E82-10067

CR-167400

AgRISTARS

FC-P1-04121
NAS9-15466

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

Foreign Commodity Production Forecasting

June 1981

Evaluation of a Segment-Based Landsat Full-Frame Approach to Crop Area Estimation

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

(E82-10067) EVALUATION OF A SEGMENT-BASED LANDSAT FULL-FRAME APPROACH TC CORP AREA ESTIMATION (Purdue Univ.) 22 p HC A02/MF A01 CSCL 02C

N82-20590

Unclas

G3/43 00067

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

FC-P1-04121
NAS9-15466
LARS 062381

Technical Report

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

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

Purdue University
Laboratory for Applications of Remote Sensing
West Lafayette, Indiana 47906

June 1981

Star Information Form

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FC-P1-04121		

4. Title and Subtitle	5. Report Date
EVALUATION OF A SEGMENT-BASED LANDSAT FULL-FRAME APPROACH TO CROP AREA ESTIMATION	June 23, 1981
	6. Performing Organization Code
7. Author(s)	8. Performing Organization Report No.
M.M. Hixson, S.M. Davis, and M.E. Bauer	062381
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 Observations Division Houston, Texas 77058	Technical Report
	14. Sponsoring Agency Code

15. Supplementary Notes

Principal Investigator - M.E. Bauer
Technical Monitor

16. Abstract

A sampling approach using cluster samples (segments) was used in LACIE to select Landsat data for classification and estimation. The selection of this sampling scheme was partially driven by the data registration technology available. As the registration of Landsat full frames enters the realm of current technology, sampling methods which utilize other than segment data should be examined.

The objective was to assess the effect of separating the functions of sampling for training and sampling for area estimation. The frame selected for analysis was acquired over north central Iowa on August 9, 1978. A stratification of the full-frame was defined. Training data came from segments within the frame.

Two classification and estimation procedures were compared: (1) statistics developed on one segment were used to classify that segment and (2) pooled statistics from the segments were used to classify a systematic sample of pixels.

Comparisons to USDA/ESCS estimates illustrate that the full-frame sampling approach can provide accurate and precise area estimates.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
Landsat, crop inventory, sampling, stratification, full-frame approach	

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

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

NASA - JSC

ORIGINAL PAGE IS
OF POOR QUALITY

I. 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. The Landsat multispectral scanner (MSS) records as a single data point (pixel) a region on the ground about one acre (0.5 ha) in size. When estimates of crop areas are desired for large regions, a statistical sampling scheme is required as it is not feasible to examine all of the pixels in the region of interest. The development of a sampling strategy which is both efficient and cost-effective is thus an important objective.

An extensive experiment, the Large Area Crop Inventory Experiment (LACIE), was conducted by NASA, the USDA, and NOAA from 1974 through 1977 (1). Its data analysis objective was to distinguish small grains from nonsmall grains using Landsat MSS data. Several other investigations have shown that the potential exists for identification and area estimation of corn and soybeans as well (2,3,4,5).

The LACIE area estimation system was based on analysis of sample segments or cluster samples (each 5 x 6 nm in size) extracted from multidate Landsat data. The selection of this sampling scheme was driven to a large degree by the data registration technology which was available at that time. Registration technology research has made considerable progress toward an operational registration capability for Landsat MSS full frames, and so we are no longer restricted to sampling small geographic regions, each of which has been separately registered. This allows us to examine the sampling efficiencies which may be introduced by using a smaller sampling unit size distributed over a larger geographic area.

One such sampling scheme, described by Bauer et al. (2), separates the functions of sampling for training and sampling for classification and area estimation. Training data were developed by photointerpretation of aerial photography taken along north-south flightlines located at intervals across the area of interest. For classification and crop area estimation, a systematic sample of pixels distributed throughout the region was used. The use of different sampling units for training and classification provides both convenience for the data analyst and high precision of the resulting area estimates.

