01

ERTS 111

E7.6-10336

CR-147219

CROP IDENTIFICATION AND ACREAGE MEASUREMENT UTILIZING LANDSAT IMAGERY

(E76-10336) CROP IDENTIFICATION AND ACREAGE MEASUREMENT UTILIZING LANDSAT IMAGERY Final Report (Department of Agriculture) 206 p HC $7.75 CSCL 02C

N76-23661

Unclas

G3/43 00336

Statistical Reporting Service
United States Department of Agriculture
Washington, D.C. 20250

MARCH 1976

1013A

CROP IDENTIFICATION AND ACREAGE MEASUREMENT UTILIZING LANDSAT IMAGERY

Donald H. Von Steen and William H. Wigton

Statistical Reporting Service
United States Department of Agriculture
Washington, D.C. 20250

March 1976

Original photography may be purchased from:
EROS Data Center
10th and Dakota Avenue
Sioux Falls, SD 57198

Prepared for: Goddard Space Flight Center
Greenbelt, Maryland 20771

#1013A, S-70251AG3

Technical Monitor: Frederick Gordon

TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle CROP IDENTIFICATION AND ACREAGE MEASUREMENT UTILIZING LANDSAT (FORMERLY ERTS) IMAGERY		5. Report Date
		6. Performing Organization Code
7. Author(s) Donald H. Von Steen and William H. Wigton		8. Performing Organization Report No.
9. Performing Organization Name and Address Statistical Reporting Service United States Department of Agriculture Washington, D. C. 20250		10. Work Unit No.
		11. Contract or Grant No. #1013A, S-70251AG3
12. Sponsoring Agency Name and Address Goddard Space Flight Center Greenbelt, Maryland 20771		13. Type of Report and Period Covered Final
		14. Sponsoring Agency Code

15. Supplementary Notes

Mr. Frederick Gordon, Technical Monitor

16. Abstract

This paper summarizes work completed by the Statistical Reporting Service, USDA, using ground observations obtained from our area frame. This frame is the key to making good estimates of crop acreages and yield as well as income and livestock. It turns out to be essential for making good use of satellite imagery as well. One critical step in using multi-spectral scanner CCT's is to identify and estimate the amount of energy for each crop that is reflected in each band. Area frame data can provide unbiased estimates of crops reflectances in the whole image, since the data were selected scientifically from the whole image.

Another equally critical step in making immediate use of satellite imagery is needed after the image has been classified. By observing how known probability data are classified or more accurately misclassified, we can adjust total image classifications so that pixels can be converted to acres in a statistically sound way.

The work presently being done by SRS directly evolved from this study. The project's intent is to reduce sampling error of the crop acreage estimates.

17. Key Words Suggested by Author	18. Distribution Statement
ERTS Data Use Acreage Estimation Crop Acreage and Production Estimates	

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	190	

"And the Lord said, 'Behold, they are one people, and they have all one language; and this is only the beginning of what they will do; and nothing that they propose to do will now be impossible for them. Come, let us go down, and there confuse their language, that they may not understand one another's speech'... Therefore, its name was called Babel, because there the Lord confused the language of all the earth..."

Genesis 11:6-9

"And we have been misleading each other ever since."

Dr. Thomas Szasz,
The Second Sin,
1974

But this is an effort to dispel some of this confusion with respect to remote sensing.

The Authors

ACKNOWLEDGEMENTS

We wish to express our gratitude to the LARS staff for their support in this project. Our special thanks go to Dr. Marvin Bauer and Ms. Jeanne Etheridge for the time and assistance rendered by them which enabled LARSYS to accomplish our needed tasks.

In addition, we tender our appreciation to our scientific monitor, Mr. Frederic Gordon for his patience during the final draft of this document.

Preface

This report provides step by step details of nearly two and one half years of work at the Statistical Reporting Service (SRS) under NASA contract AG328. The contract specified that we perform crop classification of LANDSAT data (formerly ERTS) in four states. All the classification was performed at Purdue using LARSYS. Other systems were tried, but LARSYS was flexible enough to suit our needs.

The basic objective was to evaluate LANDSAT data and to find ways to use this data to improve the present acreage estimates. This is no easy task since the current estimates are cost effective and sufficient in most areas - the exception being local estimates.

The procedures that were developed were to improve state or strata within state estimates. This project is being followed through with 1975-76 program, which is to perform wall to wall classification of LANDSAT data in Illinois, Kansas, and 44 counties in Texas.

Specifically, the objectives as presented in the original proposal are:

1. Develop methods of crop species identification from space imagery by photo interpretation and discrimination technique within the context of: (1) multiple frame sampling, and (2) an alternative approach using the techniques of double sampling. The study would compare the accuracy of results using LANDSAT imagery compared with the additional improvement using aircraft imagery when both are combined with ground data.

2. Develop methods for estimating crop acreages by extracting information from space imagery in the context of the agencies operating constraints.

The scope of this ambitious study was somewhat reduced since much of the imagery came very late in the growing season.

Less than optimum imagery was available, so less than optimum results were obtained. Nevertheless, the conclusions were that if satellite imagery were available and if software were available, LANDSAT type data could be useful and provide substantial gains in state estimates. SRS has moved ahead to build software, so that when the imagery is available, SRS will be ready to use it. However, it is vital that the data be ready for processing within 48 hours after it has been taken. Otherwise, it is of little value.

Table of Contents

TABLE OF FIGURES

Table of Figures cont.

LIST OF CONTRIBUTORS

Name	Status	Time
Donald H. Von Steen	Principal Investigator	1972-1975
Harold F. Huddleston	Principal Research Statistician	1972-1975
Paul D. Hopkins	Statistician & Data Processing	1972-1975
Fred B. Warren	Statistician & Date Processing	1972-1975
Edward Camara	Statistician	1972-1975
William H. Wigton	Statistician	1972-1975
Ronald Bosecker	Statistician	1972-1973
John E. Ridgely	Statistician	1972-1972
Ronald J. Steele	Statistician	1972-1974
Chapman P. Gleason	Statistician	1974-1975
Paul W. Cook	Statistician	1973-1975
Edward E. Burgess	Statistician	1973-1974
Clare Fisk	Writer-Editor	1975-1976
Victoria M. Posey	Typist	1972-1976

I. Introduction

The Statistical Reporting Service (SRS), U.S. Department of Agriculture, prepares estimates of crops, livestock, poultry, dairy, prices, and related agricultural topics.

Crop reports provide estimates of acreages farmers intend to plant in the coming season, the acres planted and harvested, production, disposition of the crop, and remaining stocks. Forecasts of yield and production are issued monthly during the growing season based on information voluntarily provided by farmers and from counts, measurements, and observations made in sample fields by SRS enumerators.

Livestock and poultry reports include estimates of animals on farms and ranches or in feedlots. Estimates are made of breeding and production intentions; yearend estimates cover production and disposition of major livestock and poultry species. SRS also reports slaughter numbers and meat production.

Dairy reports indicate milk cows, monthly and annual milk production, and use of milk. Production of major manufactured dairy products is reported weekly and monthly.

Price reports show prices received by farmers for nearly 200 products and prices paid for about 500 items needed for production or family living. Reports cover indexes of prices received and paid, parity prices, and season average prices of crops, livestock, and livestock products.

Other reports deal with labor and wages, fertilizer, seeds, bees and honey, mink, naval stores, stocks of major commodities, cold storage holdings, exports and other agricultural elements.

The scope of agricultural estimates has increased with the demands for information by producers, processors, manufacturers, and Government program planners, but the original goal has remained steady - to help farmers market farm products more effectively.

The launching of ERTS-1 (now LANDSAT) on July 24, 1972 opened a new potential source of agricultural data. This investigation has provided SRS with an opportunity to evaluate a different source of data relative to crop acreage estimates. In addition, there was presented the opportunity to determine whether the theory of sampling is flexible enough to utilize efficiently satellite data in conjunction with other survey procedures. If it were possible to blend these sources, a substantial increase in survey accuracy would ensue.

1/ Preparing Crop and Livestock Estimates, Statistical Reporting Service, March, 1974.

The objectives of this investigation were as follows:

1. Develop methods to identify crop species utilizing satellite and aircraft imagery.

2. Develop methods of estimating crop acreages utilizing satellite imagery.

3. Within the context of multi-stage and multiple frame sampling, develop methods of utilizing all three sources of data (ground, aircraft, and satellite) to make crop acreage estimates. Combining all three sources in a statistical model should result in a marked improvement over any one source for making crop acreage estimates.

The study areas were selected Crop Reporting Districts in Missouri, Kansas, South Dakota, and Idaho. The major crops of concern were wheat, corn, cotton, soybeans, sugar beets, potatoes, alfalfa, and grain sorghum. Some of the crops are grown in only one area while others are common to two or three. This provided the opportunity to observe crops grown under different conditions.

II. Data Acquisition

2.1 Ground Observations

In order to evaluate the new methodology, one needs independently collected (control) data. For this study, ground truth collected in the same manner as is now being used by the Statistical Reporting Service, (SRS) was used as the control data for evaluating results from both the satellite and aircraft imagery.

The thrust of the ground truth portion of the LANDSAT project is to identify the crops visible from the air on previously designated areas of land. Our ground truth identifies the crop species present and the exact location of the fields for the survey.

Throughout the growing season, the species, acreage, and condition of crops in these fields are observed periodically. This provides progressive reports about crop maturity development and a record of any changes in acreage or species. This data provides survey acreage for crops which could be compared against other sources of data and corresponding estimates.

The condition of the crop in each field is noted as supplementary information. During the processing of aerial photography and satellite imagery the condition code would, in some cases, provide some basis why a corn field was classified incorrectly.

The first enumerative survey was conducted in late May and early June of 1972 by SRS. This data was used as a source of original data and was then updated by special enumerators. 1/ However, the estimates of crop acreages generated by the JES survey included both crops already planted and crops to be planted. At the time of the enumerative survey, the wheat in Missouri might still be in the field and was recorded as such on the questionnaire. In addition, the farmer's intention to plant soybeans was recorded for that same field. The LANDSAT ground truth was only concerned with crops and ground vegetation present on the day the enumerator visited the segments. For this reason, the June Enumerative Survey (JES) acreage estimates could be different from the LANDSAT acreage estimates; however, provisions were made through the updating of JES so such differences could be measured.

The LANDSAT ground truth was also used as a training device to classify aerial photography and satellite imagery. Since the exact location of each field and the crop species present in the field was known, we could identify the field on the aerial photography or satellite imagery and train the computer to recognize and identify all similar fields. After identification, a separate estimate of acreage can be generated from these other sources of data and compared against the ground truth acreage estimates.

2.1.2 Source of Ground Data

The test areas used in this study were SRS Crop Reporting Districts (CRD) A CRD is a contiguous group of counties within a state which have similar farming activities. Generally, each state is composed of about nine such districts.

Within each of these CRD's are randomly selected areas of land (segments) that range in size from about one-half square mile to three square miles. Since the CRD's are independent strata, estimates can be made for each individual strata by multiplying the segment totals by the reciprocal of their probability of selection and summing over the CRD. For the JES and the LANDSAT study, these segments are the test sites for the classification of the aircraft and satellite imagery. The information obtained from these segments on crops present constituted our ground observations.

1/ See Appendix B for a list of terms and definitions used for the June Enumerative Survey (JES) and LANDSAT fieldwork.

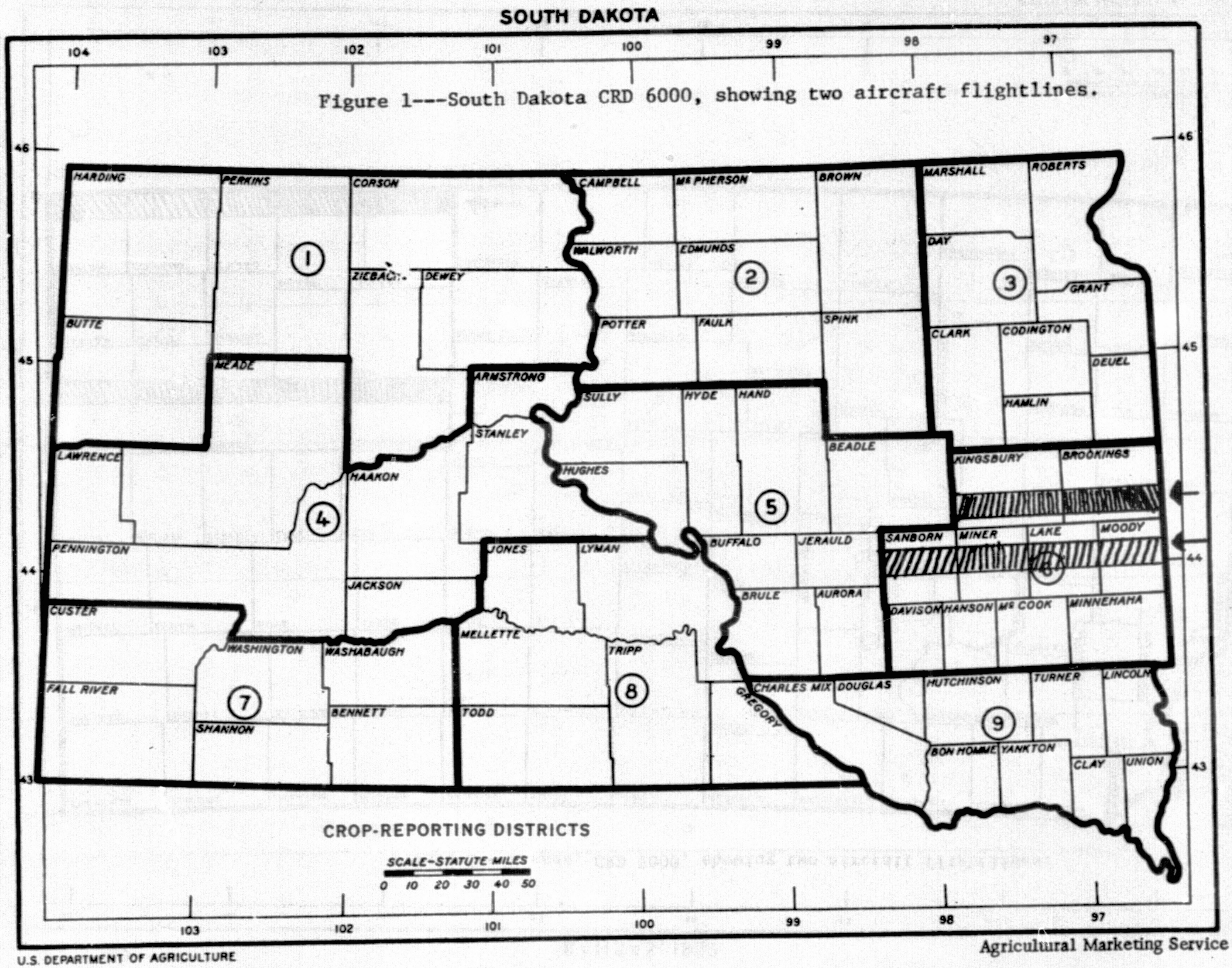
Ground data was collected for segments in CRD six in South Dakota, seven in Kansas, and nine in Missouri. In Idaho, the study area was not a CRD, but a land use stratum which included the intensive agriculture areas of Jerome, Minidoka, Twin Falls, and Cassia Counties. The study areas within each state were selected since they represented an area with a manageable volume of data and a comparable number of segments.

Table 1--States and numbers of segments in study area.

State	Number of Segments
South Dakota	50
Kansas	48
Missouri	42
Idaho	44
TOTAL	184

The four different test sites (see Figures 1, 2, 3, and 4) were selected to fulfill operational objectives. First, we wanted to monitor the progressive stages of growth and maturity of the major crop species. The original satellite launching data would have allowed monitoring crop growth from April through November of 1972. Mature wheat in Kansas could be compared to pre-headed and headed wheat in South Dakota with similar comparisons being made for other crops. Secondly, we wanted the scattered areas to help insure at least some good imagery. Imagery of cloud cover over selected areas is useless. Presumably, the distant areas would not all be engulfed with inclement weather as the aerial photography and satellite imagery were obtained. Thirdly, we wanted to answer whether or not corn in Missouri was spectrally different from corn in South Dakota, etc. Fourthly, we wanted to look at several different crops and their responses to different locational environments of soil, topography, and climate. The four State analysis gives indication of within and between State variations necessary prior to operational surveys of this type. The major crops included in the study are shown in Table 2.

Figure 1---South Dakota CRD 6000, showing two aircraft flightlines.



KANSAS, 1933

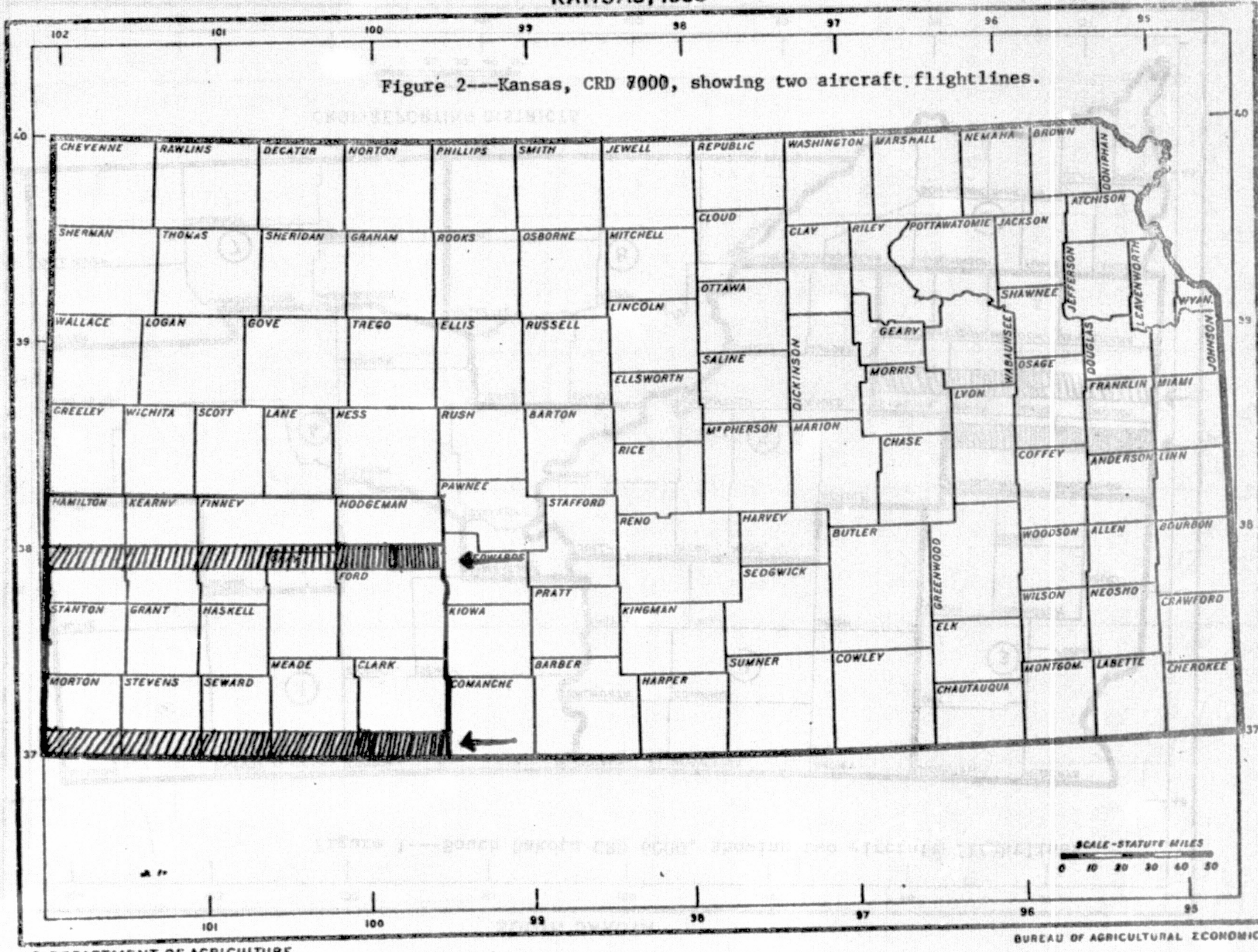


Figure 2---Kansas, CRD 7000, showing two aircraft flightlines.

REPRODUCIBILITY OF THE
ORIGINAL PAGE IS POOR

MISSOURI

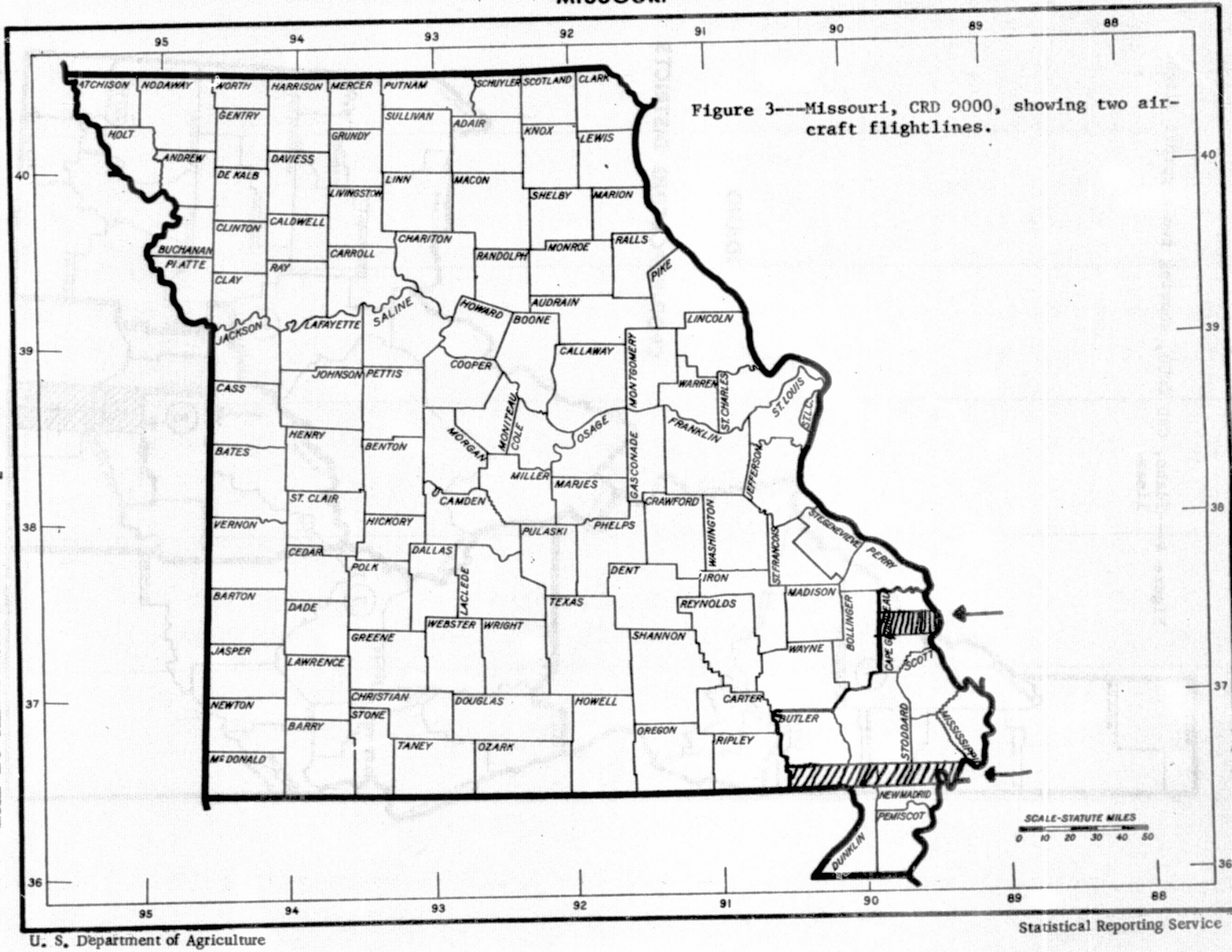


Figure 3---Missouri, CRD 9000, showing two aircraft flightlines.

Figure 4---Idaho, CRD 2000, showing two aircraft flight-lines.

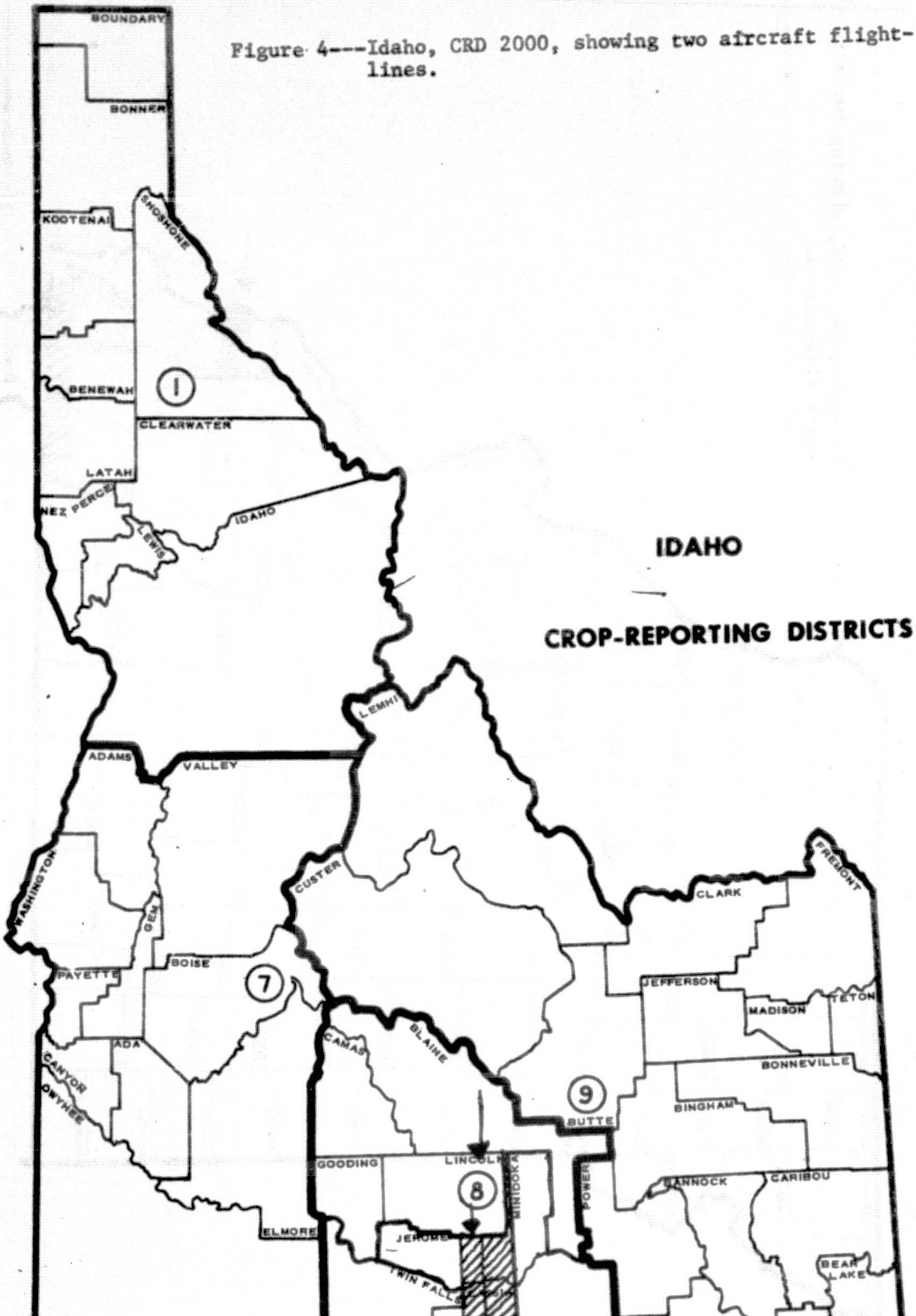


FIGURE 5 PRECEDING PAGE BLANK NOT FILMED

PRECEDING PAGE BLANK NOT FILMED

REPRODUCIBILITY OF THE ORIGINAL PAGE IS POOR

Figure 6---Ground truth record form, Kansas segment 3086, tract 3, third visit.

OMB NUMBER 40-S72045
APPROVAL EXPIRES 6-30-73
SEGMENT (3086) TRACT (03) PAGE 01 OF 01

ERTS GROUND TRUTH (76) KANSAS (47)

FIELD NO.	FIRST VISIT (JES) ACRES	FIRST VISIT CROP	SECOND VISIT JULIAN DATE (221) ACRES	CROP	CONDITN	THIRD VISIT JULIAN DATE (___) ACRES	CROP	CONDITN	FOURTH VISIT JULIAN DATE (___) ACRES	CROP	CONDITN	FINAL VISIT JULIAN DATE (___) ACRES	CROP	CONDITN
01	8*	0000	8*	OTHR	----	0*	----	----	0*	----	----	0*	----	----
02	44*	CORN	44*	CORN	TASS	0*	----	----	0*	----	----	0*	----	----
03	1*	WNWH	1*	FLOW	CLEA	0*	----	----	0*	----	----	0*	----	----
04	9*	PSTR	9*	PSTR	GREE	0*	----	----	0*	----	----	0*	----	----
05	136*	PSTR	136*	PSTR	GREE	0*	----	----	0*	----	----	0*	----	----
06	119*	FLOW	39*	FLOW	CLEA	0*	----	----	0*	----	----	0*	----	----
07	28*	WNWH	28*	GSOR	GREE	0*	----	----	0*	----	----	0*	----	----
08	150*	FLOW	125*	FLOW	CLEA	0*	----	----	0*	----	----	0*	----	----
09	135*	WNWH	135*	WNWH	HARV	0*	----	----	0*	----	----	0*	----	----
10	0*	----	25*	GSOR	GREE	0*	----	----	0*	----	----	0*	----	----
11	0*	----	80*	GSOR	GREE	0*	----	----	0*	----	----	0*	----	----
TOT	630*		630*			0*			0*			0*		

PERMISSION UNASKED

SEGMENT (3086) TRACT (03) PAGE 01 OF 01

Visit one (base) data was obtained directly from the JES questionnaire, which was completed in late May 1972 and/or early June 1972. 1/ The JES data was identified and keypunched for all "fields" in the segments. The identification of each field was required in order to delete crops which might have been reported as fields to be planted at a later date. For example, a 20 acre field could be recorded both as wheat, and also as soybeans to be planted after the wheat was harvested. Since aerial photography and satellite imagery would record only crops present, the ground observation could only correspond to what was in the field at the time of visit, and only the wheat would be punched. If the wheat field was now soybeans, this change was made during the update.

After completion of all the ground observations, the four State Statistical Offices involved in the LANDSAT study were sent an evaluation form to evaluate the computer printout recording form. From the answers to the evaluation, the following can be said:

1. The Form Printout is a workable method of collecting ground observations. There might have been a small problem orienting the enumerators to a different form than the accustomed one. However, with training, the transition was short. The enumerators were able to record acres and crop species without difficulty.

2. The crop condition codes were generally adequate, but several suggestions were made. The suggestions were a) call this "State of Growth" rather than "Condition," b) change the grain codes from "prefruit" to "blade" and "fruit" to "heading," c) remove pasture from the hays and code the pastures as lush, grazed, and range, and finally d) add the code weedy to fallow.

3. The new printout format did not create unusual editing or keypunching situations.

2.1.4 Average Field Size

Classification results from LANDSAT imagery indicate that field size may have a significant affect on how well the classification might be. Also, early reported results by other investigators suggested that relatively poor classification was obtained from fields less than 20 acres. Several inquiries to the Statistical Reporting Service for information on size of field prompted the preparation of a detailed tabulation of fields by size

---

1/ See Appendix B for a copy of a JES questionnaire and the keypunching instructions for the LANDSAT survey.

and by crop (See Tables 3, 4, 5, and 6). The data for this tabulation is from the 1972 SRS JES in the four test sites. It should be pointed out that this information only represents the four test areas.

In the Missouri test site, 28.7 percent of the fields are 20 acres or greater and account for 68 percent of the land area. Thirty-eight percent of the cotton fields are greater than 20 acres, but account for 73 percent of the reported cotton acreage. Forty-one percent of the soybean fields were 20 acres plus and represents 77.5 percent of the soybean acreage. The average size of all fields in Missouri was 17.11 acres.

South Dakota reported that 92 percent of the corn acreage and 89 percent of the oats were in fields larger than 20 acres. Overall, 52 percent of the reported fields were greater than 20 acres with an average field size of 28.74 acres. The average field size needs to be viewed with some caution in that it can be heavily influenced by large or small acreages for relatively unimportant land uses such as pasture, farmstead, etc.

Kansas showed 98.5 percent, 99.1 percent, 98.5 percent, 95.6 percent of the corn, wheat, sorghum, and alfalfa acreage respectively, were grown in fields larger than 20 acres. Field size should not be a limiting factor in identifying these crops in Kansas. Average size of all fields in Kansas was 108.31 acres.

The test area in Idaho contained some large areas of waste and pasture which influenced the average field size and the distribution. About 50 percent of the corn was planted in fields larger than 20 acres. Eighty-five percent of the barley was in 20 acre plus fields. Ninety-four percent of the potatoes were contained in fields larger than 20 acres. About 65 percent of the sugar beets were grown in 20 acres plus fields.

If field size is a factor in one's ability to do crop classification, the results in Kansas should be substantially better than in the other three states. Field shape may be a greater limiting factor than size, particularly in areas which contain irregular fields.

2.1.5 <u>Timing and Workload of Fieldwork</u>

Because of the delay in the launch of LANDSAT-1, the update surveys did not begin until August 1972. Prior to the first visits, a training school was conducted in each State involved. The training was to 1) instruct State Statistical Offices (SSO) personnel regarding enumeration, editing, keypunching, and mailing procedures, and 2) instruct enumerators regarding the collection of ground observations. 1/

1/ See Appendix A for Enumerator Instructions, for Ground Observation Editing Instructions, and B for Ground Observation SSO Keypunching Instructions.

Table 3 --Distribution of fields by size and crop for the four county test areas in South Central IDAHO, based in June Survey Data.

Crop	0-4.9 Acres				5-9.9 Acres				10-14.9 Acres				15-19.9 Acres				20-29.9 Acres				30 + Acres			
	Fields		Acres		Fields		Acres		Fields		Acres		Fields		Acres		Fields		Acres		Fields		Acres	
	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
Farmstead, etc.	138	55.9	303.9	1.2	46	18.6	288.1	1.1	18	7.3	213.1	0.8	5	2.0	81.0	0.3	14	5.7	333.3	1.3	26	10.5	24,708.8	95.3
Corn	6	10.9	15.1	1.7	13	23.6	95.1	10.9	11	20.0	134.7	15.4	11	20.0	180.0	20.6	9	16.4	214.8	24.5	5	9.1	235.6	26.9
Oats	1	25.0	2.2	4.2	-	-	-	-	-	-	-	-	3	75.0	50.5	95.8	-	-	-	-	-	-	-	-
Barley	6	7.6	13.8	0.5	9	11.4	70.3	2.3	10	12.7	116.9	3.7	15	19.0	253.7	8.1	11	13.9	257.1	8.2	28	35.4	2,414.1	77.2
Winter Wheat	-	-	-	-	11	29.7	78.7	5.0	8	21.6	91.5	5.9	1	2.7	17.0	1.1	4	10.8	88.7	5.7	13	35.2	1,284.3	82.3
Mixed Grain	10	21.3	32.3	5.2	16	34.0	115.3	18.6	9	19.2	113.7	18.3	3	6.4	50.5	8.1	4	8.5	93.0	15.0	5	10.6	216.2	34.8
Spring Wheat	5	17.2	19.3	4.4	8	27.6	51.7	11.8	8	27.6	91.9	21.1	3	10.3	53.5	12.3	1	3.5	22.0	5.0	4	13.8	198.0	45.4
Potatoes	-	-	-	-	1	5.3	8.0	0.8	2	10.5	27.6	2.8	1	5.2	18.0	1.8	1	5.3	25.0	2.6	14	73.7	905.0	92.0
Sweet Corn	2	13.3	6.5	1.3	2	13.4	12.0	2.5	3	20.0	35.1	7.3	1	6.7	16.3	3.4	2	13.3	41.7	8.7	5	33.3	369.6	76.8
Vegetables	1	100.0	1.0	100.0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Dry Beans	10	7.6	31.1	1.4	30	22.7	223.7	10.3	37	28.0	434.2	19.9	20	15.1	320.1	14.7	20	15.2	467.5	21.5	15	11.4	701.3	32.2
Dry Field Peas	3	20.0	9.9	4.1	-	-	-	-	4	26.7	49.0	20.4	4	26.6	70.5	29.3	3	20.0	77.0	32.0	1	6.7	34.0	14.2
Wild Hay	1	100.0	4.0	100.0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Alfalfa	40	15.3	132.2	2.6	68	26.0	472.0	9.4	52	19.8	621.5	12.4	34	13.0	565.5	11.3	30	11.4	729.3	14.6	38	14.5	2,486.2	49.7
Other Hay	-	-	-	-	3	60.0	21.9	31.1	-	-	-	-	1	20.0	15.6	22.1	-	-	-	-	1	20.0	33.0	46.8
Clover	-	-	-	-	2	100.0	14.8	100.0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Sugar Beets	8	11.1	25.6	1.7	13	18.0	95.8	6.5	17	23.6	193.5	13.1	12	16.7	204.9	13.8	11	15.3	274.8	18.6	11	15.3	684.5	46.3
Apples	-	-	-	-	1	100.0	7.0	100.0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Pasture	70	30.8	170.4	0.1	47	20.7	299.7	0.1	19	8.4	228.1	0.1	14	6.2	223.4	0.1	17	7.5	397.7	0.1	60	26.4	455,091.6	99.5
Cropland Pasture	5	20.0	12.1	3.3	10	40.0	72.3	19.5	5	20.0	56.5	15.2	1	4.0	19.0	5.1	-	-	-	-	4	16.0	211.0	56.9
Summer Fallow	5	12.8	16.7	0.9	8	20.5	53.4	2.9	2	5.1	28.3	1.6	2	5.1	31.0	1.7	7	18.0	177.5	9.8	15	38.5	1,509.6	83.1
Other Crops	-	-	-	-	-	-	-	-	-	-	-	-	1	100.0	17.0	100.0	-	-	-	-	-	-	-	-
Soil Improvement Crops	3	33.3	8.0	10.1	4	44.5	28.9	36.4	1	11.1	11.0	13.9	-	-	-	-	-	-	-	-	1	11.1	31.4	39.6
Idle Cropland	21	51.2	61.0	13.9	14	34.2	93.7	21.4	2	4.9	26.5	6.1	1	2.4	15.0	3.4	-	-	-	-	3	7.3	241.6	55.2
TOTALS:																								
Number of fields	335				306				208				133				133				248			
Acres	865.1				2,102.4				2,473.1				2,202.5				3,199.4				491,355.8			
Number of fields % of total	24.5				22.4				15.2				9.7				9.7				118.2			
Acres % of total	0.2				0.4				0.5				0.5				0.6				97.8			
Average size of field-acres	2.6				6.9				11.9				16.6				24.1				1,973.3			
Average Size of all fields = 367.91 acres																								

Table 4--Distribution of number of fields by size and crop for Crop Reporting District 7, based on June Survey Data - KANSAS

Crop	0-4.9 Acres				5-9.9 Acres				10-14.9 Acres				15-19.9 Acres				20-29.9 Acres				30 + Acres			
	Fields		Acres		Fields		Acres		Fields		Acres		Fields		Acres		Fields		Acres		Fields		Acres	
	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
Farmstead, etc.	41	56.9	82.5	20.0	19	26.4	106.8	25.8	5	6.9	50.0	12.1	2	2.8	34.0	8.2	3	4.2	68.0	16.5	2	2.8	72.1	17.4
Corn	2	4.2	7.5	0.2	2	4.2	14.0	0.4	2	4.2	20.0	0.5	1	2.1	18.0	0.4	2	4.1	45.0	1.1	39	81.2	3,880.4	97.4
Oats	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	100.0	40	100.0	-	-	-	-
Barely	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	100.0	152.0	100.0
Winter Wheat	5	2.5	11.8	0.1	4	2.0	27.2	0.2	1	0.5	12.0	0.1	5	2.6	83.7	0.5	23	11.6	560.8	3.4	160	80.8	15,747.8	95.7
Rye	-	-	-	-	-	-	-	-	1	100.0	10.0	100.0	-	-	-	-	-	-	-	-	-	-	-	-
Vegetables	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	100.0	40.0	100.0
Sorghum	1	1.5	2.0	0.1	1	1.5	9.0	0.1	1	1.5	14.0	0.2	5	7.4	81.0	1.1	3	4.5	73.0	1.0	56	83.6	7,138.1	97.5
Alfalfa	1	4.3	4.0	0.2	1	4.4	7.5	0.5	2	8.7	22.0	1.4	2	8.7	34.0	2.3	2	8.7	55.0	3.5	15	65.2	1,435.7	92.1
Other Hay	2	20.0	6.0	4.3	4	40.0	26.0	18.4	-	-	-	-	-	-	-	-	2	20.0	44.0	31.2	2	20.0	65.0	46.1
Sugar Beets	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	4	100.0	129.0	100.0
Pasture	-	-	-	-	3	2.6	24.0	0.1	8	6.9	91.9	0.3	3	2.6	49.4	0.2	5	4.3	129.5	0.4	97	83.6	31,190.6	99.0
Cropland Pasture	1	6.2	3.9	0.4	2	12.5	13.8	1.4	2	12.6	22.6	2.3	1	6.2	17.0	1.8	-	-	-	-	10	62.5	918.0	94.1
Summer Fallow	5	2.7	15.3	0.1	3	1.6	20.7	0.1	6	3.2	64.0	0.4	10	5.3	167.2	0.9	10	5.4	241.8	1.3	153	81.8	17,798.9	97.2
Other Crops	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	100.0	60.0	100.0
Soil Improvement Crops	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3	100.0	515.0	100.0
Idle	-	-	-	-	1	25.0	6.0	6.5	-	-	-	-	1	25.0	18.0	19.6	-	-	-	-	2	50.0	68.0	73.9
TOTALS:																								
Number of fields	58				40				28				30				52				546			
Acres	133.0				255.0				306.5				502.3				1,257.1				79,210.6			
Number of fields % of total	7.7				5.3				3.7				4.0				6.9				72.4			
Acres % of total	0.2				0.3				0.4				0.6				1.5				97.0			
Average size of field-acres	2.3				6.4				10.9				16.7				24.2				145.1			
Average Size of all fields = 108.31 acres																								

