AN ASSESSMENT OF LANDSAT DATA ACQUISITION HISTORY ON IDENTIFICATION AND AREA ESTIMATION OF CORN AND SOYBEANS

MARILYN M. HIXSON AND MARVIN E. BAUER
Purdue University/LARS

I. ABSTRACT

During the past decade, numerous studies have demonstrated the potential of satellite remote sensing for providing accurate and timely crop area information. This study assessed the impact of Landsat data acquisition history on classification and area estimation accuracy of corn and soybeans.

Multitemporally registered Landsat MSS data from four acquisitions during the 1978 growing season were used in classification of eight sample segments in the U.S. Corn Belt. The results illustrate the importance of selecting Landsat acquisitions based on spectral differences in crops at certain growth stages.

II. INTRODUCTION

Accurate and timely crop production information is a critical need in today's economy. During the past decade, satellite remote sensing has been increasingly recognized as a means for crop identification and estimation of crop areas.

An extensive experiment, the Large Area Crop Inventory Experiment (LACIE), was conducted by NASA, USDA, and NOAA during 1974 through 1977 [1]. The purpose of LACIE was to assimilate current remote sensing technology into an experimental system and evaluate its potential for determining the production of wheat in various regions of the world. In LACIE, area estimates were made from classifications of Landsat MSS data. Five by six nautical mile samples representing about two percent of the agricultural land area were selected for analysis to estimate wheat area. Segments were allocated to political units according to the historical area of wheat. The sample segments were used both for training the classifier and for aggregation to obtain area estimates. Data from four Landsat acquisitions were used in the analysis, if available. The LACIE method was generally successful in obtaining unbiased and precise area estimates.

Several investigations have shown that the potential also exists for identification and area estimation of corn and soybeans [2,3,4,5]. In one such investigation, a systematic sample of pixels spread throughout a Landsat full-frame was classified and used to make area estimates, while training data were obtained separately [2]. The pixel sampling approach was demonstrated to have the capability to produce unbiased and precise area estimates for small (e.g., county) as well as large (e.g., state) geographic areas. In these investigations, from one to four acquisitions of Landsat MSS data were used in the classifications.

The goal of any estimation procedure is to obtain an estimate which is both unbiased and precise. Numerous aspects of the crop inventory problem using remote sensing may affect the bias and precision of the estimates. Choices involving the spectral features to be measured, the sensor to be utilized, the timing of the crop observation, and the analysis methods used are all important aspects to be considered in the design of a remote sensing system. This study examines some temporal aspects of utilizing Landsat MSS data to estimate corn and soybean areas.

III. OBJECTIVES

The overall objective of this study was to assess the impact of Landsat data acquisition history on classification and area estimation accuracy of corn and soybeans. Specific objectives were to:

1. Assess the accuracy of early

CH1533-9/80/0000-0072 $00.75

Reprinted from MACHINE PROCESSING OF REMOTELY SENSED DATA, June 1980

season estimates.
2. Determine the minimum number and distribution of acquisitions necessary for accurate identification and area estimation of corn and soybeans.
3. Determine the difference in accuracy obtained using a subset of channels rather than all channels in both unitemporal and multitemporal classifications.

IV. APPROACH

Multitemporally registered Landsat-2 and -3 MSS data acquired over the U. S. Corn Belt during the summer of 1978 were analyzed. The data set consisted of eight sample segments, each 5 x 6 nautical miles in size. The locations of the test sites were selected to represent a broad range of conditions found in the Corn Belt. Two segments each were located in eastern Indiana, western Indiana, north central Iowa, and west central Iowa.

Aerial photography was acquired over the test areas and a wall-to-wall inventory of crop types in each site was subsequently conducted. Four data acquisition windows were defined based on the corn growth stage and high quality Landsat data had to be available in each of those time periods. The four time periods were: (1) preplant to eight leaves, (2) 10 leaves to tassel, (3) tassel to beginning dent, and (4) dent to mature.

A systematic sample of the inventory data was used for training and testing the classifier. The pixel at every tenth line and column of the Landsat data was examined. If that pixel fell in a field, the cover type in the field was identified from the ground inventory. The fields selected by this procedure were randomly assigned for either training the classifier or testing classification accuracy. From those fields selected for training, three sets of data were clustered: all fields of corn, all fields of soybeans, and all fields of other cover types. This procedure insures "pure" cluster classes (i.e., clusters containing pixels from only one cover type).

