AgRISTARS

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

Supporting Research

May 1982

Final Report

Remote Sensing of Agricultural Crops and Soils

Annual Technical Summary, Contract NAS9-15466

by M.E. Bauer and Staff

Purdue University
Laboratory for Applications of Remote Sensing
West Lafayette, Indiana 47907

SR-P2-04266
NAS9-15466
LARS 113081

Final Report

REMOTE SENSING OF AGRICULTURAL CROPS AND SOILS

M.E. Bauer and Staff

Purdue University
Laboratory for Applications of Remote Sensing
West Lafayette, Indiana 47907

May 1982

Star Information Form

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
SR-P2-04266		

4. Title and Subtitle	5. Report Date
Remote Sensing of Agricultural Crops and Soils	May, 1982
	6. Performing Organization Code

7. Author(s)	8. Performing Organization Report No.
M.E. Bauer and Staff	LARS 113081

9. Performing Organization Name and Address	10. Work Unit No.
Laboratory for Applications of Remote Sensing Purdue University 1220 Potter Drive West Lafayette, IN 47906	11. Contract or Grant No. NAS9-15466

12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered
NASA Johnson Space Center Earth Resources Research Division Houston, TX 77058	Final
	14. Sponsoring Agency Code

15. Supplementary Notes

Forrest G. Hall, Technical Monitor
Marvin E. Bauer, Principal Investigator

16. Abstract

This report presents a summary of the research results and accomplishments for contract NAS9-15466, Remote Sensing of Agriculture and Earth Resources, to Purdue University for the period December 1980 to May 1982. Results of sixteen tasks in the following areas are described:

- Corn and soybean scene radiation research
- Soil moisture research
- Sampling and aggregation research
- Pattern recognition and image registration research
- Computer and data base services

17. Key Words (Suggested by Author(s))	18. Distribution Statement
Remote sensing, crops and soils, spectral, reflectance, pattern recognition, digital analysis, computer data processing	

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22 Price
Unclassified	Unclassified	208	

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

NASA - JSC

TABLE OF CONTENTS

I. Introduction

INTRODUCTION

This report presents a summary of the research results and accomplishments for contract NAS9-15466, Remote Sensing of Agriculture and Earth Resources, to Purdue University for the period December 1980 to May 1982. Results in the following areas are described:

- Corn and soybean scene radiation
- Soil moisture research
- Sampling and aggregation research
- Pattern recognition and image registration research
- Computer and data base services

Unlike in previous years of the contract, this final report is a summary, rather than a detailed multi-volume report. Additional detailed information describing the various tasks, experiments, and results are available from the technical reports published during the past 18 months (see Table I-1).

Significant accomplishments over the past year include:

- Completion of the development and procurement of the multiband radiometer system for remote sensing field research. Fifteen units will be in use at eight universities, three NASA centers and one international research center during the 1982 growing season.

- Continued development of field research spectral-agronomic data base including additional spectral data, documentation and data processing and analysis software. Data were provided to eleven different research groups during 1981.

- Quantitative description of the relationship of leaf area index and percent canopy cover to the multispectral reflectance characteristics of crop canopies. These relationships led to the conceptualization of how to utilize spectrally derived inputs to crop growth and yield models. A method was developed for estimating solar radiation interception from spectral data. This variable, integrated over the season, has been found to explain much of the variation in corn and soybean yields and with satellite observations will enable crop yield models to be implemented over large areas.

- Development and evaluation of an algorithm utilizing contextual (spatial) relationships in agricultural scenes to improve classification accuracy compared to pixel-wise methods.

- Research on methods of geometrically aligning multiple images and maps of the same scene to enable multivariate analysis has led to improved capability for registration of digital imagery at NASA/JSC and LARS.

- Installation of all systems code on the AS/3000 computer at JSC and training of operations and systems groups; survey of data base management systems and installation at JSC of ADABASE and conversion of RT&E data base to run under ADABASE/NATURAL; design and partial implementation of network interfacing JSC and LARS computer systems.

Table I-1. Technical reports prepared for contract NAS9-15466, December 1980 to May 1982 (AgRISTARS/LARS Technical Report numbers).

SR-P0-00454 030181	Pixel Labeling by Supervised Probabilistic Relaxation - J.A. Richards, D.A. Landgrebe and P.H. Swain.
SR-P1-04038 021781	Soybean Canopy Reflectance as Influenced by Cultural Practices - J.C. Kollenkark, C.S.T. Daughtry and M.E. Bauer.
SR-P1-04039 021681	Canopy Reflectance as Influenced by Solar Illumination Angle - J.C. Kollenkark, V.C. Vanderbilt, C.S.T. Daughtry and M.E. Bauer.
SR-P1-04044 030381	Effects of Nitrogen Nutrition on the Growth Yield and Reflectance Characteristics of Corn Canopies - G. Walburg, M.E. Bauer and C.S.T. Daughtry
SR-P1-04079 040981	Design and Evaluation of a Pick-Up Truck Mounted Boom for Evaluation of a Multiband Radiometer System - R. Tsuchida.
SR-P1-04090 061082	Performance Comparison for Barnes Model 12-1000, Exotech Model 100, and Ideas Inc. Biometer Mark II - B. Robinson.
SR-P1-04125 042081	On the Accuracy of Pixel Relaxation Labeling - J.A. Richards, D.A. Landgrebe and P.H. Swain.
SR-P1-04139 090981	Linear Polarization of Light by Two Wheat Canopies Measured at Many View Angles - V.C. Vanderbilt, L.L. Biehl, B.F. Robinson, M.E. Bauer and A.S. Vanderbilt.
SR-P1-04140 090881	Diurnal Changes in Reflectance Factor Due to Sun-Row Direction Interactions - V.C. Vanderbilt, J.C. Kollenkark, L.L. Biehl, B.F. Robinson, M.E. Bauer and K.J. Ranson.
SR-P1-04141 060881	Application of Computer Axial Tomography (CAT) to Measuring Crop Canopy Geometry - V.C. Vanderbilt and R.W. Kilgore.
SR-P1-04148 072981	Incorporating Spatial Context Into Statistical Classification of Multidimensional Image Data - J.C. Tilton and P.H. Swain.
SR-P1-04159 062379	A Method for Classifying Multispectral Remote Sensing Data Using Context - P.H. Swain, H.J. Siegel and B.W. Smith.
SR-P1-04160 062779	Sampling for Area Estimation: A Comparison of Full-Frame Sampling With the Sample Segment Approach - M.M. Hixson, M.E. Bauer and B.J. Davis.
SR-P1-04160 091881	Sampling Landsat Classifications for Crop Estimation - M.M. Hixson, B.J. Davis and M.E. Bauer. (Revised)
SR-P1-04163 041080	Analytical Design of Multispectral Sensors - D.J. Wiersma and D.A. Landgrebe.

Table I-1. (continued)

SR-P1-04164 121280	Overcoming Accuracy Deterioration in Pixel Relaxation Labeling - J.A. Richards, D.A. Landgrebe and P.H. Swain.
SR-P1-04165 021881	Evaluation of Several Schemes for Classification of Remotely Sensed Data - M. Hixson, D. Scholz, N. Fuhs and T. Akiyama.
SR-P1-04166 042380	Contextual Classification of Multispectral Remote Sensing Data Using A Multiprocessor System - P.H. Swain, H.J. Siegel and B.W. Smith.
SR-P1-04167 060380	Parallel Processing Implementations of a Contextual Classifier for Multispectral Remote Sensing Data - H.J. Siegel, P.H. Swain and B.W. Smith.
SR-P1-04168 072180	The Development of Spectral-Spatial Classifier for Earth Observational Data - D.A. Landgrebe.
SR-P1-04169 041580	A Parametric Model for Multispectral Scanners - B.G. Mobasseri, P.E. Anuta and C.D. McGillem.
SR-P1-04170 060580	A Model of Plant Canopy Polarization Response - V.C. Vanderbilt.
SR-P1-04186 111579	Atlas of Soil Reflectance Properties - E.R. Stoner, B.F. Baumgardner, L.L. Biehl and B.F. Robinson.
SR-P1-04187 091281	Spectral-Agronomic Relationships of Corn, Soybean and Wheat Canopies - M.E. Bauer, C.S.T. Daughtry and V.C. Vanderbilt.
SR-P1-04191 111481	Variability of Reflectance Measurements With Sensor Altitude and Canopy Type - C.S.T. Daughtry, V.C. Vanderbilt and V.J. Pollara.
SR-P1-04194 101481	Multistage Classification of Multispectral Earth Observation Data: The Design Approach - M.J. Muasher and D.A. Landgrebe.
SR-P1-04197 101281	Determination of the Optimal Level for Combining Area and Yield Estimates - M.M. Hixson and C.D. Jobusch.
SR-P1-04200 071380	Spectral Properties of Agricultural Crops and Soils Measured from Space, Aerial, Field and Laboratory Sensors - M.E. Bauer, V.C. Vanderbilt, B.F. Robinson and C.S.T. Daughtry.
SR-P1-04201 071480	A Multiband Radiometer and Data Acquisition System for Remote Sensing Field Research - B.F. Robinson, M.E. Bauer, D.P. DeWitt, L.F. Silva and V.C. Vanderbilt.
SR-P1-04202 071580	Simulated Response of a Multispectral Scanner Over Wheat as a Function of Wavelength and View/Illumination Directions - V.C. Vanderbilt, B.F. Robinson, L.L. Biehl, M.E. Bauer and A.S. Vanderbilt.

Table I-1. (continued)

FC-P1-04121 062381	Evaluation of a Segment-Based Landsat Full-Frame Approach to Crop Area Estimation - M.M. Hixson, S.M. Davis and M.E. Bauer.
SR-P1-04164 052680	Extension of Laboratory-Measured Soil Spectra to Field Conditions - E.R. Stoner, M.F. Baumgardner, R.A. Weismiller, L.L. Biehl and B.F. Robinson.
SR-P1-04160 091881	Sampling Landsat Classifications for Crop Area Estimation - M.M. Hixson, B.J. Davis and M.E. Bauer.
SR-P2-04230 060980	Development of a Digital Data Base for Reflectance-Related Soil Information - E.R. Stoner and L.L. Biehl.
SR-P2-04236 030182	Spectral Estimates of Solar Radiation Intercepted by Corn Canopies - C.S.T. Daughtry, K.P. Gallo and M.E. Bauer.
SR-P2-04259	Field Measurements and Simulation Modeling of Corn and Soybean Moisture Stress - 1981 Field Studies - B.L. Blad, J.M. Norman and B.R. Gardner (CAMaC Progress Report 82-2).
SR-P2-04266 113081	Remote Sensing of Agricultural Crops and Soils - Annual Technical Summary, Contract NAS9-15466 - M.E. Bauer and Staff.
SR-P2-04301 121880	Characteristic Variations in Reflectance of Surface Soils - E.R. Stoner and M.F. Baumgardner.
SR-P2-04318 061182	Performance Evaluation and Calibration of a Modular Multiband Radiometer for Remote Sensing Field Research - B.F. Robinson, R.E. Buckley and J.A. Burgess

II. Corn and Soybean Scene Radiation Research

1. SPECTROMET CROP DEVELOPMENT STAGE ESTIMATION

Under the general objective of researching and developing models combining spectral and meteorological data for estimation of crop development stage, three tasks, described in this section, were pursued. The first addressed the question of the basic spectral separability of corn and soybean development stages. The second initiated work on the potential use of spectral-temporal vectors to estimate development stage. The third developed and evaluated several meteorological models of crop development which might be combined with a spectral model.

A. Assessing Crop Development Stage with Spectral Data

V.J. Pollara, C.S.T. Daughtry, and V.C. Vanderbilt

Introduction

Crop simulation models requiring knowledge of crop development stage are becoming increasingly important for forecasting yields of several economically important crops. Although extensive data collection programs exist in the United States through the U.S.D.A. and agricultural universities, development stage estimates are costly to compile, can be quite variable between states and may not be sufficiently accurate for some simulation models. Furthermore, many nations often have little or no system for collecting such information. Estimates of crop development stage by remote sensing techniques are potentially more cost effective for large areas than present labor intensive estimates and may provide the only estimates in many areas.

The objective of this study was to assess the spectral separability of development stages of corn and soybean crops. One unitemporal and three multitemporal methods for identifying crop development stage using multispectral data were examined.

Materials and Methods

The spectral data were Exotech Model 100A measurements in four wavelength bands (0.5-0.6, 0.6-0.7, 0.7-0.8, 0.8-1.1 μm) over field plots of soybeans and corn at the Purdue University Agronomy Farm, West Lafayette, Indiana, in 1978, 1979, and 1980. At approximately ten day intervals during the growing seasons, spectral data were collected on clear or nearly cloudless days, taking two measurements per field plot. The treatments of the soybean and corn field experiments are outlined in Table 1-1. A total of 25 different development stages for corn and 18 development stages for soybeans were recorded throughout the growing seasons.

Table 1-1. Treatments of the field experiments over which data were analyzed.

	Soybean			Corn	
	1978	1979	1980	1979	1980
Soil Type	Raub	Russell Chalmers	Toronto Chalmers	Toronto Chalmers	Fincastle Chalmers
Plant Pop. (Plants/ha.)	288,000	266,000	266,000	25,000 50,000 75,000	25,000 50,000 75,000
Row Spacing (cm)	15	25 75	25 75	76	76
Planting Dates	3 May	10 May 24 May 15 June	1 May 20 May 10 June	2 May 16 May 30 May	1 May 15 May 30 May
Cultivar	Amsoy 71	Amsoy 71 Williams	Amsoy 71 Williams	Beck 65X	Beck 65X

Raub-silt loam-mesic Aquic Argiudoll, Russell-silt loam-typic Hapludalf, Chalmers-silty clay loam-typic Haplaquoll, Toronto-fine silt-udollic Ochraquolf, Fincastle-fine silt-mesic Aeric Ochraqualf.

Results and Discussion

Graphical Analysis

When brightness is plotted versus greenness (Figure 1-1) the data scatter assumes the same tasselled cap outline described by Kauth and Thomas (1976). Figure 1-2 presents brightness vs. greenness scatterplots of corn data for individual, successive development stages. Scatterplots of soybean data grouped by development stages appear in Figure 1-3.

Estimating Development Stage with Unitemporal Data

The unitemporal method used discriminant analysis to indicate separability of development stages on the basis of spectral data acquired for each development stage. In all analyses, spectral data and all the derived spectral variables were assumed to have Gaussian characteristics and the individual covariance matrices for each group were used in the calculations; the data used to train the classifier were also used as the test data (reclassification).

With the MSS bands as variables, classification results are shown for discriminant analysis of nine training classes of corn (Table 1-2), and soybeans (Table 1-3). Misclassified data vectors are not restricted only to neighboring development stage classes. Confusion classes are often classes at the opposite ends of the growing season. Higher classification accuracy was obtained for early and late season development stages than for midseason development stages. Grouping development stages for a whole season into a smaller number of development stage classes increases overall classification performance.

It should be noted here that the classifications include all plant populations (corn) and row spacings (soybean) which strongly influence the amount of green vegetation. The results indicate that spectral response is probably more sensitive to amount of green vegetation than to development *per se*.

Estimating Development Stage with Multitemporal Data

Of the several multitemporal analysis methods tested, the best method employed an eight variable discriminant function, using the reflectance in four bands at two development stages. Four of the variables in the eight variable analyses were one multispectral observation over a field plot at the first development stage of a development stage interval. The other four variables were the multispectral observation of the same plot at the second development stage of the interval.

While conceptually simple, the eight variable analysis method becomes quite complicated when it is used to estimate crop development stages because of the many development stage classes. To better illustrate the method, consider an eight variable analysis with fewer classes, the

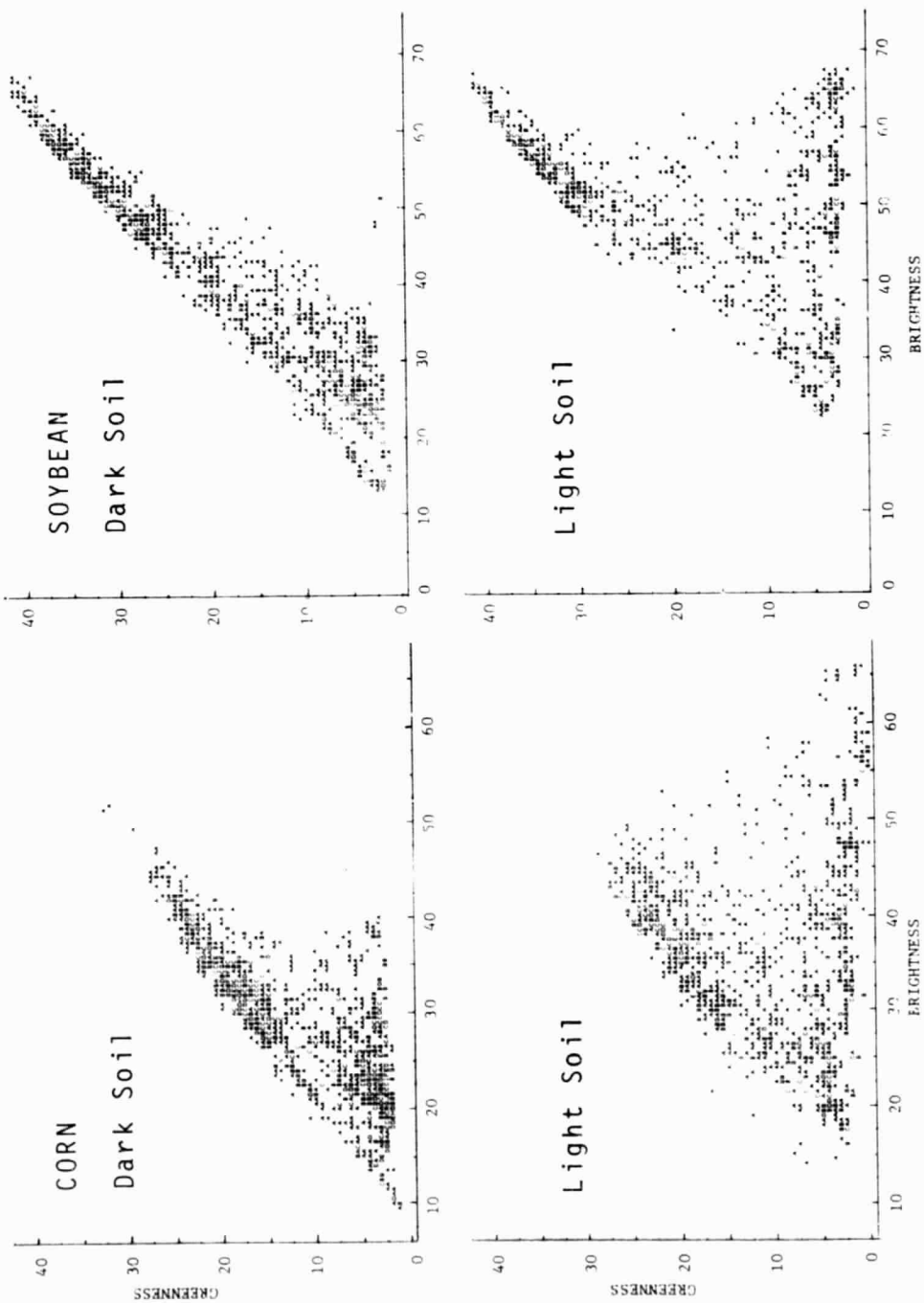


Figure 1-1. Plots of greenness vs. brightness for corn and soybean grown on dark and light soils.

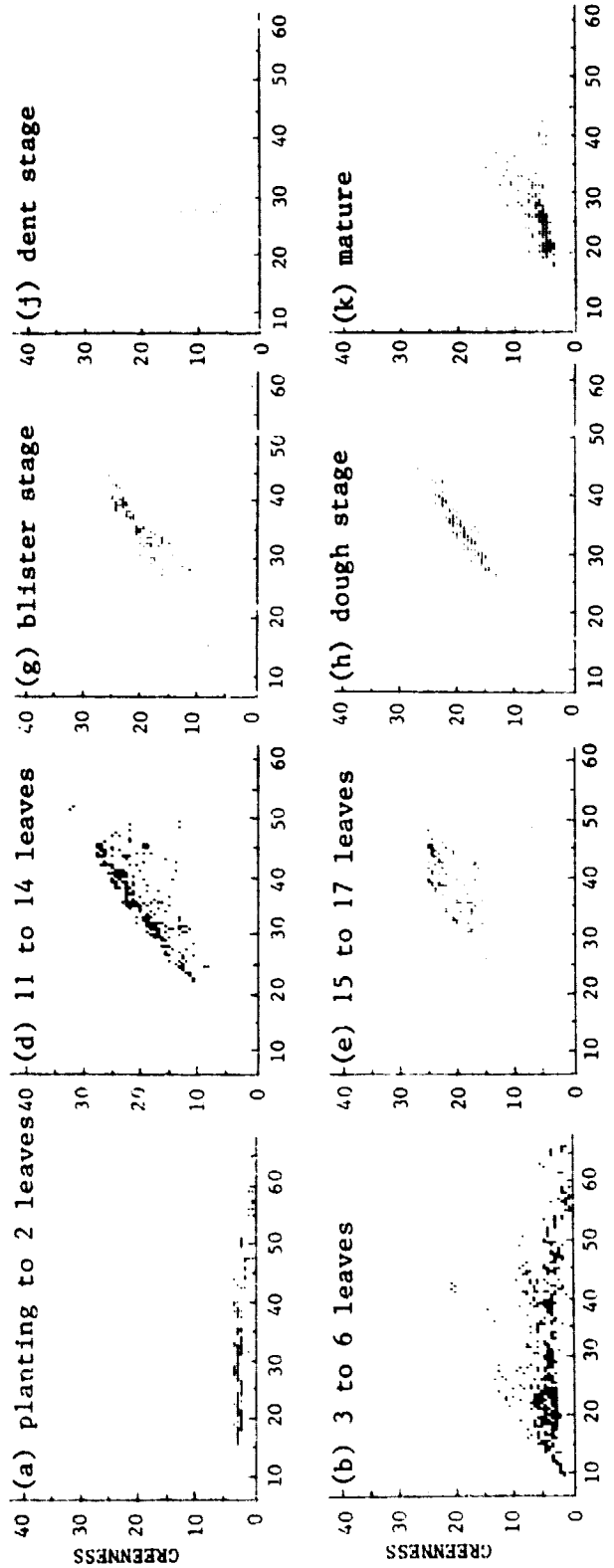


Figure 1-2. Greenness-brightness plots for successive development stages of corn.

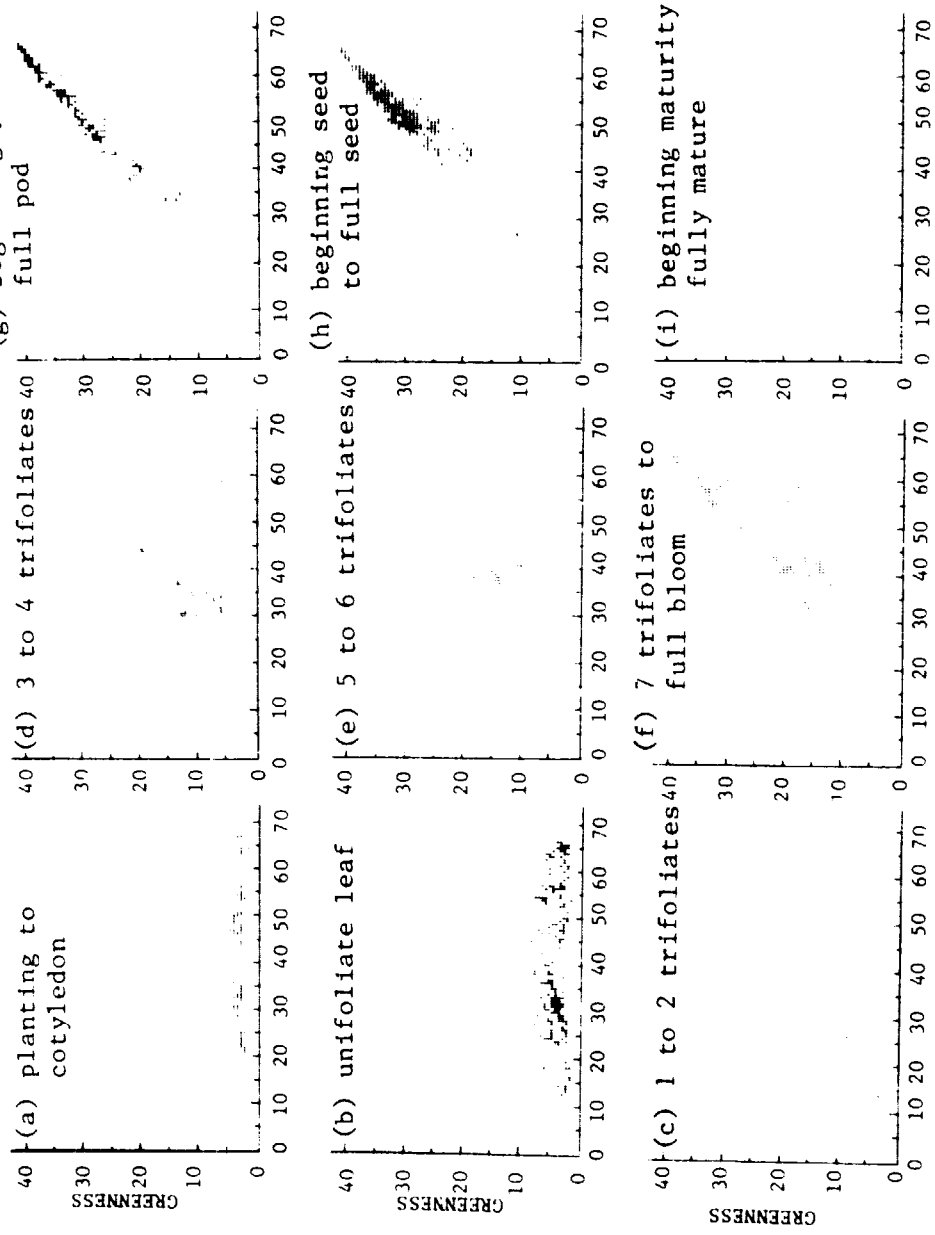


Figure 1-3. Greenness-brightness plots for successive development stages of soybean.

Table 1-2. Classification matrix for nine development stages of corn.

	PERCENT CLASSIFIED INTO CLASS									
TRUE CLASS	0-0.5	0.75-1.5	1.75-2.5	2.75-3.5	3.75-6.0	7.0	8.0	9.0	10.0	TOTAL
0-0.5	95.2	4.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	208
0.75-1.5	40.2	48.0	8.7	0.9	0.2	0.0	0.0	1.8	0.2	550
1.75-2.5	2.0	23.2	42.3	21.4	2.5	1.3	4.6	2.6	0.0	392
2.75-3.5	0.0	0.0	11.5	58.3	4.8	10.7	11.1	3.6	0.0	252
3.75-6.0	0.0	0.7	4.7	34.6	11.4	30.5	17.8	0.3	0.0	298
7.0	0.0	0.0	1.3	10.1	2.0	63.8	22.8	0.0	0.0	149
8.0	0.0	0.0	0.0	8.5	3.8	34.0	46.2	7.6	0.0	106
9.0	0.0	1.1	2.2	4.9	0.6	1.6	9.9	56.6	23.1	182
10.0	0.0	2.4	0.0	0.0	1.2	0.0	0.0	8.9	87.5	168

Table 1-3. Classification matrix for nine development stages of soybean.

	PERCENT CLASSIFIED INTO CLASS									
TRUE CLASS	0-0.25	0.5-0.75	1.0-1.5	2.0-2.5	3.0-3.5	4.0-5.0	6.0-7.0	8.0-9.0	10.0-11.0	TOTAL
0-0.25	89.2	8.8	1.6	0.0	0.5	0.0	0.0	0.0	0.0	194
0.5-0.75	44.9	26.7	21.7	4.2	1.0	0.0	0.0	0.0	1.4	285
1.0-1.5	10.5	10.5	61.3	12.1	3.2	0.8	0.0	0.0	1.6	124
2.0-2.5	1.6	10.6	19.7	39.9	19.7	6.9	0.5	0.0	1.1	188
3.0-3.5	1.2	0.6	16.9	19.3	41.6	20.5	0.0	0.0	0.0	166
4.0-5.0	0.0	0.0	7.4	8.4	16.8	45.5	13.9	7.9	0.0	202
6.0-7.0	0.0	0.0	0.5	1.4	4.7	3.7	74.8	15.0	0.0	214
8.0-9.0	0.0	0.1	1.0	0.7	2.0	11.8	32.7	50.2	1.4	697
10.0-11.0	3.2	2.0	0.8	0.0	0.4	1.6	0.0	1.6	90.3	248

problem of identifying those forest areas clearcut between two Landsat overpasses. There are only four classes to which a pixel may belong: (1) forest during first overpass - forest during second overpass, (2) forest - clearcut, (3) clearcut - forest, and (4) clearcut - clearcut. In this hypothetical example each pixel would be classified from statistics developed from training areas in the eight dimensional data. The classification results would be presented in a 4 x 4 error matrix listing the classification accuracies for the four classes above.

Table 1-4 contains the results of one eight variable discriminant analysis for soybeans. Each class is a development stage interval. Four of the variables are the reflectance observation at the beginning development stage of the interval, and the other four variables are the reflectance observation at the development stage ending the interval. Because there are 324 (18 x 18 stages) class intervals, the error matrix listing the classification results of the analysis is quite large, 324 by 324, or a total of 104976 terms. The matrix is, of course, much too large to be printed and easily understood. Thus, Table 1-4 presents only the entries from the main diagonal, a small fraction of the error matrix. These results provide an indication of how accurately classes (development stage intervals) were classified, but do not reveal the confusion classes (the off-diagonal elements of the error matrix) for those classes listed as being poorly classified.

The 83 entries in Table 1-4 are the percentages of correctly classified data vectors for each of 83 development stage intervals included as classes in the analysis. For example, 100% of the eight variable data vectors for the development stage interval 0.5 to 0.75 were correctly classified as belonging to the interval 0.5 to 0.75 and 92.0% of the data vectors for development stage interval 0.25 to 1.0 were correctly classified. For each development stage interval, the percentage of correctly classified data vectors is consistently higher for the eight variable analysis than for the other reclassification methods tested. Similar analyses for corn appear in Table 1-5.

Recommendations

The multispectral multitemporal concept of the eight variable analysis seems to have the most merit, of those investigated, for development stage estimation, especially when used in conjunction with existing crop models. The obvious next step is to extend the research, here limited to classification of the training data set to classification independent data. Further analyses are needed to determine how many data points per class are necessary to achieve repeatable, consistent classification performance and to determine the trade-off between overall classification performance and individual class accuracy. Another unanswered question is how to arrange development stages into classes to best fit the problem at hand.

Table 1-4. Classification matrix of eight variable corn data vectors.

FINAL DEVELOPMENT STAGE OF THE DEVELOPMENT STAGE INTERVAL	INITIAL DEVELOPMENT STAGE OF THE DEVELOPMENT STAGE INTERVAL																						
	0.0	0.25	0.5	0.75	1.0	1.25	1.5	1.75	2.0	2.25	2.5	2.75	3.0	3.25	3.5	3.75	4.0	4.25	4.5	5.0	6.0	7.0	8.0	9.0	10.0
0.0																									
0.25																									
0.5																									
0.75																									
1.0																									
1.25		37.8		54.9	80.5																				
1.5		88.2		30.2		38.1																			
1.75				6.7		65.6		51.3																	
2.0				19.5			25.0	39.5																	
2.25				66.7																					
2.5				16.2				54.8	85.3																
2.75				78.4																					
3.0						48.7	54.0																		
3.25																									
3.5						62.8	21.8	59.0	75.6		87.9														
3.75																									
4.0																									
4.25																									
4.5								79.2			75.0				88.2										
5.0				38.9																					
6.0																									
7.0		72.9		34.7	65.7	15.8	22.0	45.9	57.5				77.1		65.1							84.2			
8.0				45.0			25.0	44.7	75.0													54.3			
9.0		60.6	84.8	43.4	71.8	35.7	32.4	45.7	61.7				71.4									38.6	51.3	60.3	
10.0				31.4		61.5	42.6	35.9	71.2						63.0				52.9	53.2		50.1	63.3	70.7	90.9

Table 1-5. Classification matrix of eight variable soybean data vectors.

FINAL DEVELOPMENT STAGE OF THE DEVELOPMENT STAGE INTERVAL	INITIAL DEVELOPMENT STAGE OF THE DEVELOPMENT STAGE INTERVAL																	
	0.0	0.25	0.5	0.75	1.0	1.5	2.0	2.5	3.0	3.5	4.0	5.0	6.0	7.0	8.0	9.0	10.0	11.0
0.0																		
0.25	100.0																	
0.5																		
0.75	100.0	54.5	100.0	81.0														
1.0		92.0	76.0															
1.5				95.0														
2.0				52.8	88.9													
2.5				38.3			85.7	62.5										
3.0			92.0	62.9			23.3		61.9									
3.5							68.2	68.2	70.0									
4.0							70.0	44.2	48.1		69.4							
5.0	100.0	39.4	89.3	70.0	47.8		51.3	34.4	29.5	65.5	60.7							
6.0			40.0	87.9						76.7	56.5		80.6					
7.0															88.5			
8.0	96.8	57.1	78.3	35.1	57.6		47.5	81.0	28.3	60.0	34.7	53.5	85.2					
9.0	100.0	56.9	17.9	22.9	25.8	58.3	11.1	13.9	14.4	21.6	27.1	37.7	21.1	34.8	37.5	60.3		
10.0																70.7		
11.0			95.6	58.9	95.2		53.5	48.1	35.1	73.5	44.7	60.3	80.8	75.0	85.1	22.8		100.0

The three spectral variables, brightness, greenness and yellowness, probably contain all of the information needed for crop development stage estimation that is contained in the four spectral bands used for these analyses. This strongly suggests that six variable analyses similar to the eight variable analyses should be performed to reduce computing costs and the amount of data storage. Three of the variables in the six variable analyses would be the brightness, greenness, and yellowness values over a field plot at the first development stage of a development stage interval. The other three variables would be the brightness, greenness and yellowness value at the second development stage of the interval. Finally, as this research used only parametric discriminant analysis, nonparametric clustering algorithms should be tested to determine which produces more accurate classification results.

References

1. Kauth, R.J. and G.S. Thomas. 1976. The tasseled cap--graphic description of spectral-temporal development of agricultural crops as seen by Landsat. Proc. Symp. Machine Proc. Remotely Sensed Data, Purdue University, W. Lafayette, IN, pp. 4B/41-51.

B. The Use of Spectral-Temporal Trajectories for Development Stage Estimation

J.K. Ranson

In this section, we introduce a method that describes the movement of crops through spectral space with time as a set of velocity vectors. The underlying principle here is based on the "tasseled cap" concept presented by Kauth and Thomas (1976). That is, a crop starts out in feature space somewhere on a plane of soils and moves toward a point of green stuff. From this point it moves into a dimension characterized by the yellowing of the leaves as the crop scenesces. Finally, it returns to the plane of soils after harvest. The spectral variables used in this study are the Kauth-Thomas transformed features greenness, brightness and yellowness. Examples of temporal trajectories in greenness-brightness space for corn and soybean are shown in Figure 1-4.

We can describe the rate of movement or velocity in spectral space given two measurement vectors:

$$X_1 = (B_1, G_1, Y_1) \text{ and } X_2 = (B_2, G_2, Y_2);$$

where B, G, and Y are the transformed Landsat MSS features brightness, greenness and yellowness, respectively, and the subscripts refer to the time of observation. In this analysis, time can be represented by any temporal variable such as day of year, days from planting, growing degree days or development stage.

The velocity between two measurement vectors takes the form of a vector

$$\vec{X} = \left(\frac{\Delta B}{\Delta t}, \frac{\Delta G}{\Delta t}, \frac{\Delta Y}{\Delta t}\right).$$

The elements of the vector represent the change in spectral response for a given change in time between two measurement vectors.

We can assign 3-space position to the velocity vector as the vector

$$\vec{X} = (b, g, y)$$

where b, g, y are the spectral responses at the midpoints of the brightness, greenness, and yellowness trajectories. As an example, the brightness position element of $\vec{X}$ is

$$b = B_2 + (B_2 - B_1)/2.$$

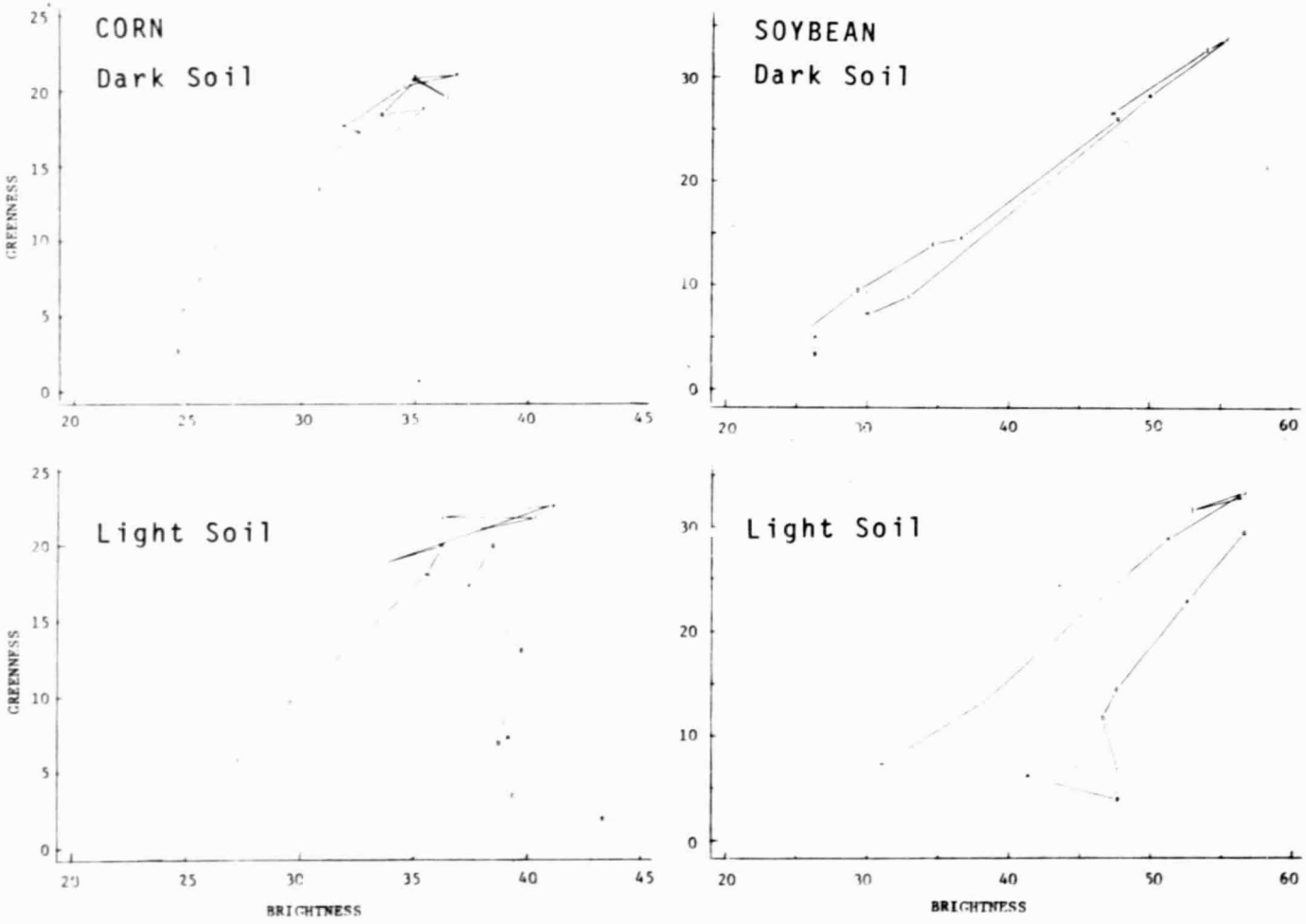


Figure 1-4. Example temporal trajectories plotted as function of development stage in greenness-brightness space for corn and soybean. Development stages are coded as in Figures 1-2 and 1-3.

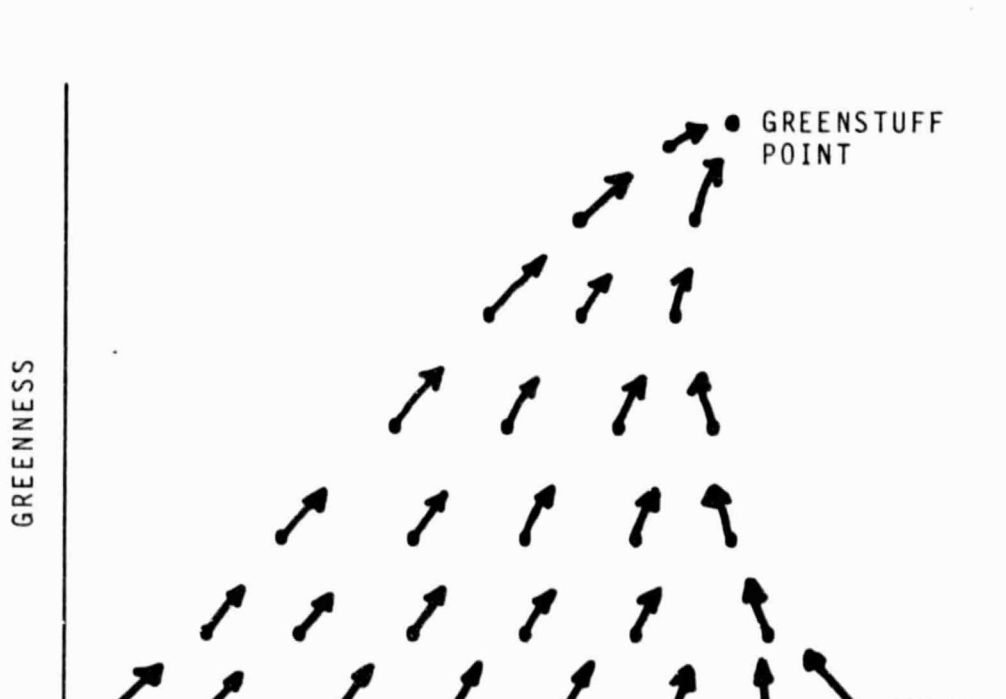


Figure 1-5. Idealized vector field of crop development stages from emergence to maximum leaf area accumulation.

The elements of X can then be modeled as a function of 3-space position:

$$\frac{\Delta\hat{B}}{\Delta t} = f(\vec{x}'), \quad \frac{\Delta\hat{G}}{\Delta t} = f(\vec{x}'), \quad \frac{\Delta\hat{Y}}{\Delta t} = f(\vec{x}').$$

This results in the expected velocity vector

$$\hat{\vec{V}} = \left(\frac{\Delta B}{\Delta t}, \frac{\Delta G}{\Delta t}, \frac{\Delta Y}{\Delta t}\right).$$

Within the limits of the transformed feature space (the tasseled cap), a vector field can be generated showing the expected trajectories at given points. An idealized representation is shown in Figure 1-5 in greenness-brightness space. The figure depicts the expected trajectories from emergence to maximum leaf area accumulation. After this point the trajectories would turn back on themselves when viewed in greenness-brightness space. In actuality, the trajectories move out of this plane into a third dimension corresponding to yellowness. Furthermore, integrating V with respect to time gives the expected location in spectral space of a given trajectory.

The above method describes the spectral-temporal trajectories of a crop within the constraints imposed by the modeling of the velocity vector. Variability in the scene due to weather or pest induced damage will result in trajectories that deviate from those expected. This is a result of changes in the spectral components caused by changes in canopy cover and canopy geometry. Soil surface moisture status also greatly alters the spectral-temporal trajectories at early development stages. A wetted soil tends to become darker so brightness is shifted downward without a concurrent change in greenness. As the soil surface dries the trajectory then returns to its normal position. The overall effect of alternately wetted and dried soils degrades efforts to model the expected trajectories, thus it may be necessary to model wet and dry soils separately. However, dry light and dark soils are adequately handled by this technique.

Our initial attempts at modeling the velocity vector were conducted with the 1979, 1980 Purdue Agronomy Farm data for corn described above. Wet soil dates were removed and an additional constraint of using only data points that were acquired within one development stage unit was imposed. This resulted in only 86 points which were insufficient for statistical analysis. In order to effectively model the trajectories an expanded data set is required. The 1981 Purdue Agronomy Farm data has recently become available and will be added to the data set for further analysis.

References

1. Kauth, R.J. and G.S. Thomas. 1976. The tasseled cap--graphic description of spectral-temporal development of agricultural crops as seen by Landsat. Proc. Symp. Machine Proc. Remotely Sensed Data, Purdue University, W. Lafayette, IN, pp. 4B/41-51.

C. Estimating Silking and Maturity Dates of Corn with Thermal Models

C.S.T. Daughtry and J.C. Cochran

Crop development involves complex physiological and biochemical processes which are influenced by the crop's environment and are still inadequately understood. Temperature and day length are the principal meteorological variables influencing crop development. Available moisture and nutrients may affect crop development in some situations.

During the past century numerous models to describe crop development as a function of thermal units have been proposed. The thermal unit concept assumes that photoperiod does not influence the rate of crop development. While thermal units are generally recognized to be superior to calendar days in predicting flowering or maturity dates, there is less than universal agreement as to which method of computing thermal units is best. Thus several methods to predict development stages should be tested and the "best" one for a particular application selected.

The objective of this research was to evaluate several methods for predicting the dates of silking and maturity stages of corn, using data acquired from an agricultural experiment station and several crop reporting districts in Indiana and Iowa.

Experimental Approach

Agronomy Farm

Agronomic and meteorological data used in the first phase of these analyses were acquired at the Purdue University Agronomy Farm in 1979, 1980, and 1981. A full season corn (Zea mays L.), Becks 65X, was planted on a Chalmers silt loam (Typic Argiaquoll) at three plant populations (25,000, 50,000, and 75,000 plant/ha) in 76-cm rows on 2, 16, 30 May 1979; 7, 16, 22, 29 May; and 11 June 1980 and 8 May, 29 May, and 11 June 1981. Development stages (Hanway, 1963) were observed once a week in 1979 and twice weekly in 1980 and 1981.

Daily meteorological data were recorded at the cooperative National Weather Service station (West Lafayette 6NW) which was within 300 m of the fields. Daily maximum and minimum air temperatures were measured in a standard shelter.

Crop Reporting Districts

The weekly average crop development stages for each crop reporting district (CRD) in Indiana and Iowa were extracted from annual summaries published by USDA (1969 to 1980). Dates on which 25, 50, and 75% of the corn crop in each CRD reached specified stages of development were interpolated from the published data.

Meteorological data consisting of daily maximum and minimum air temperatures for 1969 to 1980 were acquired for all National Weather Service cooperative stations in Indiana and Iowa. Mean daily maximum and minimum temperatures for each CRD were calculated using all stations within a CRD that had a complete weather series. The twelve years of development stage and meteorological data were divided into calibration (1969, 1971, 1973, 1975, and 1977) and test (1970, 1972, 1974, 1976, 1978, 1979, and 1980) sets.

Models and Analyses

Four thermal indexing methods and the number of calendar days after planting (DAP) were evaluated for precision and accuracy. The first index, Growing Degree Unit (GDU), accumulates thermal units for daily mean temperatures above 10C. Modified Growing Degree Unit (MGDU) index is similar to GDU but with a threshold of 30C imposed on maximum temperature. Heat Stress (HS) index (Cross and Zuber, 1972) is also similar to GDU but with a decrease in thermal unit accumulations for maximum temperatures greater than 30C. Temperature Function (FT) index (Coelho and Dale, 1980) is the mean relative growth rates for the daily maximum and minimum air temperatures.

The dates that 25, 50, and 75% of the corn acreage had been planted in CRD of Indiana and Iowa were used to start the accumulations of the thermal indexes. Dates that 25, 50, and 75% of the corn acreage in each CRD had silked or matured were the ending dates.

Results and Discussion

The means, standard deviations, and coefficients of variation (CV) for the five models evaluated at the Agronomy Farm and in Indiana and Iowa are shown in Table 1-6. When plots of corn were observed at the Agronomy Farm, the FT model had the smallest CV and DAP had the largest CV for both silking and physiological maturity. The thermal models depicted silking better than DAP over the wide range of planting dates used in the three years at the Agronomy Farm.

When models to predict development stage are compared over large areas, several factors not present in controlled experiments affect the variability. For example, length of growing season, soil type, and corn hybrids planted differ from location to location. The extraneous sources of variation should be accounted for as much as possible before the models are compared. The variations in CV and standard deviations presented in Table 1-6 were calculated using the error mean square from ANOVA with the main which removed the main effect of CRD.

Comparing models solely on the basis of the CV of cumulative thermal units for a number of environments provides an incomplete evaluation. Since CV is the ratio of the standard deviation to the mean, the effectiveness of this evaluation depends upon the range of means produced by the various

Table 1-6. Means, standard deviations (s), and coefficients of variation (CV) of thermal and DAP models for planting to silking and planting to maturity for the calibration years.

Location (n)	Statistic*	Model				
		GDU	MGDU	HS	FT	DAP
		Planting to Silking				
Agronomy† Farm (8)	Mean	831	818	796	37.8	69.7
	s	41.8	49.3	58.8	1.6	6.2
	CV, %	5.0	5.7	7.4	4.3	8.9
Indiana‡ (135)	Mean	766	759	725	36.8	66.3
	s	57.6	50.8	53.6	2.5	4.0
	CV, %	7.5	6.8	7.4	6.8	6.0
Iowa§ (108)	Mean	781	773	734	37.6	70.1
	s	58.5	48.7	49.9	2.2	3.0
	CV, %	7.5	6.3	6.8	5.8	4.3
		Planting to Maturity				
Agronomy† Farm (7)	Mean	1530	1520	1483	71.2	131.6
	s	52.2	36.9	27.4	1.9	4.8
	CV, %	3.4	2.4	1.8	2.6	3.7
Indiana‡ (135)	Mean	1436	1422	1369	69.4	122.9
	s	67.1	59.2	60.2	2.8	5.7
	CV, %	4.7	4.2	4.4	4.0	4.7
Iowa§ (108)	Mean	1376	1355	1293	65.9	120.8
	s	78.5	66.3	67.4	3.1	5.2
	CV, %	5.7	4.9	5.2	4.8	4.3

* The main effects of CRD have been removed from s and CV for Indiana and Iowa.

† Data are means of 3 planting dates in 1979 and 5 planting dates in 1980. Corn planted on 11 June 1980 did not mature before frost.

‡ Data are means of 9 CRD, 3 planting dates, and 5 years.

§ Data are means of 9 CRD, 3 planting dates, and 4 years.

models. Assuming the variance remains relatively constant, models producing the largest means will have the lowest CV.

Another way to approach this comparison of models is to use the mean cumulative units from Table 1-6 for the respective models to predict the dates of silking and maturity. Mean errors and mean absolute errors in number of days for the predicted date minus the actual date of each development stage provide more realistic evaluations of each model's ability than simply CV. Mean error ($\bar{e}$) is a measure of the bias of a model's predictions while mean absolute error ($|\bar{e}|$) measures its accuracy. The standard deviation of the absolute error ($s_{|\bar{e}|}$) provides a measure of the precision or variability of a model's errors in predicting dates of corn silking or maturity. Low variability signifies high precision.

