CR 134253

WHEAT CLASSIFICATION EXERCISE, USING JUNE 11, 1973, ERTS MSS DATA FOR FAYETTE COUNTY, ILLINOIS (FOR CITARS TASK)

William A. Malila
Ross H. Hieber
Daniel P. Rice
Jane E. Sarno

September 1973

Technical Memorandum

190100-21-R

Prepared under Contract NAS9-9784
with
NASA Johnson Space Center

ENVIRONMENTAL RESEARCH INSTITUTE OF MICHIGAN

FORMERLY WILLOW RUN LABORATORIES, THE UNIVERSITY OF MICHIGAN

Technical Memorandum No. 190100-21-R

26 September 1973
Copy Without Enclosures

WHEAT CLASSIFICATION EXERCISE, USING JUNE 11, 1973, ERTS MSS DATA FOR FAYETTE COUNTY, ILLINOIS (FOR CITARS TASK)

William A. Malila
Ross H. Hieber
Daniel P. Rice
Jane E. Sarno

INTRODUCTION

The Earth Observations Division (EOD) of the Johnson Space Center, Houston, Texas, requested a timely evaluation of ERIM's capability to use computer techniques to recognize wheat in ERTS-1 multispectral scanner (MSS) data collected over Fayette County, Illinois, on 11 June 1973. This exercise was an initial part of the CITARS (Crop Identification Technology Assessment for Remote Sensing) Task which is a joint investigation by EOD, Purdue University's Laboratory for Applications of Remote Sensing (LARS), and ERIM.

The proceduralized data processing and analysis techniques outlined for the CITARS Task (Ref. 1) were followed predominantly (with some steps being skipped because of time constraints), but the judgement of experienced analysts was required and utilized to discover and correct several problems in the data set to produce meaningful results with these techniques.

ERTS-1 data were received at ERIM from LARS on 12 September 1973. Ground truth information and aerial photography were received from EOD on 9 and 15 September. The data were analyzed and processed digitally using the ERIM multispectral software system. Telephone reports were made to EOD of preliminary results on 17 and 20 September and final results on 21 September 1973. This document describes the effort that was carried out and the reported results.

In summary, the prime emphasis was on classification of pixels in field centers, away from boundary effects. Results were encouraging in both training and test field centers for wheat and other major types of vegetation present. However, the location of fields was found to be a serious problem and it was even more difficult to select field-center pixels for fields of sizes less than 20 acres (or even larger, depending upon field shape) for use in the field-center analysis. The majority of fields in the segment are less than 20 acres in size.

An examination of the recognition maps showed that there was good recognition of woodlots and large water bodies throughout the segment. However, a sizeable number of pixels representing signals received from two or more materials along feature boundaries were misclassified. This effect was evident in wheat misclassifications along some field boundaries, as well as along the shoreline of a large body of water in the segment and along a nearby river thought to be more narrow than an ERTS spatial resolution element. Also, fields smaller than could be reliably located and used for the field-center analysis did not always appear to be classified correctly. In an attempt to quantify these boundary and small-field effects, the 19 quarter-sections for which ground truth information was available were sketched on a recognition map and the number of pixels classified as wheat were counted manually. The overall proportion of pixels classified as wheat in these quarter sections quite closely matched the actual proportion of wheat acreage, but the variance between actual and estimated quarter-section values was large (rms error approximately 1/2 the estimated overall proportion).

DATA FLOW

Data were received from both LARS and EOD for use in this exercise. A computer compatible tape containing the ERTS-1 data was received late on 12 September 73 from LARS. LARS had extracted data for a region that contains the 5 x 20-mile segment selected for the CITARS Task. Then LARS performed a rotation, de-skewing, and scaling of the data so that a line printer map has a scale of approximately 1:24,000. A second file on the tape contained un-rotated data for the segment. Digital line printer maps of rotated data for each ERTS channel accompanied the tape and were very useful in the analysis.

EOD provided ground-truth information collected by ASCS personnel for 20 quarter sections (1 was flooded) on 28 and 29 June 73, USGS topographic maps of the area, and aerial photography with annotations. On 15 September, additional aerial photography was received from EOD on which additional wheat fields to be used for test purposes had been identified by photointerpretation. Enlargements of RB-57 photographs taken in August over the segment were also enclosed and proved to be very helpful in subsequent work. No ERTS imagery was available, but it would have been helpful.

LARS determined ERTS data coordinates (scan line and point numbers) for the quarter-sections, training fields, and test fields, and telephoned them to ERIM on 13 and 14 September. These coordinates were used for initial data processing, but it became apparent that many fields were misplaced and it was necessary for us to revise them, as discussed in a later section. It was very difficult to visually locate landmarks in the line printer maps.