After refinement of the statistics was complete, the entire segment was classified using three different classification algorithms:

(1) CLASSIFYPOINTS, a per point Gaussian maximum likelihood classifier. It is a processor from LARSYS, a remote sensing data analysis system developed at LARS [6].

(2) CLASSIFY, a sum-of-normal-densities maximum likelihood classification rule which first assigns each pixel into an information category and then assigns the pixel to a spectral subclass within that category. It is a processor from EODLARSYS, developed at NASA, Johnson Space Center [7].

(3) MINIMUM DISTANCE, a linear classifier from LARSYS which assigns each pixel to the class whose mean is closest in Euclidean distance [8].

The difference in overall classification accuracies for the different classifiers was statistically significant, although most of the performances were within about 2% for all classifiers.

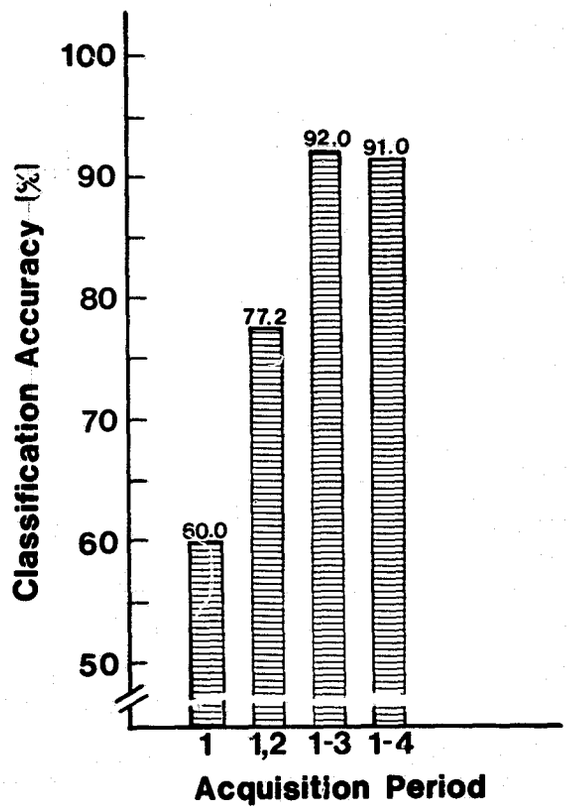

Figure 1. Overall classification performance using cumulative spectral information with a minimum distance classifier and subsets of two, four, six, and eight channels.

All possible combinations of time periods were analyzed. In multitemporal analyses using four Landsat acquisitions, prior studies have demonstrated that the use of eight wavelength bands yields classification results as accurate as using all 16 bands [9]. One visible (.6-.7μm) and one near infrared (.8-1.1μm) band were initially selected for the multidate analyses. A subset of four bands, selected from the available six or eight bands on the basis of the maximum transformed divergence value, was also used for classification in three or four date analyses.

V. RESULTS AND DISCUSSION

A. EARLY SEASON ESTIMATE ACCURACY

Classification accuracy was computed based on test field performance. The accuracies of estimates made using cumulative spectral information through the growing season are illustrated in Figure 1. Corn and soybeans were not spectrally separable using data from the first time period alone. In the Corn Belt, however, relatively accurate identification can be made of corn and soybeans together at that time. Over the same set of segments, it was found that overall accuracy classifying into two classes (corn-soybeans and other) was 92% correct, while the three-class classification (corn, soybeans, and other) was only 60% correct. The area estimates for total corn and soybeans were generally close to ground inventory estimates (Figure 2).

Consistently high classification accuracies were not obtained until an acquisition after the corn had tasseled (growth stage three) was included in the analysis. The classification accuracy did not improve by using later season information when the crops of interest had reached maturity.