Table 5—Distribution of number of fields by size and crop for Crop Reporting District 9, based on June Survey Data – MISSOURI

Crop	0-4.9 Acres				5-9.9 Acres				10-14.9 Acres				15-19.9 Acres				20-29.9 Acres				30 + Acres			
	Fields		Acres		Fields		Acres		Fields		Acres		Fields		Acres		Fields		Acres		Fields		Acres	
	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
Farmstead, etc.	71	57.7	134.1	13.1	17	13.8	107.8	10.5	16	13.0	184.5	18.0	6	4.9	97.0	9.5	8	6.5	186.0	18.2	5	4.1	315.0	30.7
Cotton	18	14.1	59.7	2.0	21	16.4	139.0	4.8	18	14.0	211.0	7.2	22	17.2	368.5	12.8	16	12.5	384.0	13.2	33	25.8	1744.5	60.0
Corn	6	20.0	16.0	4.3	8	26.7	52.9	14.3	9	30.0	106.0	28.6	1	3.3	15.0	4.1	4	13.3	960	26.0	2	6.7	84.0	22.7
Oats	1	100	4.0	100	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Barley	2	100	7.0	100	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Winter Wheat	4	4.6	9.7	0.6	15	17.2	106.0	6.0	25	28.7	278.7	15.8	13	15.0	212.2	12.1	16	18.4	362.5	20.6	14	16.1	759.6	44.9
Rye	3	42.9	7.5	13.8	3	42.8	22.0	40.3	-	-	-	-	-	-	-	-	1	14.3	25.0	45.9	-	-	-	-
Watermelons	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	100	35.0	100
Soybeans	30	16.3	71.2	1.7	30	16.3	204.5	4.9	28	15.2	328.3	7.9	20	10.9	334.4	8.0	32	17.4	757.9	18.2	44	23.9	2476.0	59.3
Sorghum	2	28.5	7.0	7.7	2	28.6	12.0	13.2	1	14.3	10.0	11.0	-	-	-	-	1	14.3	25.0	27.5	1	14.3	37.0	40.6
Wild Hay	-	-	-	-	1	100	8.0	100	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Alfalfa	2	66.7	7.0	31.8	-	-	-	-	-	-	-	-	1	33.3	15.0	68.2	-	-	-	-	-	-	-	-
Other Hay	3	23.1	9.0	3.1	4	30.8	27.0	9.3	-	-	-	-	-	-	-	-	1	7.7	24.0	8.3	5	38.4	229.0	79.3
Clover	1	20.0	4.0	5.2	2	40.0	12.7	16.6	-	-	-	-	-	-	-	-	1	20.0	20.0	26.1	1	20.0	40.0	52.1
Pasture	18	28.6	56.5	4.4	11	17.5	73.2	5.8	5	7.9	58.5	4.6	7	11.1	113.3	8.9	11	17.4	248.4	19.5	11	17.5	722.6	56.8
Cropland Pasture	3	42.8	11.0	12.4	1	14.3	6.0	6.7	1	14.3	10.0	11.2	-	-	-	-	-	-	-	-	2	28.6	62.0	69.7
Other Crops	-	-	-	-	1	50.0	8.0	24.2	-	-	-	-	-	-	-	-	1	500	25.0	75.8	-	-	-	-
Soil Improvement Crops	20	47.6	41.5	14.1	13	30.9	91.4	31.1	5	11.9	55.5	18.9	1	2.4	18.8	6.4	2	4.8	47.0	15.9	1	2.4	40.0	13.6
Idle	17	25.8	50.6	7.2	19	28.8	120.2	17.1	17	25.7	196.7	28.0	5	7.6	87.8	12.5	5	7.6	102.0	14.5	3	4.5	146.0	20.7
TOTALS:																								
Number of fields	201				148				125				76				99				123			
Acres	495.8				990.7				1439.1				1262.0				2302.8				6720.7			
Number of fields % of total	26.0				19.2				16.2				9.9				12.8				15.9			
Acres % of total	3.8				9.5				10.9				9.6				17.4				50.9			
Average Size of field-acres	2.47				6.69				11.51				16.61				23.26				54.64			
Average Size of all fields = 17.11 acres																								

Table 6--Distribution of number of fields by size and crop for Crop Reporting District 6, based on June Survey Data - SOUTH DAKOTA

Crop	0-4.9 Acres				5-9.9 Acres				10-14.9 Acres				15-19.9 Acres				20-29.9 Acres				30 + Acres			
	Fields		Acres		Fields		Acres		Fields		Acres		Fields		Acres		Fields		Acres		Fields		Acres	
	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
Farmstead, etc.	33	30.5	89.8	10.5	49	45.5	337.6	39.4	14	13.0	157.0	18.4	6	5.6	93.0	10.9	3	2.7	62.0	7.3	3	2.7	115.0	13.5
Corn	5	2.9	13.9	0.2	12	6.9	92.3	1.4	15	8.6	179.0	2.7	16	9.2	270.8	4.1	42	24.1	1016.3	15.5	84	48.3	5003.6	76.1
Oats	7	5.9	22.0	0.6	10	8.4	73.1	2.0	10	8.4	113.0	3.1	12	10.1	205.1	5.5	30	25.2	679.1	18.3	50	42.0	2611.8	70.5
Barley	-	-	-	-	5	26.3	39.3	7.0	1	5.3	14.0	2.5	6	31.6	95.0	17.0	-	-	-	-	7	36.8	411.0	73.5
Winter Wheat	1	33.3	2.5	6.7	-	-	-	-	1	33.3	11.0	29.3	-	-	-	-	1	33.3	24.0	64.0	-	-	-	-
Soybeans	-	-	-	-	1	8.3	8.5	3.7	2	16.7	20.0	8.6	3	25.0	53.2	22.9	5	41.7	113.0	48.5	1	8.3	38.0	16.3
Sorghum	2	25.0	8.0	3.9	-	-	-	-	-	-	-	-	1	12.5	15.0	7.3	3	37.5	63.0	30.7	2	25.0	119.0	58.1
Wild Hay	2	8.7	4.4	0.9	7	30.5	46.5	9.1	2	8.7	26.0	5.1	1	4.3	15.5	3.0	4	17.4	89.0	17.3	7	30.4	332.0	64.6
Alfalfa	6	6.5	19.2	0.9	9	9.9	61.2	2.8	20	22.0	233.2	10.7	13	14.3	211.0	9.6	23	25.3	519.0	23.7	20	22.0	1144.0	52.3
Other Hay	10	55.5	20.0	4.0	1	5.6	5.0	1.0	2	11.1	20.0	4.1	1	5.6	16.0	3.2	-	-	-	-	4	22.2	433.0	87.7
Flax	1	5.6	2.5	0.7	1	5.5	7.0	1.9	3	16.7	35.0	9.5	2	11.1	34.0	9.3	8	44.4	192.0	52.4	3	16.7	96.0	26.2
Pasture	6	6.8	14.5	0.3	4	4.6	25.5	0.5	12	13.6	128.0	2.8	3	3.4	46.0	1.0	9	10.2	220.0	4.8	54	61.4	4186.0	90.6
Cropland Pasture	5	10.4	15.5	1.1	6	12.5	41.7	3.0	9	18.8	109.4	7.8	4	8.3	69.0	4.9	10	20.8	227.5	16.3	14	29.2	935.3	66.9
Summer Fallow	8	12.7	22.5	1.3	4	6.4	27.0	1.6	7	11.1	75.0	4.4	10	15.9	171.9	10.0	13	20.6	303.5	17.7	21	33.3	1113.8	65.0
Other Crops	7	53.7	21.7	16.2	3	23.1	21.5	16.0	2	15.4	23.0	17.1	-	-	-	-	-	-	-	-	1	7.8	68.0	50.7
Soil Improvement Crops	-	-	-	-	1	8.3	8.0	2.8	1	8.3	13.0	4.5	3	25.0	49.0	16.8	5	41.7	122.0	41.9	2	16.7	98.8	34.0
Idle	6	23.1	16.2	3.5	2	7.7	16.8	3.6	3	11.5	33.0	7.1	4	15.4	65.0	14.1	6	23.1	135.2	29.3	5	19.2	196.0	42.4
Rye	-	-	-	-	1	11.1	5.0	2.6	1	11.1	11.0	5.6	2	22.2	31.0	15.9	4	33.4	77.0	39.5	2	22.2	71.0	36.4
Spring Wheat	1	16.7	1.3	0.9	1	16.6	8.0	5.7	1	16.7	10.0	7.1	-	-	-	-	-	-	-	-	3	50.0	122.0	86.3
Durum Wheat	-	-	-	-	1	50.0	6.0	22.2	-	-	-	-	-	-	-	-	1	50.0	21.0	77.8	-	-	-	-
TOTALS:																								
Number of fields	100				118				106				87				166				283			
Acres	274.0				830.0				1210.6				1440.5				3863.6				17094.3			
Number of fields % of total	11.7				13.7				12.3				10.1				19.3				32.9			
Acres % of total	1.1				3.4				4.9				5.8				15.6				69.2			
Average Size of Field-acres	2.74				7.03				11.42				15.56				23.27				60.40			
Average Size of all fields = 28.74 acres																								

The timetable for the collection of the ground observations was:

August 3	Enumerator training schools
August 7-11	Survey fieldwork
August 11	Enumerators mail update survey forms to SSO's
August 14-17	SSO's edit forms and keypunch data
August 17	SSO's mail forms and data cards to Washington, D.C.
August 24	Form printout run for next survey fieldwork
August 25	Printout sent to SSO's
September 9	Enumerators receive printout
September 11-15	Survey fieldwork
September 15	Enumerators mail forms to SSO's
September 18-21	SSO's edit and keypunch updates
September 21	SSO's mail forms and data cards to Washington, D.C.
September 27	Printout run for next survey fieldwork
September 28	Printout sent to SSO's
October 7	Enumerator receive printout
October 13	Enumerators mail forms to SSO's
October 16-19	SSO's edit and keypunch updates
October 19	SSO's mail forms and data cards to Washington, D.C.
October 27	Final printout run

Although the data was not summarized monthly, it would have been possible to do so after the summarization program had been implemented. After implementation of the summary program, it would have been possible to have summarized the data within 14 days from completion of fieldwork.

For each survey period, enumerators observed about 3,800 fields and recorded the data on about 1,100 forms. Because of this volume, the computer generated survey form was a necessity. The numbers of segments, tracts, and fields observed on each update survey are shown in Table 7.

Table 7--Number of segments, tracts, and fields by test site.

State	Number of Segments	Number of tracts	Number of Fields
South Dakota	50	217	860
Kansas	48	274	854
Missouri	52	284	772
Idaho	44	311	1358
TOTAL	194	1086	3844

2.1.6 Summarization of Ground Observations

Since these segments were selected at random within a CRD, an expansion is possible to estimate totals for the CRD. The following estimator could be used.

$$\hat{Y}_j = F \sum_{i=1}^{n} \hat{y}_{ij}$$

Where $F = \frac{N}{n}$ (the inverse of the probability of selection) and N = total number of sampling units in the test site, and n = the number of sampling units in the sample, and Y_{ij} is the acreage of the jth crop in the ith sampling unit.

The standard error of Y_j is $[Se(y_j)]$

$$\text{where: } Se(y_j) = \sqrt{\frac{n \sum_{i=1}^{n} (Fy_{ij})^2 - \frac{(\sum_{i=1}^{n} Fy_{ij})^2}{n}}{n-1}}$$

$$\text{The coefficient of variation (C.V.)} = \frac{Se(\hat{y}_j) \times 100}{\hat{Y}_j}$$

The update observations were summarized in the same manner as the JES. Estimates of the acreages, standard errors, and coefficients of variation by crop and date are included in Tables 8, 9, 10, and 11. The Coefficients of Variation, which are measures of the relative precision of the estimates, ranged from about 10 percent to 100 percent, depending upon the particular crop or land use being estimated. For most major crops the C.V.'s were around 16 to 30 percent. On the other hand, crops which are not very important to that area and which were found in only one field in the selected JES segments had C.V.'s of around 100 percent.

2.1.7 Flightline Ground Observations

Flightline Selection: Each of the four study areas was divided into flightlines such that all flightlines in a single study area were of the same width. The width of the flightlines was limited to the swath width of the RB-57 and U-2 aircraft photo coverage and varied from 8 to 12 miles, depending on the area. Two flightlines in each study area were then selected at random, without replacement. The approximate locations of the selected flightlines are shown in Figures 1, 2, 3, and 4.

Table 8--Estimated acres, standard errors, and coefficients of variation by crop and date, IDAHO, 1972.

Crop \ Date	June Enumerative Survey			August 7-11			September 11-15			October 10-13		
	Estimated Acres	Standard Error	Coefficient of variation	Estimated Acres	Standard Error	Coefficient of variation	Estimated Acres	Standard Error	Coefficient of variation	Estiamted Acres	Standard Error	Coefficient of variation
CORN	63,983	15,495	24.2	63,929	15,362	24.0	32,607	9,110	27.9	12,123	5,307	43.8
OATS	2,430	1,685	69.3	1,617	939	58.0	-	-	-	-	-	-
BARLEY	136,629	29,281	21.4	73,616	18,540	25.2	3,842	2,956	75.0	-	-	-
WINTER WHEAT	59,270	24,190	40.8	39,592	19,868	50.2	510	504	98.9	8,873	4,037	45.5
MIXED GRAIN	27,293	7,109	26.1	45,461	22,032	48.5	348	241	69.3	440	308	70.0
SPRING WHEAT	20,211	6,221	30.8	19,244	5,762	30.0	2,600	1,369	52.7	1,274	922	72.4
POTATOES	49,288	17,490	35.5	48,477	17,488	36.1	48,338	17,479	36.2	7,327	4,055	55.3
FRUIT	324	321	98.9	324	321	98.9	324	321	98.9	324	321	98.9
FIELD BEANS	101,069	20,836	20.6	102,904	21,384	20.8	45,767	12,856	28.1	7,683	3,474	45.2
PEAS	11,118	4,960	44.6	-	-	-	-	-	-	-	-	-
ALFALFA	230,118	29,518	12.8	220,659	27,882	12.6	227,323	28,244	12.4	227,657	28,084	12.3
OTHER HAY	3,517	1,762	30.1	4,094	1,855	45.3	8,752	3,891	44.5	8,752	3,891	44.5
CLOVER	686	679	98.9	686	679	98.9	686	679	98.9	686	679	98.9
SUGAR BEETS	68,695	16,346	23.8	68,191	16,278	23.8	69,415	16,409	23.6	67,806	16,070	23.7
FARMSTEAD, ETC.	124,706	15,998	12.8	88,622	12,165	13.7	91,217	12,332	13.5	98,390	13,007	13.2
PASTURE	233,103	40,691	17.5	261,380	45,343	17.3	263,102	45,498	17.3	253,334	45,421	17.2
FALLOW	87,434	33,947	38.8	88,729	35,632	40.2	84,689	34,752	41.0	80,790	34,770	43.0
IDLE	20,393	8,087	39.7	114,637	18,621	16.2	362,953	42,650	11.8	456,749	43,517	9.5
TOTAL	1,242,544			1,242,142			1,242,473			1,242,408		

Table 9--Estimated acres, standard errors, and coefficient of variation by crop and date, KANSAS, 1972.

Crop \ Date	June Enumerative Survey			August 7-11			September 11-15			October 10-13		
	Estimated Acres	Standard Error	Coefficient of variation	Estimated Acres	Standard Error	Coefficient of variation	Estimated Acres	Standard Error	Coefficient of variation	Estimated Acres	Standard Error	Coefficient of variation
CORN	347,849	114,470	32.9	420,127	135,917	32.4	407,164	132,121	32.4	273,442	94,154	34.4
OATS	3,492	2,432	69.6	-	-	-	-	-	-	-	-	-
BARLEY	13,268	13,210	99.6	-	-	-	-	-	-	-	-	-
WINTER WHEAT	1,435,362	229,965	16.0	12,221	12,221	100.0	284,485	194,752	68.5	2,104,732	454,589	21.6
RYE	873	869	99.6	-	-	-	-	-	-	-	-	-
VEGETABLES	3,492	3,476	99.6	3,492	3,476	99.6	3,492	3,476	99.6	-	-	-
GRAIN SORGHUM	643,962	169,342	26.3	755,179	177,470	23.5	736,193	169,471	23.0	696,388	163,769	23.5
ALFALFA	136,018	55,375	40.7	115,330	44,472	38.6	114,632	43,994	38.4	111,751	43,192	38.7
OTHER HAY	12,308	5,546	45.1	19,553	11,064	56.6	18,768	11,064	59.0	20,950	8,396	40.1
SUGAR BEETS	11,261	11,211	99.6	11,261	11,211	99.6	11,261	11,211	99.6	11,261	11,211	99.6
FARMSTEAD, ETC.	36,087	8,771	24.3	40,407	8,974	22.2	40,835	9,303	22.8	43,070	9,362	21.7
PASTURE	2,833,548	677,704	23.9	2,855,906	676,409	23.7	2,852,414	676,603	23.7	2,821,531	677,660	24.0
FALLOW	1,643,081	274,249	16.7	2,097,958	420,635	20.1	1,824,046	248,501	13.6	321,400	89,726	27.9
IDLE	8,031	4,389	54.7	817,062	114,382	14.0	869,428	124,413	14.3	758,288	106;675	14.1
TOTAL	7,128,632			7,148,496			7,162,718			7,162,813		

Table 10--Estimated acres, standard errors, and coefficient of variation by crop and date, MISSOURI 1972.

Crop \ Date	June Enumerative Survey			August 7-11			September 11-15			October 10-13		
	Estimated Acres	Standard Error	Coefficient of variation	Estimated Acres	Standard Error	Coefficient of variation	Estimated Acres	Standard Error	Coefficient of variation	Estimated Acres	Standard Error	Coefficient of variation
COTTON	528,908	78,357	14.8	486,784	73,218	15.0	486,784	73,218	15.0	360,011	62,474	17.4
CORN	67,308	16,849	25.0	65,306	18,051	27.6	63,123	18,068	28.6	36,738	11,797	32.1
OATS	728	726	99.7	-	-	-	-	-	-	-	-	-
BARLEY	1,274	899	70.6	-	-	-	-	-	-	-	-	-
WINTER WHEAT	319,997	56,649	17.7	-	-	-	-	-	-	45,672	20,547	45.0
RYE	9,917	5,669	57.2	-	-	-	-	-	-	4,549	4,537	99.7
FRUIT	6,369	6,351	99.7	6,369	6,351	99.7	6,369	6,369	99.7	6,369	6,351	99.7
SOYBEANS	759,198	144,117	19.0	1,052,448	165,294	15.7	1,046,807	165,754	15.8	1,020,987	165,473	16.2
GRAIN SORGHUM	16,559	8,308	50.2	17,286	8,432	48.8	17,286	8,432	48.8	11,646	7,217	62.0
ALFALFA	4,003	2,852	71.2	4,003	2,852	71.2	4,185	2,905	69.4	4,185	2,905	69.4
OTHER HAY	54,043	29,708	55.0	56,263	27,400	48.7	56,627	25,684	45.4	50,440	25,419	50.4
CLOVER	13,956	9.751	70.0	11,773	6,245	53.0	15,412	7,107	46.1	15,412	7,107	46.1
FARMSTEAD, ETC.	186,402	50,733	27.2	215,423	51,195	23.8	215,423	51,181	23.8	215,605	51,227	23.8
PASTURE	249,197	60,376	24.2	248,741	59,177	23.8	245,648	58,312	23.7	250,015	59,119	23.6
FALLOW	53,533	18,786	35.1	68,509	23,694	34.6	85,613	29,849	34.9	66,871	19,929	29.8
IDLE	132,523	25,566	19.3	155,668	27,506	17.7	150,937	26,852	17.8	306,060	57,034	18.6
TOTAL	2,403,915			2,388,573			2,394,214			2,394,560		

Table 11--Estimated acres, standard errors, and coefficient of variation by crop and date, SOUTH DAKOTA, 1972.

Date / Crop	June Enumerative Survey			August 7-11			September 11-15			October 10-13		
	Estimated Acres	Standard Error	Coefficient of variation	Estimated Acres	Standard Error	Coefficient of variation	Estimated Acres	Standard Error	Coefficient of variation	Estimated Acres	Standard Error	Coefficient of variation
CORN	957,449	98,744	10.3	947,272	101,201	10.7	942,467	100,559	10.7	858,267	92,184	10.7
OATS	539,315	86,785	12.4	111,660	37,787	33.8	3,203	3,195	99.7	41,176	17,773	43.2
BARLEY	81,434	28,820	35.4	15,696	12,757	81.3	-	-	-	3,640	3,630	99.7
WINTER WHEAT	5,460	3,811	69.8	3,858	3,497	90.6	-	-	-	-	-	-
RYE	28,392	12,817	45.1	9,755	7,215	74.0	5,824	5,809	99.7	23,150	9,444	40.8
SPRING WHEAT	30,474	16,216	53.2	30,765	15,664	50.9	7,426	6,099	82.1	-	-	-
DURUM WHEAT	3,931	3,921	99.7	3,931	3,921	99.7	3,931	3,921	99.7	-	-	-
FRUIT	-	-	-	1,893	1,888	99.7	1,893	1,888	99.7	1,893	1,888	99.7
SOYBEANS	33,881	17,118	50.5	33,444	15,167	45.3	33,444	15,167	45.3	26,310	13.065	50.0
GRAIN SORGHUM	28,848	13,158	44.1	33,051	13,772	41.7	11,502	5,923	51.5	11,066	5,923	53.5
ALFALFA	318,515	57,431	18.0	301,051	51,701	17.2	302,279	51,709	17.1	311,699	57,239	18.4
OTHER HAY	146,371	45,984	31.4	255,864	59,386	23.2	281,720	60,601	21.5	255,862	60,802	23.8
FLAX	53,362	19,391	36.3	60,788	21,716	35.7	15,870	11,101	70.0	7,280	5,158	70.9
FARMSTEAD, ETC.	132,147	16,318	12.3	122,405	13,698	11.2	128,885	18,553	14.4	128,594	18,634	14.9
PASTURE	879,189	122,810	14.0	854,961	120,382	14.1	845,352	119,635	14.2	845,352	119,635	14.2
FALLOW	291,854	50,598	17.3	192,963	37,458	19.4	190,924	36,820	19.3	185,683	36,365	20.0
IDLE	69,189	20,321	29.4	621,783	72,420	11.6	823,949	83,360	10.1	899,136	88,844	9.9
TOTAL	3,600,811			3,601,140			3,598,669			3,599,108		

Each flightline contained a number of sampling units (JES segments). Even though the segments already existed before the flightlines were constructed, their appearance in the sample was still random. The number of segments within each flightline varied by flightline and state - Table 12. Once it was determined which segments fell within the flightline, a count of all other possible segments in the flightline was made, thus the probability of a segment being selected is $\frac{m_i}{M_i}$ within the ith flightline.

This is a multistage sample design where the selection of flightlines is the first stage and the second stage of selection is the segments. Whereas in this particular case, maps were used as the frame to select the sample, it might have been possible to select a similar sample using ERTS imagery or aerial photography.

Table 12--Number of segments within flightline by flightline and by state.

State	Flightline	Number of segments	
		JES	Added for classifier training
Missouri	1	2	5
	2	8	6
Kansas	1	5	2
	2	9	3
South Dakota	1	4	6
	2	5	4
Idaho	1	6	6
	2	9	10

Flightline estimates for the study areas are explained below.

The estimates of totals for a two-stage sample design are as follows:

$$Y_k = \frac{N}{n} \sum_{i=1}^{n} \frac{M_i}{m_i} \sum_{j}^{m_i} y_{ijk}$$

where: $\hat{Y}_k$ is the estimate of the total acreage of the kth crop or characteristic within a study area,

N is the total number of flightlines,

n is the number of flightlines in the sample,

M_i is the total number of segments within the selected flightlines,

m_i is the number of selected segments within the selected flightlines.

The variance of $\hat{Y}_k$ is:

$$\text{var}(\hat{Y}_k) = N^2\left(\frac{N-n}{N}\right)\frac{S_{k1}^2}{n} + \frac{N}{n}\sum_{i=1}^{n}\frac{M_i(M_i-m_i)}{M_i m_i}\, s_{k2i}^2$$

where: $$S_{k1}^2 = \sum_{i=1}^{n}\frac{(Y_i-\bar{Y})^2}{n-1}$$

$$S_{k2i}^2 = \sum_{j=1}^{m_i}\frac{(Y_{ij}-Y_i)^2}{m_i-1}$$

and $$\text{C.V.} = \sqrt{\frac{\text{Var}(\hat{Y})(100)}{\hat{Y}}}$$

2.1.8 Results of Flightline Ground Observations

As would be expected from a sample of size 2 from the heterogeneous study areas, the flightline estimates in all four States were not very reliable. Coefficients of variation, the measures of precision of the estimates ranged from 20 to 100 percent. For most crop, the between flightline component of variance was the largest contributor to the total estimated variance. Therefore, if the computed variance components are any indication, the easiest way to reduce the variance would be to add more flightlines.

In several cases, the CRD estimates from the flightline ground observations compare favorably with those from the JES, but the size of the standard error would indicate that this is due to chance. Flightline estimates of total acres by crop, the estimated between and within flightline components of variance, and standard errors, and coefficients of variations of the estimated totals are shown by study areas in Tables 13-18. Generally, these computations were made only for the crops which were classified from the LANDSAT and aircraft imagery. However, some of the crops shown were not included in the LANDSAT or aircraft classification.

For Missouri, all three of the update ground surveys were tabulated for the selected flightlines. This was to correspond with the occurrence of useable LANDSAT imagery from each of the months, August, September, and October. The only significant changes in estimated totals occurred between the September 11-15 and October 10-13 ground surveys. These changes occurred as cotton and soybeans were harvested, causing the use to change from those crops to idle (stubble) land or fallow (plowed). The only flightline totals shown for the other three study areas are for the update survey periods August 7-10 (Idaho and South Dakota) and September 11-15 (Kansas).

Except for Kansas, the between flightline variance component was based on all flightlines in the study area in order to get a reasonable estimate of this variance.

The main conclusion to be drawn from the flightline ground observation analysis is that in order to get reliable estimates from this multi-stage sampling approach, more flightlines are needed but is is not necessary that they cover such a wide swath. Also, in constructing flightlines, the total size (length times width) of the flightlines should be kept as equal as possible. For example, flightline 2 in Missouri is much smaller than flightline 8. This variation in size can contribute significantly to the overall precision of the estimates.

2.2 Data Acquisition - LANDSAT Imagery

2.2.1 Objectives

Satellite imagery required for this project included both the computer compatible MSS digital data tapes and various types of photographic images.

The photographic images were required to:

1. determine if a particular LANDSAT frame was usable, and

2. to assist in locating individual test sites (segments) in the frame.

Table 13--Estimated totals, between and within flightline components of variance, standard errors, and coefficients of variation of the estimated totals by crops, Missouri Study Area, August 7-10, 1972.

Crop	Estimated Acres for CRD	Between Flightline Variance	Within Flightline Variance	Standard Error	Coefficient of Variation
Cotton	309,096	79,585,100,901	1,760,246,861	285,211	92.3
Corn	82,602	2,676,330,320	253,508,210	54,127	65.5
Fruit	33,390	929,039,587	185,838,406	33,390	100.0
Soybeans	1,533,204	1,174,865,916,260	52,150,144,244	1,107,707	72.2
Grain Sorghum	20,790	51,070,314	22,424,259	8,573	41.2
Other Hay	199,800	33,265,369,332	1,339,221,600	186,023	93.1
Clover	30,240	762,017,518	152,409,600	30,240	100.0
Farmstead, etc.	517,716	80,235,265,331	6,671,987,712	294,800	56.9
Pasture	276,606	29,869,978,560	2,352,207,811	179,581	64.9 *
Fallow	46,746	1,820,917,590	83,905,414	43,644	93.4
Idle	249,030	33,103,771,942	1,234,579,858	185,306	74.4

* The between flightline variance is based on all flightlines.

Table 14--Estimated totals, between and within flightline components of variance, standard errors, and coefficients of variation of the estimated totals by crops, Missouri Study Area, September 11-15, 1972.

Crop	Estimated Acres for CRD	Between Flightline Variance	Within Flightline Variance	Standard Error	Coefficient of Variation
Cotton	309,096	79,585,100,901	1,760,246,861	285,211	92.3
Corn	82,602	2,676,303,320	253,508,210	54,127	65.5
Fruit	33,390	929,039,587	185,838,406	33,390	100.0
Soybean	1,533,204	1,174,865,916,260	42,150,144,224	1,107.707	72.2
Grain Sorghum	20,790	51,070,314	22,424,259	8,573	41.2
Other Hay	178,200	26,452,114,920	1,471,219,200	167,103	93.8
Clover	30,240	762,017,518	152,409,600	30,240	100.0
Farmstead, etc.	517,716	80,235,265,331	6,671,987,712	294,800	56.9
Pasture	276,606	29,896,978,560	2,352,207,811	179,581	64.9 *
Fallow	68,346	526,723,656	87,015,814	24,774	36.2
Idle	249,030	33,103,771,942	1,234,579,858	185,306	74.4

* The between flightline variance is based on all flightlines.

Table 15--Estimated totals, between and within flightline components of variance, standard errors, and coefficients of variation of the estimated totals by crops, Missouri Study Area, October 10-13, 1972.

Crop	Estimated Acres for CRD	Between Flightline Variance	Within Flightline Variance	Standard Error	Coefficient of Variation
Cotton	212,742	37,700,879,084	1,196,149,683	197,223	92.7
Corn	82,602	2,676,303,320	253,508,210	54,127	65.5
Fruit	33,390	929,039,587	185,838,406	33,390	100.0
Soybeans	1,467,378	1,048,235,439,110	53,444,905,920	1,049,610	71.5
Grain Sorghum	15,066	3,694,552	19,308,022	4,796	31.8
Other Hay	178,200	26,452,114,920	1,471,219,200	167,103	93.8
Clover	30,240	762,017,518	152,409,600	30,240	100.0
Farmstead, etc.	517,716	80,235,265,331	6,671,987,712	294,800	56.9
Pasture	276,606	29,896,978,560	2,352,207,811	179,581	64.9 *
Fallow	68,346	526,723,656	87,015,814	24,774	36.2
Idle	392,130	97,687,578,082	2,542,832,141	316,592	80.7
Winter Wheat	24,804	512,493,601	102,622,253	24,804	100.0

* The between flightline variance is based on all flightlines.

Table 16--Estimated totals, between and within flightline components of variance, standard errors, and coefficients of variation of the estimated totals by crops, Kansas Study Area, September 11-15, 1972.

Crop	Estimated Acres for CRD	Between Flightline Variance	Within Flightline Variance	Standard Error	Coefficient of Variation
Alfalfa	152,390	18,578,718,288	2,437,398,690	144,969	95.1
Pasture	3,208,195	88,735,891,538	223,622,366,804	1,054,031	34.9
Corn	1,146,690	15,166,828,880	3,634,212,685	137,117	99.6
Grain Sorghum	1,146,070	129,998,137,680	182,470,351,826	558,989	48.8
Winter Wheat	29,045	674,889,620	168,722,676	29,045	100.0
Fallow	1,086,780	215,995,641,680	30,751,534,531	495,938	45.7
Sugar Beets	46,255	1,711,620,020	85,585,359	42,393	91.7

Table 17--Estimated totals, between and within flightline components of variance, standard errors, and coefficients of variation of the estimated totals by crops, Idaho Study Area, August 7-10, 1972.

Crop	Estimated four co. acres	Between Flightline Variance	Within Flightline Variance	Standard Error	Coefficient of Variation
Corn	106,909	359,570,842	489,692,090	29,142	27.3
Barley	77,572	4,533,887,611	276,569,320	69,501	89.6
Winter Wheat	39,754	1,165,824,646	109,949,320	35,718	89.9
Mixed Grain	31,713	407,075,135	65,268,794	21,733	68.5
Spring Wheat	30,090	488,902,450	83,624,281	23,928	79.5
Potatoes	109,054	1,499,274,753	465,840,844	44,326	40.6 *
Field Beans	57,071	2,023,502,496	334,116,679	48,555	85.1
Alfalfa	203,120	5,369,614,922	4,002,751,528	96,811	47.7
Sugar Beets	91,019	403,669,920	511,370,942	30,249	33.2
Farmstead, etc.	139,227	1,789,834,480	284,669,027	45.547	32.7 *
Pasture	827,398	368,272,987,764	15,481,398,570	619,479	74.9
Fallow	192,285	13,256,360,708	618,122,640	117,790	61.3 *
Idle	133,502	3,211,847,210	652,799,570	62.166	46.6

* The between flightline variance is based on all flightlines.

Table 18--Estimated totals, between and within flightline components of variance, standard errors, and coefficients of variation of the estimated totals by crops, South Dakota Study Area, August 7-10, 1972.

Crop	Estimated Acres for CRD	Between Flightline Variance	Within Flightline Variance	Standard Error	Coefficient of Variation
Corn	908,350	51,266,009,138	3,367,476,285	233,738	25.7 *
Flax	35,335	1,056,185,780	264,116,571	36,335	100.0
Fallow	111,055	1,568,220,500	616,904,097	46,745	42.1
Pasture	759,550	48,067,051,520	5,193,126,612	230,782	30.4
Sudex	61,510	1,971,303,680	310,693,399	47,770	77.7
Alfalfa	272,720	4,805,000,000	919,799,185	75,662	27.7
Idle	788,525	25,315,726,472	2,401,924,450	169,463	21.5 *

* The between flightline variance is based on all flightlines.

The computer compatible data tapes were used:

1. to generate the grey-scale computer printouts needed in locating the individual segments (and fields) within the LANDSAT frame, and

2. as data input into the computer crop classification routines.

2.2.2 Approach

Photographic imagery obtained from NASA included 70mm positive and negative transparencies and system corrected 9.5" positive B&W transparencies for all LANDSAT frames which include (1) any part of one of the four sites, and (2) any part which had less than 50 percent cloud cover. Precision 9.5 color composite photographs were also ordered, but not analyzed. Enlargements (1/250,000) of the composite photographs for selected frames were obtained from the ASCS photo lab in Salt Lake City.

System corrected MSS digital tapes were also obtained for all frames having less than 50 percent cloud cover.

2.2.3 Evaluation

We received LANDSAT 70mm transparencies and the system corrected ditigal data tapes as a standing order. The first digital data tapes were received November 1, 1972. Tapes received between November 1 and November 16 included scenes taken as early as August 15. After November 16, tapes generally were received about four weeks after the scene was taken. The initial delay in receiving data tapes was serious only in that various computer programs could not be tested operational,until at least one set of tapes has been received.

In retrospect, a more desirable procedure would have been to place a standing order for either 9.5 inch or 70mm transparencies of all LANDSAT frames which covered any part of a target area. Then, a selection of data tapes to be ordered could have been made from these transparencies. This would have effected a substantial reduction in the number of data tapes received and stored, but essentially unused because of incomplete clould-free coverage over a given site during a particular cycle.

The 1/250,000 scale color enlargements to selected LANDSAT frames were used to visually locate specific training sites in the LANDSAT frame.

2.3 Data Acquisition - Aerial Photography

2.3.1 Objectives

High altitude photogrpahy was acquired from NASA and the South Dakota Remote Sensing Institute (SDRSI) to meet the following objectives.

1. Develop methods of crop species identification from aerial photography by computer classification techniques, and compare the results with the ground data and with the results obtained using LANDSAT imagery.

2. Estimate crop acreages by expansion of classification results to the flightline level and crop reporting district level.

3. To assist in the location of segments on the LANDSAT frame or printouts.

2.3.2 Approach

Flightline Selection

Adjacent, non-overlapping flightlines were drawn on aeronautical charts to provide complete coverage of the land area within each of the four LANDSAT test site areas for this project. The flightlines constructed were 8-10 miles wide and sufficiently long to traverse the full length of the test site. Within each LANDSAT test site, two flightlines were randomly selected for aerial photography overflights. NASA provided high altitude, color positive, infrared aerial photography (9 inch format) for both selected flightlines for each LANDSAT test site. Attempts were made to coordinate overflight dates for the aerial photography with the LANDSAT imagery. NASA provided aerial photography on two separate dates for the Kansas, South Dakota, and Missouri test sites, and three dates for the Idaho test site. The South Dakota Remote Sensing Institute also provided photographic coverage (70mm color positive, infrared) for the selected flightlines in the Kansas, South Dakota, and Missouri test sites for one overflight date. Photographic check-in procedures were as follows:

1. Locate, delineate, and identify all JES segments and training segments on the aerial photography from County Highway maps.

2. Record the frame number or numbers each segment is located on.

Tables 19-22 summarizes the photographic coverage for each segment.

2.3.3 Scanning Procedures

The JES segments were scanned on a microdensitometer with an effective aperture size of 240 microns square. Reduction of the volume of data was one of the primary considerations which lead to the choice of such an aperture. Using this aperture, one data point covers a land area approximately 95 feet square on the NASA photography. Each segment was scanned with a clear, red, green, and blue color filter and in two scanning modes. Thus, multivariate observations are obtained for each data point. Prior to actual scanning of the photography, it was necessary to record coordinates of corner points of fields and field boundaries to identify training data for the classifier. A sketch of each segment was made from large aerial maps (scale: 8" - 1 mile) showing each field (small land area devoted to one crop species or agricultural practice). Field boundary coordinate information was recorded on these sketches. Figure 7 is a simplified sketch of a JES segment with field boundary coordinates recorded. Appendix D contains detailed instructions for the scanning procedures.

Data Conversion and Preparation for Classification

Output data from the microdensitometer is stored on magnetic tapes. Each file on the magnetic tape corresponds to one segment scanned with one color filter and recorded in one scanning mode. In order to obtain multivariate observations for each data point, a software program, PDSCMS (Appendix E), was developed to merge the data from several microdensitometer output files, each file corresponding to a scanning mode filter combination into one file which was compatible with the Statistical Analysis System (SAS).

In order to perform crop classification using discriminant analysis, it is necessary to "train the classifier." To facilitate automated assignment of training data for each crop class, a software program was developed. The program generated SAS program statements to assign tract and field identifiers to data points on the basis of the coordinates of each pixel. The program assigns these labels only to data points contained within user defined rectangles whose sides are parallel to the scanning axes. The tract and field identifiers were then used to merge the microdensitometer data with the ground information collected during the 1972 growing season.

The ground information that was collected monthly included the crop species and crop condition for each tract and field within the segment. The crop condition and crop species was used to form the group for classification with discriminant analysis. Thus, an observation vector in the merged data set contains the following information.

1. The value of relative light intensity for each of two scanning modes and four filter combinations,

Figure 7--Sketch of Segment Showing Field Boundaries and Crop Classes.

