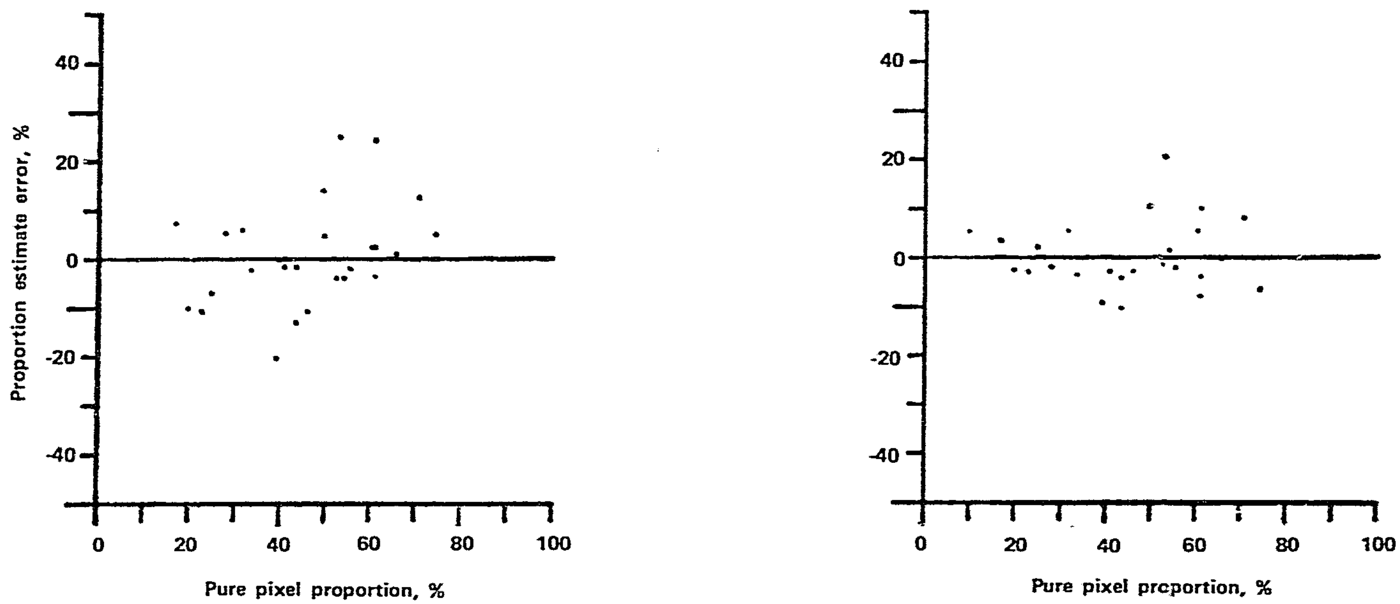
(b) Soybeans.

Figure 5.— Error in proportion estimate based on all pure pixels in scene versus true proportion. (A pure pixel is a pixel which is more than one-half pixel away from the field boundary.)

on all pure pixels in the segment as a function of the true proportion. The proportion errors for corn show the same trend to greater error with increased proportion, as seen in the type 1 dot proportion and classification results. The mean error for corn is 7.61 percent, which is consistent with the errors observed for the type 1 dot and classification estimates.

The conclusion from this analysis would be that the type 1 dots are more representative of the pure pixels in the scene than of the entire scene. Since the pure pixels are a biased sample of the segment, the proportions based on the type 1 dots and on the classification will also be biased. One way of verifying this conclusion is to compare the proportion estimates with the ground-truth proportions based on pure pixels. If the mean error, standard deviation, and mean square error are less when the pure pixel ground-truth proportion is used rather than the entire scene ground-truth proportion, then the proportions are more representative of the pure pixels than of the entire scene. Figure 6 shows the results of these comparisons. The corn estimates do not show the large positive bias evident when the entire scene proportion is used as the true proportion. The mean errors, standard deviations, and mean square errors corresponding to figure 6 are presented in table 11. The mean errors for the corn estimates are reduced from more than 8 percent to less than 1 percent. There was a slight reduction in the standard deviation. The mean square error was reduced by 50 percent or more. The results for soybeans were not as straightforward as those for corn. Although the mean square error for the type 1 dots decreased slightly when pure pixel proportions were used, the mean square error for the classification actually increased. These changes are not significant because the pure pixel and entire scene ground-truth proportions were close.

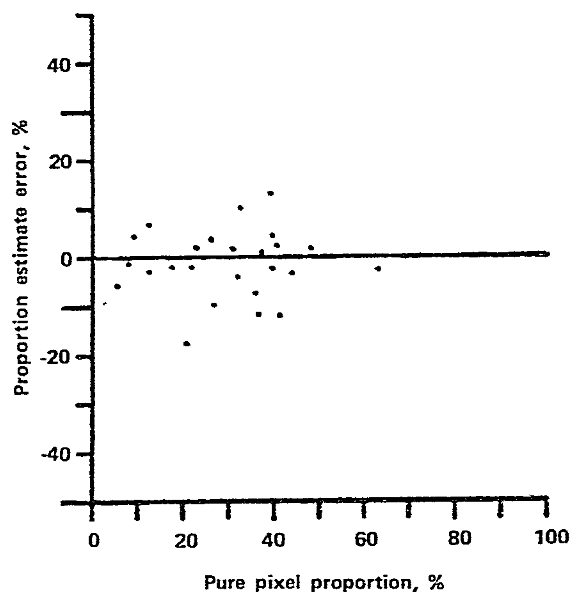
The bias and about one-half of the variability in the proportion estimates are the result of analyst dot labeling errors. A summary of the analyst dot labeling accuracy is shown in tables 12 and 13. The overall accuracy for type 1 dot labeling was 86 percent, whereas the accuracy for type 2 dot labeling was 75 percent. This is probably a consequence of the fact that all of the type 1 dots were pure, whereas type 2 dots could be impure. One can



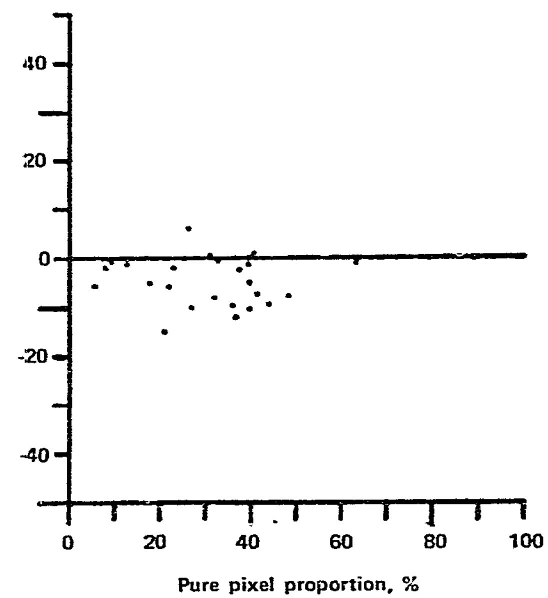
(a) Corn — type 1 dot estimate.

(b) Corn — classification estimate.

Figure 6.- Proportion estimate error using pure pixel proportions as true proportions (based on ground-truth labels for dots) versus pure pixel proportions.



(c) Soybeans — type 1 dot estimate.



(d) Soybeans — classification estimate.

Figure 6.- Concluded.

TABLE 11.- EFFECT OF USING PURE PIXEL GROUND-TRUTH PROPORTIONS ON CLASSIFICATION ERRORS

Crop	Source of classification estimate	Source of ground-truth proportion	Mean error	Standard deviation	Mean square error
Corn	Type 1 dots as random sample	Entire scene	8.48	13.19	238.9
		Pure pixels	.93	10.69	110.6
	Machine classification	Entire scene	8.21	8.98	144.7
		Pure pixels	.66	7.32	51.9
Soybeans	Type 1 dots as random sample	Entire scene	.96	8.38	68.4
		Pure pixels	-1.18	6.97	48.0
	Machine classification	Entire scene	-2.28	5.63	35.6
		Pure pixels	-4.41	4.93	42.8

TABLE 12.- DOT LABELING ACCURACY FOR TYPE 1 DOTS

Crop	Dots labeled corn	Dots labeled soybeans	Dots labeled "other"	Dots correctly labeled, %
Corn	647	34	71	86
Soybeans	54	392	52	79
"Other":				
Wheat	3	0	23	88
Oats	1	0	8	89
Grass	0	1	7	88
Hay	3	2	40	89
Pasture	7	1	138	95
Trees	6	1	142	95
Clover	0	0	9	100
Vegetable	0	0	2	100
Water	0	0	14	100
Nonagriculture	1	3	41	91
Homestead	1	0	27	96
Idle	3	2	35	88
Total "other"	25	10	486	93

TABLE 13.- DOT LABELING ACCURACY FOR TYPE 2 DOTS

Crop	Dots labeled corn	Dots labeled soybeans	Dots labeled "other"	Dots correctly labeled, %
Corn	1598	124	456	73
Soybeans	231	1014	341	64
"Other":				
Wheat	11	11	93	81
Oats	14	3	64	79
Grass	6	3	22	71
Hay	6	8	124	90
Pasture	47	18	421	87
Trees	18	8	343	93
Clover	4	2	5	45
Vegetable	0	0	9	100
Water	2	0	35	95
Nonagriculture	12	10	131	86
Homestead	7	6	95	88
Idle	21	13	119	78
Total "other"	148	82	1461	86

explain the fact that the soybean proportion estimates based on classification results were better than those based on the type 2 dots when analyst labels were used. Although the classification estimates are usually less accurate, the better labeling for the type 1 dots was enough to improve the classification results. In looking at the confusion between the categories (corn, soybeans, and "other"), it appears that there is greater confusion between corn and "other" than between corn and soybeans.

In order to determine how well the clustering algorithm is working in separating the crop of interest from a noncrop, the cluster purities were calculated for corn and for soybeans. Histograms of cluster purity are shown for corn and soybeans in figures 7 and 8. The number of clusters with given crop proportions is plotted as a function of the crop proportion. Ideally, these histograms should show two maxima (at 0 percent and 100 percent) representing pure noncrop and crop clusters. The histogram should be zero at the center. In the figures, one does see the expected two maxima with a minimum of approximately 50 percent. The crop maximum is fairly broad, but it appears that the clustering algorithm is separating crop and noncrop pixels to a certain extent.

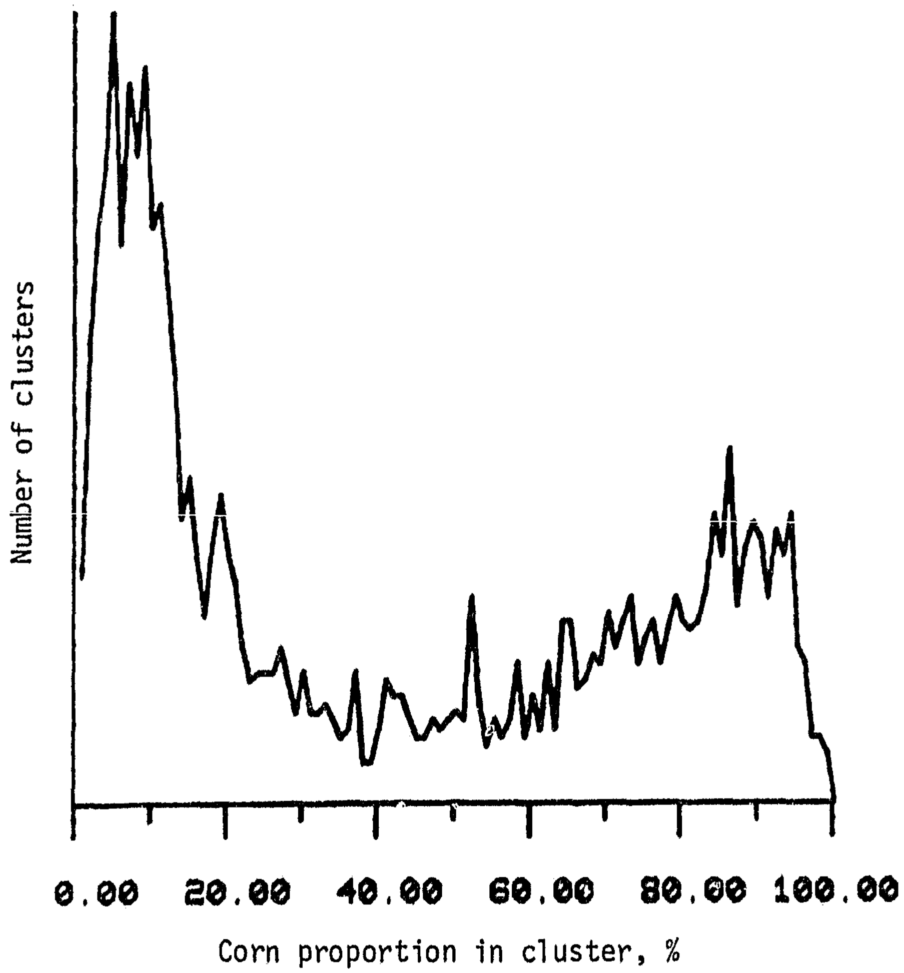


Figure 7.- U.S. Corn and Soybeans Exploratory Experiment
histogram of cluster purity for corn.

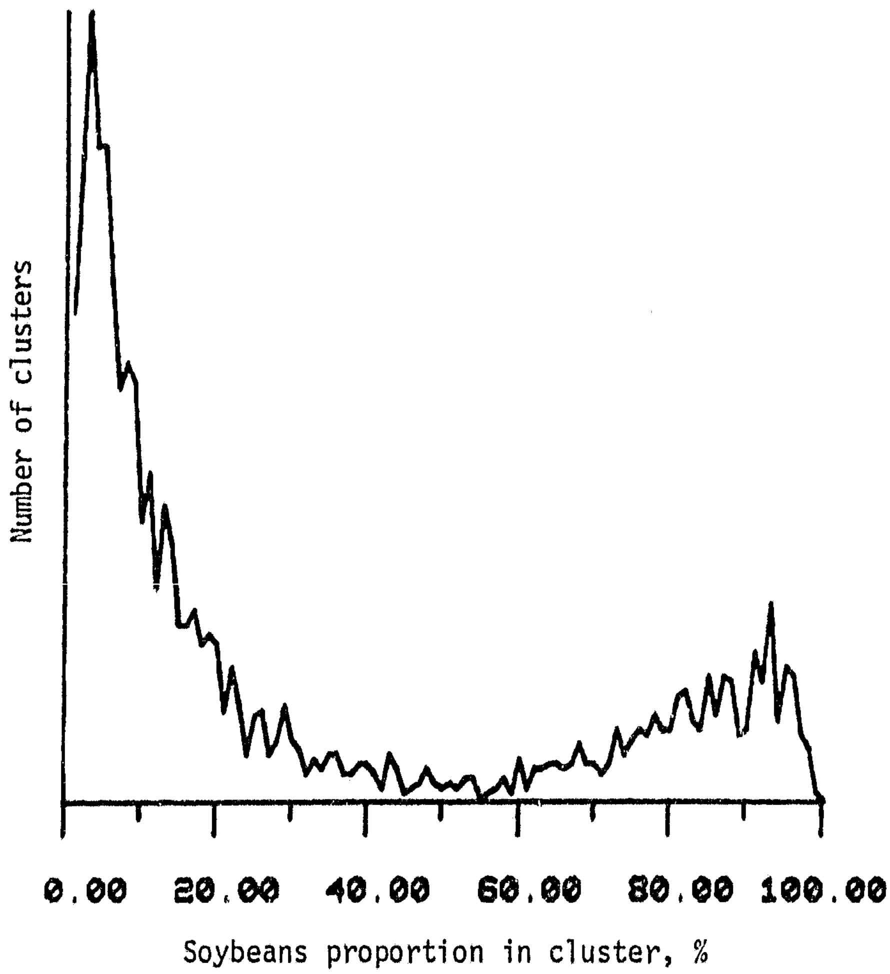


Figure 8.- U.S. Corn and Soybeans Exploratory Experiment histogram of cluster purity for soybeans.

4. SUMMARY OF RESULTS

Based on the studies presented in this document, the following conclusions can be reached:

a. The proportion estimates for corn had a bias of -4 percent with a standard deviation of 8 percent.

b. The proportion estimates for soybeans had a bias of -6 percent with a standard deviation of 7 percent.

c. The bias and about one-half the standard deviation for both corn and soybeans were the result of dot labeling errors.

d. Proportion estimates based on the type 2 dots as a random sample are as good as the final bias-corrected results.

e. The machine classification results are identical to the machine clustering results.

f. The large bias observed in the classification proportions for corn (when true labels are used) is caused by bias in the type 1 dots used as input to the classification procedure.

g. The bias in the type 1 dots was present because the type 1 dots were required to be pure.

h. Although the three groups used to process the segments were given identical training and used identical procedures, one group had significantly different dot labeling accuracy.

i. It is more difficult to label "other" dots in APU 14 than it is in APU's 25 and 28.

5. RECOMMENDATIONS

Dot labeling errors are the greatest source of error in the proportion estimates. If the quality of the proportion estimates is to be improved, the current dot labeling techniques need to be improved or an alternative for dot labeling found.

Since the machine processing used in this test does not significantly improve the accuracy of the corn and soybeans proportion estimates, the proportion estimates can be made using the labeled dots as a random sample of the segment. Alternatives to the machine processing technique used in this test should be investigated to see if a more effective technique can be found.

Since the maximum likelihood classification results are identical to the results using labeled clusters, it is not necessary to perform the maximum likelihood classification. The proportion estimates based on the clustering results should be bias corrected using a random dot set so that the kind of bias reflected in the corn proportion estimates can be reduced.

6. REFERENCES

1. Detailed Analysis Procedures for Transition Project (FY 79). LACIE-00724, JSC-13756, NASA Lyndon B. Johnson Space Center (Houston), May 1979.

2. LACIE Transition Project Accuracy Assessment Fiscal Year 1979 Interim Plan. LACIE-00634, JSC-13770, NASA/Lyndon B. Johnson Space Center (Houston), Aug. 1979.

☆ U.S. GOVERNMENT PRINTING OFFICE: 1980—671-099/25 18

APPENDIX B

CORN/SOYBEAN DECISION LOGIC DEVELOPMENT AND TESTING

AgRISTARS

FC-L0-00480
JSC-16380

A Joint Program for Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing

October 1980

Foreign Commodity Production Forecasting

CORN/SOYBEAN DECISION LOGIC DEVELOPMENT AND TESTING

C. L. Dailey and K. M. Abotteen

LOCKHEED ENGINEERING AND MANAGEMENT SERVICES COMPANY, INC.
1830 NASA Road 1, Houston, Texas 77058

Lyndon B. Johnson Space Center
Houston, Texas 77058

FC-LO-00480
JSC-16380

TECHNICAL REPORT

CORN/SOYBEAN DECISION LOGIC DEVELOPMENT AND TESTING

Job Order 73-315

This report describes labeling logic development activities performed by the classification element of the Foreign Commodity Production Forecasting project of the AgRISTARS program.

PREPARED BY

C. L. Dailey and K. M. Abotteen

APPROVED BY

R. W. Payne
for L. M. Flores, Supervisor
Design Integration Section

B L Carroll
B. L. Carroll, Manager
Commodity Forecasting Department

LOCKHEED ENGINEERING AND MANAGEMENT SERVICES COMPANY, INC.
Under Contract NAS 9-15800

For

Earth Observations Division
Space and Life Sciences Directorate
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
LYNDON B. JOHNSON SPACE CENTER
HOUSTON, TEXAS

October 1980

LEMSCO-14811

PRECEDING PAGE BLANK NOT FILMED

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
JSC-16380; FC-LO-00480		

4. Title and Subtitle	5. Report Date
Corn/Soybean Decision Logic Development and Testing	October 1980
	6. Performing Organization Code
7. Author(s)	8. Performing Organization Report No
C. L. Dailey and K. M. Abotteen Lockheed Engineering and Management Services Company, Inc.	LEMSCO-14811
	10. Work Unit No.
9. Performing Organization Name and Address Lockheed Engineering and Management Services Company, Inc. 1830 NASA Road 1 Houston, Texas 77058	11. Contract or Grant No. NAS 9-15800
	13. Type of Report and Period Covered Technical Report
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Lyndon B. Johnson Space Center Houston, Texas 77058 Technical Monitor: R. O. Hill	14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

This paper shows the development and testing of an analysis procedure which was developed to improve the consistency and objectivity of crop identification using Landsat data. The procedure was developed to identify corn and soybean crops in the U.S. Corn Belt region. The procedure consists of a series of decision points arranged in a tree-like structure, the branches of which lead an analyst to crop labels. The specific decision logic is designed to maximize the objectivity of the identification process and to promote the possibility of future automation. In this report, development and testing of the procedure are outlined and a summary of significant results is presented.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
Corn, crop identification, decision logic, hierarchy, labeling logic development, labeling procedures, soybeans	

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price*
Unclassified	Unclassified	82	

PREFACE

This report offers a detailed description of the decision logic and procedure developed for identification of corn and soybeans in the U.S. Corn Belt. Development and testing of the procedure are outlined and a summary of significant results is presented.

The development and testing of the corn/soybean decision logic procedure was a team effort which required the expertise of many individuals. The major effort of designing the hierarchical structure of the decision logic was coordinated by W. P. Palmer, who documented the initial decision logic in an internal communication (section 5). Major sections of that document are reproduced in this report. J. D. Nichols and W. L. West analyzed image and ground-truth data and constructed the cropland identification step of the decision logic. T. E. Johnson, B. B. Schroder, and R. D. Pickerel developed the initial framework for the separation of corn and soybeans using image products of the Large Area Crop Inventory Experiment. W. W. Austin aided in the analysis of spectral aids. These individuals were major contributors to the development of the corn/soybean decision logic.

The authors would like to thank the analysts from both the National Aeronautics and Space Administration and Lockheed Engineering and Management Services Company, Inc. who participated in the tests. Also, the authors wish to thank J. G. Carnes for the preliminary test results which appear in this paper.

CONTENTS

Section | Page

TABLES

FIGURES

1. INTRODUCTION

This paper shows the development and testing of an analysis procedure which was developed to improve the consistency and objectivity of crop identification using Landsat data. The procedure was developed to identify corn and soybean crops in the U.S. Corn Belt region. The procedure consists of a series of decision points arranged in a tree-like structure, the branches of which lead an analyst to crop labels. The specific decision logic is designed to maximize the objectivity of the identification process and to promote the possibility of future automation.

In prior procedures, the interpretation function was more loosely structured and many steps were very subjective. The analyst was responsible for accumulating information from various sources, assimulating and integrating the information in order to determine the most likely label for a signature. Labeling accuracies of these procedures were related to the experience of the analyst, and labeling errors were sometimes hard to diagnose.

This decision logic is a hierarchy of decisions that uses a step-by-step procedure to lead the analyst from general major land-use categories to the specific identification of corn and soybean signatures. In the first step, analysis of the signatures on the imagery is governed by answers given at decision points on the decision tree and results in the differentiation of cropland from other major land-use categories. In step two, image products are used to answer more specific questions to separate cropland into summer and nonsummer crops. In step three, summer crops are identified as definite corn and soybeans through the aid of numerical spectral information in graphic form. Any remaining signatures are labeled in step four by comparing them to definite corn and soybean profiles and choosing the label of the most similar profile. Each component of the decision logic will be further discussed in terms of its function, strengths, and weaknesses.

Two tests were performed to evaluate the decision logic. Labeling accuracies pertaining to the developmental task are summarized, and procedural problems and recommendations are discussed in this paper. The complete analysis of the accuracy of the tests is contained in an accuracy assessment report (ref. 1).

2. OBJECTIVES

This research effort was designed to develop and test a decision logic for corn and soybean identification. The objectives of the effort were to

- Define a tree-type structure of decision points that describes the image interpretation process
- Determine from all available analyst aids those to be used at various decision points
- Define a procedure so that labeling errors can be easily diagnosed
- Test the decision logic and obtain labeling results for further development

3. DATA SET

Eight segments (9- by 11-kilometer area), located in four agrophysical units (APU) of the U.S. Corn Belt, were used in developing the technique. Table 3-1 displays the segment numbers, locations, APU's, available acquisitions, and major crops. The data set is selected according to the following criteria:

a. Presence of the crops of interest (corn and soybeans)

b. Good acquisition histories

c. Availability of ground-truth data

The products available for analyst use include: (1) Landsat film products which are false color composites of three bands out of the four bands of the satellite's multispectral scanner (MSS), (2) crop calendars, (3) meteorological summaries, and (4) spectral aids in the form of plots of transformed spectral values from the MSS.

