PURDUE UNIVERSITY

Laboratory for Applications of Remote Sensing

Lafayette, Indiana

3RD QUARTERLY PROGRESS REPORT

U.S. DEPARTMENT OF AGRICULTURE

NASA PURCHASE REQUEST W12-996

Work Completed Under
Contract #12-14-100-9549(20)
at the
Laboratory for Applications of Remote Sensing
April 1, 1969 to June 30, 1969

Agricultural Applications and Requirements Programs

APPLICATION DISCUSSIONS

Discussions were held during this quarter with numerous scientists in various disciplines of the natural sciences. Interest centered around the potential use of LARS techniques and facilities for research in geology, limnology, meteorology, ecology, geography, and civil engineering.

Soil scientists associated with LARS and the Purdue University Departmen of Agronomy held informal discussions with soil scientists of the Indiana Office of the Soil Conservation Service (USDA) concerning methods of evaluating the use of multispectral data in surveying and mapping soils. Several factors were proposed for consideration in evaluating the utility of computer printouts which separate soils according to the spectral properties of the surface. It is anticipated that various techniques of evaluation will be employed during the months ahead.

INTERNATIONAL MEETINGS AND EUROPEAN ITINERARY

Plans were made in preparation for two papers and a series of seminars to be presented in Europe in August and September. A paper entitled "Automatic Identification and Mapping of Earth Surface Features" will be presented at the VIII International Quaternary Congress in Paris. "Multispectral Response Studies of Soils" will be presented at the annual meetings of the German Soil Science Society in Hannover. An invitational seminar on "Potential Applications of Remote Multispectral Sensing to Soil Classification and Mapping" will be presented before the Land and Water Development Division of FAO in Rome.

Invitational lectures will also be given at the following institutions:

International Atomic Energy Agency
Vienna, Austria

International Institute for Aerial Survey
and Earth Sciences, Delft, Holland

Netherlands Soil Survey Institute
Wageningen, Holland

University of Zagreb
Zagreb, Yugoslavia

University of Belgrade
Belgrade, Yugoslavia

Agricultural Experiment Station
Braunschweig, West Germany

Christian-Albrechts-University
Kiel, West Germany

German Geological Survey
Hannover, West Germany

Greek Nuclear Research Center
Athens, Greece

Organization of Economic Cooperation
and Development
Paris, France

The objectives of this official trip to seven European countries by M. F. Baumgardner were:

(1) To present scientific papers on aerospace remote sensing research.

(2) To present to European scientists seminars and lectures on the use of aerospace remote sensing in natural resources research.

(3) To discuss with European scientists of many disciplines the potential applications of remote sensing and automatic data

processing techniques to the development and management of agriculture, water resources, and other earth resources.

(4) To observe and study in the several countries which were visited, the present systems of soil survey and mapping; land use planning; land drainage and reclamation and agricultural research in general.

Biophysical Research Programs

MAY FLIGHT MISSION

Two flight missions were conducted during this quarter which were during the months of May and June. On May 13 some data were collected by the University of Michigan aircraft, but weather conditions deteriorated and the mission was terminated before completion. The mission was completed on May 26 and 27.

With significant effort from LARS personnel the classification of the Flight Line 21 data collected on May 26th obtained at 1200 hours was completed. These data had been collected at 4000 feet altitude, and included 3 channels in the middle infrared wavelengths between 1.0 and 2.6 micrometers. Cover types present at this time of the year included water, bare soil, wheat, hay, trees, corn, corn stubble, oats, rye, and an additional category of green vegetation which included weeds, grasses and legumes. The primary objective for the classification was to classify wheat, other green vegetation, bare soil, and water. A secondary objective was to develop a procedure for analyzing a large quantity of data; in this case, a single flight line involving approximately 24 square miles.

Using gray-scale printouts in the .62-.66 micrometer band, 237 fields were delineated. A subset of these fields were designated as training fields. Preliminary work indicated that water, bare soil, and green vegetation could be easily separated, and that there was good potential for separating wheat from other green vegetation. Preliminary work also indicated that a problem existed in the overlay technique causing misregistration of the overlay data. This was adjusted and a correction was made which allowed the overlay to be much more accurate. Two classifications were conducted; the first utilized five wavelength bands and gave an overall performance for training fields of 94 percent and for test fields of 80 percent. This classification utilized the .46-.48, .55-.58, .72-.80, 1.0-1.4, 1.5-1.8 micrometer bands. The same data was then classified using three wavelength bands, which were .40-.44, 1.0-1.4, 2.0-2.6 micrometers. This classification resulted in approximately 1 percent poorer accuracy than with the five wavelength bands. The overall performance was 93 percent for training fields and 79 percent for test fields (Tables 1-4). In this case, it appeared that the increased time to obtain the classification over such a large area using five wavelength bands did not warrant the small increased accuracy.

Conclusions resulting from this effort are: (1) The resolution of single width printouts was inadequate for easy and positive location of field boundaries where data were collected at a height of 4000 feet or higher. Resolution of double width printouts was marginal. Displays of better spatial resolution will help this problem considerably. (2) Up-to-date photography is necessary when locating field boundaries in a flight line for research purposes. (3) The techniques for handling large quantities of data in a timely manner were considered satisfactory, although some improvements will be made. One of the greatest problems was the determination of the number of sub-categories of each class and the fields which should belong in each sub-category. (4) The classification was considered to be reasonably accurate.

When gathering ground truth it was difficult to differentiate rye, oats and wheat. Flight missions early in May or after June 15 would allow more positive identification of these species.

JUNE FLIGHT MISSION

The requested June flight mission was conducted on schedule on June 25 and 26. The weather was generally good and ground truth data was collected using the photography taken on May 25. Because the photography was obtained near flight time, LARS personnel were able to partially annotate the photography before the flight mission. This made the gathering of ground truth information considerably easier. The type of crop cover, plant height, row width and row direction in the case of corn and soybeans, and special descriptions such as "weedy" or "mowed" were written on the photographs. Ground truth collection was completed during the week of the flight. Flight Line 24 was selected for detailed analysis, digitized and reformatted.

The May photography obtained during flight missions with the K-17 camera by the University of Michigan had crabbing and tilting problems. When Michigan arrived for the June mission, it was requested that the crab angle be corrected if possible. This problem could be corrected for the K-17 camera but could not be corrected for the 70 millimeter camera systems.

Because the Michigan aircraft altimeter is accurate to only + 200 feet, it was decided that in future digitization work the altitude would be determined from the photography. This altitude information would determine the frequency of digitization of the data.

SOILS RESEARCH

1968 NASA photography was studied, with particular emphasis on Flight Lines 21 and 25 in Tippecanoe County, to select intensive soil study areas for 1969 flight missions. Soils representing transitional areas (Prairie soils, outwash soils, and forest soils) were chosen. The study areas include 5 farms in Tippecanoe which are: Chitwood Farm on Flight Lines 22 and 25, Hawkins farm on Flight Line 25, Banta-Gay Farm on Flight Line 21, Prather Farm on Flight Line 21, and Dieterle Farm on Flight Line 24. The detailed study areas range in size from 60 to 100 acres. Prior to the May flight, 4 to 8 samples of surface soils were obtained from each of the fields selected for soil study. Mechanical analysis and organic matter content were determined for each sample. At the time of the scanner mission, surface soil samples were obtained in each of the fields for moisture determination to study the effects of moisture on the surface spectral properties of soil. Preliminary analysis of one of the test areas indicated there seems to be a high correlation between organic content, as well as the moisture content of the soils, in relation to the spectral characteristics as measured by the scanner.

On the basis of these preliminary results, it was decided that more detailed, quantitative analyses of the geophysical relationships between the spectral characteristics of the surface soils and their soil properties were warranted and needed. Therefore, efforts were to be concentrated on Flight Line 24, and in particular, the Dieterle Farm. At the south end of this flight line, large areas of soils had developed under deciduous vegetation. These soils have a very distinct, natural pattern which is related to the internal and surface drainage characteristics. On the Dieterle Farm, an area of approximately

1200 by 1800 feet in size was divided into a grid of approximately 100 by 100 foot areas. At the center of each 100 foot square area, 1500 grams of surface soils were obtained for laboratory analysis and study. Dr. Al Zachery of the Agronomy Department prepared a detailed soil map of the Dieterle Farm and surrounding areas. Approximately 400 acres were mapped using the conventional techniques of walking over the area, observing surface features and drawing soil boundaries on an aerial black and white photo. With this method the soil surveyor must describe the changes in the soil profile as well as the surface characteristics. Borings were made in each soil mapping unit. In the next few months, detailed analyses of the soil samples and correlations between the soil characteristics and the spectral radiance as measured by the scanner will be conducted.

A cooperative effort was also initiated between the Soil Conservation Service (SCS) in Indianapolis and LARS. SCS is interested in exploring new and faster methods for obtaining and analyzing soil data for surveying and mapping purposes. However, SCS is not a research organization and cannot participate directly in research activities. SCS officials did agree to supply to LARS soil maps for five designated areas along the Highway 37 flight line. These maps were prepared and delivered to LARS during this quarter. They will be used for comparisons of conventional maps with those obtained by computerized mapping and soil boundary delineations of scanner data.

YELLOWSTONE DATA ANALYSIS

Effort continued on a cooperative project with United States Geologic Survey (USGS) in the analysis of the Yellowstone data. The first analysis of this data had been conducted in January when Dr. Harry Smedes and Mr. Kenneth Pierce (USGS, Denver, Colorado) visited LARS. Following this initial classification, Ken Pierce revised some of the training area coordinates. A new set of statistics was obtained and a classification was conducted for the entire flight line (double width printouts) on the Yellowstone test site by LARS personnel. A second classification was conducted using wavelength bands similating those proposed to be flown in the Earth Resources Technology Satellite (ERTS). A third classification was made utilizing wavelength bands simulating those proposed for the Return Beam Vidicon system.

As a continuation of this analysis effort, a joint project between Dr. Marc Tanguay of Ecole Polytechnique, Montreal, Canada; USGS and LARS was initiated. Dr. Tanguay came to LARS in June as a visiting scientist for a two month period. Early tasks included familiarization with the categories that had been defined previously and thorough examination of the photography in the test area. A great deal of emphasis was placed on refining some of the training samples and further examination of the statistical parameters. After a thorough photo interpretation study, it was determined that some of the categories should be redefined and a new classification done using a modified set of training classes. Additional examination of the statistical parameters is being conducted. It is anticipated that this will be followed by additional classifications using the simulated ERTS and RBV wavelength bands.

OTHER ANALYSIS WORK

The classification of the Black Hills forestry data and the Cunningham forestry data was completed during this quarter.

Physical Measurements

FIELD SPECTROMETER

The main accomplishment of this quarter was the completion of the work necessary to write a thesis proposal on a field spectrometer by John Clevenger. This included a signal-to-noise ratio study over the entire spectral range from 0.35 to 2.5 micrometers; laboratory and computational analysis of slit curvatures in various prism arrangements; electronic system design and preliminary considerations of calibration schemes. The work will be reported in a thesis proposal for the Ph.D. Preliminary Exam of John Clevenger in July.

LEAF SCATTERING STUDIES

Work was started on a paper describing the leaf-scattering studies completed by Harry Breece. Some additional experimentation will prove necessary to explain spectral response anomalies beyond 800 nanometers. Further, in order to make the apparatus generally useful to life scientists not specifically familiar with electronic instrumentation, it will be necessary to plan modification and repackaging of the apparatus.

INFORMATION PROCESSING SYMPOSIUM

Three papers were invited for the LARS portion of the Purdue information Processing Symposium to cover both aircraft and satellite scanner plans. The papers to be presented orally by the authors were:

"Physical Considerations in the Channel Selection and Design Specification of an Airborne Multispectral Scanner," C. Lawrence Korb, NASA MSC, Houston, Texas.

"Design Considerations for Aerospace Multispectral Scanning Systems," C. L. Wilson, M. M. Beilock, E. Zaitzeff, Bendix Corporation, Ann Arbor, Michigan.

"Design Considerations of A Multispectral Scanner for ERTS," Oscar Weinstein, NASA GSFC, Beltsville, Maryland.

These papers will be published in the conference proceedings in August, 1969.

ELECTRONICS INTERFACE UNIT

An electronics interface unit was completed and mounted in the field van. This unit includes operational amplifiers, auxiliary power supplies, scaling potentiometers, and similar items of analog electronics for improved ease of data acquisition interfacing with both the tape and multipoint recorder.

PUBLICATION

One publication was made in this quarter, "The Physical Basis of System Design for Remote Sensing in Agriculture," Proceedings of the IEEE, April, 1969, By R. A. Holmes and R. B. MacDonald.

CONTINUING RESEARCH

Work continued on the analysis of thermal data from 1968. Modeling efforts based on percent area coverage of soil and turgid plants are yielding blackbody equivalent temperature spectra similar to actual data. In the modeling process, soil is assumed to have a sand-like emittance and a range of physical temperatures, while crop cover is assumed to have the emittance of water and a range of physical temperatures. The results are encouraging, but the need exists for good emittance data of agricultural soils. Some progress can be made by measuring spectral radiance of the undisturbed soil on a windy day when one can assume some heat approach to isothermal scene conditions. As an added part of the study, a senior project was started for this summer on thermal soil transients in lab modeling and field experiments at the Agronomy Farm.

Data Processing Programs

INTRODUCTION

Previous quarterly reports of Data Processing Programs have described efforts in two categories (1) further development and adaptation of previously defined methods for data processing which will function over larger geographical areas and (2) initial development of new computation procedures for aspects of the total remote sensing problem not treated by previous capabilities. The first category has been planning and preparation since a planned aircraft scanner mission program had not been possible since the 1966 growing season.

However, near the end of the current quarter, the first of the required new data gathering missions was carried out, and the initial testing of procedures for larger geographic areas was begun. This is reported herein. Further, preparatory work in the form of the programming of LARSYS Version 2 continues. Work on several items in category (2) continued during the quarter as reported below.

DEVELOPMENT OF LARSYS VERSION 2

The basic purpose and characteristics of the second version of the LARSYS programming system have been discussed in previous reports. The development of this system continues. The method of applying the programming philosophies discussed last quarter have been developed and critical subroutines have been written and are now being tested. A system construction program has been put on line to aid in the development of the first two major programs of direct interest to the user. These two are LARSPLAY, the aircraft data display program which replaces PICTOUT, and AIRANAL, the aircraft data analysis program which replaces LARSYSAA.

All major processors in LARSPLAY have been defined, written, and are now being checked out and finalized. The Statistics Processor in AIRANAL is at the same stage as LARSPLAY. The other three processors in AIRANAL are Feature Selection, Classification and Display and are yet to be defined.