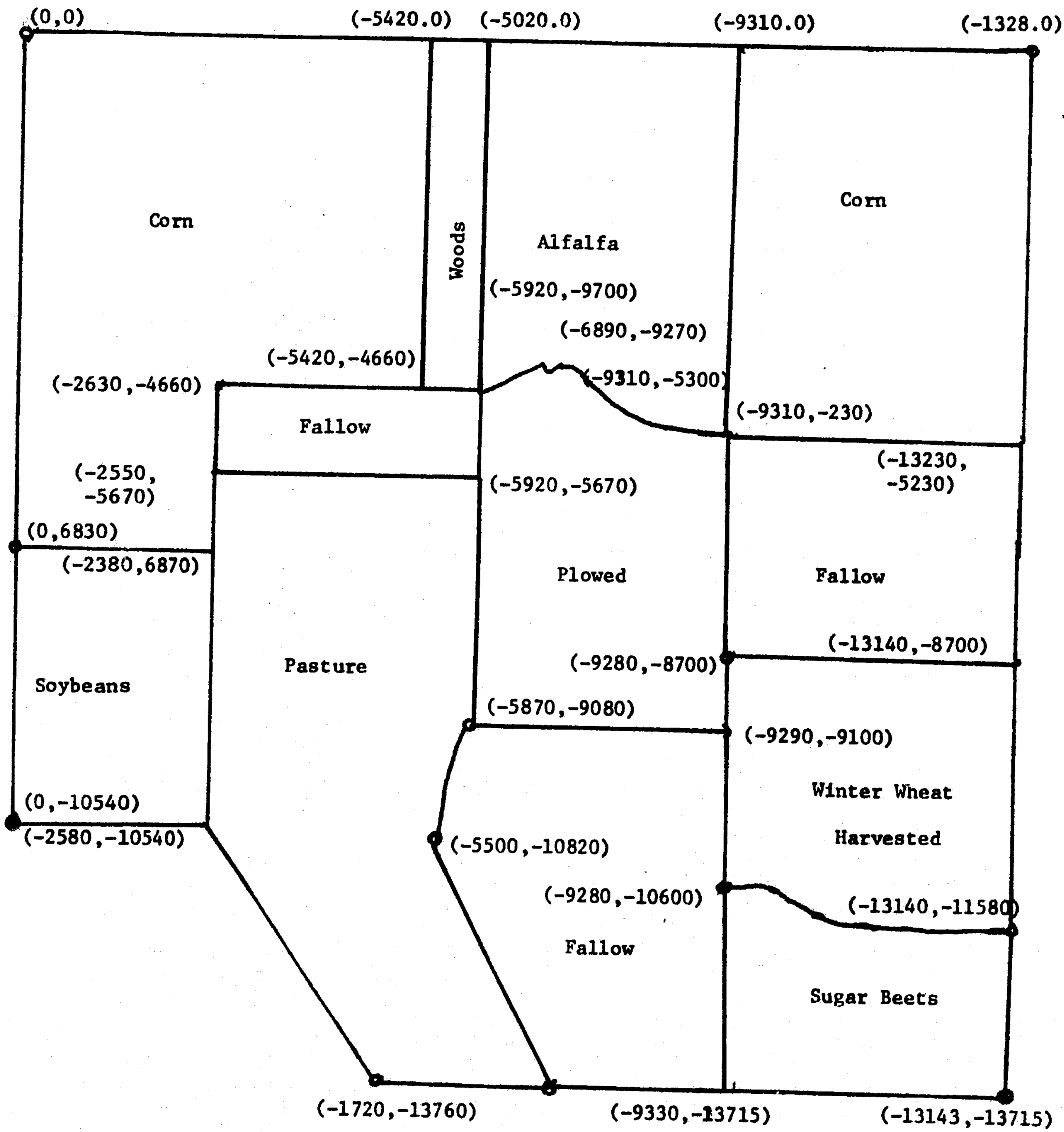


TABLE 19

MISSOURI AERIAL PHOTOGRAPHY

Mission; Date:	208; 8/28/72		211; 9/19/72		S.D.R.S.I.; 9/19-20/72
Camera, Roll	RC-8; 33	ZEISS: 34	RC-8; 42	ZEISS; 44	4 filters
Segments	Frame No.	Frame No.	Frame No.	Frame No.	Frame No.
F.L. 2					
4418	29	--	99	--	38 & 39
4420	31	55	98	25	42
F.L. 8					
4411	05	7	127		3
3412	07	--	124		28 & 29
1413	07	12	124	78	19 - 25
4414	04	6	128	84	6 & 7
1435	13	22	120	69	9
3436	10	17	122	73	2
4458	11	--	121		
4460	08	16	123	76	32 & 33
Extra					
3416	28	--	--		
4417	30	53	98		
4419	29	--	99		
3432	15	--	118		
4434	12	--	120		
4437	10	--	123		
Training					
2A1	31	55	97	23	44 & 45
2A2	30	55	98	24	47 - 53
2B	29	--	100		29 & 30
2C	29	--	99		33
2D	28	49	100		25 & 26
8A	05	6	128	85	11 & 12
8B	07	12	125	79	14 & 15
8C	11	18	121	72	37 & 38
8D	11	19	121	72	5 & 6
8E	12	--	120		41 & 43 20 & 21
8F	15	26	118	64	15 & 16

TABLE 20

KANSAS AERIAL PHOTOGRAPHY

Mission; Date: Camera; Roll Segments	208; 8/18/72 RC-8; 1 Frame No.	ZEISS; 3 Frame No.	211; 9/17/72 RC-8; 33 Frame No.	ZEISS; 35 Frame No.	S.D.R.S.I.; 8/12-14/72 4° Filters of 4 rolls Each Frame No.
F.L. 3					
4087	41	--	19	--	B26 - 31
1089	43	85	17	29 & 271	-
4101	48	95	13	20 & 280	A27 - 30
3106	37	72	23	259	B 1 - 5
4107 Noc	34	66	27	--	C40
1113	53	107	07	08 & 291	A53 - 56
4114	50	100	10	16 & 285	A34 - 38
1115	40	79	21	265	B12 - 15
3116	41	81	19	268	B22
F.L. 10					
4120	14	26	--	--	D12 - 16
3122	24	48	--	--	C23 - 26
4124	18	35	--	--	C 1 - 8
1125 Noc	Noc	--	--	--	-
4130	22	43	--	--	C17 - 19
Extra					
4088	44	--	17	--	-
Training					
3-A	50	101	10	14 & 286	A42
3-B	36	70	25	45	C36
3-C	37	72	24	260	-
3-D	40	81	20	266	B 7 & 8
3-E	40	81	20	267	B18
3-F	42	83	19	32 & 269	B39 - 51
3-G	42	83	19	32 & 269	B57 - 64
3-H	43	--	17	--	B36
3-I	43	85	17	30 & 272	A 3
3-J	43	87	17	28 & 273	A 8
3-L	46	--	14	--	A15
3-M	47	--	13	--	A18
3-P	54	109	06	07 & 293	A49 & 50
10-A	24	--	--	--	C32
10-E	9	17	--	--	D 2 - 5

Note: RC-8 and ZEISS coverage of segments 1113, 4114, and 3A are also available from Mission 217 dated 10/24/72.

TABLE 21

SOUTH DAKOTA AERIAL PHOTOGRAPHY

Mission; Date:	211; 9/22/72		211; 9/14/72		S.D.R.S.I.; 8/27/72
Camera, Roll	RC-8; 54	ZEISS; 56	RC-8; 17	ZEISS; 19	4 filters and 4 foils
Segments	Frame No.	Frame No.	Frame No.	Frame No.	Frame No.
F.L. 3				None	
3196	2934	70			46
4197	2932	66			54
1199	2934	71			50
4210	2930	62			5 & 6
F.L. 5				None	
1213	2908	18	188		26
1223	2912	27	184		14
3236	2906	14	191		35
4237	2906	--	191		32
4240	2915	--	181		8
Extra				None	
1195	2934	--			
4198	2933	69			
4208	2928	--			
4211	2928	--			
3212	2909	--	187		
4214	2908	20	188		
3222	2913	--	--		22 & 23
4224	2912	27	184		
1235	2906	--	190		
1239	2918	--	179		
4241	2918	39	179		
Training				None	
3-A3	2930	62			1
3-B-9	2933	68			53
3-C-3	2935	--			44
3-C-5	2935	72			48
3-[illegible]	2935	--			41
3-D-8	2935	74			38
5-C-2	2913	27	184		12
5-C-3	2913	29	184		20
5-C-4	2913	29	--		16
5-E-2	2908	17	189		29

TABLE 22

IDAHO AERIAL PHOTOGRAPHY

Mission; Date	72-138; 8/11/72	9/7/72	10/25/72
Camera	RC-8	RC-8	RC-8
Segments	Frame No.	Frame No.	Frame No.
F.L. 5			
8101	4702, 4812-13	3885-86, 3900-01	5565-66, 5652-53, 5820
8103	-	3881	5647
8111	4699-4700, 4814-15	3884-85, 3902-03	5650-51
3423	4816	3904-05	--
1554	4699,4814-15	3883-84	5650-51
1559	4699,4815-16	3883-84, 3903-04	5650, 5667
F.L. 6			
8094	4812	3900-01	5664, 5822
8098	4811-12	3899-3900	5817, 5663-64
8109	4813	3886, 3901-02	5665
9110	4700,4814	3884-85, 3901-02	5661, 5665-66
8113	4814-15	3902-03	5666-67
8265	4816	3904-05	5668
2332	4811-12	3899-3900	5663, 5817
8339	4816	3904-05	5668
3422	4812-13	3900-01	5664, 5821-22
Extra			
8096	4703		
8099	4701		
8102	4701		
8112	4814	3902-03	5666-67
8115	4701		
1549	4702		
1550	4702		
Training			
5-A-2	4702-03, 4810-11	3887-88, 3899	5654, 5817-18-19
5-B-2	4702,4812-13	3886-87, 3900-01	5653, 5820
5-C-2	4814-15	3884-85, 3902-03	5665-66
5-D-2	4699-4700, 4814-15	3884-85	5650-51
5-K-5	4815	3903-04	5667
5-K-6	4815	3903-04	5667
6-C-2	4812-13	3900-01	5664-65, 5821-22
6-D-1	4812-13	3900-01-02	5664-65, 5812-22
6-F-3	4813-14	3901-02	5665-66, 5821
6-F-4	4813-14	3901-02	5665-66, 5821
6-H-1	4814	3901-02	5665-66
6-H-2	4814	3901-02	5665-66
6-I-1	4814	3902-03	5666-67
6-I-2	4814	3902-03	5666-67
6-J-4	4814 -15	3902-03-04	5666-67
6-L-4	-	3902	5665-66, 5822-23, 9095

2. The x,y - coordinates,

3. The tract, field number,

4. Crop and crop condition on four month visits.

There are eight spectral variables, two spatial variables, and four label variables making up each pixel.

III. Software and Data Processing

3.1 Segment and Field Location

3.1.1 Objectives

A primary objective of this phase of the project was to develop procedures which would enable the user to locate small areas in LANDSAT images that are identified on maps. These areas must be identified with great accuracy if they are to be used either as training sites or discriminant analysis or as test sites on the estimation procedures.

3.1.2 Approach

The method used to find segments and field boundaries was mostly a manual operation. The procedure is outlined below.

1. The exact location of the individual JES segments was drawn on county highway maps.
2. The approximate locations of the JES segments on 1/250,000 scale color enlargements of the LANDSAT frame were determined by a visual comparison of the enlargement with the county highway maps.
3. Grey scale maps of large areas around the location of each segment were generated from computer compatible MSS tapes. Generally, these maps were from response band 5.
4. Visual correlation of features distinguishable on the county highway maps, on the color enlargements of the LANDSAT imagery and on the grey scale computer printouts was used to find the location of individual segments in the LANDSAT frame.

Field boundaries had been drawn on 1"/660' scale aerial photographs of the JES segments. These photographs were then used as a basis for sketching the field boundaries on the computer grey scale printouts. Next, an area definition card was punched for every scan line that crossed each field. A more detailed description of this procedures is included in Appendix C.

Two different computer programs were used to produce grey level maps. The first was called NMAP and is from the Penn State Classification System. This system had several good points.

1. It could map any combination of channels to a maximum of 16 channels.

2. It can produce grey-level maps with variable proportion of points in an interval.

3. It can use either LANDSAT or LARSYS III format tapes as input.

Some of NMAPS disadvantages are:

1. It requires a format conversion run,

2. It must do a map to obtain initial grey level response histogram.

To speed up the mapping process, a second mapping program RAD MAP was developed. It has the following advantages over NMAP.

1. It maps at a faster rate.

2. It can sample to determine the response histogram and set the grey levels accordingly.

The major disadvantages are:

1. It will only map one band at a time.

2. It is limited to LANDSAT computer compatible tapes.

3.1.3 Evaluation

The segment location procedure described here was reasonably effective in southwestern Kansas and in the Snake River Valley of Idaho. These areas were characterized by a regular 'checkerboard' road pattern, moderately large regular fields, and by a number of crops which had distinctly different reflectance patterns. We had more difficulty in east central South Dakota and in southeastern Missouri. The principal problem in South Dakota was that, at the time the LANDSAT imagery was taken crops seemed to look much the same. Also, there were not many of the distinctive field patterns as were found in Kansas. Missouri was characterized by irregular road and field patterns and by heavy woodlands which helped to hide the roads.

A more fully automated procedure is needed for any further work in this area. Among the possibilities for inclusion in such a procedure could be the following.

1. A program which could compute the approximate location of test sites in a given LANDSAT frame.

2. The use of affine transformation to locate points in small areas of the LANDSAT frame (CITARS, F.G. Hall, M.E. Bauer, W.A. MALILA).

3. A grid digitizer to convert map boundaries to a series of data points which could be converted to LANDSAT frame coordinates.

3.2 Software Implementation for Crop Classification

3.2.1. Objectives

The main objective was to find and install in the USDA Washington Computer Center a series of computer programs to perform discriminant analysis (pattern recognition). In addition, the following related objectives should be satisfied.

A) The software should be relatively easy to install and maintain.

B) The system should use a uniform control card setup for both the system and in-house developed programs.

C) The program package should be highly modular to permit experimentation.

D) The program should provide support software for data handling.

E) Programs should be easy to use and not require a lot of cumbersome vendor JCL statements.

F) The software system must be reasonably efficient. This may be in terms of fast computational algorithms and/or data reduction schemes to reduce volume.

3.2.2 Approach

There were three systems available to us that could perform the require discriminant analysis.

The first package considered was SAS [1/], (Statistical Analysis System). This system is written to run only an IBM 360/370 computer, and is distributed in both load module and source form. Installation is as simple as creating a program library or adding numbers to an existing program library. Maintenance is minimal because the authors provide all necessary program support and send updated library tapes.

The system allows the user to create his own procedures by modifying existing procedures or writing them from scratch. The SAS supervision provides software support such that all usual control card and data management features are available to the user. A user procedure is treated exactly like a normal SAS procedure.

In general, SAS is easy to use, and the SAS language permits almost unlimited manipulation of data. However, the conversion of LANDSAT data tapes into SAS observations requires considerable programming because the SAS language has no simple provision to break up a line of data into a series of SAS observations.

The original procedure DISCRIM, prints a line for every data point classified. Clearly, this is too much output for an LANDSAT file. In addition, the procedure reads the entire data set twice, once to find the calibration data, and once to classify. The procedure does not have the calibration data, nor create a SAS file of classification results.

Procedure DISCRIM was modified to create an in-house procedure that did not print the results for each point calibrated, but rather created a SAS compatible file that could be read in using the input processor.

Drs. Barr and Goodnight extended the features of the discriminant procedure in the following ways:

1. Limit the printing of point by point classification results to desired levels and always print a summary.

2. Accepted calibration and unknown data from separate files.

3. Save and reuse the calibration results.

4. Output the classified data as a SAS file for later analysis.

1/ Developed at the Pennsylvania State University, Department of Forestry, by A. J. Barr and J. H. Goodnight.

The second software package was the Penn State Classification System.[1/] This system was written in FORTRAN, and should have been easy to install. Some special input/output software has to be provided by the Penn State Computer Center. This special software was obtained from Penn State. One routine worked and one did not, but a substitute was found. The Penn State System does work now at the WCC. The point is that the Penn State System may not be completely transportable to other computing centers.

The core programs use a common set of control cards which facilitates learning to use the programs. There are some related programs that were developed by other users that do not strictly adhere to the control card setup used by the main line programs. The maximum likelihood classification software is an example.

In spite of the fact that the program is broken down into subroutines, it cannot be considered modular. There are many different subroutines called GETLIN that are used to retrieve lines of data from the file. Other critical subroutines share the same problem.

In addition, these subroutines do not provide for complete file control. Therefore, any user defined program must partially process the input file in conjunction with some version of GETLIN.

This non-modularity makes it difficult to modify or change the program.

The package does not utilize a system monitor program to manipulate data files. One must use the standard vendor JCL to create and pass files between programs and runs.

This system does use a data reduction scheme to speed up processing. Normally, an investigator is interested in only a portion of a LANDSAT image. The programs permit the user to subset the image and retain only the areas of interest. A table of contents record, preceedings, the file, permits the user access to any particular area as though he had the entire image. Unnecessary data is not processed, thus it is more efficient.

The Penn State Classification System is really a collection of main level programs that can process a common file. The major programs are SUBERTS, SUBAIR, TPINFO, MERGE, NMAP, UMAP, STATS, ACLASS, ACLUS, DCLASS, and DCLUS.

1/ Developed at the Pennsylvania State University, Department of Forestry, by Dr. F. Y. Borden and Associates.

SUBERTS and SUBAIR are used to reformat and subdivide LANDSAT and aircraft tapes into the Penn State format.

MERGE is used to combine data from different passes into temporal overlays.

TPINFO prints the heading and table of contents records from a standard file.

NMAP assigns mapping symbols to all points of specified grey levels. It is used to prepare line printer maps.

UMAP assigns mapping symbols based on contrast differences. It is also used to outline boundaries.

STATS computes calibration statistics to be used by the classification programs.

ACLASS performs a discriminant analysis of spectral signatures that have been normalized by reducing all data to a unit sphere. It is used to compensate for sun angle, and was developed for airborn scanners.

ACLUS is an unsupervised cluster analysis program which uses the angular classification algorithm.

DCLASS performs a Euclidian distance discriminant analysis of multispectral data.

DCLUS is an unsupervised cluster analysis which uses the Euclidian distance algorithm.

In addition to the above core programs, a maximum likelihood quadratic classification package was supplied as a related program. This program is not control card compatible with the core programs, but it uses the standard file.

The third software package considered was LARSYS III.[1/] The initial consideration was to install it in-house. The support group at LARS and our scientific monitors convinced us that this was beyond our means.

[1/] Developed at Purdue University, Laboratory of Applications of Remote Sensing.

LARSYS is written for a different operating system than what we have at the USDA Washington Computer Center. Conversions would be expected to take several man years, and would require some systems level programmers.

The staff at LARS has been very generous in providing both computer time and computer system personnel at various times for a period of two years.

IV. Data Analysis - Objectives and Concepts

In this section, the objectives and concepts relating to the LANDSAT investigation, both LANDSAT imagery and aircraft, are formulated. The results are presented by states, LANDSAT imagery first then aircraft photography. At the conclusion, ways to use the classification results to make acreage estimates and a method to combine data from aircraft and satellite is presented.

4.1 Crop Classification

4.1.1 Objectives

1. Investigate the use of parametric discriminate functions.

2. Estimate the rate of misclassification for each type of crop.

3. Investigate the value of temporal overlays in reducing errors of misclassification.

4. Determine differences in classification rates between states.

5. Determine differences in classification rates between months within states.

6. Evaluate the use of training data parameters from (a) one LANDSAT frame to another, and (b) in aerial photography from one flightline to another.

7. Estimate the difference in classification results between dependent and independent data used in testing.

4.1.2 Concepts

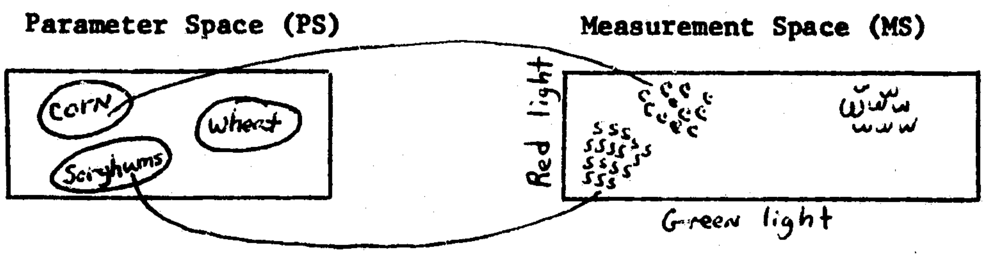
Discriminant Analysis

This background is intended to be general and enable the reader to understand the detailed computations and results that follow. Kendall and Stuart formulate Discriminant Analysis and Classification by stating...

"We shall be concerned with problems of differentiating between two or more populations on the basis of multivariate measurements... We are given the existence of two or more populations and a sample of individuals from each. The problem is to set up a rule, based on measurements from these individuals, which will enable us to allot some new individual to the correct population when we do not know from which it emanates." 1/

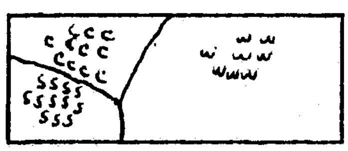
For example, the land population of interest was the Southwest Crop Reporting District (CRD) in Kansas. Wheat, sorghums, corn, oats, rye, and pasture are the major populations of interest. From every acre in the CRD, we have light intensity readings for green light, red light, and two infrared wavelengths. These light intensities are multivariate measurements that will be used to allot or classify each data point into a crop type such as corn, wheat, or sorghums. A graphical representation of the above formulation would be as follows:

Figure 8--Conceptualized mapping from agricultural fields into measurement space.



A sample of fields from each crop type is selected and their respective light intensities obtained. These sample points are plotted on a two-dimensional graph showing relative positions of each crop type in the Measurement Space (MS). The problem is to partition the measurement space in some optimal fashion so that points are allotted as nearly correct as possible. Figure 9 shows the measurement space as it might be partitioned.

Figure 9--Partitioned measurement space.



1/ M.G. Kendall and A. Stuart, The Advanced Theory of Statistics, 2nd Ed., Vol. 3, page 314.

Any point, no matter where it is in MS will be classified as one of the three crops. An unknown point where the number 1 is located in Figure 9 will be classified as wheat because wheat is probably the group to which it belongs. Likewise, a point in position 2 would be classified as sorghum and a point in position 3 would be classified as corn. A point in position 4 would also be classified as wheat, but the probability that it is actually wheat is not as great as that of a point in position 1.

There are many ways to partition a measurement space. We have done a simple non-statistical partition above, simply draw lines. Visually partitioning the measurement space may work when it is one or two dimensional, but for more than two dimensional measurement spaces, a visual partition is not possible. For most LANDSAT and aerial photography classification studies a four dimensional measurement space has been used.

The method used in this report was that of constructing contour "surfaces" in the MS. These dividing surfaces were constructed so that points falling on the dividing surface have equal probabilities of being in either group on each side. Those points not on the dividing surface always have a greater probability of being classified into the crop for which the point is interior to the contour surface. If prior knowledge of the population density function indicates that the density is multivariate normal, then a multivariate normal density distribution will be estimated for each crop. It is hoped that the data is approximately multivariate normal since only the mean vector and covariance matrix is required to estimate a discriminant function. Usually small departures from normality will not invalidate the procedure, but certain types of departures (for example, bimodal data) may be very detrimental to the statistical technique. However, the error rate and estimator properties are dependent on the assumptions of the distributions and prior information.

For example, in this study a multivariate normal density was assumed so it becomes quite simple to estimate the density functions and the discriminant scores which in turn determine boundaries.

The discriminant score for ith population is:

$$P_i \, (2\pi)^{-\frac{q}{2}} \, |\Sigma_i|^{-\frac{1}{2}} \, e^{-\frac{1}{2}(\chi-\mu_i)' \Sigma_i (\chi-\mu_i)}$$

where P_i is the prior probability for the ith crop

Σ_i is the covariance matrix (qxq) for the ith crop

μ_i is the mean vector (q length) for the ith crop

χ_{ij} is a set of measurements of an individual from the ith population. $j \neq i$ or $j = i$

or its equivalent discriminant score the $\log_{(e)}$ of S_i =

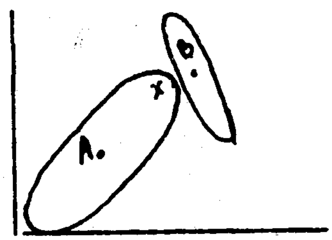
$$\log_e (P_i) - 1/2 \log_e |\Sigma_i| - 1/2 (\chi-\mu_i)' \Sigma_i^{-1} (\chi-\mu_i) .$$

The boundary between two populations is quadratic (curved) and the point χ that fall in the boundary have an equal probability of being in either population.

When an unknown land point is classified, its measurement vector is compared to the mean vector for each crop represented. The point is assigned to the crop whose mean point is "nearest" from a statistical point.

The procedure used for finding the "nearest" mean uses the Mahalanobis measure of distance, not the Euclidean. This is illustrated in Figure 10.

Figure 10--Measurement Space showing two crop density functions and an unknown point (χ).



The point χ is actually closest (Euclidean distance) to the mean vector (center point) of B. However, when one takes into account the variance and covariances, χ is found to be closest to Group A based on a probability concept and an outlier of Group B. Therefore, the point would be classified into Group A, because the probability that the point (χ) is a member of Group A is much greater than for Group B.

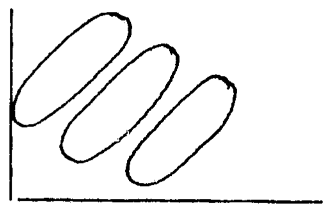
So the partitioning of the MS is done by computing the means for each crop type and using the Mahalanobis distances from this mean. This distance depends on the covariance matrix and is a measure of probability. The discriminant functions without prior probabilities are:

1) $(X - \overline{X}_i)' S^{-1}(X - \overline{X}_i)$, which is a sample estimate of $(X - \mu_i)'\Sigma^{-1} (\chi-\mu_i)$ if linear discriminant functions are used, and

2) $-1/2 \log_e |\Sigma_i| -1/2 (X - \overline{X}_i)' S_i^{-1} (X - X_i)$ if quadratic discriminant functions are used. These functions are the exponents of the density formula of the multivariate normal distribution C_{exp} - $1/2 (X - \mu_i) \Sigma_i^{-1} (X - \mu_i)$ depending on the i'th crop. If $\Sigma_i = \Sigma_j$ for all $i \neq j$ linear discriminant functions are used.

It is worth pointing out that if linear discriminant functions are used, one assumes (1) that $\Sigma_i = \Sigma_j$ and (2) that for all crops in the MS the major and minor axes are equal, and (3) the sample data of each crop has the same slope. Such an event in two-space is shown in Figure 11.

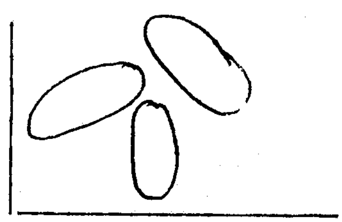
Figure 11--Measurement Space where crop types have same covariance matrix.



This space can be partitioned effectively with straight lines thus we can use linear discriminant functions.

Figure 12 shows a MS where covariance matrices are not equal, and therefore, linear discriminant functions are not appropriate. In either case, the Mahalanobis distance is used.

Figure 12--Measurement Space when crops have different covariance matrices.



In Figure 11, even though a common center point is not present, a common covariance (ellipse) matrix would be computed. In Figure 12 a different covariance matrix will be needed for each crop type. When the off-diagonal elements in the covariance matrix are unequal, the slopes of the data are different and linear discriminant functions are not appropriate.

The above techniques follow from our first assumption that the data is normally distributed in the MS. In practice, however, one does not decide what the distribution of the population density is in MS and program the correct procedure. One uses the available procedures for analyzing data. Most available programs assume multivariate normal data because the program and the calculations are greatly simplified. Thus, it becomes necessary to justify the use of these simplified programs.

In order to explain better how a parametric procedure can reduce the work load, consider that the first step in the discriminant analysis (DA) is to estimate the population density function in the MS, with a sample of points from each crop. Once these population density functions have been estimated, then partitioning the space is extremely simple.

To estimate a multivariate population density in MS for corn where we have no prior information except sample data on corn is extremely difficult. If a sample of 1000 points was available, each of these 1000 data points would need to be stored in the computer. On the other hand, if we are working with a multi-dimensional normal distribution, theory tells us that the sufficient statistics are computed (mean vector variance matrix) and stored in the computer.

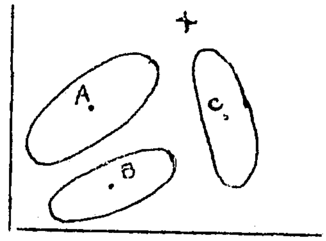
The individual data points could be discarded because no additional information about the population distribution in the MS is available in these points. (There would be information about how well the data fits the normal distribution in these 1000 data points).

Another consideration is that all the techniques we have described require independent random samples from each crop in order to estimate the population density in the MS (training data). This point is mentioned because most remote sensing analysts do not work with randomly selected points. In this study we have tried to work with randomly selected fields. However, the points within these fields are not a random sample of all possible points in a given crop, but the data are nested within fields. Consequently, the random selection is restricted to the selection of fields within the randomly selected segments.

One type of prior information that can be used in the classification procedure is the relative frequency of occurrence (prior probabilities) for each of the K populations in the total land population. For example, if 1/3 of all land is wheat, and 1/3 is pasture as it might be in parts of Kansas, this information would be used and it would effect the partitioning of the measurement space accordingly. If a crop has a high chance of selection, then the area in the MS would be increased. Conversely, if a certain crop has a very low chance of occurrence, then the area in MS would be adjusted downwards.

One last point to be covered on procedures used would be to define what is meant by thresholding. Suppose some unknown crop for which there is no sample in the original data set is to be classified. With the present system, the point will be classified as Crop A, B, or C, depending on its probability of being in either A, B, or C. For example, in Figure 13, if the probability P(AIX) that the point χ was Crop A is .01 and P(BIX) = .001, and P(CIX) = .02 the point χ would be classified as belonging to Crop C, even though the probability is only .02. It would be an outlier in MS for Crop C, and therefore, we may want to let it remain unclassified.

Figure 13--Measurement Space showing an outlier and three crop areas with 95% confidence limits.



4.1.3 Description of LANDSAT Data

The satellite data used in this report is LANDSAT Multi-Spectral Scanner (MSS) data and is described in Section 3 of Data User's Handbook. 1/

The MSS is a passive electro-optical system that can record radiant energy from the scene being sensed. All energy coming to earth from the sun is either reflected, scattered, or absorbed and, subsequently, emitted by objects on earth. 2/ The total radiance from an object is composed of two components, reflected radiance and emitted radiance. In general, the reflected radiance forms a dominant portion of the total radiance from an object at shorter wavelengths of the electromagnetic spectrum, while the emissive radiance becomes greater at the longer wavelengths. The combination of these two sources of energy would represent the total spectral response of the object. This, then, is the "spectral signature" of an object and it is the differences between such signatures which allows the classification of objects using the statistical techniques just discussed. The particular product is system corrected images refers to

---

1/ Published by Goddard Space Flight Center.

2/ Baker, J.R. and E.M. Mikhail, Geometric Analysis and Restitution of Digital Multispectral Scanner Data Arrays. LARS information note 052875.

products that contain the radiometric and initial spatial corrections introduced during the film conversion. Every picture element (pixel) is recorded with 4 variables - each variable corresponds to one of the 4 MSS bands. Table 23 shows the relationship between the MSS bands and light wavelengths.

Table 23--Sensor spectral band relationships.

Sensor	Spectral Band Number	Wavelenghts (micrometers)	Color	Band Code
MSS	1	.5 - .6	Green	4
MSS	2	.6 - .7	Red	5
MSS	3	.7 - .8	Near Infrared	6
MSS	4	.8 - 1.1	Infrared	7

The numbers are similar to transmission values - zero radiances at Step 15 which is black on positives and maximum radiance at Step 1 which is white on positives. The radiance varies linearly with gray scale stop transmission between these values with the difference between each step corresponding to 1/14th of the maximum radiance. The recording format in the CCT is 8 bits, the sensor range is 7 bits, and the actual dynamic range of usable data is between 5 and 6 bits.

The analysis was started by first locating the test and training data (ground observations with either the Penn State University program (NMAP) or an in-house program (RADMAP) that produces gray scale maps. 1/ After the ground enumeration information was located on LANDSAT CCT's, rectangular areas within fields were located and punched using the LARS field description card format. Once these cards were obtained and checked, the statistics function in LARSYS was employed to extract univariate graphs to detect bimodal classes.

In most cases, analysis proceeded from the statistics program to the Program for classify points, but with the introduction of a feature to use prior probabilities. These classifications were stored on tape by file number so the print results function could be run more than once.

4.1.4 Results

The results will be presented by state since there was a slightly different situation in each state. All LANDSAT analysis is presented first then the aircraft follows.

1/ See Section - Segment Location

Missouri LANDSAT:

The Crop Reporting District (CRD) that was the test site was in the southeast corner of the state. This area is outlined in black on the map of Missouri, Figure 3.

Summary of Results

The Missouri test site covers 4,660 square miles. There are 50 segments, each about a mile square. These segments constitute a random sample from all land areas. The ground enumeration was taken from these segments. This information was used for both training and testing.

Analysis of Missouri data was done using a tape that was assembled at LARS. The data for three dates, August 26, September 13, and October 21, 1972, were geometrically corrected then overlayed to create a tape with temporal data. Therefore, data used for analysis from three different times in the growing season was available and covered an area that contained 29 of the JES segments in this CRD. The principle results are summarized below:

1. A test was run on the covariance matrices between crops to see if they were equal. The results of this test were that they very likely were not equal. Thus, linear discriminant functions seemed inappropriate.

2. Best overall correct classification rate was 70%. This included using temporal overlays and using unequal prior probabilities.

3. Unequal prior probabilities for crops improved classification results by 10% over using the assumption of equal probabilities for crops.

4. The temporal data improved the classification by 10% even though the dates were not optimum.

5. One classification was run on data to estimate the effect of independent data. The difference was 9%, and was an over-estimate.

Data Analysis - LANDSAT

In the analysis, the equality of the covariance matrices was checked first because this is essential for the linear discriminant analysis assumptions to be valid. A test presented in Morrison's Multivariate Statistical Methods, page 152, was used to test the within crop covariance of LANDSAT data. This test is not robust with respect to certain departures from normality.

For the following example, August 26, 1972 imagery bands 4, 5, and 7 were used. The covariance matrices for cotton, soybeans, and grass were tested. The test was conducted as follows. The null hypothesis states that the covariance matrices are equal.

$$H_0: \Sigma_1 = \Sigma_2 = \Sigma_3$$

The alternative hypothesis is:

$$H_1: \Sigma_i \neq \Sigma_j \quad \text{for some } i \neq j$$

S_i is an estimate of Σ_i based on m_i degrees of freedom where i is a crop.

$$S \text{ cotton} = \begin{bmatrix} 6.76 & 7.01298 & .4914 \\ 7.01298 & 11.0889 & -5.6643 \\ .4914 & -5.6643 & 39.69 \end{bmatrix}$$

$$S \text{ soybeans} = \begin{bmatrix} 6.6049 & 8.3623 & .8265 \\ 8.3623 & 13.9876 & -6.3146 \\ .8265 & -6.3398 & 64.6416 \end{bmatrix}$$

$$S \text{ grass} = \begin{bmatrix} 5.6169 & 5.8416 & .7525 \\ 5.8416 & 9.7344 & -6.3398 \\ .7525 & -6.3398 & 40.3225 \end{bmatrix}$$

Now we form the pooled estimate of Σ.

$$S = \sum_{i=1}^{k} \frac{m_i S_i}{m} \begin{bmatrix} 6.5567 & 7.4436 & .6638 \\ 7.4436 & 12.1519 & -6.0189 \\ .6638 & -6.0189 & 50.2976 \end{bmatrix}$$

The statistic for the modified likelihood - ratio test is:

$$M = m \ln |S| - \sum_{i=1}^{k} m_i \ln |S_i|$$

$= 149.25$

Next, we form the scale factor:

$$C^{-1} = 1 - \frac{2P^2 + 3P - 1}{6(p+1)(k-1)} \sum_{i=1}^{k} \frac{1}{m_i} \frac{1}{\Sigma m_i} = .00678$$

and MC^{-1} is distributed approximately chi-squared with degrees of freedom 1/2 $(K-1)p(p+1)$ as m_i tends to infinity if H_0 is true.

$$MC^{-1} = .48.77 \quad d.f. = 12 \quad \alpha = .05 \quad \chi^2(12 \; \alpha = .05) = 22.36$$

Thus, we must reject the null hypothesis i.e. the data does not support the assumption that the covariance matrices are equal.

Therefore, the necessary assumptions for valid linear discriminant analysis are not met and better results might be attained by using quadratic discriminant functions. Generally, we used the quadratic approach on our analysis. However, it should be pointed out that upon close examination, the covariance matrices are very similar in many respects. Corresponding elements in the three covariance matrices are of at least the same order of magnitude and have the same sign. Under such conditions, it is possible to get acceptable results from a linear approach.

Conclusions of similar tests for the September 14, 1972 data were the same, the covariance matrices were unequal.

Results of the discriminant analysis (DA) are presented in a classification matrix (CM). Table 24 is an example of a CM using quadratic discriminant functions with unequal prior probabilities. The prior probabilities came from the June Survey early in the season. That is, it was not assumed that corn, cotton, soybeans, grass, and other all have the same probability of occurrence. The classification parameters were obtained from the same data that was used in the testing phase.

Although 12 bands were available, since three dates were involved, only nine were used in this study because three were of poor quality. There were two consecutive LANDSAT images that contained 29 segments. All data was used both to partition the measurement space (MS) and test the partition. The CM will be biased upward because data was used for both purposes, however, this bias should be small if ample data are available.

Table 24--Classification matrix of quadratic discriminant functions with unequal prior probabilities using data from three overflights1/, Missouri Study Area.

Group	No. of sample points	Percent Correct	Number of samples classified into				
			Cotton	Corn	Soybean	Grass	Miscellaneous
Cotton....	927	79.7	739	2	137	26	23
Corn......	58	44.8	9	26	7	14	1
Soybean...	852	71.8	99	12	612	96	23
Grass.....	240	53.3	42	1	66	128	4
Misc......	140	89.3	17	2	44	13	64
Totals....	2217		906	43	866	277	125

Overall performance 70.8 percent

1/
August 26, 1972, MSS bands 4, 5, 7
September 14, 1972, MSS bands 5, 7
October 2, 1972, MSS bands 4, 5, 6, 7

The leftmost column in Table 24 identifies the crop - cotton, corn, soybeans, grass, and miscellaneous. The next column gives the number of sample values in each of the crop classes. For example, there are 927 pixels to be classified. The next column tells the percent of these that were classified correctly as cotton (79.7%). The rest of the columns give the number of these pixels that were classified into each crop class, i.e. 739 were classified correctly as cotton, while the remainder were misclassified as follows: 2 of the 927 as corn, 137 as soybeans, 26 as grass, and 23 as miscellaneous. The overall performance in this table was 70.8 percent. To compute this figure, the correctly classified pixels were divided (the diagonal elements - 1569) by the total pixels 2217.

The prior probabilities used in this study were based on a statistical sampling of the entire land area. Data that is collected in this way enables the user to estimate the prior probability and take advantage of this procedure. Historic data could be used, but they are more difficult to justify when important changes between years are occurring.

The next table is the same as the last, except that equal prior probabilities were used.

Table 25--Classification matrix of quadratic discriminant functions with equal prior probabilities using data from three overflights 1/, Missouri Study Area.

Group	No. of sample points	Percent Correct	Number of samples classified into				
			Cotton	Corn	Soybean	Grass	Miscellaneous
Cotton....	927	74.3	689	21	83	36	98
Corn......	58	58.6	4	34	3	10	7
Soybean...	852	39.7	101	49	338	137	227
Grass.....	240	57.1	34	22	22	138	25
Misc......	140	75.0	14	5	7	9	105
Total.....	2217		842	131	453	329	462

Overall performance 58.8 percent

1/
August 26, 1972, MSS bands 4, 5, 7
September 14, 1972, MSS bands 5, 7
October 2, 1972, MSS bands 4, 5, 6, 7

Most classifications done so far by other remote sensing analysts have used this assumption that the crop classes are all equally likely to occur. Most people feel this assumption is not detrimental, however, this example illustrates that it can make a difference. Especially, if acreage for the crop classes does vary vastly or when crops are hard to distinguish. Two properties are worth noting, classification results, and the statistical properties are much better in Table 24 than in Table 25. For example, in Table 24 the total number of pixels classified as cotton is 906, compared to the actual number of 927. In Table 25, the number of cotton pixels is 842.

A similar comparison is even more drastic with soybeans. In Table 24, 866 pixels were classified as soybeans while 842 actual points were soybeans. In Table 25, there were 453 points classified as soybeans. Further, the statistical properties of the estimates are better since if the data is normal, and the prior probabilities are correct, we obtain unbiased estimates of crop categories and we can estimate the Bayes error rates (minimum error rates) using the classification.

A chi-square test for discriminatory power was run on the CM of Table 24 and 25. 1/ The null hypothesis is that the classification was done strickly at random. If the null hypothesis is correct, then the spectral information was useless as far as giving information that would help assign the data to a crop class. If the above hypothesis is correct, then the statistic $\frac{(n-e)^2}{e} + \frac{(\bar{n}-\bar{e})^2}{\bar{e}}$ has a chi-square distribution with 1 degree of freedom. Where n and $\bar{n}$ are the number of correctly classified and misclassified points respectively and e and $\bar{e}$ are the expected number of correctly classified and misclassified points under the null hypothesis.

The chi-square for Table 24 is 4626 and for Table 25 is 2782. These chi-square values with one degree of freedom are highly significant, and therefore, we conclude that the classification was not done at random. Another chi-square test based on the difference between the marginal sums and the correct number of data points in each class for Table 25 is as follows:

$$X^2_{(5)} = \frac{(906-927)^2}{927}+\frac{(43-58)^2}{58}+\frac{(866-852)^2}{852}+\frac{(277-240)^2}{240}+\frac{(140-125)^2}{125} =$$

$$.47 + 3.87 + .23 + 5.70 + 1.61 = 11.89$$

This chi-square statistic is similar to the one before, except that there are 4 degrees of freedom. $\sum_{i=1}^{k} \frac{(n-e)^2}{e_i}$ where n and e have the same meaning as before.

This chi-square value of 11.89 is significant, and therefore, the hypothesis that the marginal totals in Table 24 are estimating the actual row totals is rejected. Note that the components for grass and corn are the major contributors to the significant chi-square.

The authors know of no statistical test that compare one C.M. with another C.M., but there are two criteria that can be used to help evaluate a certain C.M. The first criterion simply assigns each misclassified point a loss of 1 and each correctly classified point as loss of 0. Under this criterion, Table 24 has a loss value of 648 and Table 25 has a loss value of 914. This criterion is crude, but it seems reasonable for our purposes to give a misclassified corn pixel the same weight as the misclassified cotton pixel.

1/ S. James Press, Applied Multivariate Analysis, pages 381-383.

The next criterion is a bit more subtle. It uses the marginal totals in the C.M. For example, in Table 24 the column sum for cotton is 906. This means that 906 pixels were classified as cotton. Actually, there were 927 cotton pixels. In Table 25, there were 842 pixels classified into the cotton group. This is not close to the correct number of 927. The marginal estimate (906) from Table 24 is within 2 percent of the actual. In Table 25, the marginal estimate of 842 or within 9 percent. Table 26 presents these estimates along with the percentages of the true value.

Table 26--Marginal estimate and difference from actual values.