PROCESSING PROCEDURES

The ERIM procedures described elsewhere for use on the CITARS Task were used predominantly. The first step was to convert the ERTS data from LARSYS-3 to ERIM format. Line printer maps also were made for ERTS Bands 5 and 7 to permit comparison with those received from LARS. An agreement between the LARS and ERIM scan line and point numbers and pattern of data values was established. We also produced line printer maps of the unrotated data and, in the opinion of several analysts, scene features appeared to be somewhat more distinct there than on the rotated data map. There also were differences in the ranges of data signal levels and symbols used to represent them, overprinting being used to produce greater density differences on the ERIM maps.

Data screening was performed by examining the line-printer maps and by computing histograms of data from the six individual detectors in each ERTS Band on the non-rotated ERTS data over the entire area represented on the tape. One scan line above the segment was found to have abnormal values for ERTS Band 4, but does not affect the conclusions of the study. The results of the histograms are presented in Table I. The variance between detectors was not great enough to warrant any special treatment of the data.

We also noted the presence of clouds and cloud shadows in portions of the data and made a new selection of levels for line printer maps based on the histograms and a listing of data values along scan lines that passed through clouds and shadows. The new maps helped in determining more quantitatively where else clouds were in the data. Cloud shadows showed more distinctly in ERTS Band 7 (values $\leq$ 11) and clouds more distinctly in ERTS Band 5 (values $\geq$ 75), although values $\geq$ 43 in Band 7 and $\leq$ 17 in Band 5 also gave reasonable indications of clouds and cloud shadows, respectively. These values will not necessarily apply to other data sets because the atmospheric conditions may differ, giving different amounts of path radiance, transmittance, and irradiance.

The next step was to extract signal statistics for the training fields whose coordinates had been forwarded by LARS. Some large variances were noted in the wheat training field statistics, but they were nevertheless combined into a single signature and rejection tests applied to each individual field, as described in the ERIM procedures (Ref. 1). All fields were accepted (at the 0.001 probability of rejection level) and the final combined signature was used for an initial classification run along with one signature for water. (The procedure calls for a run with the major crop signatures but, since only wheat was of interest in this case, a second, very spectrally different signature was used to facilitate the linear classification algorithm.)

TABLE I. SIGNAL STATISTICS FOR ERTS DATA QUALITY CHECK, 11 JUNE 1973 DATA FOR FAYETTE COUNTY, ILLINOIS

		SIGNAL STATISTICS* FOR ERTS BAND:			
		4	5	6	7
Detector 1	Mean	38.46	33.46	51.72	26.46
	Std. dev.	13.23	16.46	13.83	10.15
Detector 2	Mean	37.74	33.51	50.94	26.10
	Std. dev.	13.35	16.62	13.45	10.01
Detector 3	Mean	37.75	33.37	51.23	26.18
	Std. dev.	13.21	16.41	13.76	10.05
Detector 4	Mean	38.25	33.38	51.56	26.50
	Std. dev.	13.20	16.76	13.64	10.24
Detector 5	Mean	38.45	34.30	52.69	26.14
	Std. dev.	13.36	17.12	14.02	10.17
Detector 6	Mean	38.64	33.81	52.71	27.51
	Std. dev.	14.16	16.71	14.06	10.33

*Each entry in the table represents a value obtained from 66,500 ERTS pixels, that is, a value obtained from every sixth scan line of data over the entire un-rotated data set received from LARS for Fayette County, Illinois.

A likelihood map was generated for all points classified as wheat, and classification results were computed for each training field. An analysis of these results and the individual field statistics led us to suspect the assignment of pixels to the training and test fields. Examination of several cases on the line printer maps convinced us that the field and quarter-section boundaries all had to be checked and that most would need revision. We telephoned our conclusions to LARS and EOD, and began the process of relocation, using some of the procedures described in Reference 2.

Since it was impossible to identify the road network on the line printer maps of rotated data, we located 18 control points at positions that we could correlate with latitude and longitude coordinates on the USGS maps, with the aid of the photographs. The discernable features were primarily woodlots in ERTS Band 5 and water bodies in ERTS Band 7. A regression equation to convert from latitude and longitude to ERTS data coordinates was computed. The coordinates of the section corners were determined in latitude and longitude, converted to ERTS data coordinates, and plotted on the line printer maps. Then, quarter section lines were drawn and field-center pixels for the larger fields were selected by reference to aerial photographs that contained the current field patterns. Care was taken to avoid boundary effects, and pixels around the edges of a spatial feature usually were not selected.