When the models were used to estimate silking dates for the calibration years (Table 1-7) the thermal models were more precise and accurate than DAP at the Agronomy Farm. There were no major differences in accuracy or bias among the models for Indiana and Iowa. Rounding and truncation errors probably were responsible for the slight negative bias exhibited by all models. In Indiana silking dates were predicted with equal accuracy ($|\bar{e}|$) by all models. In Iowa the thermal models were significantly more accurate than DAP for predicting silking dates with FT being the most accurate and most precise model (1.5 $\pm$ 1.4 days). DAP models were more accurate than thermal models for estimating maturity dates at the Agronomy Farm and in Indiana and Iowa.

Evaluating a model on the same data used to develop the model tests only the goodness of fit of the model to the original data and does not evaluate the predictive ability of the model. For this more rigorous test of these models we divided our data into two series. The mean thermal units accumulated during the calibration years for each CRD were used to predict the dates of silking and maturity in the test years.

The thermal models were significantly less biased, more accurate, and more precise than the DAP model for predicting silking (Table 1-8). Predicting silking date simply as the number of days after planting presented absolute errors averaging 5.9 + 4.3 days for Indiana and 4.7 + 3.3 days for Iowa while all of the thermal models were more accurate and had absolute errors averaging 2.5 + 2.0 and 3.5 + 2.8 days for Indiana and Iowa, respectively.

The FT model showed significantly less bias ($\bar{e}$) than any other model for three of the four test cases presented in Table 1-8. The exception was for predicting silking date in Indiana when the FT model's bias was slightly greater than the other thermal models yet less than the DAP model. The absolute errors were less and thus the accuracies were greater for predicting corn maturity by the FT model than by the other thermal models. Thus the FT model appears to be superior to the DAP model and at least equal to or better than the other thermal models evaluated.

Table 1-7. Mean errors ($\bar{e}$), mean absolute errors ($|\bar{e}|$), and standard deviations of absolute errors ($s_{|\bar{e}|}$) in days for predicted minus actual date for estimating silking and maturity dates for calibration years. Data are means of 9 CRD, 3 planting dates per year, and either 5 years for Indiana or 4 years for Iowa.

Location (n)	Statistic	Model				
		GDU	MGDU	HS	FT	DAP
		Days				
		Planting to Silking				
Agronomy Farm (11)	$\bar{e}$	-0.7	-0.6	-0.6	-0.1	-0.4
	$\lvert\bar{e}\rvert$	1.0	0.9	1.1	1.3	5.6
	$s_{\lvert\bar{e}\rvert}$	0.8	0.7	1.1	1.1	3.5
Indiana (135)	$\bar{e}$	-0.1	-0.2	-0.1	-0.2	-0.3
	$\lvert\bar{e}\rvert$	2.7	2.7	2.6	2.7	2.8
	$s_{\lvert\bar{e}\rvert}$	1.9	1.8	2.0	2.0	2.4
Iowa (108)	$\bar{e}$	0.0	-0.1	-0.1	-0.1	-0.1
	$\lvert\bar{e}\rvert$	1.9	1.7	1.8	1.5	2.3
	$s_{\lvert\bar{e}\rvert}$	1.6	1.4	1.5	1.4	1.5
		Planting to Maturity				
Agronomy Farm (11)	$\bar{e}$	3.0	1.3	1.1	-0.1	-1.9
	$\lvert\bar{e}\rvert$	11.5	8.3	7.6	7.4	3.8
	$s_{\lvert\bar{e}\rvert}$	10.2	8.6	8.5	8.8	3.0
Indiana (135)	$\bar{e}$	-0.1	-0.1	-0.1	-0.2	-0.2
	$\lvert\bar{e}\rvert$	6.3	5.2	5.6	4.6	3.9
	$s_{\lvert\bar{e}\rvert}$	5.4	4.2	4.5	3.9	3.5
Iowa (108)	$\bar{e}$	-0.2	0.0	0.2	0.0	-0.1
	$\lvert\bar{e}\rvert$	6.4	4.5	5.7	4.6	3.4
	$s_{\lvert\bar{e}\rvert}$	4.7	3.8	5.0	4.3	2.4

Table 1-8. Mean errors ($\bar{e}$), mean absolute errors ($|\bar{e}|$), and standard deviations of absolute errors ($s_{|\bar{e}|}$) in days for predicted minus actual dates of silking and maturity for test years. Data are means of 9 CRD, 3 planting dates per year, and 7 years for both Indiana and Iowa.

Location (n)	Statistic†	Model				
		GDU	MGDU	HS	FT	DAP
		Days				
		Planting to Silking				
Indiana	$\bar{e}$	0.5a+	0 0b	0.0b	-0.8c	-4.1d
	$\|\bar{e}\|$	2.8b	2.4c	2.3c	2.4c	5.9a
	$s_{\|\bar{e}\|}$	2.3	2.1	2.0	2.0	4.3
Iowa	$\bar{e}$	0.7a	0.3b	0.7a	0.0c	-2.5d
	$\|\bar{e}\|$	3.6b	3.5b	3.3b	3.4b	4.7a
	$s_{\|\bar{e}\|}$	3.1	2.8	2.7	2.5	3.3
		Planting to Maturity				
Indiana	$\bar{e}$	7.4a	4.0c	5.4b	2.6d	-3.8e
	$\|\bar{e}\|$	11.0a	7.3c	8.2b	5.9d	6.7c
	$s_{\|\bar{e}\|}$	8.6	5.8	7.4	5.2	4.5
Iowa	$\bar{e}$	4.6a	1.6c	3.7b	1.2c	-3.7d
	$\|\bar{e}\|$	11.0a	8.0c	9.8b	7.8c	6.4d
	$s_{\|\bar{e}\|}$	9.4	6.9	9.1	6.7	4.7

† Within each statistic, means followed by the same letter are not significantly different at the alpha = 0.05 level using Duncan's multiple range test.

In summary, this experiment evaluated several development stage models for corn using data acquired at Purdue Agronomy Farm and data acquired by USDA for crop reporting districts (CRD) in Indiana and Iowa. Four thermal index and one days after planting (DAP) models were calibrated for each CRD using four or five years of data and were tested using seven different years of data. The thermal models were significantly more accurate than the DAP model for predicting silking and maturity dates. In most cases the FT model was equal to or better than the other thermal models.

References

1. Coelho, D.T. and R.F. Dale. 1980. An energy-crop growth variable and temperature function for predicting corn growth and development: planting to silking. Agron. J. 72: 503-510.

2. Cross, H.Z. and M.S. Zuber. 1972. Prediction of flowering dates in maize based on different methods of estimating thermal units. Agron. J. 64: 351-355.

3. Hanway, J.J. 1963. Growth stages of corn (Zea mays, L.). Agron. J. 55: 487-492.

4. USDA Statistical Reporting Services. 1969-1980. Indiana Crop and Livestock Statistics Annual Summary. Agric. Exp. Stn., Purdue University, West Lafayette, IN.

5. USDA Statistical Reporting Service. 1970-1980. Annual Crop Weather Summary. Iowa Dept. Agric., Des Moines, IA.

6. USDC Environmental Data Service. 1969-80. Climatological data, Indiana and Iowa. NOAA Natl. Climatic Center, Asheville, NC.

2. ESTIMATION OF AGRONOMIC VARIABLES ASSOCIATED WITH YIELDS OF CORN AND SOYBEANS

C.S.T. Daughtry, K.P. Gallo and M.E. Bauer

Introduction

In recent years the world's food situation has emphasized the need for timely information on world-wide crop production, but relatively few countries have reliable methods for gathering and reporting crop production information. Remote sensing from aerospace platforms can provide information about crops and soils which could be useful to crop production forecasting systems. The feasibility of utilizing multispectral data from satellites to identify and measure crop area has been demonstrated (MacDonald and Hall, 1980); however, relatively little research has been conducted on developing the capability of using multispectral data to provide information about crop condition and yield. If this spectrally derived information can be combined effectively with crop models which depict limitations imposed on crop yields by weather, then potentially much better information about crop yield and production can be gained.

Solar radiation is the basic source of energy for photosynthesis, the initial process that green plants use to convert carbon dioxide and water into simple carbon sugars. Other plant processes convert these initial products of photosynthesis into dry matter including carbohydrates, proteins, and oils. Solar radiation as an energy source for plants is available only when it interacts with leaves. Assuming other factors are not limiting, production of dry matter is proportional to the solar radiation intercepted by the canopy. Thus, important components of growth and yield are the amount and duration of plant surface available for photosynthesis.

In theory, the production of dry matter (DM) over time period t, beginning at emergence and ending at maturity, can be related to the proportion (P) of the incident light (SR) intercepted by the crop using the following equation from Steven (1982):

$$DM = \int_{t_1}^{t_2} E\ P\ SR\ dt. \qquad (1)$$

E is the efficiency of conversion of solar energy into dry matter and typically ranges from 1 to 3 g/M Joule (Monteith, 1976). This equation can be used to predict dry matter production if P is known. Current methods to measure the radiation intercepted by crops are laborous and limit use of such crop models to small research plots. If the proportion of energy available for crop growth could be estimated reliably using multispectral satellite data, then models to estimate crop production for large regions could be implemented.

An overview of this concept showing how multispectral data could be used to estimate crop yields is illustrated in Figure 2-1. Percent soil cover and leaf area index strongly influence the reflectance of radiation from crop canopies (Kollenkark et al., 1981; Walburg et al., 1981). Estimates of these canopy variables, obtained from Landsat MSS data along with solar radiation data from ground stations or meteorological satellites, will enable the amount of radiation intercepted by crop canopies to be calculated. This intercepted solar radiation variable, integrated over the growing season accounts for much of the variation in corn and soybean yields. Thus, multispectral data from satellites form the basis for estimating crop growth and yields over large areas where ground observations may be difficult or impossible to obtain.

Objectives

The overall objective of this study is to develop approaches (models) for using spectral data as a source of information for crop yield models. The specific objectives are:

1. Identify important factors determining crop yield that can be estimated from spectral data,

2. Evaluate those selected factors using spectral and agronomic data acquired in controlled experiments at agricultural experiment stations,

3. Develop models relating spectral reflectance to important biophysical characteristics of corn and soybean canopies (e.g., leaf area index, percent soil cover, and solar radiation interception),

4. Develop methods for combining spectral and meteorological data in crop yield models,

5. Extend those methods that best estimate yields at agricultural experiment stations to large areas using Landsat MSS data, and

6. Compare the results of estimating yield with and without the use of spectral data.

Materials and Methods

Experimental Data

Spectral and agronomic data used in these analyses were acquired by the field research task at the Purdue Agronomy Farm in 1979 and 1980. Bauer et al. (1979, 1980) describe these experiments in detail. Briefly, a full season corn hybrid, Becks 65X, was grown on two soil types (dark- and light-colored) at three plant populations with three planting dates in 1979 and seven planting dates in 1980.

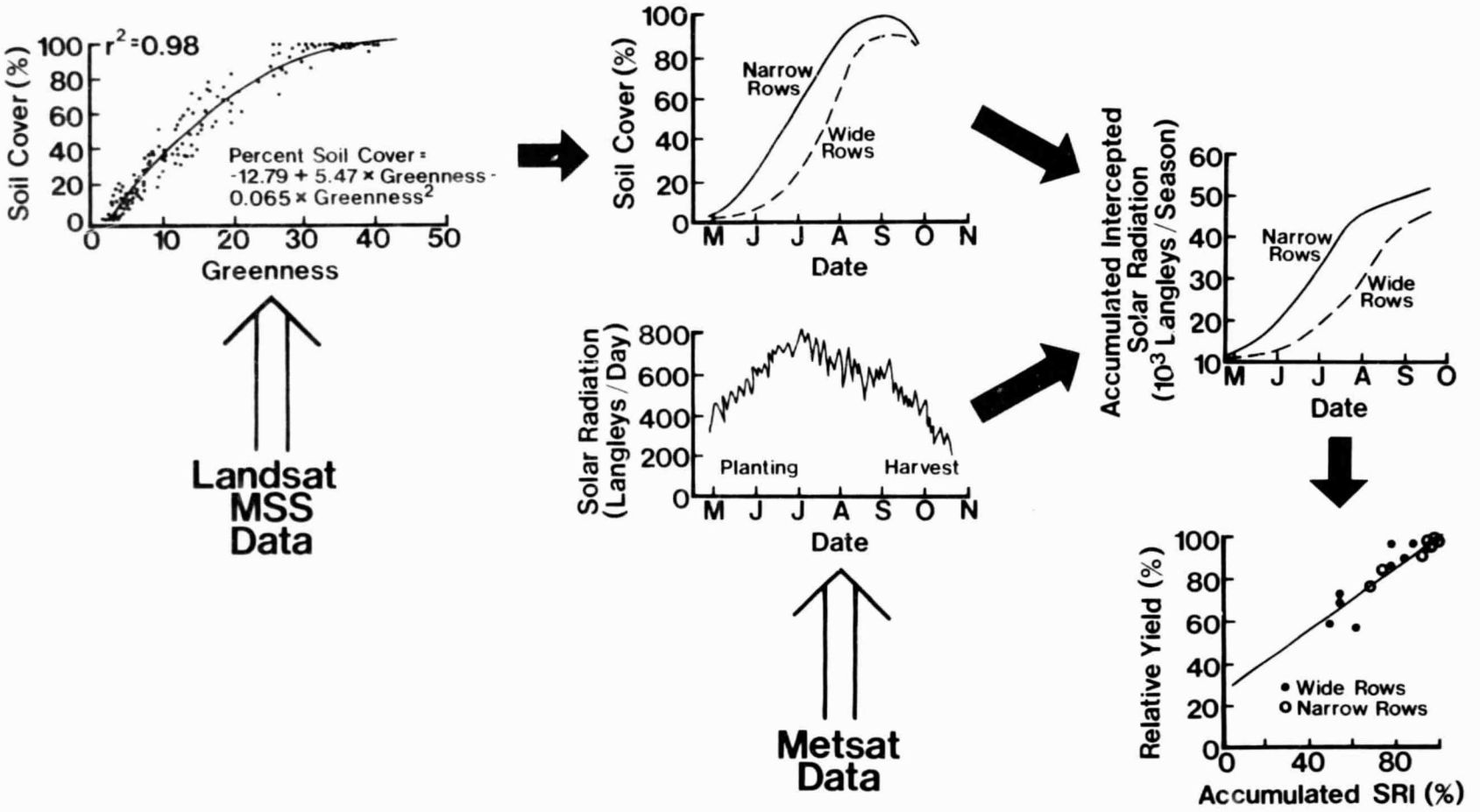


Figure 2-1. An overview of the concept of utilizing spectrally-derived estimates of solar radiation interception by crop canopies as input to crop growth and yield models.

Spectral data were acquired with a Landsat band radiometer (Exotech 100A) at approximately weekly intervals from planting to harvest. Agronomic data acquired to coincide with the spectral data included leaf area index (LAI), fresh and dry biomass, development stage percent soil cover, and grain yield.

Analysis

Two methods of estimating the proportion of solar radiation intercepted by corn canopies have been examined. First, the proportion of intercepted radiation (SRIm) may be described as a function of LAI using the following equation from Linvill et al., (1978):

$$SRIm = [1 - \exp(-0.79\ LAI)]. \qquad (2)$$

This is an application of Bouguer's Law using LAI and an extinction coefficient of 0.79. When LAI is 0, no energy is intercepted. When LAI is 2.8, approximately 90% of the visible solar radiation is intercepted by the canopy and is potentially useful to the crop.

The second method estimates SRIm as a function of the spectral variable greenness using the following equation:

$$SRIs = -0.1613 + 0.0811\ Greenness - 0.0015\ Greenness^2 \qquad (3)$$

This spectrally estimated proportion of solar radiation intercepted is called SRI to distinguish it from SRIm which is estimated using measured LAI.

SRIs predicted as a function of the spectral variable greenness (eq. 3) and SRIm predicted as a function of measured LAI (eq. 2) were calculated for each day that appropriate spectral and agronomic data were acquired and linearly interpolated for intermediate days throughout the growing season for each plot. Daily values of SRIs, SRIm, and greenness were accumulated from planting to physiological maturity. Correlations of final grain yields with these accumulated indices were examined.

The performances of spectrally estimated SRIs and measured SRIm were compared using the Energy-Crop-Growth (ECG) model (Dale, 1977) which combines the concept of intercepted solar radiation with a moisture stress term and a temperature function. The ECG model may be expressed as:

$$ECG = \sum_{i=t_1}^{t_2} (SR/600)_i (SRIm)_i (WF)_i (FT)_i \qquad (4)$$

where SR is the daily solar radiation intercepted by the canopy (Linvill et al., 1978), WF is the ratio of daily evapotranspiration to potential evapotranspiration (Stuff et al., 1978), and FT is a daily temperature function which relates growth rate to temperature (Coelho and Dale, 1980). SRIs predicted from spectral data was substituted directly for SRIm in the ECG model.

Results and Discussion

Relation of Canopy Reflectance to LAI and SRI

Leaf area index (LAI) and the proportion of solar radiation intercepted (SRIs) by these corn canopies were described as functions of several spectral variables and transformations using regression analyses. The transformations, near infrared/red ratio and greenness were highly correlated to LAI and percent soil cover but relatively insensitive to soil moisture and soil color (Kollenkark et al., 1981; Walburg et al., 1981). In each year, and for the combined data, SRI was predicted better (higher R2) than LAI, and greenness was the best spectral variable for predicting SRIs.

Figure 2-2 illustrates the relationships of LAI and SRIs to the spectral variable greenness for all plots of corn in 1979 and 1980. Undoubtedly planting date, plant population, and soil type contributed to the scatter about the regression line, but when included as terms in the regression models, they provided very little additional information. Errors in measuring LAI also account for a portion of the scatter. Nevertheless, LAI and/or SRIs predicted as a function of greenness will permit the concepts and models developed at the Purdue Agronomy Farm to be extended to large areas where only spectral data may be available.

Relation of Spectral Variables to Yield

Corn is a determinant crop; that is, corn completes its vegetative development, producing all of its leaves, then shifts to reproductive development, producing and filling its grain. If corn grain yields are strongly related to the total amount of leaf area present, then knowing the maximum LAI or maximum greenness of corn should be an important predictor of grain yields. Figure 2-3 illustrates the errors of such logic. Grain yields are plotted as a function of maximum values of greenness which occurred at tasseling. The poor relationship shown in Figure 2-3 emphasizes the limited value of a single observation of greenness for predicting corn yields.

A more important and useful indicator of grain yields is the duration of LAI and not simply the maximum LAI achieved. Variables which represent the integral of LAI over time during the growing season are accumulations of the daily values of SRIm and SRIs. These summed values of SRIm and SRIs represent the proportions of solar radiation impinging on corn field during the growing season that were intercepted by the canopy and were potentially available for photosynthesis (Figure 2-4). Both estimates of intercepted solar radiation were associated with approximately 65 percent of the variation in grain yields.

One problem in crop response to solar radiation is the confounding of solar radiation, temperature, and plant moisture stress on plant growth and yields. Thus applicability of the theoretical model presented in equation 1 is limited. Coelho and Dale (1980) combined solar radiation, temperature and moisture stress with an estimate of intercepted solar radiation (SRIm) in an Energy-Crop-Growth (ECG) model (eq. 4). This ECG model evaluates the daily effects of weather on corn growth and yield.

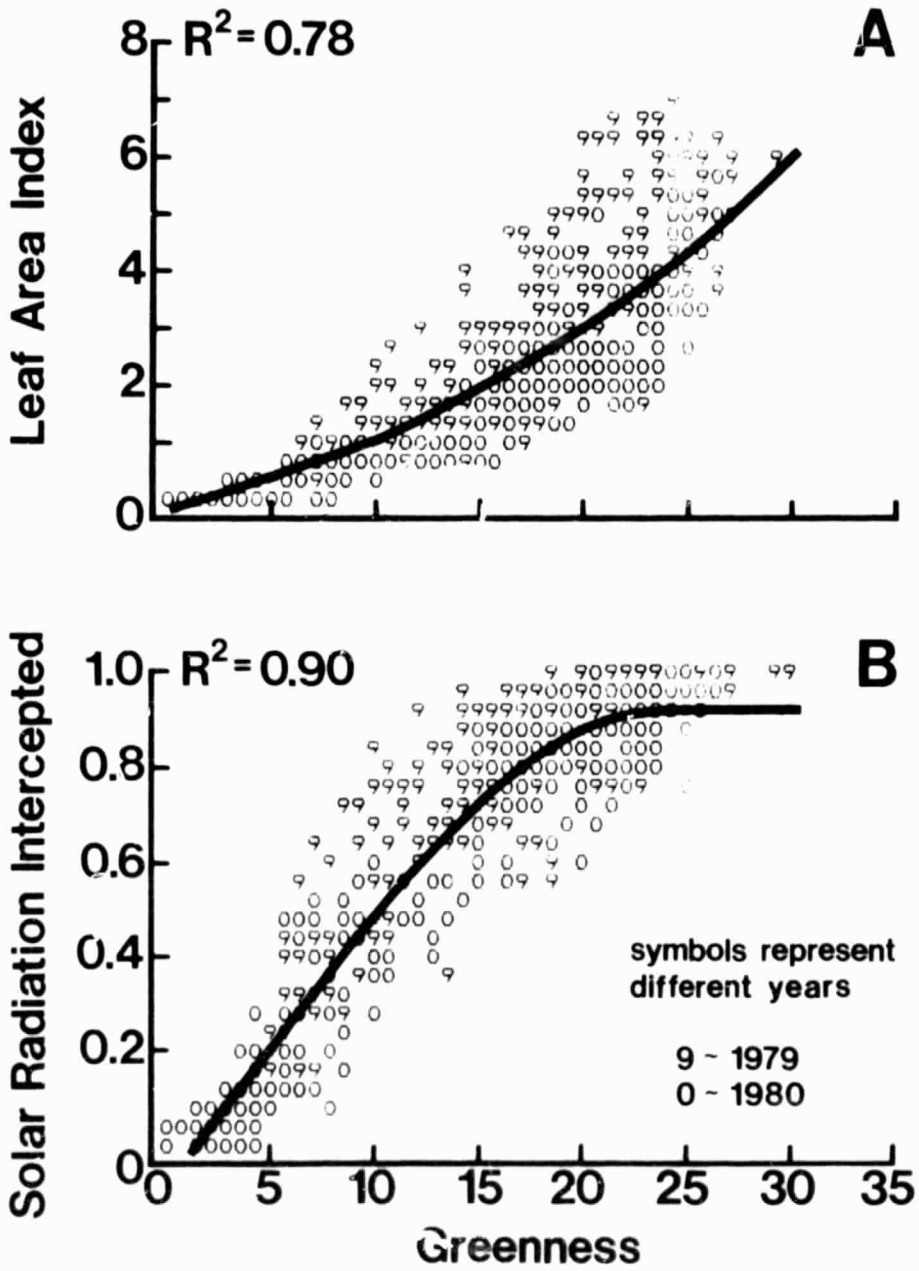


Figure 2-2. Relationships of leaf area index (LAI) solar radiation intercepted (SRIs) to the spectral variable greenness. LAI = -0.1364 + 0.0647 Greenness + 0.0047 Greenness^2. SRIs = -0.1613 + 0.0811 Greenness - 0.0015 Greenness^2.

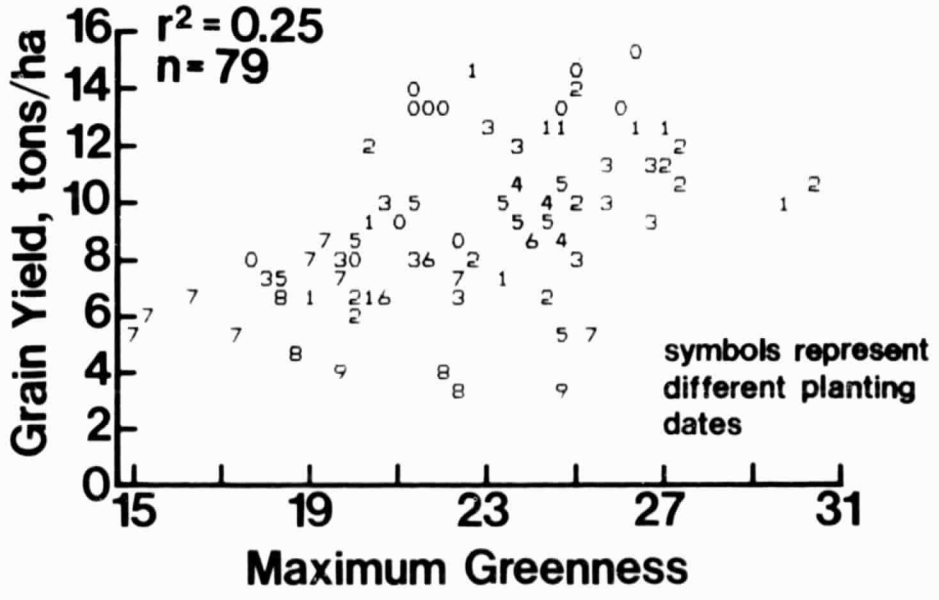


Figure 2-3. Corn grain yields as a function of maximum greenness.

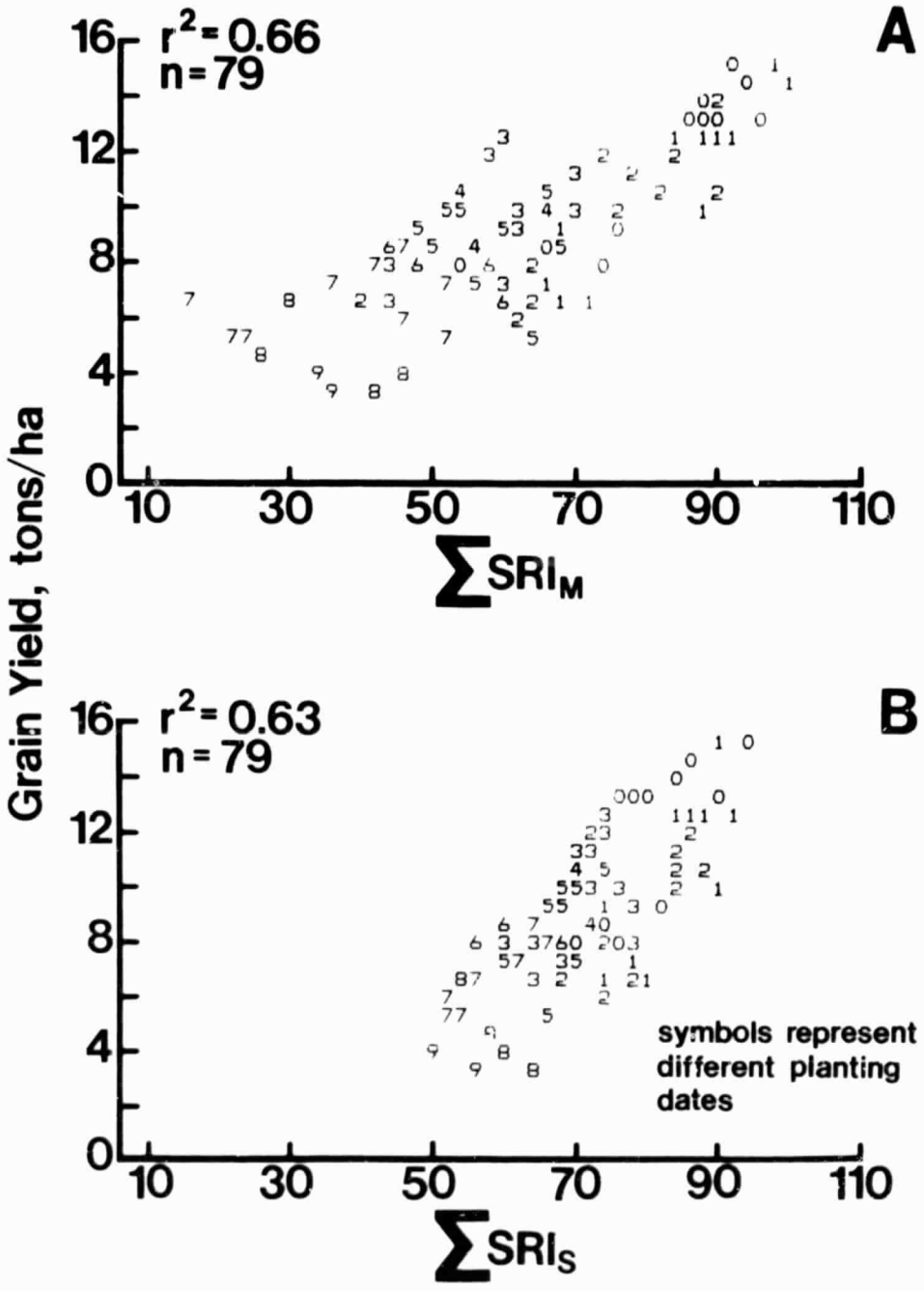


Figure 2-4. Corn grain yields as a function of the accumulated proportions of solar radiation intercepted by the corn canopies. SRIm is estimated using measured LAI and SRIs is estimated using spectral data. The symbols are relative planting dates for the two years with "0" representing the first planting date in 1979 and "9" representing the last planting date in 1980.

The sums of the daily values of ECG using SRIm or SRIs to estimate intercepted solar radiation are plotted with grain yields (Figure 2-5). The addition of meteorological data in the ECG models resulted in modest improvements in the correlation coefficients in both cases. Although the ten planting date-years represented a range of temperature and day length regimes, significant water stress was not evident among the treatments. Thus both of the intercepted radiation variables alone were nearly as highly correlated to yields as the meteorological (ECG) models (Table 2-1). In situations where moisture or temperature is limiting, the ECG models should be superior to the intercepted solar radiation models for predicting corn yields. This concept of combining spectrally derived estimates of SRIs with meteorological models permits implementation of crop growth and yield models in areas without ground measurements of LAI.

Conclusions

We conclude from this research that the concept of estimating SRIs using spectral data represents a viable approach for merging spectral and meteorological data in crop yield models and presumably in evapotranspiration models such as those of Kanemasu et al., (1976), and Ritchie (1972). This scheme is consistent with the spectral-physiological modeling approaches proposed by Wiegand et al. (1979) but avoids some of the problems associated with estimating LAI directly from spectral data. This

Table 2-1. Summary of yield models evaluated.

Model	Input Data	R^2 (N = 80)
1. Maximum Greenness	Spectral Only	0.23
2. Σ Greenness	Spectral Only	0.56
3. Σ Temperature Function	Meteorological Only	0.50
4. Σ SRI	Measured LAI	0.67
5. Σ SRIs	Spectral Data	0.64
6. Energy Crop Growth With FLAI	Met. + Measured Data	0.68
7. Energy Crop Growth With SRI	Met. + Spectral Data	0.70

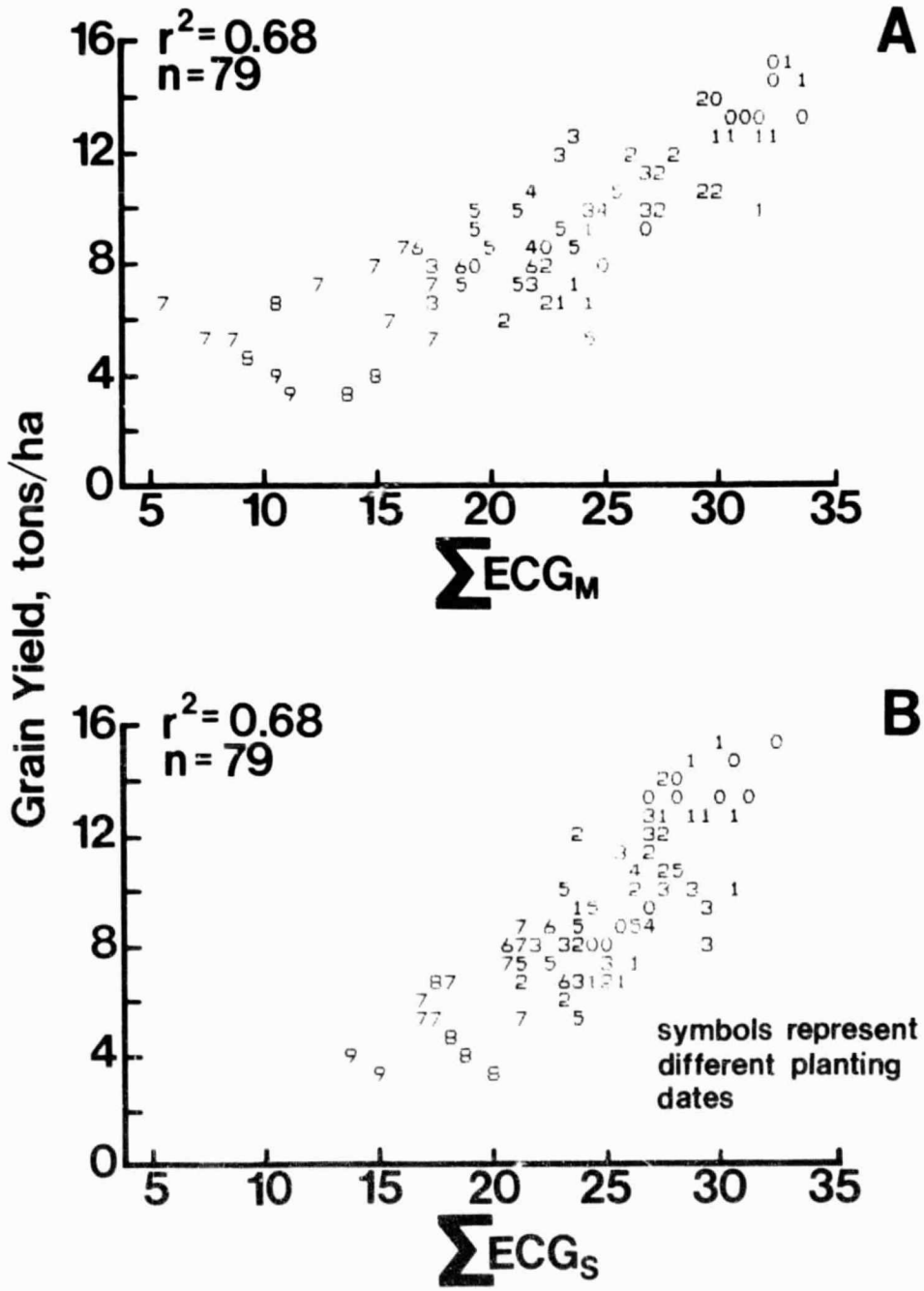


Figure 2-5. Corn grain yields as a function of the accumulated Energy-Crop Growth variable. ECGm is calculated using SRIm and ECGs with SRIs. The symbols are relative planting dates for the two years with "0" representing the first planting date in 1979 and "9" representing the last planting date in 1980.

concept may be extended to large areas using Landsat MSS data to estimate SRIs for as many fields as are of interest. Thus a spectral-meteorological system of crop forecasting could exploit the frequent temporal sampling characteristic of weather data (e.g. hourly or daily) with the high spatial resolution characteristic of spectral data from Landsat MSS and TM.

References

1. Bauer, M.E., L.L. Biehl, C.S.T. Daughtry, B.F. Robinson, and E.R. Stoner. 1979. Agricultural scene understanding and supporting field research. Volume 1. Tech. Report 112879. Lab. Applic. Remote Sensing, Purdue Univ., West Lafayette, IN.

2. Bauer, M.E., L.L. Biehl, and B.F. Robinson. 1980. Field research on the spectral properties of crops and soils. Tech. Report 112680, Lab. Applic. Remote Sensing, Purdue Univ., West Lafayette, IN.

3. Coelho, D.T. and R.F. Dale. 1980. An energy-crop growth variable and temperature function for predicting corn growth and development: planting to silking. Agron. J. 72:503-510.

4. Dale, R.F. 1977. An energy-crop growth variable for identifying weather effects upon maize yield. p. 240-248. In W. Baier, R., H. Shaw, L.M. Thompson, and R.E. Felch (ed.) Agrometeorology of the maize (corn) crop. Proc. Symp. Agrometerol. Maize (corn) crop, World Meteorol. Organ., Iowa State Univ., Ames, Iowa, USA. 5-9 July 1976. WMO No. 481. Geneva, Switzerland.

5. Kanemasu, E.T., L.R. Stone, and W.L. Powers. 1976. Evapotranspiration model tested for soybean and sorghum. Agron. J. 68:509-572.

6. Kollenkark, J.C., C.S.T. Daughtry, and M.E. Bauer. 1981. Influence of cultural practices on the reflectance characteristics of soybean canopies. Tech. Report 021781, Lab. Applic. Remote Sensing, Purdue Univ., West Lafayette, IN. (Also Agron. J., 1982, In Press)

7. Linvill, D.E., R.F. Dale, and H.F. Hodges. 1978. Solar radiation weighting for weather and corn growth models. Agron. J. 70:257-263.

8. MacDonald, R.B. and F.G. Hall. 1980. Global crop forecasting. Science 208:670-679.

9. Monteith, J.L. (ed.) 1976. Vegetation and the atmosphere. Vol. 2. Case studies. Academic Press, New York. p. 439.

10. Norman, J.M. 1980. Interfacing leaf and canopy light interception models. p. 49-67. In J.D. Hesketh and J.W. Jones (eds.) Predicting photosynthesis of ecosystem models. Vol. 2. CRC Press.

11. Ritchie, J.T. 1972. Model for predicting evaporation from a row crop with incomplete cover. Water Resour. Res. 8:1204-1213.

12. Steven, M.D. 1982. Optical remote sensing for predictions of crop yields. Proc. Soc. Photo-Optical Inst. Engr. (In Press).

13. Stuff, R.G. and R.F. Dale. 1978. A soil moisture budget model accounting for shallow water table influences. Soil Sci. Soc. Am. J. 42:637-643.

14. Walburg, G., M.E. Bauer and C.S.T. Daughtry. 1981. Effects of nitrogen nutrition on the growth, yield and reflectance characteristics of corn canopies. Tech. Report 033181, Lab. Applic. Remote Sensing, Purdue Univ., West Lafayette, IN. (Also Agron. J., 1982, In Press)

15. Wiegand, C.L., A.J. Richardson, and E.T. Kanemasu. 1979. Leaf area index estimates for wheat from Landsat and their implications for evapotranspiration and crop modeling. Agron. J. 71:336-342.

3. FIELD RESEARCH--EXPERIMENT DESIGN, DATA ACQUISITION AND PREPROCESSING

L.L. Biehl

Introduction

This section describes the results of work conducted under Task 5, Field Research - Experiment Design, Data Acquisition, Preprocessing, and Management. The objectives of this task were to acquire, preprocess, and manage the required data to support the crop identification, development stage, and condition assessment field research tasks for the corn-soybean and small grains research as described in the AgRISTARS Supporting Field Research Program Technical Plan. The emphasis of our work and this report is on the corn and soybean experiments.

Based on the previous, proven experience since 1974, there were two kinds of test sites for 1981 - controlled experimental plot sites and commercial field sites. The data from experiments in commercial field test sites provide a measure of the natural variation in the temporal-spectral characteristics of the cover type. The data from experiments in controlled plots enable more complete understanding and interpretation of the spectra collected from commercial fields. The test sites are summarized in Table 3-1.

Experiment Objectives

The experiments for 1981 include some that have been continued from previous years to sample different yearly weather patterns and several new experiments to obtain measurements over additional crops (sorghum and sunflower) and measurement/canopy variables.

The following overall objectives were selected for the experiments:

Commercial fields

Webster Co., Iowa

- To determine spectral characteristics and separability of corn, soybeans and other crops as a function of development stage and cultural practice.

- To verify spectral-agronomic relationships developed at agricultural experiment stations.

Table 3-1. Summary of 1981 field research test sites and their respective crops.

Location	Major Crop(s)
Commercial Fields	
Webster County, Iowa	Corn, soybean
Cass County, N. Dakota	Spring wheat, barley, sunflower, soybean
Agriculture Experiment Station	
W. Lafayette, Indiana	Corn, soybean, sorghum, sunflower
Manhattan, Kansas	Small grains
Sandhills, Nebraska	Corn, soybean
Brookings, S. Dakota	Small grains

Cass Co., North Dakota

- To determine spectral characteristics and separability of spring wheat, barley, sunflowers and soybeans as a function of development stage and cultural practice.

Agriculture Experiment Stations

Purdue University Agronomy Farm

- To determine relationship of crop canopy variables (development stage, LAI, biomass, soil background, etc.) to reflectance and radiant temperature of corn and soybeans.

- To determine effects of varying agronomic practices (planting date, row spacing, plant population, cultivar, soil type) on spectral response of corn, soybeans, sorghum, and sunflowers.

- To support the development of corn and soybean yield models which use spectral response as a function of crop development stage as an input.

- To determine the relationships of percent soil cover, row direction, solar illumination angle, and canopy geometry to reflectance of soybean, sorghum, and wheat canopies.

Sandhills Agricultural Laboratory

- To determine the effect of moisture stress on reflectance and radiant temperature characteristics of corn and soybeans.

Experiment Descriptions and Data Acquisition

Commercial Field Test Sites

Agronomic observations were collected every nine to eighteen days over at least 80 fields in each of the Webster County and Cass County test sites. The agronomic observations are summarized in Table 3-2. The field inventory and periodic observations were collected by USDA ESCS personnel. The detailed agronomic measurements were collected by North Dakota State University in Cass County. No detailed agronomic measurements were collected in Webster County, Iowa.

Spectral data were acquired by the helicopter-mounted FSS at two or three week intervals, coinciding with Landsat overpasses of the site throughout the growing season. Aircraft NS001 scanner data and scatterometer-radiometer data were acquired on four missions during the growing season at the Cass County test site. Side looking radar data were collected three times during the growing season. Tables 3-3, 3-4 and 3-5 summarize the spectral measurements made at the two test sites. The Barnes 12-1000 multiband radiometer was successfully flight tested on the helicopter simultaneously with the FSS, on September 17 and 18 in Webster County and September 21 in Cass County.

Procedures for operating the FSS system (and Barnes system) are similar to those for the truck-mounted systems operated at the agriculture experiment stations so that data from the different test sites can be readily compared. The FSS is operated at an altitude of 61 meters above the ground over the commercial fields. Measurements of the 6 x 12 meter canvas reflectance reference panel were made before each flight line from an altitude of six meters. The reflectance of the canvas reference panel is measured periodically at Purdue/LARS with the Exotech 20C field spectroradiometer system. The aircraft scanner system also collects data over five canvas panels, representing five different reflectance levels.

Augmenting the spectral and agronomic measurements are meteorological data. The meteorological measurements include air temperature, relative humidity, barometric pressure, wind speed and direction, rainfall observation and optical depth. A record of the irradiance was collected by a total incidence pyranometer on strip charts during the spectral data collection as a record for general sky (cloud) conditions.

Table 3-2. Summary of agronomic measurements collected at Webster County and Cass County.

Agronomic Data	
Initial Interview (of 80 fields in site)	
Crop species	Previous year field use
Acres planted	Irrigated
Planting date	Row direction
Emergence date	Row width
Seeding rate	Pesticides applied
Periodic Observations (every nine days of 80 fields in site)	
Canopy height	Disease Damage
Ground cover	Insect Damage
Canopy color	Hail damage
Development stage	Lodging damage
Surface moisture	Pesticide applied
Weediness	
Final Interview (of 80 fields in site)	
Acres harvested	Fertilizer applied
Harvest date	Second crop planted
Production	Percent moisture at harvest
Harvest method	
Detailed Measurements (of 10 fields for each major crop, Cass County, N. Dakota site only)	
Leaf area index	Percent soil cover
Plant population	Development stage

Table 3-3. Summary of spectral measurements collected at the Webster County and Cass County commercial field test sites.

Sensor System	Spectral Range
Helicopter-Mounted Radiometer	0.4 - 2.4 µm 8.0 -14.0 µm
Helicopter-Mounted Radiometer (Barnes 12-1000) (2 test flights only)	0.45- 0.52 µm 0.52- 0.60 µm 0.63- 0.69 µm 0.76- 0.90 µm 1.15- 1.30 µm 1.55- 1.75 µm 2.08- 2.35 µm 10.40-12.50 µm
NS001 Thematic Mapper Simulator (Cass County, N. Dakota only)	0.47- 0.52 µm 0.53- 0.60 µm 0.63- 0.69 µm 0.76- 0.90 µm 1.00- 1.31 µm 1.54- 1.70 µm 2.10- 2.30 µm 10.40-12.40 µm
Multifrequency Microwave Radiometer	C Band
Scatterometers	0.40 GHz 1.60 GHz 4.75 GHz 13.30 GHz
APQ-102 (Side-Looking Radar)	X Band

Table 3-4. Summary of 1981 data acquisition by the NASA/JSC helicopter-mounted field spectrometer (FSS) and multiband radiometer systems.

Measurement Date	Test Site: Webster Co. Iowa	Test Site: Cass Co. N. Dakota
week of	data acquisition date†	
April 12	---	*
19	4/24	---
26	---	5/1
May 3	---	---
10	5/14	---
17	---	5/18
24	5/30	---
31	---	6/3
June 7	6/10	---
14	*	---
21	---	6/24
28	---	---
July 5	7/10	---
12	---	7/16
19	*	---
26	---	*
Aug. 2	*	---
9	---	8/12
16	---	---
23	*	---
30	---	9/1
Sept. 6	---	---
13	9/17, 18	---
20	---	9/21
27	10/2	---
Oct. 4	---	10/6
11	---	---
18	---	---
25	10/26	---

* Mission planned, but no data collected because of inclement weather conditions.

† Data collected by both FSS and Barnes 12-1000.

Table 3-5. Summary of 1981 crop year data acquisition by the NASA/JSC aircraft systems (NS001 scanner, radiometer-scatterometers, side looking radar).

Measurement Date	Test Site: Webster Co. Iowa	Test Site: Cass Co. N. Dakota
week of	data acquisition date	
May 17	---	---
24	---	---
31	---	6/3 (1,2)
June 7	---	---
14	---	---
21	6/25 (3)	6/24 (1,2)
28	---	7/1 (3)
July 5	---	---
12	---	7/16 (1,2)
19	---	---
26	---	---
Aug 2	---	---
9	---	---
16	---	---
23	---	---
30	---	9/1 (1) 9/2 (3)

1. NS001 multispectral scanner data
2. Radiometer-scatterometer data
3. Side Looking radar data

Agriculture Experiment Stations

Purdue Agronomy Farm

Several experiments were developed at the Purdue Agronomy Farm to accomplish the objectives stated above. The experiments included treatments of cultural practices and sun angle/row direction. (Table 3-6). The turntable built during 1980 was used for the row direction experiment to allow efficient measurements over many row directions (3). Spectral and agronomic measurements (Table 3-7) were collected on every day that solar illumination conditions were favorable, i.e. no clouds over or in the vicinity of the sun.

The spectral measurements of the experiments were made by the Exotech 100 Landsat band and Barnes 12-1000 radiometer field systems. The system also includes a 35 mm camera, sighted to view the same area as the spectral sensors.

To obtain data which can be readily compared, the instrument systems are operated following defined, established procedures. The instruments are operated from aerial towers at seven to ten meters above the target to minimize any row effect and shadowing of skylight. Care is also taken to ensure that the field of view of the instrument includes only the desired subject. The routine data taking mode of the instrument is straight down for determination of bidirectional reflectance factor. Measurements of the painted barium sulfate reflectance reference panel are made at 15-20 minute intervals. Two measurements of each plot are typically made by moving the sensor so that a new scene within the plot fills the field of view.

Spectral measurements, along with agronomic and meteorological data, were acquired on each day that weather conditions permitted, Table 3-8. During 1981 over 8000 spectra of corn, soybeans, sorghum, and sunflowers were acquired on 22 days. Crop maturity stages from seedling to senescence are represented in these data.

Sandhills, Nebraska

Moisture stress experiments on corn and soybeans were conducted by the University of Nebraska at the Sandhills Agricultural Laboratory. Gradient irrigation (row 1 receives full irrigation, row 24 receives no water) treatments were employed. One set of plots received gradient irrigation during vegetative, pollenation and grain filling stages, while a second set received gradient during vegetative stages and full irrigation during pollenation and grain filling stages. Spectral measurements made with the Exotech 100 (and late in the season by the Barnes 12-1000) radiometer on twelve days between late June and mid September. Additionally, crop radiant temperatures were measured. Agronomic measurements included development stage, leaf area index, biomass, canopy structure, soil moisture, total plant water potential, plant photosynthesis, stomatal resistance, and grain yield.

Table 3-6. Summary of the 1981 field research experiments at the Purdue Agronomy Farm.

Experiments and Treatments
Corn Cultural Practices
4 Planting dates (May 8, 29, June 11, 29)
3 Populations (25, 50, 75 thousand plants/ha)
2 Soil types (Chalmers-dark, Fincastle-light)
2 Replications
Soybean Cultural Practices
5 Planting dates (May 29, June 8, 10, 29, July 13)
2 Row widths (25, 75 cm)
2 Soil types (Chalmers-dark, Fincastle-light)
2 Backgrounds (soil, wheat straw-June 10 planting date only)
2 Replications
Sorghum Cultural Practices
3 Planting dates (May 28, June 9, 30)
2 Hybrids (NK300-semi-dwarf, BR64-dwarf)
2 Replications
Sunflower Cultural Practices
3 Planting dates (May 28, June 9, 30)
2 Populations (37.5, 75 thousand plants/ha)
2 Replications
Soybean Cultivar
6 Cultivars (Hodgson 78, Amsoy 71, Williams, Cutler 71, Gnome, Elf)
2 Replications
Row Direction/Sun Angle
3 Crops (soybeans, sorghum, wheat)
72 Row directions (0,5,10,...355)
3 Backgrounds (3M Black, 3M White, Russell soil)
4-7 Solar azimuth-zenith angle ranges (time periods)

Table 3-7. Summary of measurements collected at the Purdue Agronomy Farm for the 1981 field research experiments.

Measurement Type and Description
Spectral Measurements
Reflectance factor
Exotech 100 radiometer (Landsat MSS spectral bands)
Barnes 12-1000 radiometer (Thematic Mapper bands + 1.15-1.30 μm)
Radiant temperature
Color photographs (oblique and vertical views)
Meteorological Measurements
Air temperature
Barometric pressure
Relative humidity
Wind speed and direction
Total solar incidence
Agronomic Measurements
Crop development stage
Amount of vegetation
Plant height
Percent soil cover
Number of plants per square meter
Number of leaves per plant
Leaf area index
Total fresh and dry biomass (g/m^2)
Dry biomass of leaves, stems, and heads, ears or pods (g/m^2)
Crop condition
Plant water content (g/m^2)
Presence and severity of stress
Soil background condition
Percent moisture
Roughness
Grain yield
Canopy Geometry Measurements

Table 3-8. Summary of 1981 data acquisition by field radiometer systems at the Purdue Agronomy Farm.