II. 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 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 or a systematic sample of pixels are used for estimation?

III. APPROACH

The data set available for this study was acquired over the U.S. corn and soybean production region by NASA during the 1978 crop season. For the LACIE-type sample segments (5 x 6 nm in size), Landsat data included multitemporally registered MSS data and film writer imagery (PFC Product 1) for each acquisition and segment. Color infrared prints of aerial photography with ground inventory overlays were also used. For a subset of the segments, these inventories were also available in digital format. In addition, single-date Landsat MSS frames were acquired over several sites where segments were located.

The Landsat frame selected for analysis was acquired over north central Iowa (Figure 1) on August 9, 1978, during the best time period for identification of corn and soybeans with unitemporal data (6). 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.

The data analysis procedure consisted of first defining a stratification of the full-frame. The stratification schemes considered were: (a) using the refined strata developed by NASA/JSC based on agrophysical characteristics observable from Landsat imagery such as soil type and field

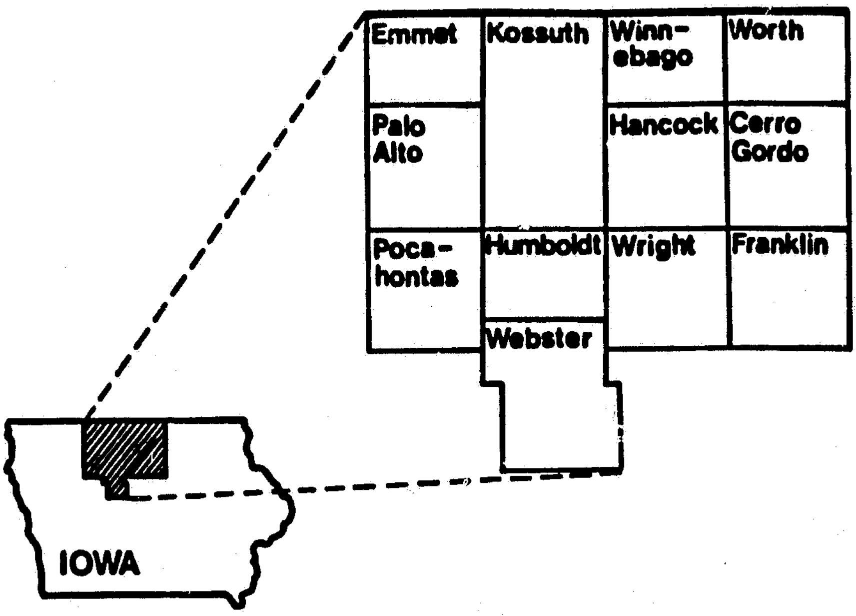

Figure 1. Twelve-county study area in north-central Iowa.

size, (b) using a subdivision of these strata to provide strata with more homogeneous yields proposed for the USDA AgRISTARS yield modeling activity, and (c) modifying those two stratification systems.

Sample segments with digital ground truth data located in the frame were used to provide training and test data. A modified supervised training appproach was used to develop statistics for each of the segments: training fields were selected on a systematic grid over the segment, and cover types were identified from ground observation data. All fields of one cover type (corn, soybeans, or "other") were clustered together.

Two sampling methods were used to select data for classification and area estimation. The first method was the method used in the LACIE project: the training statistics developed on one segment were used to classify that segment. Based on the results of classifying each segment in this manner, an area estimate was made for each county in the stratum. County estimates were defined as the average of the segment estimates within that county, as long as there was at least one segment in the county; otherwise, a ratio of the Landsat area estimates to the 1974 agricultural census estimates for counties with sample segments was used to adjust the census data for estimation of counties without sample segments.

The second sampling method used to select data for classification and area estimation was a systematic sample throughout the area of interest. The pixel at every fifth line and column throughout each county was classified, and those results were used to make county area estimates. This provided about the same sampling density as one 5 x 6 nm segment per county. The classifications were conducted using a statistics deck pooled from the segments in the stratum.