B. MINIMAL ACQUISITIONS NECESSARY

Figure 3 illustrates the overall crop identification accuracies of classifications using acquisitions from two, three, and four different time periods. A significant decrease in accuracy can be noted when the third period, tasseling to early dent of corn, is omitted from the three-date analyses. The importance of an acquisition from this time period can also be seen in examination of the two acqui-

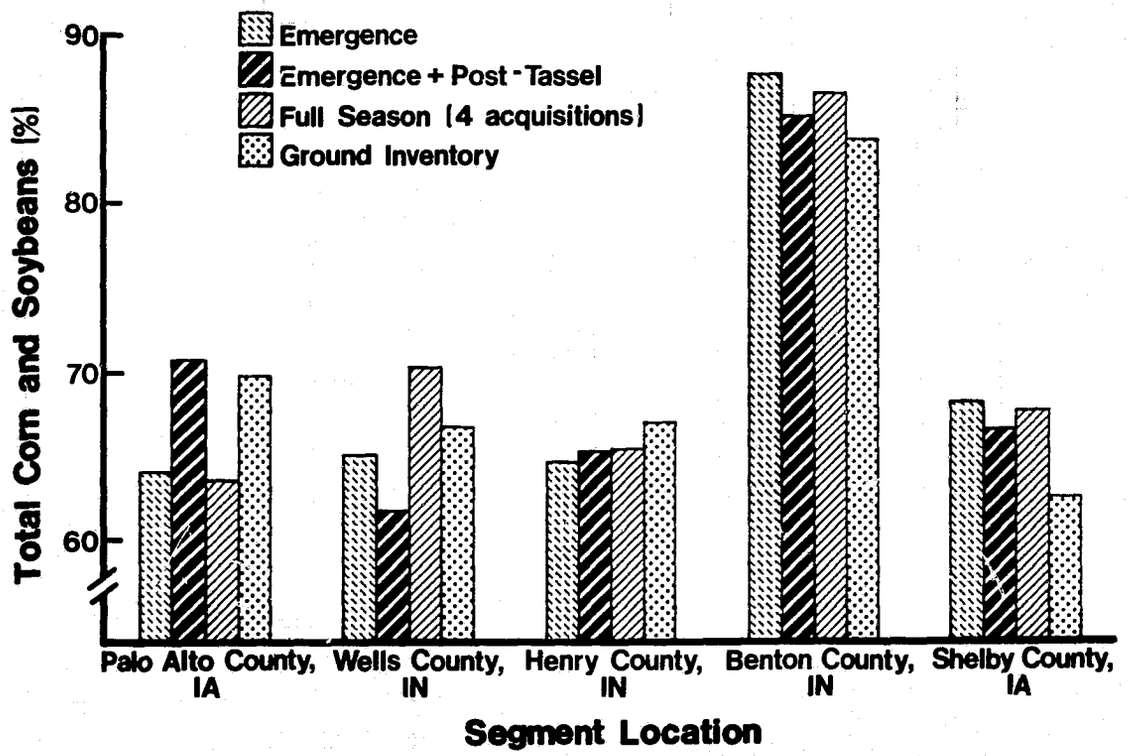

Figure 2. Comparison of classification estimates of total corn and soybean areas with ground inventory proportions.

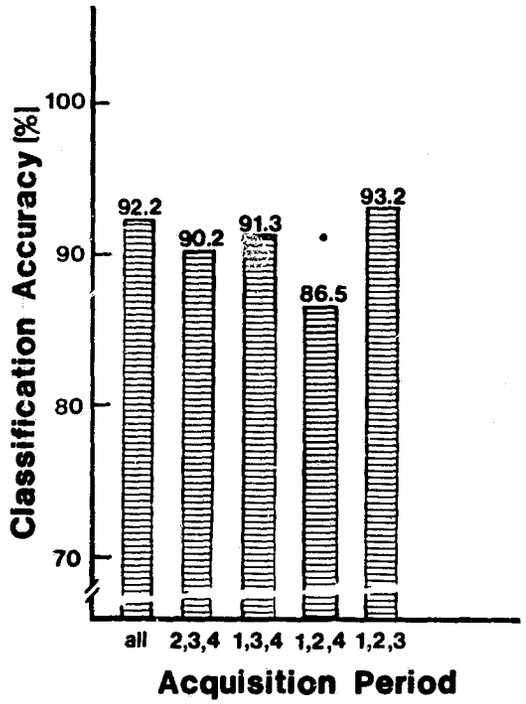

Figure 3a. Overall classification accuracies of three and four date classifications.