Measurement Date	Sensor*	Experiment						
		Soybean Cultural Practice	Corn Cultural Practice	Soybean Cultivar	Sunflower Cultural Practice	Corn Hybrid	Sorghum Cultural Practice	Row/Sun Angle
Week of		Number of data sets acquired during week						
June 14	A	1	1	1	1	1	1	-
21	A	1	1	1	1	1	1	-
28	-	-	-	-	-	-	-	-
July 5	A	2	2	2	2	4	2	-
12	-	-	-	-	-	-	-	-
19	A	1	1	-	-	-	-	-
26	A	1	1	1	1	-	1	-
Aug 2	-	-	-	-	-	-	-	-
9	A	1	1	1	1	-	1	-
16	C	2	2	2	2	-	2	-
23	-	-	-	-	-	-	-	-
30	-	-	-	-	-	-	-	-
Sep 6	C	2	2	2	2	-	2	-
13	C	1	1	1	1	-	1	-
20	B	-	-	-	-	-	-	1
27	C	1	1	1	1	-	1	1
Oct 4	C	1	1	-	-	-	1	-
11	C	-	-	-	-	-	1	-
18	C	-	-	-	-	-	-	1

* A - Exotech 100, B - Barnes 12-1000, C - Both Exotech and Barnes.

Laboratory Measurements

The painted barium sulfate reflectance reference panels used by the several truck-mounted systems were prepared and calibrated by Purdue to support the acquisition of comparable data from site to site. During this past year, fourteen panels were prepared and calibrated. The reference panels were calibrated with a bidirectional reflectance factor reflectometer (1). Reflectance measurements of the panels were obtained for illumination zenith angles from 10 to 55 degrees. The reference panels and their calibrations, Figure 3-1, were distributed to several field research teams including University of Nebraska, Kansas State University, Oregon State University, South Dakota State University, and USDA-ARS at Bushland, Texas.

Data Preprocessing

The spectral, agronomic, and meteorological data are calibrated and preprocessed into comparable formats for easy access and analysis by researchers. The spectrometer/radiometer data are preprocessed into LARSPEC format and the aircraft scanner data are preprocessed into LARSYS format.

Preprocessing of all the 1980 spectral data except for the FSS data were completed early this year. Preprocessing of the last three dates of the 1979 FSS data and most of the 1981 calibration panel data have also been completed. Preprocessing of the 1981 Exotech 100 and Barnes 12-1000 data collected at the Purdue Agronomy farm and the 1980 FSS data is underway. The preprocessing accomplishments for 1981 and the present status are summarized in Table 3-9.

Data Library and Distribution

The development of the field research data library at Purdue/LARS was initiated in the fall of 1974 by NASA/Johnson Space Center (JSC) with the cooperation of the United States Department of Agriculture (USDA) as a part of the Large Area Crop Inventory Experiment (LACIE). The purpose of the data base is to provide fully annotated and calibrated multitemporal sets of spectral, agronomic, and meteorological data for agricultural remote sensing research. Spectral, agronomic, and meteorological measurements were made primarily over wheat for three years. In 1978 and 1979 the data base was expanded to include data collected for corn and soybean experiments in Indiana, Iowa, and Nebraska, as well as from a major U.S. soils experiment. In 1980 the library was expanded again to include data collected for spring wheat, barley, sunflowers, and soybeans in North Dakota, and cotton, rice, and soybeans in Texas.

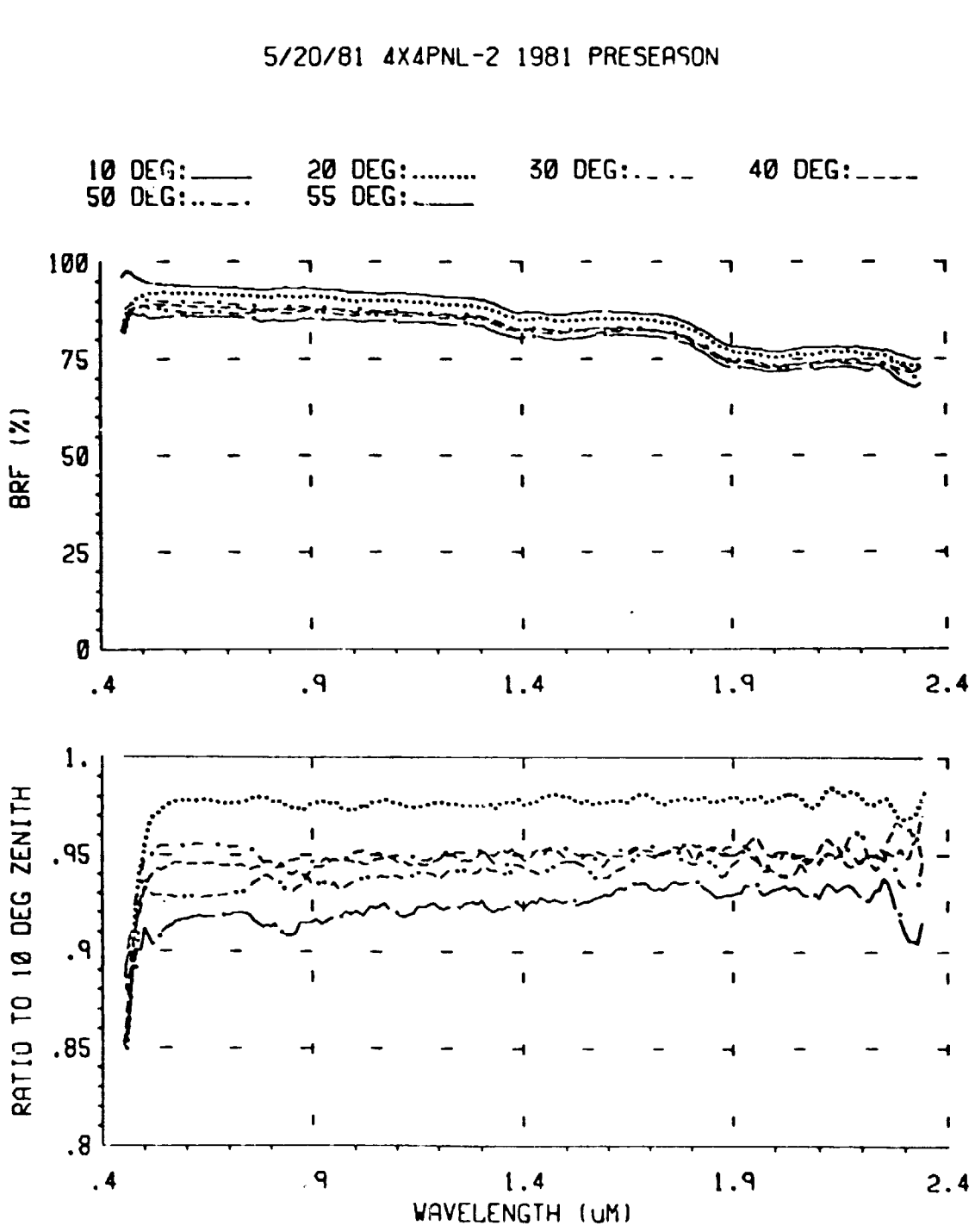


Figure 3-1. Example graph of bidirectional reflectance factor (BRF) calibration of barium sulfate reflectance reference panel that is sent with panel to field research test sites.

Table 3-9. Summary status of field research data preprocessing (1979-81 crop years) accomplishments during 1981.

Instrument/Data Type	Completed	In Processing
Aircraft Multispectral Scanner (dates/flightlines)	14/42	3/
Helicopter Mounted Field Spectrometer (dates/observations)		
Field Averages	3/112	19/
Individual scans	3/1754	19/
Helicopter Mounted Multiband Radiometer (dates/observations)		3/
Truck Mounted Field Spectrometer (dates/observations)		
Purdue/LARS Exotech 20C	20/837	---
Truck Mounted Field Multiband Radiometer (dates/observations)		
Purdue/LARS Exotech 100	18/4453	19/
Purdue/LARS Barnes 12-1000	---	11/
Laboratory Spectrometer (dates/observations)		
Purdue/LARS Exotech 20C	2/52	2/

Milestones achieved during the past year have been:

- Inclusion of much of 1980 crop year data
- Distribution of data to researchers

The data have been collected over several test sites and crops as illustrated in Table 3-10. The test sites are of two types, controlled experimental plots and commercial fields. The instruments used to collect the spectral data are listed in Table 3-11. The spectrometer data are processed into comparable units, bidirectional reflectance factor, in order to make meaningful comparisons of the data acquired by the different sensors at different times and locations (2). The multispectral scanner data are

approximately linearly related to scene radiance. The information is available for the researcher to calibrate the scanner data to in-band bidirectional reflectance factor if desired.

The Field Research Data Library Catalog summarizes the data available. The catalog includes a separate volume for each crop year during which data were collected. In the past twelve months, 42 aircraft scanner runs and more than 8,000 additional spectrometer/multiband radiometer observations have been made available to researchers. The data includes spectral observations of corn, soybeans, and wheat. A summary of the spectral data in the library is given in Table 3-12. Eleven institutions, listed in Table 3-13, have received or accessed field research data during the past year.

Hardware and Software Development Documentation

The achievements during the past year in hardware and software development to preprocess, verify, and analyze the 170,000 observations of spectrometer/multiband radiometer data and 440 flightlines of aircraft scanner data include:

- Hardware obtained to replace antiquated systems and provide new capabilities
- Implementation of wavelength correction for FSS data in LARSPEC
- Implementation of capability to plot data with different wavelength resolutions on the same graph.
- Implementation of graphics software to support development stage analysis
- Implementation of updated version of Graphics Compatibility System (GCS)
- Addition of NASA/JSC Tektronix 4002A as output device for GCS and LARSPEC

Hardware

A new 12 bit analog to digital (A/D) converter was completed this past year. It replaces an obsolete 1970 system that broke down during this past year. Also during this past year a new leaf area meter was purchased to provide a more efficient system for measurement of leaf area index. A new hard copy unit was purchased to copy graphs from the Tektronix 4054 graphics terminal. Additional hardware items include line quantum sensors to measure intercepted solar radiation and polarizer filters to make canopy polarization measurements.

Table 3-10. Summary of field research test site locations and major crops.

Test Sites		
State County	Major Crop	Crop Years
Indiana, Tippecanoe	Corn & Soybeans Winter Wheat	1978-80 1979-80
Iowa, Webster	Corn & Soybeans	1979-80
Kansas, Finney	Winter Wheat	1975-77
Nebraska, McPherson	Corn	1979-80
North Dakota, Cass	Spring Wheat Barley Sunflowers Soybeans	1980
North Dakota, Williams	Spring Wheat	1975-77
South Dakota, Hand	Spring Wheat Winter Wheat	1976-79
Texas, Wharton	Cotton Rice Soybeans	1980
U.S. & Brazil	250 Soil Types	1978

Table 3-11. Summary of major sensor systems used for field research, 1975-1981.

Platform and Sensor	Years
Spacecraft Multispectral Scanners	
Landsat 1	1975-77
Landsat 2	1975-79
Landsat 3	1978-80
Aircraft Multispectral Scanners	
24-channel Scanner (MSS)	1975-76
11-channel Modular Multispectral Scanner (MSS)	1975-79
8-channel Thematic Mapper Simulator (NS001)	1979-81
Helicopter-mounted Spectrometer	
NASA/JSC Field Spectrometer System (FSS)	1975-80
Truck-mounted Spectrometers	
NASA/ERL Exotech 20D Field System	1975
NASA/JSC Field Signature Acquisition System (FSAS)	1975-77
Purdue/LARS Exotech 20C Field System	1975-80
Truck-mounted Multiband Radiometers	
Purdue/LARS Exotech 100 Landsat Band Radiometer Field System	1977-80

Table 3-12. Summary of spectral data in the field research data library by instrument and data type for 1975-81 crop years.

Instrument/Data Type	Dates	Number of Observations or Flightlines
Landsat MSS		
Whole Frame CCT	124	124
Aircraft Multispectral Scanner	76	441
Helicopter-Mounted Field Spectrometer		
Field averages	89	7,791
Individual scans	89	132,292
Truck-Mounted Field Spectrometer		
NASA/JSC FSAS	44	813
Purdue/LARS Exotech 20C	111	7,624
NASA/ERL Exotech 20D	45	654
Truck-Mounted Field Multiband Radiometer		
Purdue/LARS Exotech 100	86	21,491
Laboratory Spectrometer		
Purdue/LARS Exotech 20C	35	1,243

Table 3-13. Recipients of field research data during 1981.

Organization	Means of Distribution
Colorado State University Ft. Collins, Colorado	Mail
Environmental Research Institute of Michigan Ann Arbor, Michigan	Mail & Computer Terminal
Goddard Institute for Space Studies New York, New York	Mail
Kansas State University Manhattan, Kansas	Mail
NASA Johnson Space Center Houston, Texas	Computer Terminal
Purdue University West Lafayette, Indiana	Computer Terminal
Research and Data Systems Lanham, Maryland	Mail
South Dakota State University Brookings, South Dakota	Mail
State University of New York Syracuse, New York	Mail & Computer Terminal
University of Nebraska Lincoln, Nebraska	Mail
USDA Statistical Reporting Service Washington, D.C.	Mail
USDA, Agricultural Research Service Phoenix, Arizona	Mail Mail

Software

LARSPEC was updated to handle the wavelength correction of the FSS data. This software is being used to evaluate the wavelength correction before modifying the FSS data on the library tapes. Also the capability was implemented in LARSPEC to allow plotting and comparing data from two different instrument systems on the same graph, Figure 3-2.

Software was also developed to support the development stage analysis. This included implementing an updated version of the Graphics Compatibility System (GCS) with enhancements developed by Westinghouse Corporation. The software for the development stage analysis allows the researcher to plot data in two or three dimensions from any defined perspective.

Software was also developed so that the NASA/JSC Tektronix 4002A graphics terminal can be used as an output device for LARSPEC and GCS on either the NASA/JSC or Purdue/LARS computers. This capability allows researchers at JSC to view plots of the spectrometer/radiometer data on a high resolution graphics terminal.

Acknowledgements

The experiment design was led by Marvin Bauer and Craig Daughtry. The plot preparation and agronomic measurements of corn, soybeans, sorghum and sunflowers were directed by Craig Daughtry. Larry Biehl was responsible for the spectral measurements. Jon Ranson, Chris Brooks, Judith Ward, Lois Grant, Kevin Gallo, Tom Bonsett, Todd Plantenga, Jim Long and Vic Fletcher assisted in plot preparation and data collection. Ching-Neu Lue, Cathy Kozlowski, Gay Benson, Pam Keuneke, and Selma Al-Abbas were responsible for data preprocessing. Jerry Majkowski, Todd Plantenga, John Cain, and Jim VanMilligan assisted with the data base management, distribution and software development. Mike Stabenfeldt, Tom Bonsett, and Mike Guba assisted with hardware development.

References

1. DeWitt, D.P. and B.F. Robinson. 1976. Description and evaluation of a bidirectional reflectance factor reflectometer. Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana. LARS Information Note 091576.

2. Robinson, B.F. and L.L. Biehl. 1979. Calibration procedures for measurement of reflectance factor in remote sensing field research. SPIE Vol. 196. Measurements of Optical Radiation, SPIE, Box 10, Bellingham, WA. pp. 16-26.

3. Bauer, M.E., L.L. Biehl and B.F. Robinson. 1980. Field research on the spectral properties of crops and soils. AgRISTARS Tech. Report SP-P0-04022 and Purdue/LARS Tech. Report 112680. pp. 111-113.

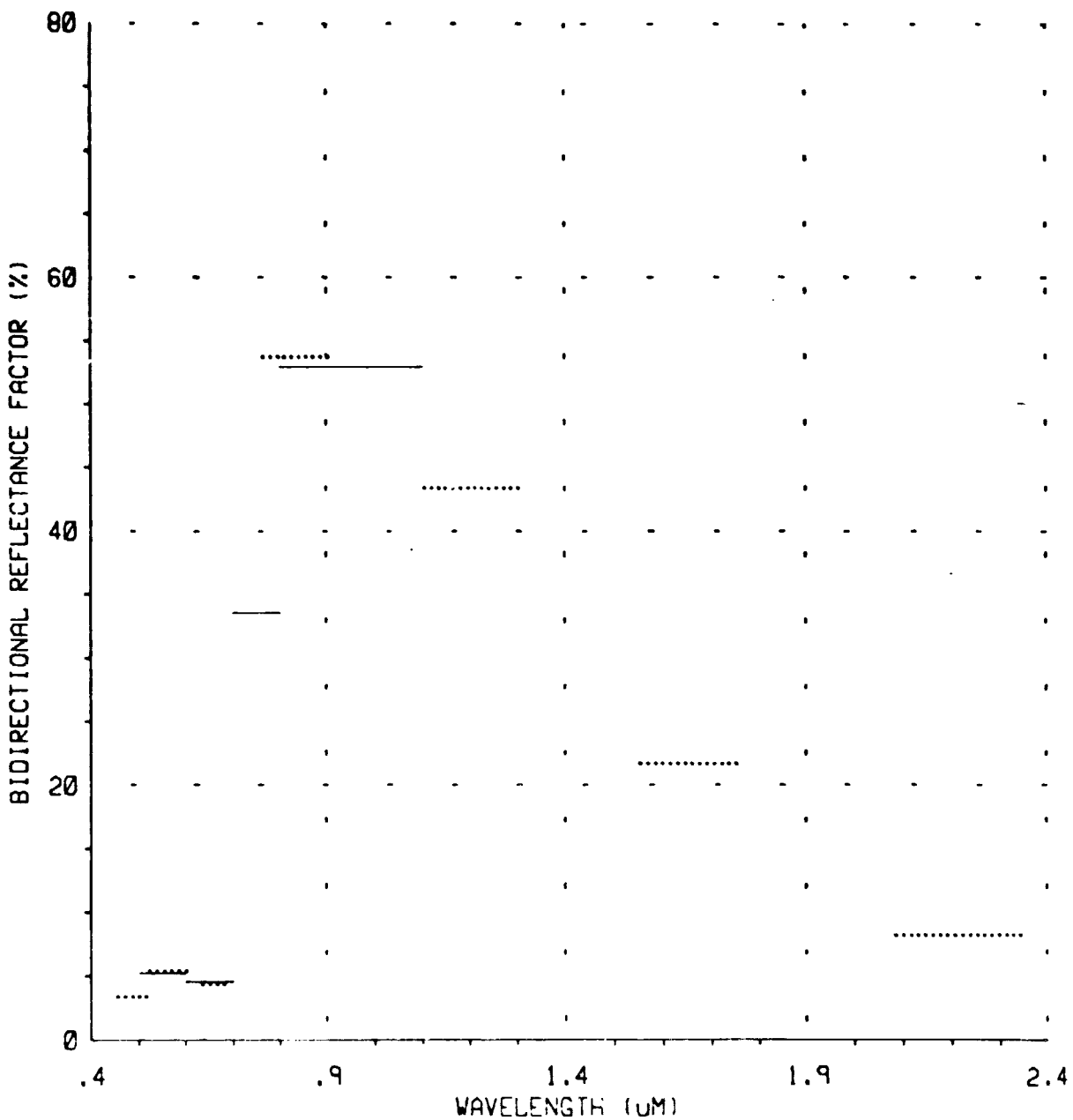


Figure 3-2. Graph of Exotech 100 Landsat band and Barnes 12-1000 Thematic Mapper band radiometer data illustrating plotting of reflectance data with different wavelength resolutions on the same graph.

4. MULTIBAND RADIOMETER SYSTEM FOR FIELD RESEARCH

B.F. Robinson

Introduction

The need for high quality field rated multiband radiometer with wide spectral coverage has increased with the growing interest in the spectral-biophysical characteristics of crops and soils. A practical means to obtain spectral data from a wider variety of subjects and to increase the number of researchers who can afford to acquire and analyze such data is to simplify the instrumentation and reduce the amount of data obtained for each observation. To achieve this, a field-rated multiband radiometer system having a limited, yet sufficient number of wavelength bands has been developed. It is:

- capable of complete spectral coverage (appropriate bands from 0.4 to 2.5 μm and a thermal infrared band);
- comparatively inexpensive to acquire, maintain, and operate;
- simple to operate, calibrate, and service;
- rugged, light weight, portable;
- complete with data handling hardware and software; and
- well documented for use by researchers.

The prototype of multiband radiometer and a high speed data logger was developed and tested under previous tasks (1,2). The radiometer, 12-1000 Modular Multiband Radiometer now commercially available from Barnes Engineering Co., Stamford, CT, has eight spectral bands (0.45-0.52, 0.52-0.60, 0.63-0.69, 0.76-0.90, 1.15-1.30, 1.55-1.75, 2.08-2.35, 10.4-12.5 μm) and is rated for operation from tripod, truck-boom, helicopter and light aircraft platforms. Using available 12 bit data loggers, the radiometer is well suited to the needs of remote sensing field researchers who require portability, spectral accuracy, well defined fields of view, appropriate dynamic range, and excellent radiometric performance in the ranges 0.4 to 2.5 um and 10 to 15 μm.

The objectives of the current task were to acquire multispectral data acquisition systems and install them at agricultural research facilities participating in the AgRISTARS Supporting Research project. The systems include a Barnes Model 12-1000 and a data logger (typically a Model 516 Polycorder manufactured by Omnidata International, Inc., Logan, UT). A pick-up truck mounted boom, an over-cab platform, a calibration platform, and calibration panels manufactured by Purdue/LARS are included in most systems (Figure 4-1). Installation and training at each site by LARS personnel were complemented by a three day workshop held at LARS in October 1981 and a two day workshop in May of 1982.

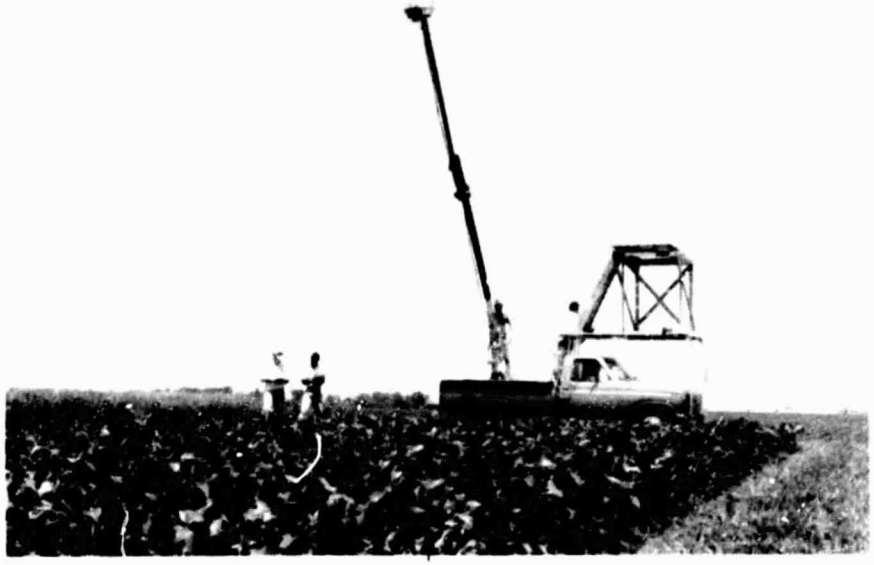


Figure 4-1. Pickup truck data acquisition system, with boom, multiband radiometer, camera, data logger, calibration tower and panel.

Hardware Testing

Laboratory and Field

The development of the multiband radiometer and initial tests were completed in 1980 (2). The prototype was returned to the manufacturer for final modifications and was tested at LARS in the spring of 1981 (3), prior to construction of the production units. Additionally, a comparison of the performance of the prototype to two other widely used instruments was conducted (4). Laboratory tests for stability and radiometric performance of the Model 12-1000 indicated that all specifications were met. Initial field tests of the units met specified performance levels and correct operation with the Polycorder 516 units was verified.

The spectral responsivity of the instrument closely matches the thematic mapper satellite scanner (with an additional band at 1.15 to 1.30 μm) and the out of band blocking is effective in all channels. Results of the radiometric tests indicate the potential use of narrow, user defined pass bands in the range from 0.4 to 2.4 μm. The fields-of-view are well suited to remote sensing field research. The acceptance of radiance from the well defined 15 degree cone enables spatial "averaging" while limiting the off-axis rays to less than 7.5 degrees. The one degree field of view may be used for more detailed measurements or for spatial averaging from an airborne platform. Of particular importance is the absence of spurious off-axis responses, which can completely invalidate many measurements. In field tests of radiometric performance, with reflectances ranging from 2.9% to 48%, the maximum fractional error was less than one percent of value.

The programability of the data loggers and their accuracy and precision was tested. It was found that the Polycorders could be programmed satisfactorily if no data memory was active. It was also found that some units produce occasional extraneous readings, but otherwise could easily measure 5 volts to 12 bit precision when a 20 millisecond integration time was used. These errata provided no serious problems and the manufacture has corrected the units free of charge.

It was found (in the field) that some of the multiband radiometers produced extraneous readings on the thermal channel when used in the presence of bright infrared targets of great extent. The problem was traced to the chopper position sensor used to obtain the reference signal for synchronous demodulation in the thermal channel. Don Dusek, USDA-ARS, Bushland, Texas and Robert Buckley, Barnes Engineering, cooperated to identify and solve this problem. A small baffle provided by Barnes Engineering was distributed along with a field modification kit. Tests using intense infrared sources did not produce the symptoms in modified units.

Helicopter

On September 16, 1981, personnel from Purdue/LARS and NASA/JSC installed the Barnes Model 12-1000 on a helicopter at the Webster County, Iowa, test site (Figure 4-2). On September 17 the first data flight was made (Figure 4-3). Additional data flights were made on September 21, 1982, at the Cass County, N. Dakota test site and on October 19, 1981 at the Wharton County, Texas test site.

The reflectance data for the commercial fields compare well with the data gathered simultaneously with the S-191H spectroradiometer. The temperature of the chopper and detectors increased monotonically at an average rate of 0.05C per minute. This indicates that the radiometer is adequately stable, even in the presence of the air flowing directly across the case.

Hardware Acquisition

To provide the needed instrumentation to facilities participating in the AgRISTARS Supporting Research project, the following equipment was acquired in 1981:

15	Barnes Model 12-1000 Modular Multiband Radiometers
12	Omnidata Model 516 Polycorders
6	Pick-up truck booms with calibration platforms
10	4 x 4 foot painted barium sulphate calibration panels
5	35 mm, 250 exposure motor-drive cameras
4	35 mm, 36 exposure motor-drive cameras

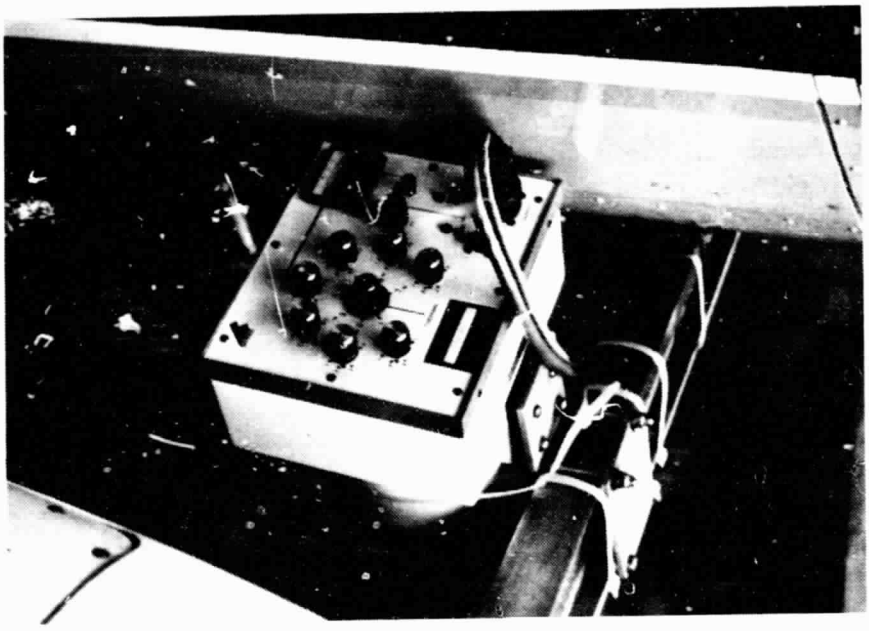


Figure 4-2. Model 12-1000 multiband radiometer mounted on helicopter for test.

Figure 4-3. Helicopter with Model 12-1000 radiometer over calibration panel.

The Omnidata Model 516 Polycorder is a portable data collection system that accepts keyboard entry of visually observed data or scans voltaged and digital signals from a variety of scientific instruments (Figure 4-4). This unit is suitable for operation from all platforms. For 12 bit accuracy the acquisition interval is about 20 milliseconds per channel which, for the eight radiometric channels, requires 0.16 seconds. In addition to time (to 0.01 second if needed) and observation number, other information may be automatically (or manually) entered prior or subsequent to the acquisition of up to 10 channels of radiometric data. The unit memory which holds about 350 observations (depending on format), may be interrogated at its front panel.

The data logger is 8 x 4.5 x 3 inches in size and weighs about 3.2 pounds with batteries. Interface to a cassette recorder and a standard RS-232 port are standard, built-in features. Because of its versatility the unit must be programmed. This is accomplished, manually, by a prompting format which is resident in the logger. Alternatively, the logger may be programmed by a development system or an assembler resident in another computer. The manufacturer of the radiometer will provide a special purpose version of the Polycorder equipped for turn-key operation with the radiometer.

Four units of a high speed data logger designed particularly for helicopter and small airplane operation are currently being constructed at LARS. All data are recorded in a 0.8 millisecond interval. Year, day of year, a six digit observation number, time (hour, minute and second), and a one digit data type code are recorded prior to recording the 12 bit radiometer data which is identified by a four bit channel code. Up to 15 channels of analog data may be digitized and stored. Data is stored in a

Figure 4-4. Model 516 Polycorder downloading to cassette tape--in the field.

128K byte CMOS memory module with data retention battery. The capacity of each memory module is sufficient to store the 60,000 target observations and 5000 calibration observations which are gathered in a typical helicopter mission. When full, the memory module may be disconnected and another module connected.

Hardware Installation

Multispectral data acquisition systems were installed at twelve sites in fall 1981 and spring 1982: Kansas State University, University of Nebraska, South Dakota State University, Oregon State University, NASA Johnson Space Center, Purdue University, International Maize and Wheat Improvement Center (CIMMYT) at Mexico City, Texas A&M University, University of Kansas, University of Minnesota, NASA Earth Resources Laboratory, and NASA Goddard Space Center. Table 4.1 summarizes the equipment at these locations. Most sites receiving a Model 12-1000 radiometer were visited by personnel from Purdue/LARS and the equipment and procedures were tested while in actual operation by the crews responsible for data acquisition. Close contact has been maintained since installation.

Field Measurements Workshops

On October 14-16, 1981, twenty-five participants from Kansas State University, University of Nebraska, South Dakota State University, Texas A&M University, North Dakota State University, University of Minnesota, CIMMYT (Mexico), Canada Department of Agriculture, Oregon State University, and USDA-ARS attended a workshop at Purdue University.

Formal presentations by Purdue/LARS agronomists and engineers were coupled with informal discussions and presentations by attendees. Topics included: fundamentals of optical radiation as applied to remote sensing field research, operation and testing of the model 12-1000 radiometer, operation and programming of the Model 516 Polycorder, using the Polycorder with the Model 12-1000, field measurement procedures, effects of sun-angle and sensor altitude, design for remote sensing experiments, and 1981 experiments at each location.

Following a trip to the Purdue Agronomy Farm for a demonstration, hands-on experience, and discussion of calibration and testing procedures, the participants returned to the conference room for presentations and discussions of agronomic measurements and procedures.

A discussion of data handling, data transfer, and Purdue/LARS processing facilities was followed by a demonstration of analysis capabilities of LARSPEC. Common goals for large scale experiments and a set of measurements which would be common all remote sensing field experiments were discussed.

On May 11 and 12, 1982, a workshop was held at Purdue/LARS for personnel from the University of Minnesota and the NASA Earth Resources

Laboratory. A total of six attendees included personnel from Rutgers and Michigan State Universities. The content of the workshop was similar to the fall workshop.

Table 4-1. Summary of multiband radiometer field research data acquisition systems in place in summer 1982.

Site	Boom Type	Logger	Camera	Number Cal Panels
NASA/JSC	Note 1	Polycorder	70mm	1
Kansas St. Univ.	KSU	Polycorder	-	1
Kansas St. Univ.	LARS	Polycorder	36MD	1
Univ. Nebraska	LARS	Polycorder	250MD	2
S. Dakota St. Univ.	LARS	Polycorder	250MD	2
Oregon St. Univ.	LARS	Polycorder	250MD	2
Purdue Univ.	LARS	Polycorder	250MD	1
Purdue Univ.	LARS	LARS	250MD	1
CIMMYT	Note 2	Polycorder	36MD	1
Texas A&M Univ.	LARS	Polycorder	250MD	1
Univ. Kansas	U Kan	Polycorder	U Kan	1
Univ. Minnesota	LARS	Polycorder	36MD	1
NASA/JSC	Note 3	Polycorder	36MD	1
NASA/ERL	Note 3	ERL	ERL	1
NASA/GSFC	GSFC	GSFC	GSFC	-

Note 1: Helicopter operation. Note 2: Special trailer boom constructed at LARS. Note 3: Hand held or tripod operation.

Acknowledgments

Marvin Bauer and Barrett Robinson (LARS), Ed Kanemasu (Kansas State Univ.), and David Pitts (NASA/JSC) planned the equipment distribution. Rudy Trabanino (NASA/JSC), Max Herridge (LEMSCO), and Barrett Robinson designed and implemented the helicopter tests. Michael Heidt (NASA/JSC) organized the data entry to the JSC computer. Barrett Robinson, Marvin Bauer, Larry Biehl and Craig Daughtry presented the workshops. Bill Kuhn is assembling the LARS data loggers.

References

1. Bauer, M.E., L.L. Biehl, C.S.T. Daughtry, B.F. Robinson, E.R. Stoner. 1979. Agricultural Scene Understanding and Field Research. Tech. Report 112879, Lab. Applications of Remote Sensing, Purdue Univ., W. Lafayette, IN (also AgRISTARS Report SR-P9-00410).

2. Bauer, M.E., L.L. Biehl, B.F. Robinson. 1980. Field research on the spectral properties of crops and soils. Tech. Report 112680, Lab. Applications of Remote Sensing, Purdue Univ., W. Lafayette, IN (also AgRISTARS Report SR-P0-04022).

3. Robinson, B.F., R.E. Buckley, J.A. Burgess. 1981. Performance evaluation and calibration of a modular multiband radiometer for remote sensing field research. Proc. Soc. Photo-Optical Instrumentation Engineers, Bellingham, WA, 308:146-157.

4. Robinson, B.F. 1981. Performance comparison for Barnes Model 12-1000, Exotech Model 100, and Ideas Inc. Biometer Mark II. AgRISTARS Tech. Report SR-P1-04090, NASA Johnson Space Center, Houston, TX.

5. EVALUATION OF LANDSAT SPECTRAL INPUTS TO CROP CONDITION AND YIELD MODELS

S. E. Hollinger and M. M. Hixson

Introduction

In recent years the trend in development of yield forecast models has been toward more causal linkages between temperature, moisture, and other weather, soil, and cultural variables and specific plant processes related to yield and growth. Research tasks were initiated at Purdue to study how spectral data could be utilized to improve crop yield estimates and to estimate development stages of corn and soybeans (Bauer et al., 1982).

Many types of information are potentially available from remotely sensed data to improve yield forecasts. These include: environmental information such as soil characteristics, meteorological conditions, and episodic events; management variables relative to field preparation and harvesting; and plant characteristics including biomass accumulations, stress effects, and development stage information.

Crop yield and production models, whether they are statistical regression or physiological models, have limited spatial resolution because of a relatively sparse network of weather stations. Landsat MSS data can improve the spatial resolution of the crop yield models by providing a digital image with 0.5 hectare resolution. Therefore, it is reasonable to evaluate the combination of a meteorological crop yield model with Landsat spectral data.

Remotely sensed data can be interfaced with physiological and combined physiological/regression (hybrid) models to serve as direct model inputs and to verify and update the model estimates. One agrometeorological yield which is appropriate as a testbed for the addition of spectral data or agronomic variables estimated from spectral data is CROPCAST, developed by the Earth Satellite Corporation (Merritt, 1981). During FY81 a joint project between LARS and EarthSat was initiated with the overall goal of determining the value of Landsat spectral inputs to crop condition and yield models. In general, LARS was to provide spectral inputs and Earthsat the yield model and crop-soil-weather data bases. Earthsat also has the facilities to collect, reduce and store the meteorological and other data needed to run the various models.

Objectives

The general objectives of the investigation are:

- to develop, implement, and evaluate approaches to use Landsat MSS data with an agrometeorological yield modeling system to improve forecasts of corn and soybean yields,

- to define ways to use an agrometeorological modeling system along with spectral data to improve the prediction of stress effects.

Approach

Preparatory tasks included the formulation of an experiment design and data analysis plan which was accomplished via meetings between the two organizations. LARS supplied to EarthSat the locations of segments, periodic observation data, universal ground truth tapes, and available full-frame Landsat data for 1978-79 to interface with their system.

All evaluations have been conducted on a field level since the density of available data is not large enough to allow for aggregation, to the county or district level.

Evaluations of spectral data to improve yield estimates include:

1. Test of the Kanemasu (1978) adaptation of the Priestly-Taylor evapotranspiration model for sensitivity to spectral and meteorological input variables and a test of the response of CROPCAST to the Kanemasu ET model.

2. Evaluation of effect of spectral albedo input to CROPCAST.

3. Examination of sensitivity of the CROPCAST model to the uncertainties associated with remotely sensed inputs.

4. Evaluation of effect of spectrally derived development stage and planting date on CROPCAST results.

5. Correlation between a spectrally derived stress index and other crop stress indices (i.e. CROPCAST, Crop Moisture Index).

Thus far, results have been obtained for the latter four objectives.

Planting Date Sensitivity

The sensitivity of CROPCAST to planting date was examined so that the effect of the errors in determining planting date on the model could be defined. This analysis helped to determine how accurate the spectral planting date model should be.

The sensitivity of CROPCAST to planting date involved running the CROPCAST system for different planting dates beginning when the crop is 10 percent planted and ending when 90 percent of the crop is planted. Simulation runs were made for every other day during the period. Four different geographical areas were used for the corn planting date sensitivity and five areas for soybeans. The four corn areas were Marion County, Georgia; Allen County, Indiana; Peoria County, Illinois; and Polk County, Iowa. The five soybean areas were the same as corn, plus Shelby County, Tennessee. A total of twelve years were included in the study.

Albedo Computation

A spectral estimate of albedo was calculated using equation (1). The equation was developed by Sabatinni et al. for calculating albedo from Landsat data.

where A is the albedo, O is the solar zenith angle; Li is the radiance in each of the four MSS bands; S is the scattered atmospheric radiation (arrow pointing up is radiation scattered to space; arrow pointing down is diffuse radation to the surface); T is the average transmittance of the atmosphere for each of the four MSS bands, and E is the integrated solar spectral radiances for each of the four bands. The scattering and transmittance values are functions of solar elevation angle only.

Development Stage Estimation

The spectral estimation of development stage was accomplished by LARS using the model developed by Badhwar and Henderson (1981). The analyses include a total of 1364 fields from 62 segments for 1978 and 1979. Planting date was estimated from the spectrally derived development stage by Earthsat using the meteorological development models.

The effect of the spectrally estimated development stage will be evaluated by comparing the results of the CROPCAST system runs with and without the spectral input. Initially comparisons of the spectrally-estimated development stage to the periodic observation data have been made.

Spectral Stress Index

A spectral stress index was obtained for each of the fields in 1978 and 1979 using the Green Index Number (GIN) developed by Thompson and Wehmanen (1980). This required the modification of the GIN algorithm to apply to the field level. This spectral stress index was compared to the stress Palmer Crop Moisture Index (CMI) and will be compared to the index obtained from the CROPCAST model. The stress indices provide a stress-no stress conclusion and were evaluated using contigency tables.

Results and Discussion

Planting Date Sensitivity

The data obtained from the planting date sensitivity runs consist of a series of points representing the within year potential yield loss due to moisture and temperature stress as a function of planting date. To analyze the data, the slope of the yield loss lines for various planting date "errors" was determined.

Using a criteria of less than 5 percent potential yield loss as an acceptable error, preliminary analysis revealed that 44 percent of the corn crop years and 15 percent of the soybean crop years would result in yield loss errors of 5 percent or greater. Table 5-1 shows the percent of time a risk of 5 percent or greater error will occur given a spectral planting date precision. From this table it can be seen that the precision in the corn planting date model must be greater than the soybean planting date model to achieve the same accuracy.

Additional analysis of the planting date sensitivity is planned. The additional analysis will be conducted the same as the analysis described above but will include a wider range of error limits and planting date precision estimates.

Table 5-1. Risk of having a 5 percent or greater error in yield prediction given a planting date accuracy.

Planting Date Accuracy (Days)	% Time When Error > 5%	
	Corn	Soybean
4	20	-
10	30	5
20	-	15

Albedo Computation

The mean albedo for 3 different segments in 1978 are shown in Figure 5-1. The significance of these plots is that albedo is different throughout the season depending upon the surface characteristics of the segment. For example, segment 824 in Iroquois County, Illinois, shows a large variability of albedo throughout the season, while segment 133 in Whitely County, Indiana, shows very little variability. The difference in the variability of albedo could easily be explained by different soil colors.

Segment 127, in Montgomery County, Indiana, shows variation throughout the growing season but is not as dramatic as segment 824. The interesting feature of the albedo plot in segment 127 is the dip in the middle of the season. This dip is abnormal and could be the result of a significant atmospheric interaction not accounted for by the albedo equation.

In sensitivity tests of the CROPCAST model to a varying albedo, the model is showing a significant response. The most significant response

occurs early in the season. Future tests using the spectral albedo input should more fully define the significance of including albedo as a spectral input.

Development Stage Estimation

A spectral estimate of the development stage has been completed for all the 1978 and 1979 segments (Figures 5-2 and 5-3). Although the analysis and evaluation is not yet completed, preliminary results show that in 1978 the corn estimates are good, but in 1979 there was a linear trend error. For soybean, in 1978 the development stage was underestimated, while in 1979 it was overestimated (Figure 5-3). The exact causes of these effects are being investigated further.

Spectral Stress Index

The Landsat estimates of crop stress were compared to the Palmer Crop Moisture Index (CMI) and an area was assumed to be stressed when the CMI was less than -1.0. Although additional tests need to be run using years when more severe stress occurs (e.g. 1980) to check the validity of the approach, in 1978 the procedure predicted 90 and 91 percent of the stress-no stress conditions correctly for corn and soybeans, respectively.

References

1. Badhwar, G.D. and K.E. Henderson. 1981. Estimating development stages of corn from development data - an initial model. Agron. J. 73:478-755.

2. Bauer, M.E., L.L. Biehl, and B.F. Robinson. 1980. Field research on the spectral properties of crops and soils. AgRISTARS tech. report SR-P0-04022. Tech. report 112680, Lab. Applic. Remote Sensing, Purdue University, West Lafayette, IN.

3. Merritt, Earl S. 1981. CROPCAST. Proc. Symp. on Machine Processing of Remotely Sensed Data, pp. 478-490. Purdue University, West Lafayette, IN, June 23-26.

4. Palmer, W.C. 1968. Keeping track of crop moisture conditions nation'de, the new crop moisture index. Weatherwise 21:156-161.

5. Sabatinni, R.R., E.S. Merritt, W.D. Hart, L.J. Heitkemper, D.L. Hlavka. 1976. A Study of Mesoscale Surface Heat and Moisture Budgets and Their Relationship to Airmass Cumulus Clouds Observed in Landsat Imagery. Final report to Goddard Space Flight Center. Earth Satellite Corporation, 7222 47th St., Washington, DC.

6. Wiegand, C.L., A.J. Richardson, and E.T. Kanemasu. 1979. Leaf Area Index Estimates for Wheat from LANDSAT and Their Implications for Evapotranspiration and Crop Modeling. Agron. J. 71:336-342.

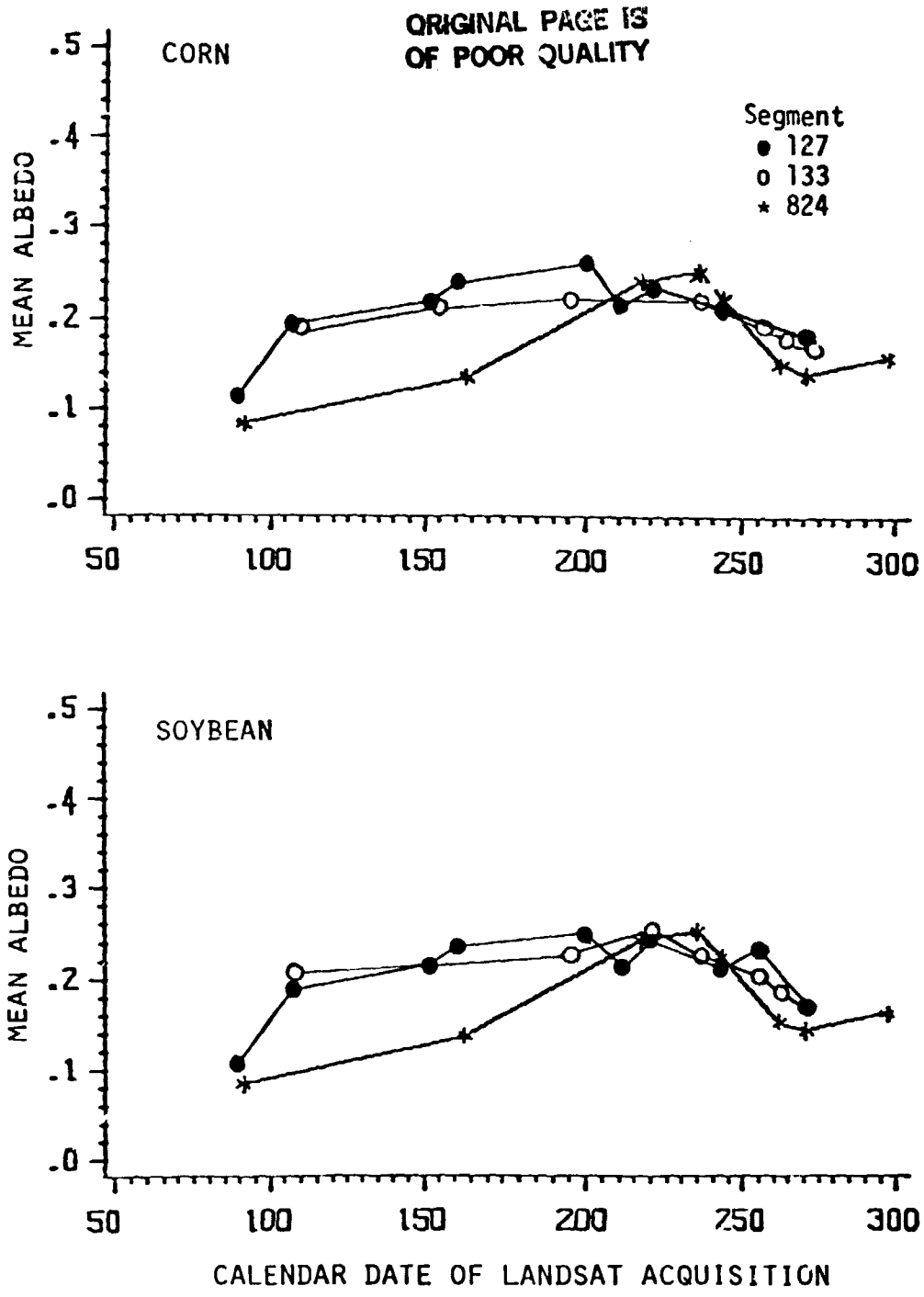


Figure 5-1. Mean albedo estimated from Landsat MSS data for three segments in 1978.

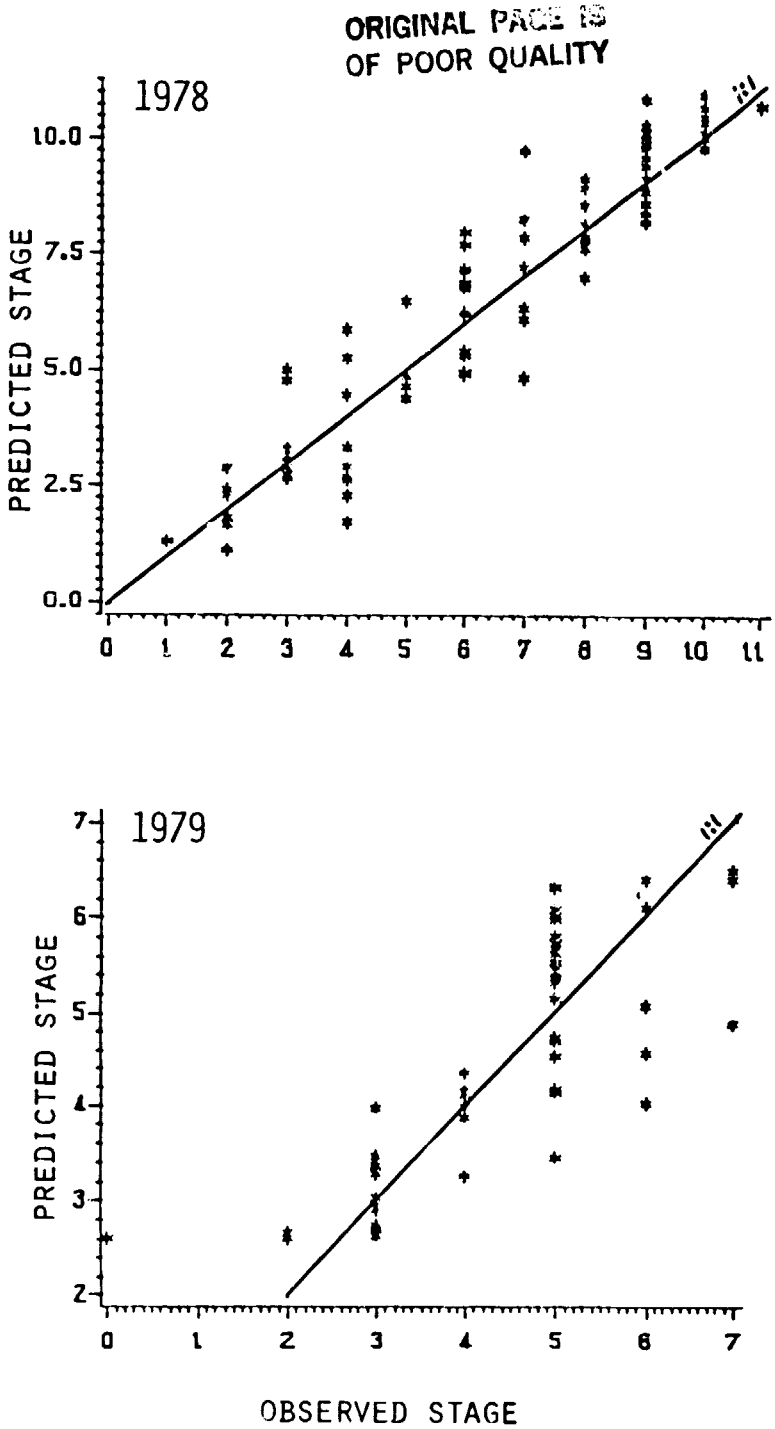


Figure 5-2. Comparison of spectral estimates and field observations of development stage for corn, 1978 and 1979.