Finally, stratified area estimation (7) was used to make estimates of corn and soybean proportions. For county estimates, the pooled error matrix for all counties in the stratum was used. The evaluation of results was based on the data analysis objective of estimation of crop areas. Thus, the accuracy of proportion estimates as well as classification accuracy was of interest.

IV. RESULTS

A. 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. Two methods were employed for pooling statistics from the five segments. In the first, fields from each segment were clustered by cover type, and then the statistics were pooled across all of the segments (Training Procedure 1). In the second method (Training Procedure 2), the fields from all segments were first pooled and then clustered by cover type.

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 than when the fields were pooled by type before clustering (Figure 2). This preference for Training Procedure 1 is again emphasized by the county results shown in Table 1. The area estimates for both corn and soybeans were closer to USDA/ESS 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.

B. STRATIFICATION METHODOLOGY

Once a method for developing training statistics had been defined, the next objective addressed was to define the geographic region to which these training statistics could apply. The statistical concept required here is stratification methodology. By the term "stratification," we refer to a subdivision of the population or universe into subgroups, each of which is relatively homogeneous with respect to a variable of interest which differs from one subgroup to another. In defining strata to determine the geographic region over which a set of statistics applies, we want to define strata where corn "looks like" corn, and soybeans "look like" soybeans. We will refer to this type of stratification as <u>spectral stratification</u>. Four spectral stratification systems were examined:

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

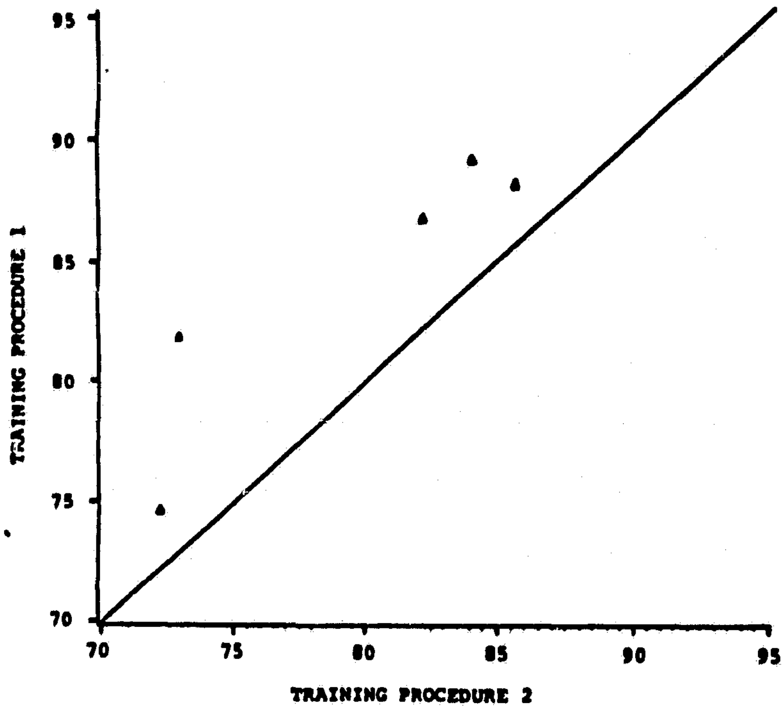

Figure 2. Comparison of the overall classification accuracies achieved using two training methods. Each point represents one sample segment. The solid line represents equal accuracies for the two methods.

Table 1. Proportion estimates of corn and soybeans made using two different training methods and compared with USDA/ESS estimates for the same region.