The extraction of field statistics was repeated with the new field coordinates for all but water, trees, and urban areas, and a new wheat signature established. Seven wheat fields were used for training (see Table II). Of these, one (Field 21-48*) was rejected by our test with the combined signature and was, therefore, excluded from the wheat signature used for all later classification. The ground-truth information we had available was for late June and did not indicate any reason for a difference between this field and the others. Subsequently, we learned from Dr. Forrest Hall of EOD that the early June ground-truth information shows that this field was substantially less mature than the others which were all in various stages of turning yellow. Figure 1 presents the signal statistics (mean $\pm$ std. dev.) for all training and test wheat fields used in this study. From this figure, it is clear that the signature for Field 21-48 departs substantially from those of the other training fields in ERTS Bands 6 and 7, owing to its greener condition. Signals for Field 32-42 are somewhat greater than the others in ERTS Band 6 and, as will be seen later, this field had only 40% of its pixels classified as wheat. One might infer that Field 32-42 was more mature than Field 21-48, but less mature than the other wheat training fields.

*Sections for this report are numbered from 1 to 100, starting in the NW corner of the segment, moving horizontally from W to E, and progressing Southward for each new line of five sections.

TABLE II. CLASSIFICATION RESULTS FOR WHEAT FIELD-CENTER PIXELS

A. RESULTS FOR TRAINING FIELDS

	Field**	Size (Acres)	Pixels Chosen	WHEAT CLASSIFICATION* Two-Signature Run # Pixels	Two-Signature Run %	Seven-Signature Run # Pixels	Seven-Signature Run %
	86-01	20	8	8	100	8	100
Used in Signature	84-11	40	9	9	100	9	100
Used in Signature	83-23	28	8	8	100	8	100
Used in Signature	25-04	20	4	4	100	4	100
Used in Signature	28-50	20	8	8	100	7	83
	32-42	24	10	10	100	4	40
	Subtotals		47	47	100%	40	85%
Rejected in Training	21-48	33	12	3	25	0	0
	Totals		59	50	85%	40	68%

B. RESULTS FOR TEST FIELDS

	Field	Size (Acres)	Pixels Chosen	WHEAT CLASSIFICATION* Two-Signature Run # Pixels	Two-Signature Run %	Seven-Signature Run # Pixels	Seven-Signature Run %
	15-A		2	2	100	2	100
	25-A		12	12	100	12	100
	58-A		3	3	100	3	100
	79-A		6	5	83	5	83
	49-A		7	7	100	5	72
	Subtotals		30	29	97%	27	90%
Suspect Field Identification	9-A/B		9	0	0	0	0
Suspect Field Identification	59-A		12	0	0	0	0
	Totals		51	29	57%	27	53%

*For a threshold corresponding to 0.001 probability of false rejection, assuming multivariate normal distributions.

**Sections for this report are number from 1 to 100, beginning in the NW corner of the segment and moving from W to E, progressing Southward line by line.