There are three types of film products: Product 1 is a simulated color-infrared (CIR) composite image using Landsat bands 4, 5, and 7 of the Landsat MSS (ref. 2); Product 2 is an enhanced image using Landsat bands 5, 6, and 7; and Product 3 is a simulated CIR composite image using Landsat bands 4, 5, and 7 with different gains and biases set to minimize color distortion. Each product is 196 pixels (picture elements) across and 117 lines down and is partitioned by a 10-by-10 grid system.

Two types of crop calendars were used. Normal crop calendars were generated for corn and soybeans within designated crop reporting districts (CRD's) in the corn belt. The calendars, as shown in figure 3-1, display the percentage (Y-axis) of a crop that is at or past a specific growth stage. The time (X-axis) is displayed in 15-day intervals throughout the growing season. These calendars are based on two or more years of historical data. Current-year crop calendars were constructed from actual field observations collected on approximately 10 fields per segment at various points throughout the growing season. The format of the current-year crop calendar is shown in figure 3-2.

TABLE 3-1.- THE DEVELOPMENT DATA SET

Segment	Location	APU	Acquisition date (Julian data)	Major crops
209	Gentry, Missouri	25	June 16 (167) July 4 (185) July 31 (212) Aug 8 (220) Aug 9 (221) Sept 4 (247) Sept 22 (265) Sept 23 (266) Oct 1 (274) Oct 19 (292)	Corn Soybeans Hay Pasture
211	Grundy, Missouri	25	June 15 (166) July 3 (184) July 21 (202) Aug 8 (220) Sept 4 (247) Sept 22 (265) Oct 1 (274) Oct 19 (292) Oct 28 (301)	Corn Soybeans Sorghum Hay Pasture
804	Marshall, Iowa	24	June 15 (166) Aug 17 (229) Sept 4 (247) Sept 22 (265) Oct 1 (274) Oct 19 (292)	Corn Soybeans Oats Pasture
824	Iroquois, Illinois	28	June 12 (163) Aug 5 (217) Aug 23 (235) Aug 31 (243) Sept 1 (244) Sept 9 (252) Sept 28 (271) Nov 2 (306) Nov 3 (307)	Corn Soybeans Oats Hay

TABLE 3-1.- Concluded.

Segment	Location	APU	Acquisition date (Julian data)	Major crops
854	Tippecanoe, Indiana	28	June 10 (161) July 26 (207) Aug 9 (221) Aug 21 (233) Aug 22 (234) Sept 8 (251) Sept 9 (252) Sept 26 (269) Sept 27 (270) Nov 2 (306) Dec 17 (351)	Corn Soybeans Clover Pasture
883	Palo Alto, Iowa	24	July 5 (186) July 23 (204) Aug 1 (213) Aug 10 (222) Sept 24 (267) Oct 20 (293) Oct 30 (303)	Corn Soybeans Hay Pasture
886	Pottawatomie, Iowa	14	June 16 (167) July 5 (186) July 23 (204) July 31 (212) Sept 6 (249) Sept 15 (258) Sept 24 (267) Oct 20 (293) Nov 7 (311)	Corn Soybeans Oats Pasture
892	Shelby, Iowa	14	June 16 (167) July 23 (204) Aug 9 (221) Sept 23 (266) Sept 24 (267) Oct 20 (293)	Corn Soybeans Oats Hay Pasture

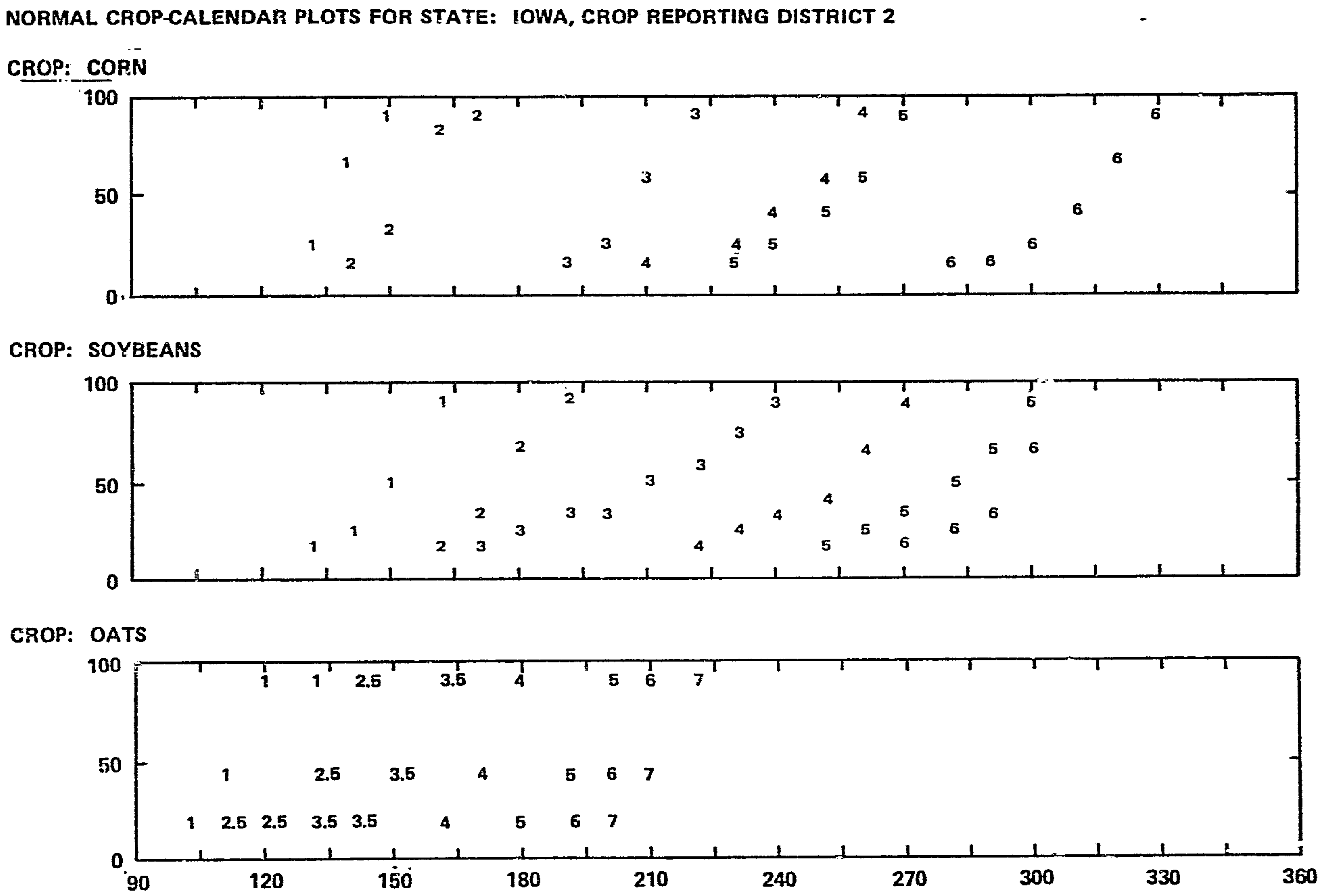


Figure 3-1.- Normal crop calendar.

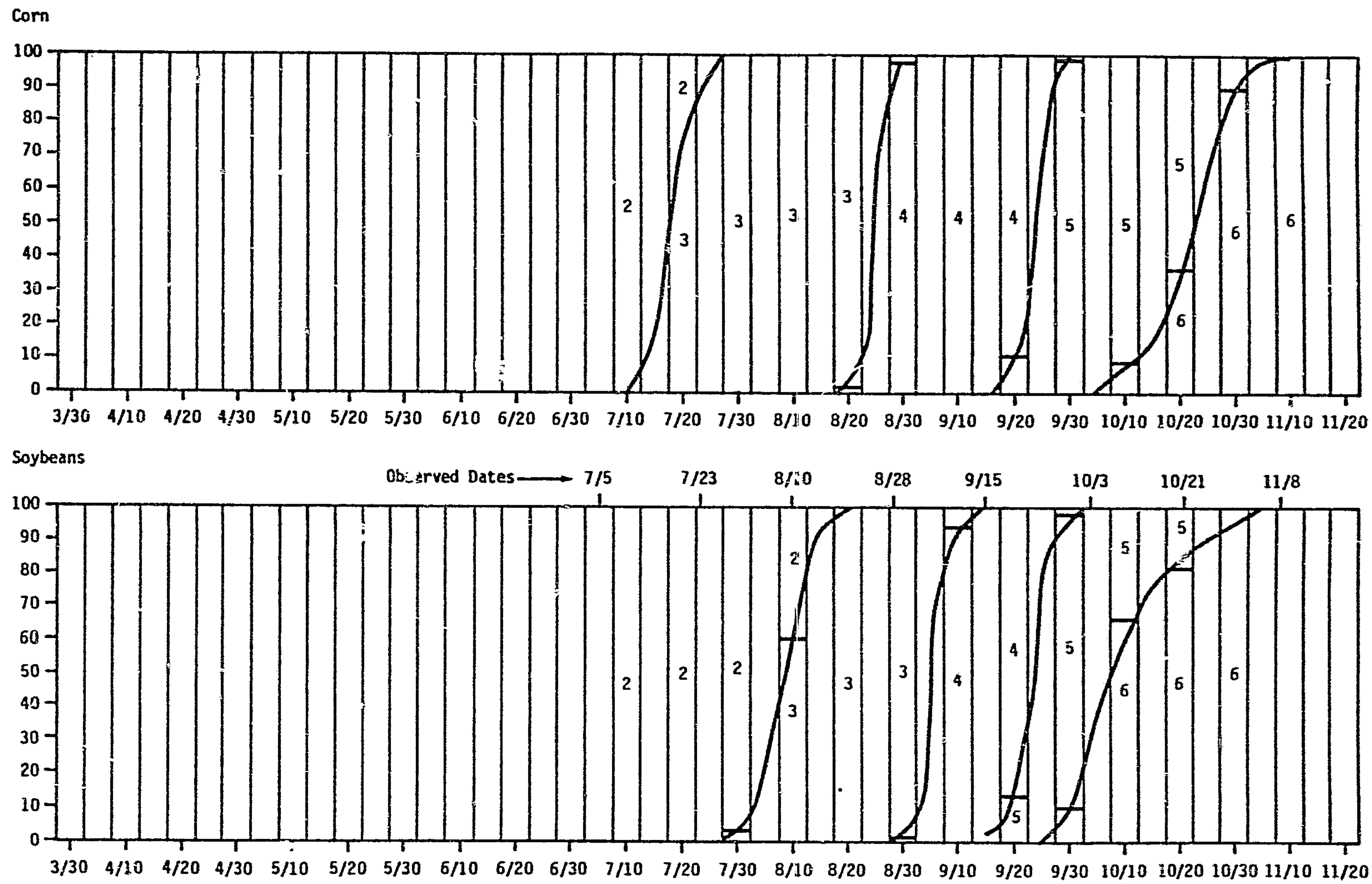


Figure 3-2.- Current year crop calendar for segment 883.

The meteorological summaries offer a synopsis of the weather at the state level and are available on a weekly basis.

Spectral aids which include scatter plots, time plots, and trajectory plots are generated before interpretation to aid in labeling. The data (209 grid intersection pixels called dots) are transformed into Kauth space before the aids are generated (ref. 3) and greenness is changed to green number by subtracting a calculated soil line (ref. 4).

The scatter plot in figure 3-3 is a graphic representation of the transformed MSS data. The typical green-number-versus-brightness scatter plot is triangular in shape. The base of the triangle contains the bare soil pixels. The distance of a pixel from the base is a measure of vegetation canopy and the distance that a pixel is from the Y-axis is a measure of its brightness. A scatter plot is generated for each acquisition in the data base.

Time plots display green number versus time and brightness versus time, as shown in figure 3-4. Two dots (pixels) are plotted per graph for every usable acquisition in the data base. Time plots show the changes in green number and/or brightness for a particular pixel over an entire growing season.

A trajectory plot displays a spectral pattern for a pixel over a period of time. It uses the same axes information as does a scatter plot, but it contains data on one pixel for up to eight acquisitions, as shown in figure 3-5.

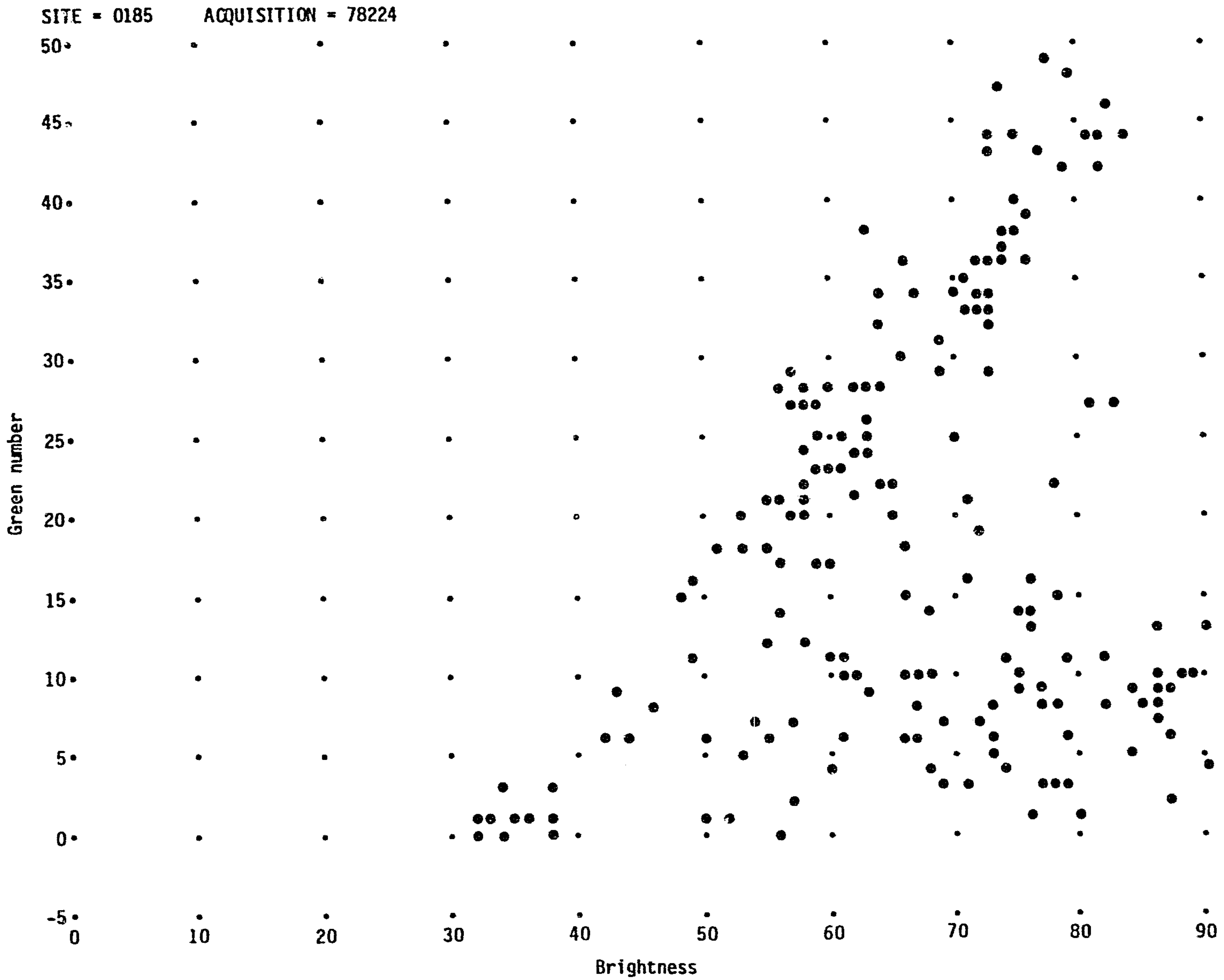


Figure 3-3.- Scatter plot.

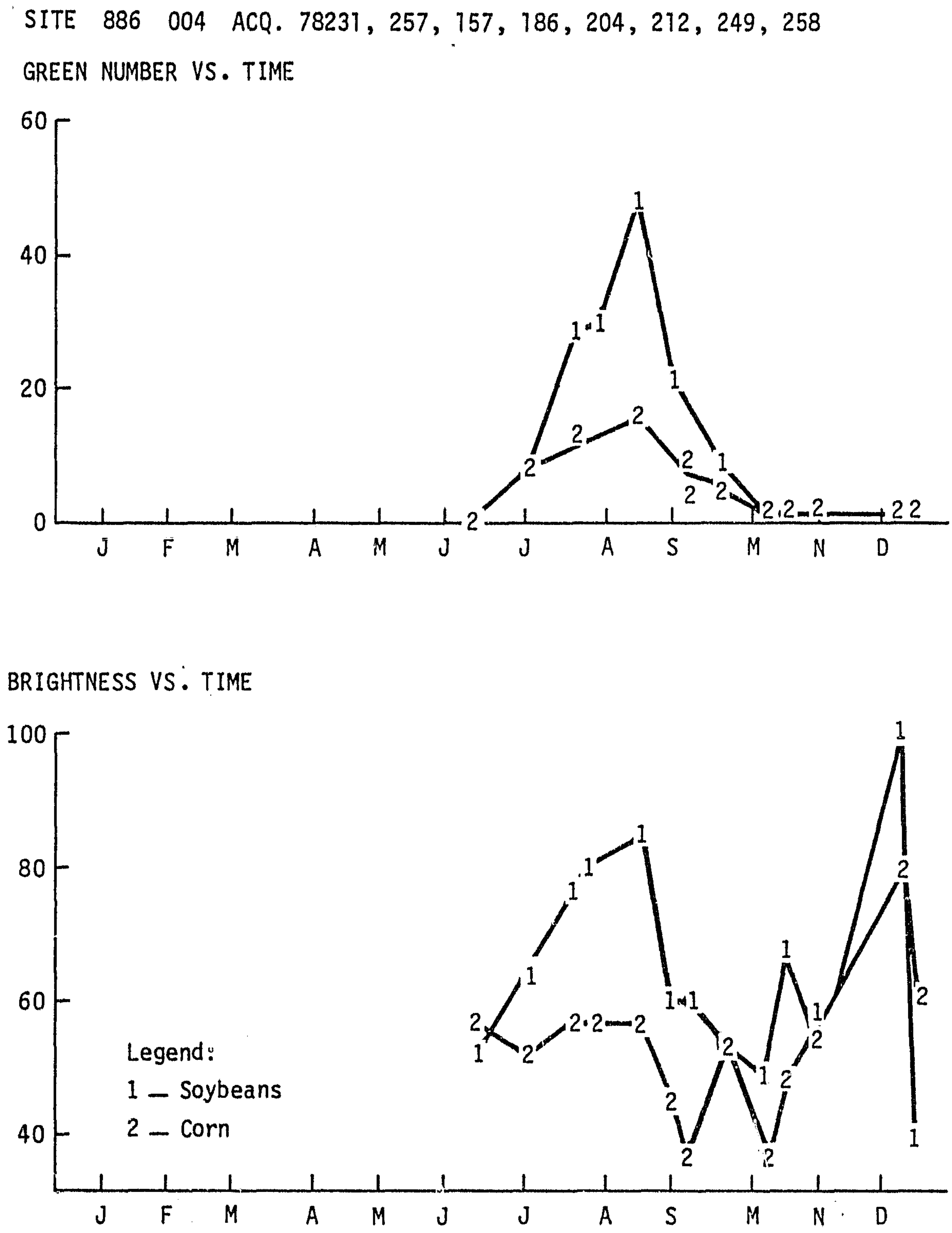
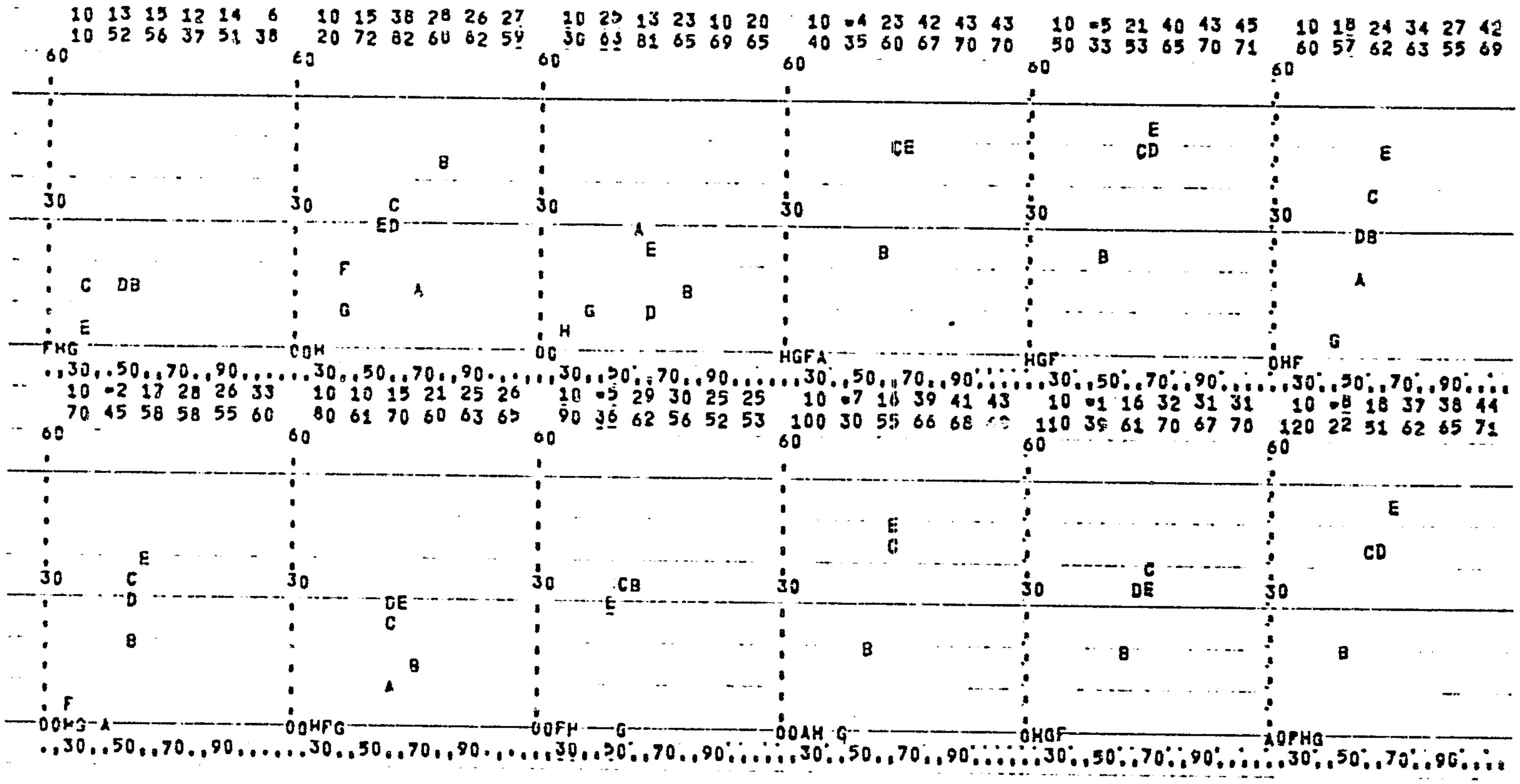


Figure 3-4.- Time plots for labeling dots.



SEGMENT	ACQ.	SYMBOL	SEGMENT	ACQ.	SYMBOL
883	78150	A	883	78222	E
883	78186	B	883	78267	F
883	78204	C	883	78293	G
883	78213	D	883	78303	H

Figure 3-5.-Trajectory plots. The grid intersection and green number and brightness values for the first five acquisitions are printed at the top of each plot. Due to the scale of the plot, values may fall in the same position on the plot and not be represented with a letter.

4. TECHNICAL APPROACH

The approach to the task (ref. 5) consisted of two phases. In the first phase, the then current procedures for labeling small grains (ref. 6) were examined for their applicability to the corn/soybeans case. Typically, these procedures consist in the examination of various alternative pieces of evidence to make a decision relating to land usage. Thus, the first step was to make this decision process more objective by eliminating the alternatives. Only one of the alternatives was selected for the decision. Then, the process was formalized by reformatting it in the form of decision points arranged in a tree-like structure. In the second phase, a separate effort was mounted to address the decision-making for the decisions that were more specifically related to corn and soybeans. These decisions were also formatted in a tree-structured approach.