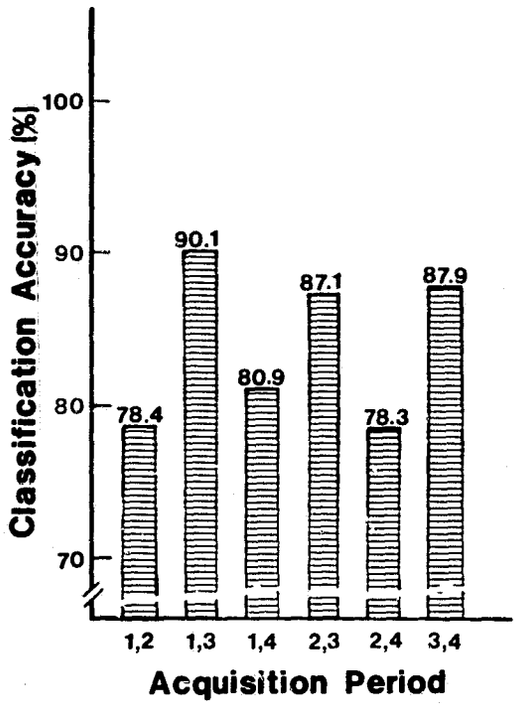

Figure 3b. Overall classification accuracies of two date classifications.

sition analyses: the three combinations using the third time period resulted in higher overall accuracies than those without that time period represented. The overall accuracy for the third period alone was only 85%, illustrating that classifications using the single best acquisition period are not as accurate as can be obtained using multitemporal spectral measurements.

The following combinations of acquisition periods had overall classification accuracies which were not substantially different: 1,2,3,4; 2,3,4; 1,2,3; 1,3,4; and 1,3. These acquisition period combinations had a range of overall accuracies of 3% while the next highest accuracy was about 3% lower than the lowest of these. These results show that acquisitions from time periods one (about emergence) and three (after tasseling of the corn) provide a minimal set for accurate identification of corn and soybeans. No combination of acquisitions which does not include period three gave as high classification performance; an acquisition from period one appears to be less critical if acquisitions from all the other periods are available. Proportion estimates from this minimal set were generally close to ground inventory proportions (Figure 2).

C. SPECTRAL BAND SELECTION

Landsat MSS channels two (.6-.7μm) and four (.8-1.1μm) from each acquisition (six for three date and eight for four date analyses) were compared with the best subset of four channels selected on the basis of the maximum transformed divergence value. The differences in accuracy and variance reduction factors were significant and, in general, use of all even numbered channels gave higher classification performances than the use of a subset of four channels (Table 1). On the average, differences were relatively small (0-5%). A large variability, however, could make loss in accuracy for a given segment with a particular combination of acquisitions be quite large (one value of 10.7% was observed). A few cases, where the subset of four channels performed better, were attributed to better defined training statistics resulting from the dimensionality reduction of the estimation problem or bad data in the omitted bands.

Table 1. Overall Classification Accuracies (percent) Obtained by the Maximum Likelihood Classifier for All Even Channels and a Subset of Four Channels (Average of Eight Segments).

Time Periods Analyzed	Subset	Even Channels	Mean Difference	Maximum Difference
1,2,3	91.2	93.6	2.4	5.5
1,2,4	86.5	86.7	0.2	-2.5
1,3,4	88.2	91.6	3.4	7.6
2,3,4	85.4	90.2	4.8	10.7
1,2,3,4	89.2	92.1	1.9	9.0

Single date classifications were conducted using two and four bands. Analyses estimating area of the two crops were not conducted using acquisitions from the first and second time periods individually, so these two time periods were not assessed. In acquisition period three, no significant differences in accuracy were found over all segments (83.1% vs. 83.0% overall accuracy). On an individual segment basis, there was a tendency (six of eight cases) for all channels to perform better. For acquisition period four alone, the even channels gave 4% higher overall accuracy on the average, keeping this trend for four of the six available segments.