County	Classification Results				USDA/ESS Estimates	
	Training Procedure 1		Training Procedure 2			
	Corn	Soybeans	Corn	Soybeans	Corn	Soybeans
Emmet	44.4	38.9	44.9	31.7	40.0	37.1
Palo Alto	42.8	41.5	42.8	35.2	41.8	38.5
Pocahontas	41.2	45.1	43.9	32.0	41.1	39.8

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 (Figure 1). After development of statistics on a segment-by-segment basis
for the ten sample segments in the stratum, the statistics decks from all segments were pooled to represent the stratum. The divergence between cluster classes was computed to determine any classes which should be pooled or deleted. The statistics for two segments, both in Webster County, were not compatible with the statistics from the other sample segments; for example, the mean vector of a class of corn in one part of the stratum was the same as for a class of "other" in another part of the stratum resulting in a divergence of zero.

Since two segments were spectrally anomalous from the rest of the segments, the county in which these two segments fell (Webster 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 (Figure 3). 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 neither the refined strata nor the refined/split strata are sufficient for spectral stratification. The strata are apparently too broad to use as spectral strata. In defining spectral strata, other factors need to be taken into

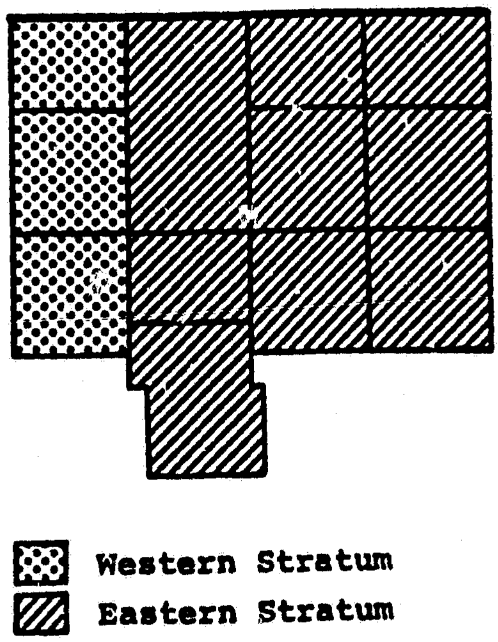

Figure 3. The counties of interest were divided into two strata by Stratification Method 3, the refined/split strata.

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.

C. COMPARISON OF PIXEL SAMPLES AND SEGMENT SAMPLES FOR CLASSIFICATION AND 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/ESS estimates.

Classification Accuracy. Classification accuracies were generally higher on the segments when statistics representing that segment alone were used in the classification (Figure 4). 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.

Figure 5 compares the classification accuracies of Methods B and C. Most segments had higher classification accuracies when two strata were used. This confirms the previous hypothesis that spectral strata are somewhat smaller than the refined strata.

Proportion Estimation Accuracy. The proportion estimates of corn and soybeans in each county are shown in Table 2 for each of the three stratification and sampling methods. Figure 6 shows the comparison between corn proportion estimates made by each of the three methods with the USDA/ESS estimate for the same county. The correlations between the Landsat and USDA estimates for corn are relatively high for all three methods (0.77 for Method A;