In order to design the structure of each step of the second phase of the study, the different land uses and crop types were observed on each of the analyst aids to identify distinctive characteristics and trends. Ground-truth information was used when analyzing the film products and the spectral aids. Ground-truth labels were obtained from an annotated aerial photograph with a registered grid overlay. The grid overlay corresponds to the film product grid. The ground-truth pixels which were used for this study spectrally and spatially represent only one category (pure pixel).

Acquisition-specific information was collected and analyzed for corn and soybeans. Appendix A contains an explanation and table of that information. These data were then used to define biowindows and image characteristics of the corn and soybeans. The spectral aids were examined for patterns which would separate corn and soybeans from each other and from other crop types (ref. 7). Then each of the analyst aids were evaluated according to their suitability for use at specific decision points. Thus, a structure was built up using these objective observations to make decisions, each of which would be an element of the structure, and each branch or set of decisions would lead the analyst to a crop identification and label.

Two tests were performed using the corn/soybean decision logic. The first experiment was designed to identify problems with the procedure and provide for improvements before further testing. Labeling accuracies and the effects of the group (analyst) and region were addressed. The second test was designed to perform a within-strata variance study and estimate sampling and classification variance. This information would then be an input to a simulated aggregation. This test allowed for the use of the labeling logic in an operational-type environment. Only preliminary labeling results have been obtained on this second test.

5. DESCRIPTION OF THE DECISION LOGIC

The procedure developed from the analysis of the analyst aids available for the eight segments uses Landsat data in both imagery format and spectral aids as input. The logic diagram that leads to land usage and crop identification consists of four steps:

Step 1 — identification of cropland

Step 2 — identification of summer cropland

Step 3 — identification of definite corn and soybean signatures

Step 4 — identification of the remaining signatures

5.1 STEP 1 — IDENTIFICATION OF CROPLAND

Step 1 consists of the series of decision points arranged in the tree-like structure (decision tree) presented in figure 5-1. All workable simulated CIR Landsat acquisitions over the segment are used to sort the signatures in the scene into land-use categories. A minimum data set of two acquisitions is necessary for use of this tree. However, the decision tree is normally used in conjunction with the subsequent steps which impose more stringent requirements on the data set. The lowest level crop(s) of interest dictate the minimum data set.

To identify the land use associated with a particular signature, the analyst follows a path determined by the decisions given at the decision points encountered. The questions asked at each decision point are keyed by number, as shown in figure 5-1, and appear in figure 5-2. Each decision point is designed to use information extracted from the imagery based on the color of the crop in an acquisition in relation to the color in other acquisitions. The pathway thus defined allows for the identification of major land-use categories. Definitions and characteristics of categories identified in this step can be found in appendix B. Since definitions from other sources (ref. 8) combine categories that are separable with this procedure or alternatively include features which are too small to be detected on Landsat

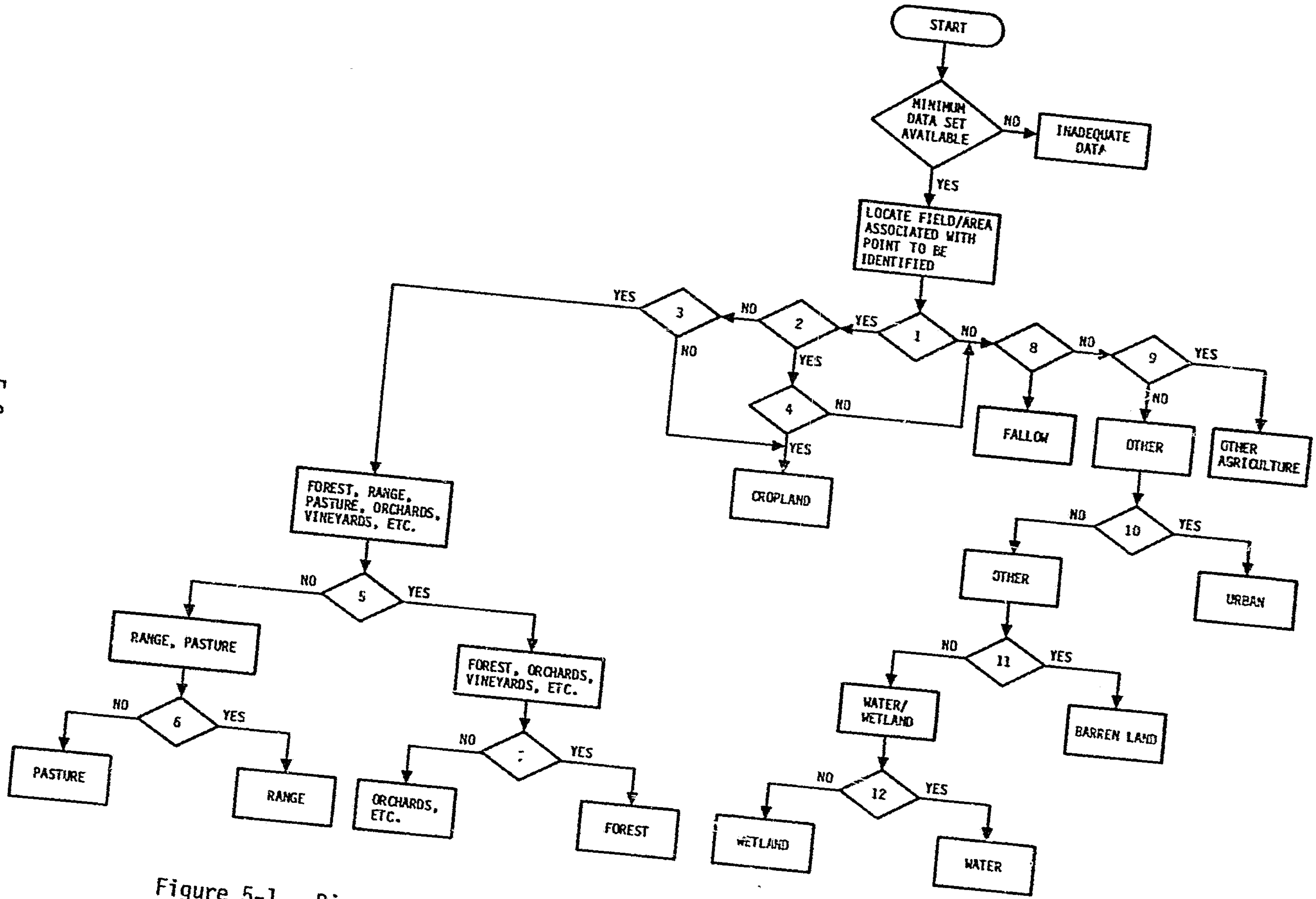


Figure 5-1.- Diagram of decision tree for major land-use categories (Step 1).

DECISION CRITERIA FOR MAJOR LAND-USE CATEGORIES

1. Is the area some shade of red (red, pink, brown, orange, etc.) on at least one acquisition?
2. Does the area appear to be water (dark blue-black to bright blue) on any of the acquisitions?
3. Is the area some shade of red on all acquisitions (i.e., no planting or harvest appearance)?
4. Is the area harvested (blue, green, white, gray, yellow) on an acquisition following the one in which it appeared red?
5. Is the area red or reddish brown throughout the year, with the color most intense during the late spring or early summer? (Some trees lose their leaves annually and may appear dark brown during the winter.)
6. Is the area large and irregular?
7. Is the area large relative to the economic endeavor of the area, along a drainage network, and bright red in late spring and early summer and reddish brown or brown at other times?
8. Is the shape of the area similar to areas that have been identified as cropland and the color green or blue (may vary from dark to light during the year) on all acquisitions?
9. Is the area small and white to dull gray?
10. Is the area irregular in shape and a constant white to mottled steel blue throughout the year?
11. Does the area appear to be constantly bright with no green vegetation and no seasonal change in shape or size?
12. Does the area appear dark blue-black to bright blue on all acquisitions? (Size and shape may change during year, but area is not seasonally wet.)

Figure 5-2.- Decision criteria questions keyed to the decision points in figure 5-1.

imagery, definition of the categories as used in the decision tree are necessary. All major land-use categories are labeled except for cropland which will be refined through further analysis. Labels are always associated with the dot which represents the area and signature being identified.

5.2 STEP 2 — IDENTIFICATION OF SUMMER CROPLAND

The signatures identified as cropland in Step 1 are separated into summer and nonsummer cropland by following Step 2. In order to perform this step, three biowindows are defined using the corn and soybean historical crop calendars, the 18-day ground truth observations, and Landsat CIR film products. (The ground truth observations are used only for development; ground truth information is not available during testing.) A biowindow is a time in the growth cycle of a crop when predictable Landsat signatures can be identified. Corn and soybean biowindows are described in table 5-1, and crop growth stage numbers for corn and soybeans are shown in table 5-2.

Figure 5-3 is a display of the crop calendar annotated with the defined biowindows. Figure 5-4 is the flow diagram for separating summer and nonsummer cropland. Fields that are bare soil (not red on imagery) on at least one acquisition in biowindow A, green vegetation (red on imagery) on all acquisitions in biowindow B, and ripe and/or harvested (not red on imagery) on all acquisitions in biowindow C are identified as summer crops. The nonsummer crop signatures are labeled at this point and the summer crop signatures are further processed in Step 3.

Dots which represent more than one signature either as a boundary between two categories or because of misregistration between acquisitions are identified and appropriately documented during this step because this is usually the last step that requires film products. Misregistered dots may be reserved for labeling in Step 4.

TABLE 5-1.- CORN AND SOYBEAN BIOWINDOWS

Bio-window	Definition[a]		Description of expected Characteristics
	Open on latest	Close on earliest	
A	C 30%>1 S 30%>1	C 80%>2 S 10%>2	Plowing, planting, pre-emergence, or very early emergence for summer crops
B	C 50%>3 S 10%>3	C 30%>5 S 10%>5	Full ground cover and green vegetation for summer crops
C	C 100%>5 S 100%>5	C 80%>6 +30 days S 80%>6 +30 days	Mature, harvest, and post-harvest for summer crops

[a]For example, entry C 30%>5 means that, according to the normal crop calendar, corn is 30 percent past stage 5 (maturity). Dates should be determined for both corn and soybeans and the latest used to open windows, the earliest to close windows.

TABLE 5-2.- GROWTH STAGE NUMBERS FOR CORN AND SOYBEANS

Growth stage number	Corn growth stage	Soybean growth stage
0	Plowing	Plowing
1	Planting	Planting
2	Floral initiation	Rapid nodal development
3	Tassel-silk	Full pod
4	Denting	Full seed
5	Maturity	Maturity
6	Harvest	Harvest

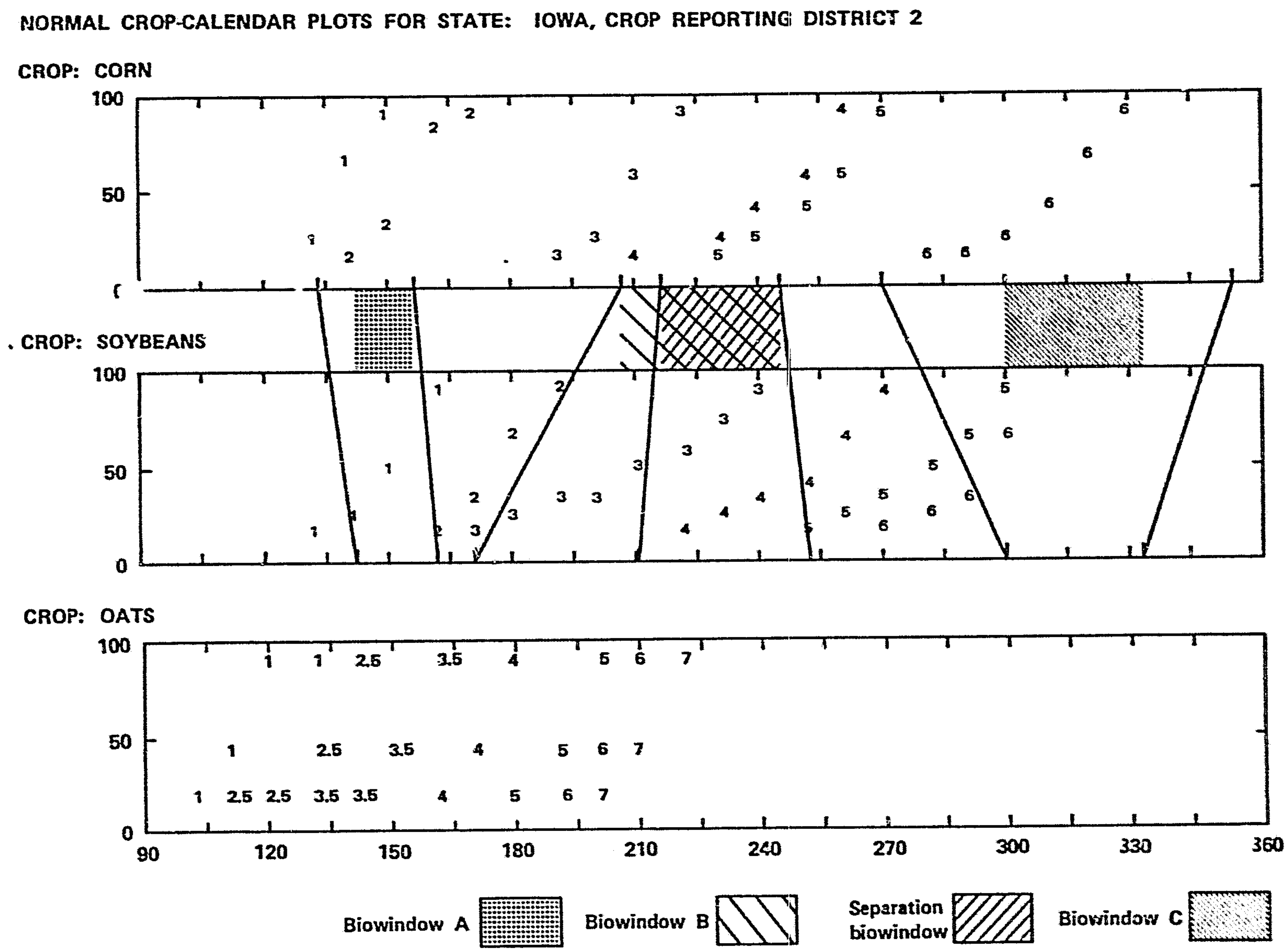


Figure 5-3.- Crop calendar annotated with biowindows.

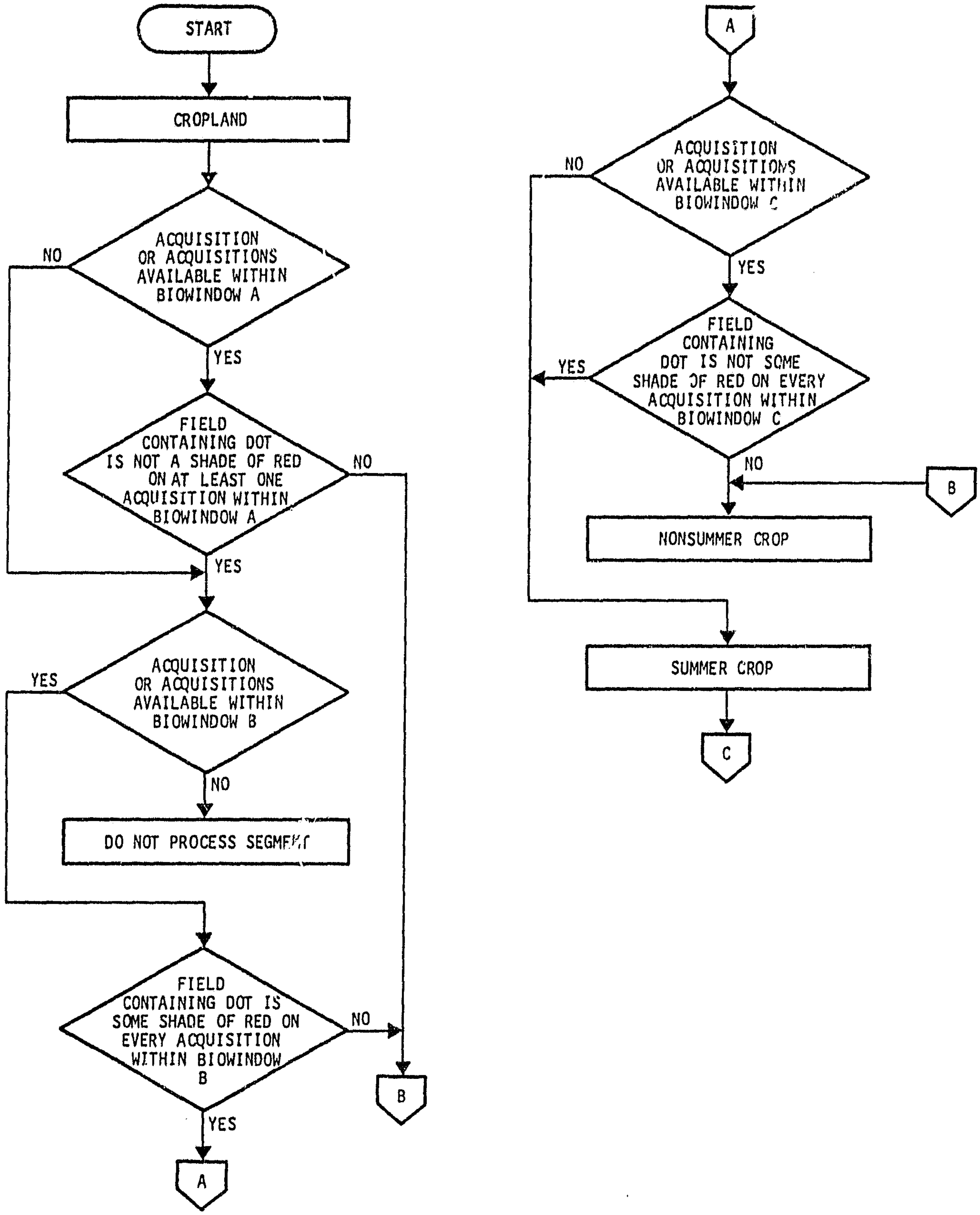


Figure 5-4. - Diagram of decision logic for summer and nonsummer cropland separation (Step 2).

C-2

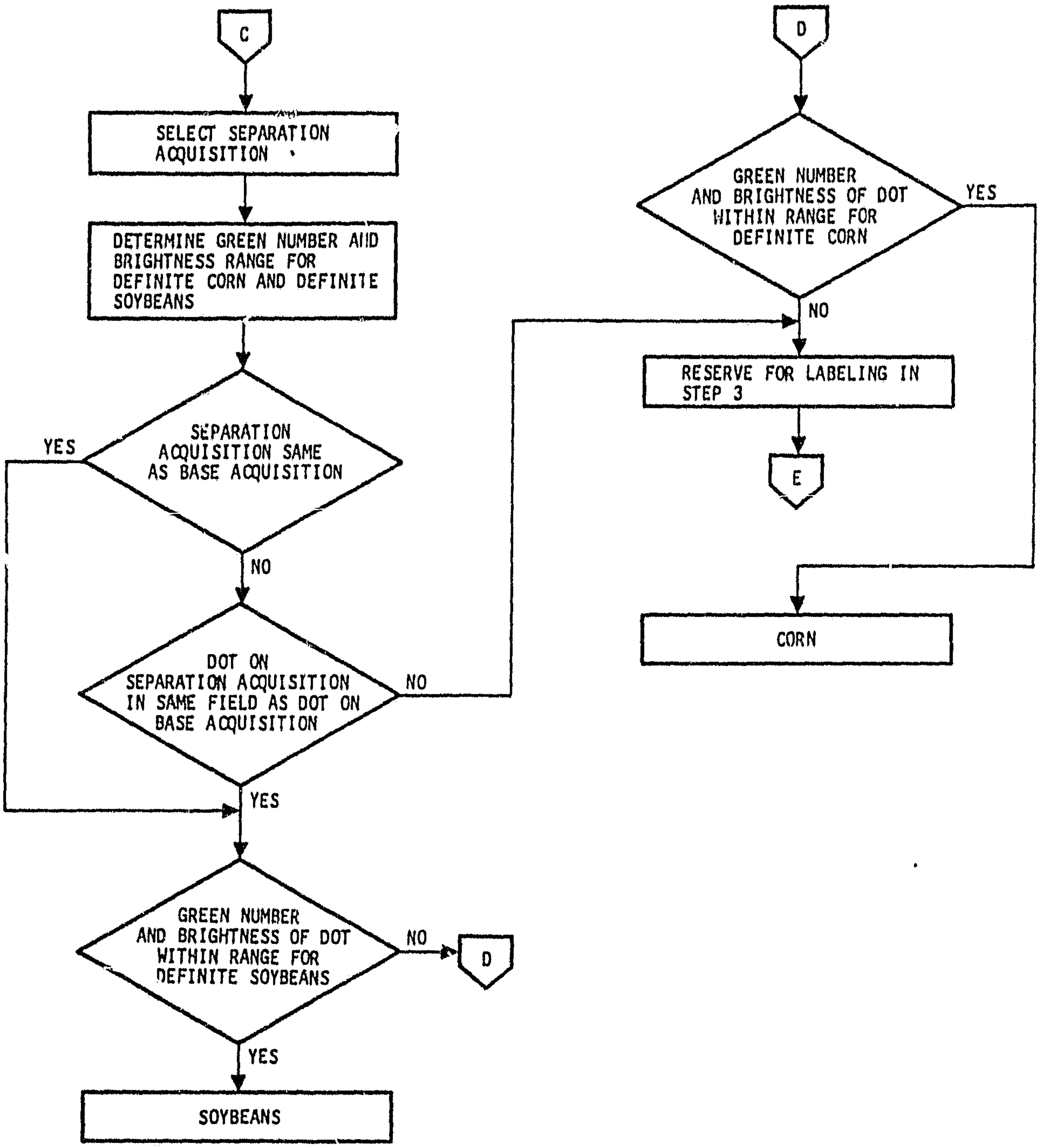


Figure 5-5.- Diagram of decision logic for identifying definite corn and soybeans (Step 3).

5.3 STEP 3 — IDENTIFICATION OF DEFINITE CORN AND SOYBEAN SIGNATURES

The logic flow of this step is diagrammed in figure 5-5. A minimum data set is required for identifying corn and soybeans. Two acquisitions are necessary, one acquisition in either biowindow A or biowindow C and one acquisition in a subset of biowindow B, called a separation biowindow, and defined as shown in the following table.

Definition		Description of expected characteristics
Open on latest	Close on earliest	
C 90%>3 S 50%>3	C 30%>5 S 10%>5	Most of the corn is in the denting stage, and most of the soybeans are in the full pod stage.

A green-number-versus-brightness scatter plot of 209 unlabeled dots selected by systematic random sampling from within the scene is generated for each acquisition in the separation biowindow. An analyst team (3 to 5 analysts) determines which acquisition has the best separation or natural break in the data. Lines are drawn through the break in the data that best separates the two groupings. One of the groupings will be associated with corn and the other with soybeans. The lines are constrained to be parallel to the x and y axes. Then, five counts are added and subtracted from the lines, as shown in figure 5-6. The shaded area accounts for areas of over-lapping categories. All summer crop dots that fall outside the limits in quadrant 1 are labeled soybeans, and all summer crop dots that fall outside the limits in quadrant 3 are labeled corn. Table 5-3, which shows the green number and brightness table generated with the scatter plot, is used to expedite this process. All dots within the limiters (shaded area) are reserved for labeling in Step 4 along with misregistered dots.