DATA COLLECTION AND ANALYSIS

Two data missions were carried out in this quarter, May and June. Data Processing Programs personnel participated in the collection of the ground truth for these flights so as to become more familiar with the stage of growth of the various ground covers. They also participated in the analysis of the data. A description of the work is included in Documentation Packages 10, 11, and 12 (Appendix A). Since this represents an initial run for data collection and analysis over larger areas, the greatest benefit from the project was the definition of new techniques to handle the larger quantities of data. An analysis of data from the June mission is not yet complete.

Analysis of data from the 1966 missions also continues. During this quarter the following flight lines were studied: C-3 and C-4 of July 1966 and C-2 of September 1966. The results of these studies are contained in Documentation Packages 6 and 9 (Appendix B).

DATA HANDLING RESEARCH

Aspects of three data handling techniques were under study during this quarter. The three are data registration, scanner angle corrections, and reformatting systems. The data registration or image overlay work has now progressed to a two dimensional research system. A paper describing this work is expected to be published in the Journal of the Society of Photo-Optical Instrumentation Engineers.

The determination of the effect of scanner sensitivity variations as a function of angle are under study. The effect of these variations on classification results is being studied experimentally. This work has now been temporarily halted and the results to data are reported in Documentation Package 7 (Appendix C).

The work on data reformatting techniques has centered on problems in locating calibration data for each scan line. It is also necessary to realign data from the scanner due to misalignments introduced by the analog tape recorder. An algorithm using correlation is being tested for this purpose.

DATA ANALYSIS RESEARCH

In addition to the aircraft data analysis activities carried out in concert with other LARS program areas and previously described, several projects are under study within the Data Processing Programs group. One is the development of the Per Field Classification Scheme which has been referred to in previous quarterly reports. Progress has continued on testing it and LARS Information Note 060569 (Appendix D) provides mathematical details and available results of tests on three of the more difficult flight lines. Also, during this quarter, the system has been implemented in a format similar to LARSYSAA so that it is readily usable by those less skilled in using computer systems. This implementation is named LARSYSPF.

Preliminary studies on the quality of fit tests and on probabilistic methods for selection of training fields were completed and documented during the early part of this quarter. These are reported in Information Notes 040469 and 040869, respectively (Appendix E).

During the later part of the quarter new work was begun to study the sensitivity of classification accuracy to the analog tape duplication process. For the first time an original analog tape became available to LARS. This tape was digitized, then sent to the University of Michigan for duplication in the standard manner. After digitization of the duplicate, both data sets will be analyzed using training areas from as near to the same location as possible. The resulting classification accuracy will be compared.

The adaptation of nonsupervised (clustering) classification algorithms to the task of finding field boundaries also continues. Development of the method has reached the point of producing the first map printouts with machine-determined boundaries. Figure 1 shows an example of such a printout together with a conventional gray-scale printout and an airphoto for comparison.

SYSTEM MAINTENANCE AND OPERATION

A routine amount of maintenance on both the hardware and software was required during the quarter. A new mode of computer operation was also used for the first time. An undergraduate student hired as a computer operator was made available two nights per week to run programs for researchers. As a result more effective use was made of the computer during the quarter and more processing carried out.

Table 1. Overall Performance results for training fields using five wavelength bands a/

LABORATORY FOR AGRICULTURAL REMOTE SENSING
PURDUE UNIVERSITY

JUNE 17,1969

CLASSIFICATION STUDY .. SERIAL NO. 616900900
CLASSIFICATION DATE .. JUNE 16,1969

RUN NUMBER-----56900070 DATE------ 5/26/69
FLIGHT LINE----PF21 TIME------1200
TAPE NUMBER---- 167 ALTITUDE-- 4500 FEET

CLASSES CONSIDERED

SYMBOL	CLASS
T	WI
-	B1
W	WH1
W	WH2
W	WH3
/	G4
/	G5
/	G10

FEATURES CONSIDERED

CHANNEL NO.	SPECTRAL BAND	
2	0.46	0.48
4	0.55	0.58
8	0.72	0.80
10	1.00	1.40
11	1.50	1.80

CLASSIFICATION SUMMARY BY TRAINING CLASSES

	CLASS	NO OF SAMPS	PCT. CORCT	NO OF SAMPLES CLASSIFIED INTO WHEA	GREE	W1	B1	THRS
1	WHEA	1827	99.8	1824	3	0	0	0
2	GREE	5369	99.9	5	5361	0	3	0
3	W1	160	100.0	0	0	160	0	0
4	B1	10155	89.6	3	1057	0	9095	0
	TOTAL	17511		1832	6421	160	9098	0

OVERALL PERFORMANCE = 93.9
AVERAGE PERFORMANCE BY CLASS = 97.3

a/ Whea, WH1, WH2, WH3 = wheat; gree, G4, G5, G10 = agricultural green vegetation; W1 = water; B1 = bare soil

Table 2. Overall performance results for test fields using five wavelength bands a/

LABORATORY FOR AGRICULTURAL REMOTE SENSING
PURDUE UNIVERSITY

JUNE 16,1969

CLASSIFICATION STUDY .. SERIAL NO. 616900900
CLASSIFICATION DATE .. JUNE 16,1969

RUN NUMBER-----56900070 DATE------ 5/26/69
FLIGHT LINE----PF21 TIME------1200
TAPE NUMBER---- 167 ALTITUDE-- 4500 FEET

CLASSES CONSIDERED

SYMBOL	CLASS
T	W1
-	B1
W	WH1
W	WH2
W	WH3
/	G4
/	G5
/	G10

FEATURES CONSIDERED

CHANNEL NO.	SPECTRAL BAND	
2	0.46	0.48
4	0.55	0.58
8	0.72	0.80
10	1.00	1.40
11	1.50	1.80

CLASSIFICATION SUMMARY BY TEST CLASSES

	CLASS	NO OF SAMPS	PCT. CORCT	NO OF SAMPLES CLASSIFIED INTO WHEA	GREE	SOIL	W1	THRS
1	WHEA	5437	61.4	3337	2100	0	0	0
2	GREE	32526	57.7	4367	18764	9395	0	0
3	SOIL	64570	92.2	1553	3504	59513	0	0
	TOTAL	*****		9257	24368	68908	0	0

OVERALL PERFORMANCE = 79.6
AVERAGE PERFORMANCE BY CLASS = 70.4

a/ Whea, WH1, WH2, WH3 = wheat; gree, G4, G5, G10 = agricultural green vegetation; W1 = water; B1 = bare soil

Table 3. Overall performance results for training fields using three wavelength bands [a]

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

JUNE 22,1970

CLASSIFICATION STUDY .. SERIAL NO. 605907200
CLASSIFICATION DATE .. JUNE 5,1969

RUN NUMBER-----56900070 DATE------ 5/26/69
FLIGHT LINE----PF21 TIME------1200
TAPE NUMBER---- 167 ALTITUDE-- 4500 FEET

CLASSES CONSIDERED

SYMBOL	CLASS
	WATER
	SOIL
	WH1
	WH2
	WH3
	G1
	G2
	G3
	G4
	G5

FEATURES CONSIDERED

CHANNEL NO.	SPECTRAL BAND	
1	0.40	0.44
10	1.00	1.40
12	2.00	2.60

CLASSIFICATION SUMMARY BY TRAINING CLASSES

	CLASS	NO OF SAMPS	PCT. CORCT	NO OF SAMPLES CLASSIFIED INTO WHEA	GREE	WATE	SOIL	THRS
1	WHEA	480	99.8	479	1	0	0	0
2	GREE	1837	97.2	8	1785	0	44	0
3	WATE	48	100.0	0	0	48	0	0
4	SOIL	2808	88.6	0	319	0	2489	0
	TOTAL	5173		487	2105	48	2533	0

OVERALL PERFORMANCE = 92.8
AVERAGE PERFORMANCE BY CLASS = 96.4

a/ WH1, WH2, WH3, Whea = wheat; G1, G2, G3, G4, G5, Gree = agricultural green vegetation; Wate = water

Table 4. Overall performance results for test fields using three wavelength bands [a]

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

JUNE 22,1970

CLASSIFICATION STUDY .. SERIAL NO. 605907200
CLASSIFICATION DATE .. JUNE 5,1969

RUN NUMBER-----56900070 DATE------ 5/26/69
FLIGHT LINE----PF21 TIME------1200
TAPE NUMBER---- 167 ALTITUDE-- 4500 FEET

CLASSES CONSIDERED

SYMBOL	CLASS
T	WATER
-	SOIL
W	WH1
W	WH2
W	WH3
/	G1
/	G2
/	G3
/	G4
/	G5

FEATURES CONSIDERED

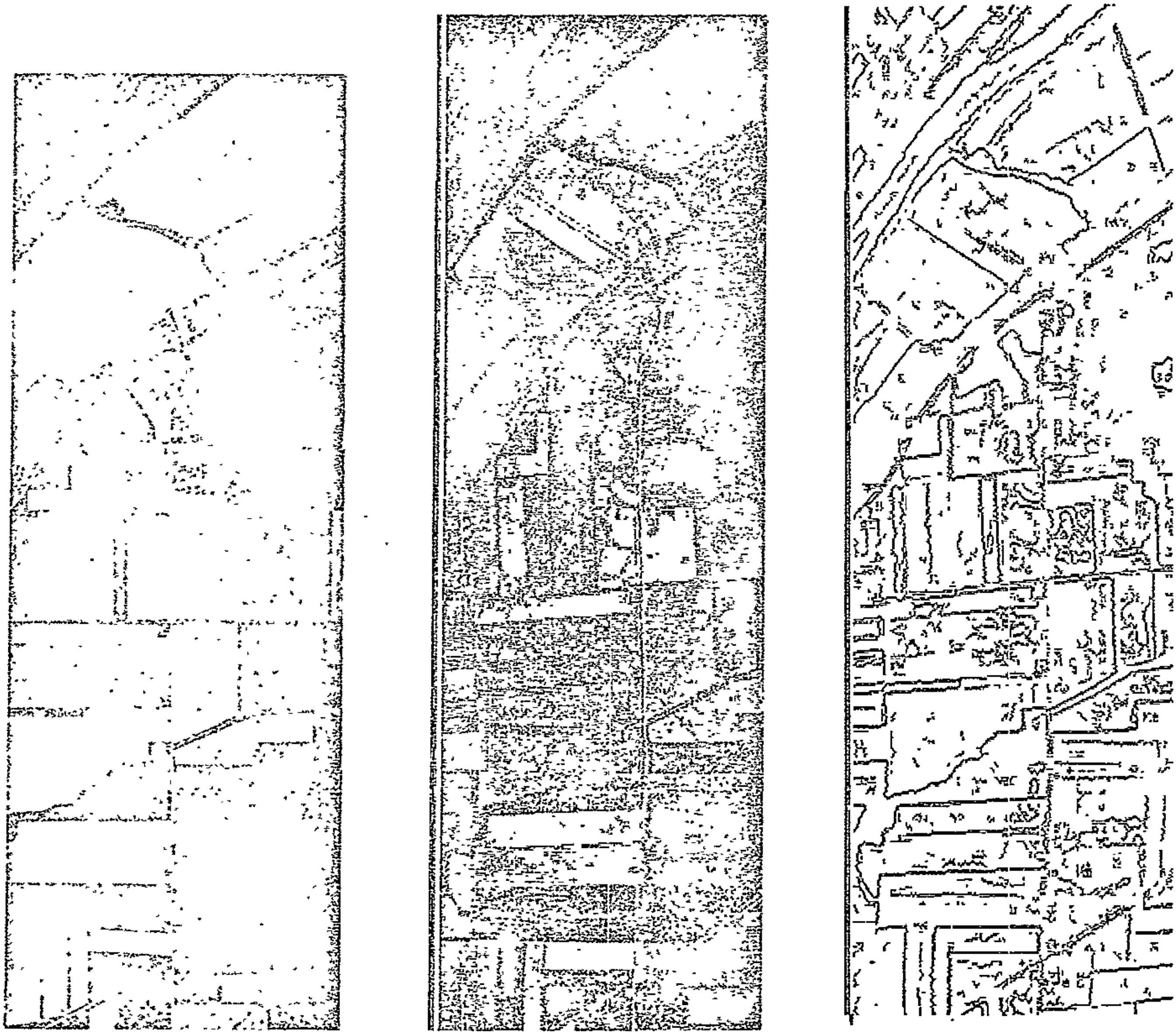
CHANNEL NO.	SPECTRAL BAND	
1	0.40	0.44
10	1.00	1.40
12	2.00	2.60

CLASSIFICATION SUMMARY BY TEST CLASSES

	CLASS	NO OF SAMPS	PCT. CORCT	NO OF SAMPLES CLASSIFIED INTO SOIL	GREE	WHEA	WATE	THRS
1	SOIL	16675	81.0	13511	3082	82	0	0
2	GREE	8225	78.2	1034	6428	763	0	0
3	WHEA	1433	58.6	0	593	840	0	0
	TOTAL	26333		14545	10103	1685	0	0

OVERALL PERFORMANCE = 78.9
AVERAGE PERFORMANCE BY CLASS = 72.6

a/ Wh1, Wh2, Wh3, Whea = wheat; G1, G2, G3, G4, G5, Gree = agricultural green vegetation; Wate = water



Panchromatic Air Photo | .62-.66μ M | Boundaries

Figure 1. Air photo, gray-scale printout and machine - determined boundaries of an agricultural acene

APPENDIX A

June 11, 1969

Documentation Package #10

1. Project Title:

Digitizing and Reformating Michigan Scanner Data of May 13, 1969

2. Experimenter:

W. R. Simmons

3. Objective:

The objective was to process Michigan multispectral scanner data of 13 May 1969. It was hoped to successfully overlay data from ends A and B of the 80° FOV scanner No. 2.

4. Description of Data:

The data was collected by the Michigan aircraft over flight lines 21 and 25 between the times 8:45 a.m. and 9:35 a.m. EDT. The aircraft altitude was 5,000 feet above the terrain with ground headings of 180° and 90° for flights 21 and 25 respectively. Weather conditions were cloudy and overcast. Thirteen channels of data plus sync were recorded on 1-inch analog tape and delivered in original form to LARS on the day of the flight. Ten data channels were from end A of scanner No. 2 and three channels from end B of the same scanner. The wavelength bands are as follows: .40-.44, .46-.48, .50-.52, .52-.55, .55-.58, .58-.62, .62-.66, .66-.72, .72-.80, .80-1.00, 1.0-1.4, 1.5-1.8, 2.0-2.6 microns. Field of view was 80° for all channels. The three channels from end B were delayed 90 scanner degrees. The data line format contained 3 calibration pulses. The first was a lamp filtered to give medium response in all wavelengths, the second was the sun sensor pulse, and the third was a lamp filtered for the UV wavelength band. Because of leakage, the UV calibration pulse did show up in wavelengths above .72 microns.