VI. CONCLUSIONS

The results of this study illustrate the importance of selecting Landsat acquisitions based on spectral differences in crops at certain growth stages. Use of a Landsat acquisition when corn has tasseled is critical, as this is the single optimal time for separation of corn and soybeans. In addition, an early season acquisition when the summer crops appear as bare soil can be beneficial in reducing the confusion between these two crops of interest and other cover types. Additional Landsat acquisitions seem to provide only a marginal amount of information for corn and soybean separability.

All available wavelength bands need not be used in the analysis. A subset of one visible and one infrared band from each date was found to produce results not significantly different from the use of all bands. Selection of a subset of these bands may also be feasible for multitemporal analysis.

VII. REFERENCES

1. MacDonald, R. B., and F. G. Hall. 1978. LACIE: An Experiment in Global Crop Forecasting. Proc. Plenary Session, The LACIE Symp., Houston, Texas, October 23-26, pp. 17-48.

2. 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.

3. Bauer, Marvin E., Jan E. Cipra, Paul E. Anuta, and Jeanne B. Etheridge. 1979. Identification and Area Estimation of Agricultural Crops by Computer Classification of Landsat MSS Data. Remote Sensing of Environ. 8:77-92.

4. Hanuschak, George, Richard Sigman, Michael Craig, Martin Ozga, Raymond Luebbe, Paul Cook, David Kleweno, and Charles Miller. 1979. Obtaining Timely Crop Area Estimates Using Ground-Gathered and Landsat Data. U.S. Dept. of Agriculture, Economics, Statistics, and Cooperatives Service. Technical Bulletin No. 1609.

5. Scholz, Donna, Nancy Fuhs, and Marilyn Hixson. 1979. An Evaluation of Several Different Classification Schemes: Their Parameters and Performance. Proc. Thirteenth Int. Symp. Remote Sens. Environ., Ann Arbor, Michigan, April 23-27.

6. Phillips, T.L. (ed.). 1973. LARSYS User's Manual. Laboratory for Applications of Remote Sensing, Purdue University,

West Lafayette, Indiana.

7. Stewart, J., and P. J. Aucoin. 1978. Earth Observations Division Version of the Laboratory for Applications of Remote Sensing System (EOD-LARSYS) User Guide for the IBM 370/148 - Volume I. System Overview. NASA, Johnson Space Center, Houston, Texas. JSC-13821.

8. Nilsson, N. J. 1965. Learning Machines: Foundations of Trainable Pattern-Classifying Systems. McGraw-Hill Book Co., New York.

9. Daughtry, C.S.T., and M. M. Hixson. 1979. Final Report, Vol. II, Processing Techniques Development, Part I: Crop Inventory Techniques. Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana. LARS Contract Report 112979.

VIII. ACKNOWLEDGEMENTS

This research was sponsored by the NASA Johnson Space Center Contract NAS9-15466.

Marvin E. Bauer, research agronomist and program leader of LARS'Crop Inventory Systems Research; B.S.A. and M.S., Purdue University; and Ph.D., University of Illinois. Dr. Bauer has had key roles in the design, implementation, and analysis phases of several major remote sensing projects including the 1971 Corn Blight Watch Experiment and the Large Area Crop Inventory Experiment. He has been the principal investigator of a Landsat investigation for crop area estimation survey. Currently he is the technical leader of the agricultural field research program at LARS.

Marilyn M. Hixson, research statistician in LARS' Crop Inventory Systems Research; B.S. in mathematics from Miami University; M.S. in mathematical statistics from Purdue University. Ms. Hixson's work at LARS has involved experiment design, data analysis, stratification, and sampling methodology. She has had a major role in the design, Landsat data classifications, and statistical analysis of results in several Landsat investigations concerning training, classification, and area estimation procedures for crop inventory. Her current work also involves sampling and aggregation approaches for AgRISTARS.

Donna K. Scholz, B.S., Geology, Indiana University, 1973, M.S., Geosciences, Purdue University, 1979, is an experienced data analyst in geology, forestry, and land use. Her responsibilities at LARS included data acquisition, quality assessment, and use of pictorial digital data. She organized the LARS data storage and retrieval system, trained new data analysts in LARSYS analysis techniques, developed and supervised the LARS monthly Short Course computer "Hands-On" project for visiting scientists. She is principal author of an instructional module "Remote Sensing for Mineral Exploration." In February 1980, she accepted a position at EROS Data Center with responsibility for the IDIMS system.