5.4 STEP 4 — IDENTIFICATION OF THE REMAINING SIGNATURES

Two methods of analyzing the remaining dots are represented in the flow diagram (figure 5-7) depending on the type of dot being labeled. If the dot is misregistered (edge dot), then the area the dot is in on the base

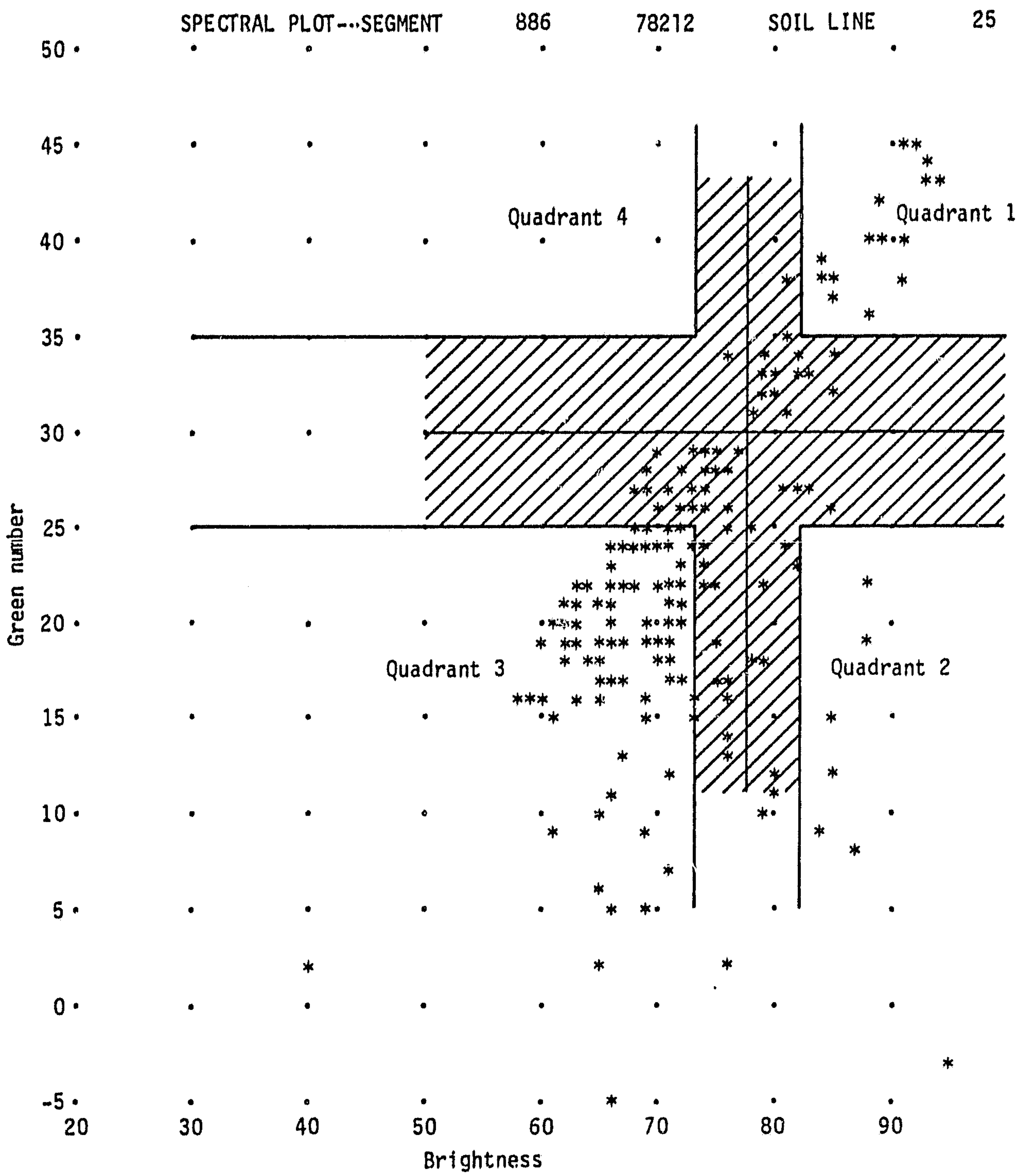


Figure 5-6.- Delineation of break in data and limiters on scatter plot for Step 3.

TABLE 5-3.- SCATTERPLOT TABLE SHOWING EXAMPLES OF STEP 3 AND 4 DOT VALUES

Dot number	Lire	Pixel	Label	Green number	Brightness number
1	1	1	**	43	[a]94
2	1	2	**	21	[b]67
3	1	3	**	16	[b]68
4	1	4	**	31	[c]81
5	1	5	**	39	[a]86
6	1	6	**	27	[c]74
7	1	7	**	-7	66
8	1	8	**	32	80
9	1	9	**	23	69
10	1	10	**	2	76
[b]11	[b]1	[b]11	[b]**	27	74
12	1	12	**	36	88
13	1	13	**	14	61
14	1	14	**	16	68
[b]15	[b]1	[b]15	[b]**	23	72
16	1	16	**	16	75
17	1	17	**	27	82
18	1	18	**	19	70
[b]19	[b]1	[b]19	[b]**	24	72
20	2	1	**	15	69
21	2	2	**	43	93
22	2	3	**	18	67
23	2	4	**	-3	95
24	2	5	**	16	76
25	2	6	**	24	69
26	2	7	**	27	67
27	2	8	**	26	74
28	2	9	**	21	61
29	2	10	**	40	89
30	2	11	**	21	72
31	2	12	**	22	67
32	2	13	**	40	91
33	2	14	**	19	66
34	2	15	**	18	70
35	2	16	**	38	86
36	2	17	**	8	87
37	2	18	**	34	85

[a]Soybeans
[b]Corn
[c]Step 4 dot

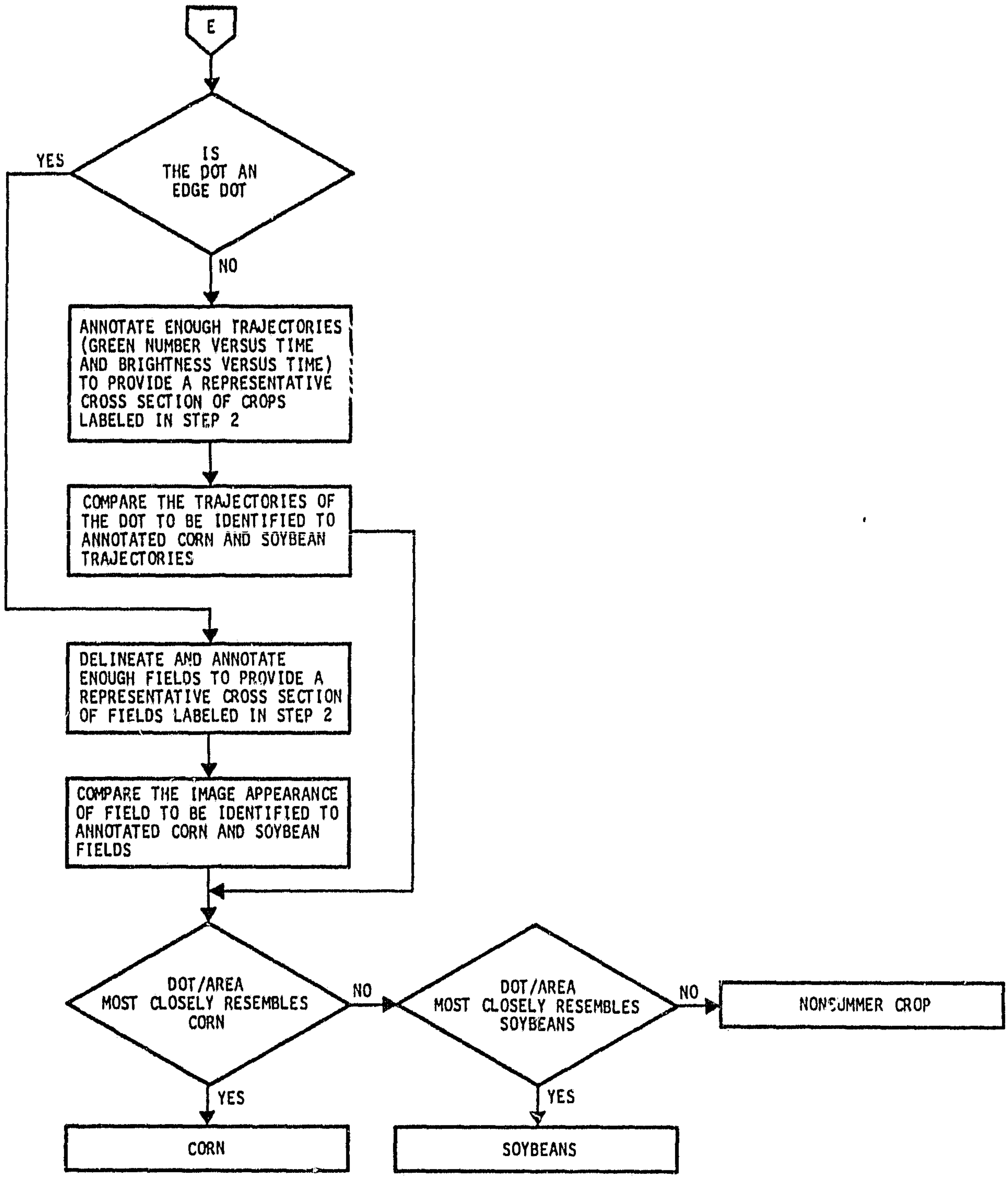


Figure 5-7. - Diagram of decision logic for labeling remaining dots (Step 4).

acquisition is compared with areas of known corn and soybeans and labeled according to the area it most closely resembles. Green number and brightness are plotted versus time for all acceptable (cloud- and haze-free) acquisitions to aid the analyst in labeling the dots that fell within the limiters. These time profiles are obtained for all previously labeled and unlabeled samples. In Step 4, the analyst compares corn and soybean profiles labeled from Step 3 with the profiles of the yet unlabeled dots. The unlabeled profiles are then labeled by assigning them the label of the most similar profile.

PRECEDING PAGE BLANK NOT FILMED

6. SUMMARY OF TESTS AND RESULTS

The two tests conducted using the corn/soybean decision logic procedure were the Multicrop Exploratory Experiment (ref. 1) and the Simulated Aggregation Test. In the first test, the objectives were to shake down the procedure and to determine if the procedure is analyst dependent. The objectives of the second test were to test the procedure that resulted after modifications based on the first test were included and to provide information such as segment number, location, acquisitions used, defined biowindows, and the separation point for the data sets used is presented in appendix C.

For the multicrop test, a rigid design plan was followed using three groups of analysts and preselected segments and acquisitions. Each segment was worked by at least two groups. In the simulated aggregation test, three analyst teams (group I, group II, and group III) were responsible for doing the entire labeling procedure including segment and acquisition selection. Of the 100 segments designated for the test, 88 met the labeling criteria. Each segment was labeled only once. Included in the second test were 23 segments from the first test which were relabeled by a new analyst team.

Overall labeling accuracies comparing analyst labels to pure small-dot ground-truth labels (ref. 9) for each test are presented in table 6-1. The better accuracies in the second test are attributed to improvements made to the procedure based on results from the first test. Also, the analyst labeled approximately 60 spectrally pure dots as opposed to approximately 140 spectrally mixed or pure dots for which labeling was required in the first test.

Although no significant difference was found, a comparison of the labeling accuracies in table 6-2 shows that the proportion of correct labels at the segment level was generally better in the second test.

During the second test, only acquisitions within a biowindow were used, and two to four acquisitions were acceptable. Preselected acquisitions used in the Multicrop Exploratory Experiment provided less than optimum data for some

TABLE 6-1.- LABELING ACCURACY FOR ANALYST LABELS COMPARED TO PURE SMALL-DOT GROUND-TRUTH LABELS[a]

Test	Accuracy[b]			Commission[b]			Ommission[b]			Confusion[b]	
	P(C/C)	P(Y/Y)	P(0/0)	P(C/0)	P(Y/0)	P(C or Y/0)	P(0/C)	P(0/Y)	P(0/C or Y)	P(C/Y)	P(Y/C)
Multicrop Exploratory Experiment	82.4	74.8	91.2	6.4	6.4	8.8	13.3	13.9	13.7	10.6	4.1
Simulated Aggregation Test	92.5	87.6	95.9	2.9	1.1	4.0	5.8	5.5	5.6	6.7	1.6

[a]Codes are C for Corn, Y for soybeans, and 0 for "other."
[b]For example, subhead P(C/C) refers to the proportion of corn to corn.

TABLE 6-2.- LABELING ACCURACY FOR TWENTY-THREE SEGMENTS PROCESSED IN BOTH TESTS

Segment	Corn				Soybeans				Other			
	Multicrop Exploratory Experiment			Simulated Aggregation Test	Multicrop Exploratory Experiment			Simulated Aggregation Test	Multicrop Exploratory Experiment			Simulated Aggregation Test
	Group I	Group II	Group III		Group I	Group II	Group III		Group I	Group II	Group III	
107	88.9	90.9		88.9	90.9	90.5		100.0	42.9	66.7		25.0
123	65.7	64.7		71.4	92.9	89.5		100.0	96.9	100.0		100.0
127	93.3	90.0	78.8	89.7	77.1	70.6	76.5	73.7	100.0	100.0	92.9	100.0
133		64.5*	25.9*	88.9		46.2*	23.1*	65.7		94.9*	97.7*	90.3
135	67.3		67.3	96.0	76.2		71.4	100.0	93.8		88.2	85.7
141	90.9		71.4	93.8	88.2		93.8	66.7	100.0		92.7	81.2
144		81.0	90.5	91.7		28.6	60.0	62.5		93.5	94.9	95.6
205	72.2	82.4		58.3	62.7	64.2		81.8	100.0	95.5		90.0
800	91.5	93.4		74.4	75.9	77.8		88.9	100.0	100.0		100.0
809	93.8		93.9	100.0	69.2		72.7	100.0	76.5		73.3	100.0
832	85.7		62.5	85.7	71.1		63.4	69.2	87.5		91.3	92.3
837	100.0	95.2	95.2	100.0	63.0	80.4	87.2	78.6	94.4	94.1	100.0	100.0
842	78.0	84.6		82.4	92.1	97.1		75.0	100.0	92.9		100.0
843	88.9	96.8		85.0	82.8	92.9		83.3	94.7	100.0		100.0
852	85.7		82.6	88.9	82.9		64.9	81.8	90.0		93.9	94.4
853	82.4		90.9	83.3	85.7		92.9	86.7	88.5		75.0	100.0
860		87.8	77.8	90.5		54.8	66.7	76.9		96.3	96.0	100.0
864		81.5	80.4	73.1		100.0	88.9	50.0		92.0	70.8	100.0
865	83.8		78.4	100.0	90.0		88.9	57.1	88.2		91.2	90.9
877	86.7	89.1		87.9	61.1	70.6		50.0	84.6	84.6		83.3
880		87.5	80.9	100.0		82.9	82.9	95.2		86.7	69.2	85.7
881	80.4	88.2	63.0	85.7	100.0	77.8	25.0	100.0	100.0	97.8	100.0	100.0
882	98.4	95.4	96.8	96.7	85.3	88.2	90.9	95.5	100.0	100.0	100.0	100.0
$\bar{x}$	85.2	85.8	77.2	87.4	80.3	75.7	71.8	79.9	90.9	93.4	89.1	91.9

*Misregistered data affected labeling accuracy.

segments because they had to be chosen before biowindow definition guidelines had been completed and before retro-ordered acquisitions were available. For example, four acquisitions were required for processing. Therefore, the fourth acquisition usually occurred outside a window, causing confusion because of mixed signatures. In some cases, acquisitions outside a biowindow were used when an equally good or better acquisition was available in the biowindow. This improvement to the test design may explain in part the better accuracies observed in the second test.

Other trends were observed during test evaluation. One observation from the first test was that, from the first to the second time a segment was labeled, accuracies increased 74 percent of the time for corn and 56 percent of the time for soybeans. This indicates that, as the analyst becomes more familiar with procedures, labeling accuracy may improve.

The labeling accuracy of group III for corn was significantly different when compared to the accuracy obtained by other groups (ref. 1). For some segments, group III picked a different separation date or differed the placement of the separation point on the scatter plot. In those cases, the inconsistencies had a definite effect on the correct identification of corn and soybeans. The overall labeling accuracies were affected negatively by this group effect.

Some problems with the procedure were identified in the procedure control reports (refs. 10 and 11) as follows:

- Although biowindow definitions were considered to be straight forward, biowindow ranges determined by two different teams sometimes varied as much as 20 days. The primary reason for the discrepancies was related to the use of the crop calendar shown in figure 6-1. This presentation of crop calendar information, depicting 10-day intervals, was not conducive to defining biowindow ranges consistently. Differences in biowindow length could seriously affect the acquisition selection.

Soybeans: Stages 2, 3, 4, and 5 were interpolated.

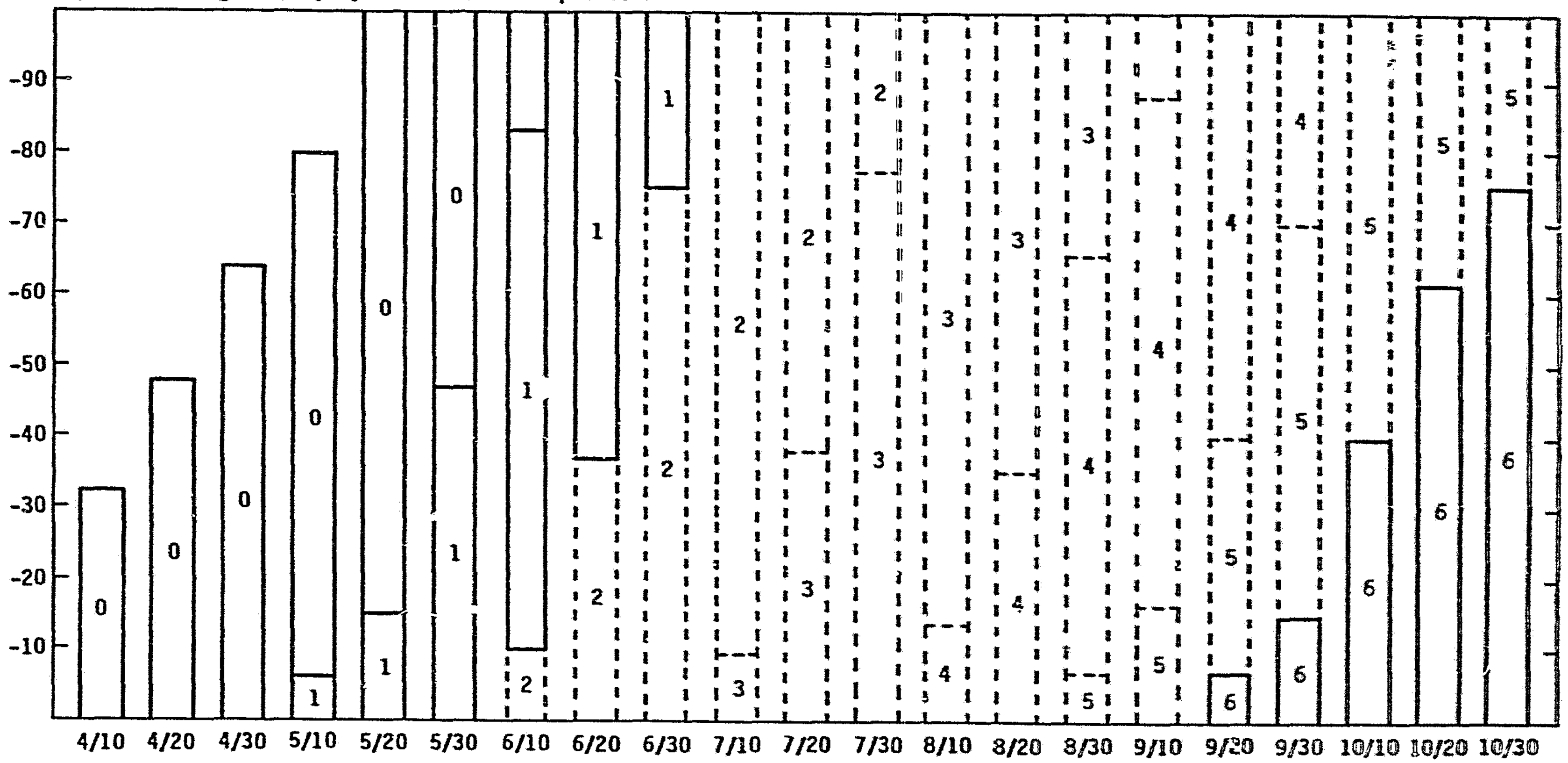


Figure 6-1.- Bar graph crop calendar.

- The spatial and color determinations which were made from the imagery introduce subjective judgments into the procedure. Identification of mixed and misregistered pixels was a difficult task to accomplish. Inconsistency was observed at two different times: by the same individual at different times and between individuals. Color determinations also differed from analyst to analyst.
- Currently, the decision logic only identifies the normal corn/soybean growth cycles. Deviations caused by double cropping, episodal events, and late and early planting were not accounted for in the decision logic.

In summary, the corn/soybean decision logic procedure was easily learned and implemented by both experienced and inexperienced analysts. The amount of time necessary to do the procedure compared favorably with other procedures. Quality assurance (Procedures Control) and error characterization functions were objective because the decision logic was systematic enough that diagnostics could be readily applied to identify the steps where labeling problems occurred. Steps which required changes and/or modifications were recognized readily. In addition, several parts of the decision logic, particularly Steps 2, 3, and 4, could be automated.

7. RECOMMENDATIONS

In order to refine the current decision logic, various actions should be undertaken:

- Normal (historical) crop calendars, which often contained interpolated data and represented only two to five years of information, should be expanded to increase reliability and should have a standard format to allow for consistent definition of biowindow ranges. Current year crop calendar information and adjustable growth models would aid in future development and more accurate biowindow definitions.
- Further study is needed to determine if incorporation of spectral aids into Step 1 and Step 2 could alleviate some of the current inconsistencies in those steps.
- Proceed to automate various parts of the decision logic. Some of the subjective decisions that an analyst is forced to make could be alleviated by using a boundary detection algorithm (i.e., BLOB, ref. 12) and a curve comparison routine (i.e., Badhwar, ref. 13). Both the biowindow definitions and the scatter plot break are conducive to automation. If a color determination scheme (i.e., Cate's color model, ref. 14) were incorporated into the procedure, then Steps 2, 3, and 4 could be completely computerized.

The corn/soybean decision logic has produced encouraging results in the U.S. Corn Belt. Further study should be done to determine if this procedure can be extended to other geographic locations. Also investigations should be done to determine if this method of crop labeling can be expanded to other crops.

8. REFERENCES

1. Carnes, J. and Baird, E.: Evaluation of Results from the U.S. Corn and Soybean Exploratory Experiment. FC-LO-00423, JSC-16339, LEMSCO-14386, July 1980.

2. GSFC: Landsat Data User's Handbook. Document number 76SDS4258 (Greenbelt, Maryland), September 1976.

3. Kauth, R. J.; and Thomas G. S.: The Tasselled Cap - A Graphic Description of the Spectral-Temporal Development of Agricultural Crops As Seen By Landsat. Proceedings of Tenth Annual Symposium on Remote Sensing of Environment (Ann Arbor, Michigan), October 1-2, 1975.

4. Wehmanen, O. A.: The Structure of Landsat Data for Wheat-Growing Region. LEC-13015, January, 1979.

5. Palmer, W.: Multicrop Labeling Decision Logic. Report generated for action document 63-1827-4845-27.

6. Detailed Analysis Procedures for Transition Project (FY 79). LACIE-00724, JSC-13756, May 1979.

7. Dailey, C.; Abotteen, K.: Corn/Soybean Spectral Aid Study. Report generated for action document 63-1287-4845-23.

8. Anderson, J. R.; Hardy, E. E.; Roach, J. T.; and Witmer, R. E.: Land Use and Land Cover Classification System for Use With Remote Sensing Data, Geological Survey Professional Paper 964. Department of the Interior, United States Government Printing Office, Washington, D.C., 1976.

9. White, T.: Dot Method of Digitized Ground Inventory Information. Earth Observations Division, Lyndon B. Johnson Space Center (Houston, Texas), February 1979.

10. Dailey, C. L.; and Abotteen, K. M.: Procedures Control Report - Simulated Aggregation Test. Report issued under action document 63-1287-4845-48.

11. Abotteen, K. M.; and Dailey, C. L.: Procedures Control Report - Multicrop Exploratory Experiment. Report issued under action document 63-1287-4845-23.

12. Kauth, R. J.; Pentland, A. P.; and Thomas, G. S.: BLOB, An Unsupervised Clustering Approach to Spatial Preprocessing of MSS Imagery, Eleventh International Symposium on Remote Sensing of Environment, Environmental Research Institute of Michigan (Ann Arbor, Michigan), vol. 2, 1977, pp. 1309-1317.

13. Badhwar, G.: A Semi-Automatic Technique for Multitemporal Classification of a Given Crop. NASA/JSC Report - (to be published).

14. Cate, R. B.; Phinney, D. E.; Kinsler, M. C.; Sestak, M. L.; Hodges, T.; and Dishler, J. J.; Interpretation of Landsat Digital Data Using a Cubic Color Model Based On Relative Energies. SR-LO-00418, JSC-13776, LEMSCO-13499, February 1980.

APPENDIX A.

OBSERVED CHARACTERISTICS OF CORN AND SOYBEANS

APPENDIX A

OBSERVED CHARACTERISTICS OF CORN AND SOYBEANS

The characteristics of corn and soybeans which were observed on the development segments are presented in tables A-1 through A-4.