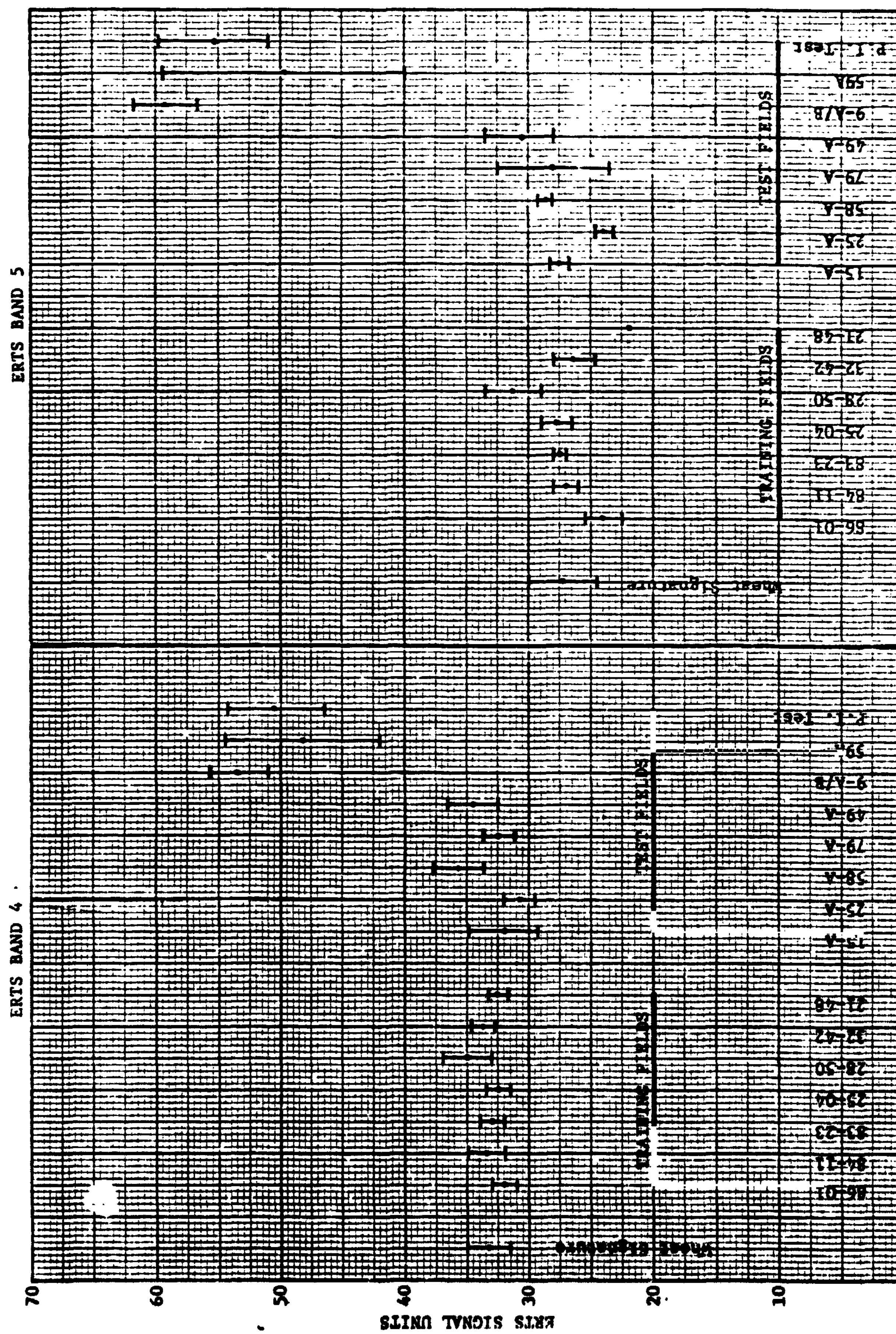

FIGURE 1. COMPARISON OF WHEAT SIGNATURE AND SIGNAL STATISTICS FROM TRAINING AND TEST WHEAT FIELDS (Part 1 of 2)