ORIGINAL PAGE
OF POOR QUALITY

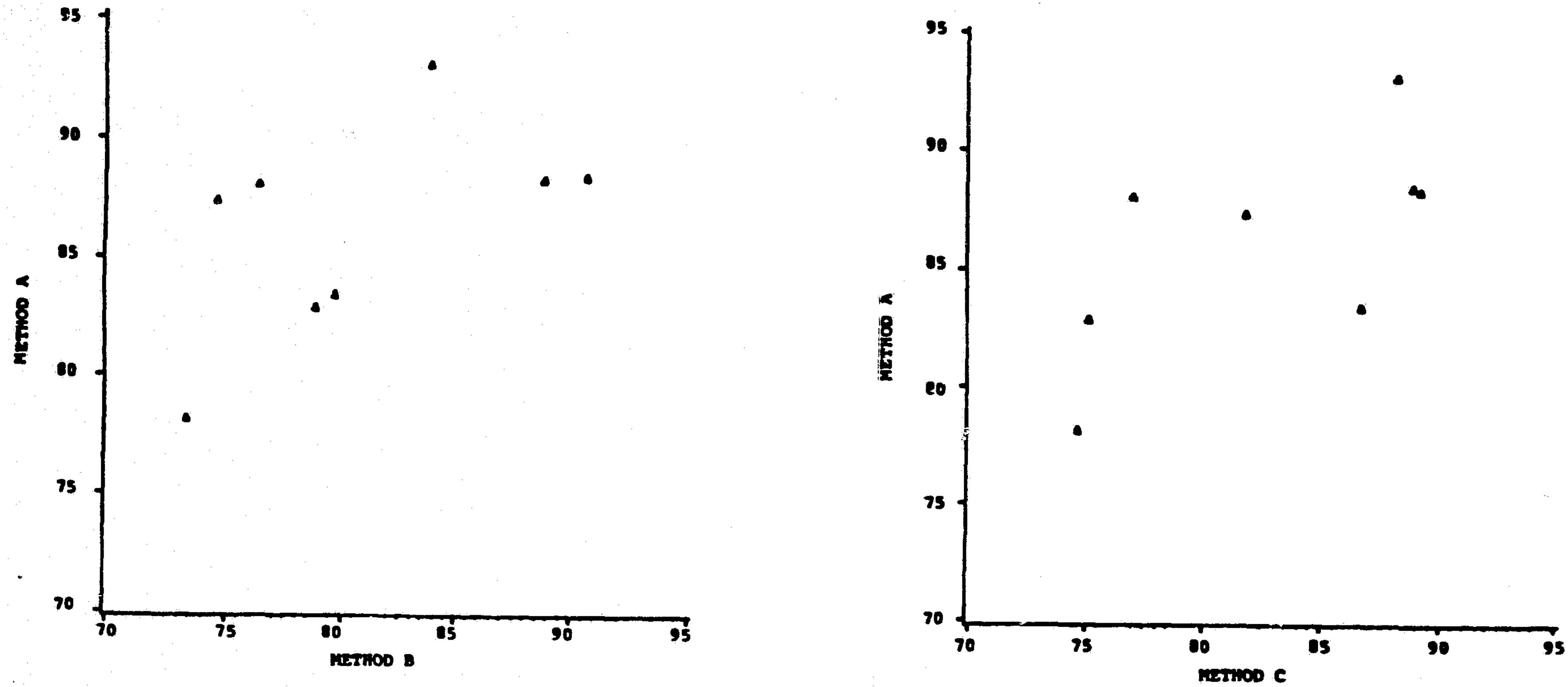

Figure 4. Comparison of the overall classification accuracies obtained using two systematic sampling methods with the accuracy of the segment method. Each point represents one sample segment.