Group	Actual	Unequal Prior Probabilities			Equal Prior Probabilities		
		Estimate	Difference	Percent	Estimate	Difference	Percent
Cotton..	927	906	21	2.2	842	85	9.2
Corn....	58	43	15	25.9	131	73	125.9
Soybean.	852	866	14	1.6	453	399	46.8
Grass...	240	277	37	15.4	329	89	37.1
Winter Wheat...	85	27	27	68.2	346	261	307.1
Odd.....	55	98	43	78.2	116	61	110.9

In every case, unequal prior probabilities were superior to the equal prior probabilities model and in some cases, substantially so. For example, the number of corn pixels for Table 25 was 131 or 125.9 percent of the difference from the actual 58. The number of corn pixels for Table 24 is 43 or 25.9 percent of the difference from the actual 58 pixels. Soybeans, a very important item, also shows a significant improvement over the equal probability model. Actually, the soybean estimate for the equal prior probability model was 46.8 percent which the estimate for the unequal prior probability model was 1.6 percent.

Next, the point classification systems were compared to the per-field classification scheme. Table 27 presents the C.M. for the per-field classifier system. With a point classification system, each point in a field can be assigned to any of the crop categories. With the sample classifier, all points in the field are assigned to the same crop class. One drawback to the procedure is that there were a large number of fields that were not assigned to a crop because the data set was not large enough. The technique requires the covariance matrix to be inverted and therefore, p+1 data points are required (where p is the number of variables). However, if enough points are present, classification performance has generally been found to be excellent.

In the work done in Missouri using the sample classifier, about 40 percent of the fields were not classified because the required number of points for the classifier (10 is this particular case) exceeded the number of points present within the defined fields. Of the total number of fields, 32.9 percent were correctly identified. Considering only those fields which were classified, 54 percent were classified correctly.

Table 27--Per-field classification matrix based on data from 3 overflights.1/

Group	No. of fields	Percent fields correct	No. of samples	COTTON	CORN	SOY-BEANS	GRASS	MISC.	NOT CLASSIFIED
Cotton	38	63.2	927	24	0	2	0	1	11
Corn..	7	14.3	58	0	1	0	1	1	4
Soy-beans.	58	25.9	852	9	3	15	3	8	20
Grass.	31	9.7	240	3	1	1	3	2	21
Misc..	9	44.4	140	1	0	1	1	4	2
Totals	143	32.9	2217	37	5	19	8	16	58

1/
August 26, 1972, MSS bands 4, 5, 7
September 14, 1972, MSS bands 5, 7
October 2, 1972, MSS bands 4, 5, 6, 7

Temporal Overlay

The next analysis investigated the value of a temporal overlay of the three LANDSAT passes. This particular data set was a temporal overlay of three LANDSAT passes. Each pass could also be compared with the three passes. However, there were 3 bad bands in the total of 12. Two poor quality bands were in the September 14 imagery and one poor quality band was in the August 26 imagery. This makes it difficult to compare the three dates since the number of bands were confounded with dates. Nevertheless, the C.M.'s for each date are presented in Tables 28, 29, and 30. These tables can be compared to the 9 band-overlay of Table 24 since they are all unequal prior probability models.

Table 28--Classification matrix using August 26, 1972, MSS bands 4, 5, and 7 with unequal prior probabilities.

Group	No. of sample points	Percent Correct	Number of samples classified into: Cotton	Corn	Soybean	Grass	Miscellaneous
Cotton...	927	60.6	562	1	311	22	31
Corn.....	58	10.3	12	6	30	2	8
Soybean..	852	86.0	70	2	733	29	18
Grass....	240	8.3	42	7	167	20	3
Misc.....	140	31.4	9	3	76	8	44
Totals...	2217		696	19	1317	81	104

Overall performance 61.5 percent

Table 29--Classification matrix using September 13, 1972, MSS bands 5 and 7 with unequal prior probabilities.

Group	No. of sample points	Percent Correct	Number of samples classified into: Cotton	Corn	Soybean	Grass	Miscellaneous
Cotton...	927	69.7	646	0	246	14	21
Corn.....	58	0.0	12	0	16	20	10
Soybean..	852	67.6	175	1	576	74	26
Grass....	240	42.1	40	0	97	101	2
Misc.....	140	22.8	14	2	82	10	32
Totals...	2217		887	3	1017	219	91

Overall performance 61.0 percent

Table 30--Classification matrix using October 2, 1972, MSS bands 4, 5, 6, and 7 with unequal prior probabilities.

Group	No. of sample points	Percent Correct	Number of samples classified into				
			Cotton	Corn	Soybean	Grass	Miscellaneous
Cotton...	927	73.2	679	6	161	59	22
Corn.....	58	12.1	30	7	14	1	6
Soybean..	852	62.4	200	7	532	76	37
Grass....	240	27.9	83	0	89	67	1
Misc.....	140	17.9	30	1	73	11	25
Totals...	2217		1022	21	869	214	91

Overall performance 59.1 percent

Table 31 summarizes these three classification matrices in 1 table.

Table 31--Comparison of multitemporal classification performance to classification of single dates. 1/ Missouri Study Area.

Group	Multitemporal	Aug. 26	Sept. 14	Oct. 2
Cotton	29.7	60.6	69.7	73.2
Corn	44.8	10.3	0.0	12.1
Soybeans	71.8	86.0	67.6	62.4
Grass	53.3	8.3	42.1	27.9
Misc.	89.3	31.4	22.8	17.9
Overall	70.8	61.6	61.1	59.2

1/ Unequal prior probabilities were used for all classification.

The same classifications were run for all dates individually except that equal prior probabilities were used.

Table 32--Classification matrix for August 26, 1972, based on MSS bands 4, 5, and 7 using equal prior probabilities.

Group	No. of sample points	Percent Correct	Number of samples classifed into				
			Cotton	Corn	Soybean	Grass	Miscellaneous
Cotton...	927	60.7	563	92	108	63	101
Corn.....	58	56.9	2	33	0	7	16
Soybean..	852	15.3	57	72	130	245	348
Grass....	240	45.4	32	41	26	109	32
Misc.....	140	62.9	11	10	13	18	88
Totals...	2217		665	248	277	442	585

Overall performance 41.6 percent

Table 33--Classification matrix for September 13, 1972 based on MSS bands 5 and 7 using equal prior probabilities.

Group	No. of sample points	Percent Correct	Number of samples classified into				
			Cotton	Corn	Soybean	Grass	Miscellaneous
Cotton...	927	60.7	563	92	108	63	101
Corn.....	58	56.9	2	33	0	7	16
Soybean..	952	15.3	57	72	130	245	348
Grass....	240	45.4	32	41	26	109	32
Misc.....	140	62.9	11	10	13	18	88
Totals...	2217		665	248	277	422	585

Overall pefformance 50.8 percent

Table 34--Classification matrix for October 2, 1972 based on MSS bands 4, 5, 6, and 7 using equal prior probabilities.

Group	No. of sample points	Percent Correct	Number of samples classified into: Cotton	Corn	Soybean	Grass	Miscellaneous
Cotton...	927	66.7	618	35	30	149	95
Corn.....	58	37.9	21	22	4	4	7
Soybean..	952	20.8	142	46	177	141	346
Grass....	240	42.5	58	9	23	102	48
Misc.....	140	60.7	20	8	8	18	85
Totals...	2217		860	120	242	414	581

Overall performance 45.3 percent

Table 35 summarizes these tables.

Table 35--Comparison of multitemporal classification performance to classifications of single dates using equal prior probabilities. 1/ Missouri Study Area.

Group	Multitemporal	Aug. 26	Sept. 13	Oct. 2
Cotton	74.3	60.7	71.4	66.2
Corn	58.6	56.9	34.5	37.9
Soybeans	39.7	15.3	28.9	20.8
Grass	57.1	45.4	44.6	42.5
Misc.	75.0	62.9	65.7	60.7
Overall	58.8	41.6	50.8	45.3

The temporal overlay classification of Table 25 shows an overall performance of 58.8 percent as compared to 41.6 percent, 50.8 percent, and 45.3 percent, respectively, for Tables 32, 33, 34. Based on these comparisons, the temporal overlay does improve the classification, However, the evaluation can become more difficult to interpret in the temporal overlay tapes because of changes in land use from one date to the next. Thus, the time of year becomes very important in areas where double-cropping is common or preparation of land follows each crop. It should be pointed out that these dates were not optimal. Other dates would have given different results.

Independent Test Data

The last exercise was completed to estimate the C.M. in Missouri on independent data. Since the number of fields and points within are small and the area covered is large, we need more training data to represent the total area. It did not seem possible to divide the set into halves and still have enough training data. It was decided to use a jacknife procedure. This procedure has the advantage of giving unbiased estimates that are simple to calculate. The data were divided into three equal subgroups, two groups were used to train with and the third group was used as a test group. This was repeated three times, each time with a different group used as test data. These three tables are presented separately, then the three are combined and presented to give an unbiased estimate of the classification matrix where independent test data is used. By using independent data, it is hoped that the bias caused by using the same data for both training and testing would be eliminated, but the variance of each item in the latter tables may be somewhat higher than those in the previous tables since a smaller data set was used.

One cotton field of 27 points was not included in any of the three groups. So the total in Table 39 is 27 pixels smaller than the total of earlier tables. Table 39 is the matrix sum of Tables 36, 37, and 38.

Table 36--Classification matrix using August 26, 1972, MSS bands 4, 5, and 7 with subgroups 2 and 3 as training data and subgroup 1 as test data.

Group	No. of sample points	Percent Correct	Number of samples classified into				
			Cotton	Corn	Soybean	Grass	Miscellaneous
Cotton	479	56.2	269	11	129	36	34
Soybean..	138	45.7	35	6	63	17	17
Grass	66	34.8	15	7	15	23	6
Misc.	68	16.2	1	4	39	13	11
Totals	751		320	28	246	89	68
Overall performance 48.7 percent							

Table 37--Classification matrix using August 26, 1972 MSS bands 4, 5, and 7 with subgroups 1 and 3 as training data and subgroup 2 as test data.

Group	No. of sample points	Percent Correct	Number of samples classified into: Cotton	Corn	Soybean	Grass	Miscellaneous
Cotton...	290	57.6	167	36	11	19	57
Corn.....	29	13.8	1	4	0	8	16
Soybean..	308	13.0	48	53	40	20	147
Grass....	42	28.6	1	11	4	12	14
Misc.....	57	78.9	0	2	8	2	45
Totals...	726		217	106	64	63	279

Overall performance 36.9 percent

Table 38--Classification matrix using August 26, 1972 MSS bands 4, 5, and 7 with subgroups 1 and 2 as training data and subgroup 3 as test data.

Group	No. of sample points	Percent Correct	Number of samples classified into: Cotton	Corn	Soybean	Grass	Miscellaneous
Cotton...	131	47.3	62	22	1	22	24
Corn.....	29	41.4	3	12	2	5	7
Soybean..	406	200	6	29	8	137	226
Grass....	132	43.2	20	27	0	57	28
Misc.....	15	0.0	5	2	0	8	0
Totals...	713		96	92	11	229	285

Overall performance 19.5 percent

Table 39--Classification matrix combining Tables 36, 37, and 38.

Group	No. of sample points	Percent Correct	Number of samples classified into: Cotton	Corn	Soybean	Grass	Miscellaneous
Cotton...	900	55.3	498	69	141	77	115
Corn.....	58	27.6	4	16	2	13	23
Soybean..	852	13.0	89	88	111	174	390
Grass....	240	28.3	36	45	19	92	48
Misc.....	140	40.0	6	8	47	23	56
Totals...	2190		633	226	320	379	632

Overall performance 34.6 percent

The comparable classification where non-independent data was used is shown in Table 40.

Table 40--Classification matrix using August 26, 1972, MSS bands 4, 5, and 7.

Group	No. of sample points	Percent Correct	Number of samples classified into: Cotton	Corn	Soybean	Grass	Miscellaneous
Cotton...	927	60.7	563	92	108	63	101
Corn.....	58	56.9	2	33	0	7	16
Soybean..	852	15.3	57	72	130	245	348
Grass....	240	45.4	32	41	26	109	32
Misc.....	140	93.6	11	10	13	18	131
Totals...	2217		665	248	277	442	585

Overall performance 43.6 percent

Anytime the results differ this much between data sets, we know the data set is either too small or the bias is large. Obviously, we have not reached the point where we have covergence of parameters based on independent and non-independent data sets. The sample sizes necessary depends on the variation in the data set and the variation in the data set is generally a function of how dispersed the data really is. One thing is certain with a small data set, either procedure may lead to erroneous conclusions.

Kansas:

The LANDSAT analysis was done on the CRD in the southwest corner of the State. Figure 2 shows the State of Kansas with the study area outlined.

Analysis of Kansas LANDSAT Data

The objective of the analysis of Kansas LANDSAT data were the following:

1. Test the covariance matrices of the most important crops to see if they were equal.

2. Compute the classification rates for the Kansas test site.

3. Compute the correlation coefficients between ground observation acreage and classified pixels.

4. Study the effect of classification in one LANDSAT frame using training parameters from an adjoining pass taken one day apart.

5. Study the classification of a Kansas county.

Approach:

1. LANDSAT imagery for the study area was too cloudy to be useful, prior to September 21, 1972. The study was based on September 21 and 22 imagery. The area of interest in Kansas was divided by two LANDSAT passes, thus the training data was also divided. Twenty-two segments were in the September 21 imagery. Seven of these segments were hidden by clouds. Therefore, 15 segments were used as training and test data.

   Since the time of year was not conducive to optimal results, a visual inspection of the grey-scale printout of MSS band 5 and ground truth was used to select particular fields to use as training fields; i.e. those fields which were partially harvested and those with a confusion of symbols were discarded. Another reason for selecting fields was to compare parameters from one pass with those from another as described in this report.

As a first step, the covariance matrices of the most important crops were compared and tested within frames and between frames. Tables 41 and 42 show the pertinent data.

The test criterion was computed and indicates that the within-crop covariances are statistically different. Also, the covariances between frames for the same crops were tested and are significantly different.

This would indicate that quadratic discriminant analysis could produce better results. In addition, a method of signature extension would be complicated if one wished to go from one frame to another.

2. The next step was to employ the quadratic classifier for the training data. The classification based on these select fields is presented in Table 43.

The overall performance was 91.2%. The classification used the standard pointwise quadratic discriminant functions found in LARSYS with the added feature of allowing unequal prior probabilities for the different crops. The unequal prior probabilities use information that is available about the likelihood of certain crops. If, for example, corn is more likely to be encountered than grain sorghum, corn is given a higher chance of occurrence. In most classifications using unequal prior probabilities done in Kansas, the prior probabilities were:

1) Alfalfa - .03
2) Pasture - .72
3) Corn - .09
4) Grain Sorghum - .16

Prior probabilities in this report were computed from a probability survey conducted by the Statistical Reporting Service in June 1972, (June Enumerative Survey).

In Table 43, the number of pixels to be classified are not proportional to the prior probabilities. The prior probabilities are based on acreage of all segments in the Crop Reporting District, and not the segments in frame 1060-16512. Development of proper prior probabilities for areas divided by LANDSAT passes presents additional problems. A better correspondence would have resulted in higher overall classification; however, 91.2% is very good.

Table 41--Covariance matrices and mean vectors for frame 1060-16512. (September 21, 1972).

		Mean	Covariance			
Alfalfa	n = 43					
		26.63	3.430			
		19.58	4.531	8.535		
		50.81	-2.357	-8.199	27.346	
		30.28	-2.751	-7.357	16.363	12.301
Pasture	n = 6378					
		29.70	10.926			
		26.36	12.975	21.821		
		56.88	10.351	12.698	22.487	
		20.07	4.405	4.332	11.388	7.339
Corn	n = 332					
		31.63	46.883			
		29.71	77.701	133.003		
		43.03	26.525	42.905	33.798	
		24.84	2.728	-6.399	11.275	10.978
Grain Sorghum	n = 508					
		32.21	115.096			
		27.32	130.402	154.965		
		43.78	78.251	85.757	76.431	
		25.65	18.089	16.152	29.548	18.198

Table 42--Covariance matrices and mean vectors for frame 1061-16570. (September 22, 1972).

		Mean	Covariance			
Alfalfa	n = 78					
		24.23	8.180			
		15.96	12.793	24.701		
		55.61	-18.345	036.494	71.234	
		34.51	-15.063	-29.604	50.802	39.313
Pasture	n = 320					
		28.62	5.290			
		25.53	6.109	11.002		
		35.98	3.534	3.061	19.272	
		19.81	1.056	0	11.213	8.237
Corn	n = 337					
		24.52	1.877			
		19.91	2.183	9.120		
		36.88	0.339	-5.114	17.056	
		22.82	-0.081	-5.291	11.039	8.820
Grain Sorghum	n = 177					
		27.16	32.718			
		22.76	49.217	77.088		
		43.69	2.100	2.865	16.646	
		27.09	-15.639	-24.393	10.975	19.448

Table 43--Classification matrix for September 21, 1972 MSS bands 4, 5, and 7, using quadratic discriminant functions with unequal prior probabilities in Kansas test site for select fields.

Class	No. of sample points	Percent Correct	Number of samples classified into				
			Alfalfa	Pasture	Corn	Sorghum	Threshold
Alfalfa..	43	100.0	43	0	0	0	0
Pasture..	172	98.3	0	169	2	1	0
Corn.....	51	90.2	0	1	46	4	0
Grain Sorghum..	78	69.2	0	10	14	54	0
Totals...	344		43	180	62	59	0

Overall performance 91.2%

A classification was then done using all identifiable fields in the 15 segments. The results of this classification are presented in Table 44. The overall performance was 90.2%.

There was a small decrease in overall performance between Table 43 and Table 44. However, a random sample of ground truth yields a better representation of all land and allows statistical inferences about the pixels.

The second pass required to cover the Kansas test site was analyzed in the same way as described above. The second scene contained 23 segments, but one of these segments fell in a non-agricultural area. In addition, to the random segments, two additional segments were selected which contained sugar beets.

Table 45 presents the classification of select fields for the second pass. The fields were selected from the grey-scale printout as described above. The overall performance was 75.5%.

Table 44--Classification matrix for September 21, 1972 imagery (MSS bands 4, 5, 6, and 7), using quadratic discriminant functions with unequal prior probabilities in Kansas test site.

Class	No. of sample points	Percent Correct	Number of samples classified into				
			Alfalfa	Pasture	Corn	Grain Sorghum	Threshold
Alfalfa..	43	93.0	40	2	0	1	0
Pasture..	6378	95.0	23	6061	123	142	29
Corn.....	332	37.7	38	110	125	59	00
Grain Sorghum..	508	64.8	38	77	60	329	44
Totals...	7261		139	6250	308	531	33

Overall performance 90.2%

Table 45--Classification matrix for September 22, 1972 imagery (MSS bands 4, 5, 6, and 7), using quadratic discriminant functions with unequal prior probabilities in Kansas test site for select fields.

Class	No. of sample points	Percent Correct	Number of samples classified into				
			Alfalfa	Pasture	Corn	Grain Sorghum	Threshold
Alfalfa..	78	84.6	66	12	0	0	0
Pasture..	230	93.0	0	214	11	5	0
Corn.....	337	65.0	0	93	219	25	0
Grain Sorghum..	177	63.9	3	34	18	122	0
Totals...	822		69	353	248	152	0

Overall performance 75.5%

Table 46 represents a classification of the second scene, using all identifiable fields. The overall performance was 65.8%. This decrease in performance could be attributed to several things. The number of crops being classified was increased from four to seven. Increasing the number of crops will reduce the performance. Secondly, there was a confusion between most crops and pasture. This could have resulted from using late September imagery; all crops are spectrally similar. Thirdly, the frequency of the data pixels presented for classification differed drastically from the prior probabilities used.

Table 47 is a classification study using the same select training fields that were used in Table 45. However, in Table 47 equal prior probabilities were applied. In Table 47, the overall performance at 79.2% is actually better than the 75.5% in Table 45. Applying prior probabilities based on all fields to a non-random selection of fields in a particular area is the cause for the lower classification in Table 45.

Table 48 presents a classification of all identifiable fields in scene 1061-16570, using equal prior probabilities. This table is comparable with the weighted classification presented in Table 46. The overall performance was increased 4.4% by using prior probabilities. When all fields are used in the classification, the total acres per crop more closely estimate the true prior probabilities of the model.

The increase caused by using unequal prior probabilities in Kansas was not as great as it had been in other areas. The smaller gain from prior probabilities is perhaps caused by the fact that the LANDSAT data contained more information; i.e., the classes were more separable. Thus, the expected gain from prior probabilities is greater in areas where classification is poorer.

3. The correlations between acres and pixels were calculated. Coordinates of ground truth segments were carefully defined. The training data from each scene were used to classify the segments in that scene. The classified pixels in the two scenes were then combined (i.e., Tables 44 and 46 were combined) and correlations with known ground truth acreage were computed.

Correlations between acreage and pixels were calculated as follows:

Total Acreage vs Total Pixel	$r^2 = .88$	$r = .94$
Pasture Acreage vs Pasture Pixel	$r^2 = .84$	$r = .92$
Corn Acreage vs Corn Pixel	$r^2 = .62$	$r = .79$
Grain Sorghum vs Grain Sorghum Pixel	$r^2 = .58$	$r = .76$

Table 46--Classification matrix for September 22, 1972 imagery (MSS bands 4, 5, 6, and 7), using unequal prior probabilities, Kansas, all fields.

Class	No. of sample points	Percent Correct	Number of samples classified into						
			Alfalfa	Pasture	Corn	Grain Sorghum	Winter Wheat	Sugar Beets	Threshold
Alfalfa.........	287	56.4	162	57	12	23	16	6	0
Pasture.........	4975	90.6	19	4508	45	44	156	0	23
Corn............	1698	40.8	1	684	693	174	99	0	0
Grain Sorghum...	2869	55.3	89	300	357	1586	265	0	4
Winter Wheat	863	13.3	14	431	16	41	115	0	4
Fallow..........	1508	64.6	10	285	44	56	134	2	3
Sugar Beets.....	25	0.0	16	2	1	1	5	0	0
Totals..........	12225		311	6267	1168	1925	790	8	34

Overall performance 65.8 percent

Table 47--Classification matrix for September 22, 1972 imagery, MSS bands 4, 5, 6, and 7, using quadratic discriminant functions with equal prior probabilities in Kansas test site for select fields.

Class	No. of sample points	Percent Correct	Number of samples classified into				
			Alfalfa	Pasture	Corn	Grain Sorghum	Threshold
Alfalfa..	78	84.6	66	11	0	1	0
Pasture..	230	75.2	3	173	38	16	0
Corn.....	337	87.5	0	29	295	13	0
Grain Sorghum..	177	66.1	14	16	30	117	0
Totals...	822		83	299	363	147	0

Overall performance 79.2%

When pixels and acreage are this highly correlated, remotely sensed data is beneficial.

4. In this study, the statistics compiled on one LANDSAT frame were used to classify points in the adjacent frame. As described earlier, two adjacent passes were used to obtain necessary coverage of Kansas. The select fields from both scenes (as described in Section A), had four classes (alfalfa, pasture, corn, grain sorghum. These four classes were also the classes for the "all fields" in frame 1060-16512. One requirement is that the same classes be used for training as those classified. The classification used the quadratic discriminant function with unequal prior probabilities.

   Table 49 presents the results of classifying the select fields in frame 1060-16512, using training statistics generated from select fields in frames 1061-16570. The overall performance was 54.4%; however, the average performance by classes 1/ was 33.3% correct classification. The 100% correct classification of the pasture class greatly influenced the overall classification.

1/ The average performance by classes is computed by averaging the percent identified for each class.

Table 48--Classification matrix for September 22, 1972 imagery, 4 bands using equal prior probabilities Kansas.

Class	No. of sample points	Percent Correct	Number of samples classified into							
			Alfalfa	Pasture	Corn	Grain Sorghum	Winter Wheat	Fallow	Sugar Beets	Threshold
Alfalfa........	287	50.5	145	18	30	9	24	4	57	0
Pasture........	4975	80.1	61	3986	371	66	340	106	22	23
Corn,..........	1698	70.3	80	267	1193	69	39	32	18	0
Grain Sorghum..	2869	42.1	496	115	620	1209	149	103	174	3
Winter Wheat...	863	23.4	20	350	50	44	202	149	44	4
Fallow.........	1508	50.5	18	208	79	120	256	762	62	3
Sugar Beets....	25	56.0	6	2	2	0	1	0	14	0
Totals.........	12225		826	4946	2345	1517	1011	1156	391	33

Overall performance 61.4%

Table 49--Classification matrix of select fields in frame 1060-16512 classification, using statistics from select fields in frame 1061-16570.

Class	No. of sample points	Percent Correct	Number of samples classified into				
			Alfalfa	Pasture	Corn	Grain Sorghum	Threshold
Alfalfa..	43	0.0	0	41	0	1	1
Pasture..	172	100.0	0	172	0	0	0
Corn.....	51	0.0	3	7	0	41	0
Grain Sorghum..	78	33.3	7	28	15	26	2
Totals...	344		10	248	15	68	3

Overall performance 54.4%

Table 50 is a classification of all identifiable fields in the segments in frame 1060-16512, using the statistics generated from the select fields in frame 1061-16570. The classifications with an overall performance of 65.5% and an average class performance of 48.5% are very good. Here again, it was the correctly classified pasture points which kept the averages high. In Table 50, more fields were classified and the influence of prior probabilities was more beneficial than in the cases where select fields were classified.

Table 51 shows a classification of select fields in frame 1061-16570, using statistics generated from all fields in frame 1060-16512. In this study the overall performance slipped to 49.0% but the average class performance was 59.1%. Classification was very good in all classes except corn, which was confused with pasture and grain sorghum. The time of year may have caused this confusion.

5. The border of Stevens County, Kansas was drawn on a grey-scale map of MSS band 5. The area was then defined on punch cards and classified. Training data for the classification were obtained from segments in the Crop Reporting District which contains Stevens County. Three of these segments were actually in Stevens County. A total of 410,505 pixels were classified which correspond to a calculated 466,560 acres in the county.

Table 50--Classification matrix of all fields in frame 1060-16512 classification, using statistics generated from "select fields" in frame 1061-16570.

Class	No. of sample points	Percent Correct	Number of samples classified into				
			Alfalfa	Pasture	Corn	Grain Sorghum	Threshold
Alfalfa..	43	65.1	28	3	0	12	0
Pasture..	6378	93.2	7	5943	11	277	140
Corn.....	332	7.5	8	79	25	204	16
Grain Sorghum..	508	28.3	16	105	75	144	168
Totals...	7261		59	6130	111	637	324

Overall performance 85.5%

Table 51--Classification matrix of select fields in frame 1061-16570 classification, using statistics generated from "all fields" in frame 1060-16512.

Class	No. of sample points	Percent Correct	Number of samples classified into				
			Alfalfa	Pasture	Corn	Grain Sorghum	Threshold
Alfalfa..	78	80.8	63	12	0	0	3
Pasture..	230	94.3	0	217	4	8	1
Corn.....	337	9.2	5	140	31	161	0
Grain Sorghum..	177	52.0	12	30	43	92	0
Totals...	822		80	399	78	261	4

Overall performance 49.0%

Alfalfa, pasture, corn, and grain sorghum were the crops classified. The following classification was obtained:

Number of Pixels	Alfalfa	Pasture	Corn	Grain Sorghum	Threshold
410,505	5,362	172,021	30,448	165,107	37,567
	1.3%	41.9%	7.4%	40.2%	9.2%

The prior probabilities as a percentage which were applied were the following:

Alfalfa	3%
Pasture	72%
Corn	9%
Grain Sorghum	16%

There is confusion between pasture and grain sorghum. Ways to use this data to produce a final estimate will be discussed in the section on estimation.

South Dakota

The test site in South Dakota is in the eastern part of the State. Figure 1 shows this Crop Reporting District.

Analysis of LANDSAT Data in South Dakota

Objectives:

The objective of this section was to determine the classification accuracy in the South Dakota test site.

Approach:

Imagery for three dates was available. However, the August and early September imagery was too cloudy to be useful. Thus, later September imagery was used. All 34 segments were contained in one LANDSAT frame (1060-16491). The segments and fields within segments were located and defined on punch cards. These segments were used for both training and classifying.

The LARS classifier with unequal prior probabilities was used. The classifier is a standard discriminant analysis.

Tabld 52 presents a classification of pixels in all segments in South Dakota. The overall performance was 30%, but the average class performance was 15%. Almost all classes in Table 52 were classified as either pasture or oats.

There were two reasons for this. First, prior probabilities used were large for pasture and oats, and second, the spectral data is quite similar at this period of time for all crops.

An attempt to improve the classification results was made by selecting fields that looked homogeneous.

These selected fields were used as training data and then classified. The results of this classification are presented in Table 53. The overall performance was 26% and the average class performance was 44%. There appears to be very little information in the data which would aid in the separation of crops. The influence of the prior probabilities again was the reason pasture and oats had high correct classification rates.

There must be reasons for the very poor classification rates. As an attempt to determine the reasons for the poor results, we have studied the means and covariances. They are in Table 54. It appears to be impossible to separate these classes with this data. Simply looking at the data does not necessarily show the true multivariate situation is four dimensional - but it does give an indication.

Summary

In South Dakota, late September imagery was used because of cloud cover in earlier imagery. Classification results were poor. Examination of Table 54 showed very little information in the data for the separation of the classes of interest. This late in the season, crops were classified as either pasture or oats.

The use of homogeneous fields selected from gray scale printouts and ground truth did not improve classification, and actually reduced the overall performance rates.

Table 52--Classification matrix for September 21, 1972 imagery (MSS bands 4, 5, 6, and 7), using unequa prior probabilities in South Dakota test site.

Class	No. of sample points	Percent Correct	Number of samples classified into										
			Corn	Pasture	Oats	Barley	Rye	Alfalfa	Flax	Sudex	Idle	Fallow	Threshold
Corn	1060	0.1	1	753	275	3	0	0	3	0	12	10	3
Pasture	812	88.4	1	718	86	1	0	0	0	0	2	4	0
Oats	243	40.3	0	142	98	0	0	0	0	0	0	3	0
Barley	97	0.0	0	77	17	0	0	0	1	0	2	0	0
Rye	16	0.0	0	15	1	0	0	0	0	0	0	0	0
Alfalfa	303	0.3	0	243	51	0	1	1	0	0	0	6	1
Flax	71	4.2	0	45	23	0	0	0	3	0	0	0	0
Sudex	55	0.0	0	47	7	0	0	0	0	0	0	1	0
Idle	18	10.5	0	14	3	0	0	0	0	0	2	0	0
Fallow	82	4.9	0	59	17	0	0	0	0	0	2	4	0
Totals	2758		2	2113	578	4	1	1	7	0	20	28	4

Overall performance 30.0%

Table 53--Classification matrix for September 21, 1972 imagery (MSS bands 4, 5, 6, and 7) using quadratic discriminant functions with unequal prior probabilities in South Dakota test site for select fields.

Class	No. of sample points	Percent Correct	Number of samples classified into					
			Corn	Pasture	Oats	Alfalfa	Sudex	Threshold
Corn.....	237	6.8	16	150	54	17	0	0
Pasture..	75	88.0	0	66	7	2	0	0
Oats.....	12	100.0	0	0	12	0	0	0
Alfalfa..	110	25.5	1	56	24	28	0	1
Sudex....	36	0.0	0	30	6	0	0	0
Totals...	470		17	302	103	47	0	1

Overall performance 26.0%

Table 54--Means and covariance matrices for crops in South Dakota on frame 1060-16491, September 21, 1972.

Corn	Means	Number 1060	Covariance Matrix			
	22.34		4.84			
	17.69		6.73	13.25		
	31.40		2.67	-0.42	33.40	
	19.38		0.37	-2.95	25.55	18.15

Pasture	Means	Number 812	Covariance Matrix			
	23.94		5.42			
	19.89		7.79	15.13		
	34.34		1.14	-1.48	29.59	
	20.85		-0.69	-3.78	18.72	13.99

Oats	Means	Number 243	Covariance Matrix			
	23.13		9.92			
	19.09		16.72	33.29		
	32.98		10.76	14.40	43.16	
	17.74		4.38	4.48	25.26	16.73

Barley	Means	Number 97	Covariance Matrix			
	24.52		5.47			
	21.46		6.25	11.15		
	30.07		5.93	5.41	25.70	
	17.51		2.65	1.54	16.87	12.53

Rye	Means	Number 16	Covariance Matrix			
	22.31		3.31			
	17.63		2.71	5.43		
	35.06		1.63	3.04	7.40	
	20.94		1.02	1.83	3.78	2.19

Alfalfa	Means	Number 303	Covariance Matrix			
	23.78		6.81			
	19.90		9.62	17.56		
	33.15		3.08	1.94	26.42	
	20.09		0.46	-1.61	16.19	12.25

Flax	Means	Number 71	Covariance Matrix			
	22.30		5.66			
	18.25		5.39	8.64		
	27.63		7.99	6.27	41.73	
	17.55		4.30	2.59	27.63	19.45

Sorghum	Means	Number 55	Covariance Matrix			
	22.51		2.79			
	17.25		3.00	6.60		
	32.15		1.44	-1.97	23.04	
	20.05		0.42	-2.38	15.76	12.74

Table 54 continued

Idle	Means	Number 19		Covariance Matrix		
	23.05		9.86			
	19.00		14.74	26.62		
	31.58		7.79	5.45	27.88	
	19.63		0.43	-3.92	14.94	11.90
Winter Fallow	Means	Number 82		Covariance Matrix		
	23.41		5.47			
	19.78		9.58	20.70		
	32.21		-1.27	-5.75	36.24	
	19.27		-2.77	-7.65	20.93	14.59

Idaho:

The test site in Idaho covers nearly four counties. The Crop Reporting District boundaries were bypassed because they did not include some areas of homogeneous types of agriculture that should have been included. Figure 4 shows the test site area.

The results are based on 42 segments in the intensive agriculture stratum in one LANDSAT frame. Two additional segments are not on this frame. The frame that contains these two segments also contains ten segments which are on the first frame. Therefore, it may be possible to use this overlapping data to calibrate from one frame to the next, or to measure the difference due to frames in the means and variance for the overlapped data. A method of using calibration or training data in one frame to adjust parameters or to classify on another frame would be valuable, since, it would increase the value of the segment data. A crop may be different over a large area because of variety, soil type, weather conditions, and state of maturity rather than technical factors associated with acquiring imagery. However, it may be possible in some areas to do signature extension and this problem should be investigated.

The data had serious banding problems. The problems seem to be most apparent in band 5, therefore, that band was left out of the first classification. Table 55 shows this first classification.

Obviously, the classification is not as good as we expected; however, by chance, one would expect only 8% correct classification for 12 crop categories. Another possible problem with the classification is that some field boundaries, sometimes, fall on adjacent points and since the pixels are partially overlapping, these border pixels may be causing some overlap of the crop categories. The grey-scale printout (Figure 14) which follows illustrates this problem.

Figure 14--Gray scale printout of a segment showing how fields are defined.

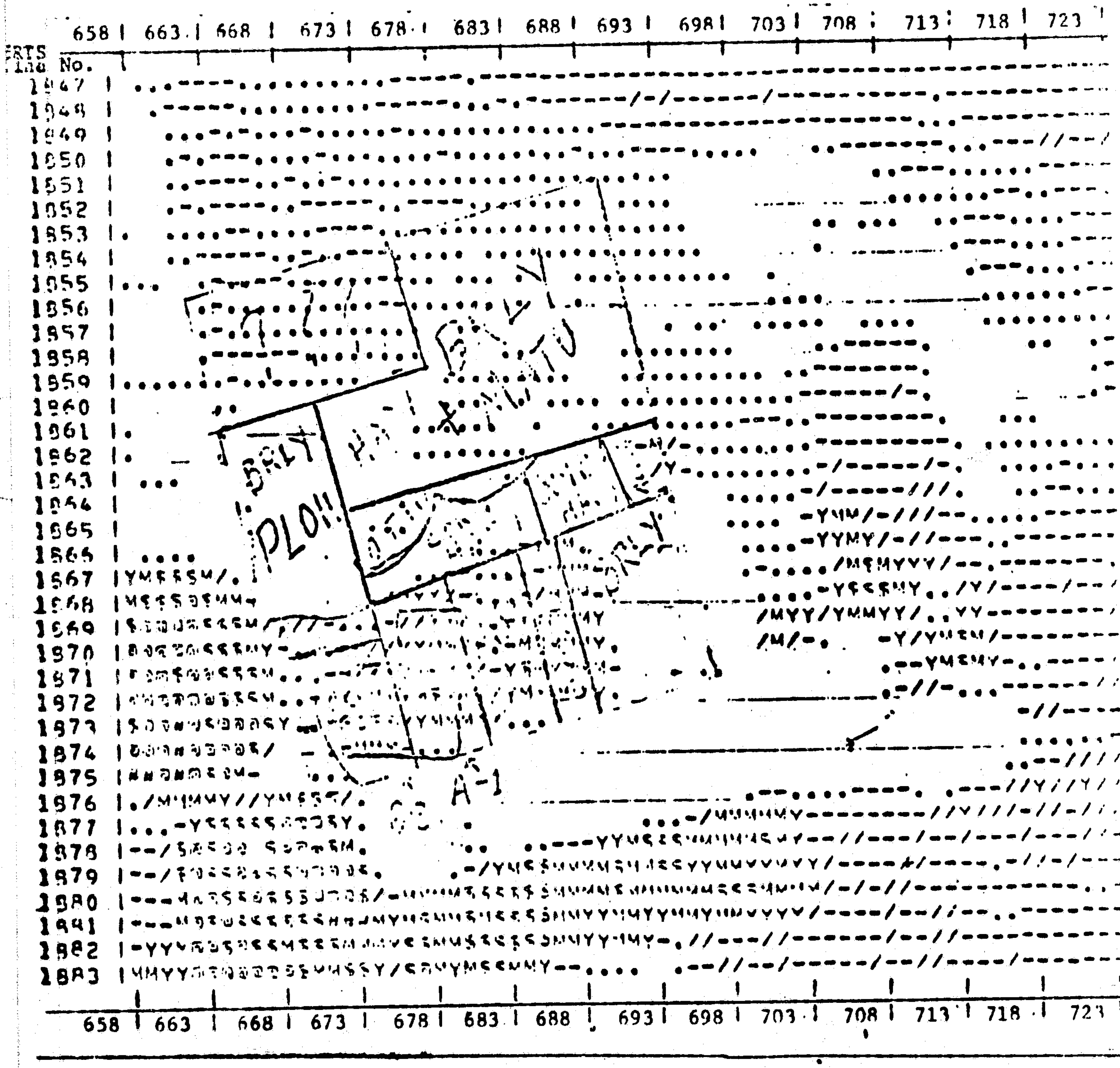


Table 55--Preliminary classification of Idaho study area data using August 1972 data bands 4, 5, and 7 and unequal prior probabilities.

	No. of samples	Percent Correct	PEAS BEANS	HARV BEANS	BRLY	ALFALFA	CORN	FALOTH	IDLE	OHAY	PASTURE	SUGBTS	POTATOES	SPWH
Peas and Beans	579	14.5	84	45	1	31	0	0	0	0	327	89	2	0
Harvested Beans	784	71.1	13	562	45	8	0	0	0	0	152	4	0	0
Barley	1019	11.5	33	271	117	27	0	2	6	0	489	64	10	0
Alfalfa	1318	17.3	57	51	2	228	0	0	6	0	527	422	25	0
Corn	542	0.0	10	21	9	119	0	0	0	0	221	161	1	0
Fallow and Other	684	0.4	14	13	3	14	0	3	33	0	575	26	3	0
Idle	206	26.7	4	10	0	1	0	1	55	0	135	0	0	0
Other Hay	11	9.1	0	0	0	0	0	0	0	0	5	3	2	0
Pasture	1484	80.7	38	25	4	78	0	2	49	1	1197	83	8	0
Sugar Beets	527	76.5	12	5	1	43	0	0	6	0	46	403	10	0
Potatoes	533	10.1	29	2	1	80	0	0	0	0	89	278	54	0
Spring Wheat	111	0.0	3	48	3	5	0	0	0	0	49	3	0	0
Total	7798		297	1054	186	634	0	8	155	1	3812	1536	115	0

Overall performance 34.7 percent

It is obvious that many groups are very similar, and therefore, misclassification is high. We will try combining several into groups based on similarity of the estimated parameters, since these initial results indicate a number of crops are not distinct.

The next classification matrix uses equal prior probabilities and is presented in Table 56. The overall classification performance is 21.8%. This points out that prior information in terms of probabilities is also important in this test area.

Since the data had serious banding problems, it was thought that perhaps this caused the extremely poor classification rates. As a result, NASA Goddard was asked to reprocess the image to remove the banding.

The image was reprocessed at considerable expense to Goddard and the classifications were again run. The results are shown in Table 57.

Table 58 is a result of combining classes after classification. It is obvious that going to fewer categories does improve the classification. However, in Idaho, where many crops are grown, the imagery must contain information that will allow users to separate the various crops. Perhaps, temporal information would improve the value of the Idaho imagery.

Results of Classification of Aerial Photography

Since aerial photography is in image form and computer techniques require digital data, it is necessary to convert the photographs to optical densities. A detailed explanation of how this is done may be found in Appendix D. The aerial photography was scanned by a Photometric Data System (now Bolen ahd Chivens) microdensitometer. This instrument records optical densities (or transmissions) of wavelengths of light corresponding to given color filters. Each time a filter is changed, however, the instrument must be recalibrated. The values recorded range from 0.00 to 4.00 in optical density. In brief, the range of values is spread between the chosen calibration point and total darkness.

Initially, the procedure for scanning segments was as follows:

An interval point within the photograph was chosen. This point was the considered lightest spot on the exposed portion of film and it was set at 0.00 on the microdensitometer scale. South Dakota, Kansas, and Idaho photography was scanned and the results brought to light problems in this technique.

Table 56--Preliminary classification of Idaho study area data using August 1972 data bands 4, 5, and 7 with equal prior probabilities.