6. Pre-Data Handling Decisions:

The LARSYS Version 1 data system can handle a maximum of 12 channels on a single data run. Since channels 1 and 2 of end B contained fair to good registration and channel 3 of end B contained poor registration it was decided to digitize the 10 end A channels and 2 end B channels.

7. Data Handling Results:

For flight lines 21 and 25 respectively, 2699 and 2308 data lines were digitized and reformatted. The two end B channels were overlaid during the reformating process. In order to evaluate the data quality gray-scale prints, graphs of data lines and graphs of lamp and sun sensor values were made.

8. Discussion of Results:

The gray-scale printouts show dark and bright areas which indicate sun illumination changes. The column graphs of the calibration lamp indicate that amplifier gains were changed for channel 11 in the middle portion of flight 21 and channel 6 at the end of flight 25. Amplifier drift is not indicated. The column graphs of the sun sensor calibration demonstrate severe sun illumination changes as well as the above mentioned gain changes in both flight lines. These graphs also show that the sun sensor level for channels 2, 3 and 5 was electronically clipped. This means the sun sensor level cannot be used for calibration purposes for the lines where clipping occurred. The data line graphs give some indication of channel-to-channel overlay precision. The data lines, examined only by eye, indicate the alignment error for the 10 end A channels to be $\pm \frac{1}{2}$ sample. The alignment error for the 2 end B channels is plus and minus 1, 2 and 3 samples. Two factors which help to explain the seemingly changing alignment error for the end B channels are: (1) the alignment technique applied in the reformating program, coupled with (2) the changing scanner frequency.

9. <u>Conclusions and Recommendations for Future Work</u>.

From a data handling viewpoint, the overall quality of the resulting data is good. The data may however, be termed poor by data analysis people because of the severe illumination changes. The difficulty in processing the data was exhausting due to the obsolescence of the reformating program, with respect to the new scanner data line formats. Format changes referred to specifically are: (1) narrowing of calibration pulses, (2) addition of the sun sensor calibration pulse and (3) addition of end B channels onto the 1-inch analog tape. In order to speed up and increase precision in processing future scanner data, the following changes are being made in the reformating program.

1. Samples from the odd and even analog tape channels in the region of the calibration pulses will be summed separately and correlated to more accurately find the odd/even head skew.
2. The odd channel sum from No. 1 above will be used to find the position of the calibration pulses.
3. The position of the pulses will be determined using a two-pass scheme.
4. The distance between the calibration pulses will be determined automatically.
5. The delay of the end B channels will be determined automatically by correlation techniques.
6. The position seeking of the calibration pulses for a data line will be aided by the position determined in the previous data line.

DOCUMENTATION PACKAGE #11

1. Project Title:

   Digitizing and Reformating Michigan Scanner Data of May 26, 1969.

2. Experimenter:

   W. R. Simmons

3. Objective:

   The objective was to process Michigan multispectral scanner data of 26 May 1969. It was further hoped that the three end B channels of scanner No. 2 could be overlaid with the end A channels with greater accuracy than had been achieved in previous runs.

4. Description of Data:

   The data were collected by the Michigan aircraft over flight lines 21 and 25 at approximately 1200 hours EDT. The aircraft altitude was 4000 feet above the terrain with ground headings of 180° and 90° for flight lines 21 and 25 respectively. The weather conditions were generally clear and sunny. Thirteen channels of data plus sync were recorded on 1-inch analog tape and delivered to LARS on the day of the flight. Ten data channels were from end A of scanner No. 2 and three channels from end B of the same scanner. The wavelength bands are: .40-.44, .46-.48, .50-.52, .52-.55, .55-.58, .58-.62, .62-.66, .66-.72, .72-.80, .80-1.00, 1.0-1.4, 1.5-1.8, and 2.0-2.6 microns. The field of view was 80° for all channels. The three end B channels were delayed 90 scanner degrees. The data line format contained registration from 80° of relative reflection from the terrain followed by as many as three calibration pulses. The calibration pulses were: (1) a lamp filtered to give medium response in all wavelengths, (2) a sun sensor pulse, (3) a lamp filtered for the UV wavelength band. The UV lamp registered in the reflective infrared band because of leakage. The syncronization channel was recorded in FM mode and it was noted the roll stabilizer pulse was delayed approximately .48 microseconds when compared to previous flights. It was later noted that the roll stabilizing scheme did not seem to be functioning properly.

6. Pre-Data Handling Decisions:

Since the LARSYS Version 1 data system can handle a maximum of 12 channels in a single run, one of the 13 analog data channels had to be dropped. The .50-.52 micron channel was dropped because its content seemed to be more similar to one of its adjacent bands than did any of the other 12 channels.

7. Data Handling Results:

For flight lines 21 and 25 respectively 2942 and 2472 data lines were digitized and reformatted. In order to evaluate the data quality gray-scale printouts, graphs of data lines and graphs of lamp and sun sensor values were made.

8. Discussion of Results:

The gray-scale printouts and the column graphs of the sun sensor calibration demonstrate very constant illumination for both flight lines. The column graphs of the calibration lamp indicate no changes in amplifier gain for either run. Certain features in the gray-scale print such as straight roads which do not appear straight in the print indicate that the roll stabilizer was not properly functioning. The graphs of certain data lines give some indication of end B data overlay error. The graphs indicate that the end A channels are overlaid with an error of less than one sample. The end B channels however seem to have errors as large as three samples. No reason for this error has been positively determined.

9. Conclusions:

The overall quality of the processed data is good from a data handling viewpoint. For future processing of scanner data having the same format, the problem of end B data overlay should be further examined.

June 20, 1969

DOCUMENTATION PACKAGE #12

1. Project Title: Ground Cover Analysis of Data Taken May 26, 1969.

2. Experimenters: K. Hunt, T. Phillips, P. Swain, R. Hoffer.

3. Objectives:

The primary objective of the first analysis of the May 1969 data was to ascertain the quality of the data taken by the Michigan Scanner System. A second objective was to develop a complete, useful technique for data analysis. The third objective was to separate the prinicipal ground covers of flight line 21 taken at 12:00 Noon, May 26, 1969.

4. Data Used:

Data digitized from original analog tapes and reformatted on May 26-27, 1969 was used. Run 56900070 covering Purdue Flight Line 21 flown at 12:00 Noon, May 26, 1969, at an altitude of 4,000 feet with a ground heading of 180° was used. This data was available in twelve wavelength bands: .40-.44, .46-.48, .52-.55, .55-.58, .58-.62, .62-.66, .66-.72, .72-.80, .80-1.0, 1.0-1.4, 1.5-1.8, 2.0-2.6 microns. The roll correction pulse was incorrect giving approximately 20° of unusable data on the left side.

The ground truth available was in the form of comments written by LARS Staff at the time of the flight on outdated photographs of the flight line. Photographs taken May 13, 1969 and in-flight scanner imagery were also referenced.

5. Description of Data:

Cover types present included water, bare soil, and green vegetation - wheat, hay, trees, stubble, oats, corn, rye, and other green vegetation. For classification purposes, however, four classes were chosen - water, bare soil, green vegetation and wheat.

6. Preclassification Work:

A. Data Editing - Several problems were encountered during the data editing which made it difficult to transfer the ground truth from photographs to the pictorial printout. All the information was available but hard to correlate because:

1. The resolution of the single-width printout was inadequate for finding field boundaries.
2. The resolution of the double-width printout was marginal for data taken at 4,000 foot altitude.
3. Data was digitized with the wrong aspect ratio because it was thought that the flight was flown at 5,000 feet and later information showed the altitude to be only 4,000 feet.
4. The ground truth was taken on outdated photographs which could not easily be compared to gray-scale printouts.

Photographs from May 13, 1969 and scanner imagery were available and quite helpful.

Data Editing Procedure:

1. Gray-scale printouts were made of the .62-.66 wavelength band.

2. Field boundaries were marked.

3. Ground truth and field numbers were transferred to the printout.

4. Test boundaries were marked.

5. Fields were addressed and punched on cards.

Area C was chosen for a first look at the data to get initial results in a short time.

B. First classification

Individual field histograms of bare soil, oats, wheat, residues, hay, and pasture indicated that we might easily discriminate bare soil, green vegetation and water. Statistics were taken only for those fields which obviously fell into one of the three categories. These fields were used as training fields. $SELECT was tried for combinations of three channels. Channels 7, 9, and 12 were computed to yield maximum divergence between the categories.

The results of this classification study were quite good, as might be expected. The overall performance of these training fields was 98.1 percent correct classification. This classification was the first indication that the data was good, but even more useful was the display map. It showed very clearly where many field boundaries were and was quite helpful with the data editing on the remainder of the flight.

C. Intensive preclassification work

A careful evaluation of the above work showed that it might be possible to separate wheat from green vegetation and it was felt that it would be desirable to extrapolate to the entire flight line. Using the same training fields, $SELECT was used to cluster the training fields into classes of water, wheat, green vegetation and bare soil. Several passes through $SELECT indicated that there was a column boundary problem due to the misregistration of the overlaid data. In order to make sure of the boundaries the top-bottom boundaries were adjusted also. PICTOUT confirmed that an adjustment was necessary to correct the boundary problem because of poor overlay.

After adjusting the field boundaries to correct the problem, $SELECT was again used to group fields into classes. This resulted in one class of water, one class of bare soil, three subclasses of wheat, and five subclasses of green vegetation. Channels 1, 10 and 12 were computed to yield maximum divergence between the classes.

7. Classification results:

Two classifications were performed on the data. A five-channel classification was done for comparison with the results of the three-channel classification. Briefly, the results of the five feature classification were:

Overall Performance (TRAIN)	93.9%
Overall Performance (TEST)	79.6%

Channels 2, 4, 8, 10 and 11 were used for classification.

Results of the three-channel classification were as follows. No threshold was used.

Overall Performance (TRAIN)	92.8%
Overall Performance (TEST)	78.6%

A tabular listing of classification above the 70% correct level is shown below.

Class	Number of test fields	Number of test fields classified correctly above 70%
Wheat	18	12
Green vegetation	100	76
Bare soil	119	87
	237	175

Discussion of Results:

Of the 161,810 points classified, 17% of those were included in test fields and 3% were included as training samples. Other samples were not used because of the lack of ground truth, the inaccuracy of the data used because of the overlay problems, and the 20° of data on the left that was bad.

There are some very good reasons why some fields were incorrectly classified. Most of the problem lies with the training fields and the ground truth. For example, six wheat fields were below 70% correct classification. Four of these six fields were non-headed wheat fields and there were no training fields of this type. One other wheat field was probably oats since the ground truth called it wheat or oats. Some green vegetation fields were probably incorrectly classified as bare soil since the growth at this time of the year is not significant enough to discriminate between the two. In the same manner there were bare soil fields classified as green vegetation, possibly because of too much growth.

9. Conclusions:

This classification did show the quality of the data to be good except for the problem of overlay in the three IR channels. Most of the incorrect classifications probably were not connected with the quality of the data. Secondly, this exercise did help to develop a technique for data analysis and much knowledge was gained about the use of LARSYSAA. It proved most important to have people with diverse backgrounds interacting as this made the work easier and much more thorough.

There are definite indications that further investigation is necessary to achieve the best possible classification on this data and for other future work. Recommendation for further work would include the following:

Further analysis

1. Discriminate between classes on the basis of percent ground cover.
2. Compare data taken May 13, 1969 versus May 26, 1969 versus May 27, 1969.
3. Extrapolate between flight lines.

Future work

1. Convert ground truth to current photographs.
2. Use color imagery as a ground truth aid.

APPENDIX B

Date: April 7, 1969

DOCUMENTATION PACKAGE #6

1. Project Title: Crops classification study of C4, July 26, 1966.

2. Experimenter: Teddy Huang

3. Objectives: The objective was to produce a crop map of C4 area by using the LARSYSAA pattern recognition programs.

4. Data Used: Aircraft Data Run No. 26600800. Data was not calibrated.

5. Description of Data: Cover type present: soybeans, corn, pasture, diverted acres, wheat, oats, and red clover.

   Cover type to be classified: soybeans, corn, pasture, which consists of pasture, diverted acres, and red clover, and stubble which consists of wheat and oats.

6. Preclassification Work: Based on the study of the histogram of every field, five subclasses were chosen to represent the soybeans, the corn, and the pasture respectively. Four subclasses were used to represent stubble. This was due to the multimodal distribution of each of the four classes. Diverted acres, red clover, and pasture were combined into a single class because of their closeness in distribution. Wheat and oats were grouped into stubble because they were stubbled in July.

   Based on the results obtained from the Selection Processor, Channel 1, 6, 9 and 11 were selected as the best four features to be used for classification. Channel 1, 6, 9, 10, 11 and channel 1, 6, 9, 10, 11, and 12 were selected as the best five and six features respectively.

7. Classification Results:

Classification results were obtained by using 4, 5 and 6 features. But the best classification results were obtained by using only 5 feature (i.e. channel 1, 6, 9, 10 and 11) and were stored on the tape associated with the serial No. 0303901202. The threshold value used for each class was set at 27.2 (or 90% level). 4832 samples were used as the training samples. The overall performance of the training field was 91.7 and average performance by class was 92.7. 81 test fields were classified. The overall performance was 79.7 and average performance by class was 79.9. The percentage of correct recognition for soybeans was 83.0% with 11% of the total soybean test samples classified as corn. The percentage of correct recognition for corn was 79.6% with 17% of total corn test samples classified as soybeans. The percentages of correct recognition for pasture and stubble were 75.7% and 81.2% respectively.

	No. of test fields	No. of test fields classified above 70%
soybeans	19	15
corn	31	23
pasture	18	13
stubble	13	11
Total	81	62

8. Discussion of Results:

Total number of RSU's (Remote Sensing Units) classified was 72270; 22353 of these were included in the test fields. About 50% of the total RSU's were not included in the test field because of insufficient ground truth information. The other RSU's were not included in the test fields because they came from small agriculture fields whose exact field boundary could not be accurately located.

9. Conclusions:

The overall classification result was satisfactory. It has been observed that the overall performance improves steadily as the number of the subclasses used to represent each main class increases.

June 3, 1969

DOCUMENTATION PACKAGE #9

1. Project Title: Crops classification study of (1) C-2, September 15, 1966, (2) C-3, July 26, 1966, and (3) C-4, July 26, 1966

2. Experimenter: Teddy Huang

3. Objectives: The objective was to evaluate the performance of per field classifier based on distance between distributions in multivariate Gaussian cases for agricultural applications.

4. Data Used: C-2 - Aircraft Data Run No. 26601540

   C-3 - Aircraft Data Run No. 26600780

   C-4 - Aircraft Data Run No. 26600800

   Data was not calibrated.