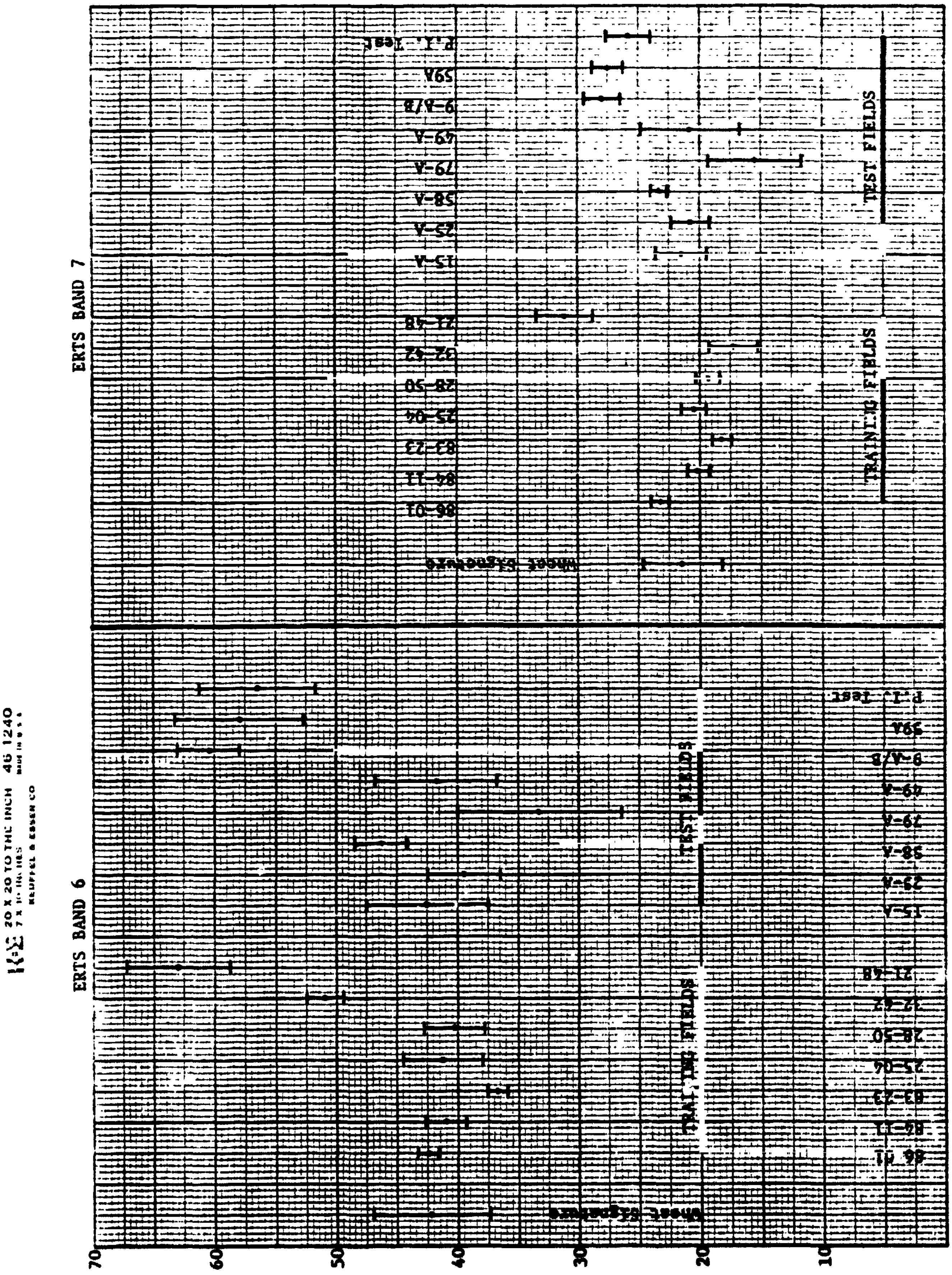

FIGURE 1. (Continued) (Part 2 of 2)

A preliminary classification run with the ERIM linear decision algorithm was made for a large area that extends somewhat beyond the segment boundaries. Two signatures were used -- wheat and water -- as before. This time, the classification results were much more consistent than for the initial run. Classification statistics were extracted for the training and test wheat fields, as well as for training fields of other classes. These results are discussed fully in the next section.

Signatures were produced for six classes, in addition to wheat, for use in the final classification run. In each instance, the signature combination and rejection-test procedure was followed. The signatures for water, trees, corn, soybeans, clover, and pasture are plotted in Fig. 2 along with the wheat signature.

The final classification run was made using seven signatures, and individual field and class results were tabulated for field-center pixels. Results within whole quarter sections also were analyzed.

RESULTS

Results are presented separately for field-center pixels, for whole quarter sections, and for the entire scene.

Field-Center Pixels

Classification results for field-center pixels in the seven wheat fields used in the training procedure and in seven test fields identified by the EOD photointerpreters are presented in Table II. Final results for the six fields used in the signature are 100% correct classifications, with the exception of Fields 28-50 (83%) and 32-42 (40%). It was noted earlier that the signal statistics for Field 32-42 lie between those of the other five fields and the seventh, immature field (21-48) which was excluded from the signature. Of the 19 wheat training pixels not classified as wheat on the seven-signature run, the 12 of Field 21-48 were called clover, 5 others pasture, 1 soybeans, and 1 not classified. Clover and pasture would likely have more green vegetation present than the ripening wheat fields.

All test wheat fields were identified by EOD photointerpreters. Three test fields were 100% classified as wheat, one 83%, one 72%, and three 0%. For the first five, the average classification accuracy was 90% of the pixels. Of the three for which no pixels were classified as wheat, only two are listed in Table II. The third is not listed because it was incorrectly identified as wheat by the photointerpreters, according to ground truth information which was withheld from them. The signal statistics (see Fig. 1) for Fields 9-A/B and 59-A are very similar to those of the incorrectly identified PI Test field and we seriously doubt that they are wheat.