For both crops, the growth stages corresponding to each acquisition are presented in terms of historical data and current-year observations. The historical growth stages are taken from CRD normal crop calendars. The observed growth stages are taken from segment crop calendars that were constructed from actual field observations collected for approximately 10 fields per segment at various times throughout the growing season.

In tables A-1 through A-4, image appearance refers to colors observed on the Product 1. The green number and brightness for corn and soybeans are presented in terms of the means and standard deviations of pure pixels.

TABLE A-1a.- OBSERVED CHARACTERISTICS OF CORN AND SOYBEANS AS A FUNCTION OF GROWTH STAGE, APU 14

For Segment 892:	June 16	July 23	Aug. 9	Sept. 4	Sept. 23	Oct. 20
Corn						
Growth stage (historical)	70%>2	60%>3	85%>3	70%>5	75%>5	20%>6
Growth stage (observed)		30%>3	100%>3	50%>5	55%>5	50%>6
Image appearance	Green, lt. red	Red	Dull red, brown	Green, brown, purple	Brown, purple, green, white	Yellow, green, white
Green number	8±3.4	25±3.4	61±2.2	7±1.9	9±2.8	4±1.3
Brightness	78±6.9	57±10.1	61±2.2	43±2.6	41±4.2	40±8.4
Soybeans						
Growth stage (historical)	30%>2	20%>3	65%>3	60%>5	65%>5	60%>6
Growth stage (observed)		100%>2	65%>2	10%>5	10%>5	100%>6
Image appearance	Green	Red	Br. red	Yellow, pink, green	Yellow, pink, green	Green
Green number	6±3.2	31±6.1	43±7	15±9.4	15±8.3	4±2.5
Brightness	83±7.6	67±8.3	83±5.4	56±4.1	54±3.8	44±5

TABLE A-1b.- OBSERVED CHARACTERISTICS OF CORN AND SOYBEANS AS A FUNCTION OF GROWTH STAGE, APU 14

For Segment 886:	June 16	July 23	July 31	Sept. 6	Sept. 15	Sept. 24	Oct. 20	Nov. 7
Corn								
Growth stage (historical)	80%>2	45%>3	75%>31	60%>4 25%>5	80%>4 50%>5	75%>5	20%>6	50%>6
Growth stage (observed)		90%>3	100%>3	100%>4	25%>5	85%>5	30%>6	85%>6
Image appearance	Green	Red	Dull red, brown	Dull red, brown	Green, white, brown	Green, brown, white	Green, white	Green, white
Green number	9±2.1	27±2.1	22±2.3	15±2.3	12±2.7	7±2.3	4±1.2	3±1.9
Brightness	68±5.8	58±3.4	67±4.9	55±2.9	44±2.8	39±2.6	43±7.6	42±8.1
Soybeans								
Growth stage (historical)	90%>1	100%>2	20%>3	75%>4	20%>5	50%>5	50%>6	100%>6
Growth stage (observed)		100%>2	10%>3	100%>4	15%>5	85%>5	100%>6	100%>6
Image appearance	Green	Br. red	Br. red	Br. red	Red, pink	Pink, green	Green, white	Green, white
Green number	6±2.1	32±2.9	35±8.2	33±5.1	25±5.2	12±4.0	3±1.0	2±.6
Brightness	72±4.3	68±4.5	81±8.2	73±3.8	56±6.8	49±2.9	44±7.6	41±2.8

TABLE A-2a.- OBSERVED CHARACTERISTICS OF CORN AND SOYBEANS AS A FUNCTION OF GROWTH STAGE, APU 24

For Segment 804:	June 15	Aug. 17	Sept. 4	Sept. 22	Oct. 1	Oct. 19
Corn						
Growth stage (historical)	65%>2	85%>3	50%>4	65%>5	90%>5	15%>6
Growth stage (observed)		100%>3	100%>4	90%>5	20%>6	86%>6
Image appearance	Green, red	Red, brown	Red, brown	Green, purple		
Green number	8±2.7	12±2.6	7±2.4	3±1.8	3±1.9	3±1.5
Brightness	46±6.7	56±1.9	51±3.5	36±2.9	38±11.9	43±8.1
Soybeans						
Growth stage (historical)	25%>2	60%>3	70%>4	30%>5	85%>5	50%>6
Growth stage (observed)		100%>3	100%>4	80%>5	45%>6	100%>6
Image appearance	Green	Br. red, orange	Br. red, pink	Yellow, pink, green	Yellow, pink, green	Green, brown
Green number	5±2.8	31±5.0	18±5.9	7±4.1	5±2.0	2±1.0
Brightness	43±5.4	77±4.3	69±4.3	43±3.4	46±7.5	41±3.3

TABLE A-2B.- OBSERVED CHARACTERISTICS OF CORN AND SOYBEANS AS A FUNCTION OF GROWTH STAGE, APU 24

For Segment 883:	July 5	July 23	Aug. 8	Aug. 9	Aug. 10	Sept. 24	Oct. 20	Oct. 30
Corn								
Growth stage (historical)	90%>3	100%>3	100%>3	100%>3	15%>4	75%>5	30%>6	60%>6
Growth stage (observed)		90%>3	100%>3	100%>3	100%>3	50%>5	40%>6	90%>6
Image appearance	Red, brown	Red, brown	Red, brown	Red, brown	Red, brown	Green, brown	Green, white	Brown, white, green, yellow
Green number	25±4.8	24±3.5	20±3.7	19±2.8	19±3.0	6±2.5	5±1.8	5±1.7
Brightness	64±3.6	61±3.2	66±3.3	60±3.3	57±3.6	40±4.2	43±8.0	41±8.2
Soybeans								
Growth stgae (historical)	90%>2	35%>3	55%>3	80%>3	85%>3	75%>5	75%>6	95%>6
Growth stage (observed)		100%>2	10%>3	55%>3	60%>3	50%>5	80%>6	95%>6
Image appearance	Red, brown	Br. red	Br. red	Br. red	Br. red	Green, brown	Green, brown	Green, brown
Green number	16±4.0	31±5.5	36±3.4	37±3.9	36±4.5	4±5.0	3±1.3	3±1.0
Brightness	58±3.6	67±4.2	79±4.3	73±3.7	70±6.0	41±2.6	36±8.2	32±6.5

TABLE A-3a.- OBSERVED CHARACTERISTICS OF CORN AND SOYBEANS AS A FUNCTION OF GROWTH STAGE, APU 25

For Segment 209:	June 16	July 4	July 31	Aug. 8	Aug. 9	Sept. 4	Sept. 22	Sept. 23	Oct. 1	Oct. 19	Oct. 20
Corn											
Growth stage (historical)	90%>2	70%>3	25%>4	50%>4	55%>4	50%>5	75%>5	80%>5		55%>6	60%>6
Growth stage (observed)											
Image appearance	Green	Green, white	Red, brown	Red, brown	Red, brown	Dull purple	Dull purple	Dull purple	Brown, green	Green, white	Green white
Green number	3±6.5	12±6.5			28±4.6	18±1.8		11±1.5	12±2.3	3±1.8	
Brightness	66±3.5	80±2.6			62±4.6	46±1.9	36±3.4	44±3.7	40±3.5		
Soybeans											
Growth stage (historical)	60%>2	100%>2	35%>3	55%>3	60%>3	55%>4	40%>5	45%>5	70%>5	50%>6	55%>6
Growth stage (observed)							10%>6	15%>6	20%>6		
Image appearance	Green	Green	Pink, red, green	Pink, red, purple	Pink, red, purple	Br. red	Br. red, pink	Br. red, pink	Pink, green white		
Green number	70±1.3	7±1.5	19±.6	29±2.5	38±2.3	27±1.5		22±1.1	21±4.3	3±.5	3±.5
Brightness	6-15	93±3.5	55±1.7	86±1.2	79±2.0	57±.4		50±1.5	70±3.0	45±1.9	44±2.5

TABLE A-3b.- OBSERVED CHARACTERISTICS OF CORN AND SOYBEANS AS A FUNCTION OF GROWTH STAGE, APU 25

For Segment 211:	June 15	July 3	July 21	Aug. 8	Sept. 4	Sept. 22	Oct. 1	Oct. 19	Oct. 28
Corn									
Growth stage (historical)	90%>2	35%>3	80%>3	45%>4	35%>5	70%>5	35%>6	50%>6	60%>6
Growth stage (observed)		100%>2	45%>3	100%>3	100%>4	25%>5	10>6	35%>6	45%>6
Image appearance	Green	Green, white	Green, red	Red, brown	Purple, brown	Green, brown	Green, brown	Green, brown	Green
Green number	1±3.8	14±6.0	19±3.1		20±2.5	10±1.8	5±2.0	1±1.6	
Brightness	76±6.1	61±5.1	68±3.3		50±2.6	38±2.1	35±4.0	31±3.2	
Soybeans									
Growth stage (historical)	50%>2	100%>2	15%>3	55%>3	60%>4	35%>5	65%>5 20%>6	90%>5 45%>6	60%>6
Growth stage (observed)		100%>2	100%>2	100%>2	90%>3	30%>4	10%>5	25%>5	65%>6
Image appearance	Green	Green, white	Green, red	Br. red, brown	Br. red	Br. red, pink	Pink, green, brown	Green white	Green
Green number	6±1.6	7±2.6	10±2.0	18±2.6	27±2.9	24±2.1	13±2.0	4±.7	1±1.1
Brightness	76±2.2	72±6.7	67±2.1	73±2.4	65±2.8	62±3.4	48±1.7	37±5.5	37±1.8

TABLE A-4a.- OBSERVED CHARACTERISTICS OF CORN AND SOYBEANS AS A FUNCTION OF GROWTH STAGE, APU 28

For Segment 824:	June 12	Aug. 5	Aug. 23	Aug. 31	Sept. 1	Sept. 19	Sept. 28	Nov. 2	Nov. 3
Corn									
Growth stage (historical)	75%>2	70%>3	35%>4	45%>4	50%>4	35%>5	50%>5	60%>6	65%>6
Growth stage (observed)		100%>3	75%>4	100%>4	100%>4	35%>5	20%>6	10%>6	100%>6
Image appearance	Red green	Red, brown, purple	Red, brown, purple	Red, brown, purple	Red, brown, green	Purple, green	Brown, green, white	Brown, green, white	Brown, green, white
Green number	7±3.1	15±3.0	12±3.3	11±4.9	8±2.5	5±2.5	5±1.5	3±.8	3±1.1
Brightness	45±6.0	63±3.2	61±3.2	54±3.3	60±3.1	42±4.4	31±3.8	21±3.8	21±4.6
Soybeans									
Growth stage (historical)	25%>2	65%>3	55%>4	75%>4	80%>4	70%>5	85%>5	100%>6	100%>6
Growth stage (observed)		55%>3	40%>4	75%>4	80%>4	75%>5	70%>6	100%>6	100%>6
Image appearance	Brown, purple	Pink, red, orange	Br. red, orange	Br. red, orange	Br. red, orange	Pink, green, white	Brown, green, white	Brown, grey	Brown, grey
Green number	4±1.6	26±6.9	28±4.3	24±6.1	20±5.6	3±1.9	4±2.0	3±1.1	3±1.1
Brightness	42±9.4	75±5.4	76±4.7	70±3.9	73±3.1	45±3.1	47±11.9	22±5.1	22±5.1

TABLE A-4b.- OBSERVED CHARACTERISTICS OF CORN AND SOYBEANS AS A FUNCTION OF GROWTH STAGE, AFU 28

For Segment 854:	June 10	July 26	Aug. 4	Aug. 21	Aug. 22	Aug. 31	Sept. 8	Sept. 9	Sept. 26	Sept. 27	Nov. 2	Dec. 17
Corn												
Growth stage (historical)	30%>2	50%>3	75%>4	45%>4	45%>4	90%>4 10%>5	15%>5	20%>5	65%>5	70%>5	50%>6	100%>6
Growth stage (observed)		50%>3	100%>3	80%>4	85%>4	100%>4	100%>4	5%>5	30%>6	35%>6	100%>6	100%>6
Image appearance	Green	Red, brown	Red, brown	Red, brown	Red, brown	Red, brown, purple	Red, brown, purple	Red, brown, green	Red, brown, green	Red, brown, green	Gray, green	Gray
Green number	5±2.1	28±3.6	25±2.8	17±2.1	16±2.0	13±3.9	14±1.8	12±1.6	8±2.1	8±1.9	3±1.2	3±.9
Brightness	73±13.0	67±3.7	61±3.3	56±2.4	54±2.2	52±2.6	55±2.3	54±1.8	42±2.0	44±2.0	31±10.8	18±4.5
Soybeans												
Growth stage (historical)	15%>2	[illegible]	100%>3	75%>4	80%>4	15%>5	30%>5	35%>5	20%>6	25%>6	50%>6	100%>6
Growth stage (observed)		100%>2	40%>3	25%>4	30%>4	100%>4	100%>4	100%>4	70%>5	75%>5	100%>6	100%>6
Image appearance	Green	Red, orange	Red, orange	Br. red, orange	Br. red, orange	Red	Br. red	Br. red	Pink, green, brown	Pink, green, brown	Green, brown	Gray
Green number	5±1.7	17±6.7	30±9.8	36±7.3	34±4.5	31±4.9	32±3.5	27±3.9	10±3.9	8±3.2	2±1.0	3±1.1
Brightness	61±12.7	65±6.2	68±8.3	76±6.8	70±4.2	70±5.3	70±3.9	69±3.1	51±3.6	50±2.4	30±6.6	17±2.7

APPENDIX B

DEFINITIONS AND CHARACTERISTICS OF DECISION-TREE CATEGORIES

APPENDIX B

DEFINITIONS AND CHARACTERISTICS OF DECISION-TREE CATEGORIES

B.1 RANGE

Range is uncultivated land that produces forage suitable for livestock grazing. Generally, it is land that is not suited for other types of agriculture, and the natural vegetation consists of predominantly grasslike plants, forbs, or shrubs. Most range in the United States is west of a north-south line that cuts through North and South Dakota, Nebraska, Kansas, Oklahoma, and Texas.

Characteristics:

1. Large and irregular in the Western United States
2. Vegetation indication varied, both within a specific area and between different areas; permanent, with some seasonal change
3. No planting or harvest
4. Coarse texture
5. Red-brown to red in summer and a shade of gray in winter
6. Can occur in conjunction with and adjacent to cropland
7. Best detected in spring

B.2 PASTURE

A pasture is a fenced or unfenced tract of land on which farm animals feed by grazing. Generally, it is a grass area, but it may also have brush and trees. This land category includes land used for feeding at a specific time in rotation with other uses; therefore, land in this situation could be pasture one year and cropland the next. It must be emphasized that the distinction between pasture and range is one of degree and location rather than of actual difference in use. Some definitions of pasture list range as a synonymous term.

<u>Characteristics:</u>

1. Shape varied; geometrical in Eastern and Central United States
2. Size small in Eastern United States, becoming larger westward
3. Easily confused with range
4. Color varied and mixed, ranging from mottled light pink or gray-brown to bright red on highly improved pastures
5. Seasonal changes; no planting or harvest unless new pasture being initiated or old one destroyed
6. Best detected in spring

B.3 ORCHARDS

An area or enclosure devoted to growing fruit, nuts, or certain forest products either as a commercial crop or for reseeding is categorized as an orchard. Isolated small enclosures used for these purposes on small farms would not be recognizable on Landsat imagery.

Characteristics:

1. Varied appearance, depending upon such variables as type of trees, spacing, age, canopy, time of year, and farming practices
2. May closely resemble forest — bright red in late spring and early summer, red-brown at other times
3. Size small in relation to forests
4. Shape and pattern generally regular
5. Area extent usually constant over long time periods

B.4 FOREST

A forest is a plant association predominantly of trees and other woody vegetation that occupies a rather extensive area.

Characteristics:

1. Shape, pattern, and size irregular
2. Generally follows terrain and drainage
3. No planting or harvest as with crops, but annual loss of leafage by certain trees
4. Area extent usually constant over long time periods
5. Bright red in late spring and early summer and reddish brown at other times; variation in intensity and shade

B.5 URBAN

This category is composed of areas that have much of the land covered by structures. It includes villages, towns, cities, strip developments, transportation and industrial areas, shopping centers, parks, cemeteries, golf courses, and sewage plants, as well as institutions that may, in some instances, be isolated from the main urban area. It also includes those areas that strictly are not urban but have been surrounded by urban development.

Characteristics:

1. Irregular in shape and area extent
2. Grid pattern within urban boundaries
3. White to a mixed mottled steel blue; constant through time
4. Texture usually extremely fine
5. Possible occurrence of irregularly shaped areas of light pink to medium red within urban area
6. Close correlation of pattern with urban outline on map
7. Transportation network associated with urban area basically white; can be constant through time

B.6 BARREN LAND

Barren land has a limited ability to support life. Generally, this is an area of thin soil, sand, or rock. Vegetation, if present, is more widely spaced and scrubby than that in the range category. Within this category are dry salt flats, sandy areas other than beaches, exposed rock, and extractive activities (e.g., strip mines, borrow pits, and gravel pits — either active or inactive) having significant surface expression (area).

Characteristics:

1. Bright and constant throughout year
2. Varied dark and light colors and tones
3. Irregular shape
4. Little or no vegetation
5. Size varied, ranging from minute (1 pixel) to extreme (1000 pixels or more)
6. No seasonal change in shape and size

B.7 OTHER AGRICULTURAL LAND

This category is for those items not classified under separate agricultural categories. It includes farmsteads, farm lanes and roads, ditches, horse farms, confined feeding operations such as beef cattle and swine feedlots, dairy operations, and large poultry farms. Generally, these items are small in area, and it is doubtful that items of this nature can be interpreted on Landsat imagery as being other than a farm or farmstead.

Characteristics:

1. Color extremely varied and mixed, white to a dirty or off white for farmsteads and related activities
2. Area extent small
3. No green vegetation
4. No planting or harvest
5. Can occur in conjunction with and adjacent to cropland

B.8 WATER

This category refers to those areas persistently water covered. It includes rivers, streams, canals, lakes (natural and manmade), reservoirs, and bays and estuaries that extend inland.

Characteristics:

1. Irregular in shape except in some cases where manmade
2. May change slightly in shape and size during year
3. Should closely resemble shape and size on map, if mapped
4. Color varied, ranging from a dark blue-black to a bright blue, but usually some shade of blue throughout year
5. Smooth and uniform texture
6. No vegetation

B.9 CROPLAND

Cropland includes all land tilled for crops, as well as cultivated wetlands such as the flooded fields associated with rice production and developed cranberry bogs.

Characteristics:

1. Distinctive geometric field and road pattern in Central and Western United States; irregular and unsystematic in Eastern United States
2. Definite seasonal and intraseasonal changes in color, generally some shade of red or red-brown during growing season
3. Variation in color and intensity with crop type
4. Planting and harvest
5. Vegetation present but not permanent
6. Best detected in summer and early fall

B.10 FALLOW

This is cultivated land that may be kept free of vegetation by such methods as plowing and disking in order to destroy weeds or to conserve a supply of moisture for a succeeding crop.

Characteristics:

1. Shape and pattern similar to areas identified as cropland
2. Planting or harvest
3. Constant blue-green in color, but may vary from dark to light during year

B.11 WETLANDS

Areas where the water table is at, near, or above the land surface for significant part of most years are categorized as wetlands. This category includes marshes, swamps, and tidal flats along the shallow margins of bays, lakes, rivers, and manmade impoundments or reservoirs, bogs, wet meadows, seasonally wet or flooded basins, playas, potholes, and wetland used for wildlife purposes. It does not include wetlands drained for any purpose or wetlands used for rice or similar types of production; these belong to other categories. Wetlands can be either forested or unforested.

Characteristics:

1. Highly varied appearance, both in color and intensity, depending upon such variables as vegetation type, wet or dry season, and winter or summer
2. Irregular in size and shape; not similar to areas identified as cropland
3. Intermittent water possible during year
4. No planting or harvest
5. Seasonally wet

APPENDIX C

DATA SETS USED IN TESTING

APPENDIX C

DATA SETS USED IN TESTING

The following tables contain the segment numbers, the state, and the APU in which the segment is located, the separation acquisition, the acquisitions used for batch processing, the biowindow ranges, the number of available acquisitions in each biowindow and the green number-brightness break in the data on the separation acquisition for all of the segments processed. Table C-1 shows the data set for the Multicrop Exploratory Experiment. Table C-2 shows the data set used in the Simulated Aggregation Test.

TABLE C-1.- DATA SET FOR THE MULTICROP EXPLORATORY EXPERIMENT

Segment number	State/APU	Separation date	Acquisitions processed				Biowindow range (no. of acquisitions)				Scatter plot break
			Base date	2	3	4	A	B	C	Separation	
107A	Ill./25	8235	8235	8208	8262	8307	(3) 144-163	(3) 207-250	(1) 300-337	(3) 221-250	31-69
[a]107B		8235	8235				(0) 145-150	(3) 205-246	(1) 297-335	(3) 220-245	33-69
123A	Ind./28	8233	8233	8161	8197	8305	(2) 146-163	(1) 208-244	(1) 283-337	(1) 223-244	37-70
123B		8233	8233				(2) 144-162	(1) 206-246	(1) 285-336	(1) 223-246	37-68
127A	Ind./28	8243	8243	8161	8269	8306	(2) 150-161	(3) 212-253	(4) 283-365	(2) 222-253	23-60
127B		8243	8243				(2) 147-161	(2) 211-243	(3) 293-334	(1) 222-243	27-65
127C		8216	8243				(2) 145-161	(2) 210-243	(3) 280-335	(1) 222-243	33-73
[b]133A	Ind./28	8233	8233	8152	8269	8314	(1) 145-161	(1) 211-242	(0) 293-345	(1) 227-242	31-67
133B		8233	8233				(1) 150-161	(1) 222-242	(0) 293-334	(1) 232-242	31-66
135A	Iowa/24	8247	8247	8130	8229	8292	(1) 146-165	(2) 210-252	(0) 312-351	(2) 225-252	18-61
135B		8247	8247				(0) 147-150	(2) 210-252	(0) 312-340	(2) 228-254	18-60
141A	Iowa/25	8221	8265	8186	8221	[c]8292	(0) 144-150	(3) 209-258	(1) 306-342	(2) 222-258	23-67

[a]The base date and acquisitions 2, 3, and 4 are the same for each processing.
[b]A misregistered date (8292) caused inaccurate labeling of this segment.
[c]Other acquisitions were available within a biowindow range.

TABLE C-1.- Continued.

Segment number	State/APU	Separation date	Acquisitions processed				Biowindow range (no. of acquisitions)				Scatter plot break
			Base date	2	3	4	A	B	C	Separation	
141B		8221	8265				(1) 144-169	(2) 207-254	(1) 309-346	(1) 222-254	30-75
144A	Iowa/25	8219	8246	8130	8264	8292*	(0) 145-147	(2) 207-240	(1) 309-336	(1) 222-240	36-76
144B		8246	8246				(0) 147-150	(3) 207-246	(2) 300-339	(2) 222-246	20-60
**202A	Mo./14	8221	8167	8221	8266*	8293	(0) 135-150	(2) 206-244	(3) 298-334	(1) 217-244	28-75
202B		8266	8167				(0) 140-150	(2) 211-252	(2) 303-354	(0) 222-252	24-60
202C		8221	8167				(0) 137-148	(2) 203-245	(4) 293-342	(1) 218-245	29-76
205A	Mo./25	8218	8218	8164	8246	8290*	(0) 142-150	(2) 201-243	(2) 295-355	(0) 224-243	36-66
205B		8218	8218				(0) 140-150	(3) 200-250	(2) 293-334	(1) 222-250	36-66
**216A	Mo./25	8220	8220	8130	8247	8292	(0) 140-150	(2) 201-245	(0) 303-354	(2) 209-245	29-68
216B		8238	8220				(0) 140-150	(3) 201-253	(0) 303-354	(2) 232-253	26-65
800A	Iowa/25	8218	8218	8164	8247	8290*	(1) 144-165	(2) 207-247	(1) 309-339	(1) 230-247	29-70
800B		8218	8218				(0) 147-161	(2) 207-251	(1) 309-339	(1) 229-251	29-68

*Other acquisitions were available within a biowindow range.
**These segments were not used in further testing due to lack of a minimum data set.

TABLE C-1.- Continued.