	No. of samples	Percent Correct	PEAS BEANS	HARV BEANS	BRLY	ALFALFA	CORN	FALOTH	IDLE	OHAY	PASTURE	SUGBTS	POTATOES	SPWH
Peas and Beans	597	25.6	148	43	1	29	19	26	109	96	12	25	59	12
Harvested Beans	784	66.1	20	518	40	15	4	18	50	7	8	1	14	89
Barley	1019	9.9	62	214	101	13	19	66	112	59	71	14	78	210
Alfalfa	1318	10.7	119	47	11	141	51	26	80	172	108	115	428	20
Corn	542	1.7	28	18	11	62	9	41	36	56	17	41	198	25
Fallow and Other	684	12.1	23	7	6	5	7	83	416	23	33	5	35	41
Idle	206	70.4	9	4	0	1	1	24	145	3	4	0	0	15
Other Hay	11	72.7	1	0	0	0	2	0	0	8	0	0	0	0
Pasture	1484	8.0	105	15	17	70	14	117	606	54	119	36	148	183
Sugar Beets	527	19.9	3	3	2	18	8	0	8	142	4	105	226	8
Potatoes	533	56.8	10	2	2	25	6	1	4	105	2	72	303	1
Spring Wheat	111	19.8	8	38	0	10	4	6	4	8	5	1	5	22
Total	7798		536	909	191	309	144	408	1570	733	383	415	1494	626

Overall performance 21.8 percent

Table 57--Classification matrix of Idaho Study Area, August 1972 Imagery Using MSS Bands 4, 5, 6, and 7, with unequal Prior Probabilities.

	NO. OF Samples	PERCENT CORRECT	PEAS BEANS	HARV BEANS	BRLY	ALFALFA	CORN	FALOTH	PASTURE	SUGBTS	POTATOES	SPWH
Peas and Beans	549	40.6	223	6	9	23	4	61	123	94	5	1
Harvested Beans	813	62.6	19	509	106	11	1	38	121	6	0	2
Barley	957	75.9	68	108	248	65	9	83	331	36	6	3
Alfalfa	1314	29.8	192	30	34	391	30	32	331	250	23	1
Corn	541	8.5	42	13	20	106	46	52	186	69	8	4
Fallow and Other	77 779	37.4	28	1	7	31	3	291	412	3	3	0
Pasture	1433	64.0	107	8	24	115	8	218	917	34	2	0
Sugar Beets	386	56.0	19	1	5	60	8	1	30	216	45	1
Potatoes	395	21.8	15	0	0	115	7	0	92	80	86	0
Spring Wheat	104	3.8	12	27	24	4	1	3	23	4	2	4
Total	7271		725	703	477	921	117	779	2566	787	180	16

Overall performance 40.3 percent

Table 58--Classification matrix of Idaho with unequal prior probability groups - Table 57 collapsed into 7 groups.

Group	No. of samples	Percent Correct	Beans	Small Grains	Corn	Fallow	Pasture	Sugar Beets	Potatoes
Beans...	1362	55.6	757	118	5	99	278	100	5
Small Grains..	1061	26.3	215	279	10	86	423	40	8
Corn....	541	8.5	55	24	46	52	287	69	8
Fallow..	779	37.4	29	7	3	291	443	3	3
Pasture.	2747	73.0	337	59	38	250	1754	284	25
Sugar Beets...	386	56.0	20	6	8	1	90	216	45
Potatoes	395		15	0	7	0	207	80	86
Totals..	7271		1428	493	117	779	3482	792	180

Overall performance 47.2 percent

It was observed that each segment had a different calibration point (lightest spot), hence, there were variations in the scanning results. As a calibration point changed, grey level readings for the <u>same</u> crop in a variety of segments, were different. In fact, when the <u>same</u> segment was scanned twice using two different calibration (light) spots, the crop signatures might not appear similar.

To overcome this defect, a new calibration technique was developed. Emphasis was placed on choosing calibration points which would produce identical results in every segment. The procedure was to focus on the clear, plastic circle which appears on each section of the film as the scanner passes across the image. This circle became 0.00 in every instance. Consequently, reliable crop data was acquired since all calibration factors were now constant in the scanning process. The state of Missouri was scanned using this improved method and the results were found to be more accurate.

Once the data has been scanned, it must be labeled for crop type. Tract and field numbers were provided by the use of a coordinate system and this data was then merged with the ground observation data. This provided crop labels. This labeled data can then be used for both computer training and testing information.

The classification procedure is explained at the beginning of this section. However, since the calibration was done using local calibration points, the classifier training and the computer evaluations were performed in two ways.

For example:

1. All data was pooled and used for both the training and testing.

2. All data in each segment was used for both training and testing one at a time.

The results were then pooled (matrix sum). The prior probabilities in each case were proportional to the training data and since this training data was used to test, it too, was proportional to the data being classified.

In the instance of the pooled training, the prior probabilities were the same for each segment. When interpreting the local training, the prior probabilities were different for each segment and depended on the data in each. For the local training, all conditions were optimal which would mean that the classification accuracy is maximal.

As a preliminary check on the effect that the different calibration points had on the data, a cluster analysis on all data was run. The means for each field were computed by segment and crop. These means were clustered using a program written by C.T. Zahn of Stanford University. 1/ The fact that the means clustered by segments rather than by crops was additional proof of the problems which had arisen because of calibration differences.

Figure 15 provides an overall state by state comparison of classification accuracy. Figure 16-19 summarize the percent correct classification for major crops in each state. These figures compare both methods of training on the same data sets; the difference lies in the results of local versus pooled training data.

Tables 59-65 give the classification matrix for both methods of training. When local training data was used for training, a classification matrix was available for each segment. These segment classification matrices were summed to obtain the final classification matrices in this report.

1/ C.T. "Zahn, Graph-Theoretical Methods for Detecting and Describing Gestalt Clusters," IEEE Transactions on Computers, Vol. C-XX, No. 1, 1971.

Figure 15--Comparison of overall percent classification by states, 1972. (///Slashes indication global classification).

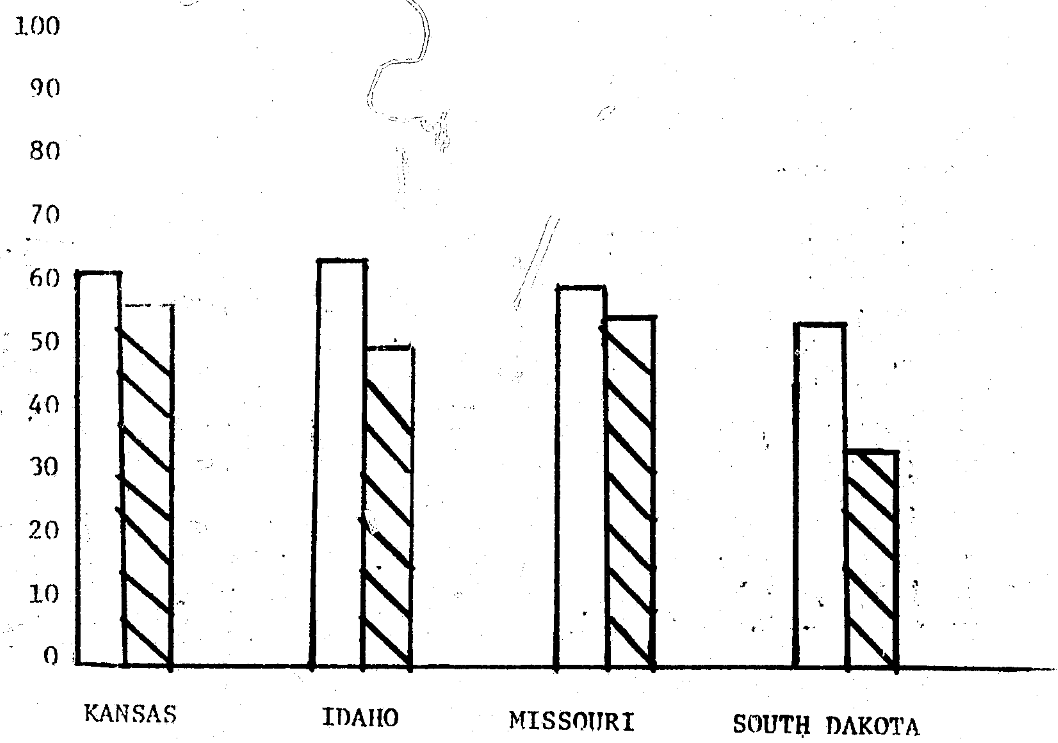


Figure 16--Comparison of classification methods by crop, Kansas August 18, 1972. (///Slashes indicate global classification.

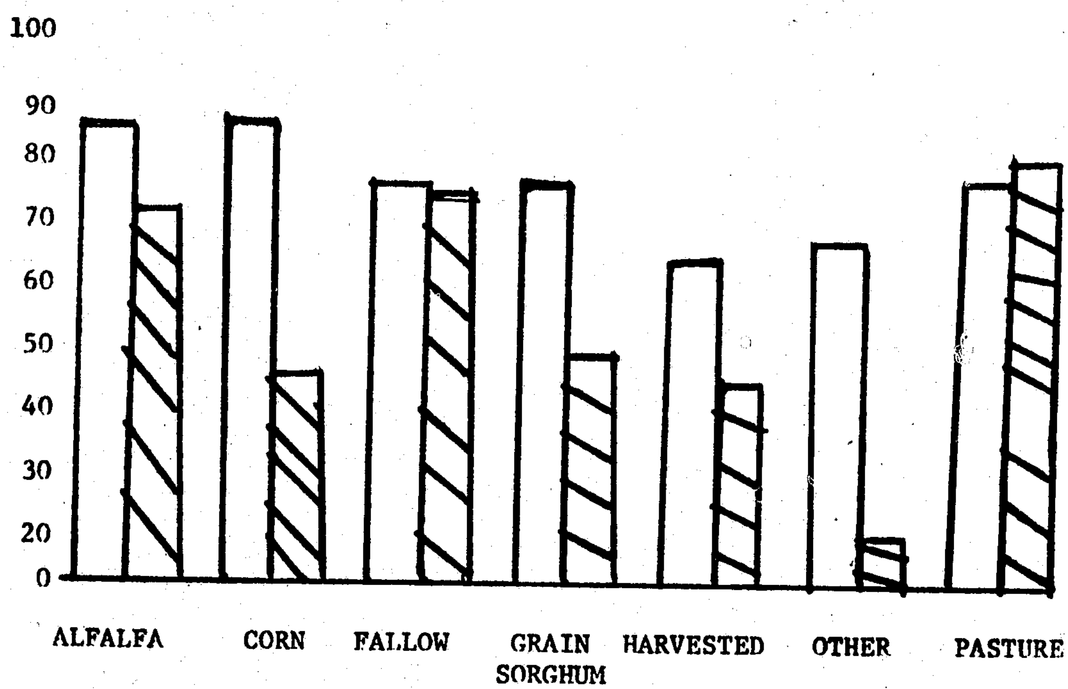


Figure 17--Comparison of classification methods by crop, Missouri, August 29, 1972. (/// Slashes indicate global classification).

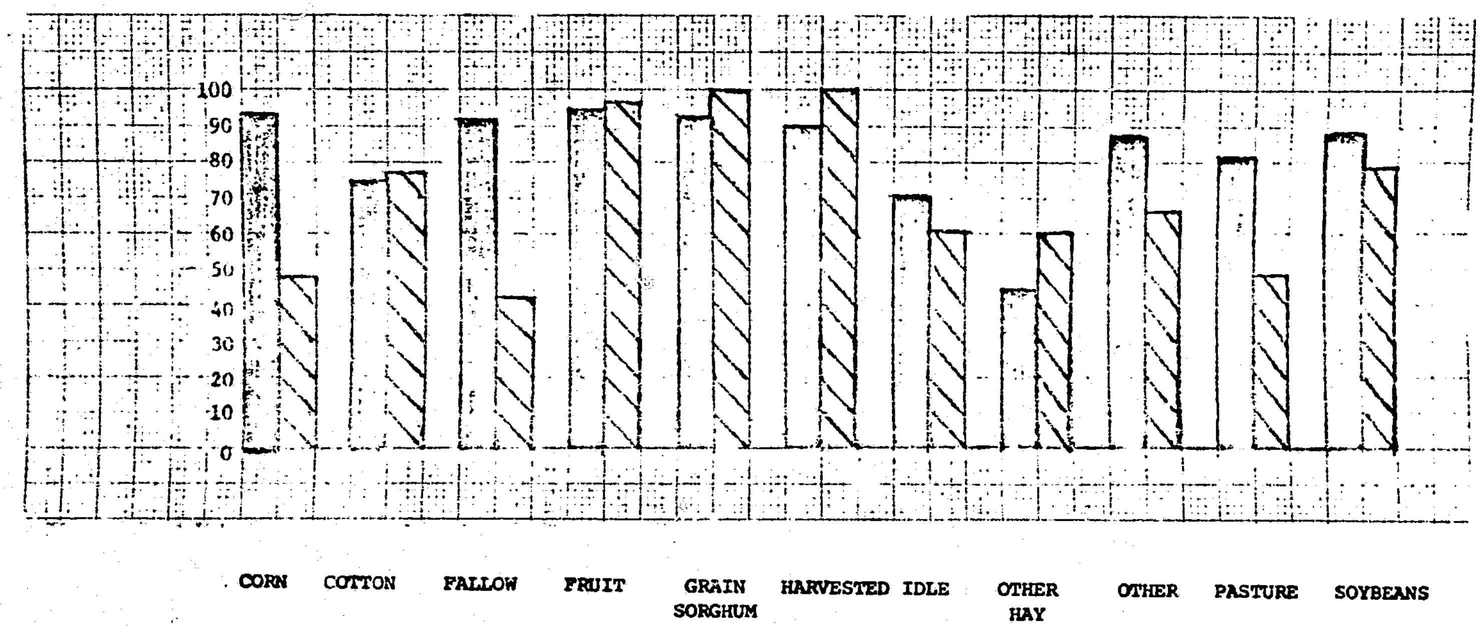


Figure 18--Comparison of classification methods by crop, South Dakota, September 23, 1972. (/// Slashes indicate global classification).

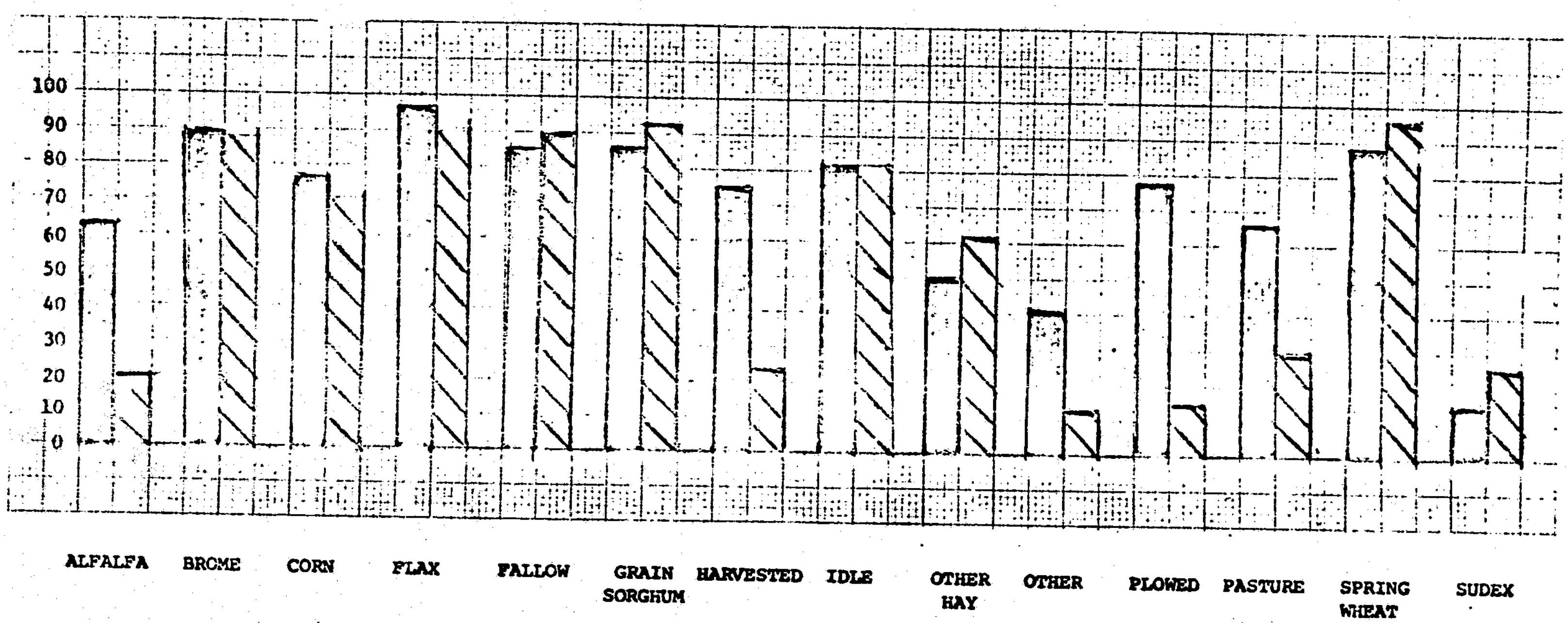


Figure 19--Comparison of classification methods by crop, Idaho, August 12, 1972.
(/// Slashes indicate global classification)

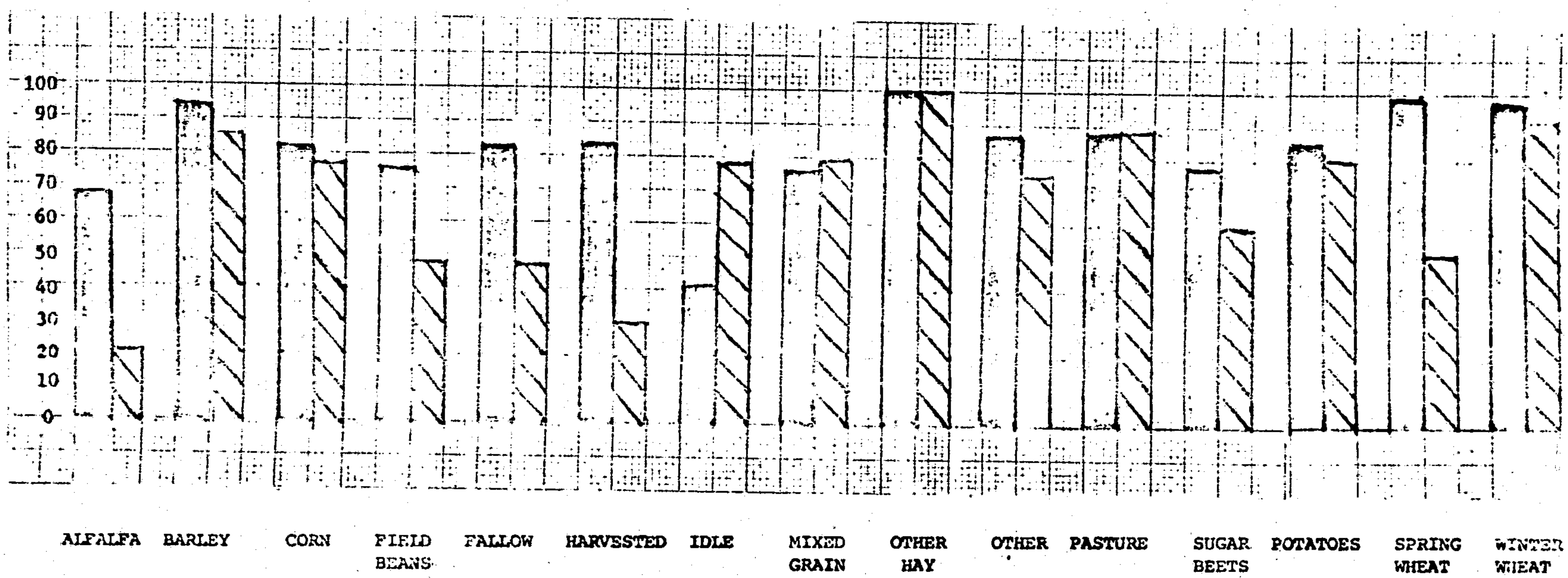


Figure 15-19 indicate that the training by segment (local) gave higher classification percentages. The differences in classification percentages can be attributed to three sources:

1. difference in calibration of data when scanned by a microdensitometer.

2. differences in the number of crop classes and the prior probabilities of each crop class.

3. differences in the variability of a local versus pooled data set.

Interpretation of Figure 15 is quite easy. Kansas (with some calibration effect and only seven crop classes) was not greatly affected by the calibration effects. However, South Dakota, with calibration differences and many crop classes was drastically effected. In Missouri, differences were slight between classification results comparing local training versus pooled training.

Table 59--Classification of flightlines 3 and 10, by segment, using quadratic discriminant functions on all eight spectral variables, Kansas aircraft data, September 1972.

Crop	Percent Correct	ALFA	CORN	FLOW	GSOR	HARV	OTHR	PSTR
ALFA	94.2	1238	0	0	21	8	36	11
CORN	93.9	0	247	0	2	3	11	0
FLOW	80.9	0	0	8383	398	1432	47	100
GSOR	82.2	4	1	51	3525	181	26	498
HARV	66.0	0	0	1031	489	4797	29	922
OTHR	70.1	37	4	13	17	14	312	48
PSTR	83.2	18	0	697	2927	1677	70	26,644
OTHERS		525	129	3186	1606	3290	829	7,166

Overall performance 80.7 percent

Table 60--Classification of flightlines 3 and 10, on all eight spectral variables, Kansas aircraft data, September 1972.

Crop	Percent Correct	ALFA	CORN	FLOW	GSOR	HARV	OTHR	PSTR
ALFA	76.0	999	14	0	107	4	20	170
CORN	41.8	170	147	0	19	0	2	14
FLOW	78.6	8	52	8141	352	708	9	1090
GSOR	46.5	102	107	81	2063	485	62	1540
HARV	41.6	16	42	1816	403	3025	26	1940
OTHR	10.6	43	5	21	126	37	47	166
PSTR	87.6	3	241	222	916	2400	180	28,071
OTHERS		335	596	2023	1402	1462	660	10,728

Overall performance 75.6 percent

Table 61--Classification of flightlines 5 and 6, by segment, using all eight spectral variables, Idaho, September 1972.

Crop	Percent Correct	Number of samples classified into														
		ALFA	BRLY	CORN	FLDB	FLOW	HARV	IDLE	MGRN	OHAY	OTHR	PSTR	SBTS	SPDS	SPWH	WNWH
ALFA	68.3	808	0	13	5	1	7	0	2	0	4	218	84	16	25	0
BRLY	94.6	2	866	0	0	5	5	8	2	0	2	5	5	3	0	12
CORN	82.6	17	4	1171	67	19	10	2	1	1	13	40	67	1	2	0
FLDB	76.3	0	0	7	371	0	0	0	0	0	1	2	40	65	0	0
FLOW	83.1	2	11	98	0	1069	74		1		4	4	16	8	0	0
HARV	83.9	1	9	0	0	26	620	1	0	1	35	27	18	1	0	0
IDLE	41.0	0	15	9	5		5	34		0	0	0	15		0	0
MGRN	76.0	6	7	1	0	6	0		76		1		2		1	
OHAY	1.00	0	0	0	0		0	0		9	0	0	0		0	0
OTHR	87.0	1	2	1	0	6	4		4	1	631	15	51	2	7	0
PSTR	87.9	71	5	35	4	30	59	0	0	0	32	2950	145	26	0	0
SBTS	78.5	0	5	60	136	17	22	7	0	0	6	5	955	3	0	0
SPDS	85.2	1	76	11	2	8	50		0		4	18	1	981	0	
SPWH	99.0	1	0	0	0	0	9	0	0	0	2	0	0	0	410	1
WNWH	97.3	0	0	0	0	0	0	0	0	0	0	0	0	0	1	36
Others		813	1016	703	623	14166	1419	129	302	36	5104	788	788	454	603	17

Overall % = 83.76

Table 62---Classification of flightlines 5 and 6, using eight spectral variables, Idaho, September 1972.

Crop	Percent Correct	Number of samples classified into														
		ALFA	BRLY	CORN	FLDB	FLOW	HARV	IDLE	MGRN	OHAY	OTHR	PSTR	SBTS	SPDS	SPWH	WNWH
ALFA	21.39	253	8	117	1	67	0	4	1	6	29	652	25	11	1	8
BRLY	86.12	1	788	0	0	5	9	35	56	0	7	5	0	0	3	6
CORN	77.74	15	6	1100	19	5	10	21	0	1	7	197	32	2	0	0
FLDB	48.35	0	1	118	235	0	11	0	0	0	0	39	55	27	0	0
FLOW	47.16	2	83	222	21	607	33	4	15	3	199	92	1	5	0	0
HARV	30.72	2	98	59	8	73	227	15	0	39	37	145	1	34	1	0
IDLE	78.31	0	0	12	0	0	0	65	0	0	4	0	2	0	0	0
MGRN	79.00	3	9	0	0	0	0	0	79	0	0	8	0	0	0	1
OHAY	100.00	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0
OTHR	75.45	0	20	3	0	16	4	6	3	2	547	114	5	1	4	0
PSTR	88.35	61	6	66	0	41	37	0	2	1	123	2966	1	53	0	0
SBTS	59.87	50	1	135	66	1	8	55	0	0	3	168	728	1	0	0
SPDS	79.95	2	89	13	71	5	8	0	0	0	2	25	5	921	11	0
SPWH	51.93	0	78	0	0	20	2	0	0	5	5	35	0	0	215	54
WNWH	91.89	0	2	0	0	1	0	0	0	0	0	0	0	0	0	45
Others		312	1513	844	453	599	919	267	640	73	4904	4918	269	316	174	52

Overall % = 66.89

Table 63--Classification of flightlines 2 and 5 by segment, using eight spectral variables, South Dakota, September 1972.

Crop	Percent Correct	Number of samples classified into													
		ALFA	BRME	CORN	FLAX	FLOW	GSOR	HARV	IDLE	OHAY	OTHR	PLOW	PSTR	SPWH	SUDX
ALFA	63.57	719	53	238	0	0	0	24	0	6	14	3	66	0	8
BRME	89.20	15	760	35	0	0	0	28	0	0	0	0	14	0	0
CORN	77.35	313	39	2681	0	1	0	81	10	23	9	43	248	1	18
FLAX	97.57	0	0	0	201	0	0	5	0	0	0	0	0	0	0
FLOW	84.87	1	0	0	0	101	0	3	0	0	0	14	0	0	0
GSOR	85.71	0	0	0	0	0	12	0	0	0	0	2	0	0	0
HARV	75.51	37	117	141	0	14	3	1421	35	6	2	48	43	23	2
IDLE	82.73	3	0	10	0	0	0	5	206	0	0	0	24	0	1
OHAY	50.57	2	0	5	1	0	0	29	0	44	2	4	0	0	0
OTHR	41.35	37	16	81	2	6	0	24	3	19	208	13	94	0	0
PLOW	78.38	8	0	12	0	77	2	85	0	4	11	747	7	0	0
PSTR	66.55	266	141	421	0	0	1	151	31	23	38	20	2246	7	12
SPWH	88.46	0	0	2	0	0	0	1	0	0	0	0	0	23	0
SUDX	15.42	135	0	21	0	0	0	3	0	3	2	15	24	0	37
Others		964	213	2305	117	276	7	1535	195	160	2123	1220	1658	14	77

Overall % correct = 71.73

Table 64 --Classification of flightlines 2 and 5, using eight spectral variables, South Dakota, September 1972.

Crop	Percent Correct	Number of samples classified into													
		ALFA	BRME	CORN	FLAX	FLOW	GSOR	HARV	IDLE	OHAY	OTHR	PLOW	PSTR	SPWH	SUDX
ALFA	16.45	186	196	412	0	10	0	24	13	6	11	3	208	0	62
BRME	88.03	28	750	7	0	0	0	19	1	3	14	3	27	0	0
CORN	71.18	138	371	2467	2	59	0	79	55	44	10	9	197	12	23
FLAX	90.78	0	4	2	187	0	0	3	0	0	2	2	6	0	0
FLOW	89.92	0	6	1	0	107	0	1	0	0	0	4	0	00	0
GSOR	92.86	0	1	0	0	0	13	0	0	0	0	0	0	0	0
HARV	24.05	68	369	381	53	58	0	455	48	35	15	75	258	42	26
IDLE	82.33	0	0	31	0	4	0	0	205	0	0	41	1	0	4
OHAY	62.07	1	2	17	0	1	0	0	0	54	2	1	7	0	1
OTHR	13.52	43	34	63	4	25	0	11	22	39	68	111	77	0	6
PLOW	15.63	15	287	41	4	296	6	32	56	21	26	149	20	0	0
PSTR	30.58	126	456	1320	24	11	1	77	97	74	60	27	1032	25	45
SPWH	96.15	0	1	0	0	0	0	0	0	0	0	0	0	25	0
SUDX	25.83	20	1	26	0	5	0	0	16	0	5	3	2	9	62
Others		666	897	3610	155	563	8	307	508	354	1386	407	1763	34	206

Overall % correct = 44.54

Table 65 ---Classification of flightlines 2 and 8, by segment, using eight spectral variables, Missouri, September 1972.

Crop	Percent Correct	Number of samples classified into										
		CORN	COTTON	FLOW	FRUIT	GSOR	HARV	IDLE	OHAY	OTHR	PSTR	SOYB
CORN	93.69	104	0	1	0	0	4	0	0	0	0	2
COTTON	75.62	6	273	0	0	1	7	25	0	2	14	33
FLOW	90.91	0	0	30	0	1	0	0	0	0	1	1
FRUIT	95.24	0	0	0	41	0	0	0	0	0	1	1
GSOR	92.86	1	0	0	0	13	0	0	0	0	0	0
HARV	89.06	1	0	0	0	0	57	5	0	0	0	1
IDLE	71.60	0	2	1	0	0	11	232	0	28	42	8
OHAY	44.28	19	0	8	0	0	0	0	182	-2	33	167
OTHR	87.47	3	2	1	0	1	0	3	0	342	10	29
PSTR	81.46	10	0	12	6	6	0	2	4	6	268	15
SOYB	88.16	6	51	5	3	5	11	3	5	26	44	1184
Others		154	200	59	68	47	61	228	217	695	485	1469

Overall % correct = 79.61

The microdensitometer can scan a photograph and obtain either density values or transmission values or both. Transmission values are functionally related to density readings by the following equation:

$$\text{Density} = \log \left(\frac{1}{\text{transmission}}\right)$$

Theoretically, all information would be contained in either mode and neither would add anything new to the data. However, in practice, this does not hold true for two reasons:

1. The scanner seems to saturate. The results of this saturation affects density measurements. It becomes difficult to differentiate between brown wheat, brown hay, harvested grains, and bare soil. When the sensor saturates, it gives similar readings even though the colors are quite different. In the use of the transmission values, correct classification is increased but lacks complete reliability.

2. An additional reason for the one mode preference is concerned with the computer operation. The computer algorithm assumes that the data is multivariate normal with equal covariance matrices. Certainly if the data was multivariate normal in the measurement space using density values, it would <u>not</u> be multivariate normal after it had been transformed by a reciprocal of the log transformation. Obviously, they could not both be distributed as multivariate normal data. Thus, it is imperative to investigate the effects of variables on classification groups.

A stepwise discriminant analysis was performed on the training data in South Dakota. The procedure used was program BMD07M of the BMD statistical package. 1/ This program performs a stepwise linear discriminant analysis with proportional group priors on the training data. Variables are entered or deleted from the discriminanting set based upon an F-test of group differences for a particular variable. The variable that has the largest pairwise group F-value is the first variable entered in the discriminant set. This procedure was executed on all eight variables and then upon the subsets of variables corresponding to transmission and density scanning mode respectively. Some of the original nine groups were pooled to eight, six, and, then, four groups and the stepwise classification was performed on the merged groups. The mergers are as follows:

1/ Biomedical Computer Programs. W. H. Dixon, Editor. Berkeley, California; University of California Press, 1973.

a) For the eight groups; "Harvested Grains" and "Harvested Row" were merged to form the classification group "Harvested.

b) For the six groups: Hay, Pasture, and Fallow were merged (a grasses type of cover), in addition to the above merger.

c) For the four groups: Corn, and Soybeans formed a group, Wheat, Pasture, and Harvested Grains formed a group, Plowed, Fallow, and Harvested Row formed a group. This merger in the above groups was a result of a cluster analysis performed on the group means.

The option was specified for the inclusion with no deletion of variables at each step in performing the stepwise discriminant analysis. Thus, supposedly one is adding more information (in the form of more variables) at each step of the stepwise discriminant analysis. The results are astonishing as we can see in Figures 20-27.

Note the following:

(1) Overall percent correct classification increases only slightly, when two variables are in the discriminanting set, irrespective of what variables are used and what the classification groups are. The contention of C. R. RAO 1/ that more variables do not necessarily mean more information and hence more discriminanting power is supported by the data.

(2) The addition of a particular variable influences one classification group greatly. For example, in Figure 20, the variable TGREEN (transmission in GREEN) has a great effect on the classification accuracy of wheat when combined with DRED (density red), and DCLEAR. However, once TRED, has entered, TGREEN's affect is diminished by the confusion variable TRED.

(3) All discriminating information is not contained in one scanning mode (four variables). For example, compare the classification curves for the group FALLOW in Figures 24 and 25 respectively. Fallow was correctly classified about 65 percent of the time when scanned in density mode but had zero recognition in the transmission mode.

(4) The overall classification accuracy increases as the number of groups is decreased, See Figures 30 and 31.

1/ Covariance Adjustment and Related Problems in Multivariate Analysis by C. R. RAO in Multivariate Analysis, editor P. R. Krishnaiah, Academic Press, 1966.

(5) The overall classification accuracy is greater for variables measured in density units, and the use of the variables measured in transmission does not improve the overall classification when only four variables are considered. This can be seen in Figures 20, 23, 26, 29.

(6) This analysis leads one to conclude that if there is interest in only one or perhaps several crop groups that a hierarchial (or layered) classifier might be the best approach to crop identification. At each stage of the hierarchy, a feature selection would be performed to maximize the particular crop or crops of interest.

A single stage classifier with all variables used clearly would not do well on the major crop Wheat in Figure 20, as evidenced by the last stage of the stepwise discriminant analysis.

FIGURE 20

Stepwise discriminant analysis, classification into nine groups, density and transmission mode, South Dakota, 1972.

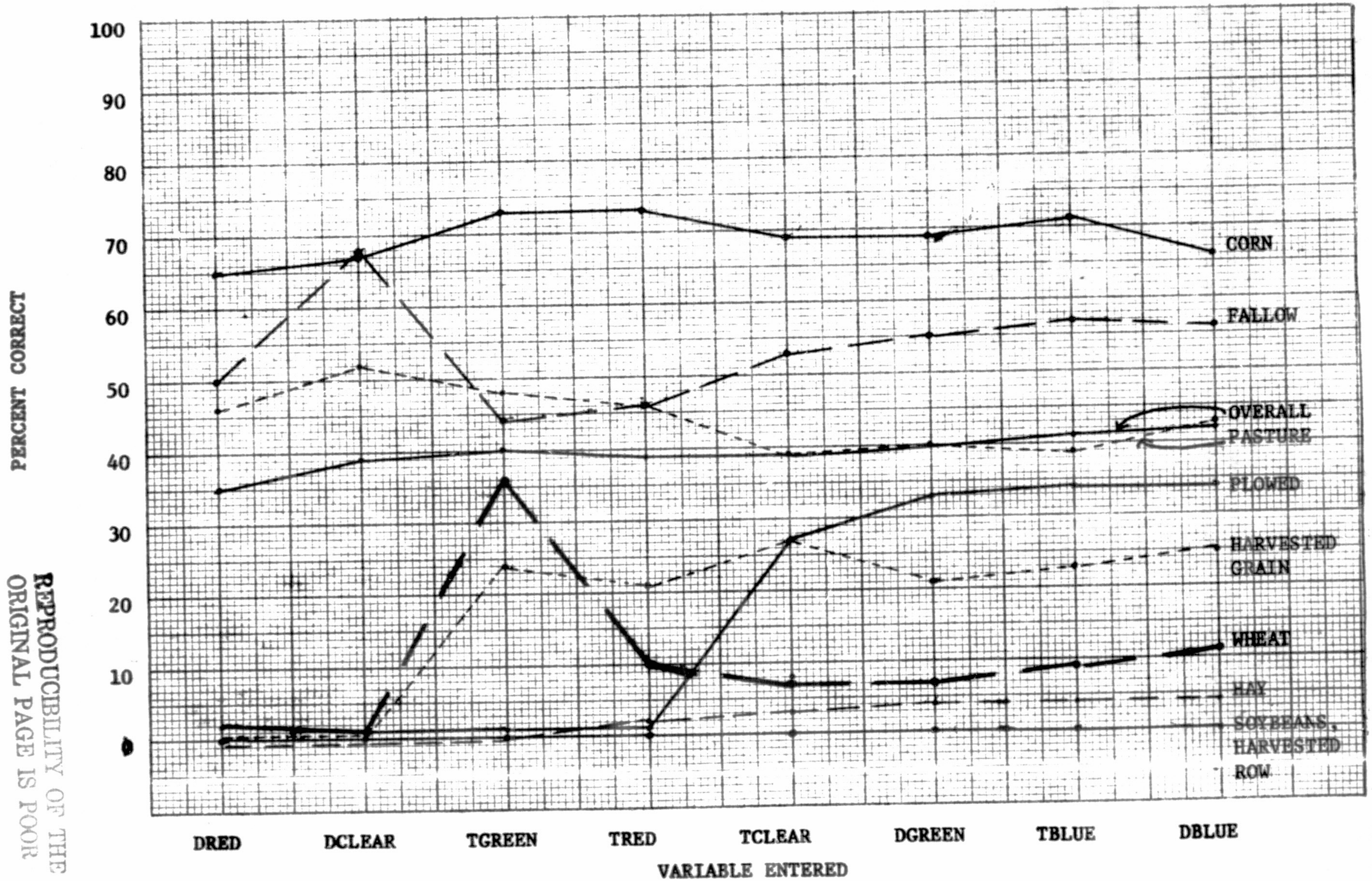


FIGURE 21

Stepwise discriminant analysis, classification into nine groups, transmission scanning mode, South Dakota, 1972.

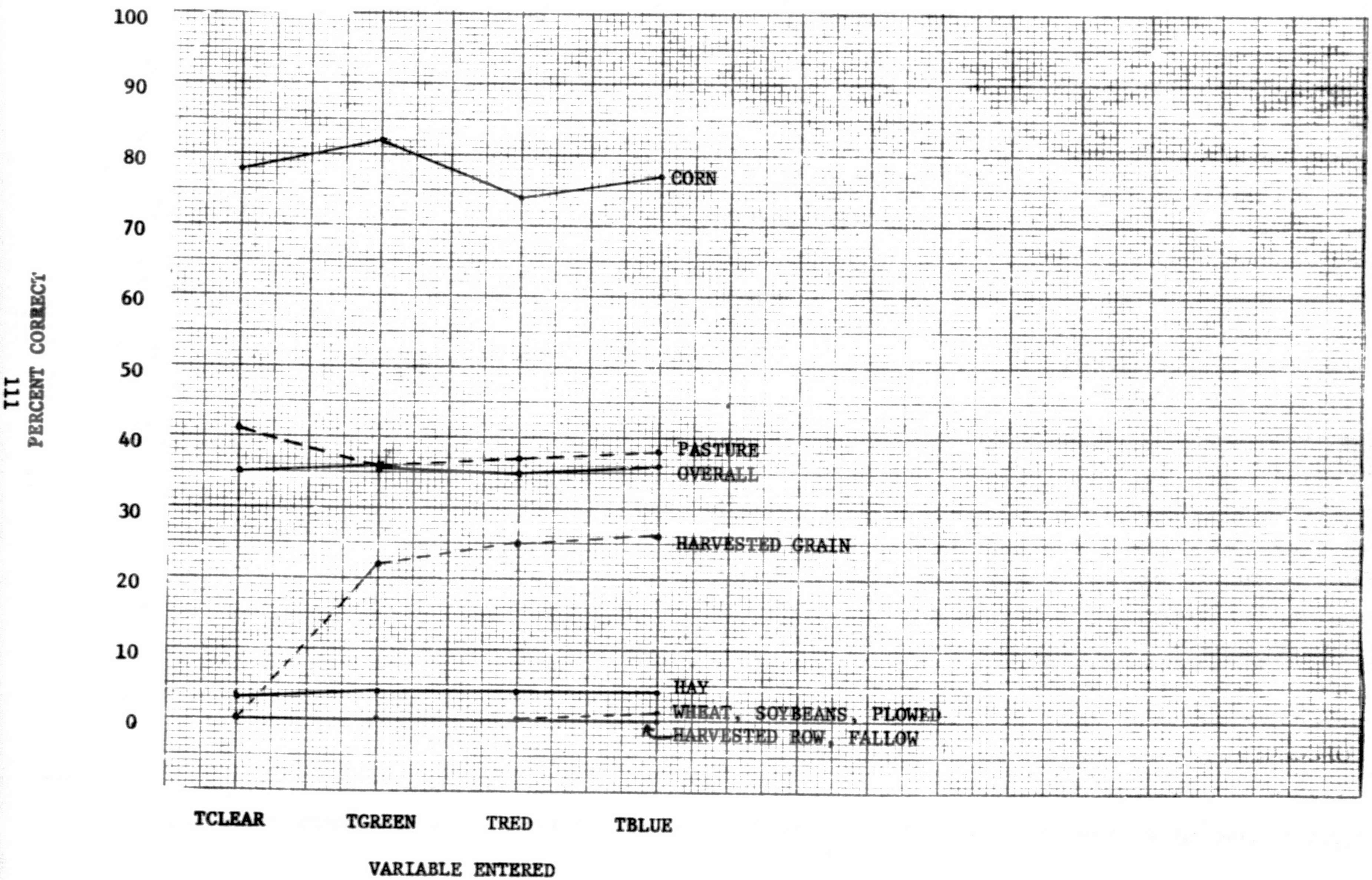


FIGURE 22

Stepwise discriminant analysis, classification into nine groups, density scanning mode, South Dakota, 1972.