5. Description of Data:

   C-2 area - Cover type present: soybeans, corn, pasture, stubble and water

   Cover type to be classified: soybeans, corn, pasture, stubble and water

   C-3 area - Cover type present: soybeans, corn, pasture, diverted acres, hay, wheat, stubble, water

   Cover type to be classified: soybeans, corn, pasture and stubble mixture, and water

   C-4 area - Cover type present: soybeans, corn, pasture, diverted acres, wheat, oats and red clover

   Cover type to be classified: soybeans, corn, pasture consisting of diverted acres, pasture and red clover, and stubble consisting of wheat, oats and stubble

6. Preclassification Work:

   Based on the study of the histogram, subclasses were chosen to represent each class considered because of various crop types and growing conditions.

C-2: Soybeans - 1 subclass
Corn - 1 subclass
Pasture - 1 subclass
Stubble - 1 subclass
Water - 1 subclass

C-3: Soybeans - 3 subclasses
Corn - 3 subclasses
Pasture and stubble mixture - 2 subclasses
Water - 1 subclass

C-4: Soybeans - 5 subclasses
Corn - 5 subclasses
Pasture - 5 subclasses
Stubble - 4 subclasses

The statistical parameters were estimated from adequate training samples.

7. Classification Results:

Area C-4

Class	No. of Fields	Number of Fields Classified into Soybeans	Corn	Pasture	Stubble
Soybeans	19	17	2	0	0
Corn	31	1	30	0	0
Pasture	18	0	0	18	0
Stubble	13	0	0	3	10

Total: 81 No. of fields correctly classified = 75

Area C-3

Class	No. of Fields	Number of Fields Classified into Soybeans	Corn	Mixture	Water
Soybeans	13	11	1	1	0
Corn	10	0	9	1	0
Mixture	18	1	0	17	0
Water	3	0	0	0	3

Total: 44 No. of fields correctly classified = 40

Area C-2

Class	No. of Fields	Number of Fields Classified into Soybeans	Corn	Pasture	Stubble	Water
Soybeans	12	8	4	0	0	0
Corn	14	1	13	.0	0	0
Pasture	13	0	0	13	0	0
Stubble	11	0	1	1	9	0
Water	3	0	0	0	0	3

Total: 53 No. of fields correctly classified = 46

8. Discussion of Results:

C-2 Area. When we classified the same set of data by using the MLDR (Maximum Likelihood Decision Rule) classifier (or LARSYSAA), the overall performance and average performance by class were 67.0 and 73.4 respectively. There were 20 of the 53 test fields which had less than 70% correct recognition. Forty-six of the 53 test fields were correctly classified by the per field classifier; this is a percentage accuracy of 86.8.

C-3 Area. When we classified the same set of data by using the MLDR classifier, the overall performance and average performance by class were 76.6 and 80.3 respectively. There were 12 of the 44 test fields with less than 70% correct recognition. Forty of the 44 test fields were correctly classified by the per field classifier; this is a percentage accuracy of 90.9.

C-4 Area. When the same set of data was classified by MLDR classifier, the overall performance and average performance by class were 79.7 and 79.9. There were 19 of the 81 test fields with less than 70% correct recognition. Seventy-five of the 81 test fields were correctly classified by the per field classifier; this is a percentage accuracy of 92.6.

9. Conclusions:

As far as the given data is concerned, the performance of the per field classifier is considered to be satisfactory. We have not made any attempt to check the assumption of multivariate Gaussian distribution to justify the simplified classification scheme used by the per field classifier but the classification results obtained encourage us to further test the applicability of the per field classifier to agricultural applications.

APPENDIX C

May 14, 1969

DOCUMENTATION PACKAGE #7

1. Project Title: To determine the effect on classification results of data correction because of scanner instrument errors as a function of angle.

2. Experimenter: Kay Hunt

3. Objectives: The motivation for this project is twofold. First, is the selection of a constructive project to acquaint the experimenter with data processing techniques used at LARS. Second, the project was selected because it is known that scanner instrument errors affect classification results.

It is the experimenter's objective to understand the scanner system through a Report of Project Michigan, correct the data, apply LARS pattern recognition techniques, and compare the results to an original classification of uncorrected data.

It is expected that the primary objective of acquainting the experimenter with data processing techniques will be fulfilled. While the complete understanding of scanner angle corrections will not be attained, the project should aid in this understanding and point to recommendations for important future work.

4. Data Used: Flight Line C1, June 1966 at 1229

5. Description of Data: The data that was used consisted entirely of crop cover from the standard training and test fields previously selected for Flight Line C1. A variety of crops were discriminated—soybeans, corn(2), alfalfa, red clover(2), oats, rye, bare soil, and wheat(2).

Before the LARS pattern recognition techniques could be used, it was necessary to correct the data in a manner which would compensate for the scanner angle error.

The data was first averaged over the entire flight run by sample for each channel. It was felt that the flight line was long enough to average out all of the inconsistencies and result in a line which would show where corrections might be needed. Through the use of PICTOUT the averaged line was then plotted for each channel. It was conjectured that these lines should be relatively straight lines except for a few samples in the center of the scan line (a road ran almost down the middle of the entire flight line). This was, in fact, the case except that the line varied from the predicted straight line along the edges of the scan line. This further indicated that the correction might be necessary. It was at this point that an empirical method other than the University

Michigan corrections (The University of Michigan developed correction factors based on laboratory experiments which were published in their Report on Project Michigan) was derived. It was shown that both of these correction factors are approximately the same over 60° of the field of view. They were quite different for the 10° edges on both sides of the 80° field of view of the scanner. Research into this inconsistency showed that the corrections recommended by Michigan were based on experiments with the scanner not mounted in the aircraft and that an obstruction of the aircraft structure could cause the effects which were noticed. Both methods are discussed below with the results tabulated in Section 7.

Technique using the University of Michigan Correction Factors: The University of Michigan devised a test fixture which pivoted a diffuse reflector with quartz line lamps about the scanning mirror axis. The scanner remained at its normal orientation, and the source which filled the field of view moved about the scanner. As the source moved, the spectrometer output pulse moved across the 'scope face. Measurements were made from scope photographs and replotted as scanner response vs. scan angle. It was from these graphs that data was taken to obtain the percent response for each data sample. Corrected data was obtained by using the following equation:

$$D_{C_{ijk}} = CO - \frac{CO_{jk} - D_{M_{ij}}}{F_{ij}}$$

where

i is a function of angle

$D_{C_{ijk}}$ = Corrected data point for the i^{th} sample, j^{th} channel, and k^{th} line

CO_{jk} = Dark level for the j^{th} channel and the k^{th} line

$D_{M_{ij}}$ = Measured data point for the i^{th} sample and the j^{th} channel

F_{ij} = Percent scanner response for the i^{th} sample and the j^{th} channel

A new tape was made containing these $D_{C_{ijk}}$ values and used for classification purposes. LARSYSAA was implemented and through the use of the $STAT and $SELECT processors it was decided to group the two red clover classes into one and use features 1, 10, 12.

Technique using average line data: Since it was felt that averaging data over the entire flight line should result in a straight line, an equation was derived to adjust the data so that upon averaging the entire flight line a straight line would result. The following equation was used to correct the data:

$$D_{C_{ijk}} = CO_{jk} - \frac{CO_{jk} - D_{M_{ij}}}{F_{ij}}$$

where

i is a function of angle

$D_{C_{ijk}}$ = Corrected data point for the i^{th} sample, j^{th} channel and k^{th} line

CO_{jk} = Dark level for the j^{th} channel and the k^{th} line

$D_{M_{ij}}$ = Measured data point for the i^{th} sample and the j^{th} channel

and

$$F_{ij} = \frac{CO_{jk} - D_{P_{ij}}}{CO_{jk} - D_{A_j}}$$

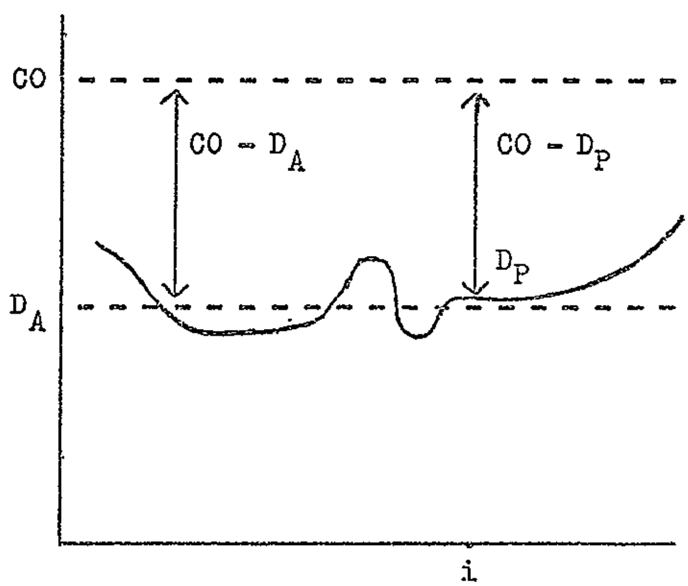
where

CO_{jk} = as above

D_{A_j} = overall average of all data points for the j^{th} channel. This may be an arbitrary value but in this case was taken to be the overall average in order to keep the correction of the data to a minimum.

D_{P_j} = calculated average for all values for the i^{th} sample and the j^{th} channel

See graph below.



A new tape was made containing these $D_{C_{ijk}}$ values and used for classification purposes. Again, LARSYSAA was implemented and through the use of the $STAT and $SELECT processors it was decided to group the two red clover classes into one class, the two wheat classes into one class, and the two corn classes into one class. Again, features 1, 10, 12 were chosen for classification purposes since they yielded the maximum divergence between the eight equally weighted classes.

Preclassification Work for Uncorrected Data: The original data was processed in the normal way using LARSYSAA. Through the use of the $STAT and $SELECT processors, there were indications that the two red clover classes should be grouped into one class. Since the divergences between the ten equally weighted categories was so similar for features 1, 10, 12 and 6, 10, 12, it was decided to classify the data twice using both sets of features.

7. Classification Results: Several classification results were performed in order to incorporate all of the preclassification results and to adequately compare the results. Only the results based on features 1, 10, 12—grouped and ungrouped—are included here since these are the best results for all three sets of data. Presented below is a tabular listing of classification about the 70% correct level. In

all cases the exact same training and test fields were used in order to compare results fairly. (Note: No grouping means only the two classes of red clover are grouped together. Grouped means each of the two classes of red clover, corn, and wheat have been grouped into three single classes.)

:T CLASSIFICATION ABOVE 70%

FEATURES 1-10-12
No Group

CLASS	Total Number of Test Fields	Number of Test Fields Classified Above 70% Uncorrected	Corrected (Using Averages)	Corrected (Using Michigan values)
CORN	9	7	7	7
WHEAT	7	6	6	6
SOYBEANS	14	12	10	13
OATS	4	4	4	4
RED CLOVER	9	6	8	6
ALFALFA	3	3	3	3
RYE	1	1	1	1
BARE SOIL	2	2	2	2
TOTALS	49	41	41	42

FEATURES 1-10-12
Grouped

CLASS	Total Number of Test Fields	Uncorrected	Corrected (Using Averages)	Corrected (Using Michigan values)
CORN	9	7	7	7
WHEAT	7	5	6	6
SOYBEANS	14	12	12	13
OATS	4	4	4	4
RED CLOVER	9	6	8	6
ALFALFA	3	3	3	3
RYE	1	1	1	1
BARE SOIL	2	2	2	2
TOTALS	49	40	43	42

It is also of interest to look at the overall performances for each of the classifications. In every instance the corrected data performed better than the original data which is what we might expect. These results are tabulated below.

OVERALL PERFORMANCE OF CLASSIFICATIONS

Training Fields

Features	Uncorrected	Corrected (Using Averages)	Corrected (Using Michigan values)
1-10-12 no group	91.2	92.5	94.1
1-10-12 grouped	92.0	92.5	94.6

Test Fields

Features	Uncorrected	Corrected (Using Averages)	Corrected (Using Michigan values)
1-10-12 no group	81.1	83.5	84.2
1-10-12 grouped	81.7	83.4	84.2

Since both methods of adjusting the data values have a much higher correction factor towards the edges of the scan line, it is also valuable to look at those individual fields which lie on the edges of the flight line. These results are presented below. There doesn't seem to be any noticeable trend in the performances which favors corrected or uncorrected data.

PERCENT CORRECT CLASSIFICATION FOR FIELDS AT THE EDGES

Features 1-10-12
No Group

TEST FIELDS	Uncorrected	Corrected (average)	Corrected (Michigan)
Field Designation			
1-15 Corn	44.5	20.9	56.5
12- 9 Corn	93.0	94.0	98.1
12-10 Wheat	91.3	99.5	82.9
36-14 Red Clover	23.2	78.1	18.6
6- 7 Red Clover	70.6	39.8	77.1
7- 2 Wheat	78.2	96.4	94.5
6-14 Wheat	100.0	97.0	99.2
TRAINING FIELDS			
36- 4 Corn	94.2	97.8	97.8
36- 9 Corn	96.3	88.9	96.3
36- 8 Corn	100.0	94.4	100.0
12- 9 Corn	97.0	90.2	97.7
7- 2 Wheat	81.3	98.0	96.7
12-10 Wheat	93.8	100.0	86.4
6-14 Wheat	100.0	100.0	100.0

Features 1-10-12
Grouped

TEST FIELDS	Uncorrected	Corrected (average)	Corrected (Michigan)
Field Designation			
1-15 Corn	36.5	12.7	48.5
12- 9 Corn	92.2	90.0	96.6
12-10 Wheat	85.5	99.5	80.2
36-14 Red Clover	28.8	78.1	19.0
6- 7 Red Clover	74.2	40.0	77.9
7- 2 Wheat	69.8	96.0	94.9
6-14 Wheat	100.0	96.7	99.2

PERCENT CORRECT CLASSIFICATION
FOR FIELDS AT THE EDGES (continued)

Features 1-10-12
Grouped

TRAINING FIELDS	Uncorrect	Corrected (average)	Corrected (Michigan)
Field Designation			
36- 4 Corn	92.0	92.8	95.7
36- 9 Corn	96.3	87.7	96.3
36- 8 Corn	100.0	86.1	100.0
12- 9 Corn	96.2	84.1	97.0
7- 2 Wheat	74.7	98.0	97.3
12-10 Wheat	88.6	100.0	83.5
6-14 Wheat	100.0	100.0	100.0

8. Discussion of Results and Conclusions:

Both correction factors increased the accuracy of classification of this data. It is difficult to come to concrete conclusions because the evidence of an experimental result over one flight line of data does not lead to conclusions; however, this project does indicate that scanner angle corrections should be investigated further and applied to the data for optimum classification.