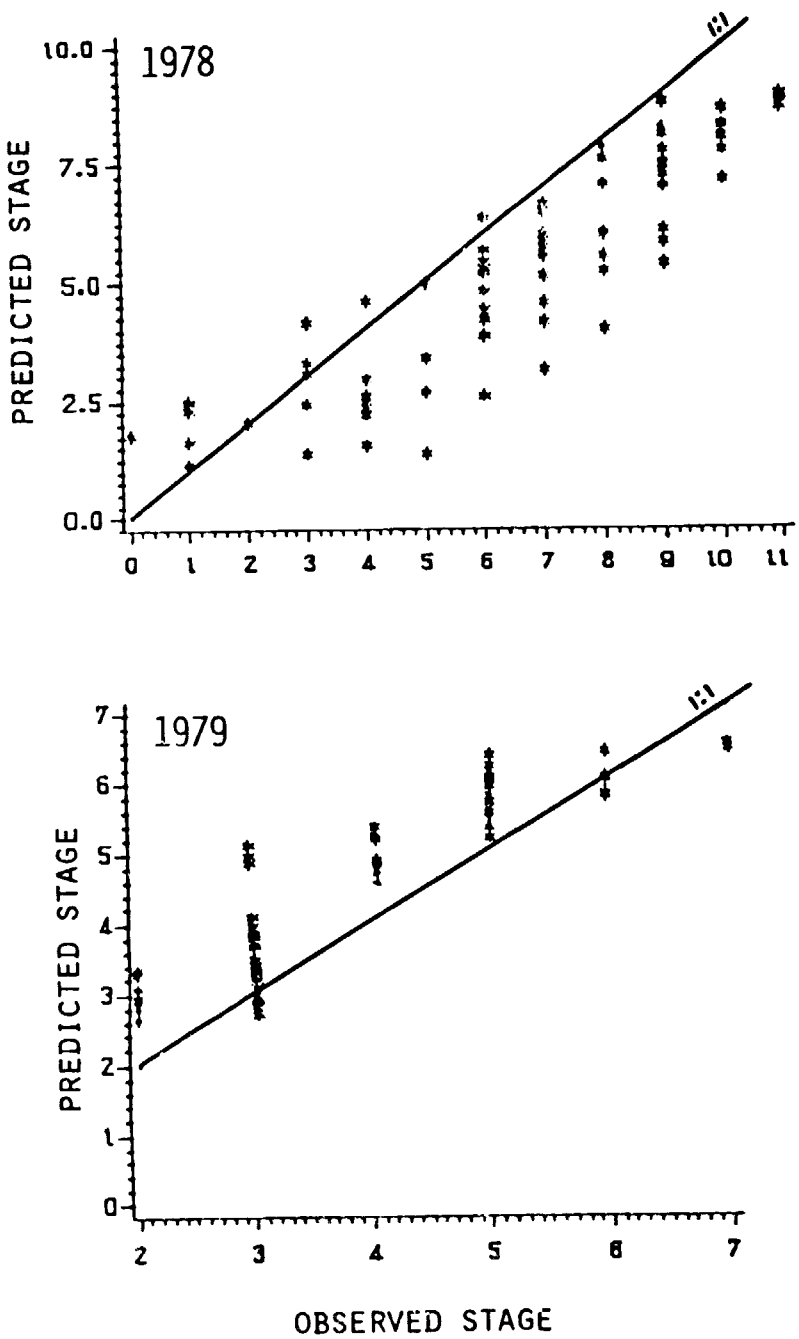


Figure 5-3. Comparison of spectral estimates and field observations of development stage for soybean, 1978 and 1979.

6. FIELD MEASUREMENTS AND SIMULATION MODELING OF CORN AND SOYBEAN MOISTURE STRESS

B.L. Blad, J.M. Norman, and B.R. Gardner
University of Nebraska

Introduction

Differences in moisture availability and stress are major sources of variation in yields of corn and soybeans. While it is generally known that moisture stress affects the spectral characteristics, as well as the growth and yield, of crops, relatively little research on moisture stress effects and relationships has been conducted and relatively little use is made of remotely sensed spectral measurements to assess crop condition. Much of the recent work in the area has been conducted by Jackson and coworkers on wheat in Arizona (Idso, 1977) and Blad and coworkers on corn in Nebraska (Blad et al., 1980).

During 1981 a subcontract was provided to the University of Nebraska (Blaine L. Blad, principal investigator) for research on the effects of moisture stress on the spectral characteristics of corn and soybean canopies. The accomplishments of the first year's research are described in AgRISTARS technical report SR-P2-04259 (Blad, Norman and Gardner, 1982) and will only briefly be described here.

Objectives

The overall objectives are to determine how plant and environmental factors, particularly moisture stress, affect the radiance characteristics of corn and soybean canopies and to determine how to interpret and use remotely sensed spectral data for stress assessment. The specific objectives are:

1. Determine the important biophysical descriptors of crop condition and how stresses, particularly moisture deficits, alter the biophysical characteristics (development and rate of growth, canopy geometry, and physiological properties) of corn and soybean canopies.

    - development, rate of growth
    - canopy geometry (e.g. LAI, percent soil cover, leaf angle distribution
    - physiological properties (e.g. pigment concentrations, stomatal diffusion resistance

2. Determine how changes in biophysical descriptors of canopy condition are manifested in the spectral responses of corn and soybean canopies in Landsat MSS and TM bands transformations.

3. Determine functional relationships between biophysical and radiometric characteristics of corn and soybean canopies as a function of moisture stress.

4. Determine sensitivity of functional relationships to soil, crop and environmental and measurement factors, including

   - soil type, moisture and surface condition,
   - row width, plant population, variety,
   - temperature, wind speed, humidity,
   - sun angle, sensor view angle.

5. Develop and evaluate models for use with Landsat data which permit interpretation of canopy spectral responses and stress levels with a minimum of ancillary data.

Technical Approach

The research being conducted by the University of Nebraska, includes a combination of field measurements and simulation modeling to (1) quantitatively assess the factors affecting canopy radiance and (2) incorporation of essential factors into models suitable for use with satellite data with a minimum of ancillary data.

Field Measurements

The field research will be conducted primarily at the University of Nebraska Sandhills Agricultural Laboratory. With sandy soils and 15-20 inches annual precipitation, there is a high probability of obtaining moisture stress. The gradient irrigation system available at this location will be used. It is designed to produce linearly decreasing amounts of water with increasing distance from the line source. A gradient from full-water to dryland can be developed across each plot. Treatments will include amount (gradient) and timing (during vegetative, flowering, and grain fill periods) of irrigation. The irrigation system used to develop the water gradients is described in detail by Elad et al. (19[illegible]).

The field measurements focus on the effect of water stress on canopy spectral responses and provide two kinds of data: (1) inputs required by the model Cupid and (2) radiance and ancillary data for improving our understanding of the interaction between radiation and vegetation and testing predictions from the model Cupid. The measurements that serve primarily as inputs to the model Cupid are: (1) a standard set of meteorological measurements to include solar radiation, air temperature, soil temperature, vapor pressure, precipitation, wind speed and wind direction; (2) soil moisture for a range of soil water conditions; (3) plant parameters such as stomatal diffusion resistance, leaf area index, plant height and leaf angle distributions; and (4) leaf photosynthesis and transpiration rates with a portable field chamber.

Radiance and ancillary measurements include (1) measurements of canopy radiance with the Barnes multiband radiometer (model 12-1000), an Exotech

100 radiometer, with multispectral scanners during aircraft overflights and with infrared thermometers; (2) visible sky radiance and canopy irradiance at five different viewing angles with a unique sensor (designed at Nebraska) that provides azimuthal averaging - this sensor will provide rapid canopy reflectance measurements so that the effect of solar zenith angle can be determined; and (3) plant water potential and crop yield.

Model Description

The comprehensive plant-environment model (Cupid) that will be used for this research is quite advanced and is regularly undergoing improvements and refinements. Cupid is a physical-physiological model that considers the processes of turbulent transfer in, above and below the canopy, radiative transfer in the canopy, and heat and water movement in the soil (Norman, 1979). The basic approach involves predicting individual leaf processes for leaves of differing orientations and positions in the canopy and then integrating these results over the canopy to predict canopy characteristics. Although the inputs required are minimal and easily obtained, the outputs can be quite detailed and complex. For example, the distribution of temperature, transpiration rate, reflected radiation of leaves of various orientations can be predicted. Profiles of air temperature, relative humidity, radiation fluxes, water fluxes, dew formation, leaf temperature, etc. can be predicted throughout the canopy and temperature and water fluxes predicted in the soil.

With the generality implicit in this model, a wide variety of results can be obtained on many variables through integration over various time and space scales. In its present form, Cupid contains predictions of bidirectional canopy reflectance (at any wavelength) and directional thermal irradiance as a function of sun angle. Cupid is also being used to obtain quantitative canopy architecture estimates of leaf area and leaf angles from radiation measurements; this is accomplished by combining Cupid with an integral inversion program. Further, three dimensional canopy architecture can be incorporated into the radiative transfer portion of Cupid to study canopy signatures under partial cover.

Summary of 1981 Results

Spectral reflectance data were collected on irrigated and non-irrigated (stressed) plots of corn and soybeans on twelve days from June to mid September in 1981. Additional data were acquired on canopy radiant temperatures, as well as soil moisture, leaf area index, plant water status, photosynthesis, transpiration, and other plant parameters.

Analysis of 1981 data is not yet complete, but preliminary results indicate that spectral data from the thematic mapper spectral bands (Barnes MMR) provide greater separability of effects due to water stress than data from the MSS bands (Exotech model 100). Additionally, the middle infrared bands appear to contain more information with respect to moisture stress than the visible or near infrared bands (Table 6-1). On September 14 the corn and soybeans exhibited a maximum range of vegetative conditions, with plants ranging from dry and brown to completely green. A large coefficient

Table 6-1. Coefficient of variation values for the thematic mapper bands for GII (non-stressed) and GGG (stressed) treatments in corn, soybean, and bare soil on September 14, 1981.

	Corn			
Wavelength (μm)	GII	GGG	Soybean	Bare Soil
0.45-0.52	6.6	14.8	32.5	6.4
0.52-0.60	8.4	14.3	29.6	5.6
0.63-0.69	9.0	35.1	49.8	4.4
0.76-0.90	8.5	13.1	10.5	4.2
1.15-1.30	6.1	6.0	9.0	4.2
1.55-1.75	6.2	17.1	20.9	4.3
2.08-2.35	9.3	33.7	43.2	2.9

of variation among the treatments indicates a high potential separating the cover types.

A new index combining thermal and spectral information (TSI) was defined. The TSI appears to be capable of quantifying persistent or severe stress but may fail to detect periodic stress events.

References

1. Blad, B.L., B.R. Gardner, D.G. Watts and N.J. Rosenberg. 1980. Remote sensing of crop moisture status. In Remotely Sensed Crop Temperature for Water Resources Management. Agric. Meteorol. Progress Report 80-5, Chap. II.

2. Blad, B.L., J.M. Norman and B.R. Gardner. 1982. Field measurements and simulation modeling of corn and soybean moisture stress - 1981 field studies. Progress Report 82-2, Center for Agric. Meteor. and Climatology, Univ. of Nebraska (AgRISTARS tech. report SR-P2-04259).

3. Idso, S.B., R.D. Jackson and R.J. Reginato. 1977. Remote sensing of crop yields. Science 196:19-25.

4. Norman, J.N. 1979. Modeling the complete crop canopy. In Modification of the Aerial Environment of Crops. eds. B.J. Barfield and J.F. Gerber. Am. Soc. Agric. Engrs., St. Josephs, Michigan. pp. 249-277.

III. Soil Moisture Research

7. IMPACT OF SURFACE SOIL MOISTURE MEASUREMENTS ON MODELING CROP GROWTH, DEVELOPMENT, AND YIELD

C.E. Seubert, C.S.T. Daughtry, S.F. Hollinger and D.A. Holt

Introduction

Assuming that radar data can provide accurate estimates of soil moisture for the top 5 cm of the soil profile, how useful will this information be for assessing crop growth, development, and yield? Research is needed (1) to define the relationship between surface and profile soil moisture contents and (2) to evaluate the usefulness of independent measurements (radar) of surface soil moisture for predicting crop growth, development and yield. Large area crop models will be used to assess surface soil moisture measurements. This report describes an initial step to evaluate the optimum number of classes needed to describe the water holding capacity of soils in Indiana and proposes future research.

Crop models for large area yield predictions frequently require a soil water balance submodel to estimate the available soil moisture throughout the growing season (Holt et al., 1979). Most soil water models require information about the available water holding capacity (AWHC) of each soil in order to estimate soil moisture. Grouping soils by AWHC produces classes that can be used in the soil water balance models to improve the accuracy of crop yield model predictions. Additionally, by combining many different soil profiles with similar AWHC into a few classes the computational time for calculating the soil water balance for a large geographical area is reduced greatly. One of the chief difficulties, however, in classifying or using information about soils is that they do not fall into discrete units, but exist in a continuum which is only artificially divided.

The objective of this initial research on soil moisture was to describe and evaluate a method for quantitatively grouping soils into available water holding capacity classes. These classes are based upon natural groupings of the available soil water holding capacity as estimated using particle size and parent material type.

Experimental Approach

Data, obtained from the Purdue University Soil Characterization Laboratory, contained laboratory analyses and field observations for 902 soil profiles representing 184 soil series (Franzmeier et al., 1977). These soils represented the full range of variability in water holding capacities present in Indiana soils. For selected soil series additional data were obtained from Soil Conservation Service soil profile descriptions.

The available water holding capacity (AWHC) is the maximum amount of water held in a soil between 0.3 and 15 bars of tension. AWHC was calculated for each soil horizon using the particle size of the soil

horizon, its position in the profile, and the underlying parent material. Available water holding capacity was reduced for greater bulk density in certain kinds of C-horizons and fragipans which limit root penetration.

Each soil profile was divided into 1.0-cm increments from the surface to 150 cm and AWHC was calculated for each 1.0-cm increment. Each profile was represented by ten 15-cm layers, subsequently referred to as layers 1 through 10. AWHC for ten 15-cm layers were calculated and expressed as cm water/15 cm of soil. The ten AWHC values for each soil profile were the ten variables used in the clustering procedure. A multivariate cluster procedure, FASTCLUS, (SAS Institute, 1981) was used to aggregate the 902 soil profiles into their natural available water holding capacity classes. Three different cluster analyses assigned each soil profile to one of 4, 8, or 12 classes based upon the calculated 10 dimensional Euclidean distances among the available water holding capacity values.

Results and Discussion

Multivariate Analyses

Table 7-1 summarizes AWHC for selected layers and indicates representative soil series in each class. Mean available water holding capacities as a function of depth for the 4-cluster analysis are shown in Figure 7-1A. The four cluster classes are distinct in their AWHC characteristics. The means of the cumulative AWHC are significantly different (Table 7-1). However, with only four classes, soils differing greatly in profile characteristics are grouped into the same class. For example, the Houghton, an organic soil, and the Ragsdale, which developed in Wisconsin age loess are both grouped into the first cluster class. Similarly, the Crider which developed in Wisconsin age loess over limestone and the Miami, which developed in moderately deep loess underlain by Wisconsin age glacial till, are two very different soils which were both included in the fourth class. Grouping all soils into only four classes produces classes containing such widely differing soils that the usefulness of this scheme is limited.

In Figure 7-1B AWHC of each class of the 8-cluster analysis is plotted as a function of depth. Eight cluster classes provided more types of AWHC curves than the 4-cluster analysis. The additional classes allowed unusual or outlier soils to be represented by separate classes. For example, classes 3 and 6 contain only eight soil profiles; however, these series differ considerably in AWHC from the other series used in the analyses.

The 12-cluster analysis (Figure 7-1C) produced a wider range of water holding capacity curves than either the 4- or 8-cluster analyses. Several series of curves differing only in the slope of their decline were identified. For example, classes 7, 8, and 11 (Figure 7-1C) all have greater than 3.0 cm of AWHC in the surface layer and decline to less than 0.5 cm/layer. Only the depth at which the change occurs varies among these classes. Classes 1, 3, 4, and 5 are basically parallel curves and differ nearly uniformly in their available water holding capacity throughout the soil profile.

Table 7-1. Available water holding capacity in 30 cm layers and representative soil series for the 4-, 8-, and 12-cluster analyses.

Cluster Class	n	Soil Layer, cm					Representative Soil Series	
		0-30	31-60	61-90	91-120	121-150	Poorly drained	Non-poorly drained
		cm H_2O/layer						
						4-Class Analysis		
1	162	6.8	6.1	5.9	6.0	5.8	Ragsdale, Houghton	Flanagan, Huntington
2	201	4.8	3.7	2.9	2.1	1.9	Gilford, Maumee	Brady, Brems
3	305	5.8	5.0	4.8	4.6	4.1	Brookston, Kokomo	Del Rey, Raub
4	234	6.4	5.6	4.9	3.3	2.5	Clermont, Pewamo	Blount, Crider
						8-Class Analysis		
1	137	6.2	5.3	5.1	3.9	2.3	Kokomo, Pewamo	Dubois, Hosmer
2	253	6.4	5.3	4.5	3.4	3.3	Brookston, Clermont	Crider, Miami
3	7	7.2	6.8	5.3	2.0	1.5	None	Muskingum
4	65	6.0	4.6	2.9	1.2	0.9	None	Belmore, Fairmont
5	253	6.7	5.8	5.7	5.7	5.4	Birds, Ragsdale	Cuba, Dana
6	1	9.0	9.0	9.0	9.0	9.0	Houghton	None
7	107	3.8	3.1	2.9	2.4	2.2	Gilford, Maumee	Brady, Brems
8	79	4.3	4.2	4.4	4.4	4.1	Bono, Lenawee	Oshtemo, Princeton
						12-Class Analysis		
1	1	9.0	9.0	9.0	9.0	9.0	Houghton	None
2	52	5.8	4.8	4.7	3.5	1.8	Pewamo	Alida, Corwin
3	65	3.2	2.7	2.7	2.4	2.3	Conrad, Maumee	Ayr, Brems
4	69	4.2	4.0	4.3	4.2	3.8	Bono, Zipp	Bronson, Oshtema
5	139	6.8	6.0	5.9	6.0	6.0	Birds, Mahalasville	Cuba, Flanagan
6	43	4.8	4.3	4.2	2.0	1.5	Gilford, Sebewa	Brady, Frederick
7	3	7.6	7.2	5.8	1.5	0.6	None	Muskingum
8	31	6.2	5.0	2.6	0.8	0.7	None	Belmore, Lucas
9	176	6.5	5.6	4.8	3.2	2.9	Rensselaer	Crider, Miami
10	232	6.4	5.6	5.4	5.1	4.3	Brookston, Clermont	Algier, Raub
11	8	5.9	2.4	0.4	0.2	0.2	None	Corydon, Fairmont
12	83	5.9	4.2	3.2	2.8	2.8	Adrian, Mussey	Blount, Frederick

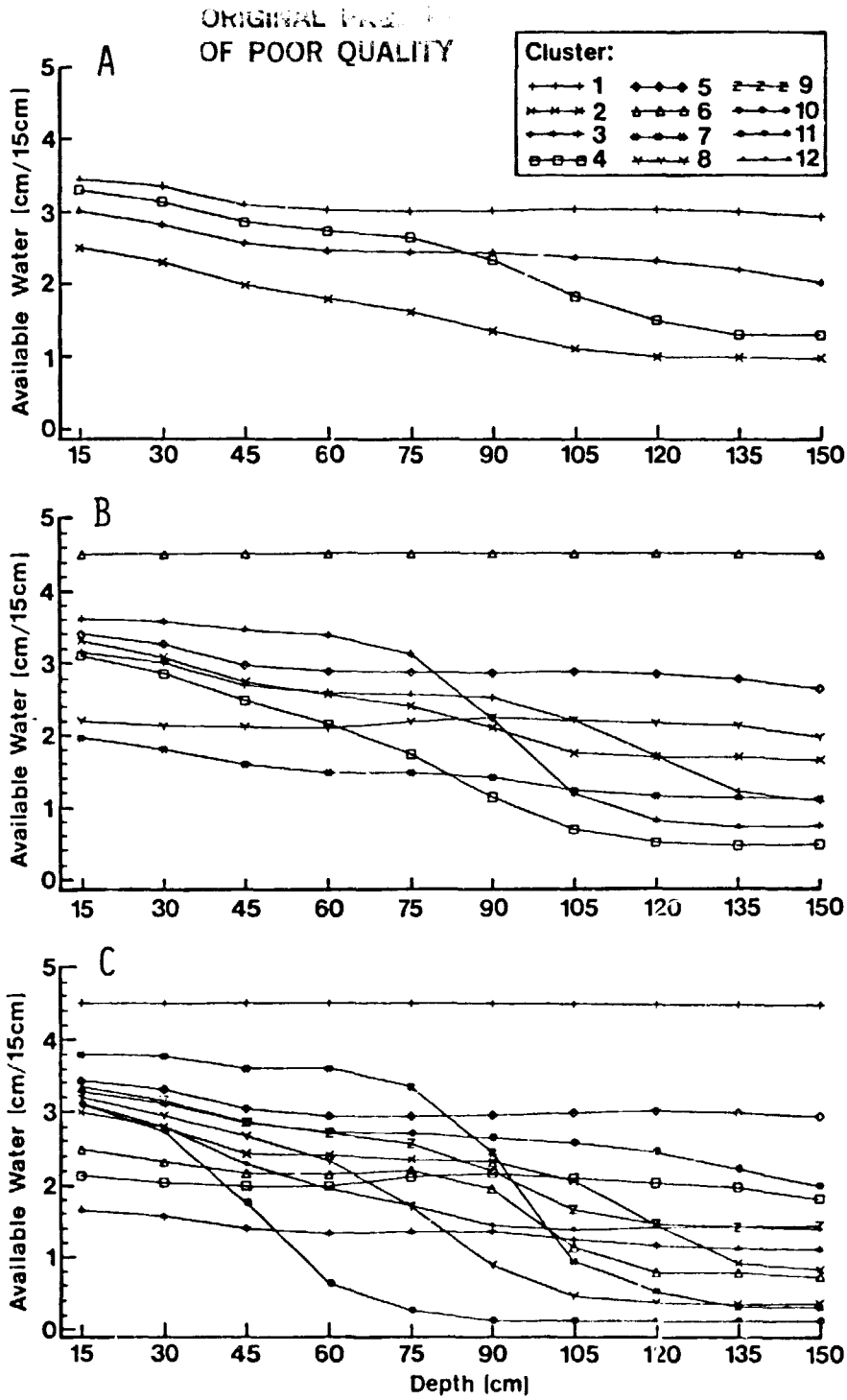


Figure 7-1. Available water holding capacity versus depth for the 4-, 8-, and 12-cluster analysis.

A quantitative evaluation of the superiority of one of these sets of cluster classes is difficult. The optimum number of classes depends on the range of data being classified. In this case the best quantitative assessment would be cluster classes which differed significantly in their soil water balance relationships. This assumes that the wide variety of soil profiles in a region could be represented adequately by the AWHC cluster classes. The best method of this assessment would be within the context of a crop yield prediction model.

Univariate Analysis

Another way to evaluate AWHC classes of the cluster analysis is to compare cumulative available water holding capacities of each class. The sum of AWHC for the first seven layers of soil (0 to 105 cm) represents the soil moisture that potentially would be available to a corn or soybean crop during flowering (Table 7-2). While this single dimensional variable does not adequately describe the complex interactions of AWHC with growing plants as well as multidimensionally based classes, cumulative AWHC should allow some quantitative comparisons among classes within each cluster analysis.

In the 4-cluster analysis the cumulative water holding capacity (0 to 105 cm depth) of each class is significantly different. This was expected since the curves (Figure 7-1A) for these four classes are simple and distinct. The multidimensional clustering of AWHC with only four classes has little advantage over the simple one dimensional classification based on cumulative AWHC.

The curves for the 8-cluster multidimensional analysis are noticeably more complex than those of the 4-cluster case. The mean cumulative AWHC of classes 3 and 5 do not differ significantly from one another (Table 7-2). However, the soils in classes 3 and 5 are very different in their AWHC profiles (Figure 7-1B) and their expected productivities for non-irrigated crops.

Based on cumulative AWHC, the 12-cluster analysis provided only one more statistically separable class than did the 8-cluster analysis. Those classes in the 12-cluster analysis which were not significantly different on the basis of cumulative water holding capacity have distinct AWHC curves (Figure 7-1C). For example, classes 5 and 7 have nearly identical cumulative AWHC at 105 cm. However, soils in class 7 are shallow soils underlain by compact glacial till with high water holding capacities only in the upper layers, while soils in class 5 are deep soils with moderately high water holding capacities throughout the profile. Similarly, classes 4, 6, 8, and 12 were not significantly different in cumulative AWHC at 105 cm although AWHC as a function of depth changes dramatically.

Conclusions

Because AWHC is difficult to visualize and compute in ten dimensional space, the total AWHC in the soil profile also was used to classify soils. While cumulative AWHC may adequately represent the total moisture supplying capacity of a soil, it does not account for changes in AWHC with depth. The

Table 7-2. Mean cumulative water holding capacity in the top 105 cm of soil for the 4-, 8-, and 12-cluster analyses.

Cluster Code Number	Number of Samples	Mean
		cm
	4-cluster analysis	
1	162	21.8 a
4	234	18.7 b
3	305	17.9 c
2	201	12.5 d
	8-cluster analysis	
6	1	31.5 a
5	253	21.0 b
3	7	20.5 b
1	137	18.7 c
2	253	17.9 d
8	79	15.1 e
4	65	14.1 f
7	107	11.0 g
	12-cluster analysis	
1	1	31.5 a
5	139	21.6 b
7	3	21.5 bc
10	232	19.9 c
9	176	18.5 d
2	52	17.4 e
12	83	14.7 f
4	69	14.5 f
6	43	14.4 f
8	31	14.2 f
3	65	9.9 g
11	8	8.7 h

Mean values followed by the same letter are not significantly different at the 0.05 level.

multidimensional approach groups soils profiles with similar AWHC as a function of depth into classes. Thus the multidimensional approach represents the AWHC profile better than total AWHC because: (i) soils with widely divergent AWHC profiles but the same total AWHC are not placed in the

same class, (ii) abrupt changes in AWHC caused by glacial till, fragipans, lithic contacts, etc. are represented, and (iii) the classes of AWHC provide valuable information for crop yield models which include soil moisture submodels. This multidimensional approach can be readily modified to include other characteristics which restrict root development. For example, percent aluminum saturation could be used to describe aluminum toxicity which would restrict root development in certain horizons of some soils (Gonzalez-Erico et al., 1979) and thus would effectively reduce the AWHC of those soils. Finally, as new crop varieties with tolerance to unfavorable soil conditions are developed, the multidimensional approach has the capability to evaluate the changes in AWHC and potential yields.

Future Work

The overall goal of this research is to evaluate radar data as a source of information for use in crop models. Specifically, this task will:

- define the relationship between surface and profile soil moisture contents, development, and yield estimates, with and without detailed soil profile information.

- determine how much additional information about soil characteristics is necessary to use measurements of the surface soil moisture effectively.

- compare results of estimating crop growth, development, and yield with and without information about surface soil moisture conditions.

To achieve these objectives the following steps are planned: (1) select soil moisture and crop yield models, (2) implement models on LARS computer system, (3) develop weather crop and soil data bases to be used in simulation, (4) run simulation studies, (5) analyze results and write technical reports.

References

1. Franzmeier, D.P., G.C. Steinhardt, J.R. Crum and L.D. Norton. 1977. Soil characterization in Indiana: I. Field and laboratory procedures. Purdue Univ. Agr. Exp. Sta. Res. Bull. 943. 30 pp.

2. Gonzalez-Erico, E., E.J. Komprath, G.C. Naderman and W.V. Soares. 1979. Effect of depth of lime incorporation on growth of corn on an oxisol of central Brazil. Soil Sci. Soc. Am. J. 43:1155-1158.

3. Holt, D.A., C.S.T. Daughtry, H.F. Reetz and S.E. Hollinger. 1979. Separating soil and water effects in large area yield prediction. Agron. Abstracts 71:12.

4. SAS Institute, Inc. 1981. SAS 79.5 Changes and Enhancements. Statistical Analysis System Tech. Rep. p. 115. Raleigh, NC.

IV. Sampling and Aggregation Research

8. EVALUATION OF A SEGMENT-BASED LANDSAT FULL-FRAME APPROACH TO CROP AREA ESTIMATION

M.M. Hixson, S.M. Davis, and M.E. Bauer

Introduction

Two general approaches to sampling Landsat MSS data for crop area estimation have been presented in the literature. One approach, used in LACIE (MacDonald and Hall, 1980), used sample segments (cluster samples) extracted from Landsat data over the region of interest for both classification and area estimation. An alternative sampling procedure separates the functions of sampling for training and sampling for classification and area estimation (Bauer et al., 1978).

Objectives

The objective of this study was to further assess the effect of separating the functions of sampling for training and sampling for classification and area estimation. This approach requires ancillary data over only a small number of areas for training, but permits classification and crop area estimation over a relatively large geographic region. Specifically, three related questions were addressed:

(a) How should training statistics be developed from the segment data to be representative of a larger area?

(b) What methods should be utilized to determine over what geographic region the training statistics apply?

(c) How does the accuracy of area estimates differ when segments compared to a systematic sample of pixels are used for estimation?

Approach

The Landsat frame (21295-16013) selected for analysis was acquired over north central Iowa on August 9, 1978, during the best time period for identification of corn and soybeans with unitemporal data (Hixson et al., 1980). The 5 x 6 nm sample segments with digital ground truth which fell within the frame were used to provide training and test data. Although the use of single-date Landsat data does not permit classification or area estimation accuracies as high as could be obtained using multitemporal data, it is expected that the relationship of accuracies among methods obtained with unitemporal data is the same as with multitemporal data.

Results

Development of Training Statistics

The first objective of this study was to examine how training statistics should be developed from the segment data to best represent a stratum. To examine this objective, a stratum containing three counties (Emmet, Palo Alto, and Pocahontas) and five sample segments was selected.

The results for all of the sample segments in the stratum showed that higher classification accuracies were achieved when the training statistics were developed on each segment and then pooled (Training Procedure 1) than when the fields were pooled by type before clustering (Training Procedure 2). The area estimates for both corn and soybeans were closer to USDA/SRS estimates when the statistics were first developed on each segment separately. The root mean square (RMS) errors are 2.6 vs. 3.3 for corn and 3.7 vs. 5.8 for soybeans. Based on the results of this study, the remaining analyses discussed in this paper will use Training Procedure 1.

Stratification Methodology

Four different spectral stratification systems were examined:

1. The refined strata, defined from agrophysical units and used for allocation of sample segments in AgRISTARS.

2. A modification of the refined strata, formed by deleting the southernmost county.

3. The refined/split strata, defined as a substratification of the refined strata for yield estimation.

4. A modification of the refined/split strata, formed by deleting the county furthest south in one of the strata.

These stratifications will be referred to as Stratification Methods 1, 2, 3, and 4.

All the counties were grouped into one stratum using Stratification Method 1. Two segments in Webster County were spectrally anomalous from the rest of the segments, so that county could not be considered to be in the same spectral stratum with the other counties. One possible reason for this is that Webster County has significantly different patterns of precipitation than the other counties. Since it is further south, it may also contain crops in different stages of development than the other counties. Thus, Webster County was deleted from the stratum to form Stratification Method 2.

Stratification Method 3 divided the region of interest into two refined/split strata. When segment statistics were pooled to create statistics for the eastern stratum, again the Webster County segments were anomalous. Thus, Webster County was again deleted, resulting in Stratification Method 4.

The results of this analysis illustrate that both the refined strata and the refined/split strata are apparently too broad to use as spectral strata. In defining spectral strata, other factors need to be taken into account, such as local weather, crop development stage, soil productivity, soil type, and confusion crops present.

Further analyses were conducted using Stratification Methods 2 and 4 only.

Comparison of Pixel and Segment Samples for Area Estimation

Three sampling schemes were compared as a basis for classification and crop area estimation in eleven counties:

- Method A: estimation based on segment training and classification (the LACIE method).
- Method B: estimation based on segment training and classification of a systematic sample of pixels throughout one stratum (Stratification Method 2).
- Method C: estimation based on segment training and classification of a systematic sample of pixels throughout two strata (Stratification Method 4).

Two types of accuracies were considered: classification accuracy and proportion estimation accuracy. Since ground data were available only on segments, classification accuracies were based on segment evaluation. Proportion estimation accuracy was evaluated on a county basis by comparison with the USDA/SRS estimates.

Classification Accuracy

Classification accuracies were generally higher on the segments when statistics representing that segment alone were used in the classification. This is to be expected since spectral confusion classes are more likely to be present in the larger geographic region of the stratum. This result probably indicates, however, that a better spectral stratification still needs to be defined. Most segments had higher classification accuracies when two strata rather than one stratum were used. This confirms the previous hypothesis that spectral strata are somewhat smaller than the refined strata.

Proportion Estimation Accuracy

Table 8-1 shows the comparison between corn proportion estimates made by each of the three methods with the USDA/SRS estimate for the same county. The correlations between the Landsat and USDA/SRS estimates for corn are relatively high for all three methods ($r = 0.77$, 0.62, and 0.83 for Methods A, B, and C, respectively). For soybeans, however, the correlations were

Table 8-1. Proportion estimates of corn and soybeans made using three different stratification and sampling methods and USDA/SRS estimates for the same regions.

	Corn				Soybean			
Stratum/County	A	B	C	USDA	A	B	C	USDA
West								
Emmet	38.9	47.3	44.4	40.0	41.4	37.0	38.9	37.1
Palo Alto	38.6	46.7	43.8	41.8	28.2	38.5	41.5	38.5
Pocahontas	39.2	46.0	41.2	41.1	35.8	41.2	45.1	39.8
East								
Kossuth	43.9	49.5	51.6	43.1	46.8	35.9	40.0	39.6
Humboldt	49.8	50.6	53.0	45.6	47.8	36.4	39.4	40.2
Winnebago	46.6	46.1	47.4	42.1	41.2	37.8	43.6	37.8
Hancock	51.5	49.7	51.9	44.4	31.0	36.7	40.5	34.8
Wright	50.6	50.8	53.6	47.2	43.1	37.2	39.6	42.8
Worth	48.1	46.2	48.2	43.0	35.1	37.7	42.8	32.6
Cerro Gordo	46.5	46.6	48.7	40.8	31.7	35.3	41.3	29.3
Franklin	46.9	50.9	53.7	44.5	32.2	34.2	38.0	30.7

much lower except for Method B ($r = 0.81$). Methods A and C had correlations of 0.51 and 0.09, respectively.

Table 8-2 compares these estimates to the USDA/SRS proportion estimates examining the root mean square (RMS) errors of the several methods. In the western stratum, Method C performed competitively with Method A (2.6 vs. 2.2) for corn, and both systematic sampling methods performed better than Method A for soybeans.

Conclusions

The potential for using pooled segment statistics for an entire stratum is indicated by the generally good classification and proportion estimation performance for both corn and soybeans in the western stratum. This type of training approach used with classification of a systematic sample of pixels merits further investigation due to the increased precision which could be obtained. In particular, the potential shown for this method should be more fully investigated using multitemporal data which should produce still higher classification accuracies and more accurate area estimates.

Table 8-2. Root mean square errors of corn and soybean proportions from USDA/SRS estimates are given for three different stratification and sampling methods.

	Corn			Soybean		
Stratum	Method A	Method B	Method C	Method A	Method B	Method C
West	2.2	5.8	2.6	6.8	0.8	3.7
East	4.5	5.1	7.3	4.3	4.1	6.9
Overall	4.0	6.0	6.4	5.1	3.6	6.2

However, a key factor in using a systematic sampling approach for area estimation has been found to be the definition of spectral strata - that region over which one set of training statistics can apply. It has been illustrated that the refined and refined/split strata based on agrophysical units are not of sufficient spatial resolution to provide a good spectral stratification. Research into the phy-ical factors defining the strata and into methods of stratification will be an important task in the development of a full-frame sampling strategy.

References

1. Bauer, Marvin E., Marilyn M. Hixson, Barbara J. Davis, and Jeanne B. Etheridge. 1978. Area estimation of crops by digital analysis of Landsat data. Photogr. Engin. 44: 1033-1043. JSC-16015.

2. Hixson, Marilyn M., Marvin E. Bauer, and Donna K. Scholz. 1980. An assessment of Landsat data acquisition history on identification and area estimation of corn and soybeans. Proc. Machine Processing of Remotely Sensed Data Symp., Purdue University, West Lafayette, Indiana, June 3-6, pp. 72-77.

3. MacDonald, R.B., and F.G. Hall. 1980. Global crop forecasting. Science 208: 670-679.

9. DETERMINATION OF OPTIMAL LEVEL FOR COMBINING AREA AND YIELD ESTIMATES

M.M. Hixson and C.D. Jobusch

Introduction

The goal of crop inventory is production estimation, not area or yield estimates alone. Production estimates can be made only at a level where area and yield strata intersect. The variance of the production estimates is dependent upon the means and variances of both area and yield in the stratum. Thus, it is important that the stratifications for area and yield estimation be coordinated, and that the levels for aggregation be selected so that acceptable variances are obtained.

In the AgRISTARS Foreign Commodity Production Forecasting (FCPF) Project, estimates are to be made for corn and soybeans area and yield in the United States as well as in Brazil and Argentina. To make production estimates, NASA provides area estimates based on analysis of remotely sensed data and the USDA provides yield estimated from a meteorological regression model. In order to obtain the most precise production estimates, the levels of estimation must be coordinated. Thus, a study to determine the precision associated with several possible levels of aggregation was conducted.

Objectives

The objective of this study was to determine the optimal level for combining area and yield estimates of corn and soybeans. Production estimates and their variances were computed for several levels of area and yield estimates, and the resulting estimates were compared.

Approach

Iowa was selected for study because it is included in the 1981 AgRISTARS pilot experiment. The year selected for evaluation ("current year") was 1978.

The level at which aggregation of area and yield to obtain production should occur is dependent upon the technology being utilized for estimation. If, for example, area or yield estimates made at a given level are biased or unreliable, then aggregation at that level would most likely be undesirable regardless of any potential gains in precision. A change in the technology utilized for estimation, however, might produce reliable estimates at the same level and be a viable candidate for aggregation. This investigation assessed the optimal level with respect to the current technology. Current technology utilizes digital analysis of Landsat MSS data on sample segments to provide area estimates; regression models developed from historical data and used with current weather data provide yield estimates. Several levels of obtaining both area and yield estimates were considered: county, refined

strata and refined split strata (Figure 9-1), crop reporting district, and state.

Estimates of yield at all the levels of aggregation were required. To do this, the variables used in the CCEA state level model were utilized (Table 9-1). Meteorological data at each of the strata levels were required. To obtain estimates at each level, a weather smoothing function (Wagner, 1971) was utilized with the cooperative meteorological station data as input.

Using the smoothed 1978 weather data and regression coefficients based on historical (1932-77) data, "current year" yield estimates were made for corn and soybeans at each of the levels of aggregation. The yield estimate ($\hat{Y}$) and its variance were computed based on the regression equations.

To compare the precision of the several levels, a yield estimate aggregated to the state level was computed. The variance of the aggregated estimates must account for both the variance and covariances of the estimates. Thus, if

$$\hat{Y}_L = \sum_i W_i \hat{Y}_i$$

then its variance can be computed by:

$$V(\hat{Y}_L) = \sum_i W_i^2 V(\hat{Y}_i) + 2 \sum_{i<j} Cov(\hat{Y}_i, \hat{Y}_j) W_i W_j$$

where $\hat{Y}_L$ is the aggregated state yield estimate, W_i is the area weight for stratum i, and $\hat{Y}$ is the regression yield estimate for stratum i.

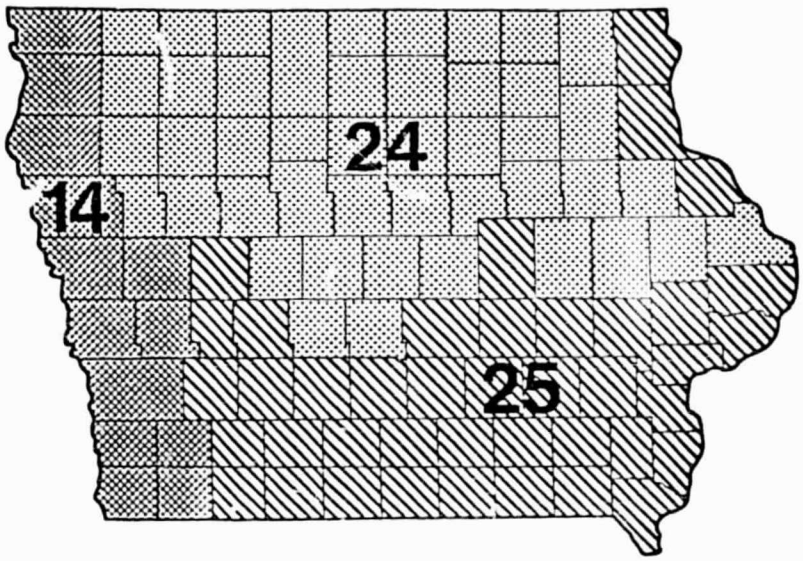
To compute production, the 1978 final area harvested estimates made by the USDA/SRS for the Iowa counties were used. The variance of the area estimates was computed using the methods described by Chhikara and Perry (1980). The number and distribution of agricultural segments in a region were obtained from NASA. For regions without Landsat imagery, the unobserved potential segments were assigned the same distribution of percent agricultural as the observed segments in that county.

For each stratum, the variance of production was computed as:

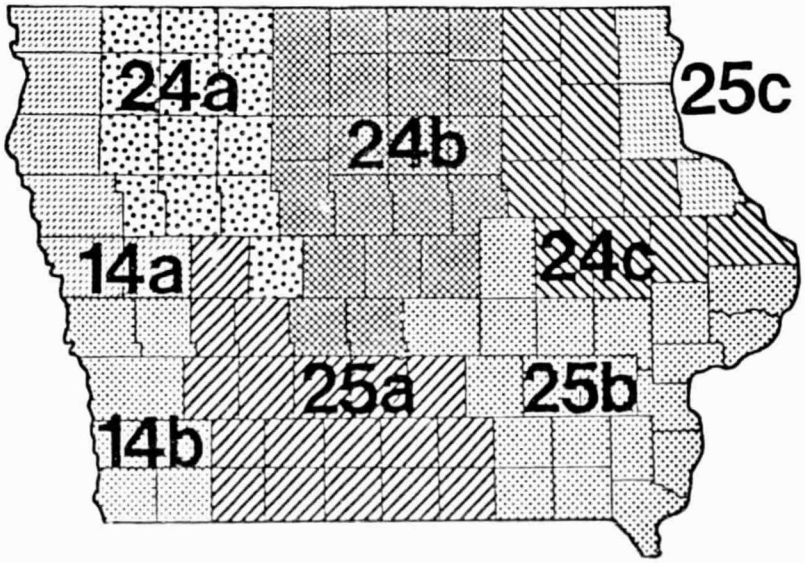
$$V(\hat{P}_i) = V(\hat{A}_i \times \hat{Y}_i) = V(\hat{A}_i) \mu^2_{\hat{Y}_i} + V(\hat{Y}_i) \mu^2_{\hat{A}_i} + V(\hat{A}_i) V(\hat{Y}_i)$$

and for the aggregated state estimate,

$$V(\hat{P}_L) = V(\sum_i \hat{P}_i) = \sum_i V(\hat{P}_i) + 2 \sum_{i<j} Cov(\hat{P}_i, \hat{P}_j).$$



REFINED



REFINED/SPLIT

Figure 9-1. Maps of the refined strata developed at NASA/JSC (top) and the refined/split strata as subdivided for the yield modeling effort (bottom).

Table 9-1. Model variables for the regressions predicting yield of corn and soybeans in Iowa.

Corn	Soybeans
Linear trend 1941-60	Linear trend 1932-74
Linear trend 1961-72	Cumulative precipitation (October - April DFN)
May temperature x precipitation	May temperature x precipitation
June temperature x precipitation	June temperature DFN*
June temperature $(DFN)^2$	July precipitation
July precipiation DFN	July temperature DFT†
July temperature DFT	August precipitation DFN
July temperature $(DFT)^2$	August precipitation $(DFN)^2$
August temperature DFT	August temperature DFT

* DFN = Departure From Normal
† DFT = Departure From Trend

Results and Discussion

Regression Analysis

At the county level, the correlations between values predicted by the regression and USDA observed values had a substantial range. Linn County, with r-square values of 0.93 and 0.92 for corn and soybeans, respectively, is fairly representative of a high correlation situation. Lyon County (r-squares of 0.78 and 0.76) is representative of a lower correlation.

In estimating the yield of corn and soybeans at the crop reporting district level, the correlations between estimated and observed values did not have such a large range as for counties. This is due (at least in part) to the smoothing effects achieved by the consideration of a larger geographic region. The results for the entire state are illustrated in Figure 9-2.

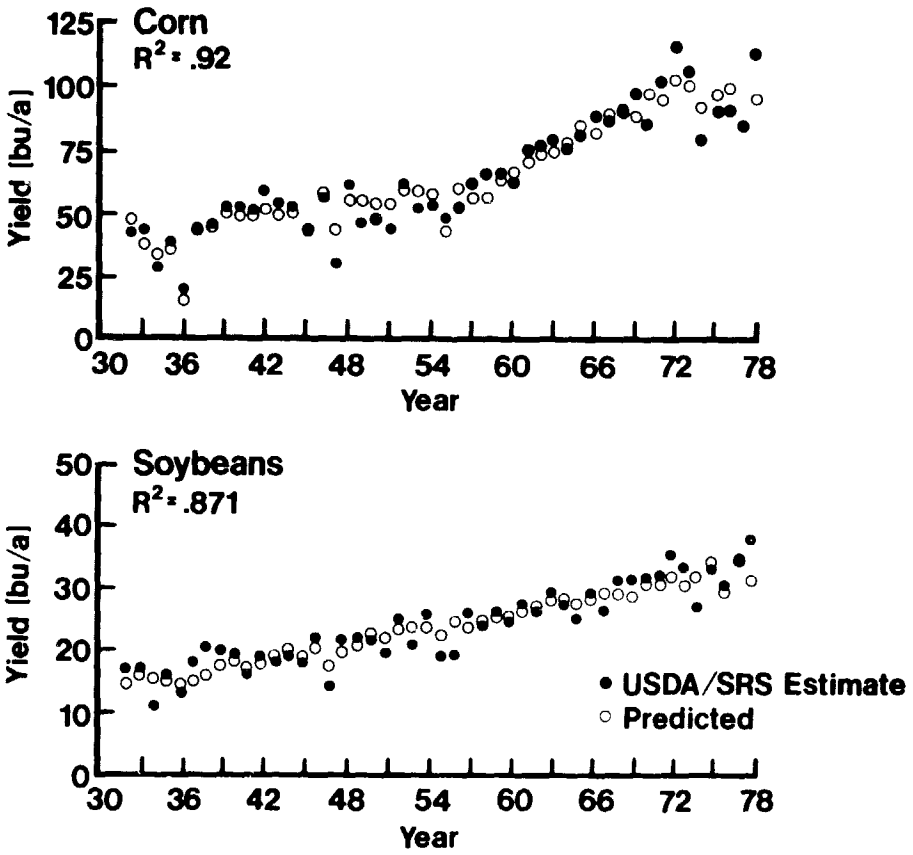


Figure 9-2. Comparison of corn and soybean yields predicted by the regression equations of historical weather data and the USDA/SRS estimates for the entire state of Iowa.

Variance of Yield Estimates

The variance of the yield estimates was computed for each stratum in each of the stratification systems from the regression equations. The aggregate of these results to the state level is illustrated in Table 9-2. For both crops, the state level had the largest variance, and the county level had the smallest variance. The refined strata had somewhat larger variances than those associated with the refined/split strata. The variances for the CRD and the refined/split strata are about the same.

Variance of Production Estimates

The variance of the production estimates computed, based on the preceeding results, is illustrated in Figure 9-3. For corn, the difference in the standard deviations among levels is not great. The differences for soybeans are quite apparent, however, with the standard deviation at the state level being approximately 50% greater than for the other levels. For both crops, the refined strata had the smallest standard deviations.

This result is somewhat surprising since the aggregated yields had shown this method to be of slightly lower precision. What is probably being illustrated, however, is the precision gained by having fewer strata. There are only three refined strata compared with eight refined/split strata and nine crop reporting districts. Due to the strata correlations, relatively more precision was obtained with the fewer refined strata.

Summary and Conclusions

Aggregation of area and yield to production at the state level was the least precise of the methods examined. The crop reporting district, refined strata, and refined/split strata had similar levels of precision of production estimates. The variances of estimates made using the refined strata were somewhat greater.

In examining the variance of yield estimates, the aggregated results from estimation at a county level showed a high precision. Unfortunately, we did not have a mechanism for the estimation of area variances at the county level. This level of estimation should be kept in mind, however, because of its high precision.

The results of this analysis can serve as a data base for further research into the most appropriate level for combining area and yield estimates to obtain production. Two aspects not considered in this study should be part of a further analysis: (1) bias introduced using estimation at any of the levels and (2) the cost of computation of area and yield estimates at the varying levels.

Table 9-2. The standard deviations of corn and soybean yield estimates obtained using each level of stratification. The individual stratum results were aggregated to the state level for comparison.

Stratification System	No. of Strata	Corn	Soybean
State	1	10.5	3.8
District	9	7.0	2.5
County	99	6.3	2.1
Refined	3	8.1	3.0
Refined/Split	8	7.2	2.5

The bias of estimates was considered in this study only to the extent that it did not appear that any of the estimates were biased with respect to the estimates made using any other level of estimation. It may be, however, that either area estimates, yield estimates, or both may have a bias when estimated at one of the levels. The area estimates are currently made at a refined stratum level; no information is available on the potential bias introduced by estimating areas on any smaller geographic region. The yield estimates are now made generally at the state level. Biases may be introduced due to the density of weather stations available for estimating the parameters of the regression equation. A technique such as was utilized in this study may be one possible solution to this problem. However, it is possible that the resulting yields should be smoothed rather than the input meteorological data since the relationship between the input data and predicted values is not linear in the input variables.

The costs of computing the area and yield components must also be considered before a final recommendation can be made. The basis for the decision will consist of consideration of the variances and standard deviations computed as a part of this investigation coupled with cost information for computation of area and yield estimates at each of the potential levels of aggregation. The analysis can be carried out based on sample survey design theory.

When this study has been completed, it may indeed be reasonable to examine some additional crops of interest such as small grains. The results for corn and soybeans were substantially different, with the level at which corn is aggregated making less difference than the level at which soybeans are aggregated.