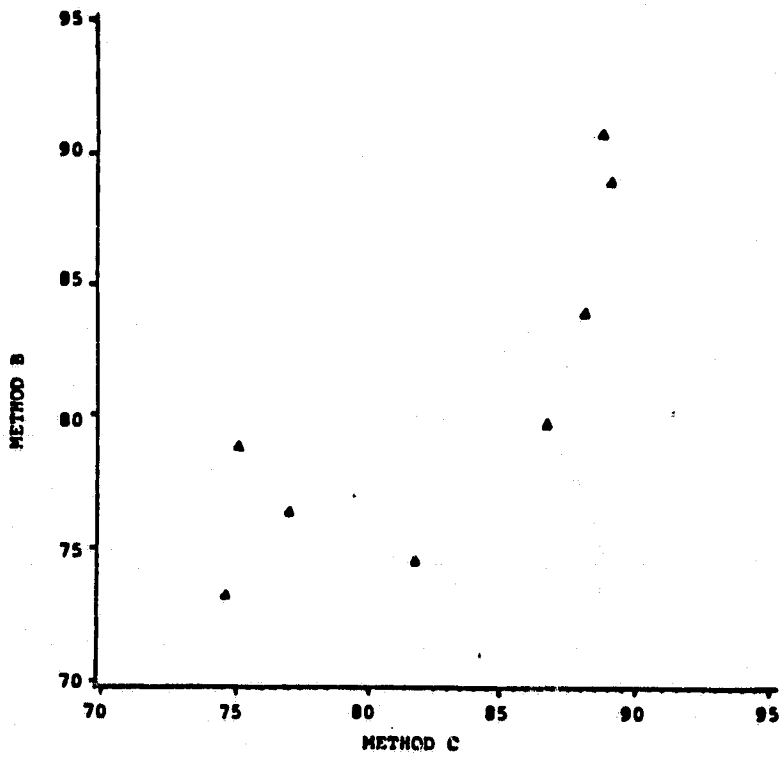

Figure 5. Overall accuracies using two systematic sampling methods.

Table 2. Proportion estimates of corn and soybeans made using three different stratification and sampling methods and USDA/ESS estimates for the same regions.

County	Method A		Method B		Method C		USDA/ESS	
	Corn	Soybeans	Corn	Soybeans	Corn	Soybeans	Corn	Soybeans
	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)
Emmet	38.9	41.4	47.3	37.0	44.4	38.9	40.0	37.1
Palo Alto	38.6	28.2	46.7	38.5	42.8	41.5	41.8	38.5
Pocahontas	39.2	35.8	46.0	41.2	41.2	45.1	41.1	39.8
Kossuth	43.9	46.8	49.5	35.9	51.6	40.0	43.1	39.6
Humboldt	49.8	47.8	50.6	36.4	53.0	39.4	45.6	40.2
Winnebago	46.6	41.2	46.1	37.8	47.4	43.6	42.1	37.8
Hancock	51.5	31.0	49.7	36.7	51.9	40.5	44.4	34.8
Wright	50.6	43.1	50.8	37.2	55.6	39.6	47.2	42.8
Worth	48.1	35.1	46.2	37.7	48.2	42.8	43.0	32.6
Cerro Gordo	46.5	31.7	46.6	35.3	48.7	41.3	40.8	29.3
Franklin	46.9	32.2	50.9	34.2	53.7	38.0	44.5	30.7