There are, in addition, some remarks that need to be made. First, in all cases the same channels were selected for classification but not the same grouping procedure. Only in using the averaged data was complete grouping indicated. Grouping the classes did not affect the overall classifications greatly except to increase the number of fields above the 70% correct level by two using averaged data. It is important to remember that with fewer classes the classification time is reduced considerably. This in itself is an improvement.

Secondly, it was hoped that investigating the results of the fields on the edges would lead to significant conclusions as to which correction factors might lead to better results. However, referring back to these results, it is found that no conclusions of any kind can be drawn. It is interesting to note that, in general, it is the corn fields which cause the results of the averaged data to drop. Further investigations should be made to find whether the cause of this is due to the correction factors or some biological property of the corn itself.

Thirdly, the method of correcting the data using averages is flight run dependent. This, of course, slows the overall procedure in comparison to using the Michigan corrections which are not run dependent. However, it might be found through repeated experiments using averaged data that these corrections also could be normalized into universal factors. Of course, this requires much further work.

In general, the results presented here are quite inconclusive but, there are definite indications that this problem is worth investigating further in order to achieve the best possible classi cations. My recommendation for further work would be the following:

1. Select more fields along the edges of the scan line using corrected data in order to compare results more completely.
2. Select many other flight runs to correct and classify.
3. Investigate the causes of the differences in the two correction factors for the 10° edges on both sides of the 80° field of view of the scanner.
4. Investigate the possibility of obtaining general correction factors using the average method which are not run dependent.

APPENDIX D

LARS Information Note 060569
Purdue University

Per Field Classifier
for
Agricultural Applications

by

Teddy Huang

I. Introduction

In accordance with general crop planting practice, if an agricultural field has a certain majority of resolution elements classified as "ω_i", then one can feel reasonably sure that the entire field as a whole is in class "ω_i". Based on this observation, we shall formulate a pattern classification procedure for agricultural applications as follows. Let $\{\omega_i\}$ be a set of pattern classes with their distribution $\{F_i\}$ respectively, and let the feature vector x be a random variable under consideration. Then given a set of points $\{x\}$ all from the same field, the problem is to decide to which F_i they belong. We shall propose a distance measure $d(F_i,F_j)$ as a suitable metric of the separation between any two distributions F_i and F_j, and we shall use this metric as a rule for decision making. Let $G(x)$ be the distribution governing $\{x\}$. We shall compare the magnitudes of $\{d(G,F_i)\}$; the distribution F_i which minimizes the distance $d(G,F_i)$ shall be presumed to be the distribution which contains $G(x)$. Therefore, the vector set $\{x\}$ is classified into one of the known classes.

If the distributions concerned are unknown, we shall empirically estimate these distributions F_i based on finite observation on $x \sim \omega_i$.

II. Definition of the Distance

Let $F_i(x)$ and $F_j(x)$ be the distributions of pattern classes ω_i and ω_j respectively, and let $p_i(x)$ and $p_j(x)$ be their probability density functions defined in Ω_x. Then the distance between distributions F_i and F_j is defined as follows:

$$d^2(F_i,F_j) = \int_\Omega (\sqrt{p_i(x)} - \sqrt{p_j(x)})^2\, dx \tag{1}$$

If we define

$$\rho(F_i,F_j) = \int_\Omega \sqrt{p_i(x)} \cdot \sqrt{p_j(x)}\, dx \tag{2}$$

we have

$$d^2(F_i,F_j) = 2 - 2\rho(F_i,F_j) \tag{3}$$

The quantity $\rho(F_i,F_j)$ expresses the correlation between distributions F_i and F_j, and we can use $\rho(F_i,F_j)$ in decision making. It is important to note that to minimize the $d(F_i,F_j)$ is the same as to maximize the $\rho(F_i,F_j)$.

III. Distance in Multivariate Cases

The distance $d(F_i,F_j)$ defined in Equation (1) is applicable to any distribution. We shall now turn to multivariate Gaussian case for agricultural applications. Let Ω_x be an n-dimensional space, and let F_i and F_j be n-dimensional Gaussian distributions with their probability density functions

$$p_i(x) = \frac{1}{(2\pi)^{\frac{n}{2}}\, |K_i|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(x - \mu_i)^T K_i^{-1}(x - \mu_i)\right] \tag{4}$$

$$p_j(x) = \frac{1}{(2\pi)^{\frac{n}{2}} |K_j|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(x - \mu_j)^T K_j^{-1}(x - \mu_j)\right] \tag{5}$$

in which μ_i and K_i are the mean vector and covariance matrix respectively. Substituting Equation (4) and Equation (5) into Equation (2), we obtain

$$\rho(F_i,F_j) = \frac{|K_i^{-1}K_j^{-1}|^{\frac{1}{4}}}{|\frac{1}{2}(K_i^{-1} + K_j^{-1})|^{\frac{1}{2}}}$$

$$\exp\left[\frac{1}{4}(K_i^{-1}\mu_i + K_j^{-1}\mu_j)^T(K_i^{-1} + K_j^{-1})^{-1}(K_i^{-1}\mu_i + K_j^{-1}\mu_j) - \frac{1}{4}(\mu_i^T K_i^{-1}\mu_i + \mu_j^T K_j^{-1}\mu_i)\right] \tag{6}$$

When $K_i = K_j = K$ (equal covariance)

$$\rho(F_i,F_j) = \exp\left[-\frac{1}{8}(\mu_i - \mu_j)^T K^{-1}(\mu_i - \mu_j)\right] \tag{7}$$

When $\mu_i = \mu_j = \mu$

$$\rho(F_i,F_j) = \frac{|K_i^{-1}K_j^{-1}|^{\frac{1}{4}}}{|\frac{1}{2}(K_i^{-1} + K_j^{-1})|^{\frac{1}{2}}} \tag{8}$$

IV. System Implementation and Experimental Results

1. Description of Data

The proposed pattern classification technique in multiclass pattern recognition has been tested on IBM System 360/Model 44 digital computer by performing crop classification experiments. The data used in these experiments was gathered by an airborne multichannel optical-mechanical scanner. The output of this particular scanner provides twelve electrical signals, each one of which is proportional to radiant energy from the scene in a different spectral band. The twelve bands cover the range from 0.4 - 1.0 microns in the visible and near infrared portions of the spectrum. By simultaneously sampling the output of twelve channels, one obtains a vector which contains all the spectral information available about a given resolution element on the ground.

From the histogram study, the data can be fairly reasonably modeled by multivariate Gaussian distributions with unequal covariance matrices. The mean vector and covariance matrix of each class can be estimated from an adequate number of training samples as the sample mean and the sample covariance respectively.

2. General Procedure

(1) From the ground truth data, we first estimated the mean vectors and covariance matrices for all classes or subclasses considered. These estimations were used to characterize the distributions of all subclasses considered.

(2) For any given agricultural field with unknown crop types present, we assumed that the distribution of the feature vector is multivariate Gaussian. Therefore, we can obtain the empirical distribution by estimating the mean and the covariance as the sample mean and the sample covariance.

(3) We calculate the correlation function $\rho(\cdot,\cdot)$ between the empirical distribution of a particular test field and each of the distributions characterizing the known classes in accordance with the formula expressed in Equation (6).

(4) The final decision is made by choosing the class which maximizes the correlation function $\rho(\cdot,\cdot)$

3. Experimental Results

In order to evaluate the performance of the per field classifier, agricultural fields from areas C-2, C-3 and C-4 were used as test fields. All 12 features were employed by the per field classifier.

(1) Area C-4, July 1966 Data

Based on the histogram study, five subclasses each were chosen to represent the soybeans, the corn and the pasture. The stubble was represented by four subclasses. This was due to the multimodal distribution of each of the four classes. The statistical parameters were estimated from 4,832 training samples. The classification results are summarized in Table 1. The same data was classified by the MLDR (Maximum Likelihood Decision Rule) classifier, LARSYSAA, using the same set of training statistics, and these results together with the per field classification results are shown in Table 2 for comparison.

Table 1

Summary of per field classification results
(Data used taken from Area C-4 in July 1966)

Class	No. of Fields	Number of Fields Classified into Soybeans	Corn	Pasture	Stubble
Soybeans	19	17	2	0	0
Corn	31	1	30	0	0
Pasture	18	0	0	18	0
Stubble	13	0	0	3	10

Total: 81

Table 2

Summary of MLDR classification results
(Data used taken from Area C-4 in July 1966)

Spectral bands used were: 0.40-0.44, 0.52-0.55, 0.62-0.66, 0.66-0.72, 0.72-0.80

Classification Summary by Test Fields

Class	No. of Samples	Percent Correct	No. of Samples Classified into Soyb	Corn	Past	Stub	Thrs	Per Field	60% Classifier
Soyb	405	88.9	360	18	27	0	0	Soyb	Soyb
Soyb	361	93.1	336	11	2	8	4	Soyb	Soyb
Soyb	220	87.3	192	28	0	0	0	Soyb	Soyb
Soyb	320	76.6	245	30	45	0	0	Soyb	Soyb
Soyb	180	87.2	157	11	4	7	1	Soyb	Soyb
Soyb	297	76.1	226	38	11	0	22	Soyb	Soyb
Soyb	192	99.0	190	2	0	0	0	Soyb	Soyb
Soyb	672	92.7	623	31	18	0	0	Soyb	Soyb
Soyb	294	86.7	255	38	0	0	1	Soyb	Soyb
Soyb	80	18.8	15	56	3	5	1	Corn	Corn
Soyb	126	42.9	54	51	17	4	0	Corn	----
Soyb	638	84.6	540	70	2	0	26	Soyb	Soyb
Soyb	406	94.8	385	14	0	0	7	Soyb	Soyb
Soyb	280	71.4	200	50	30	0	0	Soyb	Soyb
Soyb	68	98.5	67	1	0	0	0	Soyb	Soyb
Soyb	85	95.3	81	2	1	1	0	Soyb	Soyb
Soyb	304	69.4	211	60	11	1	21	Soyb	Soyb
Soyb	224	61.6	138	36	30	20	0	Soyb	Soyb
Soyb	150	84.0	126	18	0	3	3	Soyb	Soyb
Corn	224	77.7	40	174	3	0	7	Corn	Corn

Table 2 (continued)

Class	No. of Samples	Percent Correct	No. of Samples Classified into Soyb	Corn	Past	Stub	Thrs	Per Field	60% Classifier
Corn	286	89.5	27	256	3	0	0	Corn	Corn
Corn	240	95.4	11	229	0	0	0	Corn	Corn
Corn	252	73.8	62	186	0	0	4	Corn	Corn
Corn	344	92.7	17	319	2	6	0	Corn	Corn
Corn	300	93.3	20	280	0	0	0	Corn	Corn
Corn	299	88.0	36	263	0	0	0	Corn	Corn
Corn	736	91.6	47	674	15	0	0	Corn	Corn
Corn	552	88.6	50	489	13	0	0	Corn	Corn
Corn	294	88.1	32	259	0	1	2	Corn	Corn
Corn	112	85.7	16	96	0	0	0	Corn	Corn
Corn	342	78.4	17	268	57	0	0	Corn	Corn
Corn	168	52.4	70	88	10	0	0	Corn	----
Corn	285	63.9	95	182	1	0	7	Corn	Corn
Corn	102	42.2	57	43	0	0	2	Soyb	----
Corn	273	28.6	195	78	0	0	0	Corn	Soyb
Corn	348	69.0	78	240	30	0	0	Corn	Corn
Corn	100	98.0	1	98	0	1	0	Corn	Corn
Corn	416	83.9	44	349	4	19	0	Corn	Corn
Corn	800	79.5	133	636	7	24	0	Corn	Corn
Corn	182	42.3	105	77	0	0	0	Corn	----
Corn	266	37.6	162	100	0	0	4	Corn	Soyb
Corn	204	96.6	2	197	0	5	0	Corn	Corn
Corn	198	27.3	134	54	0	0	10	Corn	Soyb
Corn	252	84.5	5	213	0	34	0	Corn	Corn
Corn	418	88.8	22	371	25	0	0	Corn	Corn
Corn	210	91.0	6	191	12	1	0	Corn	Corn
Corn	168	92.9	11	156	1	0	0	Corn	Corn
Corn	90	96.7	2	87	1	0	0	Corn	Corn
Corn	90	97.8	0	88	2	0	0	Corn	Corn
Corn	324	100.0	0	324	0	0	0	Corn	Corn
Past	408	97.3	0	0	397	6	5	Past	Past
Past	63	100.0	0	0	63	0	0	Past	Past
Past	276	92.8	5	0	256	15	0	Past	Past
Past	280	76.1	36	0	213	20	11	Past	Past
Past	323	71.5	60	31	231	1	0	Past	Past
Past	190	27.4	118	0	52	20	0	Past	Soyb
Past	180	60.6	0	0	109	71	0	Past	Past
Past	360	87.5	5	0	315	40	0	Past	Past
Past	288	27.8	177	7	80	24	0	Past	Soyb
Past	165	11.5	8	2	19	134	2	Past	Stub
Past	231	98.7	0	0	228	3	0	Past	Past
Past	242	82.2	11	0	199	32	0	Past	Past
Past	665	89.9	58	3	598	6	0	Past	Past
Past	300	76.7	0	0	230	70	0	Past	Past
Past	594	81.0	5	0	481	101	7	Past	Past
Past	162	3.7	38	8	6	110	0	Past	Stub
Past	200	92.5	11	0	185	4	0	Past	Past

Table 2 (concluded)

Class	No. of Samples	Percent Correct	No. of Samples Classified into Soyb	Corn	Past	Stub	Thrs	Per Field	60% Classifier
Past	306	97.4	1	2	298	5	0	Past	Past
Stub	259	95.4	1	1	9	247	1	Stub	Stub
Stub	260	72.7	0	0	71	189	0	Past	Stub
Stub	160	22.5	6	1	115	36	2	Past	Past
Stub	171	94.7	0	0	9	162	0	Stub	Stub
Stub	266	95.1	0	0	13	253	0	Stub	Stub
Stub	180	94.4	0	0	10	170	0	Stub	Stub
Stub	68	92.6	3	2	0	63	0	Stub	Stub
Stub	220	99.5	0	0	1	219	0	Stub	Stub
Stub	378	47.6	34	0	155	180	9	Past	----
Stub	243	90.5	11	0	0	220	12	Stub	Stub
Stub	315	82.2	46	0	10	259	0	Stub	Stub
Stub	234	98.3	4	0	0	230	0	Stub	Stub
Stub	189	85.2	27	0	0	161	1	Stub	Stub
Total	22353		6563	7687	4740	3191	172		

Overall Performance = 79.7

Classification Summary by Test Classes

Class	No. of Samples	Percent Correct	No. of Samples Classified into Soyb	Corn	Past	Stub	Thrs
Soyb	5302	83.0	4401	565	201	49	86
Corn	8875	79.6	1497	7065	186	91	36
Past	5233	75.7	533	53	3960	662	25
Stub	2943	81.2	132	4	393	2389	25
Total	22353		6563	7687	4740	3191	172

Overall Performance = 79.7
Average Performance by Class = 79.9

(2) Area C-3, July 1966 Data

Soybeans, corn, stubble, pasture mixtures and water are the four classes to be classified in area C-3. Due to different crop type and growing condition of the soybeans and the corn, each of them was represented by three subclasses. The pasture and the stubble were considered as a single class because of their closeness in distribution. The statistical parameters were estimated from 3,490 training samples. The classification results are summarized in Table 3. The same data was classified by the MLDR classifier using the same set of training statistics, and these results together with the per field classification results are shown in Table 4 for easy comparison.