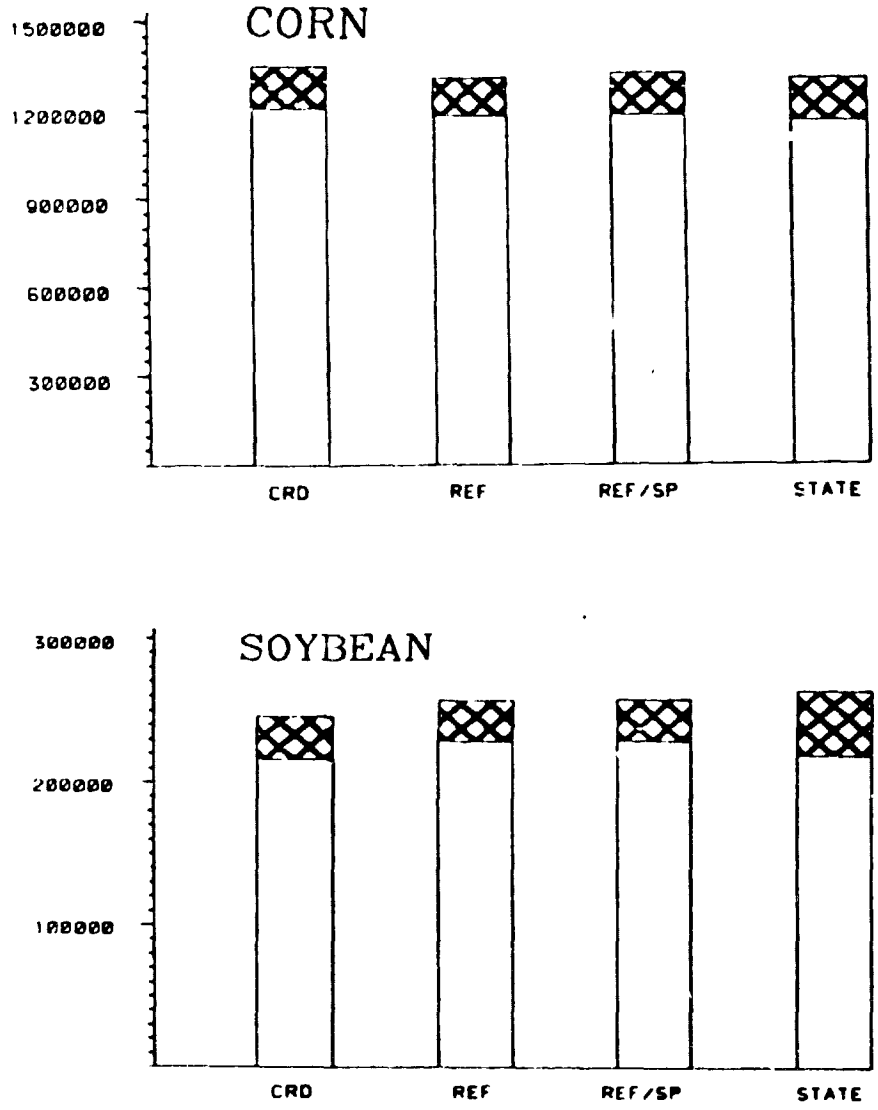


Figure 9-3. Estimated 1978 corn and soybean production for Iowa using the field size variance estimation method. Shaded area is estimated production plus and minus one standard deviation.

References

1. Chhikara, R.S. and C.R. Perry, Jr. 1980. Estimation of Within-Stratum Variance for Sample Allocation. NASA, Johnson Space Center, Houston, Texas. FC-L0-00428, JSC-16343.

2. Wagner, K.K. 1971. Variational Analysis Using Observation and Low-Pass Filtering Constraints. Master's Thesis, Dept. of Meteorology, University of Oklahoma, Norman, Oklahoma.

V. Pattern Recognition and Image Registration Research

10. A FEATURE SELECTION APPROACH FOR MULTISTAGE CLASSIFICATION

D.A. Landgrebe, M.J. Muasher, and P.H. Swain

Introduction

A number of different classifiers are now in routine use in remote sensing. Most of these classification algorithms, using pattern recognition techniques, can be regarded as "single-stage" classifiers, where an "unknown" pattern is tested against all classes using a fixed feature subset, and then the pattern is assigned to one of the classes in a single-stage classification procedure.

A multistage, binary tree classifier, on the other hand, offers several advantages over a single-stage classifier. The binary tree allows one to work with two classes at a time, making the problem easier to understand theoretically, as many theoretical results of pattern recognition apply only to two-class problems. Further, most feature selection techniques are optimal only when applied to two-class problems. As a multistage classifier allows one the use of different feature subsets for discriminating among different subsets of classes, better accuracy should be achieved as compared to a single-stage classifier, where only one feature set can be used.

One of the major needs for a decision tree classifier originates from a dimensionality problem often referred to as the Hughes Phenomenon. In the presence of a limited training sample size, the mean recognition accuracy as a function of dimensionality (number of features for our purposes) exhibits a peaking effect. Contrary to intuition, the mean accuracy does not always increase with additional measurements. Further, peaking of the curves shifts up and to the right, i.e., towards higher accuracy with increasing number of features, as the number of training samples increases, and the peaking disappears only in the case of an infinite number of training samples (complete knowledge of the underlying distributions). Any effective feature selection technique should be able to predict when/if this phenomenon occurs. This is especially true in the context of remote sensing applications because it is frequently the case that training set size is very limited.

Thus, it would be generally very useful, but especially necessary in the case of multistage classifiers, to develop a feature selection algorithm which takes into account the number of training samples used in estimating the statistics of classes and which predicts the best feature subset to be used at each node of the decision tree.

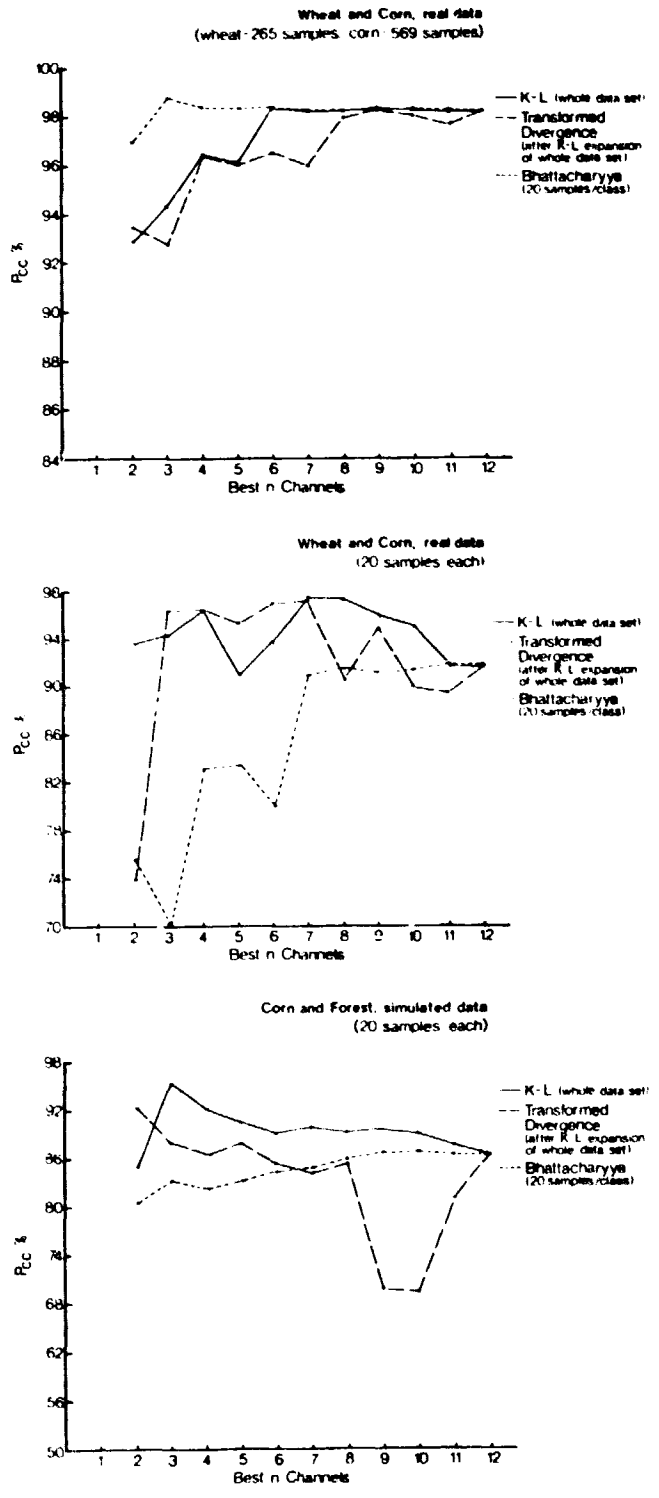


Figure 10-1. Experiments on feature selection using a large number (top) and a small number (middle and lower) of training samples.

Experiments on Feature Selection

To illustrate the feature selection problem in greater detail the results of example experiments will be given. These experiments were conducted based on several feature selection techniques to illustrate the effectiveness of several feature ordering techniques in the presence of a limited set of training samples. Classes were selected from multispectral scanner aircraft data taken on August 13, 1971. Three different techniques are compared using different numbers of training samples: the Karhunen-Loeve transformation, transformed divergence, and the Bhattacharyya distance applied after a simultaneous diagonalization technique was performed on the data to uncouple the features (see Muasher and Landgrebe, 1981). The data was classified using subsets of 2 to 12 features.

Results appear in Figure 10-1. They show that while for the larger training set case, the Bhattacharyya distance appears superior, the K-L transformation appears to be the most effective feature ordering technique in the presence of a limited set of training samples. Some reasons why these experiments turn out this way are discussed in Ref. 1.

Feature Selection Algorithm

In order to facilitate a multistage classifier design procedure, a performance estimation algorithm is developed that estimates the probability of error by computing an approximation to the area under the likelihood ratio function for two classes, taking into account the number of training samples used in estimating the probability densities of each of these two classes. Both theoretical and experimental results are reported in Ref. 1, but only experimental results will be given here to illustrate algorithm characteristics.

Figures 10-2 and 10-3 show results comparing probabilities of error predicted by the proposed algorithm as a function of dimensionality as compared to experimental observations for both simulated and real data. Both aircraft and Landsat data were used, with the experimental curves being the average classification results of five different sets of training samples with the same size.

Results indicate that the algorithm predicts the best, or near best, subset of features to be used. The algorithm results have the same shape as the trends in the corresponding experimental curve. The standard deviations plotted for the experimental curves tend to confirm the expected trend that, in general, an increase in dimensionality results in an increase in the variance of error, the increase becoming highly noticeable at high dimensionality when the randomness in the estimated statistics is large for a fixed, finite set of training samples.

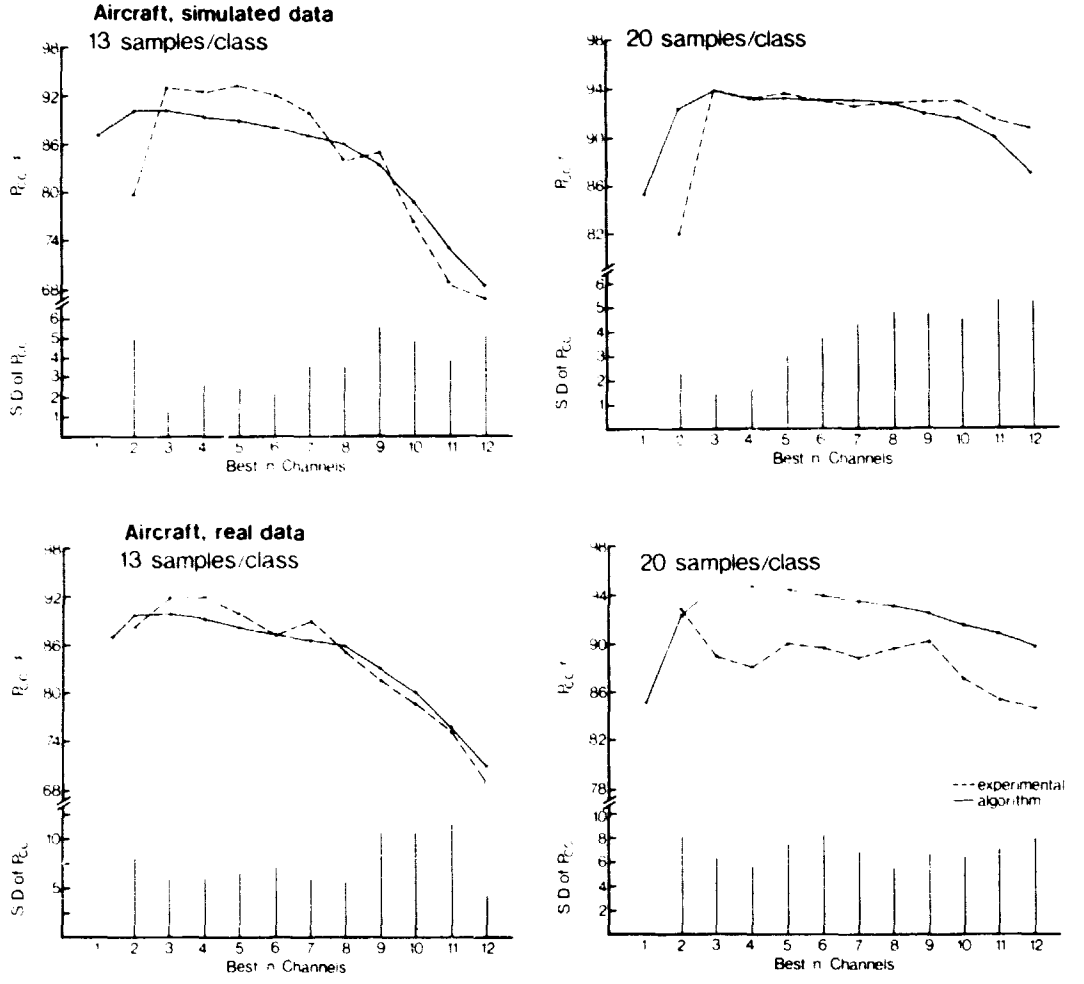


Figure 10-2. Experimental and algorithm results for simulated and real aircraft data with two sizes of training sets.

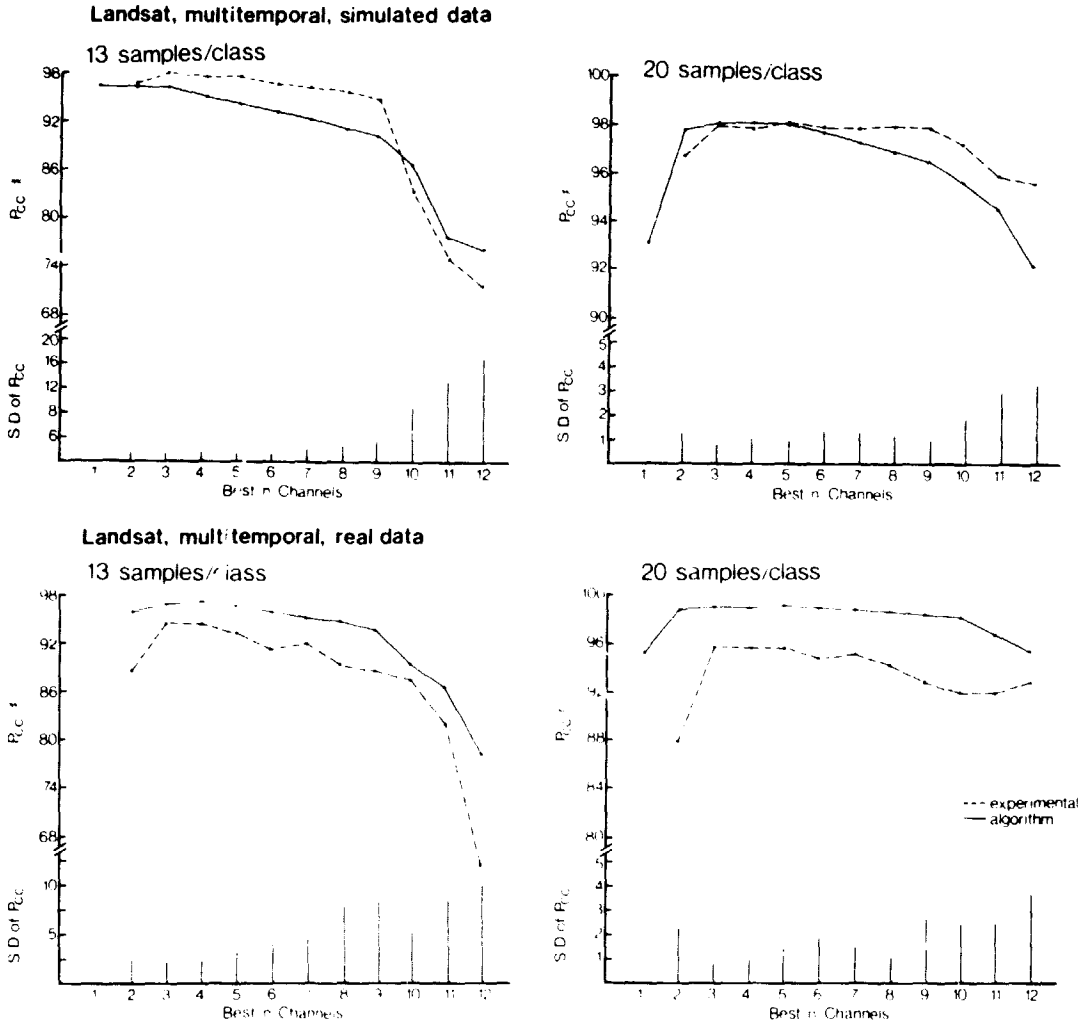


Figure 10-3. Experimental and algorithm results for simulated and real multitemporal Landsat data with two sizes of training sets.

The algorithm was then incorporated in a binary tree procedure using more than two classes. Two examples are presented: an aircraft data set with 9 spectral classes, and a Landsat data set with 3 spectral classes. Thirteen training samples per class were used, and classification results were compared to those obtained using single-stage classification for different dimensionalities.

Results are shown in Figures 10-4 and 10-5. The results indicate that the algorithm is effective in predicting the feature subsets that lead to the maximum or near-maximum accuracy possible using the K-L transformation for ordering the features. In both cases, the binary tree procedure resulted in about the maximum accuracy achieved using a single-stage classifier but with the added advantage that it provided a method for selecting those feature subsets which led to the maximum accuracy.

Conclusion

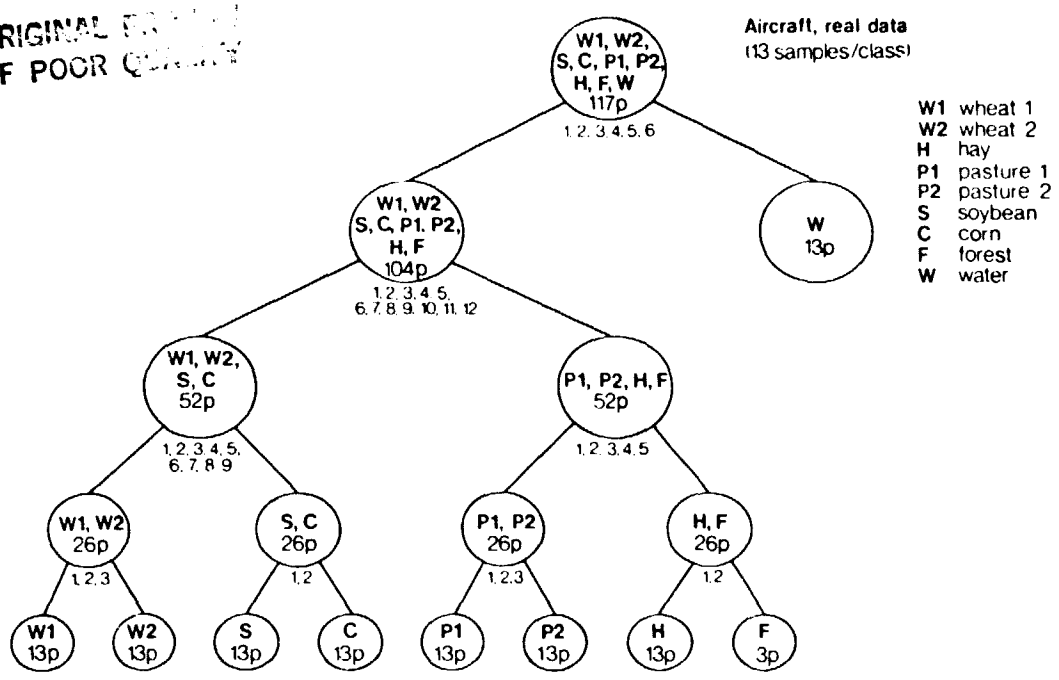
The proposed algorithm for feature selection appears to be effective in predicting the best feature subsets to use in the presence of a limited number of training samples. The algorithm is especially useful in a binary tree classification procedure, where it is shown to predict the best accuracy possible in a fairly complex data set (9 classes, 12 features).

The program provides the ability to use a small number of training samples and still get the best classification accuracy based on the available statistics. Indeed, to work with so few training samples and to achieve this accuracy with data and classes of this complexity is thought to be a significant new development in itself.

Finally, a significant factor in motivating the study of more sophisticated classification schemes such as multistage decision tree classifiers is to prepare for the more complex data sets to be available from future sensors. It is reasonable to speculate that the advantage of such classification procedures as those studied here may be even greater when more complex data sets become available.

Reference

1. Muasher, M.J. and D.A. Landgrebe 1981. Multistage Classification of Multispectral Earth Observational Data: The Design Approach. LARS Technical Report 101381, 185p. Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana.



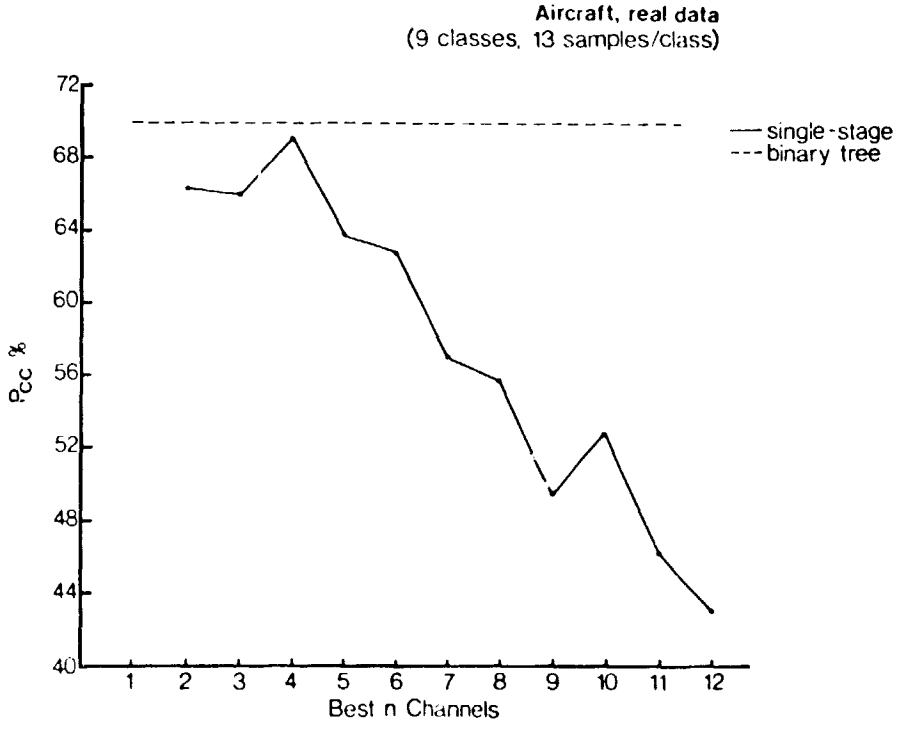
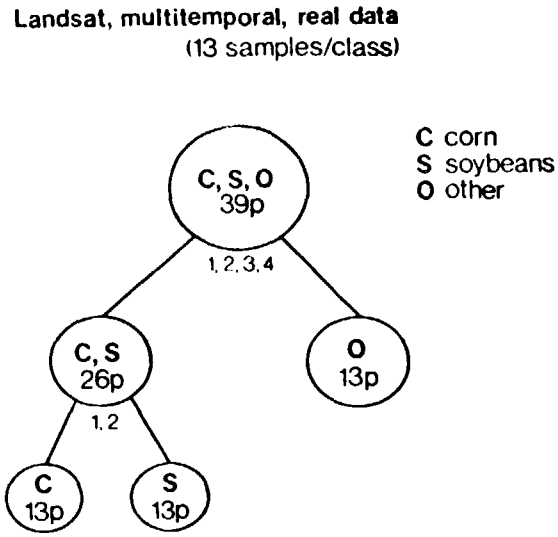


Figure 10-4. Single-stage and binary tree classification results for the aircraft data set.



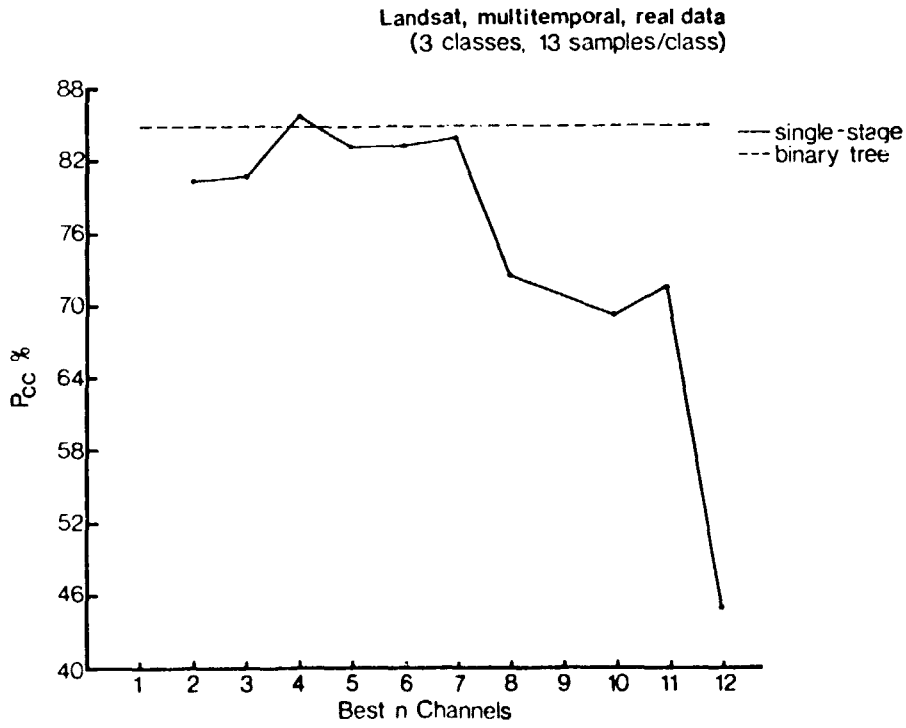


Figure 10-5. Single-stage and binary tree classification results for the Landsat data set.

11. THEORETICAL AND COMPUTATIONAL ADVANCES IN CONTEXTUAL CLASSIFICATION

J.C. Tilton, P.H. Swain, B.W. Smith and H.J. Siegel

Introduction

When multispectral scanner data are displayed in image form, a human analyst routinely uses spatial information to help decide what a particular pixel in the imagery might be. Using spatial information together with spectral information, the analyst may easily identify roads, delineate boundaries of agricultural fields, and differentiate between grass in an urban setting (e.g., lawns) and grass in an agricultural setting (e.g., pasture or forage crops) where a point-by-point classifier utilizing spectral information alone would have much difficulty in doing so.

A human analyst can use spatial information in several different ways roughly approximated by the structural, textural and contextual approaches taken in computer analysis of image data. We have pursued the contextual approach in this research. In the contextual approach, the probable classification of surrounding pixels influence the classification of a particular pixel. This approach exploits the tendency in nature for certain ground-cover classes to occur more frequently in some contexts than in others.

In this research we have developed a contextual algorithm for classifying multispectral scanner data. This algorithm exploits the spatial/spectral context of a pixel in making classification decisions. The effectiveness of the algorithm depends on accurate estimation of a statistical characterization of context, the context function. Practical application requires an intelligent implementation scheme with an eye towards reducing computation requirements and/or special processing hardware.

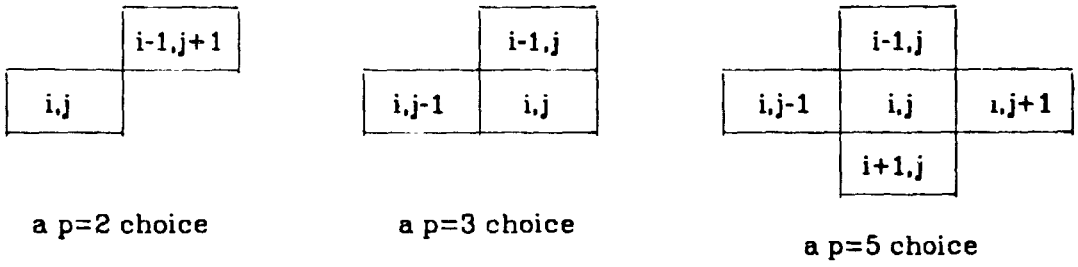
The next section briefly reviews the theoretical basis of the contextual classifier. We discuss the critical problem of estimating the context function in the following section. This discussion includes a summary of significant results from the first two years, as well as results from this year. Two methods, the ground-truth-guided method and the unbiased method, are offered as solutions to this estimation problem. After that we discuss approximate and hybird algorithms as methods for reducing computation requirements. Finally, we discuss implementation of the contextual classifier on a CDC Flexible Processor array.

Theoretical Basis of the Classifier

Consistent with the general characteristics of multispectral scanner data, we assume a two-dimensional array of $N=N_1 \times N_2$ random observations (pixels) X_{ij} having fixed but unknown classification ϑ_{ij}. The observation X_{ij} consists of n measurements from location (i,j) (usually containing spectral and/or temporal information), while the classification ϑ_{ij} can be any one of m spectral or information classes[1] from the set $\Omega = \{\omega_1, \omega_2, \ldots, \omega_m\}$.

[1]Spectral classes are spectrally discriminable subclasses of information classes (ground-cover classes of interest).

Define the spatial/spectral context of the pixel at location (i,j) as p-1 observations spatially near to, but not necessarily adjacent to, the observation X_{ij}. Call this arrangement of pixels together with X_{ij} the p-context array, several examples of which are shown in Figure 11-1. Group the p observations in the p-context array into a vector of observations $\underline{X}_{ij} = (X_1, X_2, \ldots, X_p)^T$ and let $\underline{\vartheta}_{ij}$ be the vector of true but unknown classifications associated with the observations in $\underline{X}_{ij}$. Let $\underline{\vartheta}^p \in \Omega^p$ be a vector of possible classifications for the elements of any p-context array. Correspondence of the components of $\underline{X}_{ij}$, $\underline{\vartheta}_{ij}$ and $\underline{\vartheta}^p$ to the positions in the p-context array is fixed but arbitrary except that the p^{th} component always corresponds to the pixel to be classified.



a p=2 choice a p=3 choice a p=5 choice

Figure 11-1. Examples of p-context arrays.

With these definitions, compound decision theory (with appropriate assumptions) leads to the following discriminant functions:

$$d(\underline{X}_{ij}) = \text{the action } a \text{ which maximizes}$$

$$\sum_{\substack{\underline{\vartheta}^p \in \Omega^p, \\ \vartheta_p = a}} C(\underline{\vartheta}^p) \prod_{k=1}^{p} f(X_k \mid \vartheta_k) \qquad (1)$$

where $C(\underline{\vartheta}^p)$, the "context function," is the relative frequency with which $\underline{\vartheta}^p$ occurs in the scene being analyzed. More detailed derivations of this set of discriminant functions can be found in the references [1,2,3,4].

The optimal form of $d(\cdot)$ cannot be found in practice since it depends on the unknown functions $C(\underline{\vartheta}^p)$ and $f(X_k \mid \vartheta_k)$. Methods for estimating $f(X_k \mid \vartheta_k)$ are well established from considerable experience in using the conventional noncontextual maximum likelihood decision rule [5]. During the past three years, we have studied methods for estimating the context function, $C(\underline{\vartheta}^p)$. The next section summarizes this research and offers two methods for estimating $C(\underline{\vartheta}^p)$.

Context Function Estimation

If we knew the true classification of our multispectral scanner data, we could find the context function exactly by simple tabulation from the true classification. Since the true classification is not known, we must estimate the context function from the data or, more practically, from representative portions of the data set

designated as a training set. This training set must consist of blocks of data which are at least as large as the chosen p-context array.

A straightforward way to estimate the context function is to perform a non-contextual maximum likelihood classificatior of the training sample and use the relative frequency distribution tabulated from this classification as an estimate of the context function. This "classify-and-count" method was tested on both simulated and real data [1,2,3]. While the method produced excellent results with simulated data, it gave rather poor results with real data, stimulating a search for alternative methods for estimating the context function.

Ground-Truth-Guided Method

The first satisfactory method studied we call the "ground-truth-guided" method [1,2,3,4]. In this method the non-contextual classification is constrained by ground-truth data to produce 100% correct information class classifications of the training sample. The context function is then tabulated from the training sample classifications. This ground-truth-guided method can be formulated to produce estimates of the context function in terms of either spectral classes or information classes [2,4].

Both the spectral and information class formulations of the ground-truth-guided method were tested on two 50-pixel-square Landsat data sets. One data set was a LACIE data set from Hodgeman County, Kansas, containing pasture, wheat, corn and fallow fields. The other data set was from Tippecanoe County, Indiana, containing residential and commercial areas in northern Lafayette and West Lafayette, as well as areas of forest, agriculture and water. For both data sets, the restricted spectral class classification was performed over the first 25 lines of the data set and the context function was estimated over those 25 lines. Contextual classifications of the scenes were performed and classification accuracies were evaluated over the last 25 lines as well as over the entire data set.

Tables 11-1 and 11-2 present the results from contextual classifications using four-nearest-neighbor (4nn) estimates of the context function (the p=5 choice in Figure 11-1) for both the spectral and information class formulations of the ground-truth-guided method (gtgm). These results are also compared to the accuracies obtained from uniform-priors and estimated-priors non-contextual maximum likelihood classifications. The prior probabilities for the estimated-priors non-contextual classifications were estimated by tabulation from the uniform-priors non-contextual classification. These results show that contextual classifications using the ground-truth-guided method for estimating the context function give noticeably better results than non-contextual classifications on these data sets. For these cases, the spectral class formulation of the ground-truth-guided method generally produces higher classification accuracies. However, since the spectral class estimate of the context function has substantially more non-zero elements than the information class estimate, contextual classifications using the spectral class formulation generally take over twice the computer time required for the information class formulation.

We should caution here that in these two test cases, fully one-half of the data set was used as the training set. In practical classification problems with larger

Table 11-1. Comparison of the contextual classifier using the ground-truth-guided method with non-contextual classifiers; Hodgeman County, Kansas, Landsat Data Set.

Classification	% Accuracy lines 26-50 Overall	lines 26-50 Average-by-Class	lines 1-50 Overall	lines 1-50 Average-by-Class
uniform priors	81.5	78.2	82.5	74.3
estimated priors	82.2	78.3	82.8	74.1
4nn gtgm, spectral	85.4	81.6	85.7	77.3
4nn gtgm, information	85.3	81.4	85.0	76.0

Table 11-2. Comparison of the contextual classifier using the ground-truth-guided method with non-contextual classifiers; Tippecanoe County, Indiana, Landsat Data Set.

Classification	% Accuracy lines 26-50 Overall	lines 26-50 Average-by-Class	lines 1-50 Overall	lines 1-50 Average-by-Class
uniform priors	82.7	81.7	81.8	83.4
estimated priors	84.2	82.0	83.7	83.7
4nn gtgm, spectral	88.7	91.1	89.3	90.7
4nn gtgm, information	88.2	87.3	88.2	86.2

data sets, ground-truth data is usually available for only ten percent or less (often less than one percent) of the data set. We expect that this smaller percentage of ground-truth data will decrease the effectiveness of the ground-truth-guided method. Also, contrary to what was observed here, we anticipate that the information-class formulation may perform better on larger data sets than the spectral-class formulation, since we have noted that the information-class formulation is less sensitive to estimation errors than is the spectral-class formulation [2].

While the ground-truth-guided method can produce good estimates of the context function, it suffers the limitation that it requires large areas of spatially contiguous ground-truth data. When such detailed ground-truth data are not available, some other method is needed, such as the unbiased estimator described next. This unbiased estimator does not directly require any ground-truth data. It does indirectly require ground-truth data in that it requires estimates of the class-conditional density functions, $f(X_k \mid \vartheta_k)$. The ground-truth data required for making estimates of the class-conditional density functions is minimal when compared to the ground-truth data required by the ground-truth-guided method.

The Unbiased Method

A large portion of the research effort in the second and third years of this project was directed towards developing a satisfactory method for estimating the context function that did not depend on ground truth availability. Many different tactics were attempted and reported on earlier [2,3,4,6,7]. The most satisfactory method was developed after it was noted that a problem with the "classify-and-count" method is that it produces biased estimates of the context function [2,4]. The "unbiased method" developed during the third year of this project was developed specifically to eliminate this bias in the context function estimate.

The unbiased estimator utilized by our unbiased method is presented in the statistical literature by Van Ryzin [8] and Hannon, *et al* [9]. The details of how this unbiased estimator is adapted to estimate the context function are presented in the references [2,4]. For the p=1, m class case, the unbiased estimate of the context-function vector $\underline{G} = (G(\omega_1), G(\omega_2), \ldots, G(\omega_m))^T$ can be found by solving

$$\underline{G} \triangleq J^{-1} \underline{h} \; . \tag{2}$$

The vector $\underline{h}$ is made up of the elements $\overline{h}_k(\underline{X})$, $k=1,2,\ldots,m$ defined by:

$$\overline{h}_k(\underline{X}) \triangleq \frac{(2\pi)^{\frac{n}{2}}}{N} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} f(X_{ij} \mid \omega_k) \tag{3}$$

where N, N_1 and N_2 are as defined in Section 11.1, n is the dimensionality of X_{ij} and $f(X_{ij} \mid \omega_k)$, $k=1,2,\ldots,m$, are the class-conditional densities used in equation (1). J is and $m \times m$ matrix with entries of the form

$$J_{kl} = (2\pi)^{\frac{n}{2}} \int f(X \mid \omega_k) f(X \mid \omega_l) \, dX \tag{4}$$

Assuming the ω_k are normally distributed spectral classes with respective mean vectors μ_k and covariance matrices Σ_k $(k=1,2,\ldots,m)$, we find

$$J_{kl} = \left[\det(\Sigma_k + \Sigma_l)\right]^{-\frac{1}{2}} \exp\{-\frac{1}{2}(\mu_k - \mu_l)^T (\Sigma_k + \Sigma_l)^{-1} (\mu_k - \mu_l)\} \; . \tag{5}$$

For an explicit derivation of this result see Miller [10]. When the ω_k are information classes, the J_{kl} are weighted sums of terms of the form given in (5). When the estimate is made in terms of information classes, estimates must be made of the weights used to form the weighted sum of the class-conditional normal densities of the spectral subclasses. For each information class, the weights are estimated by using the unbiased estimator with p=1 for the spectral classes which make up the information class being considered.

The unbiased estimator can be extended straightforwardly to the arbitrary p-context array case [2,4]. However, for values of p larger than about 2 or 3, the unbiased estimator must be implemented intelligently so as to minimize the amount of computer storage needed. The unbiased estimator can also be implemented adaptively, with the context-function estimate varying according to local context information. These implementation considerations are discussed in the references [2,4].

Table 11-3 presents the accuracies resulting from contextual classifications for three Landsat data sets using four-nearest-neighbor (4nn) estimates of the

context function. The results using the spectral-class formulation are shown for the whole scene (non-adaptive) version and for an adaptive version employing local context-function estimates for 25×25 pixel blocks made from the same 25×25 pixel block. The results using the information-class formulation are shown for an adaptive version employing estimates for various $n_1 \times n_2$ pixel blocks made from a $m_1 \times m_2$ pixel block centered on each $n_1 \times n_2$ pixel block. The uniform-priors non-contextual classification results are given for reference.

Table 11-3. Comparison of the contextual classifier using various unbiased estimator formulations and the uniform-priors non-contextual classifier.

Data Set	Classification	%Accuracy Overall	%Accuracy Average-by-Class
Hodgeman County, Kansas, 50-pixel-square Landsat (evaluated over lines and columns 6 through 50)	uniform-priors non-contextual	82.0	75.9
	4nn unbiased, spectral class whole image est. (nonadaptive)	83.1	75.8
	4nn unbiased, spectral class adaptive est., 25×25 from 25×25	84.0	77.8
	4nn unbiased, information class adaptive est., 25×25 from 35×35	84.0	78.0
Monroe County, Indiana, 50-pixel-square Landsat	uniform-priors non-contextual	83.1	82.7
	4nn unbiased, spectral class whole image est. (nonadaptive)	84.4	84.4
	4nn unbiased, spectral class adaptive est., 25×25 from 25×25	84.3	83.9
	4nn unbiased, information class adaptive est., 17×17 from 25×25	88.9	88.3
Tippecanoe County, Indiana, 50-pixel-square Landsat	uniform-priors non-contextual	81.8	83.4
	4nn unbiased, spectral class whole image est. (nonadaptive)	86.2	87.9
	4nn unbiased, spectral class adaptive est., 25×25 from 25×25	86.7	88.1
	4nn unbiased, information class adaptive est., 25×25 from 25×25	86.2	89.1
	4nn unbiased, information class adaptive est., 10×10 from 20×20	86.9	89.7

Figure 11-2 shows computer generated gray-scale maps of classifications of the Tippecanoe County, Indiana, Landsat data set. The contextual classification looks visually closer to the reference classification than might be expected based on the accuracy improvement over the non-contextual classifications. This is due to the tendency of the contextual information to provide a smoothing effect, making classification maps that are not only more accurate, but also more pleasing to the eye. This smoothing effect will not necessarily occur on all data sets. There is nothing inherent in the contextual classification algorithm that would force smoothing when none is called for. The smoothing effect should only occur when the contextual information so dictates.

The adaptive information-class formulation performs as well as or better than any other formulation shown. As noted earlier in the discussion of the ground-truth-guided method, the information-class formulation has the further advantage of having substantially fewer non-zero elements in the context function estimate, causing contextual classifications using an information-class formulation to require less than half the computer time required for contextual classifications using a corresponding spectral class formulation. Contrary to the ground-truth-guided method, we do not expect any degradation of performance for larger data sets.

Reducing Computation Requirements

The contextual classifier is very computationally intensive, typically requiring a large amount of computer time. One possible approach to reducing execution time would be to take advantage of special computer architectures involving multiple processing elements. We will discuss such an approach in the next section. In this section we instead explore two approximate implementations of the contextual classifier. In the "approximate algorithm" we modify the contextual decision function (equation (1)) to effect computation savings. In the "hybird algorithm" we selectively apply the contextual classifier, resorting to a non-contextual classifier whenever that classifier can "confidently" classify a given pixel.

Approximate Algorithm

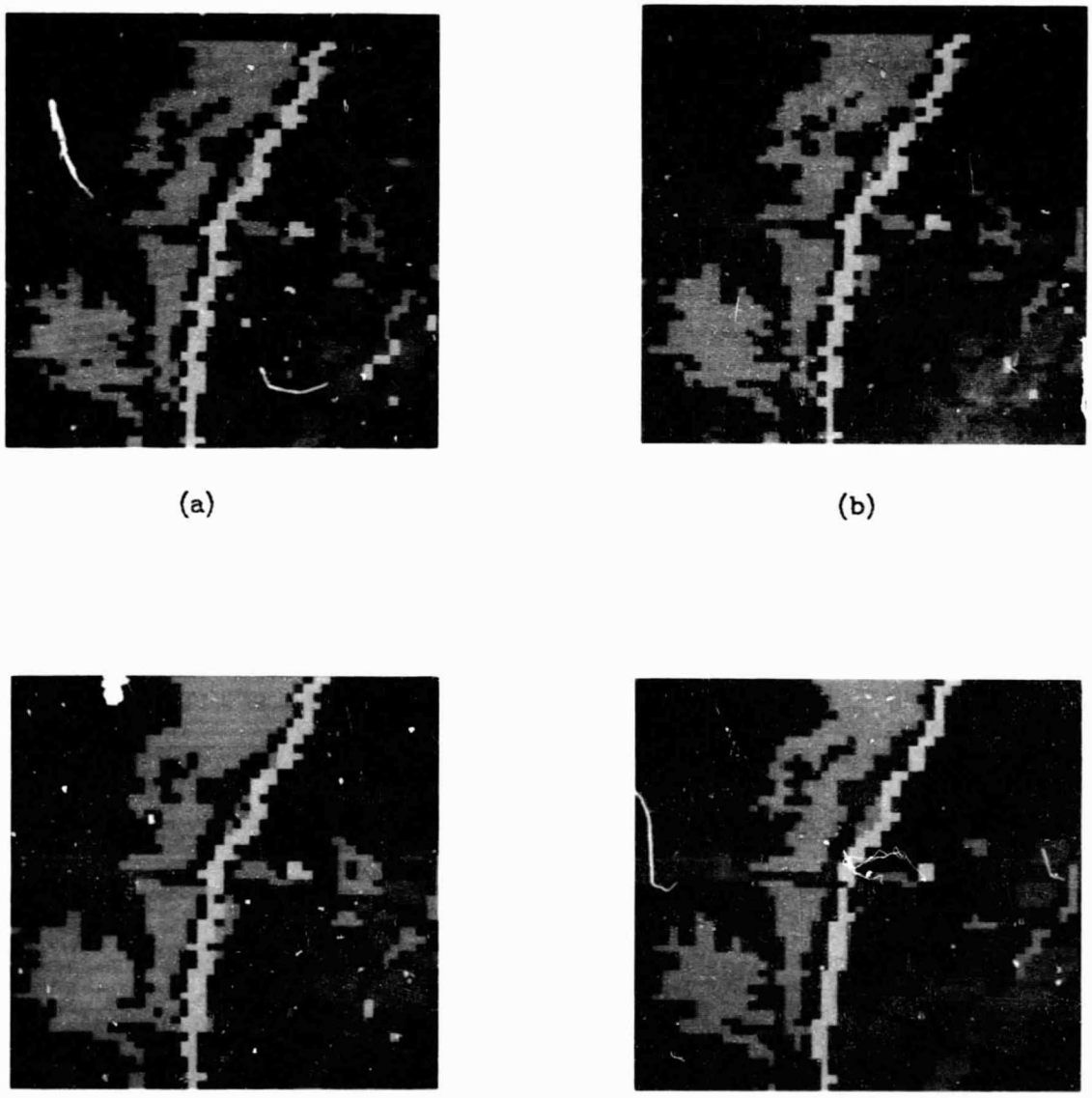
In studying the original algorithm, it was discovered that a single term of the sum in equation (1) usually dominated all other terms for each possible classification action a [2,11]. This observation suggested the following approximation to the decision rule $d(\cdot)$ given in equation (1):

$$d(\underline{X}_{ij}) = \text{the action } a \text{ which maximizes, for all } \underline{\vartheta}^p \in \Omega^p \text{ with } \vartheta_p = a,$$

$$G(\underline{\vartheta}^p) \prod_{k=1}^{p} f(X_k \mid \vartheta_k). \tag{6}$$

Comparing equations (1) and (6), we see that the summation in equation (1) is eliminated. This not only eliminates the computation required to perform the summation, but also allows the use of a discriminant function proportional to (6) which is much easier to calculate.

The approximate algorithm was compared with the original algorithm in terms of classification accuracy and computer timings on a 50-pixel-square



(a) (b)

(c) (d)

Figure 11-2. Visual comparison of classification results, Tippecanoe County, Indiana, Landsat data set. (a) uniform-priors non-contextual, (b) estimated-priors non-contextual, and (c) four-nearest-neighbor adaptive (17×17 from 27×27) unbiased estimator (d) reference classification.

simulated data set and on two 50-pixel-square Landsat data sets [2,11]; the LACIE data set mentioned earlier, and a data set containing approximately equal amounts of urban and agricultural area located to the southeast of Bloomington, Indiana in Monroe County. In terms of accuracy, the approximate algorithm performed very well when compared to the original algorithm, with overall accuracy differing by less than 0.2 percent for all data sets. The results also show that the approximate algorithm averaged less than half the computation time taken by the original algorithm.

Hybrid Algorithm

A second way to produce classifications with accuracy comparable to the original contextual classification algorthm but with less computation is to use a "hybrid" algorithm which employs a non-contextual maximum likelihood classifier whenever that classifier can classify a given pixel "confidently," resorting to the contextual classifier only on "difficult" pixels [2]. In other words, when the multispectral information alone at a given pixel is adequate to confidently classify the pixel, the contextual information is not used. Such a method should save considerable computation time, depending on the percentage of pixels that must be classified by the contextual classifier. Classification accuracy should not suffer significantly because the pixels classified "confidently" by the uniform-priors non-contextual classifier presumably would have been classified identically by the contextual classifier.

Two simple measures of the "confidence" of classification by a uniform-priors non-contextual classifier were tested this past year [2]. The results indicate that these simple confidence measures are not adequate indicators of pixels that would be classified differently by the non-contextual and contextual classifiers. It is apparent that a more sophisticated approach is needed. Such an approach would take into account the location of each measurement in the measurement space in relation to the multidimensional contours of the class-conditional density functions. A confidence (or reliability) measure of this type is suggested in Alvo and Goldberg [12]. Time did not permit further effort in this research area.

Contextual Classification on a CDC Flexible Processor System

As noted in the previous section, the contextual classifier (and even much simpler algorithms used for remote sensing data analysis) typically require large amounts of computation time. One way to reduce the execution time of these tasks is through the use of parallelism. Various parallel processing systems have been built or proposed. The CDC Flexible Processor system is a commercially available multiprocessor system which has been recommended for use in remote sensing [3,13,14].

Uniprocessor and multiprocessor implementations of the contextual classifier based on the one-by-three linear and the three-by-three non-linear neighborhood are discussed in [3]. In [3], the timings for the Flexible Processor implementation of the one-by-three linear case are presented. The Flexible Processor (FP) and the array are covered in depth in [3,13,15]. In this report, FP timing results for both the one-by-three and the three-by-three neighborhoods are given and compared to a PDP-11/70 implementation.

Linear Contextual Classifiers

Consider using an FP system to implement the contextual classifier based on a horizontally linear neighborhood of size three (Figure 11-3). Divide the A-by-B image into subimages of B/N rows A pixels long, as shown in Figure 11-4. This method of dividing the image is called "striping" and is discussed in more detail in [3]. Assign each subimage to a different FP. The entire neighborhood of each pixel is included in its subimage. No interaction between FPs is needed, i.e., each FP can process its subimage independently. A perfect factor of N improvement speedup over a single FP occurs if B is a multiple of N. The minor degradation in performance when B is not a multiple of N is discussed in [3, 15, 16].

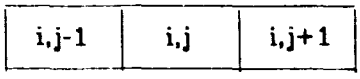


Figure 11-3. Size three linear neighborhood for pixel (i,j).

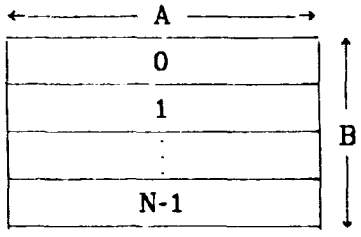


Figure 11-4. An A-by-B image divided among N Flexible Processors.