Segment number	State/APU	Separation date	Acquisitions processed				Biowindow range (no. of acquisitions)				Scatter plot break
			Base date	2	3	4	A	B	C	Separation	
**807A	Ill./25	8218	8290*	8164	8218	8272	(0) 140-151	(1) 211-242	(2) 293-334	(0) 222-242	22-59
807B		8218	8290				(0) 142-157	(1) 206-247	(2) 293-334	(0) 224-244	29-68
809A	Ill./25	8218	8218	8164	8244	8290*	(0) 140-161	(2) 211-253	(2) 293-334	(1) 222-253	34-64
809B		8218	8218				(0) 141-160	(2) 206-250	(2) 291-332	(1) 222-250	36-65
832A	Ind./28	8232	8232	8160	8268*	8304	(2) 150-161	(1) 222-242	(2) 293-334	(1) 232-242	32-61
832B		8232	8232				(2) 145-161	(1) 213-244	(2) 285-344	(1) 228-243	34-65
837A	Ind./28	8251	8270*	8180	8198	8234	(0) 150-161	(5) 211-253	(1) 283-334	(4) 222-253	22-62
837B		8234	8270				(0) 145-161	(5) 206-244	(1) 284-334	(3) 223-244	24-67
837C		8234	8270				(0) 145-161	(4) 208-247	(1) 283-334	(3) 223-247	23-66
842A	Ind./28	8232	8268*	8160	8232	8304	(1) 144-162	(1) 208-249	(3) 288-334	(1) 209-249	27-68
842B		8232	8268				(1) 148-164	(1) 209-248	(3) 293-354	(1) 227-248	27-68
843A	Ind./28	8232	8232	8160	8197*	8269*	(2) 143-162	(1) 209-249	(2) 291-342	(1) 227-249	31-68

*Other acquisitions were available within a biowindow range.
**These segments were not used in further testing due to lack of a minimum data set.

TABLE C-1.- Continued.

Segment number	State/APU	Separation date	Acquisitions processed				Biowindow range (no. of acquisitions)				Scatter plot break
			Base date	2	3	4	A	B	C	Separation	
843B		8232	8232				(2) 146-162	(2) 209-251	(2) 290-343	(2) 224-251	32-69
852A	Ind./28	8232	8232	8160	8268*	8304	(2) 145-163	(2) 211-250	(3) 291-334	(2) 227-250	31-74
852B		8232	8232				(2) 150-171	(2) 211-253	(3) 293-334	(2) 232-253	29-71
853A	Ind./28	8232	8232	8160	8268*	8304	(2) 145-161	(2) 208-250	(2) 293-344	(2) 227-250	28-66
853B		8232	8232				(2) 148-171	(2) 211-253	(2) 293-344	(2) 232-253	26-65
860A	Ind./28	8232	8304	8160	8197*	8232	(2) 147-161	(2) 192-243	(1) 283-342	(1) 227-243	32-65
860B		8251	8304				(2) 150-161	(1) 222-243	(1) 293-334	(1) 232-243	26-59
864A	Iowa/14	8231	8267	8150	8186	8231	(1) 144-156	(3) 207-252	(0) 300-336	(1) 224-252	38-64
864B		8231	8267				(2) 132-159	(2) 210-246	(0) 300-340	(1) 225-246	39-61
865A	Iowa/14	8231	8231	8150	8186*	8267	(2) 141-162	(2) 210-252	(0) 300-342	(2) 225-252	36-63
865B		8231	8231				(1) 144-162	(2) 207-258	(0) 301-345	(2) 225-258	38-62
877A	Iowa/14	8222	8222	8141	8186*	8267*	(2) 141-159	(2) 207-252	(0) 300-336	(2) 222-252	26-64

*Other acquisitions were available within a biowindow range.

TABLE C-1.- Continued.

Segment number	State/APU	Separation date	Acquisitions processed				Biowindow range (no. of acquisitions)				Scatter plot break
			Base date	2	3	4	A	B	C	Separation	
877B		8222	8222				(2) 141-156	(2) 206-248	(0) 297-335	(2) 222-248	28-69
**878A	Iowa/24	8186	8186	8266	8293	8311	(0) 141-150	(0) 207-252	(1) 300-342	(0) 225-252	33-58
878B		8186	8186				(0) 143-160	(0) 208-251	(1) 300-334	(0) 225-251	33-59
880A	Iowa/14	8231	8231	8150	8186	8267	(1) 141-162	(2) 207-252	(1) 300-339	(1) 225-252	36-58
880B		8231	8231				(1) 143-162	(2) 208-262	(1) 300-335	(1) 225-262	37-60
881A	Iowa/14	8222	8222	8141	8186*	8267	(1) 141-162	(2) 210-261	(0) 300-337	(1) 225-252	31-66
881B		8222	8222				(1) 141-160	(2) 208-251	(0) 300-336	(2) 222-251	30-63
881C		8222	8222				(0) 143-159	(2) 208-252	(0) 300-339	(1) 225-252	28-61
882A	Iowa/24	8222	8222	8150	8186*	8293*	(1) 142-157	(3) 207-246	(2) 298-333	(2) 216-246	28-64
882B		8222	8222				(1) 144-159	(3) 207-246	(2) 302-333	(2) 217-246	29-64
882C		8222	8222				(1) 144-160	(3) 211-245	(2) 300-334	(2) 219-245	26-61

*Other acquisitions were available within a biowindow range.
**These segments were not used in further testing due to lack of a minimum data set.

TABLE C-1.- Concluded.

Segment number	State/APU	Separation date	Acquisitions processed				Biowindow range (no. of acquisitions)				Scatter plot break
			Base date	2	3	4	A	B	C	Separation	
**891A	Iowa/14	8249	8258	8168	8168	8267	(0) 150-165	(2) 208-261	(0) 300-345	(1) 220-253	21-67
891B		8249	8258				-(0) 144-162	(1) 225-253	(0) 301-344	(1) 225-252	20-65

**These segments were not used in further testing due to lack of a minimum data set.

TABLE C-2.- DATA SET FOR THE SIMULATED AGGREGATION TEST

Segment number	State/APU	Separation date	Acquisitions processed: Base date	2	3	4	Biowindow range (no. of acquisitions): A	B	C	Separation	Scatter plot break
107	Ill./25	8235	8235	8307			(0) 142-161	(5) 204-250	(1) 293-	(4) 217-250	31-68
109	Ill./25	8234	8234	8306			(0) 145-159	(3) 202-234	(1) 288-325	(2) 216-234	26-68
112	Ill./25	8244	3244	8307			(1) 143-159	(4) 206-251	(1) 293-332	(0) 246-251	26-62
113	Ill./25	8235	8235	8307	8163		(2) 140-163	(1) 201-241	(1) 281-323	(1) 211-241	29-73
114	Ill./28	8235	8235	3271	8208		(0) 140-156	(3) 203-239	(1) 286-322	(2) 211-239	26-71
115	Ill./25	8235	8235	8306	8163		(1) 145-159	(4) 203-234	(2) 291-329	(1) 222-234	28-72
120	Ind./28	8233	8233	8306	8161		(1) 145-164	(4) 207-253	(1) 288-344	(2) 225-253	32-69
123	Ind./28	8233	8233	8305	8197	8161	(2) 145-163	(1) 208-245	(1) 283-334	(1) 222-245	28-71
127	Ind./28	8216	8243	8306	8161	8216	(2) 147-161	(2) 210-248	(4) 283-	(1) 225-248	32-72
133	Ind./28	8233	8233	8314	8152		(1) 147-161	(1) 222-242	(0) 293-334	(1) 227-242	35-70
134	Iowa/24	8247	8247	8310	8130		(1) 143-161	(1) 209-247	(1) 300-336	(1) 222-247	23-65
135	Iowa/24	8247	8247	8292	8130		(0) 146-150	(2) 213-261	(0) 310-339	(2) 225-261	17-59
136	Iowa/25	8221	8221	8166	8311		(0) 144-150	(3) 208-255	(1) 306-339	(2) 220-255	34-73
137	Iowa/25	8229	8229	8311	8130		(0) 148-151	(2) 207-252	(3) 303-336	(2) 225-252	26-70

TABLE C-2.- Continued.

Segment number	State/APU	Separation date	Acquisitions processed: Base date	2	3	4	Biowindow range (no. of acquisitions): A	B	C	Separation	Scatter plot break
138	Iowa/25	8246	8246	8300	8156		(1-156) 144-150	(2) 207-240	(0) 303-336	(1) 222-240	23-65
139	Iowa/24	8231	8231	8311	8150		(2) 144-156	(3) 209-249	(2) 306-336	(1) 225-249	27-61
141	Iowa/25	8220	8220	8130	8311		(1-130) 145-150	(2) 206-255	(1) 309-339	(1-220) 222-255	27-70
142	Iowa/24	8231	8231	8303	8141		(1) 138-157	(3) 204-242	(1) 300-332	(2) 216-242	30-59
144	Iowa/25	8246	8246	8300	8130		(0) 142-150	(2) 207-241	(0) 309-336	(1) 223-241	27-69
145	Iowa/25	8220	8220	8311			(0) 144-150	(3) 207-255	(1) 306-339	(3) 222-255	33-69
183	Minn./24	8221	8221	8293			(0) 141-161	(2) 211-259	(3) 283-334	(1) 217-259	32-66
184	Minn./24	8247	8247	8157	8302		(1) 144-158	(4) 205-259	(1) 300-326	(1) 236-259	28-64
201	Mo./25	8220	8220	8292	8138		(0) 140-148 (1-137)	(5) 201-248	(0) 294-363	(3) 221-248 (1-218)	22-65
205	Mo./25	8246	8218	8164	8246	8308	(1-164) 140-150	(3) 199-248	(2) 294-359	(1-246) 223-248	42-75
206	Mo./14	8229	8229	8292			(0) 137-147	(4) 203-245	(1) 293-344	(3) 218-245	28-70
212	Mo./14	8238	8238	8166	8292		(1) 146-178	(5) 217-248	(1) 293-332	(3) 227-248	24-64
217	Mo./14	8247	8247	8130	8302		(1-130) 138-148	(1-247) 203-244	(2) 298-353	(1-247) 218-244	26-67
233	Ohio/63	8248	8248	8302			(0) 145-161	(1) 208-245	(2) 285-334	(1) 222-245	31-66

TABLE C-2.- Continued.

Segment number	State/APU	Separation date	Acquisitions processed				Biowindow range (no. of acquisitions)				Scatter plot break
			Base date	2	3	4	A	B	C	Separation	
238	Ohio/28	8249	8249	8150			(1) 144-159	(2) 206-259	(0) 287-332	(2) 218-259	29-62
800	Iowa/25	8246	8218	8164	8246	8308	(1-164) 147-162	(3) 207-249	(1) 309-360	(1) 225-249	24-64
802	Iowa/25	8219	8219	8156	8318		(1) 147-160	(2) 207-254	(1) 305-340	(1) 228-254	30-73
806	Ill./25	8219	8219	8156	8308		(1) 141-161	(1) 201-252	(2) 290-326	(1) 217-252	29-71
809	Ill./25	8244	8218	8244	8307		(0) 140-157	(2) 206-248	(1) 293-331	(1) 222-248	20-54
820	Ill./25	8235	8235	8306			(1) 141-160	(4) 212-242	(1) 291-322	(3) 227-242	28-72
821	Ill./25	8235	8235	8163	8307		(2) 140-163	(4) 213-243	(1) 283-324	(3) 227-243	21-71
822	Ill./28	8234	8234	8306	8252		(0) 143-160	(3) 213-241	(1) 293-325	(1) 227-241	24-63
825	Ill./28	8234	8234	8153	8297		(1) 142-161	(2) 211-242	(3) 293-324	(2) 232-242	27-64
826	Ill./28	8234	8234	8117	8306		(1-117) 142-161	(3) 212-242	(1) 292-325	(2) 227-242	25-67
827	Ill./28	8243	8243	8306			(0) 142-161	(4) 213-243	(1) 292-324	(2) 227-243	23-63
828	Ill./28	8234	8234	8271	8163		(1) 142-166	(3) 213-236	(0) 293-349	(1) 227-236	29-65
830	Ill./28	8234	8234	8305			(0) 141-160	(3) 212-242	(1) 291-322	(2) 227-242	27-65
832	Ind./28	8232	8232	8313	8151		(2) 145-166	(1) 213-246	(2) 291-349	(1) 230-246	35-69
834	Ind./28	8251	8251	8305	8160		(3) 145-161	(3) 212-253	(2) 283-342	(3) 224-253	23-62

TABLE C-2.- Continued.

Segment number	State/APU	Separation date	Acquisitions processed				Biowindow range (no. of acquisitions)				Scatter plot break
			Base date	2	3	4	A	B	C	Separation	
835	Ind./28	8232	8232	8160	8304		(2) 145-161	(1) 211-243	(2) 283-343	(1) 225-243	30-63
836	Ind./28	8234	8234	8306			(0) 145-161	(4) 206-248	(3) 285-336	(2) 223-248	24-66
837	Ind./28	8234	8234	8306			(0) 145-161	(5) 205-245	(1) 290-334	(3) 224-245	21-66
839	Ind./28	8233	8233	8305	8160		(3) 147-161	(2) 209-248	(1) 291-344	(2) 224-248	36-67
840	Ind./28	8233	8233	8313	8160		(2) 145-162	(3) 210-250	(3) 292-345	(3) 227-250	25-64
842	Ind./28	8232	8232	8313	8160		(1) 145-162	(2) 210-252	(3) 293-	(2) 229-252	26-68
843	Ind./28	8232	8232	8313	8152		(3) 147-161	(2) 209-248	(2) 293-	(2) 232-248	32-69
847	Ind./28	8232	8232	8151	8313		(2) 138-163	(1) 211-249	(3) 293-349	(1) 227-249	35-67
848	Ind./28	8233	8233	8305	8160		(3) 145-161	(2) 207-243	(2) 283-339	(2) 222-243	30-73
849	Ind./28	8233	8233	8305	8151		(4) 146-162	(2) 209-243	(1) 282-334	(2) 220-243	28-66
851	Ind./28	8234	8234	8306	8198	8126	(0) 144-161	(4) 208-249	(1) 283-334	(3) 222-249	28-64
852	Ind./28	8232	8232	8313	8151		(2) 145-161	(2) 210-253	(3) 293-	(2) 225-253	29-74
853	Ind./28	3232	8232	8313	8151		(2) 145-162	(1) 208-245	(2) 293-	(1) 227-245	29-69
855	Ind./28	8233	8233	8305	8161		(2) 146-161	(4) 211-243	(1) 283-336	(3) 224-243	36-69

TABLE C-2.- Continued.

Segment number	State/APU	Separation date	Acquisitions processed				Biowindow range (no. of acquisitions)				Scatter plot break
			Base date	2	3	4	A	B	C	Separation	
856	Ind./28	8234	8234	8306	8161		(2) 148-161	(3) 210-242	(2) 284-334	(1) 223-242	28-65
860	Ind./28	8232	8232	8304	8152		(3) 145-161	(2) 214-247	(1) 288-340	(2) 232-247	36-68
862	Iowa/24	8231	8231	8303	8150		(1) 142-156	(2) 207-249	(1) 300-336	(1) 222-249	34-59
863	Iowa/24	8221	8221	8311	8150		(1) 144-162	(1) 207-252	(2) 300-354	(0) 225-252	32-68
864	Iowa/24	8231	8267	8150	8231		(2) 138-162	(2) 208-252	(1) 300-336	(0) 225-252	33-58
865	Iowa/14	8231	8231	8150	8267		(2) 144-162	(2) 207-252	(0) 300-336	(2) 225-252	29-54
866	Iowa/24	8221	8221	8311	8150		(1) 144-159	(3) 208-246	(2) 301-334	(3) 218-246	31-65
867	Iowa/24	8222	8222	8150	8303		(1) 144-156	(3) 208-244	(2) 303-333	(2) 216-244	26-63
868	Iowa/24	8221	8221	8166	8302		(0) 144-147	(2) 206-252	(2) 299-336	(2) 222-252	27-63
869	Iowa/24	8221	8221	8311			(0) 144-147	(1) 204-248	(1) 298-334	(1) 221-248	26-62
870	Iowa/24	8221	8221	8311			(0) 147-153	(1) 207-252	(1) 300-348	(1) 222-252	33-69
871	Iowa/24	8221	8221	8311			(0) 144-150	(1) 207-251	(1) 298-336	(0) 225-251	32-65
874	Iowa/24	8221	8221	8311			(0) 144-162	(1) 207-252	(1) 300-338	(1) 214-252	27-64

TABLE C-2.- Continued.

Segment number	State/APU	Separation date	Acquisitions processed: Base date	2	3	4	Biowindow range (no. of acquisitions): A	B	C	Separation	Scatter plot break
875	Iowa/24	8221	8221	8293			(0) 142-160	(1) 207-251	(1-293) 300-337	(1-221) 225-251	30-66
876	Iowa/14	8231	8231	8141			(1) 141-159	(1) 207-252	(0) 300-336	(1) 233-252	32-63
877	Iowa/14	8231	8231	8141	8267		(1) 144-161	(2) 208-251	(0) 300-336	(0) 244-251	28-53
880	Iowa/14	8231	8231	8150	8267		(1) 144-162	(2) 207-250	(0) 300-339	(1) 225-250	30-56
881	Iowa/14	8231	8231	8267	8141		(2) 144-162	(3) 209-252	(0) 300-336	(1) 225-336	42-69
882	Iowa/24	8231	8231	8150	8303		(1) 144-159	(3) 213-248	(2) 300-342	(2) 218-248	35-63
884	Iowa/24	8231	8231	8311	8150		(1) 144-157	(3) 208-244	(3) 302-332	(2) 218-244	31-53
885	Iowa/24	8222	8222	8153	8303		(1) 144-162	(3) 206-245	(2) 300-333	(2) 219-245	29-66
887	Iowa/14	8222	8222	8159			(1) 142-146	(3) 203-242	(2) 301-335	(2) 212-242	31-69
888	Iowa/14	8231	8231	8294	8150		(2) 141-156	(2) 206-246	(1) 300-335	(1) 214-246	31-67
890	Iowa/24	8231	8231	8141	8303		(1) 141-159	(2) 207-252	(1) 300-336	(2) 221-252	31-61
894	Iowa/24	8221	8221	8311			(0) 143-148	(1) 205-249	(1) 300-335	(0) 223-249	32-65
895	Iowa/14	8231	8231	8159			(3) 141-159	(3) 207-250	(1) 300-335	(2) 222-250	32-64
896	Iowa/[illegible]	8231	8231	8150			(1) 143-159	(2) 208-250	(0) 303-335	(1) 224-250	33-67

TABLE C-2.- Concluded.

Segment number	State/APU	Separation date	Acquisitions processed				Biowindow range (no. of acquisitions)				Scatter plot break
			Base date	2	3	4	A	B	C	Separation	
897	Iowa/14	8222	8222	8141			(3) 141-161	(2) 207-250	(0) 300-336	(2) 222-250	28-68
898	Iowa/24	8221	8221	8311			(0) 142-147	(1) 208-248	(1) 300-335	(1) 223-248	26-66
899	Iowa/24	8221	8221	8311			(0) 141-156	(1) 207-249	(1) 299-336	(1) 221-249	27-63
1872	Minn./24	8222	8222	8150			(1) 141-150	(4) 204-250	(0) 294-321	(3) 213-250	32-65

NASA-JSC

APPENDIX C

SEGMENT-LEVEL EVALUATION OF THE SIMULATED AGGREGATION TEST: U.S. CORN AND SOYBEAN EXPLORATORY EXPERIMENT

AgRISTARS

FC-L0-00493
JSC-16820

A Joint Program for Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing

Foreign Commodity Production Forecasting

October 1980

SEGMENT-LEVEL EVALUATION OF THE SIMULATED AGGREGATION TEST:

U.S. CORN AND SOYBEAN EXPLORATORY EXPERIMENT

S. A. Davidson

Lockheed Engineering and Management Services Company, Inc.
1830 NASA Road 1, Houston, Texas 77058

Lyndon B. Johnson Space Center
Houston, Texas 77058

PRECEDING PAGE BLANK NOT FILMED

1. Report No. FC-LO-00493; JSC-16820	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Segment-Level Evaluation of the Simulated Aggregation Test		5. Report Date October 1980
		6. Performing Organization Code
7. Author(s) S. A. Davidson, Lockheed Engineering and Management Services Company, Inc.		8. Performing Organization Report No. LEMSCO-15116
9. Performing Organization Name and Address Lockheed Engineering and Management Services Company, Inc. 1830 NASA Road 1 Houston, Texas 77058		10. Work Unit No.
		11. Contract or Grant No. NAS 9-15800
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Lyndon B. Johnson Space Center Houston, Texas 77058 Tech. Monitor: R. O. Hill		13. Type of Report and Period Covered Technical Report
		14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

An evaluation of the corn and soybean proportion-estimation accuracy and dot labeling accuracy of the Simulated Aggregation Test, U.S. Corn and Soybean Exploratory Experiment, is presented. These results are in turn compared with the corn and soybean proportion-estimation accuracy and dot labeling accuracy of the Classification Procedures Verification Test.

17. Key Words (Suggested by Author(s))	18. Distribution Statement		
U.S. Corn and Soybean Exploratory Experiment Classification Procedures Verification Test, proportion-estimation accuracy, dot labeling accuracy, corn and soybean labeling procedure			
19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price*
Unclassified	Unclassified	25	

PRECEDING PAGE BLANK NOT FILMED

FC-LO-00493
JSC-16820

SEGMENT-LEVEL EVALUATION OF THE SIMULATED AGGREGATION TEST: U.S. CORN AND SOYBEAN EXPLORATORY EXPERIMENT

Job Order 74-402

This report describes Accuracy Assessment Activities of the Foreign Commodity Production Forecasting project of the AgRISTARS program.

PREPARED BY

S. A. Davidson

APPROVED BY

M. D. Pore, Supervisor
Accuracy Assessment Section

B. L. Carroll, Manager
Commodity Forecasting Department

LOCKHEED ENGINEERING AND MANAGEMENT SERVICES COMPANY, INC.

Under Contract NAS 9-15800

For

Earth Observations Division

Space and Life Sciences Directorate

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

October 1980 LEMSCO-15116

CONTENTS

TABLES

FIGURES

1. INTRODUCTION

The Simulated Aggregation Test (SAT): U.S. Corn and Soybean Exploratory Experiment was executed (1) to determine the labeling accuracy obtainable with the current corn and soybean labeling procedure and to determine the crop proportion-estimation errors of the resulting proportion estimates; (2) to compare the corn and soybean labeling procedure utilized in the SAT with that utilized in the Classification Procedures Verification Test (PVT) via a comparison of the labeling accuracy and the proportion-estimation errors of the two procedures; and (3) to test the aggregation logic for obtaining crop area and production estimates at state and regional levels. This report presents the results of (1) and (2).

The design of the SAT called for three analyst-interpreter (AI) groups (two from NASA and one from Lockheed) to label 50 to 70 Type I dots on each of 88 segments located in 5 agro-physical units (APU's) in 6 states of the U.S. Corn Belt. Each segment was to be labeled once only using a modified version of the corn and soybean labeling procedure utilized in the PVT (refs. 1 and 2).

Of the 88 segments labeled, 23 were a subset of the 29 blind sites processed in the PVT; 35 were additional blind sites; and the remaining 30 were nonblind sites. All the 23 segments in the SAT that were also processed in the PVT (hereafter referred to as Group 1 segments) had digitized ground truth available. Of the additional 35 blind sites (hereafter referred to as Group 2 segments), 18 had digitized ground truth available, and the remaining 17 had 400-dot ground truth available.

Since the NASA groups had already seen the ground truth for the Group 1 segments, it was stipulated that these 23 segments would be processed by the Lockheed group. Otherwise, there were no constraints on the assignment of segments to the AI groups. Table 1-1 shows the assignment of the blind sites to the APU's and AI groups.

TABLE 1-1.- ALLOCATION OF BLIND SITES TO GROUP AND APU

Group	APU						
	14	24		25		28	
A	888	142	890	137		120	855
	895	866	1872	145		825	
	897	871				827	
		875				840	
						848	
						851	
B	887	134	894	138		826	
	896	183				836	
		862				847	
		867				849	
		874				856	
C	864	135		107	809	123	842
	865	184		141		127	843
	877	870		144		133	852
	880	882		205		828	853
	881			800		832	860
						837	

2. ANALYSIS OF THE SIMULATED AGGREGATION TEST

Analyses were made to investigate the crop proportion-estimation accuracy and dot-labeling accuracy in the SAT as well as to compare the crop proportion-estimation accuracy and dot-labeling accuracy of the SAT with that of the PVT.