Table 3

Summary of per field classification results
(Data used taken from Area C-3 in July 1966)

Class	No. of Fields	Number of Fields Classified into			
		Soybeans	Corn	Mixture	Water
Soybeans	13	11	1	1	0
Corn	10	0	9	1	0
Mixture	18	1	0	17	0
Water	3	0	0	0	3
Total:	44				

Table 4

Summary of MLDR classification results
(Data used taken from Area C-3 in July 1966)

Spectral bands used were: 0.66-0.72, 0.72-0.80, 0.80-1.00

Classification Summary by Test Fields

Class	No. of Samples	Percent Correct	No. of Samples Classified into					Per Field	60% Classifier
			Soyb	Corn	Mix	Water	Thrs		
Soyb	90	0.0	0	0	87	0	3	Past	Past
Soyb	270	24.4	66	204	0	0	0	Soyb	Corn
Soyb	105	100.0	105	0	0	0	0	Soyb	Soyb
Soyb	121	84.3	102	19	0	0	0	Soyb	Soyb
Soyb	168	75.6	127	41	0	0	0	Soyb	Soyb

Table 4 (Concluded)

Class	No. of Samples	Percent Correct	No. of Samples Classified into Soyb	Corn	Mix	Water	Thrs	Per Field	60% Classifier
Soyb	90	38.9	35	45	10	0	0	Soyb	----
Soyb	238	98.7	235	3	0	0	0	Soyb	Sovb
Soyb	70	68.6	48	22	0	0	0	Soyb	Sovb
Soyb	80	83.7	67	13	0	0	0	Soyb	Soyb
Soyb	182	68.1	124	58	0	0	0	Soyb	Sovb
Soyb	660	55.9	369	137	154	0	0	Soyb	----
Soyb	117	11.1	13	104	0	0	0	Corn	Corn
Soyb	189	79.9	151	38	0	0	0	Soyb	Soyb
Corn	196	89.8	20	176	0	0	0	Corn	Corn
Corn	140	93.6	7	131	2	0	0	Corn	Corn
Corn	171	31.6	117	54	0	0	0	Corn	Soyb
Corn	91	9.9	82	9	0	0	0	Corn	Soyb
Corn	100	91.0	9	91	0	0	0	Corn	Corn
Corn	143	58.0	60	83	0	0	0	Past	----
Corn	70	70.0	21	49	0	0	0	Corn	Corn
Corn	66	89.4	7	59	0	0	0	Corn	Corn
Corn	120	93.3	8	112	0	0	0	Corn	Corn
Corn	91	100.0	0	91	0	0	0	Corn	Corn
Mix	120	100.0	0	0	120	0	0	Mix	Mix
Mix	360	100.0	0	0	360	0	0	Mix	Mix
Mix	230	100.0	0	0	230	0	0	Mix	Mix
Mix	275	100.0	0	0	275	0	0	Mix	Mix
Mix	182	75.8	44	0	138	0	0	Mix	Mix
Mix	90	100.0	0	0	90	0	0	Mix	Mix
Mix	80	91.2	7	0	73	0	0	Mix	Mix
Mix	80	100.0	0	0	80	0	0	Mix	Mix
Mix	91	63.7	18	15	58	0	0	Soyb	Mix
Mix	100	92.0	7	1	92	0	0	Mix	Mix
Mix	126	100.0	0	0	126	0	0	Mix	Mix
Mix	132	0.0	25	107	0	0	0	Mix	Corn
Mix	169	99.4	1	0	168	0	0	Mix	Mix
Mix	96	100.0	0	0	96	0	0	Mix	Mix
Mix	168	100.0	0	0	168	0	0	Mix	Mix
Mix	117	100.0	0	0	117	0	0	Mix	Mix
Mix	117	100.0	0	0	117	0	0	Mix	Mix
Mix	192	100.0	0	0	192	0	0	Mix	Mix
Water	52	100.0	0	0	0	52	0	Water	Water
Water	28	96.4	0	0	1	27	0	Water	Water
Water	21	90.5	0	0	1	19	1	Water	Water
Total	6394		1875	1662	2755	98	4		

Overall Performance = 76.6

Classification Summary by Test Classes

Class	No. of Samples	Percent Correct	No. of Samples Classified into Sovb	Corn	Mix	Water	Thrs
Soyb	2380	60.6	1442	684	251	0	3
Corn	1188	72.0	331	855	2	0	0
Mix	2725	91.7	102	123	2500	0	0
Water	101	97.0	0	0	2	98	1
Total	6394		1875	1662	2755	98	4

Overall Performance = 76.6

Average Performance by Class = 80.3

(3) Area C-2, September 1966 Data

Sovbeans, corn, pasture, stubble and water were classified in area C-2. The statistical parameters were estimated from 3,094 training samples. The classification results are summarized in Table 5. The same set of data was classified by the MLDR classifier with the same set of training statistics, and the classification results together with the per field classification results are shown in Table 6.

Table 5

Summary of per field classification results
(Data used taken from Area C-2 in Sept. 1966)

Class	No. of Fields	Number of Fields Classified into				
		Soybeans	Corn	Pasture	Stubble	Water
Soybeans	12	8	4	0	0	0
Corn	14	1	13	0	0	0
Pasture	13	0	0	13	0	0
Stubble	11	0	1	1	9	0
Water	3	0	0	0	0	3
Total:	53					

Table 6

Summary of MLDR classification results
(Data used taken from Area C-2 in Sept. 1966)

Spectral bands used were: 0.40-0.44, 0.48-0.50, 0.52-0.55, 0.62-0.66, 0.66-0.72

Classification Summary by Test Fields

Class	No. of Samples	Percent Correct	No. of Samples Classified into						Per Field	60% Classifier
			Soyb	Corn	Past	Stub	Wtr	Thrs		
Soyb	114	95.6	109	4	1	0	0	0	Soyb	Soyb
Soyb	560	77.5	434	114	8	0	0	4	Soyb	Soyb
Soyb	96	95.8	92	3	1	0	0	0	Soyb	Soyb

Table 6 (continued)

Class	No. of Samples	Percent Correct	No. of Samples Classified into Soyb	Corn	Past	Stub	Wtr	Thrs	Per Field	60% Classifier
Soyb	594	16.0	95	479	3	14	0	3	Corn	Corn
Soyb	126	88.1	111	0	15	0	0	0	Soyb	Soyb
Soyb	306	94.1	288	18	0	0	0	0	Soyb	Soyb
Soyb	135	97.0	131	1	0	0	0	3	Soyb	Soyb
Soyb	338	8.6	29	303	0	5	0	1	Corn	Corn
Soyb	338	12.1	41	280	17	0	0	0	Soyb	Corn
Soyb	342	92.7	317	7	7	0	0	11	Corn	Soyb
Soyb	342	89.8	307	35	0	0	0	0	Soyb	Soyb
Soyb	513	0.0	0	456	47	10	0	0	Corn	Corn
Corn	95	96.8	0	92	0	3	0	0	Corn	Corn
Corn	324	69.1	96	224	3	1	0	0	Corn	Corn
Corn	340	87.6	40	298	1	1	0	0	Corn	Corn
Corn	836	93.8	0	784	1	51	0	0	Corn	Corn
Corn	140	92.9	6	130	0	4	0	0	Corn	Corn
Corn	110	21.8	76	24	7	3	0	0	Corn	Soyb
Corn	216	93.1	12	201	3	0	0	0	Corn	Corn
Corn	242	83.1	1	201	8	32	0	0	Corn	Corn
Corn	64	95.3	0	61	0	3	0	0	Corn	Corn
Corn	72	56.9	26	41	1	0	0	4	Corn	----
Corn	84	16.7	58	14	2	0	0	10	Soyb	Soyb
Corn	286	68.5	60	196	18	0	0	12	Corn	Corn
Corn	441	90.9	21	401	9	0	0	10	Corn	Corn
Corn	468	65.6	0	307	33	125	0	3	Corn	Corn
Past	76	81.6	0	1	62	13	0	0	Past	Past
Past	72	94.4	0	0	68	4	0	0	Past	Past
Past	630	91.4	19	0	576	29	0	6	Past	Past
Past	735	92.5	23	16	680	16	0	0	Past	Past
Past	255	96.5	9	0	246	0	0	0	Past	Past
Past	396	88.6	0	0	351	45	0	0	Past	Past
Past	240	92.1	0	0	221	19	0	0	Past	Past
Past	288	59.7	0	0	172	115	0	1	Past	----
Past	210	9.0	0	3	19	188	0	0	Past	Stub
Past	72	63.9	0	1	46	20	2	3	Past	Past
Past	176	65.9	0	6	116	54	0	0	Past	Past
Past	198	51.5	0	72	102	13	5	6	Past	----
Past	260	71.5	2	21	186	12	1	38	Past	Past
Stub	123	95.1	0	0	6	117	0	0	Stub	Stub
Stub	615	91.9	0	0	50	565	0	0	Stub	Stub
Stub	450	0.7	8	12	352	3	0	75	Stub	Past
Stub	132	42.4	0	2	74	56	0	0	Stub	----
Stub	120	28.3	0	19	61	34	0	6	Stub	----
Stub	420	0.5	0	418	0	2	0	0	Corn	Corn
Stub	288	95.1	0	2	12	274	0	0	Stub	Stub
Stub	285	98.6	0	0	4	281	0	0	Stub	Stub
Stub	459	27.9	0	7	324	128	0	0	Stub	Past
Stub	700	88.9	0	2	62	622	0	14	Stub	Stub
Stub	100	38.0	0	1	58	38	1	2	Past	----

Table 6 (concluded)

Class	No. of Samples	Percent Correct	No. of Samples Classified into Soyb	Corn	Past	Stub	Wtr	Thrs	Per Field	60% Classifier
Water	68	100.0	0	0	0	0	68	0	Water	Water
Water	39	97.4	0	0	1	0	38	0	Water	Water
Water	27	100.0	0	0	0	0	27	0	Water	Water
Total	14956		2411	5257	4034	2900	142	212		

Overall Performance = 67.0

Classification Summary by Test Classes

Class	No. of Samples	Percent Correct	No. of Samples Classified into Soyb	Corn	Past	Stub	Wtr	Thrs
Soyb	3804	51.4	1954	1700	99	29	0	22
Corn	3718	80.0	396	2974	86	223	0	39
Past	3608	78.9	53	120	2845	538	8	54
Stub	3692	57.4	8	463	1003	2120	1	97
Water	134	99.3	0	0	1	0	133	0
Total	14956		2411	5257	4034	2900	142	212

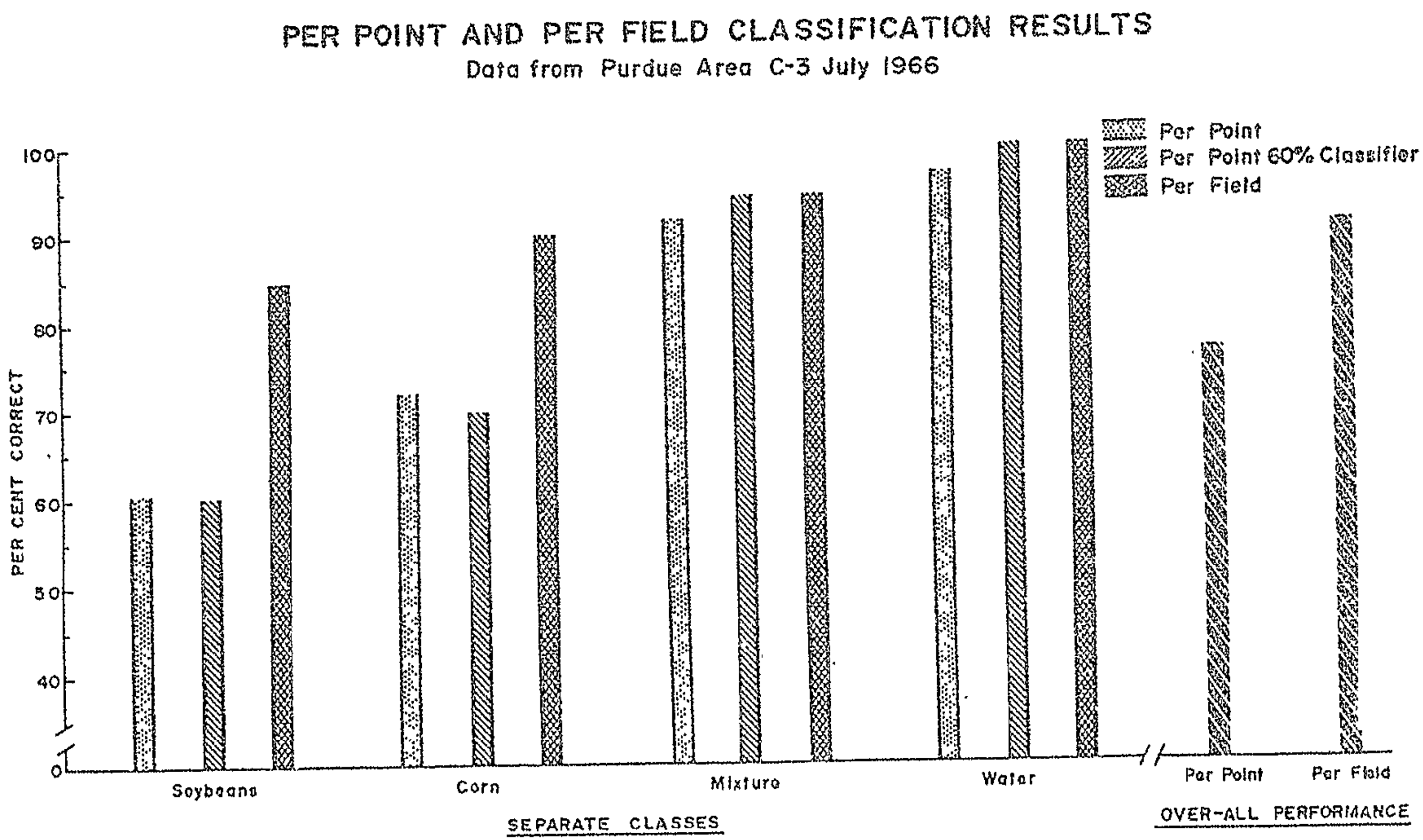
Overall Performance = 67.0
Average Performance by Class = 73.4

V. Summary

The results are summarized in the bar graphs of Figures 1, 2, and 3. As far as the given data is concerned, the performance of the proposed per field classifier is considered to be very satisfactory. We have not made any attempt to check the assumption of Gaussian distribution to justify the simplified classification scheme employed in this report, but the classification results encourage us to further test the applicability of the per field classifier to agricultural applications.