An FP micro-assembly language version of the algorithm stated in [15] was coded and debugged and is on file at LARS. Because each FP is microprogrammable, determining program correctness and analyzing the execution time is done through the use of the micro-assembler and simulator [15]. Execution times per pixel vary because all floating point operations are done in the software. For example, a floating point addition may take anywhere from 5 to 50 machine cycles depending on the particular (non-zero) numbers being added. This represents a cost differential of 10:1. The final timings are, thus, severely dependent on the a priori probabilities and the image data itself. In order to partially overcome this problem, the timings for the contextual classifier were obtained from a sample image consisting of over 400 pixels.

The pixel measurement vectors, covariance matrices, logarithms of the determinants of the covariance matrices, a priori probabilities, and temporary variable array are all stored in the large file [3,13,15]. In this way, each FP has all the information it needs for performing the classification on its subimage.

For many possible $\underline{v}^p$ in Ω^p, the value of the context function, $C(\underline{v}^p)$, may be zero or effectively zero. If the number of the non-zero values of the context function is small (less than 50%), and if the classification values in the p-context array, $\underline{v}^p$, associated with each value of the context function can be stored in the space of one floating point number (32 bits), then any algorithm that stores all a priori probabilities will waste space [3]. This is the case in the Landsat data used for this experiment. Each non-zero value of $C(\underline{v}^p)$ is stored as two 32-bit quantities. The first 32-bit quantity contains information about the class of each pixel within the p-context array. For example, if G(3,3,2) is non-zero, the word preceding it is a representation (catenation) of 3,3, and 2. This allows $\left\lfloor \frac{32}{p} \right\rfloor$ bits per class, i.e., up to $2^{\left\lfloor \frac{32}{p} \right\rfloor}$ classes. The second 32 bits is the value of $C(\underline{v}^p)$ itself. Only the non-zero $C(\underline{v}^p)$ values are stored, so only the non-zero $C(\underline{v}^p)$ values affect the computation time.

For the purposes of testing the FP implementation of the one-by-three linear contextual classifier program, measurement vectors from 30 rows of 16 pixels were classified. Each measurement vector consisted of four 32-bit floating point representations of 8-bit integers. All data was stored in the large file. The data set consisted of a four-class subset of the data used in [1]. To provide a basis for comparison, a similar contextual classifier was run on a PDP-11/70 over the same test data. It was found that lack of exponent range in the 11/70 floating point hardware required extra handling. FP floating point algorithms are implemented in the software, so a 14-bit exponent was used to overcome this problem. A description of the floating point software is available in [15]. The "e" calculations are based on those in [17] and are on file at LARS. Twenty non-zero $C(\underline{v}^p)$s were chosen for the benchmark tests. Running under the above constraints, the single FP classifier took .035 secs./pixel, while the PDP-11/70 required .050 secs./pixel.

Using .05 secs. per pixel as the PDP processing time and .035 secs. per pixel as the single FP processing time, a 16 FP configuration would perform contextual classifications at a rate of 457 pixels per sec., as opposed to 20 pixels per sec. for a single PDP-11/70. There are, of course, cost differences between these two systems; however, the purpose here is to show the gains made possible by a multiprocessor system.

Non-linear Contextual Classifiers

Next, consider the three-by-three non-linear neighborhood. First, the single FP timings are considered, then the timings for an array of N FPs are considered.

The eight-nearest neighbor contextual classifier is similar to the previously described linear case. Differences arise in the calculation of the discriminant function, the method of updating the data for a given window, and the method of data storage. The general approach for this case is discussed in [3].

Timings run from Landsat data from [1] show that, on the average, the FP implementation of the four-class, size nine square neighborhood contextual classifier with all data entries and a priori information stored in the large file requires .137 secs./pixel. A PDP-11/70 implementation of the same algorithm

requires .34 secs./pixel. Tests for the 11/70 were run with 50 non-zero $C(\underline{v}^p)$s and 4 spectral classes on 52 lines of 16 pixels. A 30-line-by-16 pixel subset of the above image was used to derive the FP timings for a 52-line image. Pixels on the top and bottom lines (and right and left columns) of an image are not classified, and thus do not appear in the number of classified pixels. As a result, for the first and last rows (and first and last columns) of an image, the classifier must calculate the class conditional probabilities for these pixels without ever classifying them. Only the non-zero $C(\underline{v}^p)$s are stored, so only the non-zero $C(\underline{v}^p)$s affect computation time.

Using .34 secs. per pixel as the PDP processing time and .137 secs. per pixel as the single FP processing time, a 16 FP array would perform contextual classifications at a rate of approximately 116 pixels per sec., as opposed to the 6 pixels per sec. rate of a single PDP-11/70. This assumes, however, that all needed data is stored in the large file, a somewhat unrealistic assumption. The use of the bulk memories for storing and sharing data is discussed in the next two sections.

Processing of Images With Large Numbers of non-zero $C(\underline{v}^p)$s

If the number of non-zero $C(\underline{v}^p)$s is too large for all of them to fit in the register files, bulk memory can be used to store the overflow $C(\underline{v}^p)$s. The width of the bulk memory is 16 bits. Each $C(\underline{v}^p)$ is composed of two 32-bit quantities; one is a concantentation of the classes associated with the p-context array, $\underline{v}^p$, and the other is the $C(\underline{v}^p)$ value itself [5]. Thus, the 64-bit $C(\underline{v}^p)$ entry can be accessed with four reads from bulk memory. One of the special features associated with an FP is that every time a read from bulk memory is performed, the pointer to bulk memory is automatically incremented [3]. A read from bulk memory is accomplished in two steps [3,13]. First the read must be initialized and second (after $.250\mu-secs.$) the data must be read from the bulk memory (Z) register [3,13]. On the surface, it would appear that a 16 bit read requires four clock cycles; however, this is not the case. The read can be initialized in parallel with other operations, thus no time is lost due to the initialization. An FP can wait for the data or it can execute other instructions in the mean-time. Thus, the total cost of a read from bulk memory is one machine cycle per 16 bits. The cost, then, of accessing a $C(\underline{v}^p)$ entry from the bulk memory is $.500\mu-secs$. To perform the corresponding operation from the large file requires $.250\mu-secs$.

As an example, use the benchmark 8-nearest neighbor non-linear p-context array. Allow all 50 of the $C(\underline{v}^p)$ entries to be stored in bulk memory. The total time spent accessing the $C(\underline{v}^p)$ entries is:

$$.500\frac{\mu-secs}{C(\underline{v}^p)entry} \times \frac{50\ non-zero\ C(\underline{v}^p)entries}{pixel} = 25\frac{\mu-secs.}{pixel}$$

Only half of this time; however, represents additional processing time. It was earlier stated that $.250\mu-secs./C(\underline{v}^p)$ entry was required to fetch the $C(\underline{v}^p)$ entry from the large file. Thus, the additional processing time required to process a $C(\underline{v}^p)$ entry stored in bulk memory is:

$$.250\frac{\mu-secs}{C(\underline{v}^p)entry} \times \frac{50\ non-zero\ C(\underline{v}^p)entries}{pixel} = 12.5\frac{\mu-secs.}{pixel}$$

When this is compared to the 137,000 $\mu-secs./pixel$ required for classification, this time represents a negligible cost.

Processing of Images in Bulk Memory

If an image is small, data vectors may be stored in the large file. This was the method used to acquire the timings presented. For actual images, however, the large file is too small to hold the image data. Pixel measurement vectors can be stored in bulk memory. There is, however, an additional cost associated with reading pixel measurement vectors from bulk memory. Pixel data is represented as a one-by-four vector of 32-bit floating point numbers. It was earlier stated that a 16 bit read from bulk memory requires the same amount of time as a 32-bit read from the large file. Thus, reading a 32-bit number from bulk memory will require twice as much time as a corresponding read from the large file. Reading a data vector from the large file will require four reads, or $.5\mu-secs./pixel$ = [4 $cycles\ x\ .125\mu-secs./cycle$]. Reading the same data from bulk memory will require $1\mu-secs./pixel$, or an additional processing time of $.5\mu-secs./pixel$. This is minimal when compared with the $137{,}000\mu-secs./pixel$ processing time associated with the eight nearest-neighbor contextual classifier.

A 16 FP System

Consider the problem of using N (<=16) FPs together to do contextual classification with a square size 9 neighborhood. Assume the image data is stored in the bulk memories. The approach taken is to divide the image among the FPs using the "striping" method (Fig 11-4). Each FP classifies the pixels in its own subimage. Because the p-array is non-linear, FPs will have to communicate to share subimage edge data [3]. Thus, some way to achieve this sharing is necessary.

The speed at which the contextual classifier runs depends on the floating point algorithms which are implemented in the software. This can cause a bottleneck in the processing if one FP is required to wait for another. Synchronization can require large amounts of time if the full 16 processor array is used, so asynchronous processing at the instruction level is necessary.

An FP is capable of addressing up to three channels of 16-by-128K bytes of bulk memory each [3,13]. The sharing of bulk memory is a scheme that can be used for shared data. One possible implementation is shown in Figure 11-5. Assume each FP will classify the pixels in B/N rows (Figure 11-4). If border areas are stored in the shared memory banks, a processor will begin processing in banks of bus 1. Processing will continue through half the banks in bus 1 to bank 0 on bus 2. After all the data in the banks on data bus 2 have been processed, processing will continue to the banks on bus 3.

Allowing 25% of FP i's data to be stored in the shared banks on bus 1, 50% of the data to be stored in the local banks on bus 2, and 25% of the data to be stored in the shared banks on bus 3, no contention will occur. Consider that for processor i to "catch up" with processor i+1, processor i will have to process more than 75% of its data in the time that it takes processor i+1 to process 25% of its data. Contention is not a problem. Details of this are in [3].

An FP will be allowed to address only half of its memory banks at one time. This is done to facilitate double buffering. The other half of the memory banks will be accessible by the host. This allows, for example, the FP to be classifying the current image while the host unloads and stores the results of the previous classification and then loads the next image to be processed.

Processing of Large Images

Assume that an FP is configured as previously described. If the image to be processed will fit into bulk memory, the image can be processed according to the "striping scheme" discussed earlier. There is, however, another problem that can arise. An image may be too large to fit in the bulk memory.

Assume that there are L bulk memory banks per FP for data, separate from the bulk memory banks for the $C(\underline{v}^p)$s. If an image will not fit into the L bulk memory banks, the image should be divided according to the "striping" scheme discussed earlier (Figure 11-4); however, the host will transmit only portions of the image that will fit into L/2 bulk memory banks per FP. Figure 11-6 is an example of how an image is divided and processed. The FPs process subimages from right to left. Each subimage is processed in a zig-zag fashion. The class-conditional densities of the pixel measurement vectors for the leftmost two columns of data are saved, as they are needed to process the next subimage. The exception to this rule is the last subimage. While the FP is processing one subimage in one half of its memory, the host can be loading the next subimage into the other half of the bulk memory. This will overlap the FP operation with the host's operation.

Since the floating point operations are done in software, an FP processing its portion of the image may finish before the rest of the processors. With the FPs running asynchronously, it is theoretically possible for a given FP to eventually get two subimages ahead of its neighboring FPs. Subimage edge data would be destroyed for the neighboring FPs if the host were to load new data into the shared memory banks before the neighboring two FPs had finished with the old data. To prevent this from happening, after an FP processes two subimages, it must wait for the other FPs to finish. As was stated above, the possibility for the destruction of data is quite minimal.

When an FP finishes writing results into a bank of bulk memory, it signals the host to read all necessary data from that memory bank, even though it will need to read data corresponding to the subimage edge pixels from that bulk memory bank to process the next subimage. Since a read is non-destructive, the host reading from bulk memory will not hamper an FP reading from the same bulk memory bank. All FPs accessing a given bulk memory bank must set flags in bulk memory before the host can write to this bank. This will prevent the host from overwriting data that is still in use.

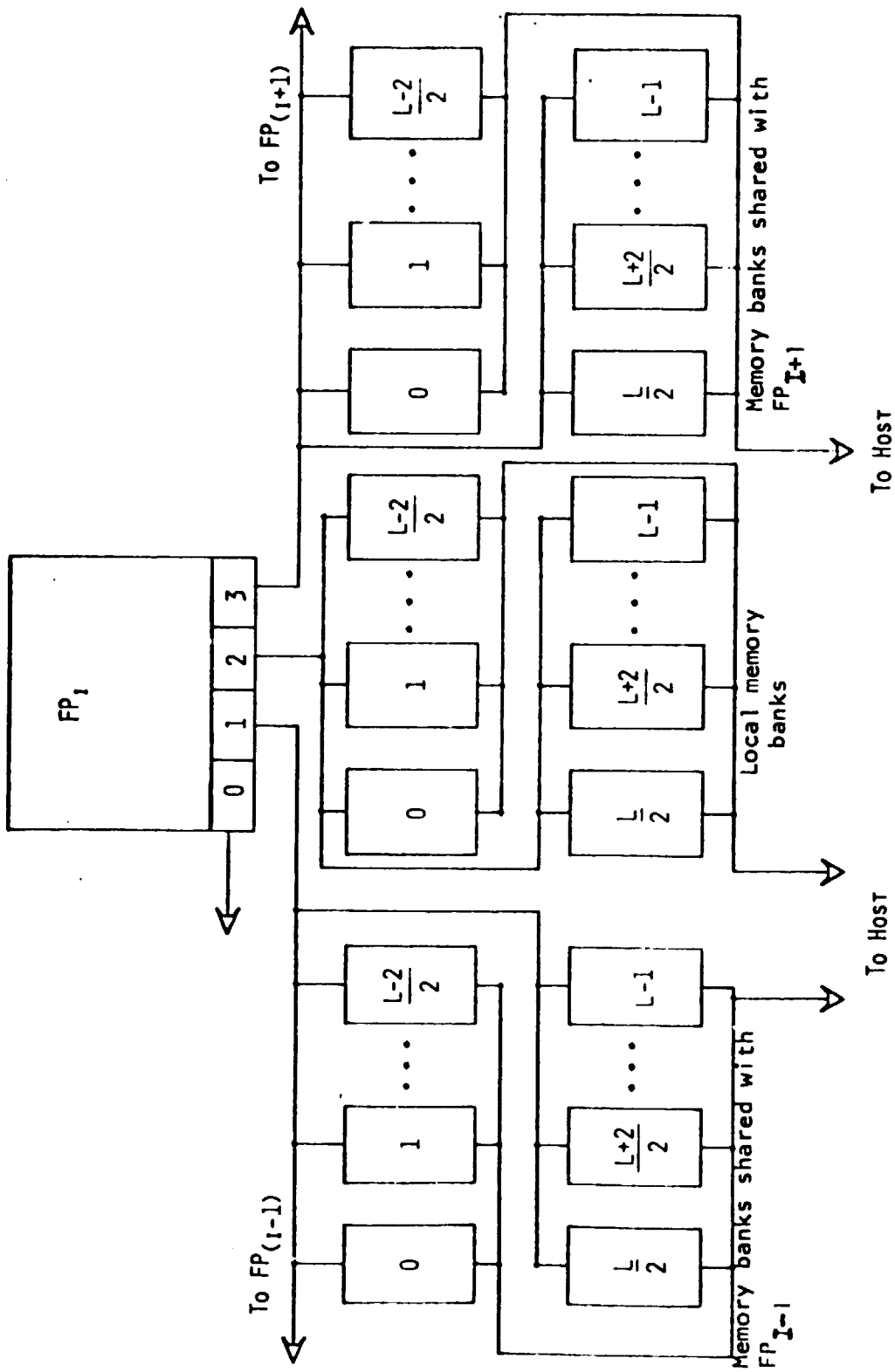
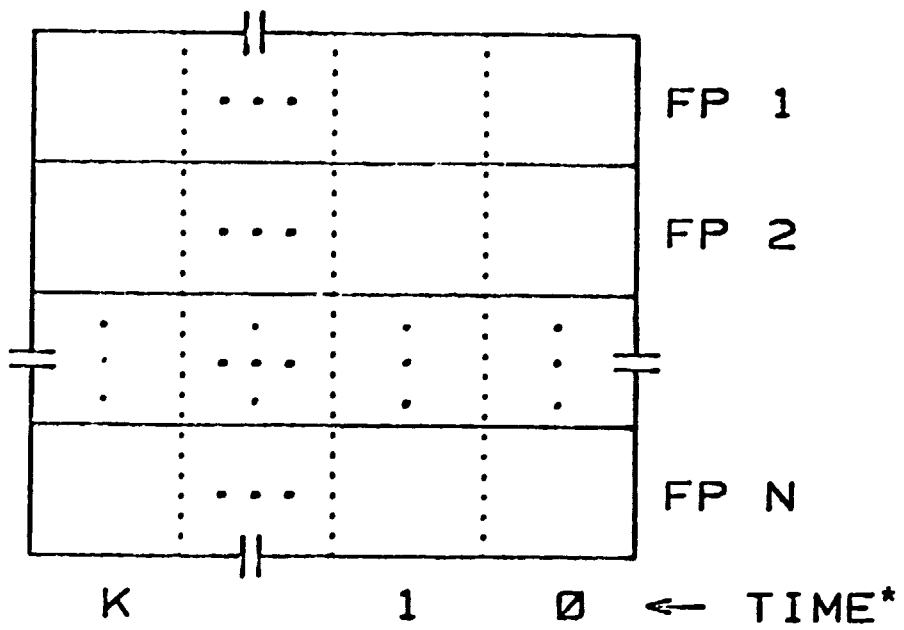


Figure 11-5. Potential memory organization for striping scheme.



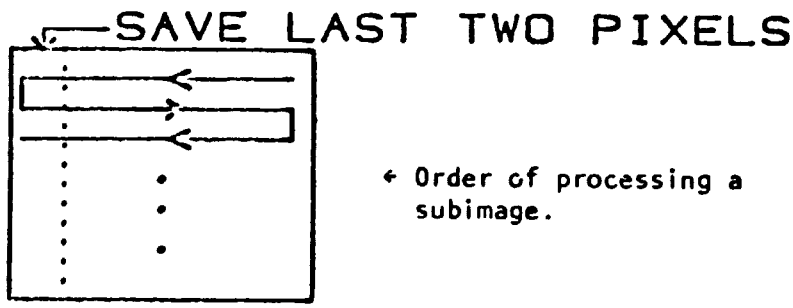


Figure 11-6. Order of processing of a large image that will not fit into bulk memory. (*-Dotted lines indicate the amount of image that will fit into bulk memory. FPs process from right to left.)

The organization of the FP system given above will allow contention-free sharing of data. This means that N FPs will be able to operate approximately N times faster than one FP. Furthermore, the double-buffering of the bulk memories will allow the loading of images to be processed and storage of results to be overlapped with the classification operation of the FPs.

Summary

In this research we have developed a general contextual algorithm for classifying multispectral scanner data and have explored efficient implementations of this algorithm. This contextual classifier depends only on general contextual information, and can, in principle, be used to advantage with any multispectral image data set. In the limited tests performed to date, the contextual classifier has consistently produced classifications that are noticeably more accurate than those produced by a non-contextual maximum likelihood classifier.

A key input to the contextual classifier is the context function (a quantitative characterization of context). Previous to this year, we had developed the "ground-truth-guided" method for estimating the context function. During this past year we developed the "unbiased" method for estimating the context cases for cases where the substantial ground-truth data required by the ground-truth-guided method are not available. Besides having minimal ground truth requirements, this unbiased method can be straightforwardly formulated in an adaptive mode where the context-function estimate varies according to local contextual information.

The contextual classifier is very computationally intensive as originally formulated. "Approximate" and "hybrid" algorithms were explored as efficient implementations of the contextual classifier on serial computers. In addition, parallel implementations on a CDC Flexible Processor system was explored. Here implemetations using N FPs in parallel were described for one-by-three and three-by-three neighborhoods and timing results were compared with a PDP-11/70 implementation. Cases with large numbers of non-zero context funtion values were considered and the additional processing requirement for reading bulk memory was found to be minimal. An organization of FP systems which allow contention-free sharing of data was discussed. Using this organization, N FPs can operate approximately N times faster than one FP, significantly reducing computation times for contextual classification.

With the completion of this portion of research, we can now consider practical implementations of the contextual classifier which can be used effectively with data sets larger than thee 50-pixel-square test sets employed to date. The larger scale implementation could use either the ground-truth-guided method or and adaptive version of the unbiased method for estimating the context function. In either case, the information-class formulation of the approximate algorithm should be employed to reduce computation requirements. If subsequent research into the hybrid algorithm produces favorable results, this algorithm should also be incorporated into the implementation [2]. For optimal efficiency, the contextual classifier should be implemented on several flexible processors in parallel. However, for moderately-sized data sets, an intelligent implementation on a serial computer could be effective for analysis tasks where high classification accuracy is required.

References

1. Swain, P.H., S.B. Vardeman and J.C. Tilton. 1981. Contextual Classification of Multispectral Image Data. Pattern Recognition, Vol. 13 No. 6, pp. 429-441.

2. Tilton, J.C. 1981. Incorporating Spatial Context Into Statistical Classification of Multidimensional Image Data. Ph.D. Thesis, School of Electrical Engineering, Purdue University, West Lafayette, Indiana 47907. Also available as: AgRISTARS Report SR-P1-04148 and LARS Technical Report 072981, Laboratory for Applications of Remote Sensing (LARS), Purdue University.

3. Anuta, P.E., D.A. Landgrebe, H.J. Siegel and P.H. Swain. 1980. Vol. III: Data Processing Research and Techniques Development, Laboratory for Applications of Remote Sensing (LARS), Purdue University, West Lafayette, Indiana 47907, November 1980. LARS Contract Report 112880.

4. Tilton, J.C., P.H. Swain and S.B. Vardeman. 1981. Contextual Classification of Multispectral Image Data: An Unbiased Estimator for the Context Distribution. Proceedings of the 1981 Machine Processing of Remotely Sensed Data Symposium, June 1981, pp. 304-313. IEEE Catalog No. 79CH1637-8 MPRSD. Also LARS Technical Report 070381.

5. Swain, P.H. and S. M. Davis, ed. 1978. Remote Sensing: The Quantitative Approach. McGraw-Hill International Book Co., New York.

6. Tilton J.C., P.H. Swain and S.B. Vardeman. 1980. Context Distribution Estimation for Contextual Classification of Multispectral Image Data. Proceedings of the 1980 Machine Processing of Remotely Sensed Data Symposium, June 1980, pp. 171-180. IEEE Catalog No. 80CH1533-9 MPRSD.

7. Tilton, J.C., and P.H. Swain. 1981. Contextual Classification of Multispectral Data. Proceedings of the International Geosciences and Remote Sensing Symposium, June 1981, pp.285-290. IEEE Catalog No. 81CH1656-8.

8. Van Ryzin, J. 1966. The Compound Decision Problem With m × n Finite Loss Matrix. Annals of Mathematical Statistics, Vol. 37, pp. 412-424.

9. Hannan J., D. Gilliland and S. B. Vardeman. (in preparation). Empirical Bayes and Compound Decision Theory: A Survey and Annotated Bibliography.

10. Miller, K.S. 1964. Multidimensional Gaussian Distributions. John Wiley and Sons, Inc., New York, pp. 24-26.

11. Tilton, J.C. 1980. Contextual Classification of Multispectral Image Data: Approximate Algorithm. Laboratory for Applications of Remote Sensing (LARS), Purdue University. AgRISTARS Report SR-P0-00491; also LARS Technical Report 081580.

12. Alvo, M. and M. Goldberg. 1981. A Measure of Reliability for Classification of Earth Satellite Data. IEEE Transactions on Systems, Man and Cybernetics, Vol. SMC-11, No. 4, pp. 312-318, April 1981.

13. Control Data Corp. 1977. Cyber-Ikon Image Processing System Concepts. Digital Systems Division, Control Data Corp., Minneapolis, MN, Jan. 1977.

14. Control Data Corp. 1977. Cyber-Ikon Flexible Processor Programming Textbook, Digital Systems Division, Control Data Corp., Minneapolis, MN, Nov. 1977.

15. Smith, B.W., H.J. Siegel, P.H. Swain. 1980. A multiprocessor implementation of a contextual image processing algorithm. Laboratory for Applications of Remote Sensing(LARS), Purdue University, W. Lafayette, IN 47907, Jul. 1980. AgRISTARS Report SR-PO-00474; also LARS Technical Report 070180.

16. Swain, P.H., H.J. Siegel, B.W. Smith. 1980. Contextual classification of multispectral remote sensing data using a multiprocessor system. IEEE Trans. Geosc. and Remote Sensing, Vol. GE-18, pp. 197-203, Apr. 1980.

17. Hart, J.F., et al. 1968. Computer Approximations. John Wiley and Sons, Inc., New York.

12. COMPARISON OF EDGE DETECTION METHODS FOR LANDSAT IMAGERY

Paul E. Anuta and C. Pomalaza

Introduction

The edge detection task was defined to evaluate several popular or new edge detection algorithms in the context of RBV imagery. The algorithm would, of course, be applicable to any similar type of imagery. The factors considered in selection of algorithms were numerical complexity, nature of the particular analysis, and quality of the results obtained on test data on areas that contain typical field patterns.

Algorithms chosen for analysis were (1) Sobel Gradient, (2) Kirsch Gradient, (3) Frei and Chen Algorithm, and (4) Hueckel Algorithm. The algorithms and test data are first described and the experimental results are then briefly presented.

Descriptions of Algorithms

The gradient type algorithms tend to use a common neighborhood:

$$\begin{matrix} b_1 & b_2 & b_3 \\ b_4 & b_5 & b_6 \\ b_7 & b_8 & b_9 \end{matrix}$$

denote b, the image intensity vector, as $b = (b_1 \ldots b_9)$ The application of most algorithms involves the scalar product of b with particular weighting vectors.

Sobel Gradient [1]

$$dx = b.w_1$$
$$dy = b.w_2$$

where w_1 and w_2 are defined as

$$w_1 = \begin{matrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{matrix} \qquad w_2 = \begin{matrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{matrix}$$

Kirsch's Gradient [2]

Evaluate the "contrast" function,

$$\max\left(1, \max_{i=1}^{8} |b.w_i|\right), \text{ where } w_1 \ldots w_8 \text{ are}$$

$$\begin{matrix} 5 & 5 & 5 \\ -3 & 0 & -3 \\ -3 & -3 & -3 \end{matrix} \qquad \begin{matrix} 5 & 5 & -3 \\ 5 & 0 & -3 \\ -3 & -3 & -3 \end{matrix} \qquad \begin{matrix} 5 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & -3 & -3 \end{matrix} \qquad \begin{matrix} -3 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & 5 & -3 \end{matrix}$$

$$\begin{matrix} -3 & -3 & -3 \\ -3 & 0 & -3 \\ 5 & 5 & 5 \end{matrix} \qquad \begin{matrix} -3 & -3 & -3 \\ -3 & 0 & 5 \\ -3 & 5 & 5 \end{matrix} \qquad \begin{matrix} -3 & -3 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & 5 \end{matrix} \qquad \begin{matrix} -3 & 5 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & -3 \end{matrix}$$

Frei and Chen Algorithm [3]

- Define an "edge" subspace by finding a set of orthogonal basis vectors. $(w_1 \ldots w_e)$.
- Complete the basis with 9-e "nonedge" basis vectors.
- Project the image subarea intensity values b onto the edge subspace.
- Evaluate the angle between b and its projection onto the edge subspace by

$$\theta = \arccos\left(\sum_{i=1}^{e} (b.w_i)^2 \Big/ \sum_{j=1}^{9} (b.w_j)^2 \right)^{\frac{1}{2}}$$

- The image subarea is considered as containing an edge element if θ is small, i.e., thresholding the value of

$$\sum_{i=1}^{e} (b.w_j)^2 \; / \; (b.b)$$

- The nine orthogonal set of basis vectors for the Frei and Chen method is:

Edge subspace

isotropic	1	$\sqrt{2}$	1	1	0	1
average	0	0	0	$\sqrt{2}$	0	$-\sqrt{2}$
gradient	-1	$-\sqrt{2}$	-1	1	0	-1
ripple	0	-1	$\sqrt{2}$	$\sqrt{2}$	-1	0
	1	0	-1	-1	0	1
	$-\sqrt{2}$	1	0	0	1	$-\sqrt{2}$

Line subspace

line	0	1	0	-1	0	1
	-1	0	-1	0	0	0
	0	1	0	1	0	-1
discrete	1	-2	1	-2	1	-2
laplacian	-2	4	-2	1	4	1
	1	-2	1	-2	1	-2

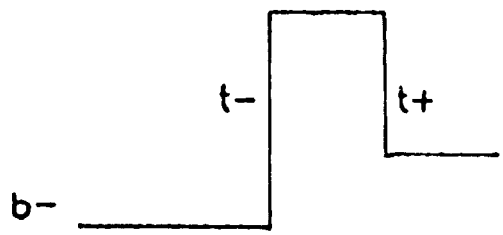
Average

1	1	1
1	1	1
1	1	1

Hueckel Algorithm [4,5]

- Optimal values of an ideal edgeline to the image intensity values in a small circular neighborhood.

- The ideal edge is determined by a 6-tuple of parameters: Three parameters determine the intensity levels (b-, t-, t+ as shown in figure below) and the other three parameters determine the position, orientation, and width of the line.



- The fitting process consists of determining the value of the six parameters for a best fit with the image intensities, i.e., when,

  $N = ||b - S(\text{tuple})||$ is minimum
  S is an ideal edge image.

- The minimization process is approximated by expansion of the input image disk and the edgeline in an orthogonal Fourier series. The minimization is then approximated by choosing a tuple such that,

$$N = \sum_{\ell=0}^{8} (a_i - s_i)^2$$

where a_i are the coefficients of expansion for the image and s_i are the coefficients for an ideal edgeline.

- The reasons for using only the first nine terms of the expansion are:

  (a) Higher order terms correspond to noise in the image and should be ignored.

  (b) An analytical solution to the minimization problem is found using nine terms.

Analysis

The four algorithms were implemented and tested on RBV data from areas in South Carolina and Iowa. These selections were based on data availability and were tested as acquired. The final test area was the Webster County, Iowa, AgRISTARS Segment 893; however, comparisons also were made on an area near Des Moines. The RBV imagery exhibited a speckled noise much like that of radar imagery. These artifacts are thought to be due to analog telemetry and preprocessing problems. Since other data sets could not be acquired and inspected in the time frame of the study, the data on hand were used. In order to reduce the noise in the image, the following procedures were applied:

Linear Filtering. This is a common method to improve the S/N. However, due to the low-pass frequency characteristic of almost all common linear filters, the edges will be smoothed, making their detection more difficult when applying some of the algorithms discussed before.

Non-Linear Filtering. Some non-linear filters have the property of increasing the S/N, preserving at the same time the edges of the image. One of these filters is the median filter 6 . To compare the effect of this filter, a 3-pixel by 3-pixel window linear box filter was applied to the Segment 893 data along with the median filter, also using a 3x3 neighborhood. The same test area was used in the registration task, so some of the figures are common to both sections. Figure 13-3 (see Section 13) is the original RBV data for Segment 893. Its poor quality is due to a combination of factors, including low signal-to-noise ratio, shading in the data, and shading and poor print quality in the dot printer used to make the image. Figure 12-1 shows the area processed with the linear box filter, a median-filter, and after a second pass with the median filter. The double-median filter results were judged best for use in edge detection. The figures do not depict the results as well as desired, but the doubly median-filtered result was judged most effective. Note the considerable reduction in variability in the field in the second median with road and field edges still sharp. The linear filter result is still quite noisy and the fields show some blurring.

It is to this last image that the various algorithms for edge detection were applied. The results of using the Sobel gradient are shown in Figure 12-2 where the gradient values have not yet been thresholded to obtain a binary image. Results obtained by the Kirsch and Frei and Chen algorithms are indistinguishable from the one shown in Figure 12-2. The computational time required on an IBM 3031 computer for the four algorithms for an area 300 x 300 pixels were: Sobel - 30 CPU seconds, Kirsch - 60, Frei and Chen -

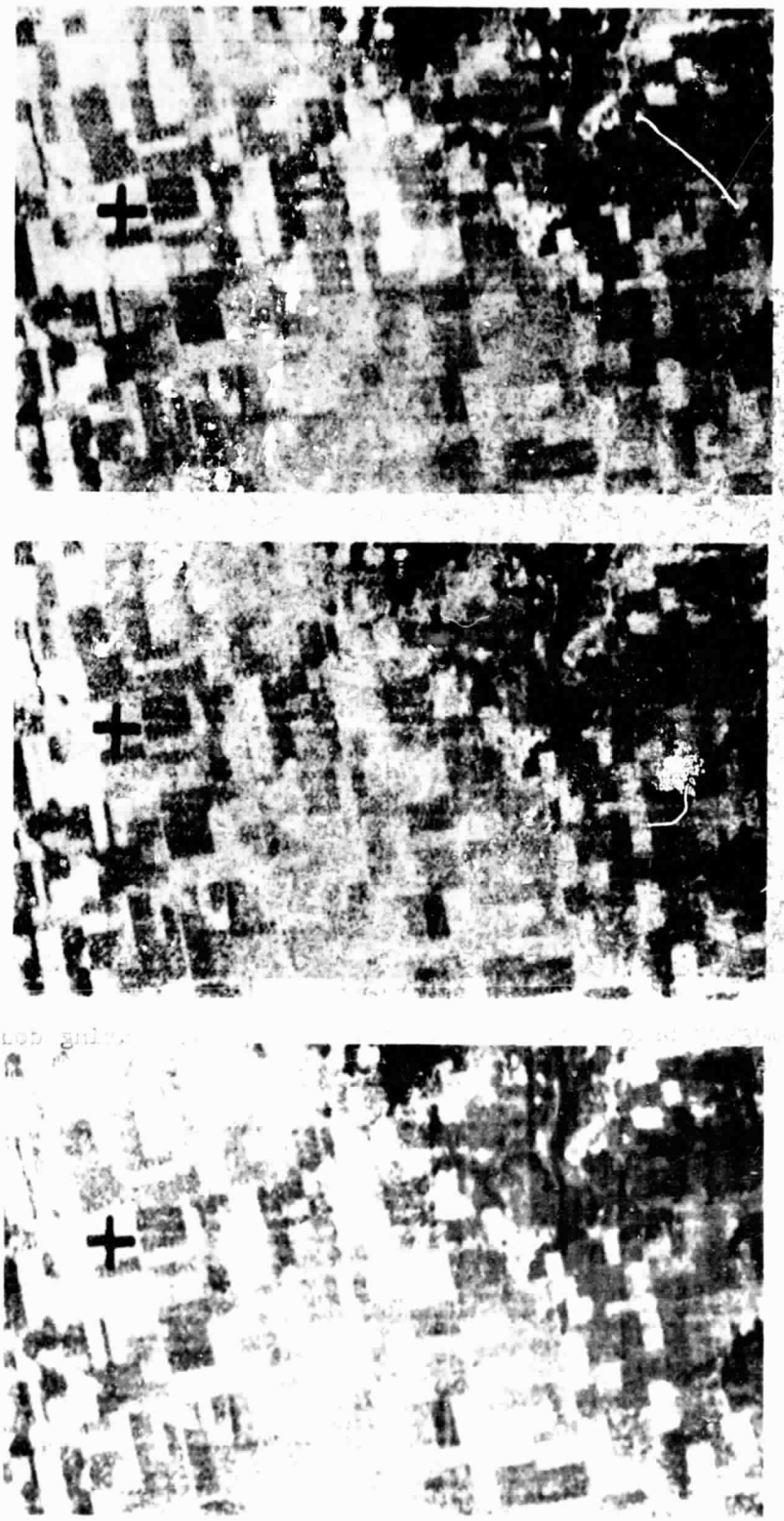


Figure 12-1. RBV imagery of segment 893, Webster County, Iowa, processed by three filter methods.

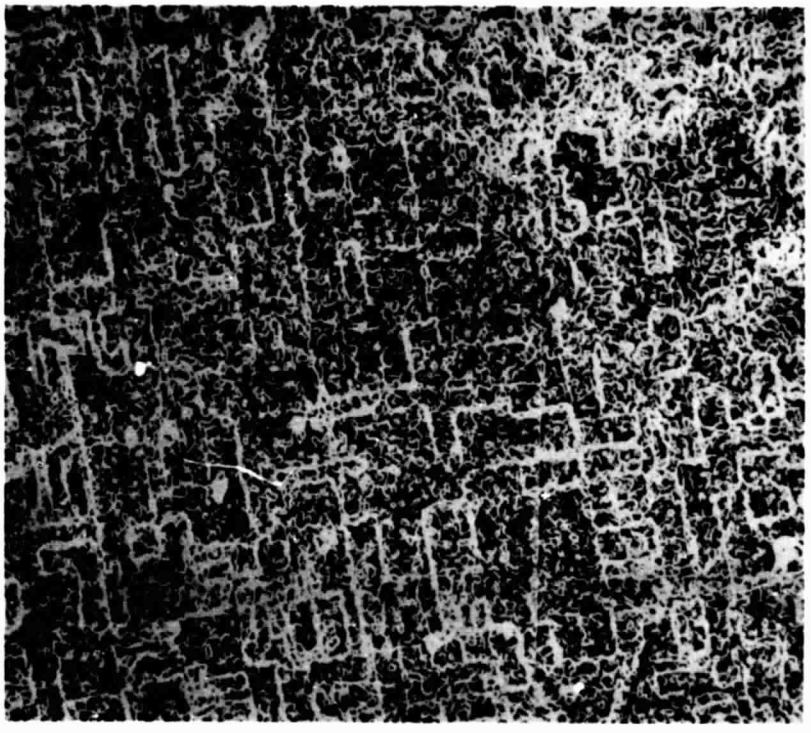


Figure 12-2. RBV imagery processed by Sobel gradient operator using doubly median-filtered data.

40, and Hueckel - 50. Taking into account that the computation of the Sobel gradient requires less CPU time than the other algorithms, it was concluded the Sobel gradient method is the most attractive of the gradient algorithms. Thresholded versions of Figure 12-4 are shown in Section 13 and more clearly reveal the actual edge structure obtained.

The application of Hueckel's algorithm seems now more appropriate for a localized edge detection and/or location problem rather than an image preprocessing step to produce an enhanced image. The output of this algorithm is a sextuple of parameters that defines the intensity levels, position, location and width of the ideal edge that best fit the circular neighborhood where a successful edge detection has been made. An image was not reconstructed from the test. By using these parameters, the resolution of the edge location is available at a subpixel level. Registration applications could take advantage of these results especially when registering images that have different resolutions. The need to incorporate a line-following algorithm in order to process a scene that has several edge-line patterns could make the application of this method more expensive in terms of computer time than the use of the previously mentioned algorithms, followed by some type of postprocessing graph interpolation procedure.

Summary

The Sobel, Kirsch, and Frei and Chen algorithms give basically the same type of image. A threshold operation to obtain a binary image would provide a more interpretive edge image. The value of the threshold depends on the particular characteristics of the scene being analyzed. One method that could be used is the threshold selection procedure developed by Gurney (7). Also some postprocessing algorithms to interpolate graphs and remove isolated noisy segments are necessary.

Hueckel's algorithm has the advantage of providing an equation for the edge-line pattern detected inside the area of analysis (disk of pixels). This equation can define the location of the edges and lines within a subpixel resolution which could be used for registration purposes. For the processing of a scene, a line-following algorithm has to be used in combination with the edge-line detection program in order to obtain a continuous type image.

References

1. Rosenfeld, A. and A. Kak. 1976. Digital Picture Processing. Academic Press, New York.

2. Kirsch, R. 1971. Computer determination of the constituent structure of biological images. Comput. Biomed. Res. 4:315-328.

3. Frei, W. and C. Chen. 1977. Fast boundary detection: a generalization and a new algorithm. IEEE Trans. on Computers 26:988-998.

4. Hueckel, M.H. 1973. A local visual edge operator which recognizes edges and lines. J. Assn. Comput. Mach. 20:634-647.

5. Hueckel, M.H. 1975. An operator which locates edges on digitized pictures. J. Assn. Comput. Mach. 18:113-125.

6. Frieden, B.R. 1976. A new restoring algorithm for the preferential enhancement of edge gradients. J. Opt. Soc. Amer. 66:280-282.

7. Gurney, M.G. 1980. Threshold selection for line detection algorithms. IEEE Trans. Geosc. and Remote Sens. GE-18:204-211.

13. REGISTRATION OF DISSIMILAR DATA

P.E. Anuta and F. Davallou

Introduction

Improved registration capabilities are under development at NASA/JSC for Landsat MSS to MSS data registration. Extensive improvements were made in the methods of correlation and warp function estimation for segment (5 by 6 mi.) size areas. It is desired by JSC that additional data types be registered by expanded versions of the system to extend the selection of data available to application projects. Three tasks were defined and are reported on here to support the expansion of JSC registration capabilities.

1. MSS/RBV Registration. The RBV sensor on Landsat-3 provides high-resolution panchromatic imagery coincident with the MSS data. The nearly double resolution of the RBV could prove helpful in identifying small features and improving the location of boundaries relative to the MSS. Research on image preprocessing and correlation methods and software for registration of these data types is reported.

2. MSS-Map Registration. Registration of the temporal sequence of MSS data to a map cartographic reference will allow rectification of segments to specific map projections via the JSC processor. Map digitization, feature generation, and correlation studies are discussed in this task report.

3. Evaluation of JSC Ground Truth Registration System. Ground truth field boundaries obtained from aerial photography are registered to MSS data using a separate system built for the JSC cartographic technology laboratory. This task is a review of the existing system to determine if changes should be made to improve performance.

Registration of MSS and RBV Data

Registration of these data types will provide a high-resolution panchromatic band which may be useful in scene object location and edge detection. Furthermore, this research applies to the general problem of registering images having different resolution, different spectral bands, and different signal-to-noise properties. Future data sources, such as Thematic Mapper and SPOT, will provide high-resolution imagery which may be required to be registered with lower-resolution imagery. Thus this research is considered to be in support of the problem of registering such dissimilar types.

The problem is first addressed from the theoretical point of view in which optimum filters are determined that will minimize the variance of the registration estimate. This solution is then implemented along with other methods and compared using test data from an AgRISTARS segment. Results are presented and the structure for software for implementing the RBV-MSS registration operation on the JSC registration process is presented.

Solution for Optimum Registration Processor

A model which represents the registration case at hand is shown in Figure 13-1. It is assumed that the input is a high-resolution image and the lower-resolution image is generated by convolving an IFOV function with the high-resolution image. These preprocessing filters are applied to the two images and the results correlated to produce estimates of misregistration.

The solution of Riemer [1] and Svedlow [2] is a deconvolution filter followed by a prewhitening filter for the low-resolution image and a prewhitening filter for the high-resolution image. The filters are of the form:

$$h_1(x,y) = h_D(x,y) \circledast h_w \quad (x,y)$$

$$h_2(x,y) = h_w(x,y)$$

where:

$h_w(x,y)$ is a prewhitening filter which is the reference image invariant part of the matched filter

$h_D(x,y)$ is the IFOV deconvolution filter

For a change image with negative exponential autocorrelation function (representative of Landsat temporal change) with parameters α and β, the prewhitening filter is:

$$h_w(x,y) = \frac{1}{2A\sqrt{\alpha\beta}} \left[\alpha\beta\delta(x,y) + \beta\frac{1}{dx} + \alpha\frac{d}{dy} + \frac{d^2}{dxdy} \right]$$

The IFOV deconvolution filter can be solved for using various approaches. The work of Riemer [1], Chu [3], and Dye [4] is an example. A typical filter derived by Riemer for the Landsat case is shown in Figure 13-2. This assumes a truncated Gaussian IFOV with 80 meter half-power point. This filter was used in the experiments performed as part of this task.

Experimental Evaluation of Registration Methods

A number of preprocessing and correlation methods have been developed for registration of MSS imagery which are based on cost and ease of evaluation criteria rather than on strictly optimum theoretical results. These are derived through engineering judgment and often approximate but do not equal optimum performance.

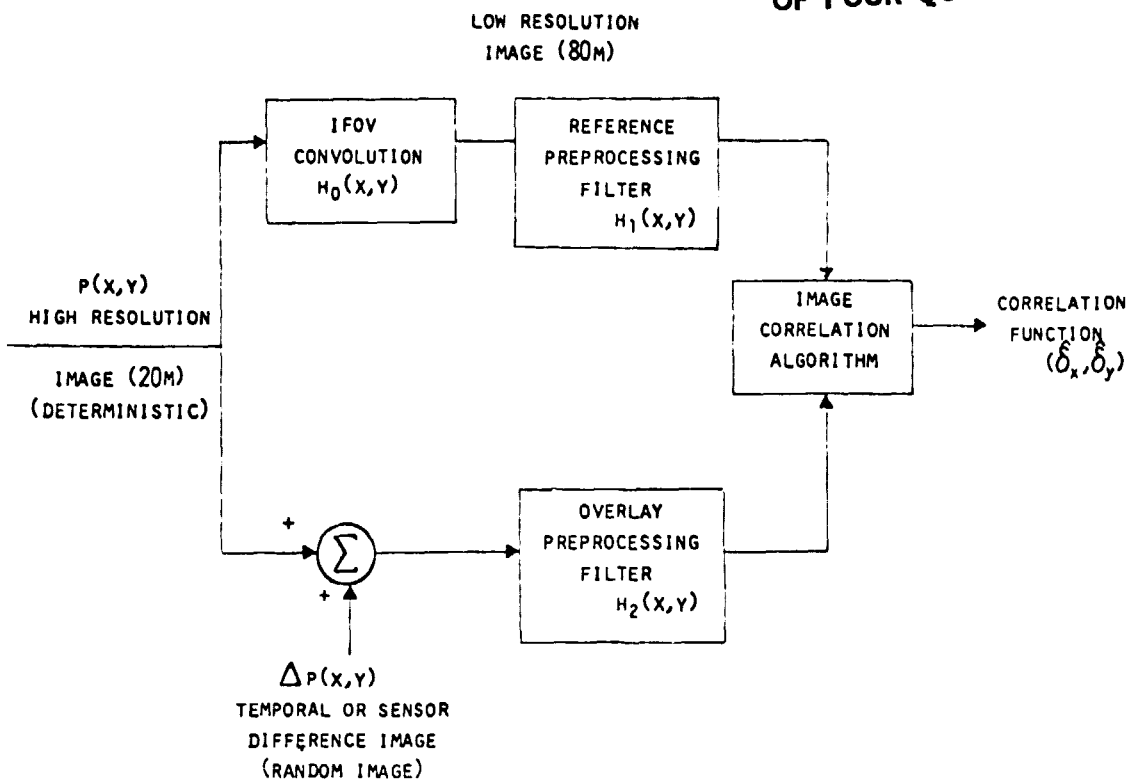


Figure 13-1. Model for analysis of registration of data with different spatial resolutions.

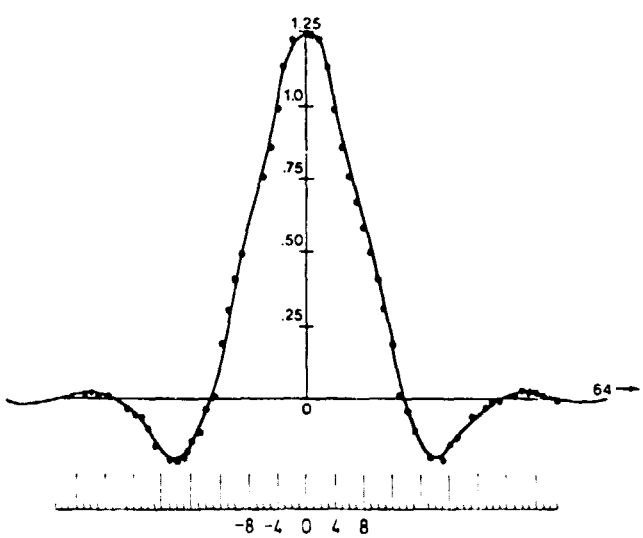


Figure 13-2. Example of IFOV deconvolution filter.

The registration tests were run on data from AgRISTARS Segment 893 near Fort Dodge, Iowa. RBV and MSS data could not be obtained for the same day due to data acquisition problems. However, a very close match was obtained

from overlap coverage nine days separate in time. The RBV data are from Frame 830922-16095 on September 12, 1980 and the MSS data are from Frame 27051-16204 on September 3, 1980. The MSS data are of good quality but the RBV data are of lower quality than expected. The RBV data have a pronounced speckled effect due to telemetry and ground-processing problems and because the signal-to-noise ratio was low. Also, shading was apparent from place to place over the subframe. These data are the best available with MSS data, so they were used for the evaluation.

The problem of the differing resolutions of the sensors was addressed first. The MSS IFOV is nominally 80 meters with nominally 57 meter square pixels in the Landsat-3 data. The RBV data have nominally 19 meter pixels with the actual resolution being in the area of 40 meters. In order to match resolutions for registration, the MSS data were interpolated using cubic Lagrangian interpolation by a factor of three to produce 19 meter pixels. This is a low-cost method of achieving the higher pixel resolution and it was of interest to test this approach since the IFOV deconvolution methods discussed above are much more costly.

Figure 13-3 contains a photographic reproduction of a Varian printer grayscale image of the RBV data for the test area. The poor quality of the image is due both to Varian printer problems and RBV data problems. The field and road structure is clearly evident, however, and the sharper quality of the edges is readily apparent even with the problems. Figure 13-4 is a Varian printer reproduction for MSS band 5 cubic interpolated to 19 meter pixels. Figure 13-5 is the same for band 6. Note the much broader width of roads and field edges. The signal-to-noise ratio of band 5 also is not particularly high in this frame.

The three images shown form the basis for all experimental work done on the project. Although the optimum correlator solution indicates that the best preprocessor is a gradient operator feeding a linear correlator, practical considerations dictate that binary or edge no-edge correlators be used in actual implementations due to computing-cost considerations. Thus the test images were processed with a gradient type preprocessor and the result was thresholded to produce a 0-1 binary edge image. Results from the edge detection task, plus other research results indicate that the Sobel edge operator is very effective and at the same time low cost. Thus the Sobel operator was chosen for this project and applied to the data. The RBV data were first doubly-median filtered to reduce the noise problems. The threshold was chosen by examining edge histograms and choosing samples from edge areas. The thresholds tend to place 10 to 15% of the edge pixels above the threshold.

The problem of differing spectral bands of the two sensors was approached by combining edges from MSS bands 5 and 6. The RBV spectral band is .50 to .75 μm, MSS band 5 is .6 to .7 μm, and Band 6 is .7 to .8 μm. Combining bands 5 and 6 is a reasonably good approximation. Roads tend to show up well in band 5 and field edges are very strong in band 6. The RBV image contains both clear roads and field edges. Figure 13-6 contains the edge image for the RBV data and Figure 13-7 contains the combined band 5 and 6 edges, both using the Sobel gradient.