2.1 CROP PROPORTION-ESTIMATION ACCURACY IN THE SIMULATED AGGREGATION TEST

Initially, a linear model of the form

$$\hat{P}_{ijk} - P_{ijk} = u + A_i + G_j + (AG)_{ij} + \varepsilon_{(ij)k}$$

was assumed where

$\hat{P}_{ijk}$ = the proportion estimate of the crop of interest for the k^{th} segment of the i^{th} APU, labeled by the j^{th} group

P_{ijk} = the corresponding ground truth proportion

u = the overall mean difference

A_i = the effect of the i^{th} APU (fixed)

G_j = the effect of the j^{th} group (random)

$(AG)_{ij}$ = the interaction of the i^{th} APU and the j^{th} group (mixed)

$\varepsilon_{(ij)k}$ = the random error resulting from the k^{th} segment of the i^{th} APU, labeled by the j^{th} group, assumed $NID(0,\sigma^2)$.

However, for the crops of interest (corn and soybeans), the model accounted for less than 29 percent of the observed variation. (Table 2-1 gives the coefficient of determination, R^2, for each crop.) Hence, the analyses were performed without regard to APU or group effects.

Plots of ground truth proportions (abscissa) versus crop proportion-estimation error (ordinate) are displayed in figures 2-1(a) for corn and 2-1(b) for soybeans. Overestimation of corn and underestimation of soybeans are clearly evident, a pattern that also emerged in the PVT (ref. 3).

TABLE 2-1.- COEFFICIENT OF DETERMINATION FOR EACH CROP OF INTEREST

Crop	Coefficient of determination, percent
Corn	28.4
Soybeans	25.4

Table 2-2 presents the mean error, the standard deviation of the error, the mean square error, and the 95 percent confidence intervals of the mean error for the corn and soybean proportion estimates. Since neither confidence interval contains zero, the mean proportion-estimation error for both corn and soybeans is significantly different from zero ($\alpha = 0.05$), with corn overestimated an average of 4.58 percent per segment and soybeans underestimated an average of 7.81 percent per segment.

Table 2-3 indicates that the overestimation of corn is due largely to an overestimation in the Group 2 segments, whereas for soybeans, the mean errors for the Group 1 and Group 2 segments are essentially equal.

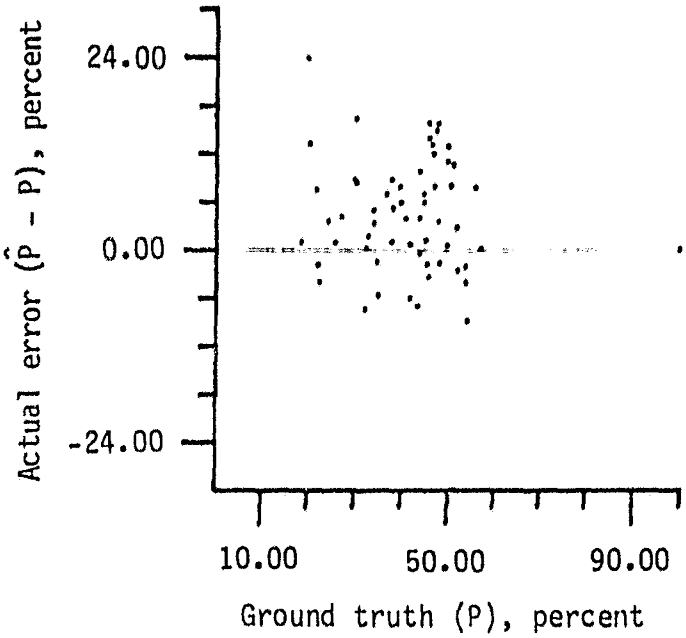
2.2 COMPARISON OF THE CROP PROPORTION-ESTIMATION ACCURACY OF THE SIMULATED AGGREGATION TEST WITH THE CLASSIFICATION PROCEDURES VERIFICATION TEST

The comparison of the SAT with the PVT was made in two parts:

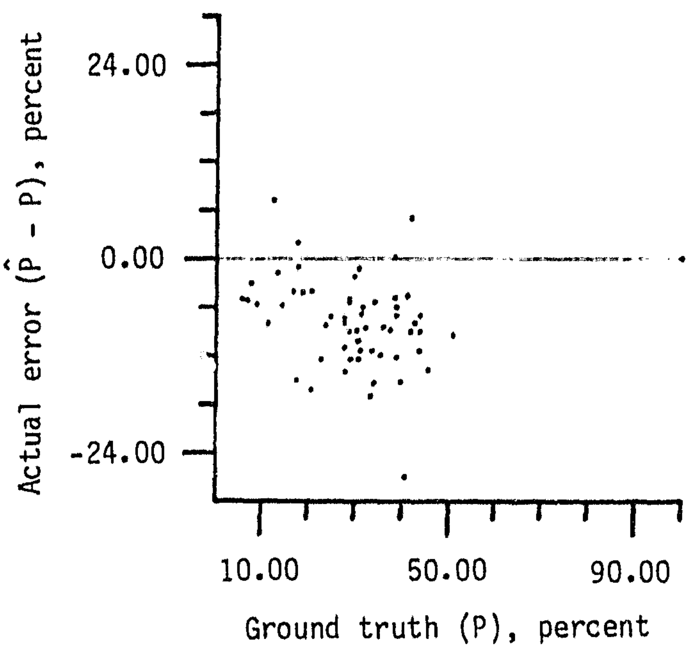
1. A paired comparison of the Group 1 segment proportion-estimation accuracy with the PVT proportion-estimation accuracy.
2. A comparison of the Group 2 segment proportion-estimation accuracy with the PVT proportion-estimation accuracy.

2.2.1 PAIRED COMPARISON OF THE GROUP 1 SEGMENTS WITH THE CLASSIFICATION PROCEDURES VERIFICATION TEST

Since the segments of the PVT were labeled by at least two AI groups whereas the Group 1 segments were labeled only once, it was necessary to compare the



(a) Corn.



(b) Soybeans.

Figure 2-1.- Crop proportion-estimation accuracy for the SAT.

TABLE 2-2.- CROP PROPORTION-ESTIMATION ACCURACY FOR THE SAT

Crop	Mean ground truth proportion, percent	Mean error, percent	Standard deviation of mean error, percent	Mean square error	95 percent confidence intervals of mean error
Corn	40.58	4.58	6.95	68.38	[2.80, 6.36]
Soybeans	29.67	-7.81	5.57	91.54	[-9.24, -6.38]

TABLE 2-3.- CROP PROPORTION-ESTIMATION ACCURACY OF THE PVT AND THE SAT

Test	Corn			Soybeans		
	Mean error, percent	Standard deviation, percent	Mean square error	Mean error, percent	Standard deviation, percent	Mean square error
PVT	2.43	10.00	103.8	-4.67	6.33	61.0
SAT	4.58	6.95	68.4	-7.81	5.57	91.5
SAT Group 1, [a]23	1.88	6.52	44.1	-8.10	4.71	86.8
SAT Group 2, [b]35	6.35	6.73	84.3	-7.62	6.13	94.7

[a]Number of blind site segments in the SAT that were also processed in the PVT; referred to in text as Group 1 segments.

[b]Number of additional blind sites in SAT; referred to in text as Group 2 segments.

absolute value of the proportion-estimation error (absolute error) of each Group 1 segment with the mean absolute error of the corresponding PVT segment by means of the difference: mean absolute error minus absolute error.

The hypothesis of a mean difference of zero versus all alternatives was then tested ($\alpha = 0.05$). The results, displayed in table 2-4, show no significant difference in the proportion-estimation accuracy of corn; however, soybeans were underestimated to a significantly greater degree in the Group 1 segments (a mean difference of -2.60 percent).

2.2.2 COMPARISON OF THE GROUP 2 SEGMENTS WITH THE CLASSIFICATION PROCEDURES VERIFICATION TEST

The analysis for the comparison of the Group 2 proportion-estimation accuracy with the PVT proportion-estimation accuracy consisted of testing the hypothesis that the mean error of the PVT segments minus the mean error of the Group 2 segments was significantly different from zero ($\alpha = 0.05$) versus all alternatives. Table 2-5 displays the results of this test. Corn was overestimated to a significantly greater degree and soybeans underestimated to a significantly greater degree in the Group 2 segments.

2.3 LABELING ACCURACY OF THE SIMULATED AGGREGATION TEST

Tables 2-6(a) through 2-6(c) display, for all blind sites for the Group 1 segments and all blind sites for the Group 2 segments, the percentage of a given crop category labeled "corn," "soybeans," and "other" (neither corn nor soybeans). With errors of omission being essentially equal for corn and soybeans, the confusion errors for Group 1 and Group 2 together [table 2-6(a)] indicate that the AI groups could recognize corn signatures more readily than soybean signatures. This failure to discriminate soybeans from corn is due to late planting of soybeans, making the signatures of these late planted soybeans spectrally inseparable from corn. As a result, corn is overestimated and soybeans underestimated.

TABLE 2-4.- PAIRED COMPARISON OF THE CROP PROPORTION-ESTIMATION ACCURACY OF THE GROUP 1 SAT SEGMENTS WITH THE PVT SEGMENTS

Crop	Mean difference (PVT and Group 1 SAT), percent	Standard deviation, percent	Standard error of the mean, percent	95 percent confidence intervals
Corn	2.01	5.69	1.19	[-0.32, 4.34]
Soybeans	-2.60	4.53	0.94	[-4.44, -0.76]

TABLE 2-5.- COMPARISON OF THE PROPORTION-ESTIMATION ACCURACY OF THE PVT SEGMENTS WITH THE GROUP 2 SAT SEGMENTS

Crop	PVT mean error (standard deviation), percent	Group 2 SAT mean error (standard deviation), percent	Difference of mean errors, percent	Standard error of difference, percent	95 percent confidence intervals
Corn	2.43 (10.00)	6.35 (6.73)	-3.92	1.94	[-7.72, -0.12]
Soybeans	-4.67 (6.33)	-7.62 (6.13)	2.95	1.38	[0.25, 5.65]

TABLE 2-6.- DISTRIBUTION OF LABELS WITHIN EACH GROUND TRUTH CATEGORY

(a) All SAT blind sites

Ground truth	Label			Ground truth proportion, percent
	Corn, percent	Soybeans, percent	Other, percent	
Corn	92.58	1.62	5.80	43.36
Soybeans	6.87	87.58	5.54	30.25
Other	2.92	1.14	95.93	26.39

(b) Group 1 blind sites

Ground truth	Label			Ground truth proportion, percent
	Corn, percent	Soybeans, percent	Other, percent	
Corn	88.25	1.77	9.98	44.00
Soybeans	7.97	83.33	8.70	26.93
Other	3.69	2.35	93.96	29.07

(c) Group 2 blind sites

Ground truth	Label			Ground truth proportion, percent
	Corn, percent	Soybeans, percent	Other, percent	
Corn	94.89	1.54	3.56	43.03
Soybeans	6.39	89.46	4.15	31.99
Other	2.45	0.41	97.14	24.99

The drop in labeling accuracy from the Group 2 segments to the Group 1 segments [tables 2-6(b) and 2-6(c)] is accompanied by a small increase in confusion errors (6.39 to 7.97 percent for soybeans and 1.54 to 1.77 percent for corn), and a rather large increase in errors of omission (4.15 to 8.70 percent for soybeans and 3.56 to 9.98 percent for corn). In other words, the discrimination between corn and soybeans of the Group 1 segments was at approximately the same level as that of the Group 2 segments. However, the separation of corn and soybeans from "other" was not done as well on the Group 1 segments as on the Group 2 segments.

The discrepancy in labeling accuracy between Group 1 and Group 2 segments is difficult to explain. Those AI groups labeling the Group 2 segments had previously used, in the PVT, a corn and soybean labeling procedure similar to the one used for the SAT. On the other hand, the AI group labeling the Group 1 segments had never used a corn and soybean labeling procedure. This observation seems to indicate that labeling accuracy is a function of familiarity with the labeling procedure. However, any effect induced by familiarity with the labeling procedure would be totally confounded with any effect induced by the segments.

Relating the labeling accuracy of the Group 1 and the Group 2 segments to their respective proportion-estimation accuracies (table 2-3) shows that even though the labeling accuracy of corn and soybeans is higher for the Group 2 segments, the proportion-estimation accuracy of corn in the Group 2 segments is much worse than that of the Group 1 segments. Also, the proportion-estimation accuracy of soybeans is only slightly better.

This discrepancy in labeling is a result of the reduction in omission errors for the Group 2 segments and the spectral inseparability of some soybeans from corn due to late planting of soybeans. This inseparability of soybeans from corn results in an underestimation of soybeans and an overestimation of corn for both groups of segments. The decrease in omission errors for corn in the Group 2 segments, however, further inflates the estimate of corn. The decrease in omission errors for soybeans appears to have little influence on reducing the underestimation of soybeans, indicating that committing soybeans with corn has a greater impact on soybean proportion-estimation accuracy than the mislabeling of soybeans as "other."

2.4 COMPARISON OF THE DOT-LABELING ACCURACY OF THE SIMULATED AGGREGATION TEST AND THE CLASSIFICATION PROCEDURES VERIFICATION TEST

Dot-labeling accuracy for the PVT, the Group 1 segments, the Group 2 segments, and the Group 1 and Group 2 segments combined is displayed in table 2-7. Overall, the labeling accuracy of the SAT improved over that of the PVT, with the labeling accuracy of the Group 2 segments contributing the most to this improvement. However, since dot-labeling accuracy data at the segment level was available only for the Group 1 segments, it was not possible to determine if the improvement in labeling accuracy for the Group 2 segments was significant.

The labeling accuracy of each Group 1 segment was compared with the mean labeling accuracy of the corresponding PVT segment by subtracting the Group 1 figures from the corresponding PVT figures. The null hypothesis of a mean difference of zero was tested against all alternatives ($\alpha = 0.05$). The results are given in table 2-8.

Since each of the 95 percent confidence intervals contains zero, the null hypothesis that the mean difference in labeling accuracy between the PVT segments and the SAT Group 1 segments is zero could not be rejected.

2.5 ANALYST-INTERPRETER LABELED, TYPE I DOT PROPORTION ESTIMATES

Crop proportion estimates of corn and soybeans were made for each blind site by using the proportion of dots labeled corn and the proportion of dots labeled soybeans. Figures 2-2(a) for corn and 2-2(b) for soybeans display plots of ground truth proportions versus the dot proportion-estimation error.

In table 2-9, the mean errors of the machine-classified estimates and the dot estimates are displayed. For both corn and soybeans, the Type 1 dots, as a random sample, produced smaller estimation errors, with the dot-estimation

TABLE 2-7.- DOT-LABELING ACCURACY FOR THE PVT AND THE SAT

Test	Crop		
	Corn, percent	Soybeans, percent	Other, percent
PVT	86	79	93
SAT Group 1	88	83	94
SAT Group 2	95	89	97
SAT	93	88	96

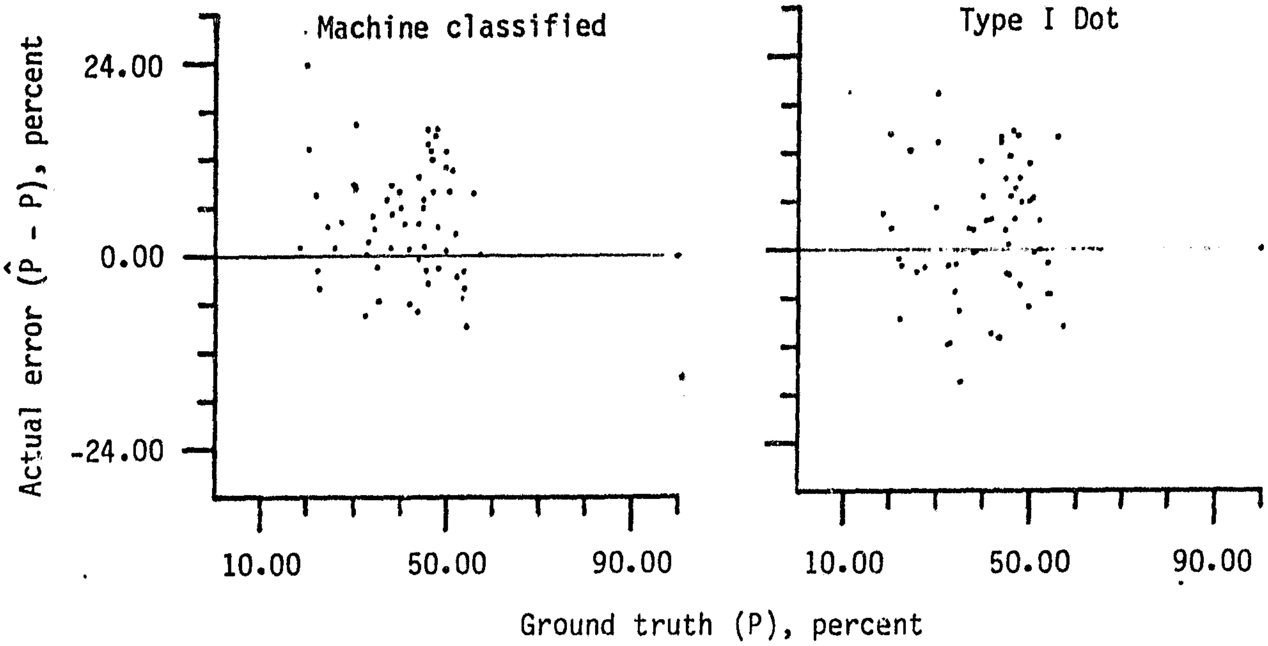
TABLE 2-8.- COMPARISON OF THE PVT AND THE SAT GROUP 1 LABELING ACCURACY

Crop	Mean difference (PVT and Group 1 SAT), percent	Standard deviation, percent	Standard error of the mean, percent	95 percent confidence intervals
Corn	-3.47	11.05	2.36	[-8.10, 1.16]
Soybeans	-2.95	20.14	4.29	[-11.36, 5.46]
Other	-1.73	11.11	2.37	[-6.38, 2.92]

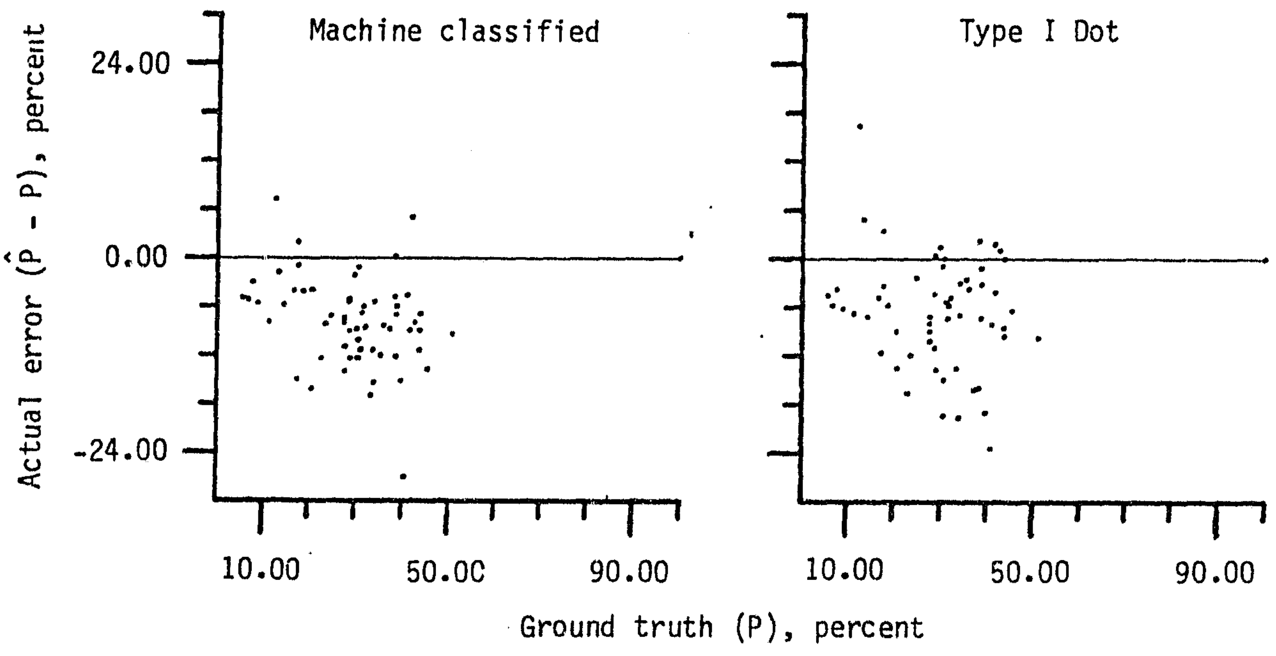
TABLE 2-9.- CLASSIFICATION ERRORS OF THE SAT

Source of classification	Corn			Soybeans		
	Mean error, percent	Standard deviation, percent	Mean square error	Mean error, percent	Standard deviation percent	Mean square error
Machine classification	[a]4.58	6.95	68.38	[a]-7.81	5.57	91.54
Type 1 dots as random sample	1.91	8.32	71.72	[a]-6.62	6.91	90.86

[a]Significantly different from zero ($\alpha = 0.05$).



(a) Corn.



(b) Soybeans.

Figure 2-2.- Comparison of machine-classified estimates with AI-labeled, Type 1 dot proportion estimates.

error for corn not significantly different from zero, although the estimate of soybeans is biased. However, the mean square errors for the two types of classification are not appreciably different, indicating that if the dot estimates are not better than the machine-classified estimates, then certainly they are no worse.

To compare the types of classification, two procedures were used. The first procedure, utilizing the binomial test, was to investigate whether or not one type of classification tended to yield superior estimation accuracy over the other. The first step in this procedure was determining the proportion of segments for which the dot estimates produced smaller, absolute deviations from ground truth. (See "Improved," table 2-10.) Then the null hypothesis that this proportion was not significantly different from 50 percent ($\alpha = 0.05$) was tested. For both corn and soybeans, the null hypothesis was not rejected. In other words, machine classification is no more likely to yield accurate estimates than a random sample of Type 1 dots.

To further qualify the comparison, the mean improvement of machine-classified estimates over dot estimates (see table 2-10) was obtained by finding the mean, on a segment-by-segment basis, of the absolute deviation from ground truth of the machine-classified estimate minus the absolute deviation from ground truth of the dot estimate. The null hypothesis of no significant improvement ($\alpha = 0.05$) was tested. The null hypothesis could not be rejected.

Thus, machine classification does not improve upon a random sample of Type 1, analyst-labeled dots whether measured as a reduction of mean square error, a likelihood of yielding more accurate estimates, or a mean difference in estimation accuracy.

TABLE 2-10.- PROPORTION-ESTIMATION ACCURACY IMPROVEMENT USING ANALYST-LABELED, TYPE 1 DOTS AS A RANDOM SAMPLE

Corn		Soybeans	
Improved, percent	Mean improvement, percent	Improved, percent	Mean improvement, percent
45	-1.20 [a][-3.00, 0.6]	52	0.59 [a][-0.57, 1.75]

[a]Ninety-five percent confidence interval for the mean improvement.

3. SUMMARY OF RESULTS

The following results emerged from the evaluation of the SAT:

1. Corn was significantly overestimated on an average of 4.58 percent per segment (standard deviation, 6.95 percent), and soybeans were significantly underestimated on an average of 7.81 percent per segment [standard deviation, 5.57 percent (table 2-2)].

2. When comparing the proportion-estimation accuracy of the Group 1 SAT segments with the PVT segments, no significant difference emerged for corn; however, soybeans were underestimated to a significantly greater degree in the SAT segments (table 2-4).

3. When comparing the proportion-estimation accuracy of the Group 2 SAT segments with the PVT segments, corn was overestimated to a significantly greater degree and soybeans underestimated to a significantly greater degree in the SAT segments (table 2-5).

4. The labeling accuracy of the Group 2 segments was higher than that of the Group 1 segments as a result of fewer corn and soybean dots being mislabeled as "other" in the Group 2 segments [tables 2-6(b) and 2-6(c)].

5. In the SAT, more soybeans were labeled corn than corn, soybeans. This was caused by the spectral inseparability of late planted soybeans from corn [tables 2-6(a) through 2-6(c)].

6. The spectral inseparability of late planted soybeans from corn resulted in the overestimation of corn and underestimation of soybeans.

7. Since fewer corn and soybean dots were mislabeled "other" in the Group 2 segments (as compared with the Group 1 segments), the estimation of corn was further inflated, although the reduction in mislabeling had little effect on the soybean proportion estimates [tables 2-6(b) and 2-6(c)].

8. Overall, labeling accuracy in the SAT improved over that in the PVT. However, there was no significant difference in labeling accuracy between the PVT and Group 1 segments (tables ^ 7 and 2-8).