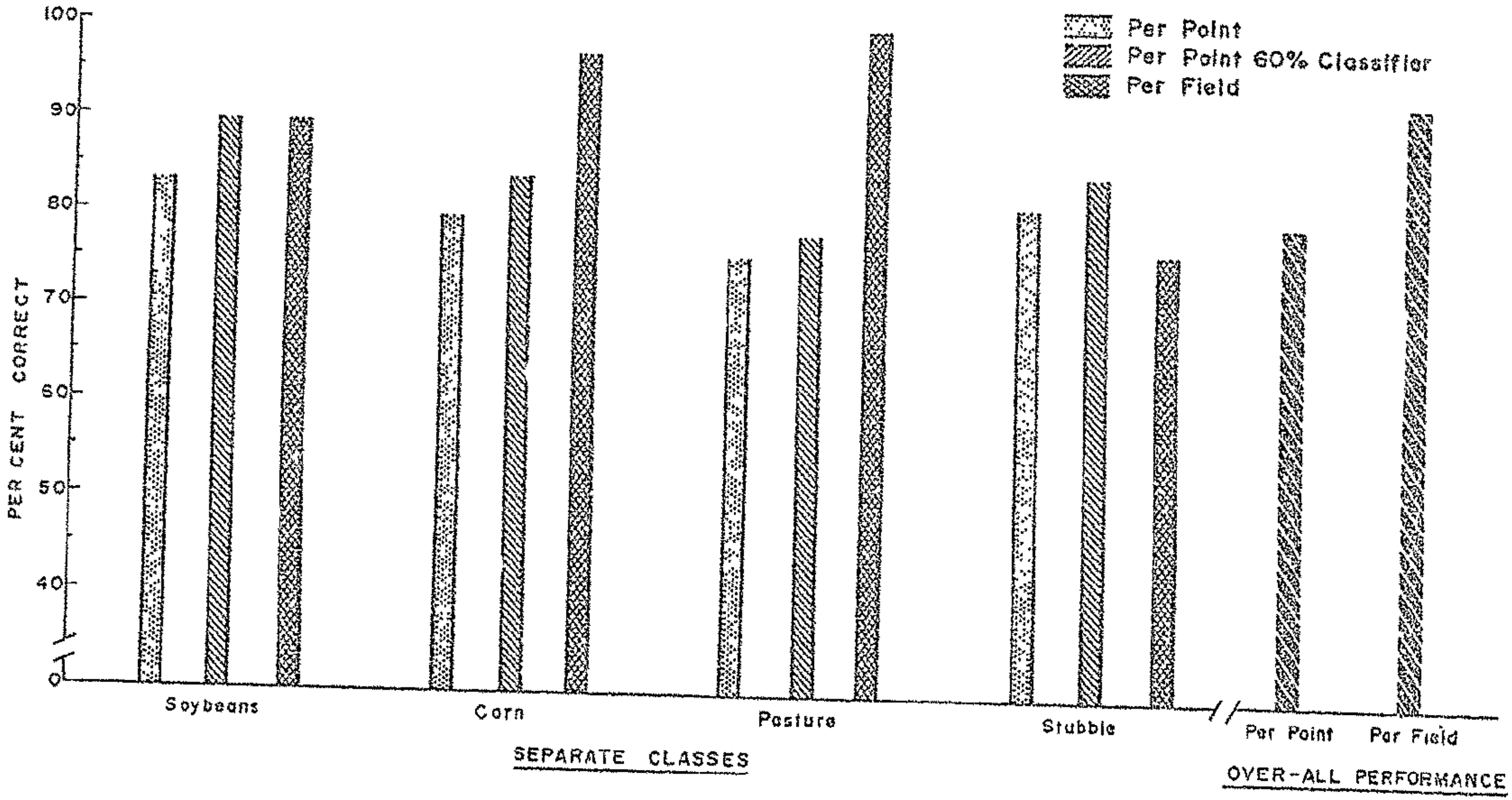
REFERENCE

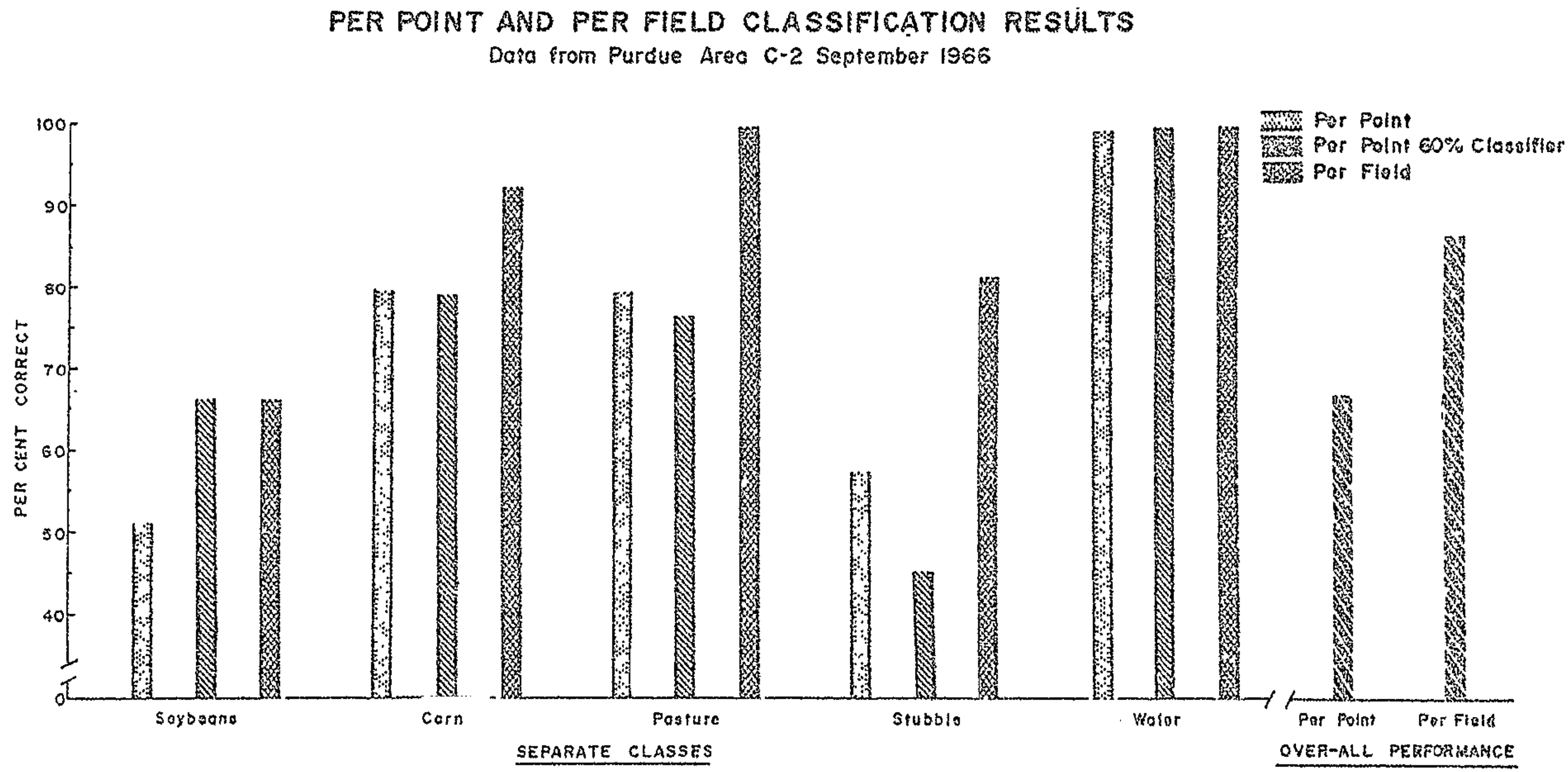
1. K. Matusita, "Classification Based on Distance in Multivariate Gaussian Cases," Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1, Part 1, 1965.


PER POINT AND PER FIELD CLASSIFICATION RESULTS
Data from Purdue Area C-3 July 1966
Per Point
Per Point 60% Classifier
Per Field
PER CENT CORRECT
100
90
80
70
60
50
40
0
Soybeans
Corn
Mixture
Water
Per Point
Per Field
SEPARATE CLASSES
OVER-ALL PERFORMANCE

PER POINT AND PER FIELD CLASSIFICATION RESULTS

Data from Purdue Area C-4 July 1966




PER POINT AND PER FIELD CLASSIFICATION RESULTS
Data from Purdue Area C-2 September 1966
Per Point
Per Point 60% Classifier
Per Field
PER CENT CORRECT
100
90
80
70
60
50
40
0
Soybeans
Corn
Pasture
Stubble
Water
Per Point
Per Field
SEPARATE CLASSES
OVER-ALL PERFORMANCE

APPENDIX E

LARS Information Note 040469
Purdue University

TESTING THE GAUSSIAN ASSUMPTION ON AIRCRAFT DATA

When feature selection and classification are done using LARSYSAA, the user makes the assumption that the data to be classified comes from a multivariate Gaussian distribution. This assumption is inherent in the algorithms used and if the data is far from Gaussian, results are unpredictable.

An experiment was conducted to test the assumption that aircraft data is Gaussianly distributed in each individual feature. Notice that the assumption that must be satisfied is that of multivariate Gaussian. If the data is Gaussian in each feature, this does not guarantee that it will be multivariate Gaussian. However, if it is not Gaussian in any particular feature, it is not multivariate Gaussian. Since no such test for multivariate data is known, it was felt that it would be useful to test the unvariate assumption.

The method used is that of the Chi-Square Goodness of Fit test. The data is histogrammed and a vector of observed values (OBS(I)) is formed. A vector of expected values corresponding to the observed values is calculated. Given the mean and standard deviation of the data, the expected (EXP(I)) number of samples falling in each bin (given a Gaussian distribution) is calculated. If k bins are formed, the test values

$$\chi^2 = \sum_{i=1}^{k} (OBS(I) - EXP(I))^2/EXP(I)$$

If χ^2 is less than $\chi^2_{1-\alpha}$ with k-3 degrees of freedom, the hypothesis that the data is Gaussianly distributed can be accepted with probability 1-α. Notice that with a constant number of degrees of freedom, the smaller the χ^2 value, the better the data fits the hypothesized density. (This test can be used to hypothesize any density.)

A program was written (CHi2-Serial No. DA 0011) to analyze aircraft data in this manner. Training fields and classes and large areas (500 lines) were tested. (See program output in file.)

The following is a sample output.

Class stub
No. of samples = 960
Channel 12
Mean 166.92
St. Dev. 4.56

(bin)	EXP	OBS	CHI
155	4.3	5	0.1
157	9.9	11	0.2
159	25.4	35	3.9
161	53.6	55	3.9
163	93.9	109	6.3
165	136.1	138	6.4
167	163.2	172	6.8
169	162.0	164	6.9
171	133.1	127	7.2
173	90.6	79	8.6
175	51.0	41	10.6
177	23.8	15	13.8
179	9.2	5	15.7
225	3.9	4	15.7

XCHI = 15.7 with 11 degrees of freedom

If that χ^2 value with 13 degrees of freedom is looked up in a Chi-Square table, it can be seen that the hypothesis can be accepted with probability .10. After testing a random sample of fields and classes in this manner, it was found that accepting the hypothesis with even that much probability was unusual. Most times, the χ^2 values obtained were much bigger and off the table.

A closer look at the observed and expected values for each bin shows that most discrepancies between the two occur at each end. The aircraft data does not have enough of a range to be really Gaussianly distributed.

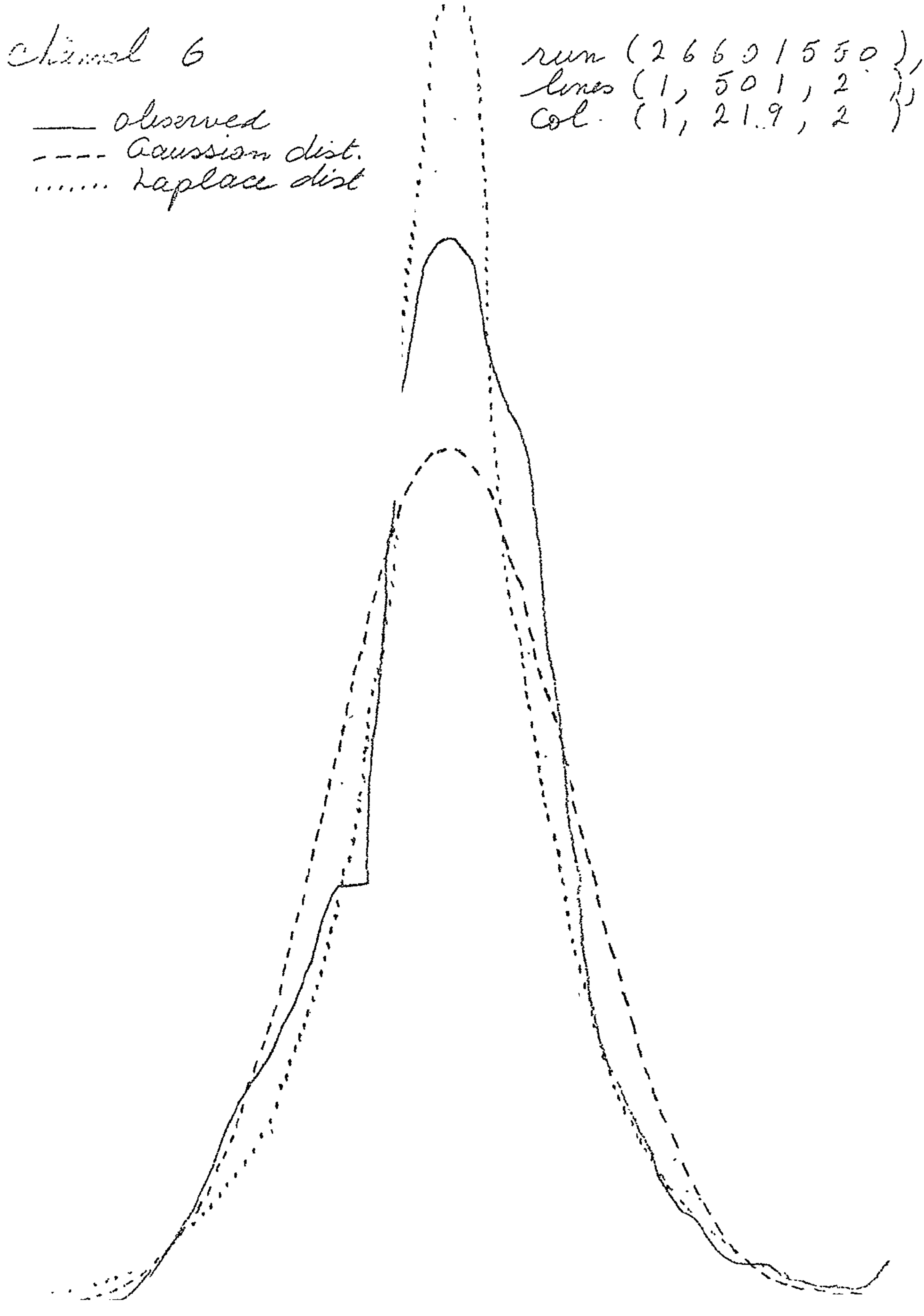
A Laplace distribution was also fitted to the data. A Laplace density resembles a Gaussian distribution in general shape but has a much higher kurtosis (peak), and a smaller range. x is said to be distributed as a Laplace distribution if