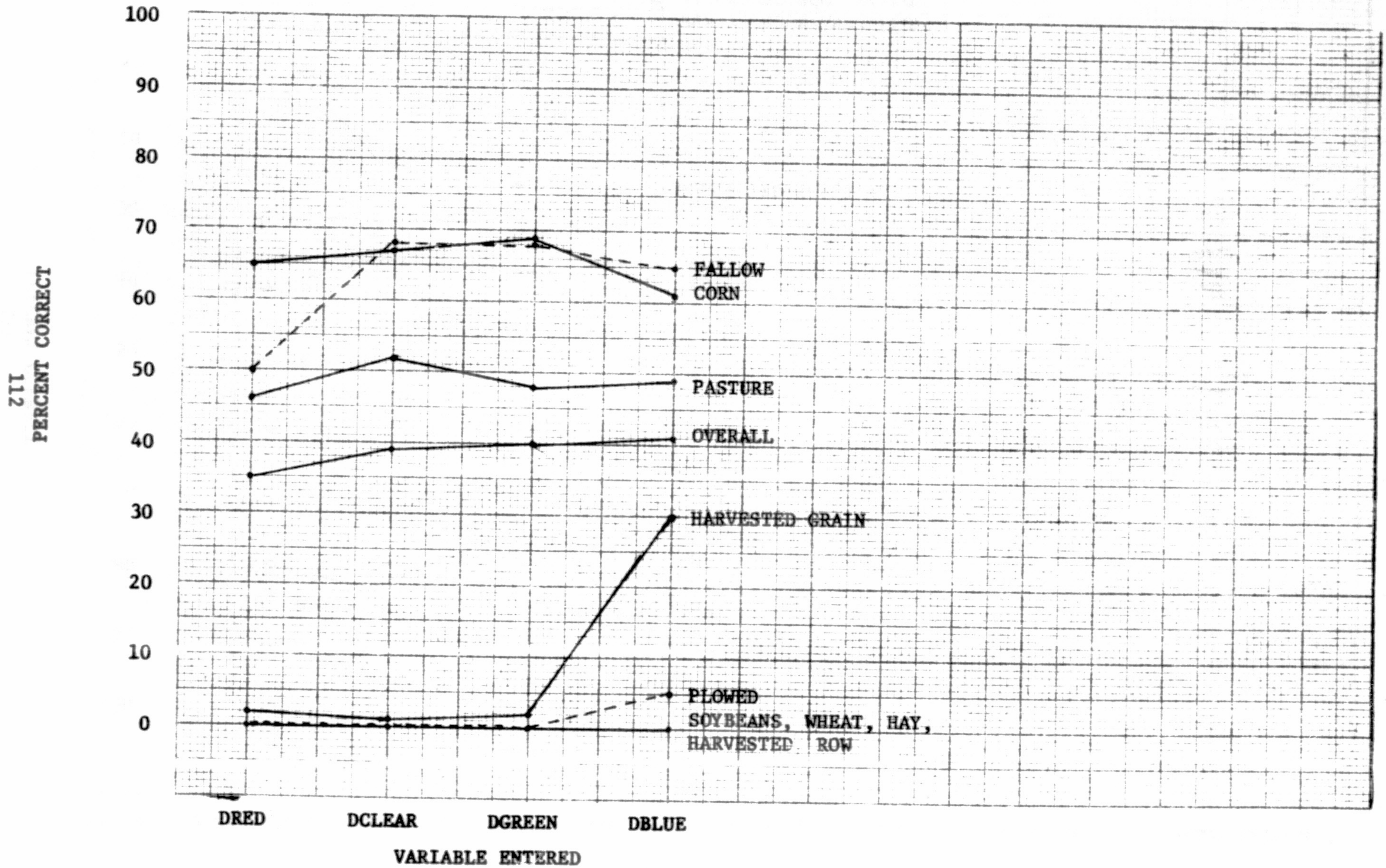


FIGURE 23

Stepwise discriminant analysis, classification into eight groups, all variables, South Dakota, 1972

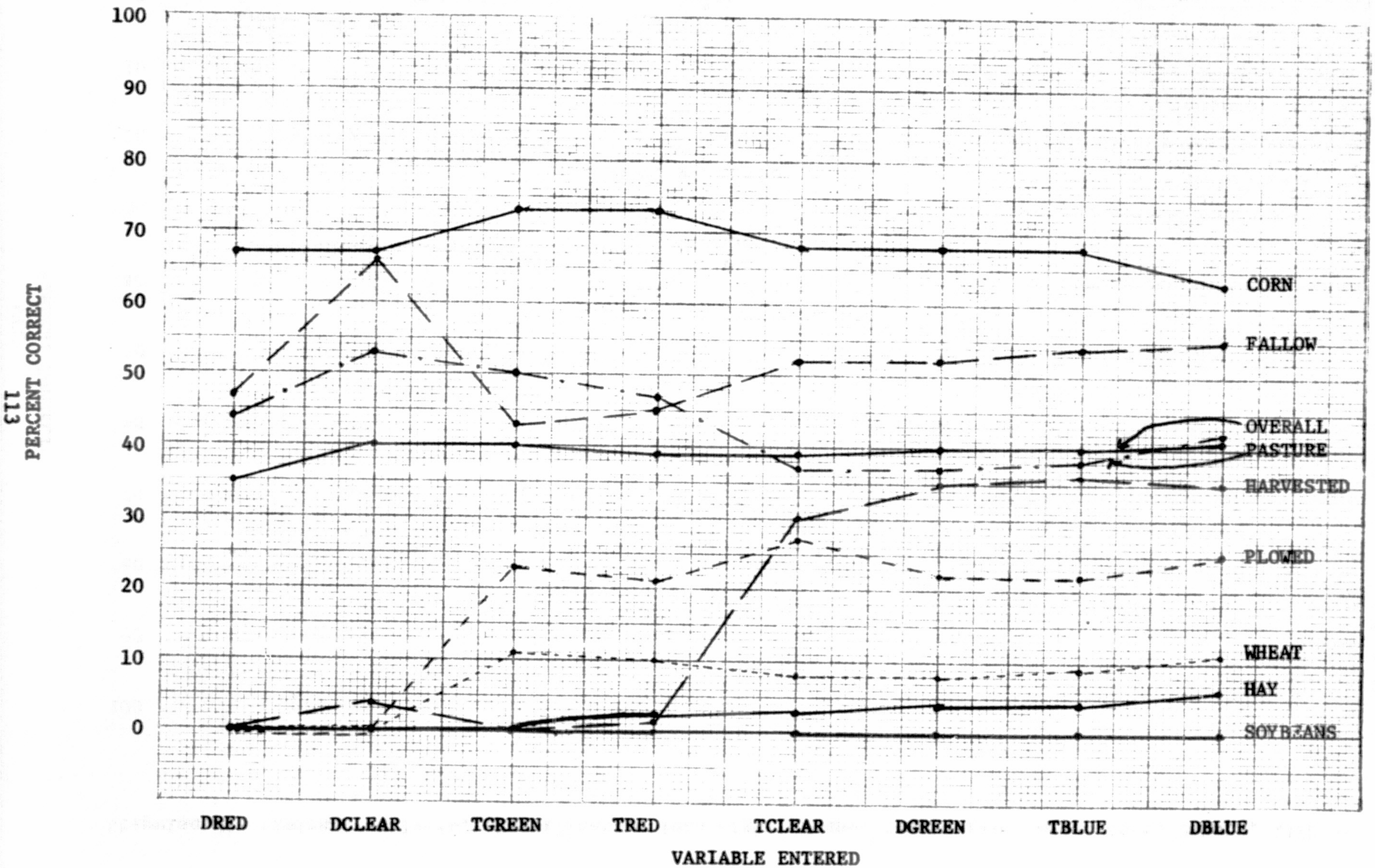


FIGURE 24

Stepwise discriminant analysis, classification into eight groups, transmission mode, South Dakota, 1972.

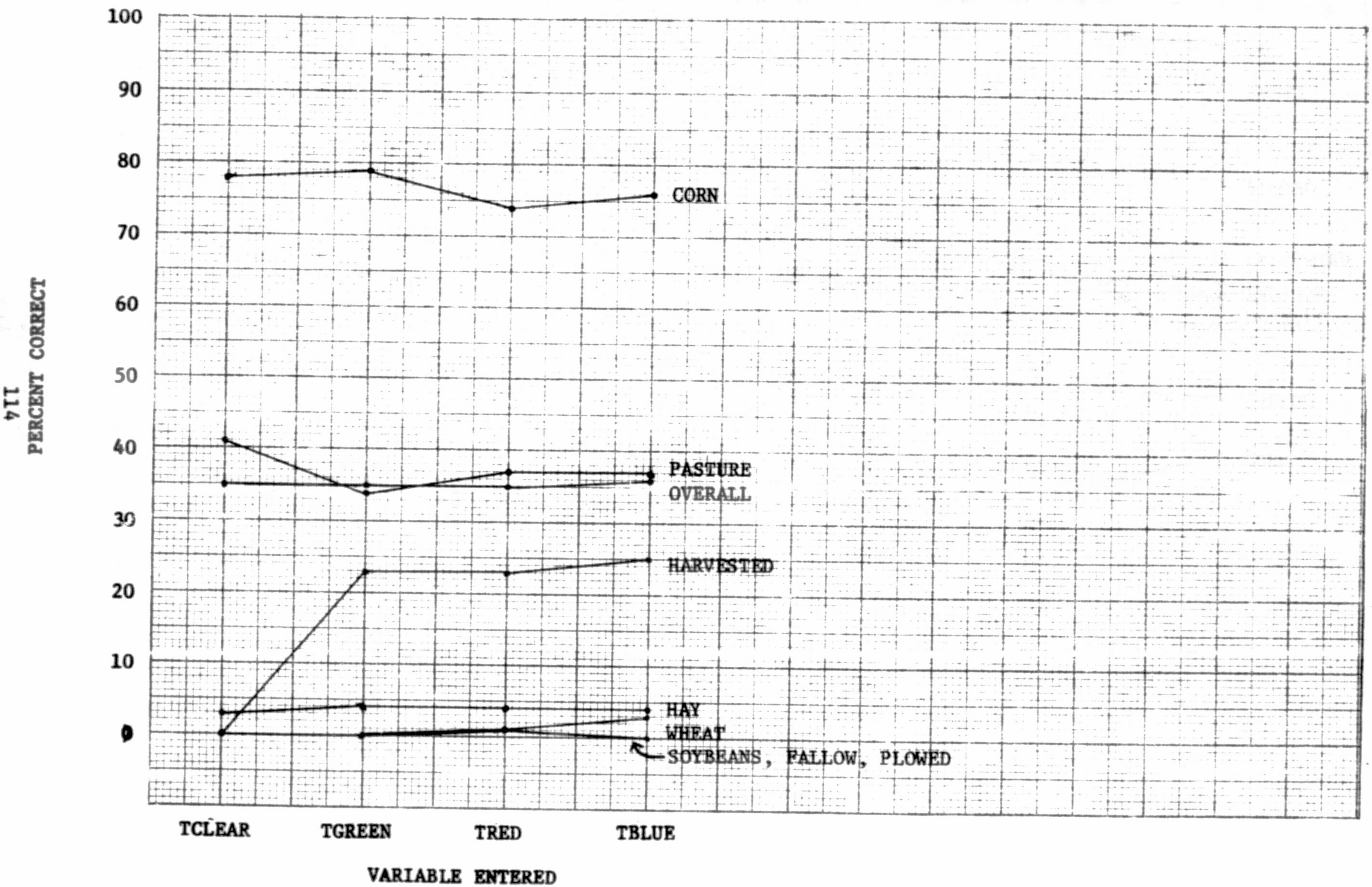


FIGURE 25

Stepwise discriminant analysis, classification into eight groups, density mode, South Dakota, 1972.

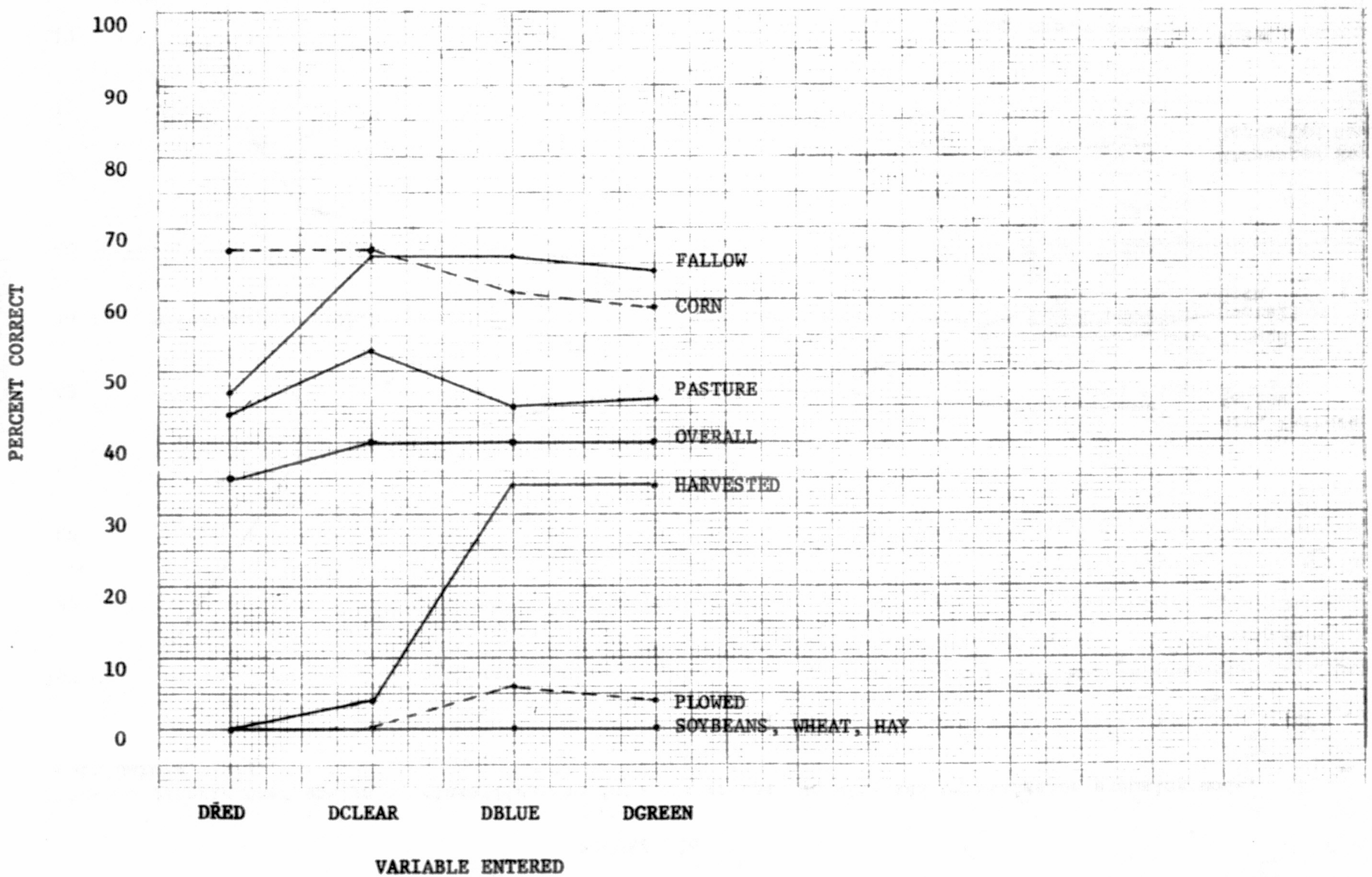


FIGURE 26

Stepwise discriminant analysis, classification into six groups, density and transmission scanning mode, South Dakota, 1972.

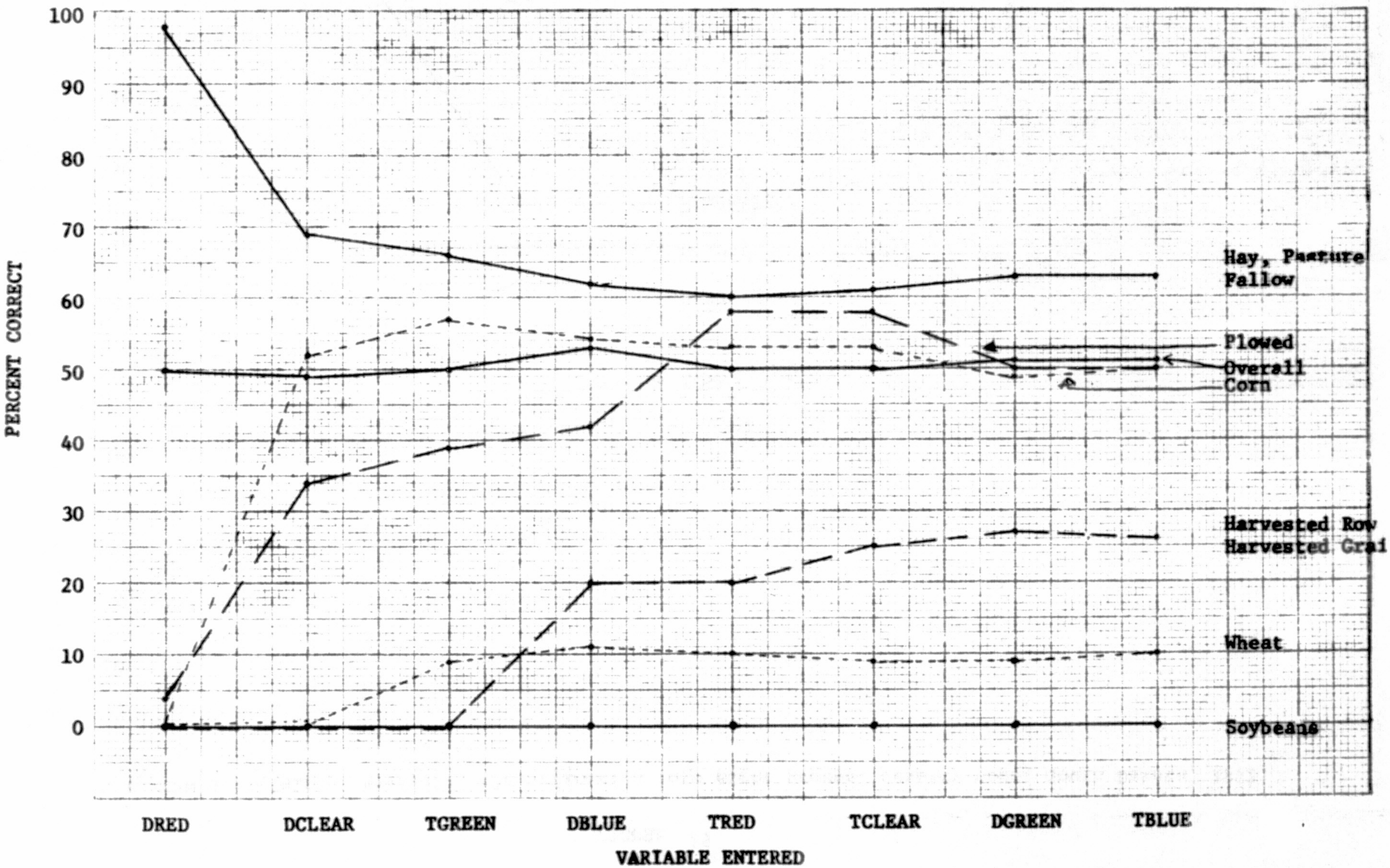


FIGURE 27

Stepwise discriminant analysis, classification into six groups, transmission scanning mode, South Dakota, 1972.

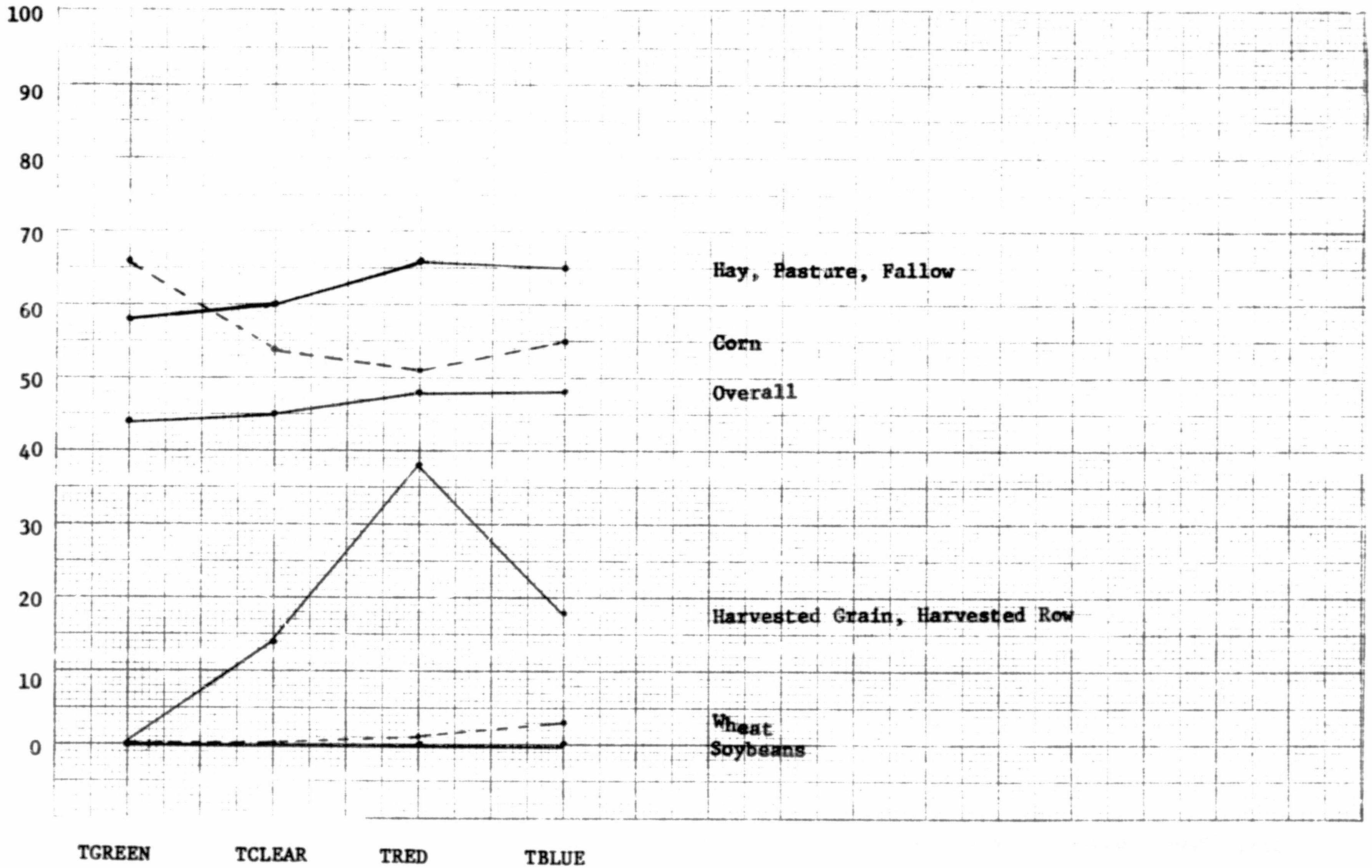


FIGURE 28

Stepwise discriminant analysis, classification into six groups, density scanning mode, South Dakota, 1972.

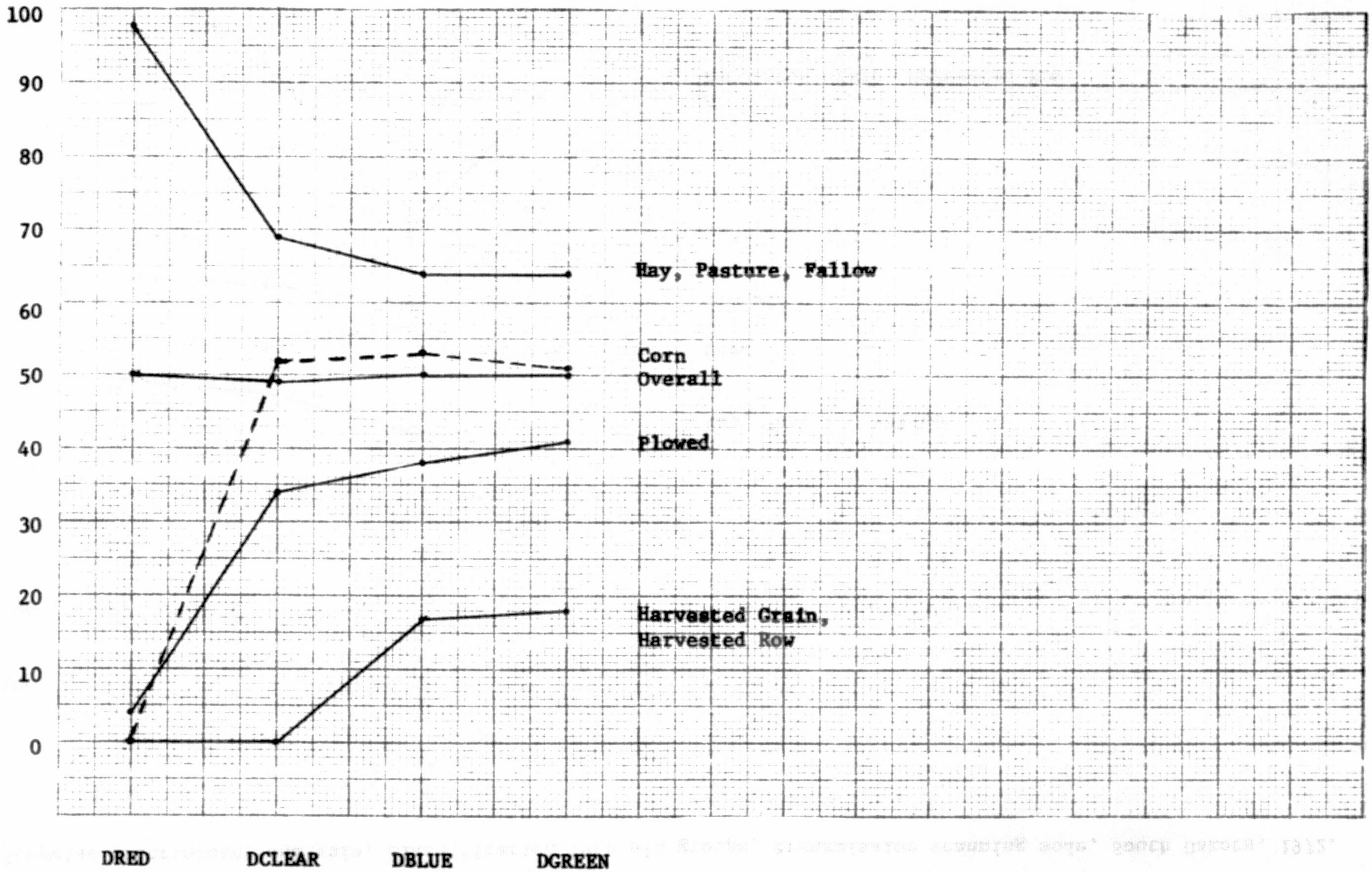


FIGURE 29

Stepwise discriminant analysis, classification into four groups, density and transmission scanning mode, South Dakota, 1972.

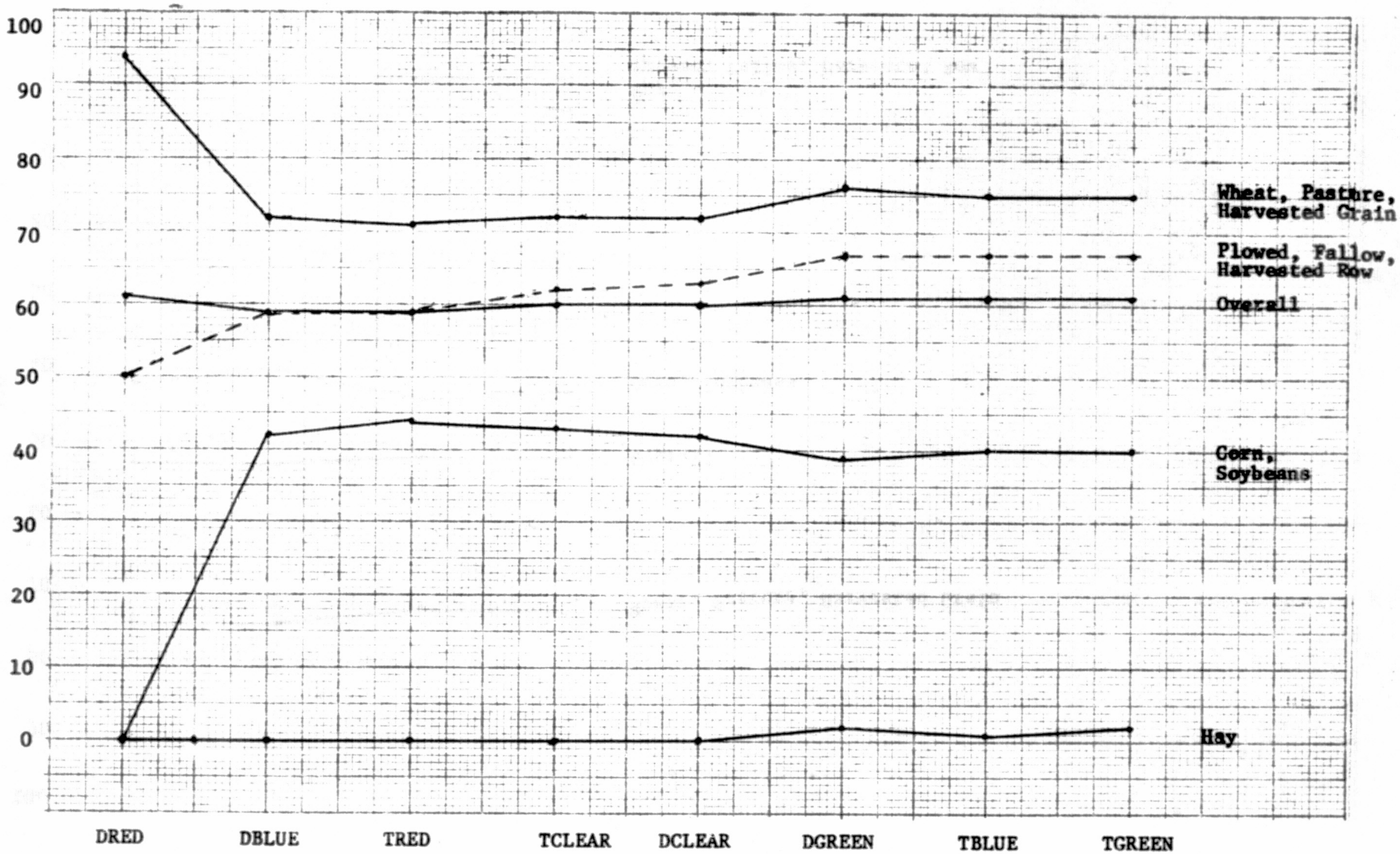


FIGURE 30

Stepwise discriminant analysis, classification into four groups, transmission scanning mode, South Dakota, 1972.

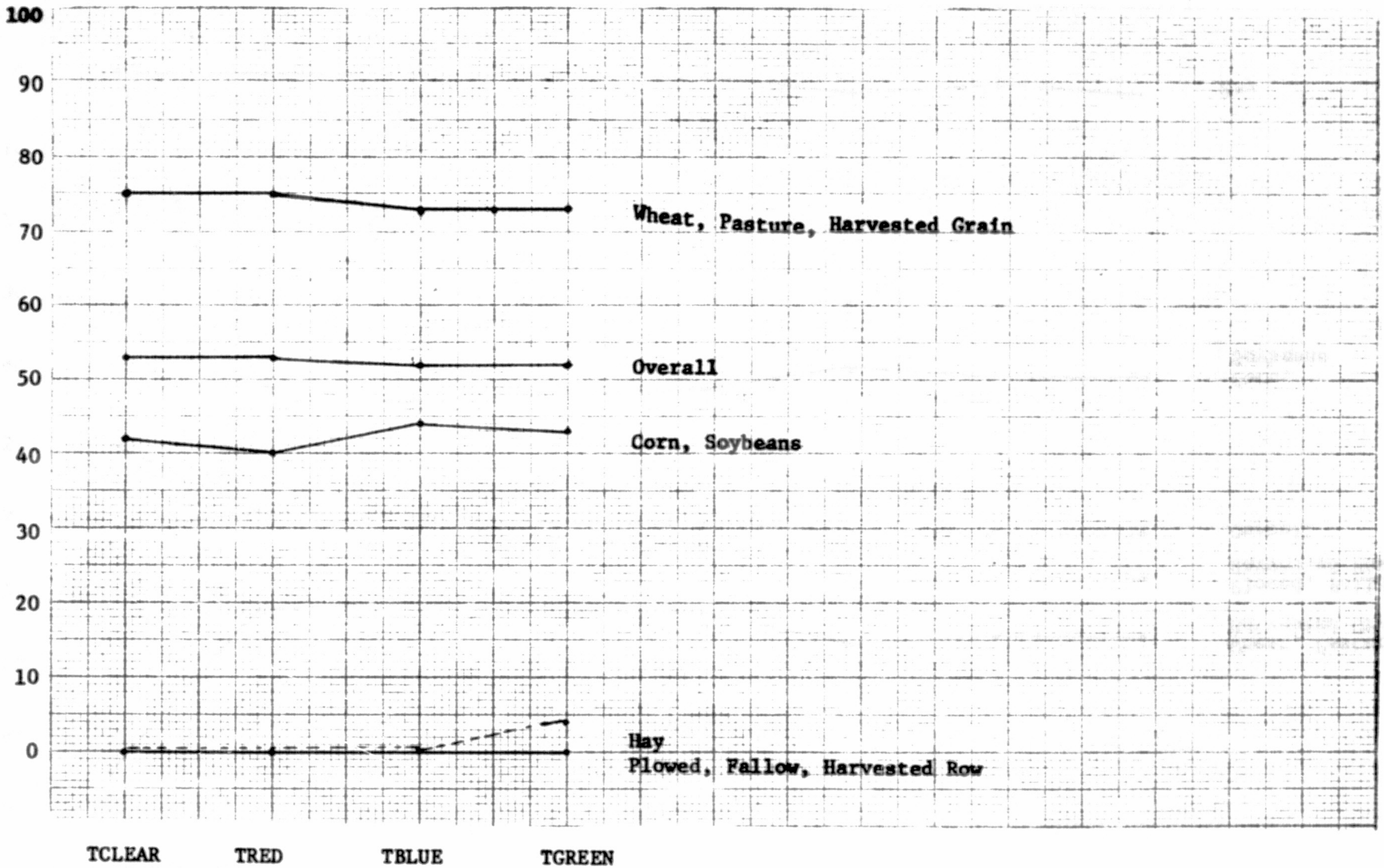


FIGURE 31

Stepwise discriminant analysis, classification into four groups, density scanning mode, South Dakota, 1972.

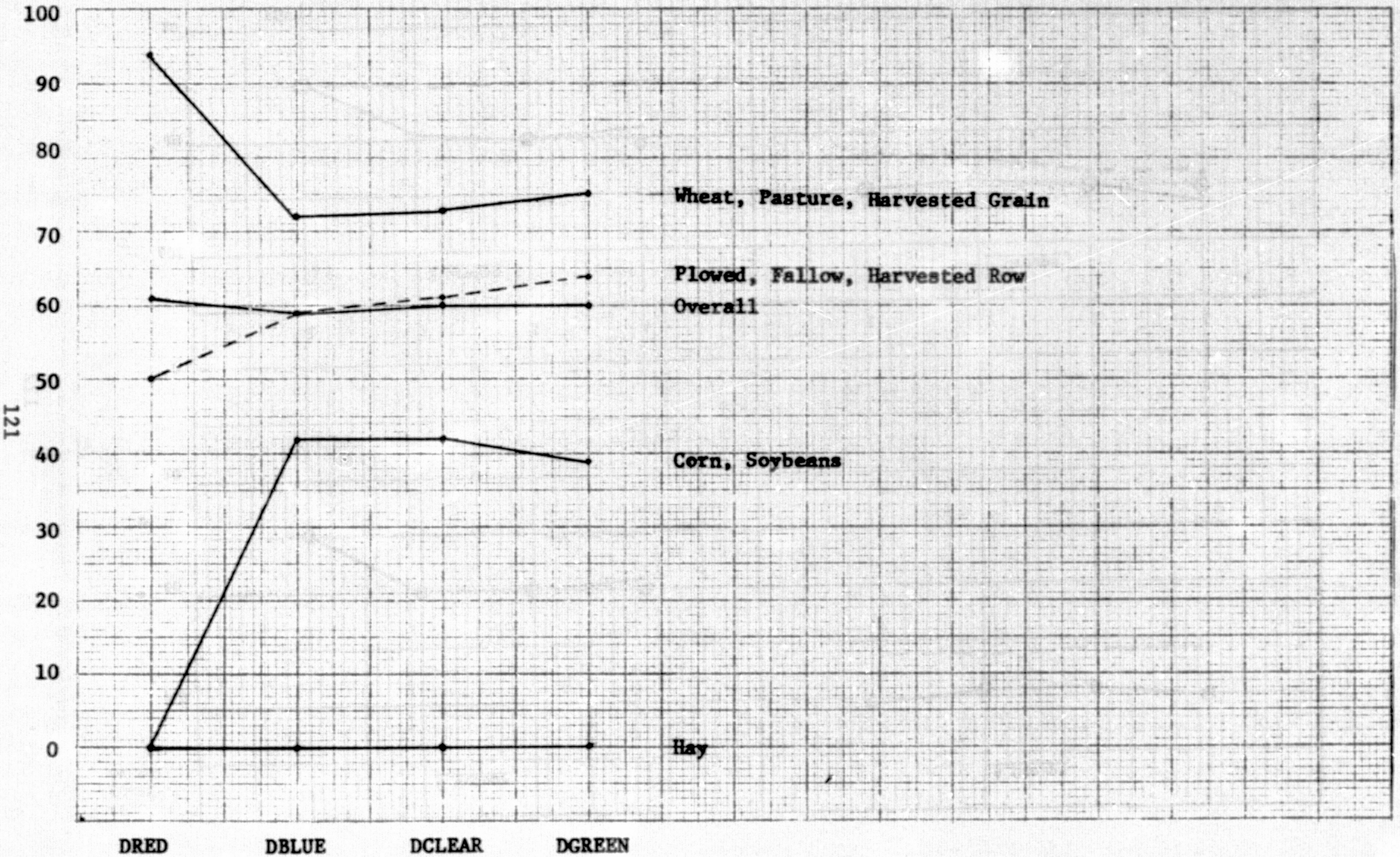
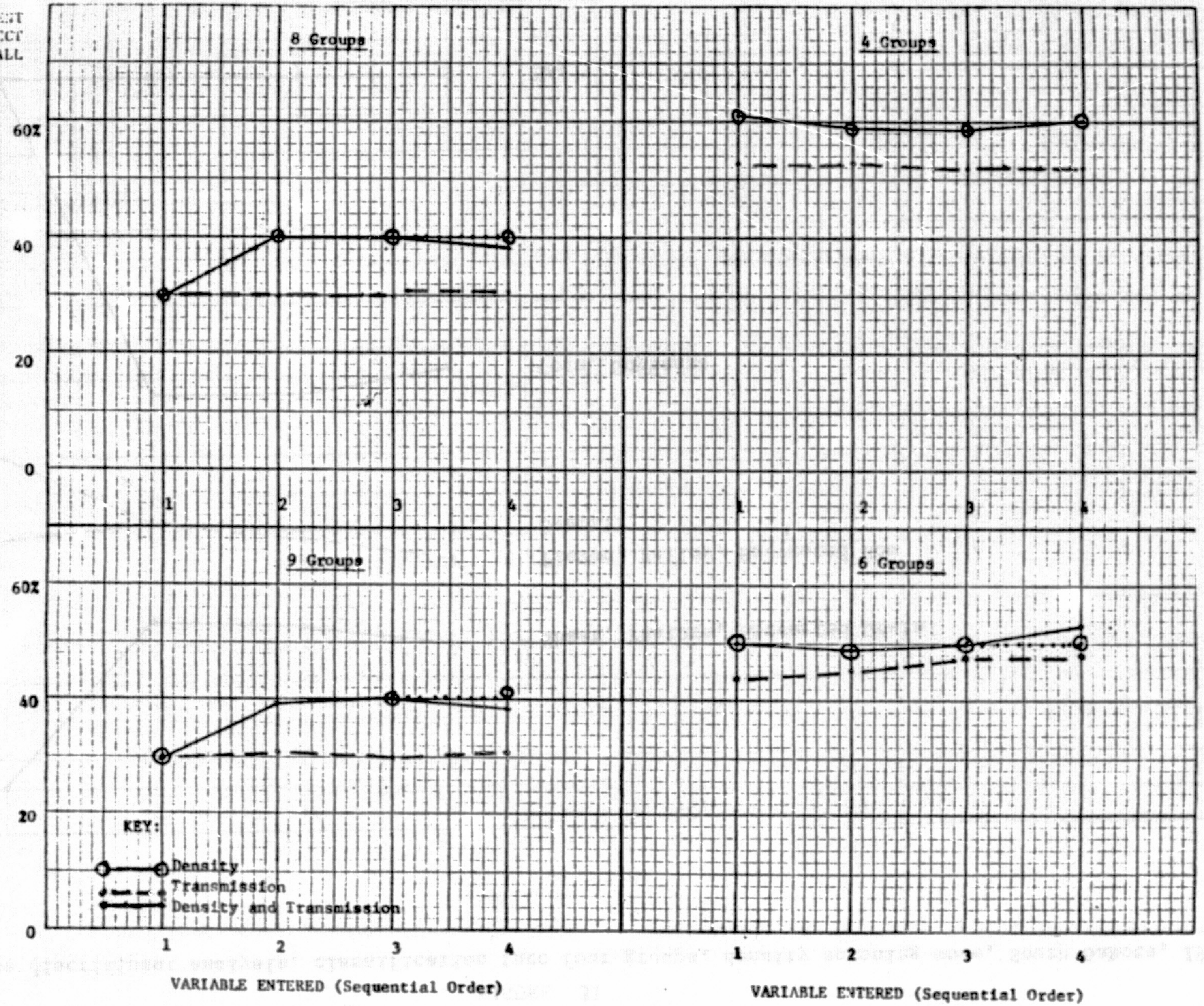


Figure 32--Overall Classification Accuracy By Number of Groups and Measurement Mode for Four Variables



4.2 Crop Acreage Estimation

The objective of this section is to present a procedure that will use classification results to produce an area acreage estimate. The regression technique presented may not be appropriate for users with different ground data. This technique requires that a random subsample of the total of all segments be selected for ground observations.