9. When comparing machine-classified estimates with estimates based upon a random sample of Type 1 dots, machine-classified estimates did not improve upon the Type 1 dot, random sample estimates (tables 2-9 and 2-10).

4. RECOMMENDATIONS

An alternate machine classification technique should be developed since the procedure used in this experiment did not improve upon a random sample of analyst-labeled, Type 1 dots. Methods should also be developed to compensate for the adverse effect that late planted soybeans have upon corn and soybean proportion-estimation accuracy.

5. REFERENCES

1. Abotteen, K. M.; and Dailey, C. L.: Procedures Control Report - Multicrop Exploratory Experiment Test. Report issued under Action Document 63-1827-4845-38, NASA/JSC (Houston), May 22, 1979. (Unpublished.)

2. Daily, C. L.; and Abotteen, K. M.: Procedures Control Report - Simulated Aggregation Test. Report Issued Under Action Document 63-1827-4845-48. NASA/JSC (Houston), July 13, 1979. (Unpublished.)

3. Carnes, J. G.; and Baird, J. E.: Evaluation of Results of U.S. Corn and Soybeans Exploratory Experiment — Classification Procedures Verification Test. NASA JSC-16339, LEMSCO-14386, April 1980.

APPENDIX D

TEST OF GROUPED OPTIMAL AGGREGATION TECHNIQUE

APPENDIX D

TEST OF GROUPED OPTIMAL AGGREGATION TECHNIQUE

The objective of this simulation study was to conduct a simulated test with two sub-objectives: first, to evaluate the Multicrop Allocation Procedure (MAP) of H. O. Hartley et al. (ref. 1), and second, to evaluate the Grouped Optimal Aggregation Technique (ref. 2). Since one of the major goals in the AgRISTARS program is to extend the technology developed during the Large Area Crop Inventory Experiment (LACIE) for wheat production to the estimation of production of several crops, the need for a MAP is apparent.

In the MAP, the allocation problem is formatted in terms of nonlinear programming. The actual process used was minimization of the total sample size using a Lagrange Multiplier technique, subject to the constraints that the sample C.V.'s for each crop not exceed a given value (in this case 5 percent).

The Grouped Optimal Aggregation Technique is designed to improve upon the aggregation scheme used in LACIE by using a weighting scheme which combines contextual information (neighboring strata) with the target strata information by giving more weight to the proportion estimates of strata with plentiful data and less weight to the estimates of strata with little data.

The simulation was performed in August and September of 1980 by A. H. Feiveson at the Johnson Space Center, Houston, Texas, and the methods used and results obtained are described in this appendix.

D.1 BACKGROUND

The study was based on corn and soybeans acreage and production statistics for 1978 in Iowa, Illinois, and Indiana. These three states were stratified into a total of 12 acreage strata, each representing the intersection of an APU (agro-physical unit) with a state. A total of 204 segments were then allocated to 12 strata using the MAP, with the goal of achieving a 5-percent C.V. for both corn and soybeans productions in the three-state region. The strata and number of segments allocated to each appear in figure D-1.

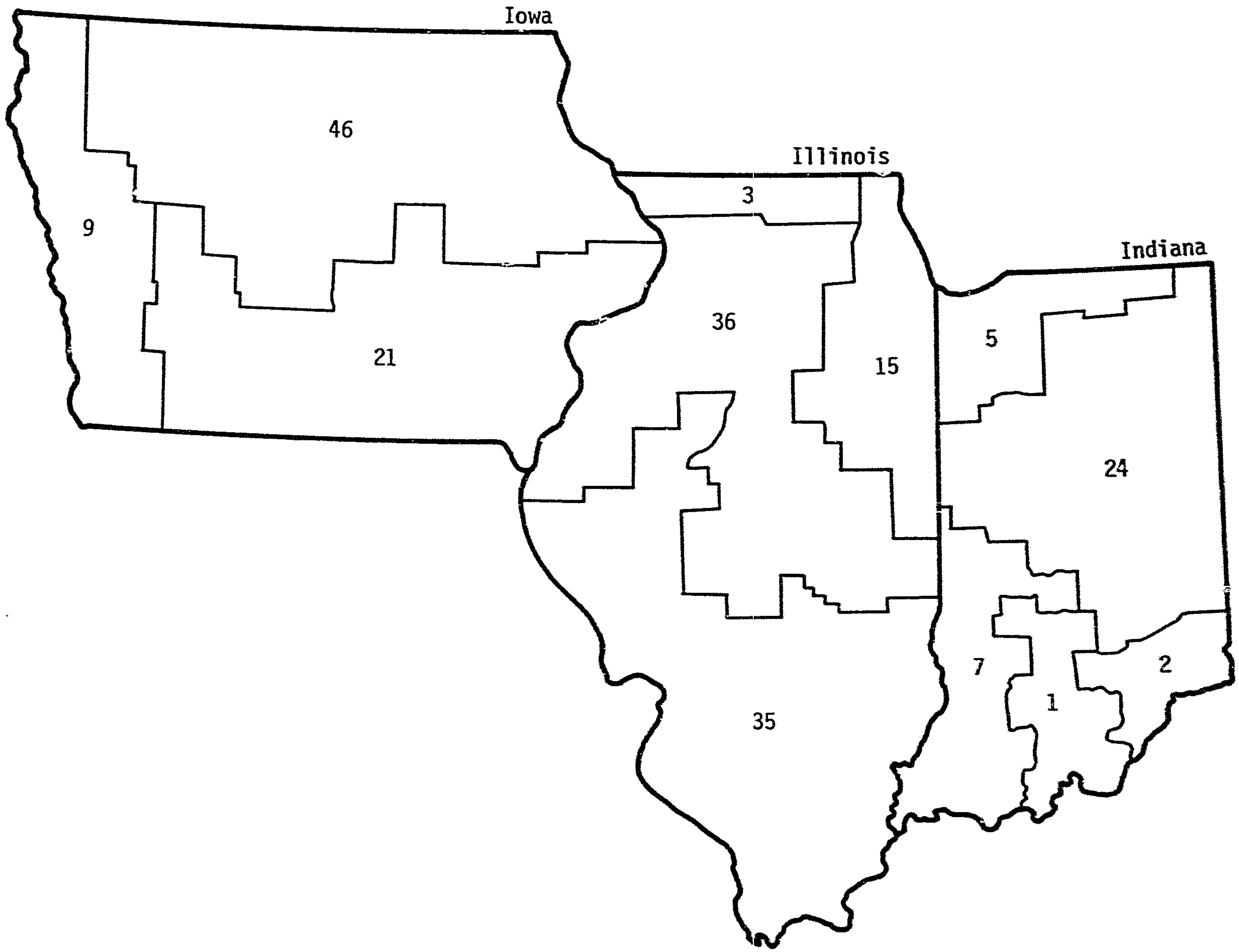


Figure D-1.- Allocation of segments to strata.

Each entire state was one yield stratum. That is, yield numbers were given for Iowa, Illinois, and Indiana, based on the actual yield in 1978 for each of these three states.

D.2 MONTE CARLO SIMULATION

Three types of simulations were performed: yield, cloud cover, and segment-level proportion estimates.

D.2.1 YIELD SIMULATION

Each time a simulation run was performed, a yield estimate was generated for each state. The procedure was simply to use the known yield for 1978 as the mean and the NOAA yield model variance as the variance of a normal distribution. A pseudorandom number from this distribution was then selected by the computer and this number was fed into the Grouped Optimal Aggregation Technique as the yield number.

D.2.2 CLOUD COVER SIMULATION

The simulation was run using five acquisition rates, namely, 10 percent, 25 percent, 50 percent, 75 percent, and 100 percent. For a particular acquisition rate r, each segment was "acquired" with probability r or "not acquired" with probability 1-r. In this study a simple but rather unrealistic assumption was made that each segment would be acquired or not acquired independently of any other segment. Thus, the number of segments acquired in an acreage stratum, X, follows the following binomial distribution, where N represents the number of segments allocated to the stratum:

$$Pr(X = x) = \binom{N}{x} r^x (1 - r)^{N-x}$$

$$x = 0,1,2,\cdots,N \qquad (2.1)$$

D.2.3 PROPORTION ESTIMATE SIMULATION

For each segment that was "acquired," a crop proportion estimate, p, was simulated. The expected value, μ, of p was taken to be the actual stratum proportion for 1978, and the variance, σ^2, was taken to be the sum of the classification variance and the sampling variance. The former was estimated from actual Landsat segments that had been worked by analysts, and the latter was estimated using the within-stratum variance estimation model (ref. 3).

This second variance was estimated for each acreage stratum, while the first was considered constant over all strata.

The distribution of p was a mixture of a discrete and continuous distribution as described below. Since Landsat segments occasionally contain none of the crop of interest, the establishment of p as zero or positive had to be determined. The probability of a zero proportion estimate, say α, was taken to be the probability that a normally distributed random variance having mean μ and variance σ^2 would be less than or equal to zero (see figure D-2).

Once α was determined for the stratum, the proportion was assigned the value zero with probability α. If p was not zero, its value was selected randomly from a beta distribution with parameters a and b (chosen so that the distribution of p, which is a mixture of a continuous and discrete distribution, would have mean μ and variance σ^2). A typical beta density is depicted in figure D-3.

D.2.4 DESIGN OF THE EXPERIMENT

A total of 1000 runs of the simulation were performed — 100 for each of the 10 combinations of acquisition rate and crop type. The simulation layout is depicted in table D-1.

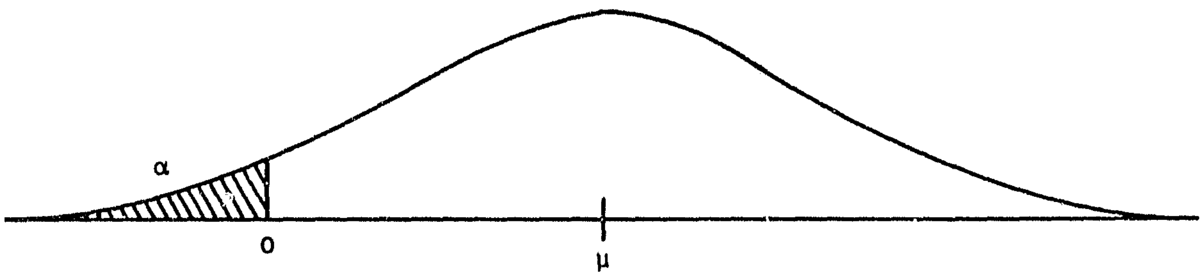
D.3 RESULTS OF THE SIMULATION

In this section, the questions that the simulation was designed to address and the results of the simulation study are presented.

TABLE D-1.- SIMULATION LAYOUT

Acquisition rate, %	Crop type		Total
	Corn	Soybeans	
10	100	100	200
25	100	100	200
50	100	100	200
75	100	100	200
100	100	100	200
Total	500	500	1000

Note: Entries in the table denote the number of simulations performed.



μ = mean of simulated proportion estimate
= stratum crop proportion in 1978
σ = probability simulated proportion estimate equals zero

$$= \int_{-\infty}^{0} (2\pi\sigma)^{-1} \exp\left[\frac{-1}{2\sigma} (t-\mu)^2\right] dt,$$

where σ = standard deviation of p.

Figure D-2.- Determination of the probability a proportion estimate is zero.

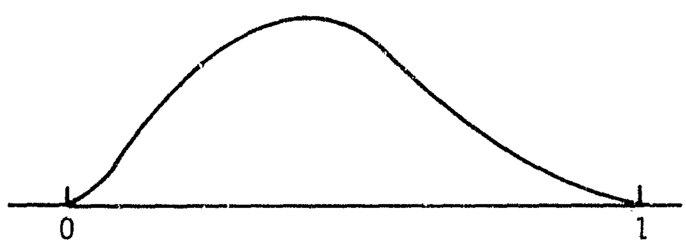


Figure D-3.- Typical beta density.

D.3.1 QUESTIONS ADDRESSED

The simulation was performed in an attempt to answer the following questions:

a. Does the MAP provide a 5-percent C.V. for production of each of the two crops for which the segments were allocated?

b. Does the Grouped Optimal Aggregation Technqiue provide unbiased acreage and production estimates for each state and for the 3-state region?

c. Are the variance (C.V.) estimates computered by the Grouped Optimal Aggregation Technique correct?

d. Is the Grouped Optimal Aggregation Technique robust against loss of data?

The following sections show that the answer to each of these questions is affirmative.

D.3.2 MULTICROP ALLOCATION

Table D-2 illustrates the effectiveness of the MAP in meeting the goal of a 5-percent C.V. for production of each crop. Note that C.V.'s are somewhat higher for individual states than for the 3-state region. This can be explained by noting that the goal of the allocation was to provide a 5-percent C.V. for the entire region, not for any individual state. The entries in the table indicate the sample C.V.'s computed from the 100 simulations on each crop type with 100-percent acquisitions.

D.3.3 UNBIASED AGGREGATIONS

Table D-3 shows the relative bias of the aggregated production and acreage estimates at the state and at the 3-state level for corn and soybeans at both the 100-percent and 10-percent acquisition rates. Clearly, no detectible bias exists at the 100-percent acquisition rate, and the small bias seen for soybeans at the 10-percent acquisition rate could easily be due to chance. In fact, none of the biases are significantly different from zero (statistically) at any reasonable significance level. Hence, the conclusion is that no procedural bias has been detected in the Grouped Optimal Aggregation Technique.

TABLE D-2.- SAMPLE C.V.'s

State	Sample C.V.
Corn	
Illinois	0.060
Indiana	.071
Iowa	.071
All 3 states	.047
Soybeans	
Illinois	0.070
Indiana	.087
Iowa	.092
All 3 states	.052

TABLE D-3.- RELATIVE BIAS OF AGGREGATED PRODUCTION ESTIMATES

State	Acquisition rate, 100%	Acquisition rate, 10%
Corn		
Illinois	-0.001	-0.002
Indiana	.000	- .006
Iowa	.001	.009
All 3 states	.000	.002
Soybeans		
Illinois	0.003	-0.014
Indiana	-.007	- .009
Iowa	.000	- .023
All 3 states	.000	- .016

D.3.4 VARIANCE ESTIMATES

Table D-4 shows the average of the estimated C.V.'s computed by the Grouped Optimal Aggregation Technique over 100 simulations for corn and soybeans at 100-percent acquisition rates. From this table it is apparent that the variance estimation procedure used in the Grouped Optimal Aggregation Technique provides good variance (C.V.) estimates.

D.3.5 MISSING DATA

A consistent problem inherent in aerospace remote sensing is nonresponse due to cloud cover. One of the main reasons for developing the Grouped Optimal Aggregation Technique was to provide an improved method of handling nonresponse. It is, of course, unreasonable to expect any aggregation procedure to perform as well with missing data as with complete data; however, a robust procedure can be expected to provide C.V.'s which are approximately proportional to $n^{-1/2}$, where n is the sample size. Figures D-4 and D-5 give C.V.'s for production and acreage as computed from the simulation results for corn and soybeans over the 3-state region. Also shown is $kn^{-1/2}$, where k is chosen such that $kn^{-1/2} = .05$ at the 100-percent acquisition rate. These figures show that the Grouped Optimal Aggregation technique is quite robust against nonresponse.

D.4 CONCLUSIONS

We have seen that the MAP provides a good allocation for multiple crops surveys, at least in the two-crop case. The Grouped Optimal Aggregation Technique was seen to give unbiased acreage and production estimates, provided the input segment proportion estimates are unbiased. The Grouped Optimal Aggregation Technique gives good variance estimates, and it is seen to be robust against nonresponse. On the basis of this simulation study, it is therefore recommended that the MAP and Grouped Optimal Aggregation Technique be used as baseline procedures in the 1981 experiments.

TABLE D-4.- AVERAGE OF ESTIMATED C.V.'S PRODUCED BY GROUPED OPTIMAL AGGREGATION TECHNIQUE AND SAMPLE C.V.'S

State	Sample C.V.	Average of estimated C.V.
Corn		
Illinois	0.060	0.053
Indiana	.071	.068
Iowa	.071	.064
All 3 states	.047	.040
Soybeans		
Illinois	0.070	0.075
Indiana	.087	.096
Iowa	.092	.085
All 3 states	.052	.053

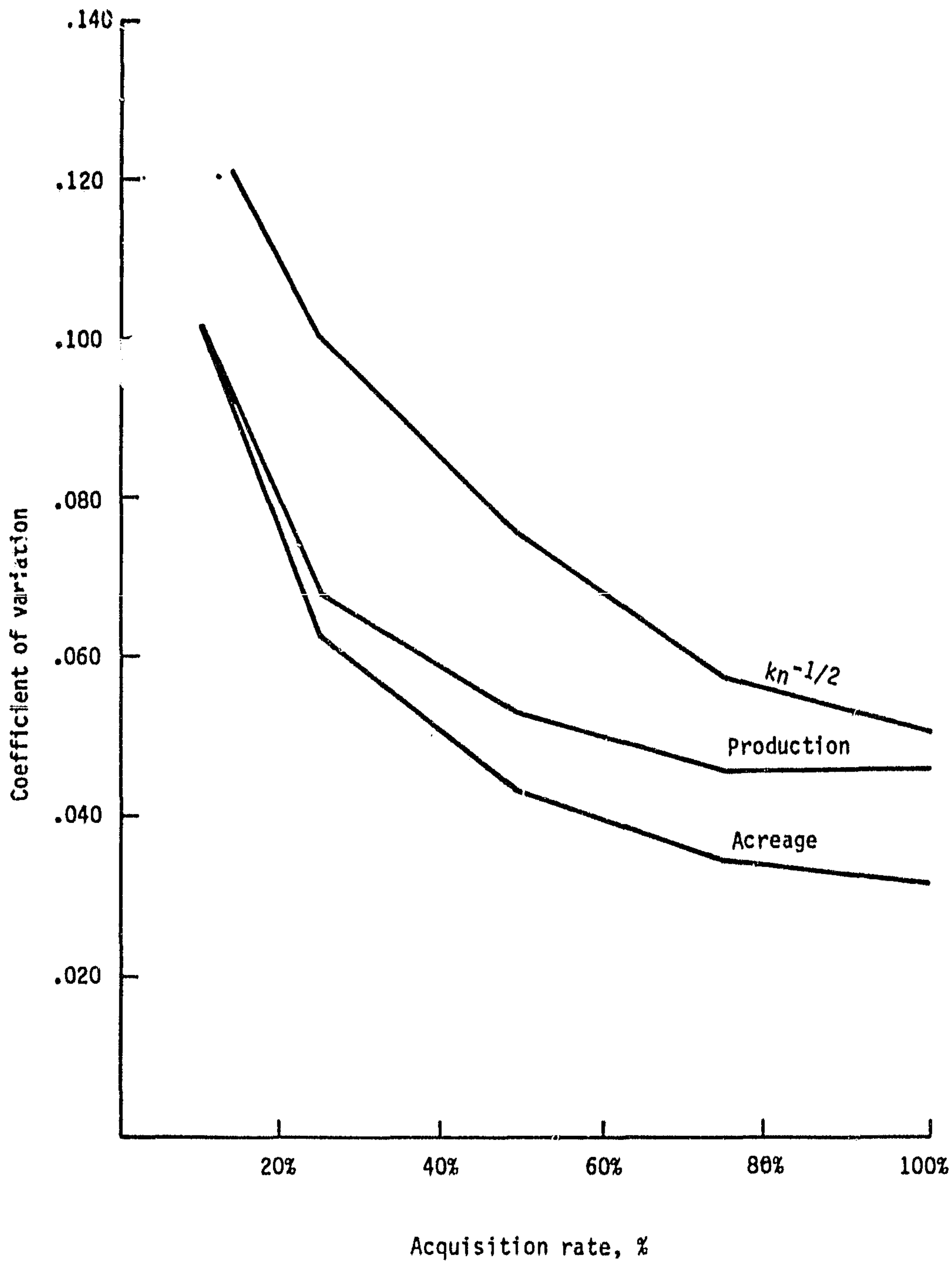


Figure D-4.- Effects of nonresponse on C.V. for corn in the 3-state region.

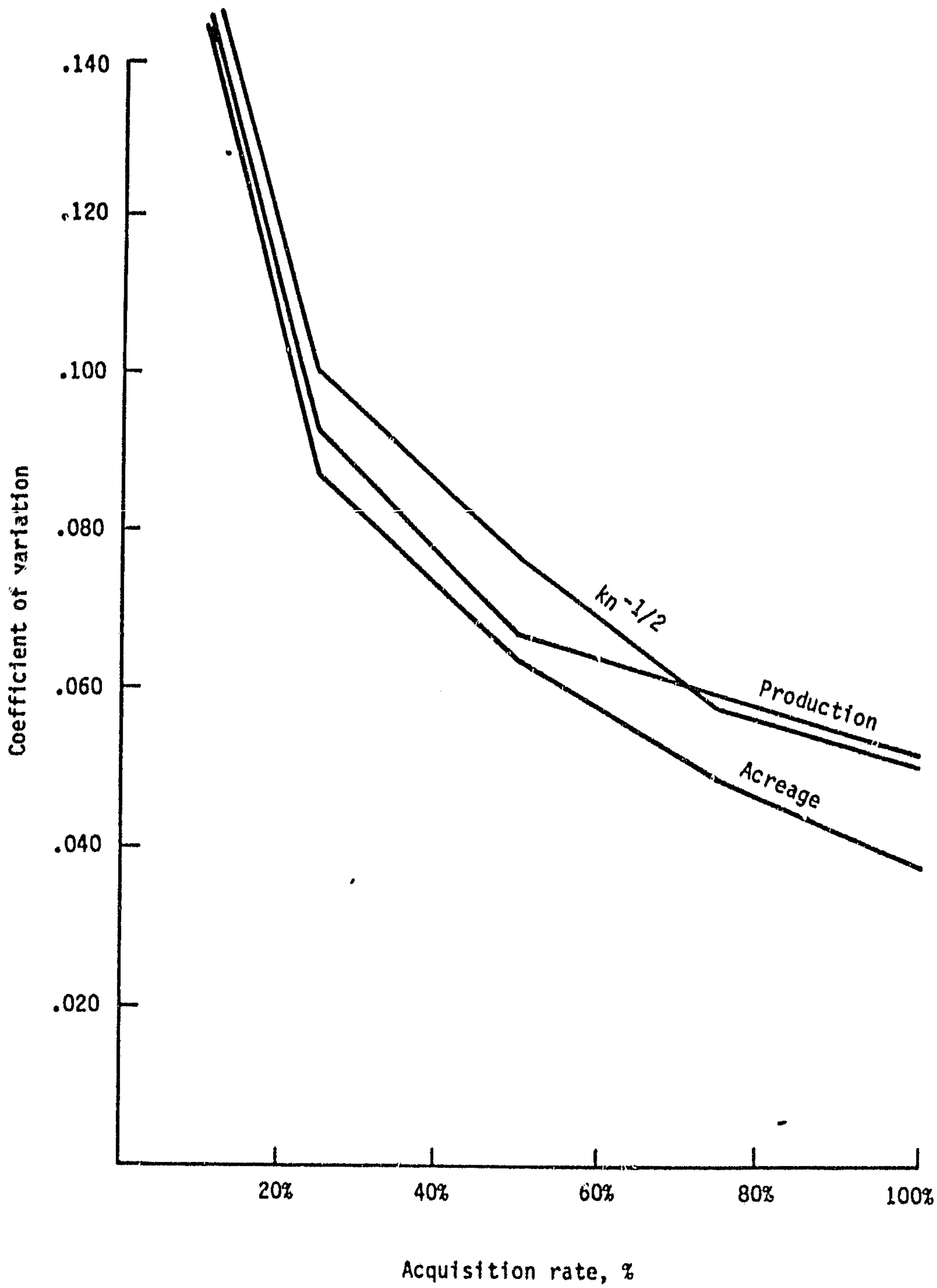


Figure D-5.- Effects of nonresponse on C.V. for soybeans in the 3-state region.

REFERENCES

1. Hartley, H. O.; Hughes, T. H.; and Sielken, R. L.: A Computer Implementation of the Multicrop Sampling Strategy, Technical Report - 27, Institute of Statistics, TAMU, May 1979.

2. Feiveson, A. H.: Weighted Ratio Estimation for AgRISTARS. Draft Technical Memorandum, NASA-JSC, October 1980.

3. Chhikara, R. S.; and Perry, C. R.: Estimation of Within Stratum Variance for Sample Allocation. Technical Report, JSC-16343, NASA-JSC, July 1980.

U.S. GOVERNMENT PRINTING OFFICE: 1981--569 011 2410