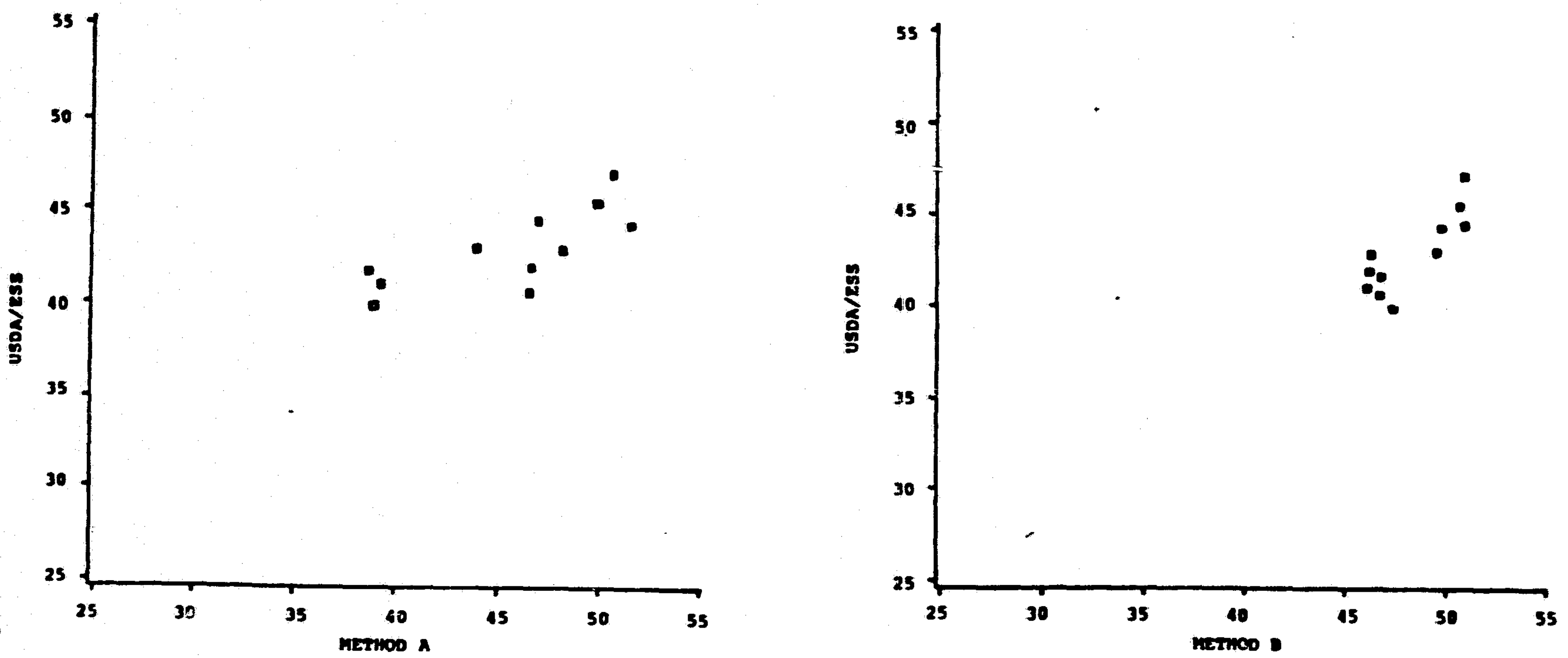

Figure 6. Comparison of the corn proportion estimates for each county made by each of the three stratification and sampling methods with the USDA estimates for the same county.

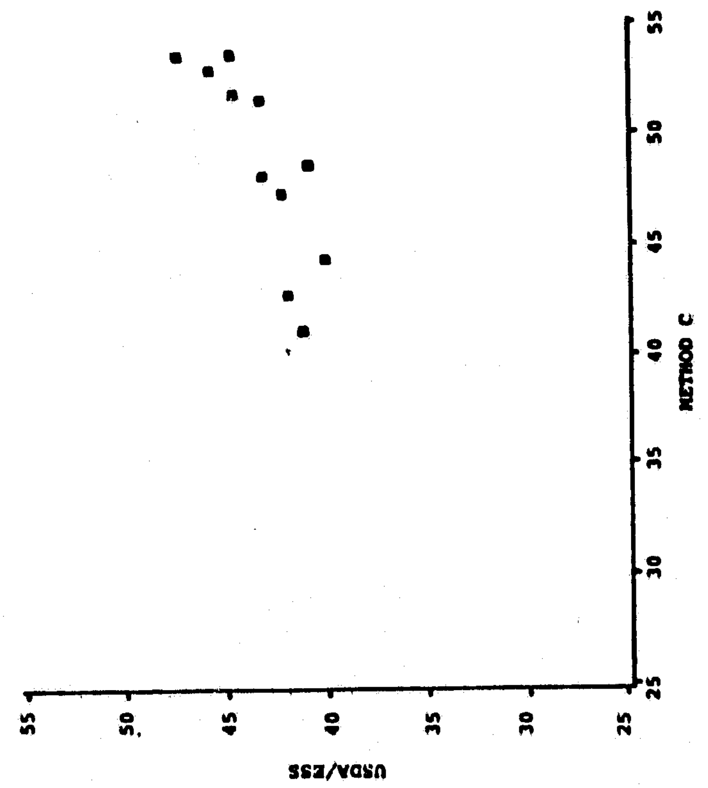

Figure 6. Continued.

ORIGINAL PAGE IS
OF POOR QUALITY

0.62 and 0.83 for Methods B and C, respectively). For soybeans, however, the correlations were much lower except for Method B (R=0.81). Methods A and C had correlations of 0.51 and 0.09, respectively.