It is assumed that classification errors will be substantial, that is, perfect classification is not possible, and unbiased classification is not probable. Unbiased classification means more than that the classification errors simply balance. It means that the prior probabilities used are correct and the data are multivariate normal.

If unbiased classification were possible, we could use pixel counting techniques as estimators.

We know that the prior information was not exact and further that the data are not multivariate normal. Some delicate adjustments are necessary to produce an unbiased estimator and in order to make this adjustment, we will use the fact that a random subsample of segments has been selected for ground observations.

The first step is to estimate the linear relationship between total crop acres and total crop pixels inside the segment. This information must come from the ground truth segments and the relationship must be applied to the segments that were not selected for ground observations. An example of how the procedure would work follows. It turns out to be illuminating, but the estimates are poor because the relationships that are established in the ground observation segments do not represent the population that is being estimated.

This data came from the Southwest Crop Reporting District in Kansas.

The correlation coefficients squared (r^2) between the items of interest are presented in Table 66.

The relationship between acres on the ground and points classified corresponding to the same on the ground area can be established on a per segment basis.

Table 66--Source, r^2, $\overline{Y}$, $\overline{X}$, Var(Y), Cov(XY), and Var(X).

Source	r^2	$\overline{Y}$	$\overline{X}$	Var(Y)	Cov(XY)	Var(X)
Total acres (Y) versus total pixels (X)	.95	1843	1841	2,401,627	2,716,190	3,242,228
Alfalfa acres (Y) versus alfalfa pixels (X)	.01	39	223	7,187	-2,417	9,302
Pasture acres (Y) versus pasture pixels (X)	.89	728	890	1,467,689	1,325,965	1,348,245
Corn acres (Y) versus corn pixels (X)	.76	145	69	61,931	23,668	11,850
G. Sorghum acres (Y) versus G. Sorghum pixels (X)	.53	171	404	70,505	115,948	656,917

The model that will be used to represent the relationship is:

$$\hat{y}_i = \bar{y}_i + b_i \, (\overline{X}_{total\ i} - \overline{X}_{sample\ i})$$

where $\hat{y}_i$ is the adjusted acreage estimate for the i^{th} crop.

$\bar{y}_i$ is the average number of acres of the i^{th} crop in the selected segments.

b_i is the regression coefficient for the i^{th} crop estimated by:

$$\frac{\sum_{j=1}^{N} x_{ij}\, y_{ij}}{\sum_{j=1}^{n} x_{ij}^2} = \frac{\text{cov}(xy)}{\text{var}\ (x)}$$

where $\overline{X}_{total\ i}$ is average number of pixels of i^{th} crop in all segments in a county.

$\bar{x}_{sample\ i}$ is the average number of pixels in the selected sample for the i^{th} crop.

The estimator y_i is the adjusted average number of acres in the average segment. To get an estimate of the total y_i, would be multiplied by the total number of segments in the population (N).

The error of the regression estimator is written as:

$$Var(\hat{Y}) = \frac{S^2_{yi} \quad (1-r^2)}{n}$$

where $Var(\hat{Y}_i)$ is the variance of the final adjusted estimator of the average segment of the i^{th} crop.

S^2_{yi} is the adjusted between segment sums of squares for the i^{th} crop.

r^2 is the correlation coefficient squared between the number of acres in the segment and the computer classified number of pixels in the segments for the ith crop.

n is the number of degrees of freedom in the estimator.

Since the estimator for the total number of acres in the county is $N(\hat{y}_i)$, the variance of the total is N^2 times $Var(\hat{y})$.

The regression estimator above is the best in terms of lowest bias and smallest variance. Other estimators of the regression type such as, ratio estimators and difference estimators may be quite good in special cases. The regression estimator has definite advantages over the other two types of estimators just mentioned.

In Stevens County, Kansas, each pixel was classified. There were 410,505 pixels in the county and 468,000 acres. Each pixel represents 1.1401 acres. Actually, the county boundaries were approximated and this introduces a small amount of error. Out of the total of 410,505 pixels, the following pixels were classified as:

1.)	Alfalfa	5,362
2.)	Pasture	172,021
3.)	Corn	30,448
4.)	Grain Sorghum	165,107
5.)	Other	37,567

The first step is to put these pixels into a per segment basis. There were 280 segments in the county so the average segment contains 1,466 pixels for all land uses. The other averages were:

1. Alfalfa 19.2
2. Pasture 614.1
3. Corn 108.7
4. Grain Sorghum 590.0
5. Other 134.0

Since the relationship between alfalfa acres and alfalfa pixels is quite poor, we shall demonstrate the procedure using pasture data.

The pasture acreage estimate for Stevens County using ERTS data is:

$$y_{pasture} = 430 + .9835(614 - 714) = 332$$

$$\hat{Y} = (280)(332) = 92,960 \text{ acres for Stevens County.}$$

$$Var(\hat{Y}_{acres}) = \frac{(1467,689)(4)(1-.89)(280)^2}{4(5)} = 3,164,337,484.$$

Standard Error = 56,252.4 C.V. = 60.5

The estimate and variance without using LANDSAT data are 120,400, and 23,013,363,520, respectively:

$$\text{where } V(\hat{y}) = \frac{1,467,689}{5}(280)^2 = 23,013,363,520$$

$$\text{and C.V.} = \frac{151,702}{120,400} = 126\%$$

Table 67 shows acreage estimates with variance and coefficients of variation for various crops with the aid of LANDSAT data.

Table 68 shows acreage estimates, variances, and C.V.'s for Stevens County, disregarding LANDSAT data.

The first point is that the variances of the estimates that use LANDSAT depend on the variance of the ground observations, the correlation of LANDSAT data with ground observations and the sample size. If the correlation is very high as with pasture, it is possible to produce an accurate estimate only if the ground observation is accurate. For example, no alfalfa was observed in the ground truth segments. Even though the computer was trained with alfalfa from outside the county and 5262 pixels were classified into the alfalfa category for Stevens County, the relationship was bad, and the ground observations were poor, and therefore, the estimate is bad and the C.V. very large.

These estimates and estimates of the variance were computed for two sample sizes. There were really three segments in Stevens County, and one of those was not used because of location problems. These numbers used the two segments left in Stevens County, the relationship for all 17 segments, and the total Stevens Company classification data. However, variances and C.V.'s were figured for samples of size 5 and 10.

If total aircraft classification were available for the same area, the model would be as follows:

$$\hat{y} = \bar{y} + b_1 (\bar{X}_1 - \bar{x}_1) + b_2 (\bar{X}_2 - x_2)$$

The variance would be similar to the previous formula:

$$\mathrm{Var}(\hat{y}) = \frac{S_y^2 (1-R^2)}{n}$$

where R^2 is the multiple correlation coefficient squared and n is the number of degrees of freedom left in the estimator.

Table 67--Acreage estimates, variances, coefficients of variation for sample sizes of 5 and 10, using LANDSAT data.

Crop	Acreage Estimate	Sample of 5 segments		Sample of 10 segments	
		Variance	Coefficients of Variation	Variance	Coefficients of Variation
Alfalfa........	0	111,565,238	∞	55,782,619	∞
Pasture........	92,960	2,531,469,978	54.1%	1,265,734,994	38.3%
Corn...........	78,764	223,058,739	19.4%	116,529,370	13.7%
Grain Sorghum..	150,689	519,593,648	15.1%	259,796,824	10.7%

Table 68--Acreage estimates, variances, coefficients of variation for sample segments of size 5 and 10, without the aid of LANDSAT data.

Crop	Acreage Estimate	Sample of 5 segments		Sample of 10 segments	
		Variance	Coefficients of Variation	Variance	Coefficients of Variation
Alfalfa........	0	112,692,160	∞	56,346,080	∞
Pasture........	120,400	23,013,363,520	126.0%	11,506,681,760	89.1%
Corn...........	65,520	971,078,080	47.6%	485,539,040	33.6%
Grain Sorghum..	321,840	1,105,518,400	14.3%	552,759,200	10.1%

V. Cost Analysis

This section is presented to provide cost information relative to various sources of data collection. It is documented so that as technology is improved, the cost of developing an integrated data collection system can be realistically evaluated.

However, the cost data cited reflects only the conditions under which this project was completed. It is to be expected that new technology will change some of these costs in the future.

Cost of Ground Data

The cost of ground data can be broken into collection costs and summarization costs. The data collection costs include:

a. pre-survey planning and materials preparations,

b. enumerator training schools, and

c. enumerator fieldwork.

The summarization costs include:

a. collection, edit, and keypunch time for Washington, D.C. and State Statistical Office personnel, and

b. programming and summarization costs. These costs pertain to all four test sites and are as follows:

5.1 Data Collection

1) Survey Planning and Materials Preparation		
Research and Development		
Salaries	$1,342.00	
Travel costs (map preparation salaries)	263.67	
Programming Costs		
Salaries	849.59	
Computer costs	1,259.11	
		$ 3,714.37
2) Enumerator Training Schools		
Instructors	901.84	
Salaries	477.44	
Travel		
Enumerators		
Salaries	530.00	
Travel	210.00	
		$ 2,119.28

3)	Enumerators Fieldwork		
	Salaries	$4,931.65	
	Travel	3,044.00	
			7,975.65
Total Data Collection Costs			$13,809.30

5.2 Data Summarization

1)	Collection Edit and Keypunch Costs		
	SSO Salaries	2,524.05	
	Research and Development Salaries	6,816.68	
			$ 9,340.73
2)	Programming and Summarization Costs		
	Salaries	2,632.80	
	Computer Costs	1,281.49	
			$ 3,914.29
Total Data Summarization Costs			$13,255.02
Total LANDSAT Ground Truth			$27,064.32

It should be noted that the above cost data are for the update work conducted in August, September, and October. The costs of the regular June Enumeration Survey (JES) are not comparable since in addition to observing and recording ground cover, the JES records crop intentions and livestock numbers. Estimates of these costs can be derived, however, by using enumerator time and mileage costs. Mileage rates and hourly wages applied against the miles driven and hours worked together give a total cost estimate by segment. This comparison follows:

5.3 JES Fieldwork Costs

A. Time

District	State	Time	#Segs.	$/hour	
9	Missouri	6.42 hr/seg.	52	$3.30	$1,101.67
6	S. Dakota	4.80 hr/seg.	50	3.30	792.00
7	Kansas	8.93.hr/seg.	48	3.30	1,414.51
2	Idaho	5.75 hr/seg.	44	3.30	834.90
	Total		194		$4,143.08

Time cost per segment = $21.36

B. Mileage

District	State	Miles	#Segs.	$/mile	
9	Missouri	99.98 m/seg.	52	.11	$ 571.89
6	S. Dakota	80.86 m/seg.	50	.11	444.73
7	Kansas	136.81 m/seg.	48	.11	722.36
2	Idaho	82,85 m/seg.	44	.11	400.99
	Total Mileage Cost				$2,139.97
					$6,283.05
	Mileage cost per segment		$11.03		

C. Total Time and Mileage

Total time and mileage cost/segment $32.39

5.4 Update Fieldwork costs (3 visits)

A.	Salaries	4,931.65	
B.	Travel	$3,044.00	
C.	Total Time and Mileage ($7,975.65/3=$2,658.55)		$7,975.65

Total update time and mileage costs per segment $41.11

Total update time and mileage costs per segment per visit $13.70

The difference between $6,283.00 and $2,658.55 represents the additional costs of $3,624.50 needed to locate the June Segment Operators, secure livestock data and farm labor data. This LANDSAT update fieldwork only included locating the segments and recording the crops present and their conditions. The operators were not contacted unless the enumerator could not view the fields from the road.

Tables 69 through 76 show detailed time and mileage data for the study sites.

Table 69--Missouri 1972 JES Time and Mileage Data

Dist	Number Segs	Visits/ seg	Hours/ seg	Miles/ seg
10	60*	1.43	4.90	70.63
20	51*	1.75	4.70	81.08
30	49	2.04	4.91	82.78
40	50	1.94	5.30	86.78
50	60*	1.93	5.34	80.90
60	39*	1.82	6.95	85.08
70	46	1.52	5.19	61.04
80	42	1.71	5.55	94.52
90	52	2.23	6.42	99.98

*Not all segments in this district had cost data reported.

Table 70--South Dakota 1972 JES Time and Mileage Data

Dist	Number Segs	Visits/ seg	Hours/ seg	Miles/ seg
10	31	1.87	7.93	107.16
20	46	1.67	5.33	93.20
30	42	1.83	5.74	88.19
40	31	1.90	8.19	122.23
50	42	1.90	6.96	99.90
60	50	k.92	4.80	80.86
70	22	1.86	8.32	131.23
80	35	1.63	7.23	101.14
90	51	1.82	4.65	75.98

Table 71--Kansas 1972 JES Time and Mileage Data

Dist	Number Segs	Visits/ seg	Hours/ seg	Miles/ seg
10	42	2.26	6.99	130.52
20	54	2.04	5.67	94.43
30	50*	2.58	5.74	88.94
40	40	2.25	8.16	141.90
50	56	2.46	6.81	127.21
60	53*	1.87	4.92	69.00
70	48	2.79	8.93	136.81
80	60*	1.93	6.68	105.83
90	53*	1.81	4.69	82.34

Table 72--Idaho 1972 JES Time and Mileage Data

Dist	Number Segs	Visits/ seg	Hours/ seg	Miles/ seg
2	54	1.54	5.75	82.85

Table 73--Time and mileage data for Idaho by enumerator.

Enumerator identification	JES segments completed	Average Number visits per segment	Average hours per segment	Average miles per segment
6	3	3.00	3.78	104.67
12	8	3.00	8.83	125.00
18	10	2.10	3.44	56.70
19	14	3.29	7.41	80.07
30	13	1.54	4.19	80.23
33	6	2.83	5.93	71.00
Totals:	54	2.54	5.75	82.80

Table 74--Time and Mileage data for Missouri by enumerator.

Enumerator identification	JES segments completed	Average number visits per segment	Average hours per segment	Average miles per segment
1	15	1.33	4.89	58.53
2	15	1.47	4.19	80.93
3	6	2.50	3.38	87.5
4	17	1.76	3.66	59.18
5*	6	1.83	7.77	118.33
6	9	1.89	7.09	93.78
7*	12.5	1.68	4.69	66.64
8	13	1.77	7.06	78.92
9	10	1.40	4.99	78.10
10	11	1.73	4.44	109.91
11	16	2.12	4.64	78.56
12	15.7	1.27	5.06	63.76
13	17	1.41	4.50	57.53
14	13	2.15	6.27	109.23
15	13	1.08	4.94	56.23
16	11	1.45	5.11	81.18
17	15.5	2.58	5.97	115.10
18	14	2.43	7.01	93.21
19*	14	1.21	3.93	51.00
20	11	1.55	6.66	90.09
21*	13	2.00	5.58	78.46
22	7	3.57	9.18	99.71
23*	13	1.92	5.47	87.62
24	11	3.27	10.27	163.91
25	14	2.07	4.79	90.36
26	12	1.67	4.60	64.17
27	14	2.43	6.70	102.57
28	14.3	2.10	4.70	103.08
29	9	1.67	8.78	67.67
30	16	1.94	5.07	78.00
31	9	2.56	6.13	114.33
32	18	1.28	4.56	46.50
33	9	1.22	4.28	62.11
34	10	1.50	5.38	77.80
35	17	1.41	3.76	65.29
36	8	1.75	6.12	130.25
TOTALS:	449	1.82	5.42	82.22

* Supervisors

Table 75--Time and mileage data for South Dakota by enumerator.

Enumerator identification	JES segments completed	Average number visits per segment	Average Hours per segment	Average miles per segment
1*	2.0	1.03	1.04	37.59
2*	3	1.33	5.50	111.33
3*	8	1.75	4.35	73.12
4	13.8	1.88	6.96	151.52
5	8.6	1.74	6.25	87.79
6	7	1.43	8.36	80.14
7	14	2.07	5.91	117.03
8	6	1.50	4.79	46.50
9	19	2.16	4.13	80.21
10	15	1.93	5.51	95.93
11	9	1.33	7.19	69.78
12	7	1.86	5.17	63.43
13*	13	1.85	3.87	79.54
14	11	1.45	5.18	71.91
15*	7.4	1.08	2.90	62.03
16	10.5	1.43	5.09	76.86
17	13	2.38	9.29	144.69
18	12	2.50	1	159.00
19	25.4	2.24	4.72	82.72
20	15	1.53	7.01	90.07
21	11	1.73	4.67	71.55
22	14	1.64	4.90	133.29
23	15	1.60	5.61	70.33
24*	8.7	1.84	8.33	126.78
25	13.1	2.14	7.36	129.01
26	8.3	2.29	7.14	135.54
27	13	1.54	7.81	83.85
28	15	2.00	7.48	95.33
29	15	1.80	5.38	77.67
30	5	1.40	6.02	33.00
31	9	1.56	5.81	113.67
32	2.3	2.17	5.20	149.13
TOTALS:	350	1.83	6.27	95.91

* Supervisors

Table 76--Time and mileage data for Kansas by enumerator.

Enumerator identification	JES segments Completed	Average number visits per segment	Average hours per segment	Average miles per segment
1	8.9	2.70	7.15	105.73
2	18	2.00	5.21	105.11
3	15	2.40	5.58	89.67
4	12	2.33	4.63	78.17
5	20	1.90	5.99	111.00
7	14	2.71	5.29	98.07
8	9.8	1.63	9.29	134.69
9	11	2.09	6.33	69.18
10	9	1.56	7.96	135.56
11	4	1.75	7.96	80.00
12*	3.4	2.94	11.06	272.94
13	19	1.95	4.82	74.37
15	16.9	2.84	9.09	180.65
16	14	3.5	8.78	142.36
17*	4	1.25	6.69	116.75
18	11	1.73	4.83	58.64
19	12	2.42	8.38	162.08
20	12	3.25	8.68	132.92
21	14	2.64	5.08	108.07
22	3	1.67	5.67	98.67
23	12	2.25	7.85	122.00
24	14	2.71	4.67	90.00
25	16	2.44	4.60	99.69
26	13	1.85	6.50	108.77
27*	5	3.00	15.92	260.20
28	16	2.12	5.43	89.12
29	12	1.75	5.62	69.25
30	1[illegible]	2.08	7.25	117.25
31	1[illegible]	1.60	4.76	65.73
32	14	1.21	4.23	69.14
33	9	3.11	7.62	114.67
34	7	3.29	10.86	173.86
35	11	2.82	9.20	151.09
36	14	1.64	7.02	93.57
37	11.5	1.48	4.85	81.04
38	17.5	2.46	5.15	107.43
39	15	1.53	5.29	91.40
40	11	1.73	5.93	54.91
TOTALS:	456	2.21	6.44	107.10

* Supervisors

5.2 Aircraft Cost Analysis

NASA provided the following estimates for aircraft costs:

U-2 operational costs are $2,150 per hour with coverage of about 400 nautical miles per hour. Coverage is 14.8 nautical miles on a side per scene.

Scenes per hour $= \frac{400}{14.8} = 27.03$

Cost per scene $= \frac{\$2{,}150}{27} = \$79.63 = \$80$

For the study areas, the acquisition costs average about $60 per segment.

The activities and the approximate time and costs required to prepare the aircraft data for crop classification are:

		Average time per segment
Sketch segment and record field boundaries		37 min.
Microdensitometer scanning		33 min.
Recording and keypunching input data for field extraction		
	Total man hours	1.83 hours
	Cost/man hours	$4.50
	Average cost/segment	$8.23

ADP Costs

PDSCMS data conversion

Field extraction

Total ADP costs/segment

The average cost per segment for data preparation $29.23

The costs of crop classification varies with the size of segment, but in order to have a comparable cost with ground observations, it is presented on a per segment basis. The average cost per segment for crop classification was about $81 segment. The average cost per data point is about 3 cents per point.

The total aircraft survey costs were about $170 per segment. This compares with $47 per segment per visit for the ground observations.

This analysis deals primarily with the time and costs required for scanning the aerial photography and converting the data into a form suitable for crop classification by discriminant analysis in the Statistical Analysis System (SAS).

Time and cost data were collected as follows:

1) Pre-scan setup: the time (man minutes) required to locate the segment on the microdensitometer, sketch the segment, record field boundary coordinates and define the microdensitometer scanning parameters.

2) Scanning: the time (man minutes) required for system analog calibration and microdensitometer scanning with each of the four filters in density and transmission units.

3) Data preparation for field extraction: the time required to record and keypunch input data for the field extraction program.

4) PDSCMS data conversion: ADP costs for converting the microdensitometer output data to SAS compatible data.

5) Field extraction: ADP costs for assigning crop classes, tract and field identifiers to individual pixels on the basis of ground observations utilizing pixel coordinate information.

Several factors contributed to the substantial differences between states for the average cost per segment. The differences for pre-scan setup times can be attributed to two primary factors:

1) different microdensitometer operators. A new operator was in training while scanning South Dakota, and had gained in experience when Missouri was scanned.

2) the relative difficulty recording field boundary coordinates for each state. South Dakota and Missouri were most difficult because of many small field sizes, followed by Idaho, with Kansas least difficult.

New field boundary coordinate recording procedures were implemented near the end of the Idaho scanning and were subsequently employed while scanning the Missouri photography. Due to operator differences, it is difficult to objectively assess the effectiveness of the new procedures. Subjectively, it is believed the new procedures will reduce pre-scan setup time by 10-20% and data preparation for field extraction by 25-40%.

Scanning time remains fairly constant between states (the large difference in South Dakota is attributable to a new operator in training on the microdensitometer). Small differences are a function of the number of segments and average size of each segment.

Between state differences in automated data processing costs are a function of the number of segments, average size of each segment, and the number of tracts and fields within each segment.

5.5 Computer Costs

Processing LANDSAT data and digitized aircraft data requires enormous amounts of computer time. The following table shows the cost of computer time for processing at the Washington Computer Center for various broad classes of processing.

DEVELOPMENT	$ 6,631
GROUND DATA	$ 2,915
MAPS FOR SEGMENT LOCATION	$ 9,227
AIRCRAFT DATA ANALYSIS	$30,142
TOTAL	$48,915

Development costs include converting software to run at WCC, maintenance, developing original programs and overhead.

Ground data costs are for building and maintaining, and summarizing of ground data.

The MAPS were grey level maps of LANDSAT CCT's for segment location.

The aircraft analysis cost includes charges for conversion of microdensitometer data, building the data, files, and runs used to determine the best analysis procedure, the discriminant analysis, and the combination of the satellite results, aircraft results, and ground data.

APPENDIX A

ERTS ENUMERATORS INSTRUCTIONS

ERTS [1/] ENUMERATORS INSTRUCTIONS

1.1 What you will do:

You are one of about 16 enumerators in four states (Kansas, Missouri, South Dakota, and Idaho) employed to obtain "ground truth" about crop species, acres and crop condition. Briefly, your job is to update information from the June Enumerative Survey (JES) by verifying crop species and acres and observing crop condition during July, August, September, and October. Your field verifications and observations are to be recorded on the Earth Resource Technology Satellite (ERTS) Ground Truth Printouts.

1.2 Equipment and Supplies

USDA identification card
aerial photos
aerial photo mailing boxes
county maps
CEF-201's
ERTS Ground Truth Printout
large envelopes for mailing completed forms
motor vehicle accident report kit
ball point pen
lead pencil, plus red, orange, and yellow colored pencils
clipboard
highway maps
Julian dates.

1.3 Mailing and survey dates

After each survey is completed you will mail the updated printout and your CEF-201's to the SSO in the envelopes provided. All other materials used during the survey will be retained until the final survey period Your final mailing will include the updated printout, CEF-201, aerial photos and county maps, plus any other surplus materials.

The survey periods and mailing dates are as listed:

Survey period		Enumerators mailing date
August	7-11	on or before August 11
September	11-15	on or before September 15
October	10-13	on or before October 13

[1/] At the time this manual was written, LANDSAT was called ERTS, acronym for Earth Resources Technology Satellite. ERTS was never changed to LANDSAT because this manual was never used after 1972.

1.4 Terms and definitions

The regular enumerative survey definitions hold for this survey

A. Segment - land area outlined in red on aerial photos and county maps. Each segment is identified by a permanently assigned 4-digit number. See the "Survey Enumerators Handbook" for discussion on use of aerial photos and locating segments.

B. Tract - an area of land inside the segment which is under one management, Each tract is identified by a letter code A, B, C, etc. on the aerial maps and by the corresponding numeric code on the form printout (i.e. A = 01, B = 02, C - 03, etc.). Tract boundaries and letter codes are drawn in blue pencil inside the segment on the aerial photo.

C. Field - a continuous area of land inside a tract which is devoted to one crop or land use. Each tract on the aerial photo is divided into fields during the enumeration of segment acreage in late May or early June. Fields are numbered and their boundaries outlined in red pencil.

D. Farm Operator - the person who is responsible for the day-to-day decisions for a tract.

Part II - The Survey

2.1 Purpose of the survey

The purpose of the Earth Resource Technology Satellite (ERTS) program is to:

a) investigate and evaluate the use of space imagery to identify crop species.

b) investigate ways of using space imagery to improve agricultural statistics.

Through ground truth obtained during July, August, September, and October we will be able to check and verify the accuracy of satellite imagery (500 miles) and high altitude photography (60,000 feet) as a method of measuring crop acres which in turn will be used to generate an expanded estimate of crop acreage.

Ground, high altitude, and satellite estimates of acreage will be obtained and compared against collection costs to indicate a cost-information ratio. Trained enumerators as yourselves will collect the ground information from the field. Trained photo interpreters will record species and acreages for high altitude photography in the Washington, D.C. office. Computers will be used to analyze satellite imagery in the Washington, D.C. office. Cost information for each method of collection will be retained and compared versus the accuracy of reliability of each method of data collection.

2.2 The sample

The segments selected for this survey were selected to provide different crops in different locations. A different mix of crops will be found in Idaho versus Kansas versus Missouri versus South Dakota. How do sugar beets compare with potatoes in Idaho or grain sorghum in Kansas? Will spring wheat be distinguishable versus winter wheat? Does corn in Missouri look the same as corn in South Dakota? The information collected will provide answers to these types of questions. Additionally, with the distant geographic areas, inclement weather should not cover all the test sites and limit the quality of all imagery on a particular survey.

We use the JES segments since they represent 100% coverage of the areas in question. If bad weather renders some of the aerial photography or satellite images useless, we will attempt to develop reliable estimates for the other areas based on the ERTS ground truth. This may become a multiple frame model for acreage estimation.

2.3 Survey forms

For the second visit you will be provided a printout listing in segment and tract order fields, acreages and crops from the JES (first visit). On this visit you will note the condition of the crop on the printout. On succeeding visits the printouts will show fields, acreages, crop and conditions for each earlier visit.

The name and addresses of the operators from the JES will be provided on separate sheets of paper grouped by segments. These will be for use when it is necessary to locate the tract operator for permission to view fields not observable from public roads.

Part III - Field Observations

3.1 Locating the segment

Locate the segments you will be visiting on the county maps, then plot them on a highway map. Plan your journey to observe these segments with minimum mileage and travel time for each day's journey.

3.2 Recording observations for segment

When the printout and maps are updated on the monthly visit, use the color codes listed below for field boundaries and numbers.

Second visit	- red dashes
Third visit	orange dashes
Fourth visit	yellow dashes

Note: We will only mark corrections on the map. Incorrectly drawn fields will not be erased. There may be no new dates for the survey duration if fields are drawn correctly from the JES and there are no acreage changes.

For this survey we are not interested in transfer of ownership etc., except to know whom to contact for enumeration purposes. Our concern is enumerating the land use and crop development condition of the segment.

Step 1. Verify that you are looking at the correct photograph(s) and the correct printout by locating landmarks on the map and locating the segment and tract number on the printout. Record the Julian date on all N of N pages of the tract printout.

Step 2. Verify that the field is drawn correctly on the maps by a) looking at the field defined by the map and b) deciding whether the map accurately shows the field with respect to common landmarks. If the field cannot be observed from public roads contact the tract operators and request permission to observe the fields in question, then write on the bottom of the printout whether permission was secured or refused. (By default, the printout will write unasked unless permission is noted as secured or refused).

If permission is refused, record observations for fields observable from public roads. Enter refused (code - RFSD) for the fields not observable from public roads in acres, crop and condition columns. If the map is correct go on to Step 3. If the map is incorrect, redraw the fields using the correct color scheme before beginning Step 3. Do not erase any previous survey boundary lines.

Step 3. For the given field number, check the acres listed on the printout versus the map and your own best estimate of the actual field acreage in whole numbers. If the acres are the same as the previous visit check (✓) the space for acres. Where a correction is necessary (i.e. --- an error has beem made or an obvious change in acreage has occurred since the previous visit) check with the operator for the corrected acres or record your own acreage observations where checking is not possible and write a note explaining the change. See Figure 1 for examples of corrections and changes.

Step 4. If the crop is the same as the previous visit check (✓) the crop code. If a change has occurred record the corrected crop code for that field.

Step 5. Using the guide from Section 3.3, write the condition of the crop on the printout in the space provided. Write a note to explain any situation or our condition codes do not accuractly describe.

Step 6. Repeat steps 1-5 until all fields are completed, then check that all N of N pages for a tract listing are present and complete.

Step 7. Repeat steps 1-6 until all tracts and segments are completed.

3.3 Crop codes and conditions

Since we will be looking at aerial photography and satellite imagery, we need to know the crop species and the condition code that best described and coded appropriately on the printout. In order to code the condition properly you must observe the total area in the field which would be covered by the crop and then give a subjective evaluation of the crop development as well as a recommendation for action to be taken.

Crop	Situation	What to do
Alfalfa	part down, part cut	condition = cut or down, whichever portion is the larger.
All crops	very poor stand, dry etc.	condition - what a normal healthy crop would look like with a note.
Applies to most	two similar species but different planting dates	draw in new fields boundaries and properly number and classify the new field.
Grass waterways	located on natural boundaries or can accurately be drawn on the map	draw in new field and classify as OTHR and specify on printout with a note.
Drowned out areas	located near natural boundaries or can accurately be drawn on the map	draw in new field and classify as OTHR and specify on printout and a note.

ERTS Editing Instructions

1. Edit the Julian date to correspond to the actual field visit.

2. Check the acres for a given field number versus the previous recorded acres.

   A. Do not edit where column is checked (✓).

   B. Dn case the acres differ from the previous visit:

      1) If a new field is created or acres for given fields are adjusted, the acres should be adjusted to total the previous acreage total or a note shoudl explain the total acres change. Check that the new field is correctly numbered.

      2) Where an error occurred on the previous enumeration, an enumerators note should explain the correction. With an explanation, the correction will be punched. With no explanation talking to the enumerator or statistics judgement will appropriate to edit in corrected acres.

   C. Check (✓) the acres column where the tract was a refusal and the field not observable.

3. Check the crop codes for correctness.

   A. Do not edit where the column is checked (✓).

   B. Edit out where they are unrecognizable.

   C. Correct the code when it is a change from the previous month and incorrectly written.

   D. Enter RFSD where the tract was a refusal and the field not observable.

4. Check the condition code against nearby fields.

   A. Edit to compare with fields in the tract or segment where the condition is not entered.

   B. Check with the enumerator where the condition is not entered and there are no comparable fields in the tract or segment.

   C. Edit cut where condition is the sure as the previous month.
   NOTE: On the first visit there must be an entry for every field.

D. Enter RFSD where the tract was a refusal and the field is not observable.

5. Code the permission

   A. Unasked = 0

   B. Secured = 1

   C. Refusal = 2

ERTS Mailing Instructions

Send the edited printout and the punched cards Air Mail Special Delivery in "Special C" envelopes and "Special C" card mailer.

Each envelope and card mailer should be marked in the lower left hand corner as follows:

REPORT: ERTS Ground Truth

STATE: Your State (99) (1 of 2)

Secure each envelope and card mailer with a strand of filament tape each way around the envelope and card mailer.

Send the aerial photos as follows:

Research and Development Branch
SRS of USDA
Room 4837 South Bldg.
Washington, D.C. 20250

APPENDIX B

State Office Instruments

TERMS AND DEFINITIONS USED FOR JUNE ENUMERATIVE SURVEY (JES)

AND ERTS FIELDWORK

A. SAMPLE:

Information for the ERTS Survey is obtained from a small sampling of the total land area in four States. Small areas of land have been selected at random for this survey. Each area to be enumerated has been outlined in red on the county highway maps and aerial photographs which you are supplied. Every acre has one and only one chance to be selected in the sample.

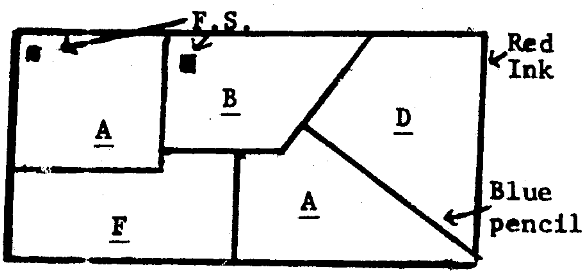
B. SEGMENT:

Segments are land areas outlined in red on aerial photos. Segments generally range in size from one-half square mile to three square miles. A few are larger or smaller depending on locations. Segments are identified by a permanently assigned number.

C. TRACT:

A TRACT is an area of land inside the segment which is under one operation. This tract may consist of agricultural land, non-agricultural land, residential areas, or some other land use. Examples of tracts are as follows:

(1) An occupied house and land in segment operated by the person in charge. Examples are Tracts A and B. Notice that Tract A has land at two locations in segment.

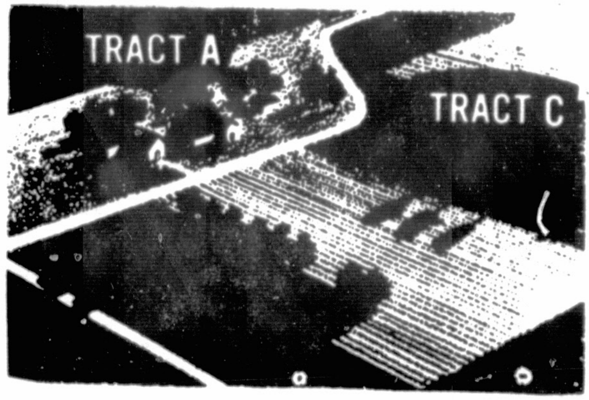


Segment 3189

(2) A farm operator living in the segment on a dwelling where he is not the person in charge, and who has no other land in the segment. See F.S. above. If all the land he operates is outside the segment, he is still a resident farm operator and should be assigned 1.0 acres of land for the tract. He lives in this dwelling.

(3) Any area of land in the segment under one operator who does not reside inside the segment. Tracts D and F are examples.

The boundaries of each tract will be outlined in blue pencil and each tract will be identified by a code letter. If a tract consists of more than one separate parcel of land, all parcels will be identified with the same letter; i.e., all of the land inside the segment that is operated by one person will be reported under one tract code.



D. FIELD:

A field is a continuous area of land inside a tract which is devoted to one crop or land use. Each tract will be divided into fields by you during this Survey. Each field will be outlined in red pencil and assigned a number.

E. FARM:

A farm consists of the area or areas of land both inside and outside of the segment boundaries under one management on which there were crops, livestock, poultry, or some sales of agricultural products at some time in 1971 or 1972.

F. OPERATOR:

The OPERATOR is the person who is responsible for the day-to-day decisions for the tract and total land operated.

If the tract contains a farming operation, the operator could be the owner, hired manager, cash tenant, or sharetenant. If a person operates farmland as a hired manager or partner and also operates land for himself as a separate farm, the managed or partnership land should be separated and assigned another tract code.

If the land is rented to others or worked by others on shares, the tenant or renter is considered the operator of the rented land.

UNITED STATES DEPARTMENT OF AGRICULTURE
Statistical Reporting Service
O. M. B. NUMBER 40-R2766
Approval Expires 4/30/74 480

State	District	Segment	Tract

PART A - 3 (Mo.)

JUNE 1972 ACREAGE, LIVESTOCK & LABOR

ENUMERATIVE SURVEY

Use this questionnaire only if operator lives INSIDE the SEGMENT.

Facts about your farm or ranch will be kept CONFIDENTIAL and used only in combination with similar reports from other producers.

SEGMENT NUMBER ______________________ TRACT CODE LETTER: ______________________

NAME OF RESIDENT OPERATOR: ______________________

ADDRESS: ______________________
(Route or Street)

____
(City) (State) (Zip)

TELEPHONE NUMBER: ______________________ COUNTY: ______________________

NAME OF FARM: ______________________

DATE: ______________________

Name and Address of PARTNERS: ______________________
(Record partnership operations as a separate tract.)

1. How many acres are inside these boundaries drawn on the photo (or map)?.. [ ]

2. Will any acres INSIDE these boundaries be IRRIGATED during 1972?

YES ( ) Ask irrigation questions
NO ( ) Skip irrigation questions

Now I would like to ask you about each field in the tract, the total acres in each field, and the crop or land use in 1972. For crops, I will ask for acres planted and to be planted for harvest this year. If two crops will be harvested from the same field, we should list both of them.

A-3

SECTION A - ACREAGES OF FIELDS AND CROPS IN TRACT

FIELD NUMBER...				1	2	3	4
1.	TOTAL ACRES IN FIELD			.	.	.	.
2	CROP OR LAND USE (Specify)						
3.	TWO CROPS HARVESTED FROM THIS FIELD?			No ( ) Yes ( )	No ( ) Yes ( )	No ( ) Yes ( )	No ( ) Yes ( )
3a.	Has this field been planted?			No ( ) Yes ( )	No ( ) Yes ( )	No ( ) Yes ( )	No ( ) Yes ( )
4.	ACRES IRRIGATED AND TO BE IRRIGATED?			.	.	.	.
5.	FARMSTEAD, ROADS, DITCHES, WOODS, WASTE, ETC			.	.	.	.
6.	PASTURE	Permanent - Not in Crop Rotation		842 .	842 .	842 .	842 .
		Cropland - Used Only for Pasture		845 .	845 .	845 .	845 .
10.	WINTER WHEAT	Planted		540 .	540 .	540 .	540 .
11.		For Grain		541 .	541 .	541 .	541 .
12.	RYE	Planted		547 .	547 .	547 .	547 .
13.		For Grain		548 .	548 .	548 .	548 .
14.	OATS	Planted		533 .	533 .	533 .	533 .
15.		For Grain		534 .	534 .	534 .	534 .
16.	BARLEY	Planted		535 .	535 .	535 .	535 .
17.		For Grain		536 .	536 .	536 .	536 .
18.	CORN	Planted		530 .	530 .	530 .	530 .
19.		For Grain		531 .	531 .	531 .	531 .
20.	SORGHUM	Planted		603 .	603 .	603 .	603 .
21.		For Grain		604 .	604 .	604 .	604 .
22.	OTHER USES OF GRAINS PLANTED — Acres abandoned, cut for hay, etc.		Use				
			Acres	.	.	.	.
23	HAY	Cut and to be cut	ALFALFA AND ALFALFA MIXTURES	653 .	653 .	653 .	653 .
			OTHER HAY Kind				
24.			Acres	65— .	65— .	65— .	65— .
25.	SOYBEANS			600 .	600 .	600 .	600 .
26.	TOBACCO			67— .	67— .	67— .	67— .
27.	PEANUTS			690 .	690 .	690 .	690 .
28.	RICE			605 .	605 .	605 .	605 .
29.	COTTON, UPLAND	Planted		524 .	524 .	524 .	524 .
30.		Abandoned		523 .	523 .	523 .	523 .
35.	IRISH POTATOES			552 .	552 .	552 .	552 .
36.	OTHER CROPS	Name					
		Acres planted or in use		.	.	.	.
38.	CROPS PLANTED FOR SOIL IMPROVEMENT ONLY — No other use during 1972			856 .	856 .	856 .	856 .
39.	IDLE CROPLAND - Idle all during 1972			857 .	857 .	857 .	857 .

Special Keypunch Instructions

1. Punch 76 in column 1-2 for all cards.

2. Face page: Punch identification as appears on face page upper right hand corner.

3. Page 2

   a) Punch field numbers as they appear at the top of page.

   b) Item 5: Leave crop code blank and punch acres 'as is.'

      Item 6-9: Punch code and acres 'as is.'

      Item 10-21: Punch code and planted acres only.

      Item 22: Skip.

      Item 23-39: Punch code and acres 'as is.'

4. Page 3 on

   a) Punch field numbers as they appear at the top of page.

   b) Other Land: Leave crop code blank and punch acres 'as is.'

      Permanent Pasture and Cropland Pasture: Punch crop code and acres 'as is.'