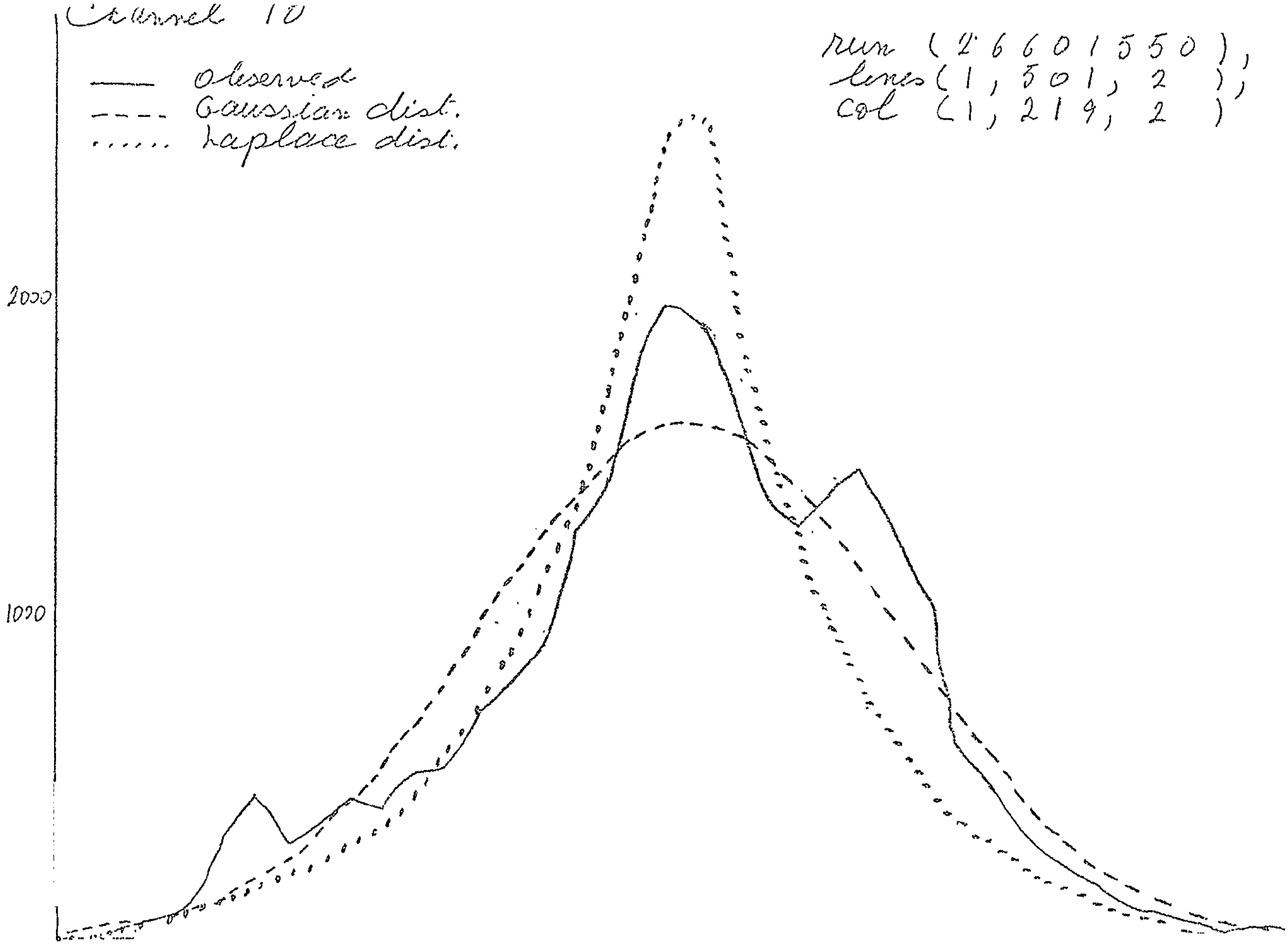
$$p(x) = \int_{-\infty}^{\infty} \frac{1}{2\beta} \cdot e^{-\frac{|x-\alpha|}{\beta}} \qquad -\infty < x < \infty$$

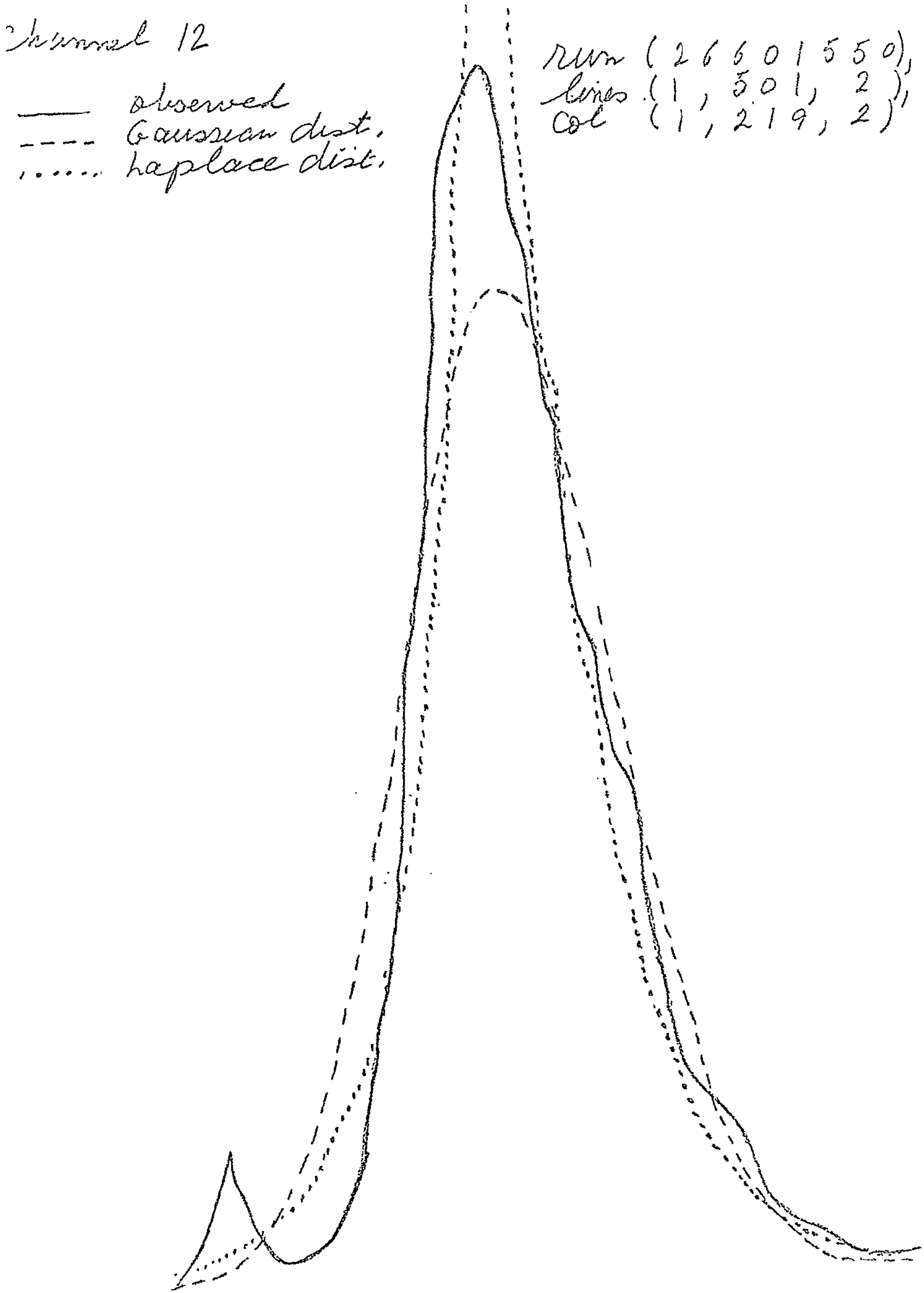
where α is the mean and $2\beta^2$ is the variance.

Results were about of the same magnitude. A Laplace distribution fit the data better in some channels, worse in others. However, the hypothesis that the data was distributed as a Laplace distribution could not be accepted with greater probability. The accompanying graphs show an area of 500 lines of run 26601550 (September 1966 - C3). The observed data lies in between the Gaussian and Laplace distribution. In certain places (channel 10) it is bimodal.

The aircraft data analyzed was found not to be Gaussianly distributed. Nothing can really be said about the population to which it belongs. However, because of the small range of the data, it is doubtful whether the population is Gaussian. It is therefore possible that errors occurring between classes that are close together are due to non-Gaussian data.


channel 6
run (26601550),
lines (1, 501, 2),
col (1, 219, 2)
—— observed
---- Gaussian dist.
...... Laplace dist


Channel 10
observed
Gaussian dist.
Laplace dist.
run (26601550),
lines (1, 501, 2);
col (1, 219, 2)
2000
1000


Channel 12
observed
Gaussian dist.
Laplace dist.
run (26601550),
lines (1, 501, 2),
col (1, 219, 2)

LARS Information Note 040869
Purdue University

An Automated Method of Choosing Training Fields

by

Danielle R. Bernstein

When a flight line of data is to be analyzed using LARSYSAA, a set of representative training fields must be chosen. This set must be able not only to recognize (classify) itself but to recognize other fields (test fields) as well. Therefore it is important that an automatic method be devised to give the user an initial set of training fields. If necessary, this set can be refined in subsequent classifications to make it more representative of the test fields.

An experiment was conducted to evaluate such a method. A program was written (PPS - Serial No. DA 0013) to choose training fields within each class with probability proportional to size (PPS).

The fields are chosen in the following manner. The number of samples in each field is calculated (Mi) and a sum total is kept (ΣMi). Let the following table be an example.

Unit	Size Mi	ΣMi	Assigned Range
1	3	3	1- 3
2	1	4	4
3	11	15	5-15
4	6	21	16-21
5	4	25	21-25

A random number is chosen between 1 and 25. Suppose it is 19. In the sum, number 19 falls in unit 4 which covers numbers 16 to 21 inclusive. With this method of drawing, the probability that any unit is selected is proportional to the size of the unit. Sampling is done without replacement and therefore unit 4 cannot be picked again. In addition, the program will optionally scale the training fields to a desired number of samples (through manipulation of the line and column intervals) so that no one field will have an undue influence on the statistics of the training classes.

With this method, notice that:

1) The number of fields and the range of their sample size are selected by the user.

2) The training fields are chosen from all the fields which can be outlined and for which there is ground truth, i.e. from all fields to be used as test fields No attempt is made to delete the nonuniform or atypical fields nor are the boundaries changed so that only the center part is used.

3) The fields are chosen from each class to be separated (not subclasses) and the histograms of the training classes may not be unimodal. However, this results in fewer training classes and faster classification time.

Two runs (September, 66-C3 and C4) were each analyzed twice (see computer output in file) in this manner and compared with their analysis done previously in the conventional manner (see Information Note D022469).

C3 - Five classes were to be separated; soybeans, corn, stubble, forage and water. Water consisted of only one segment which was not picked by PPS. The following number of fields were chosen initially.

Class	Training	Test
Soybeans	4	10
Corn	5	13
Stubble	4	10
Forage	4	9
Water	1	1

Histograms revealed that although none of the classes were unimodal, the stubble and forage classes were so bimodal that some adjustment had to be made. One field was removed from the stubble class and forage was divided into two subclasses; FRG1, and FRG2 made up of two fields each. The feature selection processor showed that there was poor separation between soybeans and corn but adequate-to-good separation between the rest of the fields. Features 1, 6, 9; 10 were chosen.

The classification results for training classes were 80.7% for overall performance compared with 91.9% for the previously documented classification. This is to be expected as the training classes were not unimodal and not picked for their uniformity. The test class results were 72.4%. This compares favorably with 71.0% obtained in the previous classification. This would tend to indicate that the training fields, when picked at random, do not result in worse classification of the test set than when carefully chosen through several iterations. The latter (previously documented) classification took three or more iterations to arrive at the final set of training samples. Another classification of C3 using five classes gave similar results - 81.8% for training fields and 69.6% for test fields.

C4 - C4 was also classified twice in this manner. Only 4 classes (and no subclasses) were needed; soybeans, corn, stubble and forage. The original classification resulted in 91.0% overall recognition of training classes and 73.1% for test classes. The classification, with training fields chosen with PPS, used channels 1, 6, 10, 12. The classification resulted in 80.6% overall performance for the training classes and 69.6% for test classes. For the second classification of C4, the same classes and features 1, 6, 9, 10 were used. Training class recognition was 85.2% and test class recognition was 68.0%.

Although test field recognition was a little lower in C4, this decrease has to be judged in the light of the number of classifications performed to arrive at the final results. When classifying aircraft data in the conventional manner, the user is essentially working with his training fields. He refines his set and tries to improve his training fields classification performance. Only when he is satisfied with this performance, does he really concentrate on his test fields. However, it is test field accuracv that is important and which must be improved.

LARSYSAA users are encouraged to employ this program to obtain at least a starting training set, "Mild" multimodality can be ignored. In this way, fewer subclasses will be used and higher accuracy with fewer iterations will result.