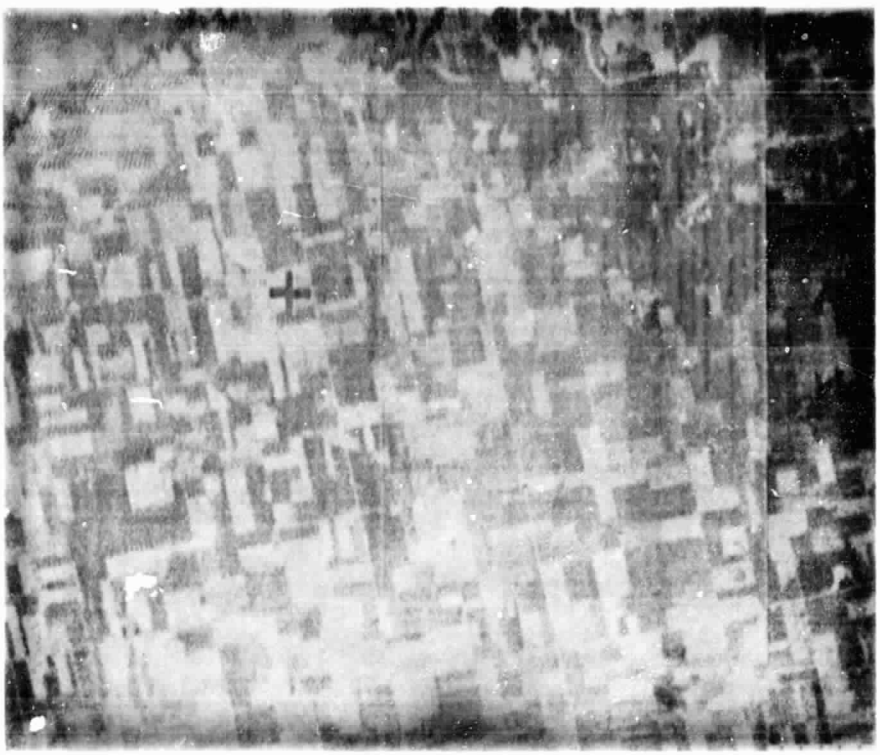


Figure 13-3. RBV imagery for Segment 893, September 12, 1980.

Figure 13-4. MSS band 5 imagery interpolated to 19m using cubic interpolation, September 3, 1980.

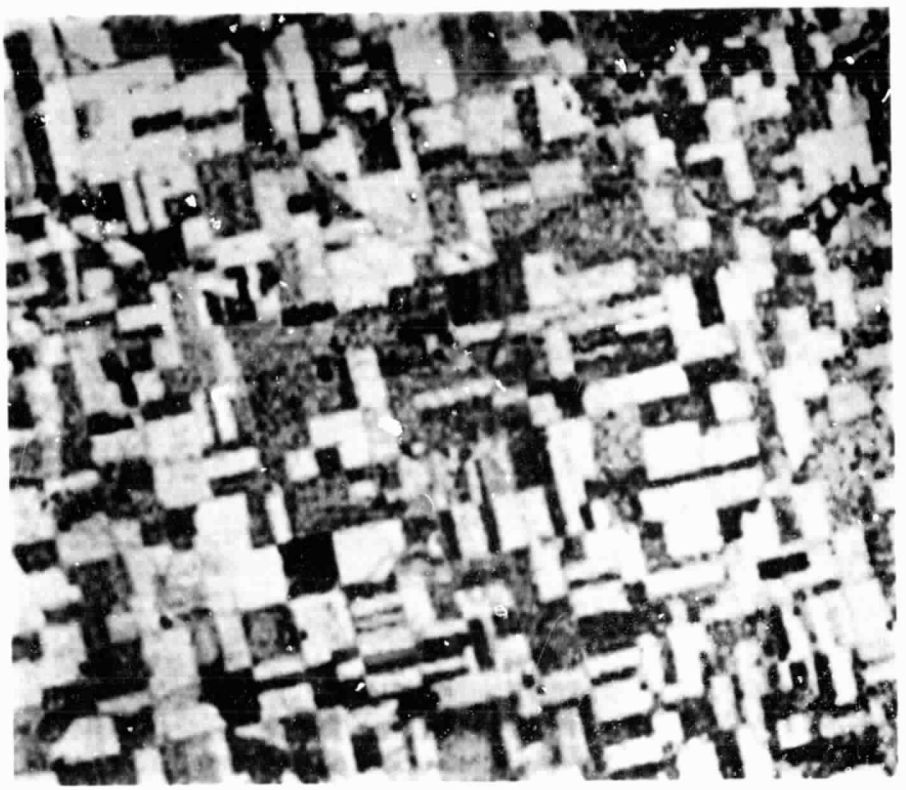


Figure 13-5. MSS band 6 imagery interpolated to 19m using cubic interpolation, September 3, 1980.

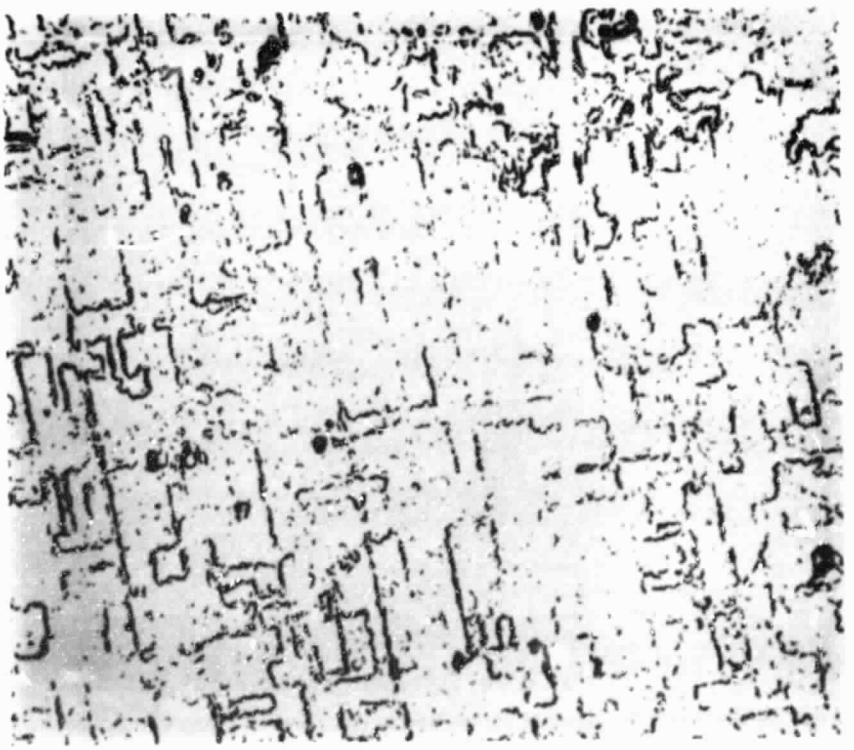


Figure 13-6. RBV edge image obtained from thresholded Sobel gradient.

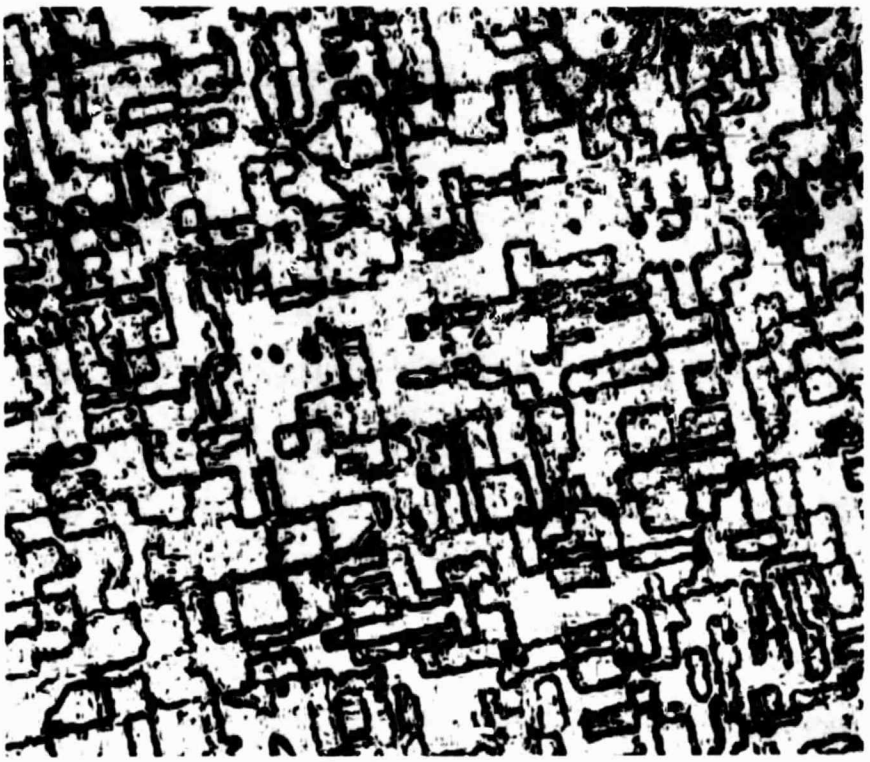


Figure 13-7. MSS band 5 and 6 edges from thresholded Sobel gradient.

Correlation tests were run on the edge images using the correlation coefficient to enable a correlation number between -1 and 1 to be obtained. As noted above, in an operational implementation, a binary logic and counter arrangement would be used for maximum efficiency; but in the research case, the more complex expression can be used. The non-thresholded data were also correlated to give a comparison with the thresholded case. The MSS bands 5 and 6 edges were not combined for these cases, so the correlations were done separately. Tables 13-1 and 13-2 contain the results. The analysis assumes the scales are the same for the MSS and RBV, so the position of the correct shift peak in the correlation function would be the same in all tests. To verify this, a geometric analysis was made of the RBV versus MSS images. Three control points were accurately chosen in RBV and MSS printouts and an affine function was fit to these points. Three points provide six values so the six parameters can be computed. The affine form is:

$$x^1 = a + bx + cy \quad \text{(Lines)}$$

$$y^1 = d + ex + fy \quad \text{(Cols.)}$$

Table 13-1. Correlation results for MSS bands 5 vs. RBV, segment 893, non-thresholded Sobel edges.

ρMAX	ΔL	ΔC	UNIQUE PEAK	ρMAX	ΔL	ΔC	UNIQUE PEAK
.29	-3	3	Yes	.14	-2	5	Yes
.17	-6	8	Yes	.16	7	2	No
.08	-7	2	No	.24	-8	-2	Yes
.08	-4	2	No	.07	-4	4	Yes
.22	1	6	No	.22	-8	2	Yes
.14	-2	4	Yes	.42	-2	3	Yes
.07	4	3	No	.08	-6	7	No
.14	-8	3	Yes	.13	-5	-6	No
.08	-4	2	No	.25	-8	2	Yes
.28	-7	4	No	.09	-4	3	No
.10	4	3	Yes	.10	-8	4	No
.10	-4	6	No	.03	-8	1	No
.11	-8	4	No	.42	-4	2	Yes
.30	-3	3	Yes	.09	-1	6	Yes
.12	1	4	Yes	[Mean ρ = .16]			

* For Unique Peak results only

	ΔL	ΔC
MEAN*	-3.6	3.33
STD*	3.94	2.12

% Acceptable: 52%

These six parameters provide specification of translation in two dimensions, rotation, scale in two dimensions, and skew. The results were a = 23, b = -109, and all others were zero. This verifies that there is no rotation, scale difference, or skew between the RBV and MSS imagery to six decimal places. The constant shift of 23 lines and -109 columns was used in all the correlations to place the correct peak at the center of the correlation function output.

The correlation results for thresholded RBV and combined MSS band 5 and 6 edges are presented in Table 13-3. A slightly different evaluation method was used. If peaks were at the lag limit, they were not used and if a secondary peak was observed closer to the center (true) position, it was used instead of the primary peak. The standard deviations are large in either case, indicating that accuracy would not be particularly good for RBV, MSS registration. Errors in the region of three to four pixels could be expected using the preprocessing described.

Table 13-2. Correlation results for MSS band 6 vs. RBV, segment 893, non-thresholded Sobel edges.

ρMAX	ΔL	ΔC	UNIQUE PEAK	ρMAX	ΔL	ΔC	UNIQUE PEAK
.25	-6	2	Yes	.20	-8	2	Yes
.21	-8	8	Yes	.19	-3	1	Yes
.18	-6	5	Yes	.08	-8	-4	No
.12	-8	3	Yes	.17	-8	0	Yes
.09	-5	-7	No	.16	-8	-4	Yes
.21	-6	4	Yes	.13	2	5	Yes
.24	-1	6	Yes	.15	-4	6	Yes
.17	-7	3	Yes	.17	-7	-5	Yes
.13	-8	3	Yes	.15	-7	1	Yes
.18	-8	-1	No	.12	-4	0	Yes
.29	0	7	Yes	.19	-8	1	Yes
.11	-1	5	Yes	.09	3	8	Yes
.28	-8	1	Yes	.14	-8	-5	No
.22	-4	7	No	.29	1	4	Yes
.19	-1	6	Yes	[Mean ρ = .18]			

* For Unique Peak results only

MEAN*	-4.63	2.88
STD*	3.13	3.16

83% Successful

In an attempt to decrease the variance of the registration error, the IFOV deconvolution operation discussed above was run on the MSS data to hopefully sharpen the edges. Only the band 6 edges were obtained from this process. Figure 13-8 contains the thresholded Sobel edges for the deconvolved band 6 data. Note the much narrower edges. This operation was considered very promising and correlation tests with the RBV edges were conducted to see if results were improved. The correlation results are presented in Table 13-4. It can be seen that the line standard deviation is lower and the column value is about the same. However, the % acceptable was 100% and the degree of correlation was much higher. It is concluded that the IFOV deconvolution is desirable for this difficult registration task. The MSS deconvolution approach was thus selected for the processor to be implemented on the JSC system.

Figure 13-9 is a block diagram of the steps recommended for the RBV-MSS registration process. Software compatible with the JSC processor necessary for carrying out these steps is a product of this study.

Table 13-3. Thresholded MSS of combined bands 5 and 6 and RBV correlation results, segment 893.

ρMAX	ΔL	ΔC	ρMAX	ΔL	ΔC
.2	1	2	.11	1	-1
.09	2	4	.09	1	6
.08	--*		.09	--	
.09	-6	6	.10	4	-1
.16	--		.11	-5	-1
.07	5	2	.10	-7	-8
.12	--		.11	4	-2
.08	-6	6	.19	-5	7
.08	-6	6	.15	-2	3
.13	-3	0	.10	-1	3
.14	5	3		--	
.11	--		.12	0	0
.18	-1	6	.11	4	1
.17	1	3	.10	5	-3
			.11	-6	0

Line Mean	-.65	% Acceptable	79%
Col. Mean	1.83	Aver. ρMax	.12
Line Std. Dev.	4.16		
Col. Std. Dev.	3.51		

*Blank entries indicate shift was at limit of lags and considered unacceptable.

MSS-Map Registration

The second task in the registration research involves developing techniques for correlating MSS image data and imagery derived from topographic maps. The purpose is to improve the accuracy with which remote sensor imagery can be rectified to a geographic reference. The approach used in this study is to digitize topographic maps and create an image format which can be correlated with MSS imagery. It is hoped that the sharp edges in the map will reduce the variance of the registration relative to image-to-image correlation. Also, the positions of map features are known precisely by survey.

The original intent was to digitize map separations obtained from the USGS for the map image scene. The separations were not received by the end of the contract year. Expecting late acquisition of the separations, we asked an artist to trace a synthetic map image from a topographic map and

Figure 13-8. MSS band 6 imagery deconvolved, interpolated to 19m, Sobel edge processed, and thresholded.

used it as the map image for experimental purposes. Segment 893 near Ft. Dodge, IA, was again chosen as the test site since all needed maps and data were available from other parts of the study. Figure 13-10 is a photo of the topographic map mosaic of Segment 893 and Figure 13-11 is the artist's extraction of the features from the map likely to be seen in the MSS data.

Figure 13-12 is a printout of the digitized hand-drawn map. This is the image to be correlated with the MSS data. Note the noisiness in some areas of the image. This is due to variations in the transparency used in the film-scanning process and can be removed by improved photo processing and by thresholding the digital image.

Correlation tests were carried out on the digitized map and MSS data. The results were inconclusive and as it is expected that this task will continue in the following contract year, no further results are presented here.

Technical Review of Automatic Registration System for the Cartographic Technology Library

The third task of the registration research required that the subject registration system be reviewed and evaluated and recommendations made for changes which would include latest technology if needed. The descriptive document used in the evaluation is cited as reference 5.

Table 13-4. Correlation results for RBV and IFOV deconvolved MSS band 6 data, segment 893, thresholded Sobel edges.

ρMAX	ΔL	ΔC	ρMAX	ΔL	ΔC
.29	-1	3	.29	1	5
.15	-2	-3	.10	5	6
.21	-6	6	.11	0	4
.18	-6	6	.37	-6	4
.17	8	-5	.20	-6	-1
.24	-3	7	.13	-2	2
.24	2	8	.19	1	8
.21	-7	6	.18	-1	-1
.18	-6	6	.14	0	5
.17	-6	3	.15	0	8
.31	2	8	.18	0	6
.16	-1	7	.26	-8	3
.19	-4	4	.25	-2	2
.26	-2	7	.32	2	7
.17	3	-3			

Line Mean	1.55	% Acceptable	100%
Col. Mean	4.07		
		Aver. ρMax	.21
Line Std. Dev.	3.74		
Col. Std. Dev.	3.55		

Review of System Structure

The system consists of two main processors: (1) The Landsat processor and (2) the Registration processor. The first produces edge images from the four Landsat bands and delivers these to the registration processor. The registration processor carries out the correlation of the ground truth data and Landsat edge imagery and computes a least-squares fit to a warp function to the correlation-derived control points so that precise registration may be achieved. Numerous subroutines exist to carry out the many functions required.

A brief overview of the system indicates that it is very well organized and documented. The robustness and efficiency of the code cannot be evaluated in this effort; rather, specific algorithm approaches are reviewed and alternate advanced technology approaches are suggested when they are deemed appropriate.

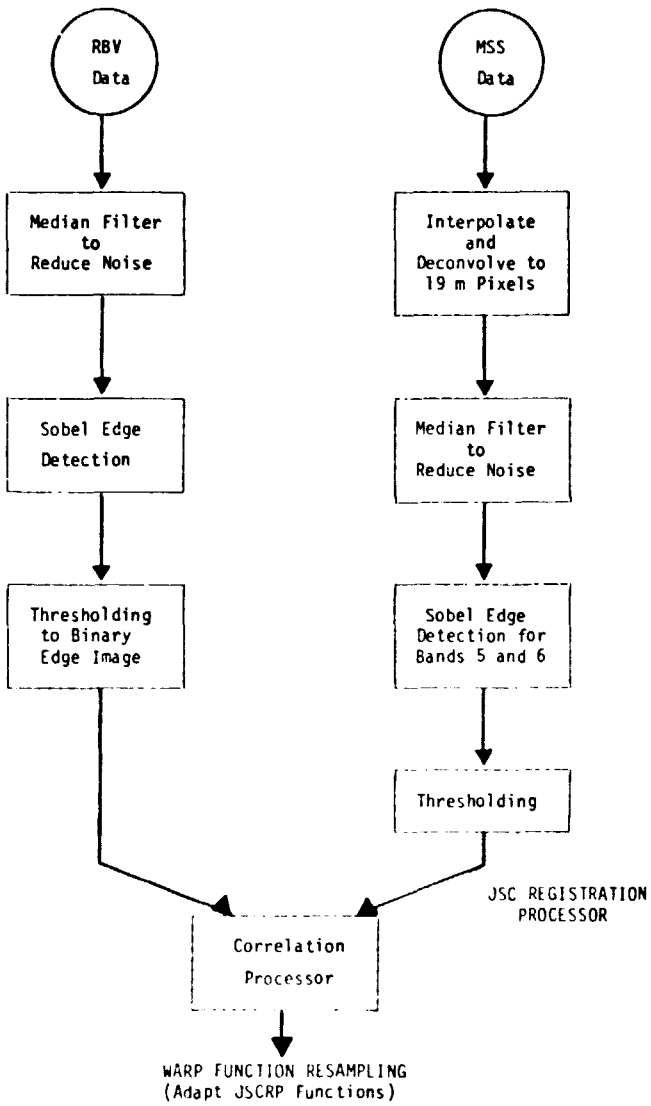


Figure 13-9. RBV-MSS preprocessing steps for registration using JSC processor.

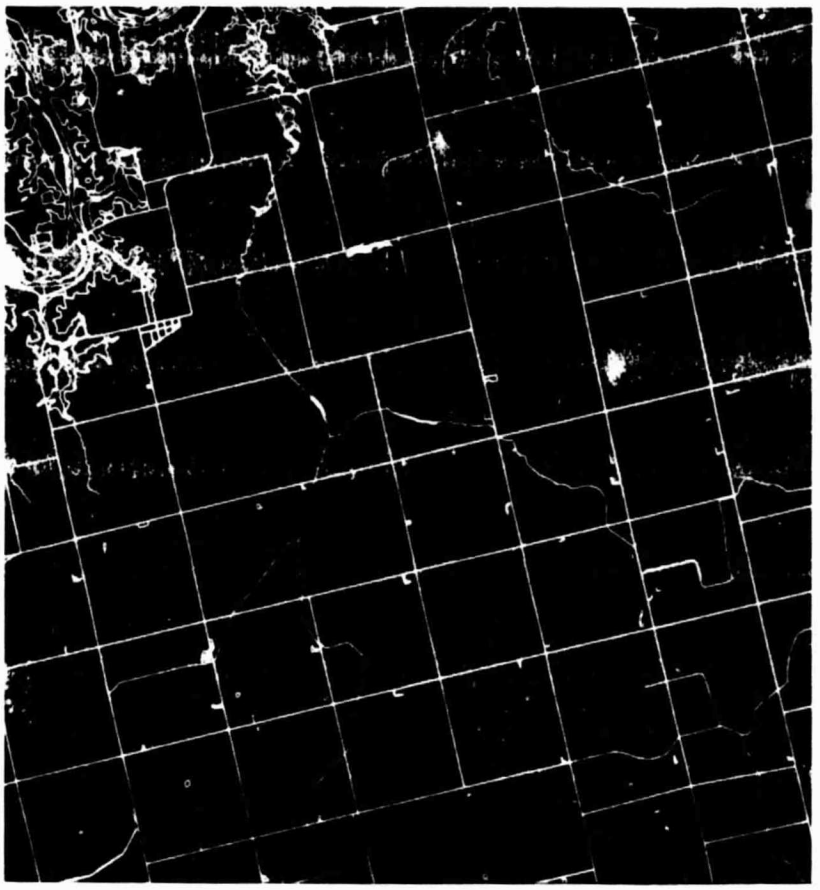


Figure 13-10. Topographic map mosaic of segment 893, Webster County, Iowa.

Specific Comments on the Processors

The Landsat Processor creates an edge image for each channel and combines edges from two channels to produce the output edge images to be used by the correlator. The edge images are created by a gradient operator, the output of which is thresholded such that 15% of the gradient values are above the threshold. For the combined edges, the threshold is varied 2% to thin the edges and noise pixels are removed.

Comments on this algorithm are that it is ad hoc but conforms to standard engineering practice and good results are generally obtained. Roads show up well in Channel 2 (band 5) and field edges show up well in Channels 3 and 4 (bands 6 and 7). The basis of the test for 90% of gradient pixels for band 7, being less than 20, is not stated. If this fails, band 6 is used. This cannot be evaluated. Either 6 or 7 should give good field edges. There are a number of edge-detection and thresholding algorithms which are well known in the image-processing field. Some were tested in the current year's research and it was found that the Sobel algorithm performed as well or better than others. Thresholding approaches by Gurney [6] and Nack [7] are attractive Nack [7] are attractive but were not tested in the current research.

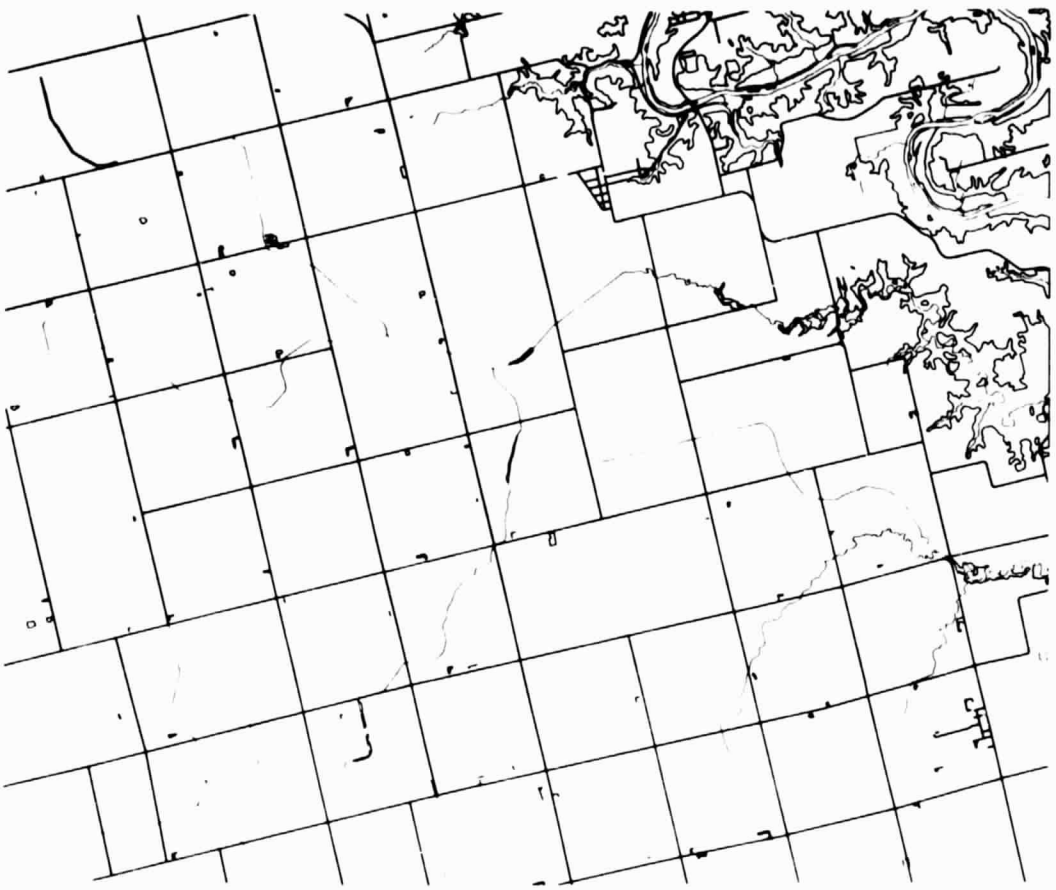


Figure 13-11. Artist's extraction of map features seen in MSS data of segment 893.

Improved results would likely be obtained if the Sobel detector were used. It cannot be stated whether alternate thresholding methods would yield improvement. The cited Gurney and Nack methods should be investigated if possible.

The results of the edge detection research (Section 12) should be cited again here since significant thinning of MSS edges was observed. The use of IFOV deconvolution and median filtering resulted in much sharper edges for MSS band 6 imagery and this may result in more accurate registration of ground truth edges. The cost of the processing would be higher but the deconvolution needs to be done only at the correlation locations. The deconvolution involves about a 9x9 pixel integration area, which is significant; but if accuracy is increased, it may prove worthwhile. Further consideration should be made of this technology since efforts were made in the design to thin edges by raising thresholds. This indicates an awareness by Lockheed of the problem, but we feel that deconvolution of the MSS imagery would be a more robust approach.

The Registration Processor uses the ground-truth correlation and least-squares processor to achieve registration. An iterative sequence is

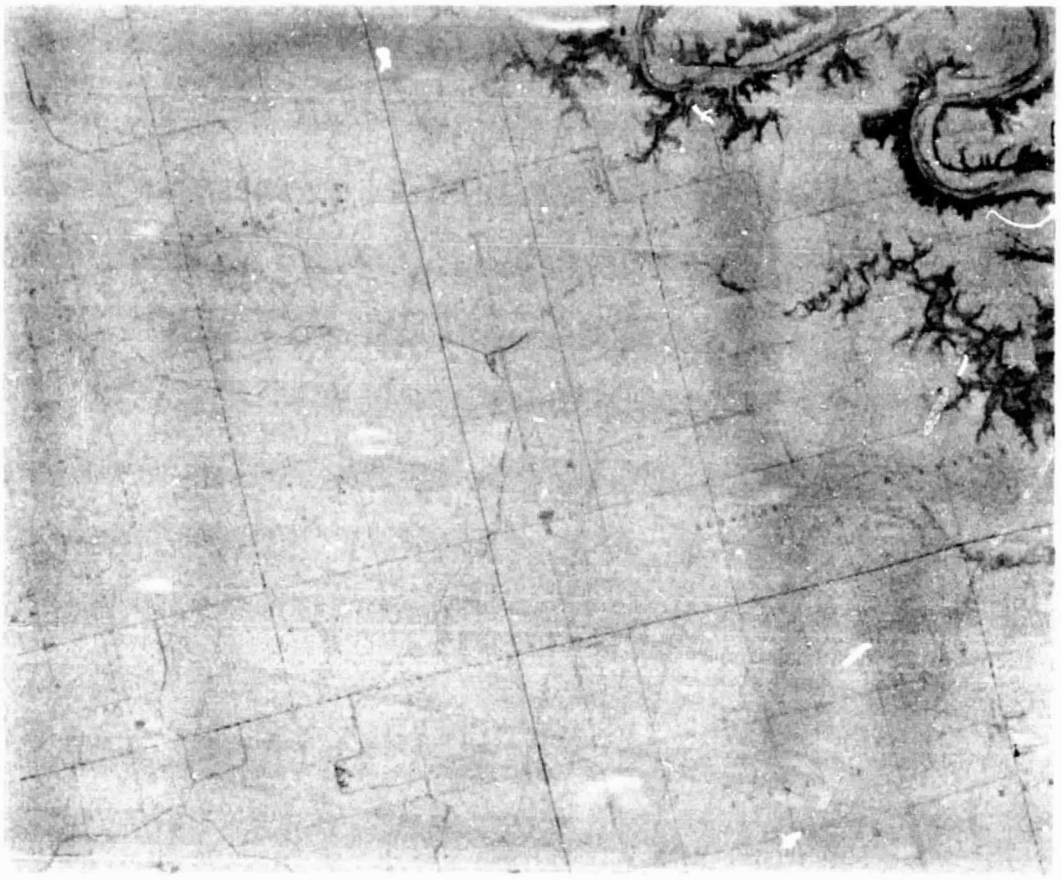


Figure 13-12. Printout of digitized hand-drawn map shown in Fig. 13-11.

used to converge on increasingly accurate registration coefficients. Two converging sequences are used, one for sites having a "normal" number of boundaries. The second is for sites having few boundaries. The algorithm continues if any control point shifts more than one pixel; otherwise it stops. Different window sizes are used and different amounts of shifts are taken in the converging sequence.

It is not clear from the write-up what the rationale or theory is behind the iteration. This appears to be an ad hoc approach and impossible to evaluate without test evidence.

The Ground Truth Processor transforms boundary pixels from the digitized ground-truth field edges to points in the Landsat grid. The transformation used is the colinearity transform from photogrammetry. The pixel size in the ground truth is 6 to 8 times finer than Landsat. A nearest-neighbor assignment of the boundary pixel to the Landsat grid is used.

Two comments arise here. The colinearity transform is used for aerial photographs to compensate for tilt of the photo plane with respect to the scene plane. It is thought that most or all of this distortion is taken out

when the segment photographs are made, so it may not be necessary to use colinearity. If the photogrammetric distortions are small, then the simple affine transformation can be used and the parameter estimation process would be much simpler. It should be determined if the full twelve coefficients are necessary for this process, especially since registration is evidently done at Landsat resolution so the high accuracy of the photo geometry is not even used. The second point deals with the grid used and relates to the suggestion to use deconvolution. By using a finer grid, more of the photo accuracy would be used and edge-thinning after deconvolution of the MSS would allow more accurate boundary registration.

The Correlation Processor carries out the correlation of photo and MSS edges. Control points are searched for near nodes of a regular grid. Correlation is, of course, a logical "AND" operation and summing of the number of matched "ones." A peak detection algorithm employs three quality evaluations and selects the peak accordingly. A straightforward, least-squares procedure is then used to obtain the coefficients for the photogrammetric distortion function.

The first two parts of the process seem to employ many ad hoc steps and the third is straightforward. Searching near the nodes of a regular grid for successful correlations is judged to be a sound approach and no modifications are suggested. The three criteria used for peak evaluation are logical and we believe will work well. Other methods use the variance of the pedestal around the peak relative to the peak value as a measure of sharpness; this may be considered, but it is felt that the methods used will be satisfactory.

This system is well structured and although many of the methods appear to be ad hoc, they are thought to be robust and should work well. Only performance evaluations on actual data will reveal actual capabilities and such data are needed for further evaluation.

References

1. Riemer, T.E. and C.D. McGillem. 1974. Optimum constrained image restoration filters. Tech. Report 091974, Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, IN.

2. Svedlow, M.E., C.D. McGillem and P.E. Anuta. 1976. Analytical and experimental design and analysis of an optimal processor for image registration. Tech. Report 090776, Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, IN.

3. Chu, N.Y. and C.D. McGillem. 1978. Methods and performance bounds for constrained image registration. Tech. Report 061678, Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, IN.

4. Dye, R.H. 1975. Restoration of Landsat images by discrete two-dimensional deconvolution. Proc. Tenth Intl. Symp. on Remote Sensing of Environment, Ann Arbor, MI, Oct. 6-10, 1975.

5. LEMSCO Staff. As-built design specifications of the automatic registration system for the cartographic technology laboratory. Report JSC-17017, Lockheed Engineering and Management Services Co., Houston, TX.

6. Gurney, C.M. 1980. Threshold selection for line detection algorithms. IEEE Trans. Geoscience and Remote Sens. GE-18:204-211.

7. Nack, M.L. 1975. Temporal registration of multispectral digital satellite images using their edge images. Proc. AAS/AIAA Astrodynamics Specialist Conference, Paper No. AAS75-104.

14. WORKSHOP ON KEY ISSUES IN ANALYSIS OF REMOTE SENSING DATA

P. H. Swain

Introduction

The year 1981 found the remote sensing community assessing the results of completed applications-oriented tests of the remote sensing technology and looking ahead with great anticipation to new opportunities for advancing the technology and broadening its use. For example, the Large Area Crop Inventory Experiment (LACIE) 1 and the development of a Forest Resources Information System (FRIS) for commercial application 2 had demonstrated the capabilities and limitations of the mid-1970's technology. The future availability of new sensors, including the Thematic Mapper and the French SPOT multispectral sensor, plus the anticipation of renewed research support from NASA through a new fundamental research program provided motivation for understanding clearly both the current status of the technology and the directions which future research must take to best utilize remote sensing.

These considerations stimulated the convening of a Workshop on Key Issues in the Analysis of Remote Sensing Data at Purdue University, June 22-23, 1981, in conjunction with the 1981 Symposium on Machine Processing of Remotely Sensed Data. Jointly sponsored by Purdue's Laboratory for Applications of Remote Sensing (LARS) and NASA, the workshop had the following objectives:

1. To assemble experts in remote sensing and related information-processing and image-processing technologies for the purpose of making an up-to-date assessment of the state-of-the-art of machine analysis of remote sensing data.

2. To determine the nature of the key research problems remaining as barriers to broader and more effective use of machine analysis of remote sensing data.

3. To produce a report for use by interested researchers and potential research sponsors detailing the findings and recommendations of the workshop participants.

To achieve these objectives, invitations to participate in the workshop were extended to several well-established scientists and engineers in the field from universities, research institutions, and government. The workshop also was publicized in the widely mailed preliminary program of the Machine Processing Symposium. Thirty-six participants were on hand when the workshop was called to order.

To establish a common point of departure for the meeting, the report entitled "Basic Research Planning in Mathematical Pattern Recognition and Image Analysis," by Jack Bryant and L.F. Guseman of Texas A&M University, was mailed to those who registered in advance and was distributed at the conference to all others who registered. The report summarized the

conclusions of a NASA-commissioned working group charged with defining a fundamental research program in image processing for remote sensing. As such, it provided a natural starting point for the discussions planned for the workshop.

Sessions of the workshop focused on:

* Data Bases and Image Registration, including presentations on Data Bases for Remote Sensing, Image Preprocessing Operations, and Map-Oriented Considerations.

* Advanced Technology, including presentations on Advanced Digital Systems, and Artificial Intelligence Methods.

* Information Extraction, including presentations on Classification, and Classifier Training Considerations.

Each session had a reporter assigned to record and summarize key points in the presentations and the associated discussion periods. The workshop ended with general comments from Mr. R.B. MacDonald of NASA/JSC, representing the workshop cosponsor, concerning the near and intermediate term outlook for support of fundamental research in remote sensing.

Summary of Workshop Proceedings

With regard to data bases and image registration, it was surprising to find a great deal of disagreement on the degree to which improved registration and rectification of data are required. There seemed to be a general consensus that research is needed in

* Improved platform control and sensor modeling to reduce the need for rectification and registration.

* Modeling atmospheric effects and the atmosphere point spread function.

* Acquisition and utilization of digital terrain data.

* Understanding how to quantify the real needs of the user/application for precision rectification and registration of the data and the degree to which analysis results and user end products are affected by errors in registration and rectification.

The areas of advanced technology which were considered seemed to be perceived as somewhat divergent with respect to their prospects for near-term applicability to remote sensing of renewable resources. The emergence of parallel processing systems, capitalizing on the shrinking size and cost of digital computers, was recognized as having great potential for amplifying the rate at which digital imagery can be processed; some systems already exist to do this. General applicability of this form of advanced digital technology may follow from successful research in the direction of

* Memory architecture and management strategies for interfacing parallel processing systems and high-volume, high-dimensional remote sensing data.

* Understanding the theoretical speedup limitations of parallel systems and the concomitant implications for the cost versus benefit tradeoffs involving such systems.

It was further recommended that

* A prototype parallel system using contemporary technology should be assembled to demonstrate the theoretical models and validate performance predictions.

Artificial intelligence, the aim of which is to find ways to make computers perform tasks normally thought of as requiring human intelligence, could eventually lead to automation of the process of obtaining high-level information from pictorial data. Looking at the steps conventionally followed in proceeding from a scene to a description of a scene by way of remote sensing and computer processing, it was observed that artificial intelligence research could contribute to

* Development of scene and sensor models which will allow reduction of raw image data to a form free of incidental variations ("noise" of various forms) without needing local ground truth or ancillary data.

* Development of scene models and analytical mechanisms which will facilitate both the representation and manipulation of information available from a scene (e.g., graph structures and machine-implemented reasoning processes).

But there was some skepticism with respect to the near-term applicability of artificial intelligence research results in remote sensing. Some feel that a more fruitful approach would be to concentrate on facilitating interaction of the human analyst with his data. Still, given that the potential payoff of success in the artificial intelligence domain is very great, near-term progress is hardly a fair criterion for prioritizing fundamental research needs.

In the information extraction sessions, a recurrent notion was "mixtures." More specifically, two significant issues were how to deal with mixtures of dissimilar data in multitype data bases, and how to resolve ambiguities resulting from mixture pixels (often boundaries) in image data. The latter problem stands as a serious barrier to improved spectral classification accuracy and proportion estimation accuracy and is widely recognized as requiring concerted attention. The former represents more an opportunity than a barrier, a source of information about the observed scene which the technology has only begun to exploit. Specific research issues identified include

* Quantifying the effects of mixture pixels on classification and proportion estimation accuracy; finding effective ways to resolve uncertainties arising from the presence of mixture pixels.

* Development of more effective and efficient sampling techniques for classifier training, classifier evaluation, and area/proportion estimation.

* Determining meaningful ways to evaluate and compare alternative methods for proportion estimation.

* Development of effective formalisms for characterizing and differentiating among spatial patterns in complex scenes.

* Development of statistical models and classification methods applicable to data sets with components from greatly different sources.

Conclusions

Overall, the panel of experts did not take issue strongly with any aspect of the Bryant/Guseman report. Quite appropriately there was a strong tendency to focus sharply on basic understanding as opposed to, say, algorithm development. Specifically, the discussions highlighted the need for:

1. Understanding and modeling the physical phenomena which produce deleterious abberations in remote sensing image data.

2. Quantification of user needs for precision in image registration and rectification in order to understand the real value of these operations and impact of residual errors.

3. Understanding the real potential of parallel computing systems for improving the processing efficiency of large remote sensing data sets.

4. Understanding how images capture useful information, how humans extract that information through reasoning processes, and how computers might emulate these processes.

5. Understanding the impact of mixture pixels on scene analysis results and exploration of new approaches for dealing effectively with them.

6. Modeling relationships among diverse data sources and understanding how useful information may be extracted from these relationships.

Acknowledgements

The assistance of Shirley M. Davis of LARS and John M. Almon of the Continuing Education Administration, Purdue University, in coordinating physical arrangements for the workshop is gratefully acknowledged.

NOTE: A more complete digest of the workshop proceedings is available as LARS Technical Report 062481, Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, IN 47907.

References

1. MacDonald, R.B. and F.G. Hall. 1980. Global crop forecasting. Science 208:670-679.

2. Barker, G.R. Resource information needs in industry and the role of remote sensing. 1981 Proc. Machine of Remotely Sensed Data Symposium, Purdue University, West Lafayette, IN, p. 9 (abstract only).

3. Bryant, J. and L.F. Guseman, Jr. 1980. Basic research planning in mathematical pattern recognition and image analysis. Depart. Mathematics, Texas A&M University, College Station, TX.

VI. Computer and Data Base Services

15. COMPUTER PROCESSING AND DATA BASE SUPPORT

J.L. Kast, R.A. Garmoe, S.K. Hunt, L.A. Kraemer, and J.C. Cochran

Overview of Objectives and Approach

For the past four years investigators at the NASA Johnson Space Center (JSC), Purdue University and several other JSC research contractors have shared the computer facility at Purdue's Laboratory for Applications of Remote Sensing (LARS) as their primary research computing environment. The Computer Processing and Data Base Support task has supported JSC's remote sensing research programs by:

* Providing access to a modern computer facility designed and implemented to support remote sensing research needs;

* Providing training in the use of the hardware and software available on the computer system;

* Providing consulting support for the remote users of the facility;

* Developing and acquiring software supporting and enhancing the utilities;

* Demonstrating the benefits that may be derived when geographically dispersed centers working on a common problem share a computational environment.

For each of the past four years usage has more than doubled. As a result, during early 1981 JSC acquired an AS/3000 to augment the computer resources provided by Purdue/LARS. Purdue was contracted to install and maintain on the AS/300, the LARS versions of the VM370/CMS operating system, a number of Purdue developed utility routines and a number of commercially available programs packages.

The objective of the Computer Processing Support Task has been to provide JSC and its associated research community with a shared data processing environment for researching remote sensing. Benefits of this environment include:

* The opportunity to better mold geographically-dispersed research groups into a more informed and integrated research team;

* A mechanism for efficient transfer of information between research centers, NASA, and other participating government agencies;

* Faster, less redundant software development;

* Faster transfer of newly developed analysis techniques and research results to and from participating research groups;

* Effective, streamlined provision of systems programming, data acquisition, data access and specialized services.

Numerous instances where these benefits have been realized have been demonstrated over the three years of experience with this task.

A second major objective during 1981 was to install and maintain a compatible operating system and software environment on the newly acquired AS/3000 computer at JSC. LARS assumed the responsibility for the installation and maintenance of the software environment on the AS/3000 in January of 1981. In addition LARS was to train JSC personnel in the operating procedures on the AS /3000, train systems programmers in system maintenance and assist users in the transition between machines. The AS/3000 at JSC was to be networked with the IBM 4341 at Purdue and a PDP 11/34 at Purdue to form the Earth Resources Data Applications network (ERDAnet).

A third objective of the 1981 program was planning for the integration of data archives and significant software capabilities at Johnson Space Center and Purdue with their Atmospheric and Oceanographic counterparts at Goddard Space Flight Center and the Jet Propulsion Laboratory (JPL). This network of NASA centers was to be the first step in the development of an Application Data Service (ADS) for researchers under NASA's Office of Space and Terrestrial Applications (OSTA).

The Computer Processing and Data Base Support task has a very broad scope. Detailed discussion of subtasks in this report is organized under headings of Hardware Installation and Support, System Maintenance, Software Package Maintenance, Data Base Management, and Application Data Service Development.

To build a suitable environment for the implementation, evaluation and exchange of remote sensing data processing technique, Purdue has concentrated its efforts on providing access to suitable hardware, software utilities and data. Purdue has also provided consulting and training support to foster user knowledge of the tools available on the ERDAnet.

Hardware Installation and Support

The major tasks supported under this heading were the installation of the NAS AS/3000 at JSC, the replacement of the IBM 3705 communications controller at LARS with a Comten 3670, the replacement of the IBM 3031 at LARS with an IBM 4341 and the expansion of the disk system at LARS.

Hardware at JSC

Hardware and Software Benchmark Tests

During the contract year LARS personnel assisted JSC personnel in the preparation of hardware and software benchmark tests. The prepared tests

were run on the LARS 3031 to provide a basis of comparison for the remainder of the tests.

A packaged set of test program was developed to test the hardware system with the type of programs that would be run on the JSC machine after it was delivered. This package was designed to be run by the bidders who need not understand the procedures involved. The test results were written back to tape so they could be evaluated. The package was run on the LARS system to test it and to provide a basis of comparison for the results returned by the bidders.

The test package developed above was run by LARS personnel to test the effects of the operating system and system load on system performance. The test systems were: VM/CMS on idle and busy machines, VM/VS1 on idle and busy machines, and VS1 on an idle machine. The test results indicated that VM/CMS always performed better than VS1 under comparable loading conditions. In addition, VS1 always performed better under VM than when it was run on the real machine. The parameters measured were elapsed time and total CPU time used.

Selection Support

When JSC negotiated the contract with the successful bidder, LARS personnel provided systems expertise to JSC personnel. This support was provided via telephone conversations and by LARS personnel traveling to Houston to attend the meeting in an advisory capacity to JSC personnel.

Installation

After the AS/3000 computer was selected by JSC, LARS personnel generated a VM system for the selected hardware configuration. This system was tested on the LARS system in a virtual machine configured to have the same virtual devices as the AS/3000. The system was delivered to JSC and installed after the AS/3000 was delivered. Other software products were installed on the LARS system and then transferred to the AS/3000 including:

* VM/370 Rel 6.0 and Basic System Extension Program Product (BSEPP 2.0)

* Interactive Problem Control System Extended (IPCS/E)

* Remote Spooling Communication System Networking 2.0 (RSCS 2.0)

* Directory Maintenance Program Product (DIRMAINT 1.0)

* University of Waterloo text processing system (SCRIPT 3.6)

* FORTRAN compiler and library (FORTRAN H (EXTENDED))

On site support for the AS/3000 was provided for three weeks during the installation period. Plans for the support of the LARS supplied software and the LARS support of the vendor software were made during this time. After the computer was installed, the LARS version of the VM/370 operating system was installed and verified by LARS. Program packages were then installed and verified by the responsible product managers from LARS. The

software installation was performed with no difficulties. In fact, the system was installed and operational before the vendor turned the AS/3000 over to JSC personnel.

Systems Training

To enable JSC personnel to support the AS/3000 system, systems training was supplied to JSC three times during the contract year. The first session was provided at LARS and covered the steps necessary to generate and install the CP and CMS components of VM/370. All steps from planning the initial installation of the system to installing corrective service received from IBM were covered.

Additional training was provided in Houston during the installation of the AS/3000. A review of the training provided at LARS and more information on how to maintain the program packages were provided.

An additional training session was provided to JSC personnel at LARS during October. This session was designed to cover many of the same areas as the first two sessions. The major reason for providing this session was because of turnover in previously trained JSC staff.

Operator Training

LARS personnel provided operator training for JSC and contractor personnel. The first session was a one week, hands-on course for the Lockheed operations supervisors. These personnel came to LARS and attended presentations on how LARS operates the system. They then assisted in operation of the system and learned the special procedures required for each of the shifts including system startup and shutdown.

Additional operator training was provided during the installation of the AS/3000. In addition, a LARS shift supervisor traveled to Houston to provide operator training during the first week of AS/3000 operation.

Post-Installation Support

Post-installation support was provided by LARS personnel via trips to Houston to plan system upgrade and maintenance activities. Support was provided via telephone conversations whenever problems occurred that JSC or contractor personnel were not able to handle.

Hardware at LARS

The major hardware efforts at LARS during the past year have been made to reduce the system cost significantly and to provide better support.

IBM 4341 Installation

In July the IBM 3031 at LARS was replaced by an IBM 4341 computer with four megabytes of memory. This change was made to reduce the cost of providing computer service to LARS projects. The 4341 is about 15 percent slower than the IBM 3031. However, the additional two megabytes of memory

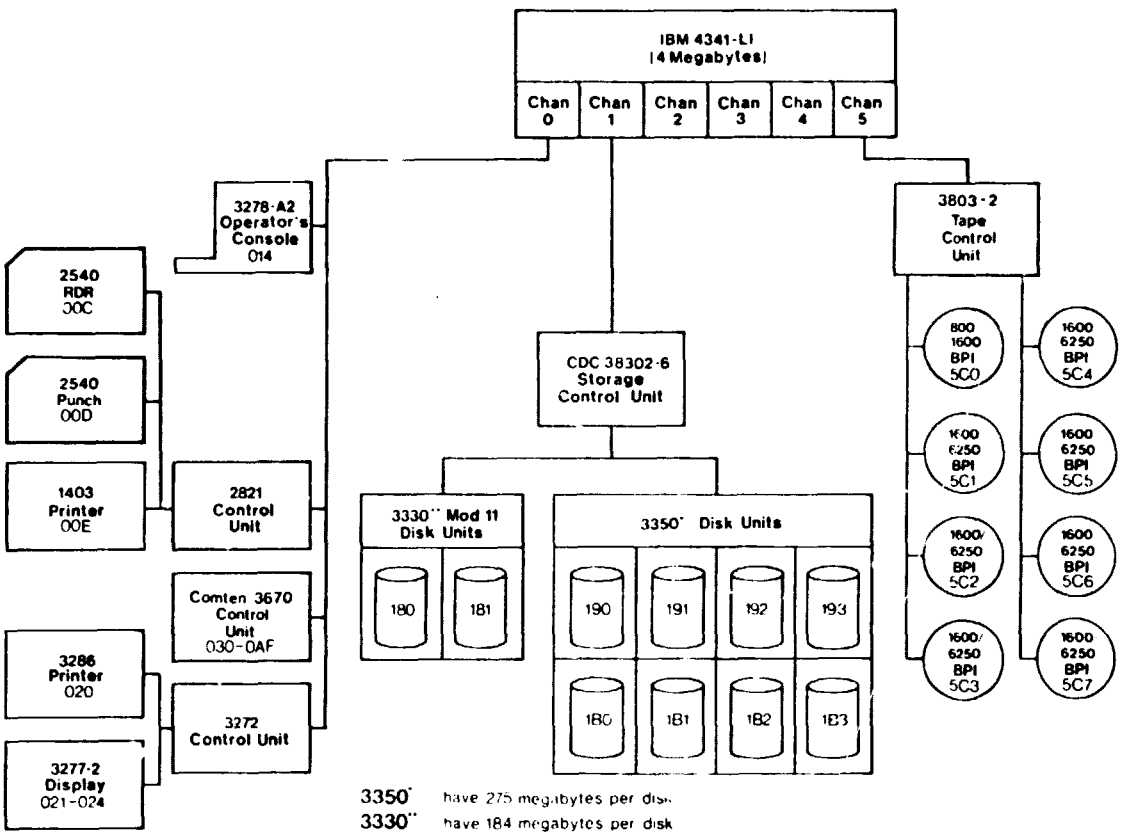


Figure 15-1. The hardware configuration of the LARS computer system. During FY81 the disk space was doubled and the IBM 4341 replaced the IBM 3031.

reduce the paging load on the system. The net result of the change is a machine effectively as powerful as the 3031 at approximately one-third the cost of the 3031. The system configuration is illustrated in Figure 15-1.