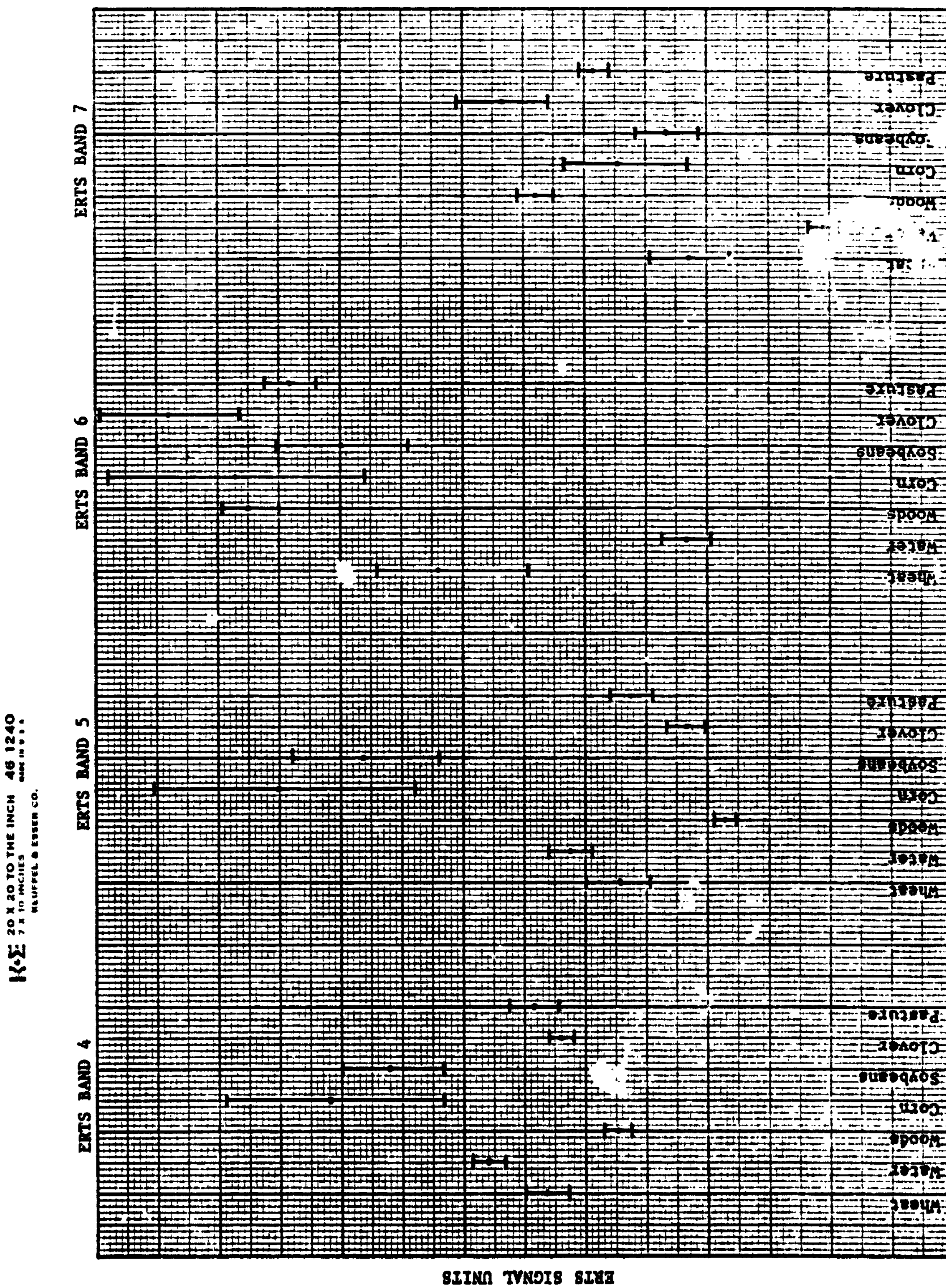

FIGURE 2. COMPARISON OF WHEAT AND OTHER-CLASS SIGNATURES
(FAYETTE COUNTY, ILLINOIS; ERTS DATA COLLECTED 11 JUN 73)

Over one-third of the scene was classified as wheat in the two-signature run (See later section for results over entire scene). The seven-signature run reduced the recognized proportion of wheat in the scene to a more reasonable value. There also was a slight decrease in the number of correctly classified field-center wheat pixels when the other signatures were added.

It is of interest to examine how field-center pixels for other classes were classified on the two runs. Table III shows that only a very few of the other-class field-center pixels were misclassified as wheat, even on the two-signature run. Time and ground-truth constraints kept us from establishing a separate set of test fields for the other categories but, in view of their rejection during the two-signature run, it is not believed to have been an important step here. The other classes used for the final classification run were corn, soybeans, clover, pasture, trees, and water; they were employed primarily to reduce the total number of scene pixels improperly classified as wheat from the number obtained for the two-signature run. A rectangle over the city of Vandalia was selected as a test section, but no city signature was employed for the classification runs.

The classifications for the other-class field-center pixels were largely correct, not surprising in view of the fact that they were used for training. The exceptions were corn and soybeans which were frequently confused with each other. This confusion too is predictable in early June when most of the corn and soybean fields were bare soil. Note the footnote in Table III regarding pixels for one water training area. The boundary problem was more apparent on our line printer map for Band 7 with its additional levels to help distinguish cloud shadows. This example points out the need to stay at least one pixel away from edges in the selection of field-center pixels and the need for display levels tailored to the scene features being distinguished.

The city test area, classified mostly as wheat in the two-signature run, was classified mostly as pasture and soybeans in the seven-signature run, with a representation of all vegetation classes. Pixels from a city or suburban area represent many different mixtures of objects and ground covers.