      After Cropland Pasture through Sorghum for Grain: Punch crop code and planted acres only.

      Alfalfa Hay through Idle Cropland: Punch crop code and acres 'as is.'

5. Verify.

Note: a) Punch acres to one decimal without the decimal point.

b) Right justify and punch lead zeros for all numbers.

c) There will be only one code and one acreage figure punched per field number. The proper code and acreage will always be the first entry in a column for any field number.

ERTS SSO Keypunch Instructions

1. Do not punch blanks or edited out data fields.

2. Punch only current survey data.

3. Punch permission code only on the first card.

4. Punch the first four alpha characters of the recorded condition.

APPENDIX C

Grey-Scale Map Computer Program

WMAP

This program will:

1. Map directly from LANDSAT MSS Bulk data tapes (either the original) non-labeled tapes or standard label copies).

2. Map from any one of the four MSS response bands (LANDSAT channels 4, 5, 6, or 7).

3. Compute a histogram of a sample of a designated area and compute grey scale boundaries for the mapping from this histogram. 1/ The user may specify as many as 16 grey scale classes. The user may also specify:

   a. as to whether or not the program will assign boundaries so that each grey scale class will contain about the same number of data points,

   b. If the number of data points in a class will be proportional to the square roots of the percentage distributions of the different response levels found, or

   c. that the program use limits which are defined by the user.

4. For very large areas, will map about 14,000 characters a second (CPU time) on the WCC IBM 370-168. For smaller areas, e.g. 100 lines and 100 columns the mapping rate decreases to around 5,000 characters per second.

The USER MUST:

1. Specify the response band to be mapped (default is LANDSAT Band 5.

2. Specify the number (k) of grey scale divisions to be used in the mapping (default is 9).

3. Specify a printable character for each grey scale division.

4. Specify the location of the areas to be sampled for the frequency tabulation and/or to be mapped.

1/ If the total number of data points in the sample area is less than 10,000, then the histogram will include all data points in the designated area.

Control cards required for each run are:

1. A CLASS card which will define:

   a. the response band to be mapped (punch in column 14).

   b. the number ($k \leq 16$) of grey-levels to be used in the mapping (punch in columns 15-16, right justified).

   c. the string of printable characters to be used in the mapping (punch these in consecutive one-column fields starting with column 18. The first character will be used for the lowest level set of response values. Blanks in the string will cause a blank to be printed for that level(s) on the map).

   If mapping in LANDSAT band 7 (MSS channel 4), any data points having values of 1 to 5 (deep water) and 6 to 9 (shallow water) will be assigned the characters punched in columns 18 and 19 of the class card. Therefore, when mapping in band 7, the user should specify (k+2) printable characters on the CLASS card.

2. At least 1 SAMPLE/MAP AREA block card.

   A SAMPLE AREA card defines an area on the tape which is to be sampled for the frequency distribution to be used in determining the class levels for the printout. The first card after the CLASS card will always be treated as a SAMPLE AREA card. Any later card which has a '1' punched in column 20 will also be used as a SAMPLE AREA card. Any SAMPLE AREA card which has a '1' in column 24 will also be treated as a MAP AREA card.

   The format for the SAMPLE AREA card is:

C.C.	
1-4	the number of scan lines to be skipped.
5-8	the number of the last scan line in the desired area.
9-12	the number of data points to the left of the desired area.
13-16	the number [1/] of the last data point to be included in the desired area.

[1/] Columns should always be numbered in conformance with the LARS System whereby data points 1-804 are on tape 1, 805-1614 on tape 2, 1615-2424 on tape 3, and 2425-3228 on tape 4.

20 a '1' (optional if first SAMPLE AREA card).

32 a '1' (optional, to be used only if class limits are to be assigned by means of the square root transformation).

The MAP AREA card defines an area for which a grey-scale printout is to be produced. As with the SAMPLE AREA card, a single MAP AREA card can define an area as large as the tape itself (1/4 of an LANDSAT frame) or anything smaller. However, the output will be in 120 column strips. The format for the MAP AREA card will be the same as for the SAMPLE AREA card EXCEPT that:

1. A '1' is also punched in column 24 (optional unless a '1' has been punched in column 20).

2. A '1' in column 28 indicates that the user has inserted a 'LIMITS' card after that MAP AREA card.

LIMITS Card

The LIMITS card enables the user to specify the class limits to be used for a particular map area, regardless of the values computed by the program. The values established by a LIMITS card will continue to be used for succeeding map areas until the next LIMITS card is read.

The values to be punched on the LIMITS card will be the upper boundaries of the k grey-scale divisions. They are to be punched in consecutive four digit integer fields, from smallest to largest.

JCL and Control Card Sequence

for Program WMAP in USDA Washington Computer Center

Label parameter is for non-labeled LANDSAT tape

```
/*
as many additional SAMPLE and/or MAP area cards as desired
initial SAMPLE area card, may also be a MAP area card
a CLASS card
//GO.SYSIN DD *
//GO.FT10F001 DD SYSOUT=(c,,8431),DCB=RECFM=FBA
//G(.FT09F001 DD SYSOUT=(c,,8431),DCB=RECFM=FBA
//          BLKSIZE=3320),LABEL=(,NL,,IN),BOL=SER=  6   ,DSN=
//GO.FT08F001 DD UNIT=2400,DISP=(OLD,PASS),DCB=(RECFM=U,
//          EXEC RADLGO,P=RADMAP
    job cards
```

APPENDIX D

Detailed Instructions for

Microdensitometer Scanning of Aerial Photography

1. Load aerial photography on manual film transport and locate frame containing desired segment.

2. Be certain there is a frame gap (gap between two adjacent frames) within the stage travel limits. Calibration is to be performed on this gap.

3. Rotate the stage to align the scanning axes with section lines and/or major roads.

4. Determine the easternmost point, relative to the stage, that is to be scanned. Move the stage until the verticalcross hair on the viewing screen is aligned on that point. Set X to zero on the Digital Coordinate Readout System (DCRS).

5. Determine the northernmost point, relative to the stage, that is to be scanned. Move the stage until the horizontal crosshair on the viewing screen is aligned on that point. Set Y to zero on the DCRS.

6. Advance teletype paper to beginning of next page. With teletype on "LOCAL," enter segment number and the words "corner coordinates."

7. Return teletype on "ON LINE" and enter "E" command.

8. Record field boundary coordinates as follows:

   a. Sequentially number all corner points of fields and other field boundary points as needed on the sketch of the segment.

   b. With the microdensitometer in "AUTO" operation model and the stage control motors off, manually move the stage to field boundary point 1. Press "INIT" button to record the coordinates of that point on the teletype terminal.

   c. Do (B) for every field boundary point sequentially.

   d. With teletype on "LOCAL," record the time required to record field boundary coordinates.

9. Advance teletype paper to beginning of next page. With teletype on "LOCAL" record the segment number, Julian date of photography, and magnetic tape file numbers for the next 8 files.

10. Locate the lightest area on the clock in the margin between frames contained in the frame gap and record the coordinates of this point. This will be the calibration point for all scans on this segment.

11. Enter "U1" command after teletype has been placed in "ON LINE" mode. The "U1" command allows users to define scanning parameters after being promoted by the computer as follows:

Prompt	User Response
XDIR	-
YDIR	-
DELTAX	240ƀ
XTRAV	x
YSTEP	240ƀ
NO. SCANS	y
SCAN TYPE	R
SPEED	255ƀ
BBACKUP?	N

where ƀ = space bar

$x = \max x_i$

$y = \frac{\max y_1}{240} + 1$

Where $(x_i y_i)$ are field boundary coordinates displayed on the DCRS in microns.

12. Enter "I" command on teletype to record identification information for this segment in the following format:

Col. 1 - -"C", "R", "G", or "B" corresponding to color filter in use.

Col. 2-5 - segment identification number

Col. 6-8 - three letter state abbreviation

Col. 9-11 - Julian date of photography

Col. 12-14 - "DEN" or "TRA" corresponding to scanning mode (density or percent transmission)

13. Perform analog system calibration (Section 4.5.3.3, Microdensitometer Operation Manual, P. 4-18) at the calibration point (See step 10).

14. Enter "CS" command on teletype. Make sure stage control motors are turned on.

15. Change color filter and repeat steps 12-14 until the segment has been scanned with all four color filters.

16. With teletype on "LOCAL," record time required to scan the segment.

APPENDIX E

A PROGRAM TO CONVERT PDS MICRODENSITOMETER SCAN LINES INTO SAS COMPATIBLE OBSERVATIONS

This program is designed to convert a Photometric Data System (PDS) microdensitometer scan into a Statistical Analysis System (SAS) compatible multivariate observation. Up to 4 scans of the same area may be included in the SAS observation.

The user controls the number of scans (normally 1 for each filter) to be used in building the multivariate observations. The microdensitometer scans are read in serially and saved on temporary files. After all the data for a given picture section (pisect) has been read in, the temporary files are rewound and read back a line at a time, and a SAS observation produced for each point in the line. Each observation consists of data from corresponding points from all scans used.

The program is divided into 3 phases: (1) parameter phase, (2) read phase, and (3) combine phase. The normal operation of the program is to go from phase 1, to phase 2, to phase 3, and repeat as desired.

Parameter phase:

Allows the user to define the initial settings from all counters, and indicators used during the read and combine phases. If fatal errors occur during the run, control reverts to the parameter phase for an error scan of all remaining control cards, but no data will be processed.

Read phase:

During the read phase, microdensitometer scans are read in and stored on temporary files. During this process, the PDS 9-track format is converted to a 8 bit internal IBM notation. If the data were scanned in a raster or right edge scan, it would be converted to a left edge scan. The user, however, may elect to cancel this option and accept the data in the order scanned. While in read phase, all parameter definition cards are ignored. If an attempt is made to read more than 4 scans, the combine phase is automatically entered.

Combine phase:

This phase combines the results of the read phase. Corresponding points from each read file are included in each SAS observation produced. The data from the reads are put in correspondence with the data items in the SAS observation set. If these are fewer than 4 scans to be combined, the trailing data items are assigned the missing value. The coordinate values and pixel serial numbers are computed and assigned as each observation is produced. At the conclusion of this phase, control reverts to the parameter phase, and new parameter settings will be accepted.

NUMERIC VALUE REPRESENTATION

The microdensitometer output is a digital representation of an analog signal. The amount of light passing through a sample is converted into a voltage by a photo-multiplier tube. If transmissions are being recorded, the voltage is routed to the panel display meter and then to the analog to digital (A/D) converter. If optical densities are being recorded, the voltage is first sent to a logarithmic converter before going to the panel display meter and then to the A/D converter.

The A/D converter produces a positive integer value that represents the voltage. The input range of the A/D converter is 0.00 to 5.12 volts in .005 volt increments. The digital output ranges from 0 to 1024, or 200 times the voltage input. It is important to remember that these values could be either transmission or density depending on the calibration settings.

When the digital output from the A/D converter is stored in the computer (PDP8) it is multiplied by 2 and is now 400 times the value shown on the panel meter. This is done to reduce the effect of noise contamination. Some noise could result from the fact that the microdensitometer actually takes discrete readings from a continuously varying function.

The data values are recorded in a 9-track tape format. The PDP8 computer is a 12 bit word machine with 6 bit bytes and is not directly compatible with the 9-track 8 bit byte tape format. Therefore, 2 zero pad bits are appended to each PDP8 byte as it is written in a 9-track format. Physically, the data on tape has the format shown below:

ppsddddddppddddddn

where p represents the pad bits appended to fill the 9-track tape format,
s is the PDP8 sign bit and is normally 0,
d represents one of the 10 data bits from the A/D converter,
n represents the noise bit position, normally 0.

In reconstructing the microdensitometer data back into a useable form, the program allows the user two choices. By default, values will be produced from storage type data. Optionally, actual panel display values may be generated.

Storage data has been reduced to a form which is suitable for bulk storage. Each value is reduced to an 8 bit integer and requires exactly 1 byte of storage. This is the form used by ERTS, LARSYS, and the Penn State Classification System.

The numeric range of the integer valued data is from 0 to 255. Approximate panel values may be derived by multiplying a storage value by .02. At first, it may seem that we are discarding valid data, but this is not so if we consider the accuracy of the microdensitometer.

The microdensitometer specifies linearity of ±.02 density or .5% transmission, and that the drift for a 10 hour period is less than ±.02 density or less than 1% transmission. This means that a recorded value could differ from the true value by as much as .04 density or 1.5% transmission. The stored values will resolve density to the nearest .02 units and transmission to the nearest .4% (.3921569), which is within the limits of the equipment.

The PanelData option allows the reconstruction of exact panel readings as shown by the panel display meter. The data accuracy implied is beyond the capability of equipment, but is should be useful in checking machine specifications.

X Y COORDINATE SYSTEM

The program assumes a generalized coordinate reference system. The x,y coordinates are signed integers, with (0,0) as the default origin. The x ordinate is the element index, and the y ordinate is the line index. The program always assigns the algebraically smallest x,y value to the pixel in the northwest corner (upper left). The x ordinate increases as the scan moves to the east (right), and the y ordinate increases as the lines move south (down).

The PDS microdensitometer normally scans lines in a raster (back & forth) with the direction of scan alternating, and can scan lines from top to bottom or bottom to top. The Photometric Data System Conversion to Microdensitometer Scan (PDSCMS) program has the ability to determine the scanning directions, and use this in the coordinate assignment algorithm. Thus, regardless of how the points are scanned, the above defined coordinate reference system is valid.

The program computes the coordinates during the combine phase. The coordinates of the physically first point are computed and assigned to that point. If this point is not the northwest corner point, the coordinate of the northwest corner point are derived. The program prints out the northwest corner coordinates as the first x and y ordinates.

The above described coordinate reference system may seem unduly complicated, but it (1) sets up a reference system that is both hardware and software compatible, and (2) permits full use of the microdensitometer scanning ability.

Display devices such as line printers and CRD devices, display data from left to right and top to bottom. The natural order of computer indexing is from smallest to highest. Thus, after coordinates are assigned. data points may be sorted by coordinate and they will be in the natural order for computer processing regardless of how scanned.

The user may have several scans from a scene with the microdensitometer defining the origin at each pisect. The conversion software would call that point (0,0) by default. Later, the user may wish to restore or assign relative position of pisects by relocation. The user could also move the origins of all pisects from the microdensitometer (0,0) setting to any arbitrary point (n,n).

The user may have the microdensitometer scan several pisects from a scene relative to a common origin. The conversion software will compute initial coordinates for each pisect using the microdensitometer supplied locations. Thus, the resulting pixel coordinate will preserve the relative spatial location of the pisects relative to the scene origin. Later, the user may wish to perform an origin transformation, and spatially locate this scene relative to any other independently scanned scene.

SAS OBSERVATIONS

Each observation produced has 11 items as follows:

SCENE-NAME — 1-8 characters left justified with trailing blanks in bytes 5-12.

This name is used to identify a collection of pisects (picture sections). If the user fails to supply a valid name, the program will use the current date in the form mm/dd/yy by default.

PISECT-NAME — 1-8 characters left justified with trailing blanks in bytes 13-20.

This name is used to identify a pisect within a scene. A new name is supplied for each pisect processed. If the user fails to supply a valid name, the program will use the current value of the system clock in the form hh•mm•ss by default.

GROUP-NAME — 1-8 characters left justified with trailing blanks in bytes 21-28.

This name is used to identify calibration data. A null or 'blank' name indicates unknown data. The discriminant function, uses named groups as training, and classifies unknown data. If the user fails to supply a valid name, the program supplies the null or 'blank' name by default.

IDENT-NAME — 1-8 characters left justified with trailing blanks in bytes 29-36.

This name is used to establish user identity of unknown data. A null or 'blank' name indicates that the user does not know or cannot identify the item. Valid ident-names are taken from the set of group names. The discriminate function would use the ident-name to check classification accuracy. Of the user fails to provide a valid name, the program supplies the null or blank' name by default.

XORD — integer binary in bytes 37-40.

This is the relative position of the SAS observation within a line of data. It always gives relative element position within its own pisect, and depending on user options may be positional relative to an entire scene or group of scenes.

YORD — integer binary in bytes 41-44.

This is the relative line position of the SAS observation. It always gives relative line position within its own pisect, and depending on user options may be positional relative to an entire scene or group of scenes.

PSN integer binary in bytes 45-48.

This is the pixel serial number assigned by the program. Pixels are serialized in order processed in the combine phase. Unless directed otherwise, pixels are serialized for the entire run starting with 1. The serial number may be signed.

PIXF1V real binary in bytes 49-52.

This is the microdensitometer value for the first scan read for the current pisect. It will never be assigned the missing value. 1/

PIXF2V real binary in bytes 53-56.

This is the microdensitometer value for the second scan read in for the current pisect. If there was no second scan, it takes on the missing value.

PIXF3V real binary in bytes 57-60.

This is the microdensitometer value for the third scan read in for current pisect. If there was no third scan, it takes on the missing value.

PIXF45V real binary in bytes 61-64.

This is the microdensitometer value for the fourth scan read in for the current pisect. If there was no fourth scan, it takes on the missing value.

The program writes the SAS compatible file in binary (unformatted) variable blocked spanned mode. (RECFM=VBS). Because SAS includes the record description word as part of the record, the byt locations of all items have been offset by 4 bytes in the above description.

1/ The missing value is a floating point -0, or in hexidecimal 80000000.

CONTROL CARDS

The program uses 14 different control cards. Most of them are optional because the program will supply default values when the user does not. Each control card is divided into 3 major fields as follows: (1) Key word or op-code in columns 1-8; (2) parameter field in columns 11-50; and, (3) comments field in columns 51-80.

There are 4 classes of control cards, depending on the kind of action to be performed. Each class is described separately below:

Class 1 - Run Cards

These cards set indicators that remain in effect for the duration of the run or until redefined during the run. All run cards are optional.

EDGE Card

cols 1-8 EDGE

This card causes the program to convert to all scans to a left edge scan. This effectively removes the raster produced by the back and forth microdensitometer scanning motion. All lines are running from right to left are turned around. If an EDGE card is not supplied, it is assumed.

ASIS Card

cols 1-8 ASIS

This card causes the program to accept the data points in the order scanned. However, the x,y coordinate assigned are computed based on line direction. If the pixels are sorted based on the x,y coordinates, a normal picture will be produced. That is, the true northwest corner point has the algebraically smallest coordinates, and the southeast corner has the algebraically largest coordinates. If an ASIS card is not supplied, EDGE is assumed by default.

ABL Card

cols 1-8 ABL

This card causes the program to accept microdensitometer data sets that have identified with blank or first character blank labels. By default such scans are rejected as a fatal error. Note that once turned on this option cannot be rescinded during a computer run.

VALUE Card

cols 1-8 VALUE

cols 11-18 STORAGE / PANEL

This card allows the user to select the type of numeric values to produce for the SAS file. Storage values are normalized floating point integers, range 0 $\leq$ value 255. Panel values are also normalized floating point, but is the microdensitometer A/D converter output expressed as a display panel number. The range is 0.000 $\leq$ values $\leq$ 5.115, in increments of .005. A storage value is numerically 50 times the panel value with the decimal fraction truncated.

Class 2 - Scene Cards

These cards set parameters that apply only to the scene about to be processed. They are automatically cleared to default values after a STACK control card. All scene cards are optional.

SCENE Card

cols 1-8 SCENE

cols 11-18 1-8 character name left justified with trailing blanks used to identify a group of pisects. The contents of columns 11-18 are placed in the scene-name field of the SAS compatible record. If the user does not make a scene, the program supplies the current date by default.

PSN Card

cols 1-8 PSN

cols 11-15 signed integer constant starting serial number.

This card can be used to extend the serialization of previous computer runs. If the user does not supply a starting serial number, a value of 1 will be used by default. The STACK control card resets PSN to 1.

ORIGIN Card

cols 1-8 ORIGIN

cols 11-15 signed integer constant x coordinate offset.

cols 16-20 signed integer constant y coordinate offset.

This control card is used to provide origin translation of each pisect processed. The coordinates of the first point are computed and the offset applied. It may be used to relate the pisects from the current scene to those in a previous or subsequent scene. This feature may be useful when the data are from sequential scenes such as aircraft photography.

Class 3 - Pisect Cards

These cards set parameters that apply only to the pisect about to be processed. They are automatically cleared to default values after the COMBINE or STACK control card. All pisect cards are optional.

PISECT Card

cols 1-8 PISECT

cols 11-18 1-8 character name left justified with trailing blanks.

The contents of columns 11-18 are saved in the pisect name in the SAS Compatible Record. It serves to identify pisects within scenes. If the user does not supply a PISECT card, the program uses the current value of the system clock by default.

GROUP Card

cols 1-8 GROUP

cols 11-18 1-8 character name left justified with trailing blanks.

The contents of columns 11-18 are placed on the group field in the SAS Compatible Record. A non-blank name indicates that this pisect contains calibration data for a specific groups. If the user does not supply a group name, the program inserts a blank name by default.

IDENT Card

cols 1-8 IDENT

cols 11-18 1-8 character name left justified with trailing blanks.

The contents of columns 11-18 are placed in the ident-name field of the SAS Compatible Record. A non-blank name indicates that the user has identified the points in this pisect as belonging to the specified group. If the user does not supply an IDENT, the program inserts blanks by default.

RELOCATE Card

cols 1-8 RELOCATE

cols 11-15 signed integer constant representing the northwest x ordinate.

cols 16-20 signed integer constant representing the northwest y ordinate.

The northwest corner pixel will be assigned the coordinates given on this card. All subsequent pixels will be assigned coordinates relative to these. Thus, any pisect can be arbitrarily moved in space. By default, absolute relocation will not be performed.

This card overrides the origin transformation in effect for each pisect for which relocation is performed. The origin transformation will be performed for each pisect not relocated.

Class 4 - File Manipulation Cards

These control cards cause data to be moved from one file to another, and to perform some transformations in the process. These cards are required as specified below.

READ Card

cols 1-8 READ

cols 11-50 1-40 character name left justified with trailing blanks.

This card causes the program to read in 1 PDS microdensitometer scan, stored on a temporary file. One read card is required for each scan to be included in a SAS observation. When a read card is processed, while the program is in the parameter phase, control is switched to the read phase. No more parameter cards will be honored until control reverts back to the parameter phase.

Up to 4 consecutive read cards will be honored. If a 5th read card is encountered, the program will combine the 4 scans already stored on temporary files, and then scan the remaining control cards for errors. No more data will be transferred. Either an end-of-file, a combine card, or a stack card must follow read cards.

The 1-40 character name is used for label checking as follows:

(1) If the name is absent or begins with a blank, the program assumes that no label checking is to be performed, and whatever file it finds is assumed to be correct.

(2) If a name is present, it must match the label put in the scan line by the microdensitometer operator. Label checking is performed up to the first blank character in the supplied name. Thus, if the user has a common prefix for a series of scans, he may use an abbreviated label to verify that the correct scans are being processed. If the label check fails, no more files are processed, but the remaining control cards are checked for errors.

COMBINE Card

cols 1-8 COMBINE

This card causes the program to combine the results of the previous reads and add the results to the SAS compatible data set being built. If n scans are being combined, exactly n-1 combine cards are required. The last combine card in the control card stream is optional as any uncombined reads are automatically combined at end-of-file. At the end of a combine operation, the program returns to the parameter phase and will accept parameter control cards.

STACK Card

cols 1-8 STACK

This card is the same as combine in that the results of the previous reads are combined and concatenated to the SAS compatible data set being built. In addition, the data set is endfiled and the scene and pisect indicators cleared to default values. Any control statements following a STACK control card cause PDSCMS to start a new SAS compatible file. This new file may be stacked or separated, depending on the JCL used for the run.

Both STACK and COMBINE cards may be used in the same run, providing at least 1 read operation is performed between them. If a STACK card would be the physically last control card,it can be omitted.

EXECUTING THE PDSCMS PROGRAM

The PDSCMS program is executed by using the RADLGO procedure. The PDS microdensitometer tape is read from unit 8, and the converted file is written on unit 9. Program control cards are read from SYSIN.

The microdensitometer output is a series of stacked data sets on magnetic tape. The program reads as many data sets from the stack as directed by READ control cards by incrementing the unit 9 FORTRAN Sequence Number. Each READ control card requires a unit 9 DD JCL statement with an appropriate sequence number. The data set sequence number in the labels parameter points to the particular scan to be processed by the READ command.

```
//FT08F001  DD  LABEL=(i,NL,,IN)        for first READ card
//FT08F002  DD  LABEL=(j,NL,,IN)        for second READ card
//FT08F003  DD  LABEL=(k,NL,,IN)        for third READ card
      .
      .
      .
//FT08Fnnn  DD  LABEL=(n,NL,,IN)        for nnn'th READ card
```

The letters i, j, k, nnn represent the data set sequence number of the tape file to be processed. They point to the i'th, j'th, k'th, and n'th data set respectively.

The converted SAS file is written on Unit 9 in FORTRAN binary (unformatted) mode as either a single data set or a series of stacked (separated) data sets. Stacking is performed by incrementing the Unit 9 FORTRAN Sequence Number. The DD statement parameters determine if stacking or separation is being performed.

```
//FT09F001  DD  LABEL=p                 initial output from PDSCMS
//FT09F002  DD  LABEL=q                 after the first STACK card
//FT09F003  DD  LABEL=r                 after the second STACK card
      .
      .
      .
//FT09Fmmm  DD  LABEL=s                 after the (mmm=1)'th STACK card.
```

The letters p, q, r, s represent the data set sequence number on the tape being produced. If the data sets were being written on disk, separated names would be required.

SAMPLE JCL USING TAPE INPUT & OUTPUT

```
//XO   EXEC   RADLGO,
//            P=PDSCMS
//GO.FT08F001  DD  DISP=OLD,UNIT=2400,DCB=(BLKSIZE=6400,RECFM=U,BUFNO=1),
//                 VOL=SER=URxxxx,
//                 LABEL=(i,NL,,IN)
//GO.FT08F002  DD  DISP=)LD,UNIT=2400,DCB=*,FT08F001,VOL=REF=*.FT08F001,
//                 LABEL=(j,NL,,IN)
               .                                                            .
               .  as many ft08fyyy dd statements as required: extra ones do no harm.
               .                                                            .
//GO.FT08Fnnn  DD  DISP=OLD,UNIT=2400,DCB=*,FT08F001,VOL=REF=*.FT08F001
//                 LABEL=(k,NL,,IN)
//GO.FT09F001  DD  DISP(,PASS),UNIT=2400,DCB=(BLKSIZE=6400,LRECL=32000,RECFM=VBS,
                   BUFNO=1),
//                 DSN=dsname,
//                 LABEL=(p,,,OUT)
//GO.FT09F002  DD  DISP=(OLD,PASS),UNIT=2400,DCB=*.FT09F001,VOL=REF=*,FT09F001,
//                 DSN=*.FT09F001,
//                 LABEL=(q,,,OUT)
               .                                                            .
               .  as many FT09Fyyy dd statements as required: extra ones do no harm.
               .                                                            .
//GO.FT09Fmmm  DD  DISP=(OLD,PASS),UNIT=2400,DCB=*.FT09F001,VOL=REF=*.FT09F001,
//                 DSN=*.FT09F001
//                 LABEL=(r,,,OUT)
      PDSCMS       control cards
/*   EOJ.
```

SAS PROCESSING THE COMPATIBLE FILE

JCL Requirements

In order to process the compatible file with the SAS program, an additional DD statement is required by the RADSAS procedure. This statement is required to point to the file to be used. In the following JCL, the PDFSILE 1/ DD statement is used to gain access to the converted PDS data.

```
//S  EXEC  RADSAS
//PDSFILE  DD  DSN=dsname,DISP=OLD,UNIT=2400,VOL=SER=xxxxxx
//SYSIN  DD  *
      .
      .  sas program statements
      .
/*  EOJ.
```

In the above example, the converted file is assumed to reside on magnetic tape as a single unstacked data set. If the file is not on magnetic tape, or is passed from a previous job step, an appropriate alteration in the PDSFILE DD statement will be required.

If the stack option has been used to stack or separate scan pictures, a separate DD statement is required for each stacked data set to be read in during a given SAS run. If the data sets are stacked on tape, extra DD statements may be left in the job stream whether needed or not. The following JCL illustrates the set up the stacked data sets on tape.

```
//ST  EXEC  RADSAS
//STACK1  DD  DISP=OLD,UNIT=2400,
//            DSN=dsname,VOL=(,RETAIN,SER=xxxxxx),
//            LABEL=p
//STACK2  DD  DISP=OLD,UNIT=2400,DSN=*.STACK1,VOL=(,RETAIN,REF=*.STACK1),
//            LABEL=q
        :as many stack DD statements as may be needed: extra ones do no harm.
//STACKn  DD  DISP=OLD,UNIT=2400,DSN=*.STACK1,VOL=(,RETAIN,REF=*.STACK1),
//            LABEL=r
//SYSIN  DD  *
        :sas program statements
        .
/*  EOJ.
```

1/ The user may substitute any name for PDSFILE, but that name must also be used in the SAS INPUT statement.

If the converted files are separated on disk, a file must exist for each DD statement in the SAS step. If both PDSCMS and SAS are executed in the same job, SAS DD statements may point to PDSCMS DD statements that were not used in the PDSCMS step. However, if the SAS program is run as a separate job, all the converted files referred to by DD statements must actually exist in order to prevent JCL errors.

SAS Program Statements

The SAS program must be directed to use the PDSFILE DD Statement for its input. The model statements given below can be used to read in all the items from the converted file.

```
DATA;
INPUT  DDNAME=PDSFILE  SCENE $ 5-12  PISECT $ 13-20  GROUP $ 21-28
IDENT $ 29-36  XORD  IB 37-40  YORD  IB 41-44  PSN  IB  45-48
PIXF1V  RB  49-52  PIXF2V  RB 53-56  PIXF3V  RB 57-60
```

The user may not wish to read in all the items. Those items not wanted may be omitted from the list in the input statement. The following statement shows how to read in only the data from the first and third read cards.

```
DATA;
INPUT  DDNAME=PDSFILE  PIXF1V  RB 49-52  PIXF3V  RB 57-60; PIXF4V RB 61-64;
```

In order to read stacked or separated data sets in to the SAS system, the user must provide a separate INPUT statement for each separated file. Each data set referred to by the INPUT processor must actually exist. Data sets that do not exist or have never been created cause SAS to abend.

The following example illustrates a simplified method of reading redundant type data sets by using a SAS macro.

```
MACRO  WHATEVER  SCENE  $  5-12  PISECT $ 13-20  GROUP $ 21-28
IDENT $ 29-36  XORD  IB 37-40  YORD  IB 41-44  PSN  IB 45-48
PIXF1V  RB 49-52  PIXF2V  RB 53-56  PIXF3V  RB 57-60  PIXF4V RB 61-64%
```

(other SAS statements could also be included in the macro to perform special transformations, range checks, etc.)

```
DATA  STK1;  INPUT  DDNAME=STACK1  WHATEVER;
DATA  STK2;  INPUT  DDNAME=STACK2  WHATEVER;
    .
    .  as many statements as required:  extra ones must be removed.
    .
DATA  STKn;  INPUT  DDNAME=STACKn  WHATEVER;
```

The PDSCMS program assigns the missing value to the PIXFiV elements for which there was no corresponding read card. The user can do 1 of 4 things with missing value: (1) accept data with missing values and let SAS handle them, (2) do not read in the pixel filter values that are missing, (3) convert the missing value to some neutral value, or (4) identify and take special action for missing items.

Sample Program to Convert Missing Values

```
PIXF2V=PIXF2V+0;
PIXF3V=PIXF3V+0;
PIXF4V=PIXF4V+0;
```

DATA CONVERSION

Microdensitometer data is expected to be used from a storage format which is an 8 bit integer value from 0 to 255 inclusive. Storage data can either represent densities (logarithmic response), or transmission (linear response). Simple linear transformations are required to reduce storage values into the corresponding panel meter value, optical density, or percent transmission.

Storage values can be converted directly into corresponding panel meter values by multiplying by .02. 1/ The resultant is either an optical density or transmission value, depending on the microdensitometer calibration settings when the scan was performed.

When the microdensitometer is calibrated to record densities, the panel value is optical density. Storage values are increments of .02 density units with a valid range from 0.00 to 4.00 inclusive. Density readings larger than 4.00 constitute an overflow condition because they are beyond the specified range of the equipment.

When the microdensitometer is recording transmissions, the stored data represents an incremental percent transmission that is dependent on the gain setting during calibration. Normally, the gain is set at 5.10 to give maximum range and accuracy to the transmission levels. The incremental step is then .3921569% transmission.

In addition, it may be useful to convert the storage data into, from logarithmic densities into linear transmissions and vice versa. In the following relationships, the transmission calibration (Gain) is assumed to be 5.10. The density is always calibrated to 0.

The following symbols are used in the equations that follow.

SD density (logarithmic) storage value	$0 \leq SD \leq 200$
ST Transmission (linear) storage value	$0 \leq ST \leq 255$
G Gain setting for transmission	nominal value 5.10
PT Percent transmission	$0 \leq PT \leq 100$
OD Optical density	$0 \leq OD \leq 4.00$

1/

Described in the numeric representation section.

The relationship between optical density and transmission is:

$$\text{Density} = -\text{Log}_n\ (1/\text{Transmission})$$

If we impose on this basic relationship, the requirement that 100% transmission is 0 density and 0% transmission is 4.00 density, the equation can be rewritten as:

$$OD = 2 = \log_{10}(PT)$$

$$PT = 10 ** (2 - OD)$$

Note that the relationship of 0% transmission = 4.00 optical density requires a mathematical impossibility, namely $\text{Log}_{10}(0) = -2$, and $10^{-2} = 0$. These conditions are definitional and are imposed by the resolution limits of the electronic circuiting in the microdensitometer. During computer processing this limiting point requires special handling. Computationally, the valid conversion ranges for percent transmission and optical density are:

$$0 < PT \leq 100$$

$$4.00 > OD \geq 0$$

Also, be aware that 4.00 optical density can be transformed into the computationally valid percent transmission value .01. If storage transmissions are being produced, the minimum storage value is .39% and is larger than 01. An attempt to produce a storage value for .0% transmission will result in a 0 value.

Because in the density to transmission, computations can be performed over the entire density range, it is possible to computationally extend the valid transmission range beyond 2.3 optical density. An image is digitized in densities and the corresponding percent transmission computed. Thus, a percent transmission values less than .30, can be used in computations, but cannot be produced by the microdensitometer, nor stored in standard form.

The equation to convert stored density data into optical density is:

$$OD = SD * .02$$

The equation to convert stored transmission data into percent transmission is:

$PD = ST * .3921569$	when $G = 5.10$
$PT = ST * (2/G)$	$0 < G < 5.10$

The following transformations are used to convert logarithmic values into linear values and vice versa.

To convert stored density into percent transmission use:

$$PT = 10 ** (2 - SD * .02)$$

To convert stored density into stored transmission use:

$$ST = 10 ** (2.40654 - (SD * .02)) \quad G = 5.10 \text{ implied}$$

To convert Optical Density into stored transmission use:

$$ST = 10 ** (2.40654 - OD) \qquad G = 5.10 \text{ implied}$$

To convert stored transmission into optical density use:

$$SD = (2 - \log_{10} (ST * .3921569)) * 50 \quad G = 5.10$$

$$SD = (2 - \log_{10} (ST * (2/G))) * 50 \quad 0 < G < 5.10$$

To convert percent transmission into stored density use:

$$SD = (2 - \log_{10} (PT)) * 50$$

APPENDIX F

Field Extraction Program

Version 1 and 2

Introduction:

This program generates SAS program statements and control cards for PDSCMS to facilitate conversion and identification of microdensitometer data into final form suitable for discriminant analysis. It is a special purpose program with few options and little in the way of error checking. It is the user's responsibility to make certain the input data is in the correct form, as described in the input section. There are two versions of the program. The major difference between the two versions is the input required for each. Thus, the input section of this paper is divided into two sections, one for version 1, and the other for version 2. Any other differences between versions will be noted in the appropriate sections. The output from the two versions is identical.

JCL Requirements:

```
//jobcard
/*ROUTE PUNCH LOCAL
//   EXEC  RADGO,
//   P=RSFEP1              THIS CARD FOR VERSION 1.
//   P=RSFEP2              THIS CARD FOR VERSION 2.
//GO.FT08F001  DD  SYSOUT=B
//GO.FT09F001  DD  SYSOUT=A
//GO.SYSIN  DD  *

        {input cards}

/*  EOJ.
```

Output: Output is routed through logical units 6, 7, 8, and 9. Logical units 6 and 9 are for printed output, units 7 and 8 for punched output. The printed output on unit 6 consists of job processing information and images of PDSCMS control cards. The PDSCMS control cards are punched from unit 7. SAS program statements for field extraction are punched from unit 8 and printed from unit 9.

PDSCMS Control Cards:

1. SCENE state name (8 character maximum)

2. PISECT segment number (4 characters)

3. READ label for clear filter

4. READ label for red filter

5. READ label for green filter

6. READ label for blue filter

7. STACK or COMBINE — see PDSCMS 2.1.0 for effects of each.

Two sets of control cards are punched for each segment; one set for data scanned in the density mode, one set for data scanned in the transmission mode. The labels on the READ cards will match the identification label on the microdensitometer tape only if those identification labels are in the following form:

col 1: 'C', 'R', 'G', or 'b' corresponding to the filter in use.
col 2-5: segment number
col 6-8: first three characters of state name
col 9-11: julian date of photography
col 12-14: 'DEN' or 'TRA' corresponding to scanning mode
col 15-40: any other information desired by the user

Input: The first data card is the same for both versions. It must be in the following form:

col 1: blank
col 2-23: &NUMBER OFSEGS=xx,&END

where xx - the number of segments to be processed.
For each segment, the following cards are required:

VERSION 1:

1. State, segment number, and STACK or COMBINE

Col. 1-12: state name, left justified.

The first three characters are used to create label information for READ cards for PDSCMS. The state name is also output as the SCENE identifier for PDSCMS.

Col. 15-18: segment number.

Used for label information for READ cards, and output as PISECT identifier.

Col. 21-28: STACK or COMBINE (See PDSCMS 2.1.0 for effect of each).

2. Scanning information in the following form:

Col 1: blank
Col 2-54: &INFO PHOTO=www,DELTAX=,xxx,DELTAY=yyy,NOFLDS=zzz &END

where www - julian date of photography
xxx - delta x for scanning
yyy - y step for scanning
zzz - number of fields in this segment to be processed. This must equal the number of FID cards for the segment.

3. FID cards: corner coordinates, tract, field and crop identifiers.

col 1-3: FID

col 5-11: x_1

col 12-18: y_1

col 19-25: x_2

col 26-32: y_2

col 33-39: x_3

col 40-46: y_3

col 47-53: x_4

col 54-60: y_4

where (x_1,Y_1)=N.E. corner of field

(x_2,y_2)=N.W. corner of field

(x_3,y_3)=S.E. corner of field

(x_4,y_4)=S.W. corner of field

col 61-62: two digit integer corresponding to tract identification.

col 63-64: two digit integer corresponding to field identification.

col 65-72: 8 character crop identifier.

col 73-80: 8 character crop identifier

The effect of each FID card is to create a SAS program statement which will append tract, field, and crop identifiers to <u>most</u> data points within the quadrangle specified by the corner coordinates on the FID card. Not all points will be identified since boundary points are deleted and the program operates only on rectangular areas parallel to the scanning axes which are contained within the specified quadrangle. The assumptions are also made:

1. $|\min (x_2,x_4)| \quad |\max (x_1, x_3)|$

2. $|\min (y_3,y_4)| \quad |\max (y_1, y_2)|$

3. (x_i,y_i) i = 1, 2, 3, 4, are measured in microns.

4. No origin offset will be used in PDSCMS.

Restriction number 3 on the preceeding page can be bypassed. If (x_i, y_i) are in pixel coordinates as produced by PDSCMS, then specify DELTAX=1, DELTAY=1, on the scanning information card, rather than their true values. Irregular fields (non-rectangular) may be split by the user into two or more rectangular fields parallel to the scanning axes in order for the maximum number of points in the field to be identified.

<u>Version 2</u>:

1. State, segment number, and STACK or COMBINE in same format as Version 1.

2. Scanning information in the following form:

   col 1: blank

   col 2-64: &INFO PHOTO=www,DELTAX=xxx,DELTAY=yyy,NOFLDS=zzz, NOPNTS=ttt,&END

   where www, xxx, yyy, zzz are as defined in Version 1, and ttt is the number of corner points in the segment.

3. Coordinates for each field corner point in the segment.

   col 4-10: x_i

   col 14-20: y_i

   col 24-27: i

   where (x_i, y_i) is the ith corner point in the segment. These cards must be in order from the smallest to largest i, where i = 1,2,3,...,n

4. SFID cards: subscript of corner points, tract, field and crop identifiers.

   col 1-4: SFID
   col 11-13: i
   col 16-18: j
   col 21-23: k
   col 26-28: l
   (where (x_i, y_i) = N.W. corner, (x_j, y_j) = N.E. corner, (x_k, y_k) = S.W. corner, (x_l, y_l) = S.E. corner)
   col 31-32: integer tract identifier
   col 35-36: integer field identifier
   col 39-46: eight character crop identifier
   col 49-56: eight character crop identifier

Implementation of Version 2 considerably reduces setup time for scanning, and time required to record field identification for keypunching. By entering the E command on the microdensitometer, then positioning the stage at a field corner point and depressing the PROG INIT button, the coordinates of that point are printed out on the teletype. Field corner point coordinates can then be keypunched directly from the teletype printout. On the sketch of the segment, it isno longer necessary to record the coordinates for that point, merely record the subscript for that point. Then on the SFID keypunch form,it isonly necessary to record the subscript for each corner point, not the full set of coordinates. This should reduce the man-hours required for each of these steps by better than 50%.