Comten 3670

In March, the IBM 3705 communications controller was replaced with an NCR Comten 3670. The Comten 3670 has higher reliability than the IBM 3705 and has a much more powerful set of software available for it.

The current configuration of the Comten includes 256k bytes of memory, two A-MIMs, one B-MIM and two DLC-MIMs for a total line capacity of sixty four asynchronous lines and sixteen synchronous lines.

Disk System Changes

The disk system at LARS was changed by adding an additional CDC 33502 disk storage unit and removing one CDC 3330-11 unit. The IBM 3803 controller and 3350 disk storage unit were also removed. The total capacity of the system is now 2,900 megabytes of on-line storage.

Tape System Changes

The seven track tape drive has been removed from the system. One of the two tape controllers will be removed and the remaining controller will be upgraded to handle 6250 bpi density. There will be one drive with 800/1600 bpi capability and seven drives with 1600/6250 bpi. All eight drives will operate at 125 ips. This will provide a greatly expanded data transfer capability.

This configuration change has already started and is expected to be completed by the end of April, 1982.

System Maintenance

This section discusses the system software changes made on both the LARS and EODL systems.

VM/370 Release 6.0 and BSEPP 2.0

The current version of the operating system for both systems is VM/370 Release 6.0 and BSEPP 2.0 Program Level Change (PLC) 13. Current plans are to install VM/SP (System Product) on the LARS system in the second quarter of FY82. There are no plans to install VM/SP on the JSC system.

RSCS Networking 2.0

The Remote Spooling Communication Subsystem was installed on the LARS system early in the contract year in preparation for the installation of the AS/3000. This program product is a greatly improved version of the standard RSCS delivered with VM/370 Release 6. The system is easier to maintain and has improved function over the standard version.

For the purposes of JSC, the most important new function is the ability to connect IBM and IBM-compatible hosts into a data network. RSCS Networking 2.0 allows users on two or more hosts to transfer files and messages between systems. In addition, users can query the status of the hosts in the network and the status of users on each host.

RSCS Networking 2.0 was installed on the AS/3000 when it was delivered to JSC. One 9600 bps line is dedicated to RSCS for file transfers.

Directory Maintenance

The Directory Maintenance Program Product (DIRMAINT) was installed on both systems during the contract year. This system simplifies the management of the user directory for VM. The system administrator has control of all aspects of the directory including user ID creation and deletion and resource allocation. In addition, users are given the ability to control the detailed configuration of the virtual machine. This includes the ability to change passwords for the ID and minidisks, setting the default memory size, establishing links to minidisks of other users if the passwords are known and setting the user distribution code.

Code was added to DIRMAINT to allow it to send messages to users and administrators using the LARS-developed MAIL system instead of transferring spools files. This prevents unexpected spool files from appearing in the user's virtual reader. A set of programs was developed to format the user directory into reports for the system administrator and generate accounting data for disk usage.

Continuous System Modeling Program

The Continous System Modeling Program (CSMP) is still maintained on LARS system for use by JSC personnel.

LARS Developed Utilities

A large number of LARS developed utilities were installed on the AS/3000 including Accounting, MAIL, SRTNEWS, LISTFILE, TAPOP, GETDISK, enhancements to the EXEC processor, etc. LARS has maintained this code on both systems.

Processing Trouble Reports

The LARS developed problem reporting system (TROUBLE) is used to receive and track trouble reports for both systems. When TROUBLE reports are processed, mail is sent to the submittor acknowledging receipt via the LARS MAIL system. Mail is sent nctifying the user when a workaround or correction is developed. Summaries are sent to the system managers at LARS and JSC weekly giving the number of opened, closed and remaining reports for each system. The summary contains the text of the original TROUBLE report and the status and tracking information that is appended to each report.

Development of HELP Files

In conjunction with JSC personnel, HELP files are being generated that will assist new users in obtaining information about the specifics of the hardware and software configuration for both systems. HELP files are also being generated that will describe the function and administrative policies for both sites.

Development and Installation of BATCH 2.0

The most significant new system software is the LARS-developed BATCH 2.0 system. This system is designed to be much more user friendly and to have better reliability and usability than the previous system. The system is designed to make generating and submitting a batch job very similar to performing the same work interactively. An extensive set of HELP files and interactive aids were developed to assist the user in creating and submitting jobs. Commands exist to allow the user to query the status of jobs and change scheduling parameters of jobs after they have been submitted. Operator commands allow the system operator to control the batch system.

User guides have been written and made available to all users at both sites. Operator guides are also available which describe the operation of BATCH 2.0 and the operator commands that are used to control the system. Functional specifications and installation manuals are also available.

BATCH 2.0 was installed on the LARS system in August for user testing. The last of the old batch system machines were removed from the LARS system on November 1. BATCH 2.0 was installed on the JSC system in September for user testing. Both systems are available at JSC and there is no current schedule for removal of the old system.

Software Package Maintenance

Accomplishments regarding software packages during the past contract year have mainly been in the areas of software maintenance, consulting, and accessibility. These three areas were of high priority in providing software services. Figure 15-2 shows the software packages now resident on the LARS computer.

Software Packages

For supported software packages on the LARS system, documentation in the form of User's Manuals is available in terminal areas at each research site. LARS tests the software packages with the installation of each new package update and after installation of new operating system releases. An individual at LARS has been assigned the responsibility for the maintenance and testing of each package. This individual is also able to serve as a consultant in the characteristics and uses of the packages.

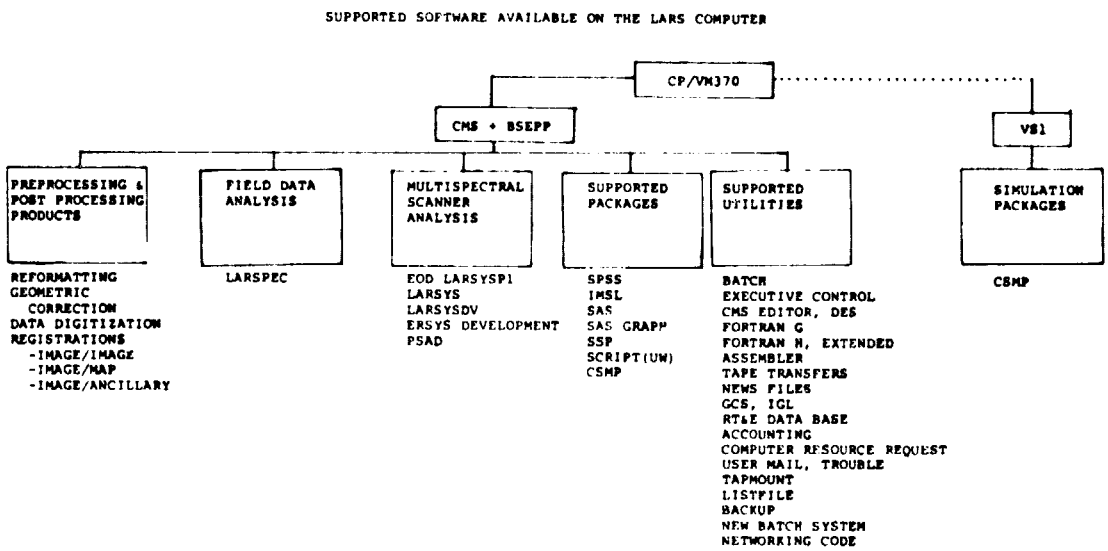


Figure 15-2. Software packages currently available on the LARS computer.

Statistical Software

SAS provides a wide range of statistical procedures, an ability to read complex files, and report writing capabilities. SAS/GRAPH is an interactive computer graphics system for producing high-resolution plots, charts, and graphs including three-dimensional response surface plots. Release 79.5 was installed as a test version in March 1981 and became the default version in July 1981. Use of SAS has grown rapidly this past year. This growth was aided by increasing knowledge of SAS and by training sessions provided at LARS and JSC.

In July SAS became available through the new LARS BATCH system. This, coupled with a rate reduction in April of over 50%, made SAS and SAS/GRAPH much more attractive to use. Also available in January 1981 was a tape/slide module "Introduction to SAS," which provided additional educational benefits.

Edition 8.0, revision 1 of IMSL (International Mathematical and Statistical Library) was received, installed, and tested during the third quarter of FY81.

Use of SPSS, the Statistical Package for Social Sciences has decreased as users have learned about the power and versatility of SAS.

Purdue/LARS has provided a statistical consultant to assist users of SAS, SPSS, and IMSL. Development or conversion of special purpose statistical programs may also be requested. If a user encounters a problem of a statistical nature, he may contact the consultant.

Word Processing

SCRIPT, the University of Waterloo's text processing software, was installed on the LARS computer. In October 1981, version 3.7 (81APR24) was made available. This version contains SCRIPT, SYSPAPER, and SYSPUB. SYSPAPER and SYSPUB are powerful macro sets used to produce form papers, theses, etc. A new SYSLABEL macro package is available in Version 3.7 to process form letters and mailing labels from a driver file and an address file. Also available is processing of variable width character sets as well as other minor additions and changes to the previous SCRIPT package.

Graphics Software

The USMA Graphics Compatibility System (GCS) is a set of FORTRAN callable subroutines providing two-dimensional and three-dimensional graphics capabilities. In general the 2D version is a proper subset of the 3D version but requires less core space and time to execute a job than the 3D system.

Graphics devices currently supported on the LARS/GCS system are:

- Line Printers
- Alpha-numeric terminals
- Varian Electrostatic Printer/Plotter
- Tektronix 4054 graphics terminals

- Tektronix 4002A graphics terminals
- Printronix dot matrix printer/plotter

The latter two devices were added during 1981 while the Decwriter graphics was removed due to lack of need and space.

A major revision to give the GCS system at LARS greater compatibility between the 2D and 3D systems was also made during 1981. This revision involved the inclusion of software obtained from Westinghouse Electric Corporation with expansions such as contour plotting, surface plotting, nineteen additional text fonts, polygonal shading, major/minor tic marks, additional 3D axis generation and viewing, and creation of regularly gridded surfaces from irregularly distributed points.

LARSPEC

The LARS Spectral Analysis System (LARSPEC), has three overall capabilities: printing and punching statistics of wavelength bands; plotting spectral information; printing and punching identification record information.

Various major upgrades were made to the LARSPEC system during 1981. In July a capability was added to allow the plotting of observations with differing wavelength resolutions on the same graph. Also LARSPEC was given greater flexibility to plot any type of wavelength data by using tables, as well as generating wavelength data as in the past. This also allowed correction of wavelength calibration errors in some of the FSS data.

In September the LARSPEC BATCH command was upgraded to use the new LARS Batch system and provide greater flexibility while in the LARSPEC Batch environment.

Finally LARSPEC was interfaced to the new GCS system to give access to the expanded capabilities of GCS as well as providing graphics on the Printronix dot matrix printer.

Future plans for LARSPEC include improvements and expansions to the loading and testing of LARSPEC, to field data base access and querying and to band averaging capabilities

Software Utility Reports

In order to facilitate and manage operations involving multiple sites, LARS has written a few specialized utilities. Tape transfer software has been written to pass the contents of tapes to and from LARS and remote sites. The User Group Accounting System was implemented to allow JSC administrators to flexibly define sub-account (User Groups) and monitor resource consumption for User Groups, Branches and in aggregate. The Resource Request System was initiated to provide a "paper trail" for all computer resource requests (IDs, disks, tapes, etc.) and to keep track of the status of all requests. This system replaced phone calls as the primary means to request directory changes.

Tape Transfer

Operation of the tape transfer software (TAPTRAN) was running smoothly and software modifications had been made to refine messages and information to the user. The capability was expanded to include remote ports other than JSC. In September 1981, the batch machines TAPTRAN and TAPERIM were removed from the system to encourage users to use the new LARS BATCH capability. In order to continue the tape transfer facility, two execs were written, "TSEND" and "TRECEIVE". To send a tape, type: TSEND TAPENO ID. TAPENO is the number of the tape to be sent and ID is the account identifier where the spooled punch file will be sent.

To receive information onto a new tape, type: TRECEIVE TAPENO SPID. TAPENO is the receiving tape to be written on, and SPID is the spool identification number of the tape file which was sent to the receiving ID's reader.

Usage Reporting

The JSC user group reporting system has been revised and is functioning smoothly. Reports are produced weekly and monthly by User Group for each ID. These reports contain detailed information regarding local terminal, tape units, priority service, basic service, and statistical services resources used.

The generation and distribution of all reports is controlled through named execs. On April 1, 1982, entire operational control for these reports was transferred to personnel at JSC.

Resource Request System

The Resource Request System allows users to view request forms and is designed to enable a remote site coordinator to request resources of a LARS computer facility. User ID's may be established or changed, tapes may be purchased, and ring-in authorization may be established for tapes through the resource request system. All of the work involved with this system is to be done online and will likewise be reported to the appropriate personnel through online methods.

The documentation for the system has been updated over the past year to include information for the requestor, the user and the site coordinator. This documentation is available as three separate SCRIPT files on the LIBRARY ID. These files are REQUEST, STATUS, and PERUSE. The REQUEST file contains information for the User and Requestor to request resources of the LARS computer facility. STATUS contains information for the Purdue/LARS site coordinator to act upon the requests that have been made. PERUSE is designed to enable the site coordinator to query any of the requests that have been made to the resource request system.

The PERUSE subsystem was added to the Resource Request System this past year. Individual requests or select groups of requests may be reviewed through this system. This entire subsystem revolves around one command - the "QUERY" command. Also added as a new capability this year is the function of editing a request.

Data Base Management

To support the research needs of the JSC research community over the past year, LARS has acquired additional data sets for the inclusion in JSC data bases. Improving the existing RT&E Data Base, implementing new capabilities, and creating and maintaining identical data bases at LARS and EODL have highlighted this past year. LARS has also inverted large meteorological data sets and created a disk directory and on-line query system. Finally, Purdue was responsible for identifying, evaluating, and selecting a commercial data base management system that satisfied the system and data requirements at JSC.

With the realization that JSC's data management tasks would require frequent data reorganizations and a number of additions of collateral data sets, a decision was made to investigate the acquisition of a Data Base Management System as a joint Purdue/JSC project.

RT&E Data Base

Most of the work done on the RT&E Data Base this year was devoted to adding new data sets to the library, adding new capabilities to existing software, installing the disk directory and current software on the new EODL system, and improving the existing data.

Receive, Verify and Enter Data Sets

The following data sets were received at LARS for inclusion into the RT&E data base:

Crop Year 1980	US
Crop Year 1980	Argentina
Crop Year 1980	Ground Truth
Crop Year 1979	Ground Truth

The addition of these data sets has enabled LARS/EODL users access to approximately 62,000 LACIE sized images (Figure 15-3). The data, stored on high density magnetic tapes, are assigned tape slots in the LARS library. Each tape is tested for its format and readability. After a complete and correct data set has been received, both the RT&E data base and the SUBSET data base must be updated to include the new information. A similar process takes place at the EODL site before the data base update is complete. Finally, the data is released to the user community.

EODL Installation

The RT&E data base and its associated software was extensively tested and successfully installed on the new AS/3000 at JSC in late January of 1981. All software on the AS/3000 is identical to the LARS version. New capabilities added to the data base and its software include the ability to detect which system the user is on, request the proper tape number for that site, search on the JSC tape number, and query on the location or site of the data.

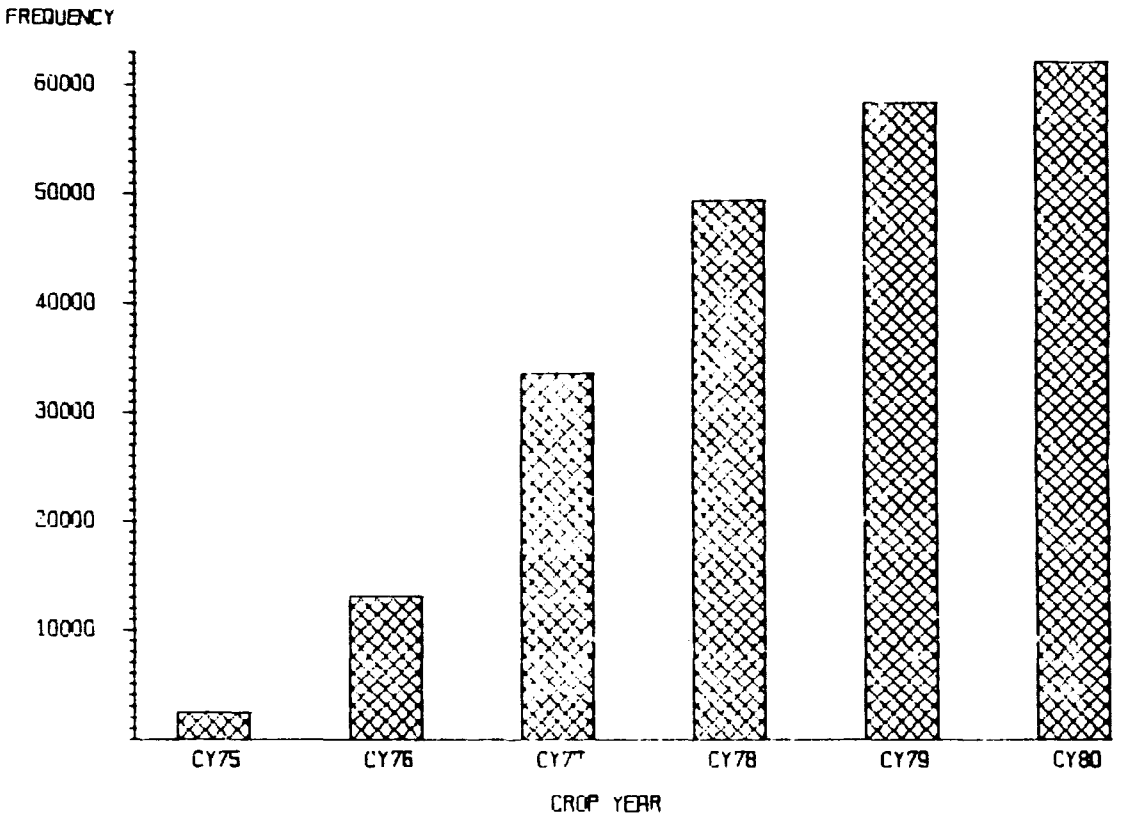


Figure 15-3. Growth (cumulative acquisitions) in LACIE size images of RT&E data base.

Data Base Improvements

Several data base improvements took place over the past fiscal year. A major improvement was the identification of the political location of every segment in the RT&E data base based on its given latitudinal/longitudinal coordinates (Figure 15-4). A numeric code identifies the country where the segment is located. If the segment falls within the United States, the location is further broken down to state, county, and crop reporting district coded identifiers.

A comprehensive study was done on the feasibility of correcting the sun elevation and sun azimuth angles that appear erroneously for various crop years in the RT&E image tape header records. It was concluded that, given the acquisition date and nominal time, and the latitudinal/longitudinal coordinates of the segment center, the algorithms developed at LARS can compute the sun elevation and sun azimuth angles to within 3 degrees of the original value. For the majority of the test cases (using Goddard Registration Data), the computed value was identical to the original value.

LARS deleted acquisitions from the RT&E data base in 1981. Segments were ordered for crop year 1979 and stored at JSC. Later, NASA found the segments were not registered over the requested area. The segments were reordered and added to the original set, but the previous acquisitons were not eliminated. All tapes were transmitted to LARS and entered into the data base. The redundant data was detected and LARS provided NASA with a duplicate acquisition data listing to help identify the retro-orders. LARS later received from NASA a list of duplicates and deleted the acquisitions no longer needed. A tabular report of segments that had locational coordinate changes over crop years was also provided. This information was important for considering renumbering segments that move over time. Currently, there are over 780 such segments.

Software Improvements

Necessary querying capabilities have been incorporated into the system of accessing the RT&E data base directory. Additional programs to report all queried segment information have been written. SEGALL, a Fortran interface subroutine that will return all the information available for the desired acquisitions, has a wall-to-wall ground truth equivalent called GTALL. DETAIL is a RT&E data base program that provides the user with querying capabilities of SUBSET and the detailed description of acquisitions provided with SEGALL. It is designed to run at the terminal. FINDALL is a program that allows the user to query the data base with a segment number and crop year, and output may be routed to the terminal, the virtual printer, or any CMS disk in write access (Figure 15-5). ALLGT is the Ground Truth equivalent to FINDALL.

Additional programs were written to allow for rapid transfer of magnetic tape data between LARS and EODL, utilizing the spooling area as a temporary storage location. TSEND sends the data stream through the network, and TRECEIVE writes to the designated receiving tape.

ORIGINAL PAGE IS
OF POOR QUALITY

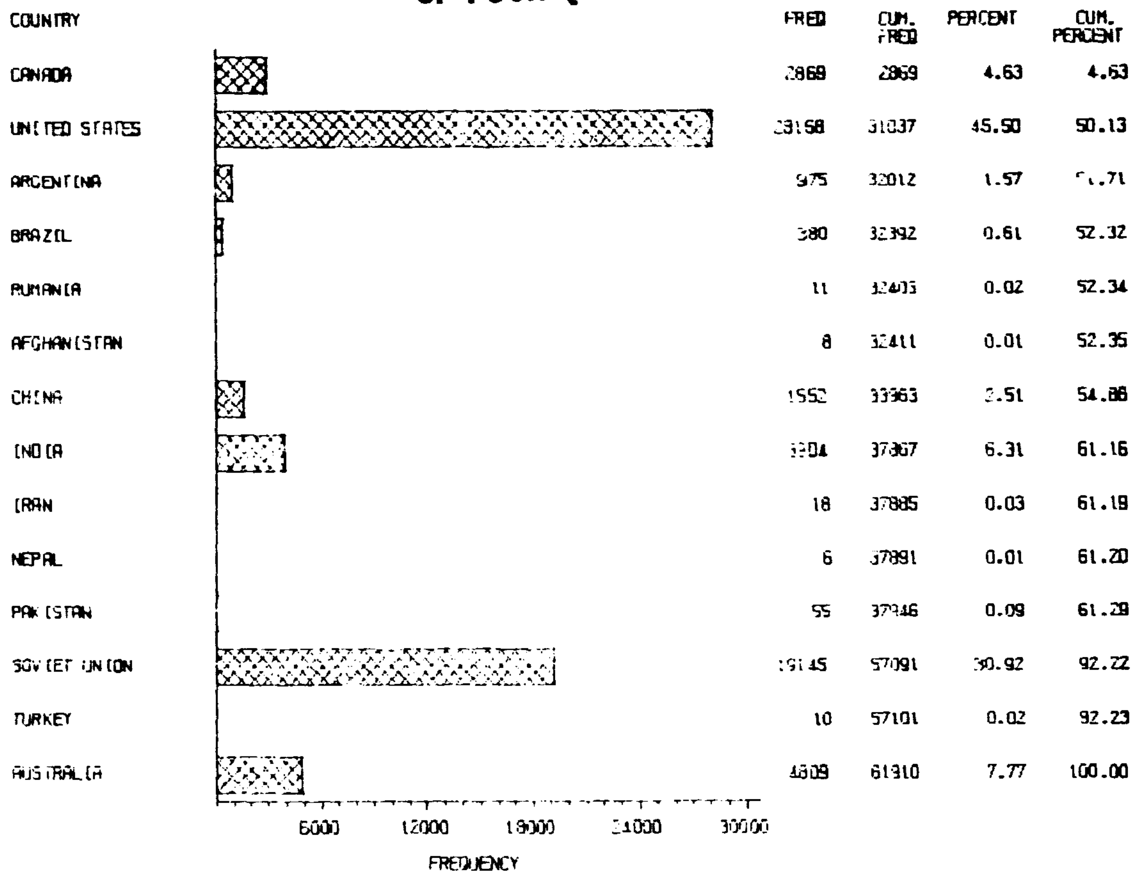


Figure 15-4. Locations by country of segments in the RT&E data base.

```
CONVERT1         LABORATORY FOR APPLICATIONS OF REMOTE SENSING      JAN  15,1982
COCHRAN                          PURDUE UNIVERSITY                    09 29 32 AM

  DATA LOCATED AT:   BOTH                                          CHANNELS:   4
                                    DATE OF COLLECTION:      78089    COLUMNS: 196
  SITE   TAPE FILE DENSITY          TIME OF COLLECTION:  15:22 GMT      LINES: 117
 ------- ----- ---- -------        SEGMENT INITIATED:  12 MAY 80   1ST CHAN.:   1
  LARS   7451  40    1600          ENTRY IN DATABASE:  12 MAY 80  LAST CHAN.:   4
  JSC    3266  40     800          GODDARD PROCESSING:  2 FEB 79     SEGMENT: 127
 ARCHIVE 78090 40     800          DATE OF UNLOADING:   NOT SET      CROP YR:  78
                                    LATITUDE:   40 13 N      LANDSAT NUMBER:   2
  CHANNEL    BIAS    GAIN          LONGITUDE:   86 51 W      SOIL GREENNESS:  -7
  -------    ----    ----             SENSOR:  ERTS MSS     PROCESSING FLAG:   0
    1     -3185.3    9.9               ORBIT:  6467         PEAK SHARPNESS: 0.5
    2     -3251.8    6.3             COUNTRY:  UNITED STATES  SUN ELEVATION:  40
    3        10.3    4.9               STATE:  INDIANA          SUN AZIMUTH: 123
    4         6.7   12.1              COUNTY:  TIPPECANOE       CLOUD COVER:   0

  PEAK TO BACKGROUND RATIO:  1.48       REFERENCE SCENE-FRAME ID:   62351525
     XSTAR HAZE PARAMETER:   0                    SCENE FRAME ID:   61811522
      DATA CLASSIFICATION:   0          CROP REPORTING DISTRICT:          40
```

Figure 15-5. Example output from ⸢ ⸣DALL program which allows user to query the RT&E data base.

Weather Data Base

NASA has archived a large collection of historic meteorological (MET) synoptic data. This data covers world weather readings from 1966 to 1975. The data is stored on date and time keys. Each one-year collection of reports requires 24 magnetic tapes. To query one MET reporting station for an entire year, one would have to search every record on all 24 tapes. There are approximately 11 million reports per year.

LARS has been responsible for converting this data to a specified format. The desired sort keys are, in order: block number, reporting station number, date, and time. Processing of the MET weather data has been completed for the years 1974, 1975 (test case), 1976 and 1977.

Software

The designing of a disk directory and data accessing system has been completed. There are some querying capabilities, and routines that will mount and position data tapes and read desired weather reports.

METRED is a query program that searches the weather data base on block, station, and year, and produces an easily readable report of the results. Subroutine METALL returns the station description and tape file location for a given block and station number of all months in either a specified year or all years available. Subroutine METMNT is used to mount and position a synoptic data tape so the reports for the desired month can be returned to the user. METGET is a subroutine used to return synoptic reports for the block, station, year, and month that was used when calling METMNT. STATESET has been updated to include graphic locational output of state related MET stations (Figure 15-6).

Documentation

All user callable MET data base routines are fully documented internally. User documentation also exists in the form of LARS abstracts. These abstracts have been entered onto the system in SCRIPT and are available to all users on the JSCDISK 39C. The Meterological Data Base Software User's Guide may be accessed with the command, 'SCRIPT METHELP'.

NOAA Co-op Weather Data

There are 173 co-op weather tapes that were ordered from NOAA by NASA that span the years 1878 to 1980. NASA has slotted the tapes and sent duplicate copies to LARS to be slotted. LARS has been responsible for verifying the data for readability and testing its values on the JSC computer. At the same time, information such as yearly unique state station numbers and their tape and starting record location, have been extracted for inclusion in a directory. During the tape verification process, LARS detected 14 erroneous tapes, and has since processed 12 acceptable replacements. LARS has also documented a list of data errors and procedures to correct the data, and has forwarded them to NOAA.

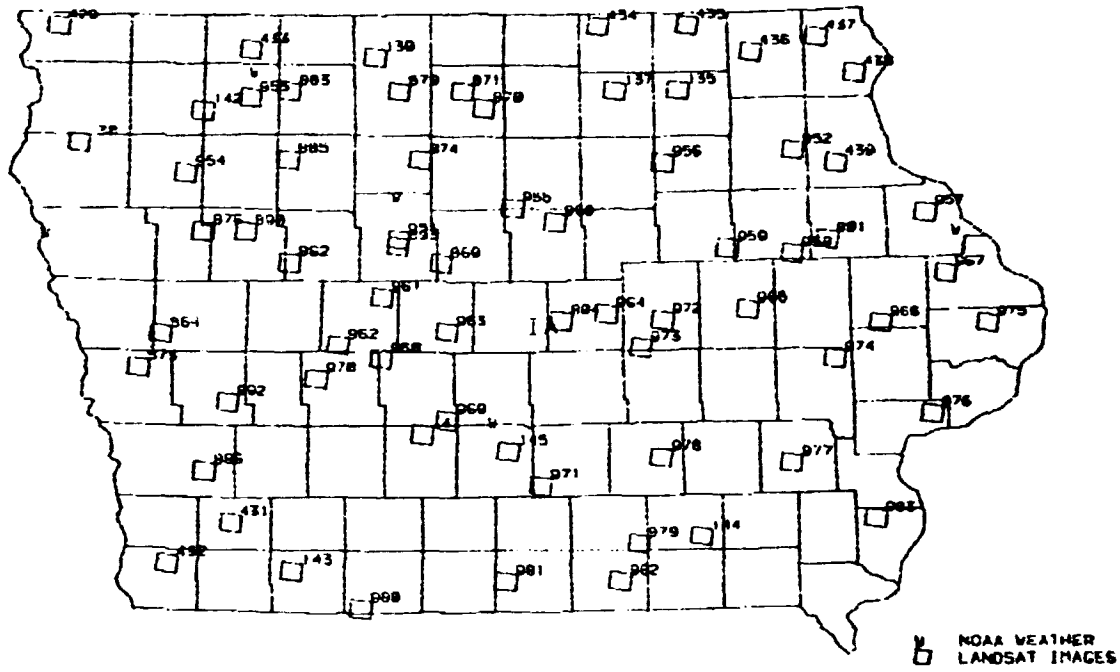


Figure 15-6. Example for Iowa of graphic output of STATESET program for locating MET stations.

Data Base Management System Evaluation

Prior to the Data Base Management System (DBMS) evaluation task, Purdue participated in a data base workshop sponsored by NASA at JSC. Data base experts from industry and government also attended. Following a review of the EOD system and data requirements, each participant discussed his facility's data base environment. At the conclusion of the workshop, EOD decided to investigate further the applicability of commercial DBMS's to their need. Later, Purdue was assigned to evaluate the DBMS market.

LARS began the evaluation by reviewing DATAMATION magazine, COMPUTERWORLD newspaper, and the DATAPRO manual of computer software. Eight vendors were selected and invited to present their product at JSC. The vendors and their products were: Cincom Systems - TOTAL, Cullinane Database Systems - IDMS, Infodata Systems - INQUIRE, Intel Corp. - SYSTEM 2000, International Business Machines - SQL/DS, International Data Base Systems - SEED, Mathematica Products - RAMIS II, and Software AG - ADABAS. Each vendor was allotted four hours for a product introduction and presentation, question and answer period, and an optional demonstration.

Following the presentations, Cullinane, Intel, and Software AG were invited to Purdue University for a more thorough examination of their products. Purdue staff as well as representatives from NASA and LEMSCO participated. Each vendor was allotted one day for his product. In

addition, vendors were encouraged to install their package on the LARS computer. Software AG and Intel loaned LARS a version of their systems for a one month trial period. A written evaluation of the various data base products was prepared in early 1982. Subsequently, ADABASE has been installed on both the IBM 4341 at LARS and the AS/3000 at JSC.

Applications Data Service Development

As hardware and data become less expensive, it is wise to maximize the efficiency of those specialists who can conceptualize and implement new technologies. This was NASA's primary motivation for investigating the development of an Applications Data Service (ADS). ADS was intended to provide a NASA-wide catalog of geographically dispersed data archives together with a network supporting communication, archive interrogation and data transfer. NASA designated three applications programs - Earth Resources, Atmospheres, and Oceans - as pilot testbeds for ADS. These programs are monitored at Johnson Space Center (JSC), Goddard Space Flight Center (GSFC), and the Jet Propulsion Laboratory (JPL), respectively. Each pilot center was to develop and test ADS concepts within the context of its application research. In addition the three ADS pilots were to bring their technical experts together to review progress, establish standards and prepare to interconnect pilot systems so the data could be shared.

For over five years the earth resources research community successfully shared the data processing facilities of Purdue/LARS. Based upon an increased demand for computing and the favorable experience of sharing the LARS facility, the recently acquired AS/3000 at JSC was integrated with the IBM 4341 and DEC PDP 11/34 systems at LARS. Together with terminals at a number of remote research sites, these expanded facilities comprise the Earth Resources Data Applications network (ERDAnet). See Figure 15-7. The ability for users to transfer message and data files between ERDAnet computers and to interrogate the status of remote IBM host computers was implemented using commercially available software.

In addition Purdue University agreed to support the ADS project with the experience gained through its development of an Agricultural Data Network (ADN) for Indiana. ADN allows county extension agents throughout Indiana and agricultural specialists at Purdue University to exchange information and query data bases interactively on over 130 PDP, IBM and CDC computer systems.

During contract year 1981 plans were made for:

1. The development and upgrade of ERDAnet capabilities to conform with those envisioned for ADS.

2. The expansion of ERDAnet to include additional host computers and to be available to additional research sites.

3. The test, development, evaluation and documentation of hardware, software, procedures and techniques for the interconnection of ADS pilot systems.

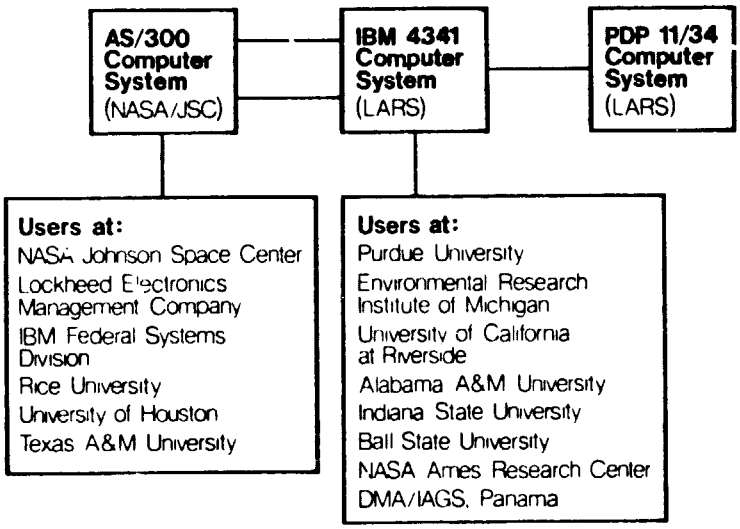


Figure 15-7. The new AS/3000 computer at JSC and the IBM 4341 at LARS were joined to form the Earth Resources Data Applications network (ERDAnet).

In development of these plans LARS participated in several major ADS activities. In May, three Purdue employees attended an ADS standards workshop representing JSC. The survey of system capabilities presented at that workshop greatly understated the accomplishments and experience of the earth resources community. Over the next two months much time was spent documenting the status of and plans for the earth resources data processing environment.

Having the experiences of ERDAnet and ADN to draw from allowed the Earth Resources ADS pilot to propose a method to quickly interconnect the three ADS pilots in order to more rapidly realize the benefits of ADS. Details of this proposal were reviewed in Houston in late August and presented at NASA headquarters in early September. As a result of favorable reaction to the Earth Resources proposal, for ADS, NASA Headquarters scheduled a planning session for early November to make the other NASA centers aware of the requirements to achieve a network.

During early October LARS hosted Phil Yu from GSFC to discuss the networking plan, demonstrate a number of networking capabilities developed at Purdue and review Goddard's progress on their "Transportable Application Executive" and "Remote Services Subsystem."

LARS prepared presentation materials for the November planning meeting. However, due to NASA rebudgeting, JSC's participation in ADS was cut and the planning meeting was cancelled at the end of October.

16. REGISTRATION OF NS-001 SCANNER DATA

B. C. Kozlowski

The NS001 Registration Task involved the registration of seven data acquisitions over test sites in Webster County, Iowa and Cass County, North Dakota. Four sets of Webster County data and three sets of Cass County data were registered to USGS maps with a pixel resolution of 30m x 30m. All data was collected using the NS001 sensor carried aboard aircraft flying at an average altitude of 24,000 feet. Table 16-1 presents the site, date, time, and altitude of each data acquisition.

The data to be registered contained both radiometric and geometric distortions. Radiometric problems were introduced from bad data lines, sun angle effects, poor data quality, and cloud cover; geometric problems included along track sampling rate, scanner view angle, site coverage, and stability of scanner platform. Preprocessing corrected for bad data lines by substitution of the nearest good lines; for sun angle effects by adjustment of radiometric values according to mean angle responses; and for along track sampling rate ("double scanning") by dropping every other scan line.

Grayscale plots of the data sets after the above corrections were applied are presented in Figures 16-1 and 16-2. No corrections were made for the cloud cover present in two of the data sets (Webster Co. 7/1/80 and Cass Co. 6/17/80) nor for the poor data quality in the 2.08-2.35 micrometer band. No corrections were possible for poor site coverage. Particularly in the Webster Co. 7/1/80, 8/6/80, 9/10/80 and Cass Co. 6/17/80 data sets the aircraft flying to one side of the site caused extreme scanner view angle distortions (Figure 16-1). By far the most severe geometric distortions were due to the lack of stability in the scanner platform. Roll and pitch of the aircraft during data collection caused high frequency distortions in all the data sets. Roll caused the across track distortion noticable in north/south roads "jittering" left to right. Pitch caused across track distortions making some features appear longer or shorter in the north/south direction than they actually were. Yaw caused the greatest distortion evident in the curvature of linear features (roads) and the nonpreservation of orthogonal relationships (road intersections).

Geometric distortions were removed through a multiple step process. First, where the roll and pitch distortions were non-systematic and high frequency, corrects were performed. Road "jitter" was removed by straightening an along tract feature known to be linear (e.g. a north/south road which is straight on the USGS map). Where noticable distortions occurred in along track distances (due to changes in aircraft pitch attitude), scan lines were resampled. After these processes were completed, the image was segmented into blocks and a separate registration function calculated, then applied to each block. The registration equation selected was linear, quadratic, or cubic depending on the characteristics of the

particular image. This approach was necessary to compensate for continued presence of non-systematic, high frequency distortions in the image.

The Webster County 9/10/80 data (Figure 16-1) was severely distorted geometrically and is a good example for illustration of the entire registration process. Figure 16-3 shows the data as received. The data after preprocessing to remove bad data lines and sun angle effects is illustrated in Figure 16-4. Figure 16-5 shows the data after a north/south road was straightened to compensate for aircraft roll and after scan lines were resampled to compensate for aircraft pitch. The final registered data set is shown in Figure 16-6. For the registration, a quadratic function was used based on a least squares fit of 27 checkpoints per block. Checkpoints were taken from quarter sections in the data and on USGS maps. Evaluation of the registration accuracy is under way at this time.

Software written for this task included programs for the IBM 4341 and the PDP 11/34. For the PDP 11/34 routines were written to transfer graphics data between the PDP and the Comtal Vision One/20 color display. For the IBM machine, programs were written to convert Comtal graphics points to data set X,Y coordinates. These X,Y coordinates became input for the 'straightening' routine and for the block registration routine, both written specifically for this task.

Table 16-1. Data sets registered of aircraft scanner data.

Site	LARS Run Number	Date	Time	Altitude
				feet
Webster Co., Iowa	80002702	7/01/80	21:19	24,000
	80002902	8/06/80	16:06	23,000
	80003102	8/19/80	16 37	24,000
	80003302	9/10/80	16:28	23,900
Cass Co., North Dakota	80003602	5/15/80	15:05	24,000
	80003802	6/17/80	18:32	25,000
	80004002	8/13/80	16:04	25,000

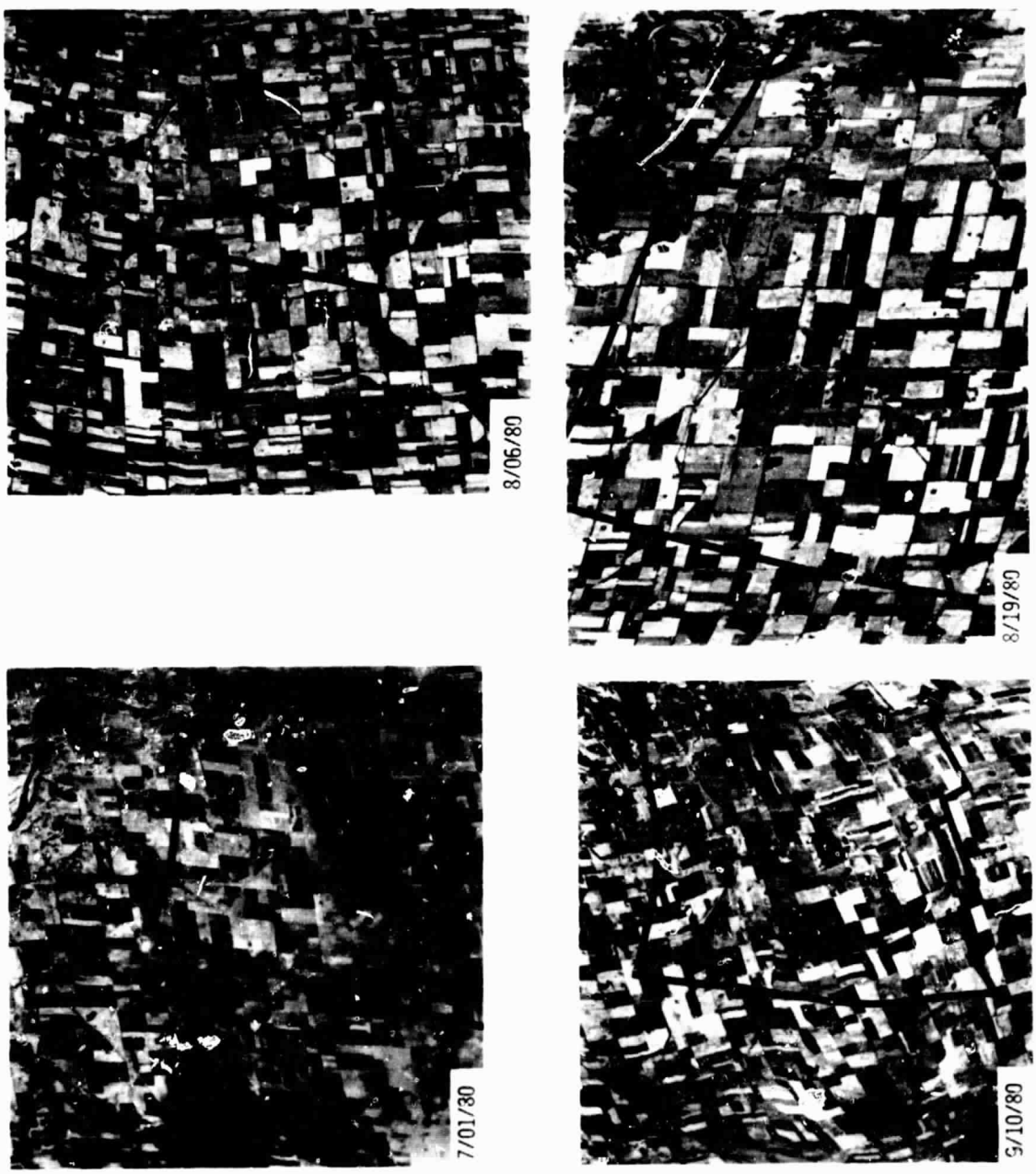


Figure 16-1. Four dates of unregistered NS001 scanner data for Webster County, Iowa test site (1.0-1.3 μm wavelength band).

Figure 16-2. Three dates of unregistered NS001 scanner data for Cass County, North Dakota test site (1.0-1.3 μm wavelength band).

Figure 16-3. Webster County scanner data (1.0-1.3 μm wavelength band) for September 10, 1980 as received; every sixth scan line and third sample (column) shown.

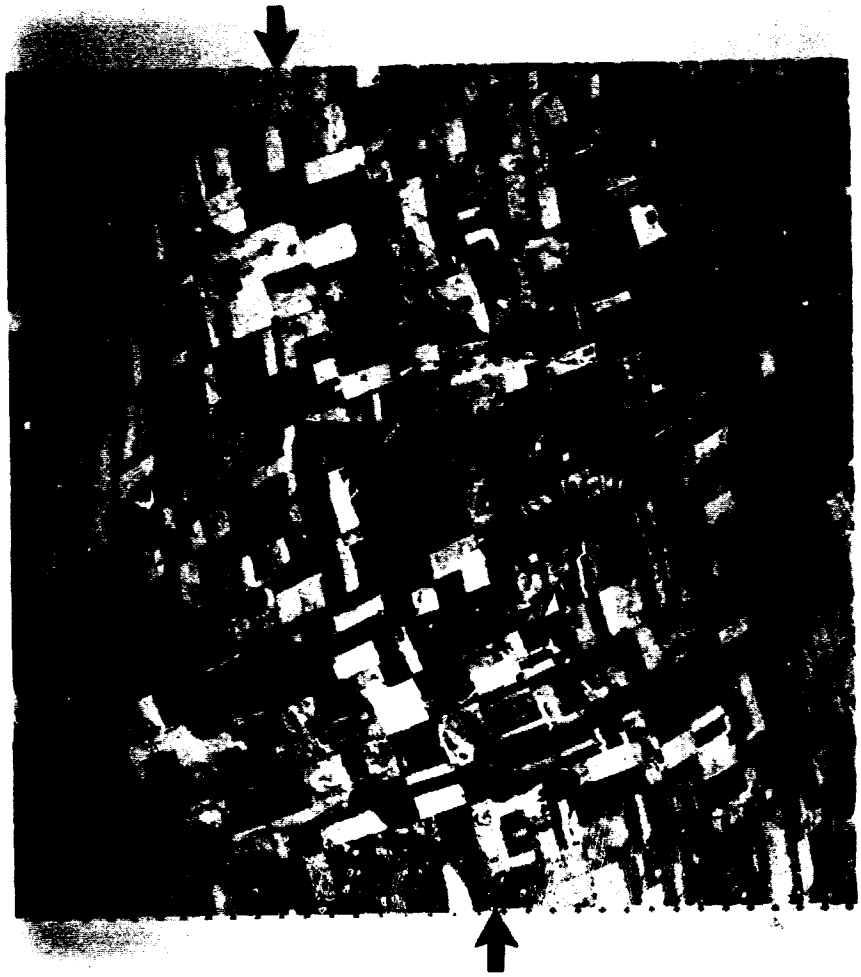


Figure 16-4. Same data as in Figure 16-3 after sun angle correction; arrow indicates road to be straightened.

C-3

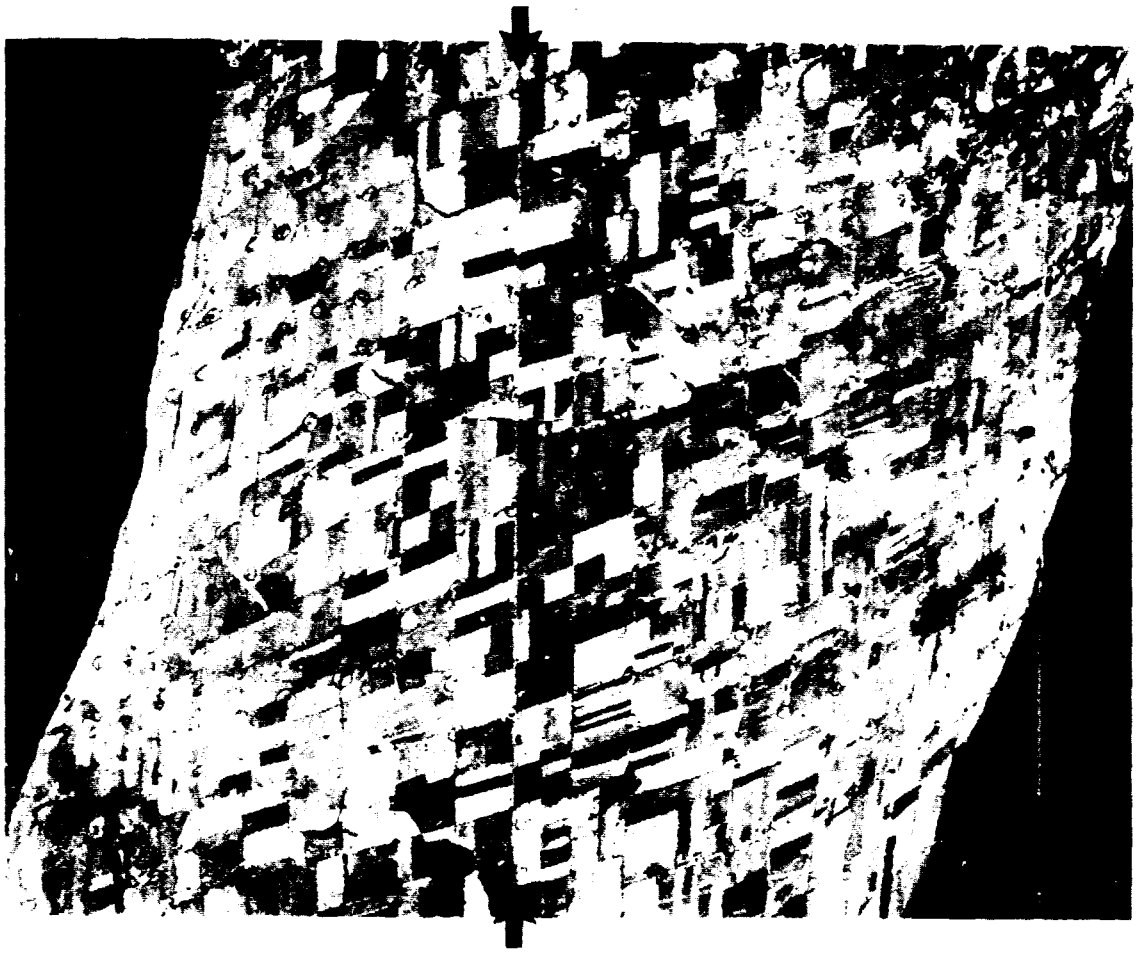


Figure 16-5. Same data as in Figure 16-4 after a north-south road was straightened to compensate for aircraft roll and after scan lines were resampled to compensate for aircraft pitch.

Figure 16-6. Geometrically and radiometrically corrected and registered data for September 9, 1980 (1.0-1.3 μm wavelength band).