Table 3 compares these estimates to the USDA/ESS proportion estimates by 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. In the eastern stratum, however, Method A performed better than Methods B and C for corn and better than Method C for soybeans.

The results here indicate the potential for using pooled statistics from segment data to represent a spectral stratum. The results from the western stratum illustrate that a good spectral stratification can provide area estimates that are as accurate or more accurate than segment-based estimation. In addition, the precision of the estimates made from the systematic sample will be greater.

The eastern stratum results, on the other hand, show a general degradation in accuracy when the systematic sample is utilized. We believe this is due to one of two causes: first, only three sample segments were available to provide training data for the eight counties in the stratum, so the spectral subclasses in the stratum may not be well-represented; second, the geographic extent of the eastern stratum (eight counties) is relatively large and may be too broad for a good spectral stratification. The results indicate that both of these potential causes may be contributors to the lowered accuracy. The lack of training data may be a factor since the single stratum accuracy (eight training segments) was higher for both crops than the two stratum accuracy (three training segments). The hypothesis that the eastern stratum is too broad is based on the fact that neither systematic sampling method provided accuracies as good as the segment-based estimation method.

V. SUMMARY

The potential for using pooled segment statistics for an entire stratum is indicated by the generally good performance for both corn and soybeans in the western stratum. This type of training approach used with classification of a systematic sample of pixels seems to merit further investigation due to the variance reduction benefits which could be obtained. In particular, the potential shown for this method should be more fully investigated using multitemporal data which should produce

ORIGINAL PAGE IS
OF POOR QUALITY

Table 3. Root mean square errors of corn and soybean proportions from USDA/ESS estimates are given for three different stratification and sampling methods.

	Corn			Soybeans		
Stratum	Method A	Method B	Method C	Method A	Method B	Method C
Western	2.2	5.8	2.6	6.8	0.8	3.7
Eastern	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

still higher classification accuracies and more accurate area estimates.

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 physical factors defining the strata and into methods of stratification will be an important task in the development of a full-frame sampling strategy.

VII. REFERENCES

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

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 Sennsing 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. Department of Agriculture, Economics, Statistics, and Cooperative Service. Technical Bulletin No. 1609.

5. Hixson, Marilyn, Donna Scholz, Nancy Fuhs, and Tsuyoshi Akiyama. 1980. Evaluation of Several Schemes for Classification of Remotely Sensed Data. Photogr. Engin 46:1547-1553.

6. 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 Symposium, Laboratory

for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana, June 3-6, pp. 72-77.

7. Heydorn, R.P., R.M. Bizzell, J.A. Quirein, K.M. Abotteen, and C.A. Sumner. 1978. Classification and Mensuration of LACIE Segments. Proc., The LACIE Symposium, NASA, Johnson Space Center, Houston, Texas, October 23-26, pp. 73-86. JSC-16015.

DISTRIBUTION LIST

NASA Johnson Space Center

SA:	W. Rice
SK:	R. Erb
SG:	R. MacDonald
SH:	J. Erickson
SG3:	F. Hall
	C. Hallum
	A. Houston
	A. Feiveson
	D. Pitts
	D. Thompson
	K. Henderson
	G. Badwahr
	J. Paris
	R. Heydorn
SH2:	M. Trichel
SH3:	J. Dragg
	R. Bizzell
	R. Eason
	V. Whitehead
JM6:	Technical Library (4)
AP:	Public Affairs Office

NASA Headquarters

ER-2: P. Thome

NASA Scientific and Technical
Information Facility

R. Horvath:	ERIM
W. Malila:	ERIM
R. Cicone:	ERIM
F. Pont:	ERIM
R. Kauth:	ERIM
C. Hay:	Univ. Calif.-Berkeley
R. Sielken:	TAMU
T. Lee:	LEMSCO
J. Lundgren:	LEMSCO
J. Malin:	LEMSCO