<u>Whole Quarter Sections</u>

As noted earlier, the field-center pixels for large fields were accurately classified, but there were indications that some of the smaller fields and areas along boundaries between fields were misclassified. A section-by-section examination was made of a likelihood map for all pixels assigned to the wheat category by the classification algorithm. On this map, Enclosure 1, the

TABLE III. CLASSIFICATION RESULTS FOR NON-WHEAT FIELD-CENTER PIXELS

				WHEAT CLASSIFICATION*			
				Two-Signature Run		Seven-Signature Run	
Class	Field	Size (Acres)	Pixels Chosen	# Pixels	%	# Pixels	%
Corn	19-70	38	16			0	0
	18-06	26	7	0	0	0	0
	72-05	40	12	0	0	0	0
	72-11	22	6	0	0	0	0
	75-30	54	15	0	0	0	0
	82-66	32	20	0	0	0	0
	83-30	63	12	1	9	0	0
	Total		88	1	1%	0	0%
Soybeans	18-12	32	27**	0	0	0	0
	21-41	26	8	0	0	0	0
	25-01	25	9	0	0	0	0
	28-51	20	8	0	0	0	0
	32-43	34	12	0	0	0	0
	32-45	40	5	0	0	0	0
	40-66	26	8	0	0	0	0
	46-52	41	12	0	0	0	0
	40-70	25	15	3	20	1	7
	07-45	50	12	7	58	4	33†
	Total		116	10	9%	5	4%
Clover	83-25	20	6	0	0	0	0
	07-43	32	4	1	25	0	0
	18-05	37	7	3	43	1	14
	Total		17	4	24%	1	6%
Trees	3		60	0	0	0	0
	2		60	0	0	0	0
	1		60	2	3	0	0
	Total		180	2	1%	0	0%
Pasture	44-51	10	7	7	100	0	0
	21-46	11	2	2	100	0	0
	Total		9	9	100%	0	0%
Water	1		32	0	0	0	0
	2		64	0	0	0	0
	3		24	3	12	3	12††
	Total		120	3	3%	3	3%
City			252	149	59%	29	12%

(See next page for footnotes)

(Footnotes for Table III)

*For a threshold corresponding to 0.001 probability of false rejection, assuming multivariate normal distributions.

**Extends across quarter-section boundary.

†Listed as bare soil on 6/28/73.

††The points designated as water for this test area were found to include some that contained the shoreline. The rejection of pure water was total.

exponent of the wheat likelihood function was quantized into 12 bins, each corresponding to a different interval of probability of false rejection (under the assumption of multivariate normality). The intervals and symbols used are listed in Table IV. Symbols 3-9 correspond to those accepted by a 0.001 probability of rejection threshold. A color overprint feature was used to designate those pixels used for training and test for the different classes.

The following observations summarize the results of our examination of the likelihood map in the quarter sections. Most wheat fields appear to be recognized quite well, except several which are very narrow, the training fields (21-48 and 32-42) already discussed, and 56-04 (weeds and wheat). There are spotty misclassifications of wheat in several fields of pasture, clover, and grasses, and solid misclassifications in a few. Two rye fields are called wheat. The most common occurrences of wheat misclassifications appear to be for boundary elements, like between trees and fields or between two types of fields. A few misclassifications are in the vicinity of farmsteads. The reader must bear in mind that it is not possible to determine the exact location of individual pixels within the quarter sections relative to the various field boundaries, or even the exact locations of the quarter-section boundaries themselves.

To obtain a quantitative assessment of the wheat recognition over whole quarter sections that include field boundaries, a manual quarter-section-by-quarter-section count was made of all wheat assignments with probability of false rejection of 0.001 or greater. The results were converted to proportions of wheat in each quarter section and are presented in Table V, along with proportions determined from the ASCS ground-truth information. The average proportion determined from the ERTS MSS data is 13.9%, compared to 12.9% from the ground truth information. This is good agreement; however, the large variance (rms error = 7.1%) of the estimates also should be considered. The misclassifications of wheat roughly balance the missed classifications of wheat; this situation probably would not obtain in general.

Entire Scene

Recognition maps were produced for the entire scene for both classification runs. The scene contains roughly 190 sq. mi of area which includes the 5 x 20 mi segment. The overall proportions of the various classes in the scene are listed in Table VI. It can be seen that the recognized percentage of wheat in the scene for seven signatures is less than half that for two signatures, the new value being much closer to the proportion in the quarter sections. For a 0.001 threshold, 15.8% of the scene was classified as wheat on the seven-signature run.

TABLE IV. INTERVALS AND SYMBOLS USED FOR WHEAT LIKELIHOOD MAP

Probability of Rejection*, Lower Bound	Map Symbol
0.70	9
0.50	8
0.10	7
0.05	6
0.01	5
0.005	4
0.001	3
0.0005	2
D	1
C	0
B	*
A	-

*Assuming multivariate normal distributions. Bounds D through A are for progressively smaller values.

TABLE V. COMPARISON OF WHEAT PROPORTIONS IN QUARTER SECTIONS AS OBTAINED FROM ERTS DATA AND FROM GROUND TRUTH

Quarter Section	No. Wheat Fields	Ground Truth Acreages Wheat	Total	% Wheat	Percent Wheat from ERTS MSS Data
7	0	0	160	0	13.0†
18	2	13	157	8.3	14.0
19	1	7	160	4.4	2.1
21	2	45	158	28.5	11.8
25	2	35	160	21.9	12.4
28	2	36	157	22.9	21.7
32	1	24	155	15.5	8.5
40	2	4	145	2.8	3.0
44	2	29	158	18.4	18.8
46	0	0	153	0	17.3††
56	3	44	165	26.6	10.7
64	3	31	151	20.5	10.9
72	1	13	140	9.3	14.9
73	0	0	174	0	15.2*
75	0	0	178	0	8.0
82	3	39	174	22.4	14.0
83	1	28	159	17.6	15.3
84	1	40	178	22.5	13.2**
86	1	20	160	12.5	24.2
100		(Flooded)			
			Average	12.9	13.9
			RMS Error =		7.1

†Misclassifications primarily along a narrow timber strip and along border between a clover and bare soil field which might be separated by a narrow strip of other material.

††Misclassifications primarily in hay and clover fields.

*The identity of fields 42 and 44, classified as wheat, is not clear from ground truth forms for late June.

**Rye fields are classified as wheat.

TABLE VI. SUMMARY OF CLASSIFICATIONS OVER ENTIRE 190-SQ-MI SCENE

Class	PERCENTAGE CLASSIFIED* Two-Signature Run	Seven-Signature Run
Wheat	38.9%	15.8%
Water	0.3%	0.3%
Trees	N/A	9.8%
Corn & Soybeans	N/A	40.7%
Clover & Pasture	N/A	20.1%
Rejected	60.8%	13.3%

*With Threshold For 0.001 Probability of Rejection.

When the recognition map for seven signatures (Enclosure 2) is compared with an aerial photograph of the scene, one is struck by the accuracy of the recognition of trees and woodlots.

The large water reservoir in the center of the segment is also recognized well, except near its edges where water and tree mixtures are frequently classified as wheat or are rejected as being none of the materials. Also, points along the river that runs along the right side of the segment are frequently classified as wheat. The reason for these misclassifications can be understood after an examination of the signatures for wheat, water, and trees. In each ERTS Band, the wheat signature lies between the other two, so mixtures of water and trees would tend to look like wheat.

There also is scattered wheat recognition on the boundaries of cloud shadows.

CONCLUSIONS

The ERTS signals from ripening wheat appear to be significantly different from those of most other types of vegetation present in the Fayette County, Illinois, on June 11, 1973, and appear to be amenable to detection by automated data processing techniques. High classification accuracies were achieved for field-center pixels of the large fields in the segment, but a greater sample is recommended before definite conclusions about wheat detectability are drawn.

Problems of misclassifications were identified along boundaries between differing fields (or trees) and between vegetation and water, and in certain fields of grasses (clover, hay, and pasture).

One large field of immature wheat was spectrally different from the ripening wheat and was classified as clover. Small wheat fields of less than 20 acres, particularly long and narrow ones, often were not classified as wheat, presumably because few if any ERTS resolution elements contained pure wheat, unmixed with an adjacent crop type.

A problem of incorrect assignments of pixels to training and test fields invalidated initial work and required a re-assignment of them with a computer-aided technique for correlating ERTS data coordinates with Earth coordinates.

Acknowledgements

The authors express their appreciation to ERIM Computer Operations personnel for their cooperation in expediting the processing of the data and to Mr. Richard F. Nalepka for his consultation and advice. Also, the data and information provided by LARS and EOD and their cooperation were essential to the study.

References

1. Appendix A of Quarterly Progress Report No. 190100-17-L (6 Sept. 73) for May, June, and July 1973 on Contract NAS9-9784 with NASA/JSC.

2. "Correlation of ERTS MSS Data and Earth Coordinate Systems" by William A. Malila, Ross H. Hieber, and Arthur P. McCleer, ERIM Report No. 193300-18-SA/J, August 1973, to be presented at Purdue Conference on Machine Processing of Remotely Sensed Data, October 16-18, 1973.