E 7.4 - 1 0.3 2 1

CR-136812

A STUDY OF THE UTILIZATION OF ERTS-I DATA FROM THE WABASH RIVER BASIN

SIX MONTH PROGRESS REPORT

DAVID A. LANDGREBE & STAFF

Purdue University

Laboratory for Applications of Remote Sensing

1220 Potter Drive

West Lafayette, Indiana 47907

Jan 1974

~~July 1973~~

Type II Report for Period July I-December 31, 1973

Prepared for

GODDARD SPACE FLIGHT CENTER

GREENBELT, MARYLAND 20771

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle A STUDY OF THE UTILIZATION OF ERTS-1 DATA FROM THE WABASH RIVER BASIN; SIX MONTH PROGRESS REPORT		5. Report Date **January 1974**
		6. Performing Organization Code
7. Author(s) D.A. Landgrebe & Staff of Purdue Lab for Applications of Remote Sensing		8. Performing Organization Report No.
9. Performing Organization Name and Address Purdue University Laboratory for Applications of Remote Sensing 1220 Potter Drive West Lafayette, Indiana 47906		10. Work Unit No.
		11. Contract or Grant No. NAS5-21773
		13. Type of Report and Period Covered **Type II** July 1, 1973 to December 31, 1973
12. Sponsoring Agency Name and Address Dr. Arthur Fihelly NASA Goddard Space Flight Center Greenbelt Road Greenbelt, Maryland 20771		
		14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

Work performed during the third six month period of the study is described. Eight projects are discussed: Five ERTS data applications experiments and three supporting technology tasks. The most significant results were obtained in the Water Resources Research, Urban Land Use Mapping, and Soil Association Mapping Projects. ERTS data was used to classify water bodies to determine acreages and high agreement was obtained with USGS figures. Quantitative evaluation was achieved of urban land use classifications from ERTS data and an overall test accuracy of 90.3% was observed. ERTS data classifications of soil test site were compared with soil association maps scaled to match the computer produced map and good agreement was observed. In some cases the ERTS results proved to be more accurate than the soil association map. These results plus progress on the five other projects are presented in the report.

17. Key Words (Selected by Author(s)) Digital ERTS-1 Imagery Analysis Earth Surface Feature Classification	18. Distribution Statement
19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price*
---	---	---	---

A. Objectives

The objectives of the work reported herein are outlined in the Data Analysis Plan for the ERTS Investigation "A study of the Utilization of ERTS-1 Data from the Wabash River Basin", ERTS-1 proposal No. SR049. The general objectives are: (1) to evaluate the applications of ERTS-1 measurements which have been appropriately reduced for use in specific earth resources problems, and (2) to determine the desirable measurements needed in future earth resources systems.

B. Scope of Work

There are five scientific investigations which are being pursued to evaluate the applications of ERTS-1 measurements to specific earth resources problems. To further support these objectives four specific supporting technology tasks are also included. The nine tasks are all based on the use of digital computer techniques, including the LARSYS multispectral analysis system, for studying ERTS data in digital form.

C. Conclusions

The report is an interim progress report and no final conclusions on the study are appropriate at this time. Significant results in the area of water resources, urban land use and soil association mapping are presented.

iv TABLE OF CONTENTS

I. INTRODUCTION

This report describes work performed during the third six-month period (July 1 - December 31, 1973) of the Purdue University-LARS ERTS-1 Wabash Valley Study. The study consists of nine projects as described in the Data Analysis Plan and progress and results are presented for eight of these. The ninth project was completed in the previous period and discussed in the June 1973 Type II Report.

Section II presents progress and results from the Crop Species Identification Project. ERTS data from Southeastern Missouri and Southeastern Idaho were analyzed to determine the accuracy with which major crop species could be identified using computer techniques. A project to analyze Indiana data from August 1973 is described. Section III describes advances in the Water Resources Research Project. ERTS-1 MSS data and ERIM aircraft scanner underflight data were analyzed to estimate the area of various water bodies. Look-sun angle effects are observed and discussed for the aircraft data. Section IV presents results of the Urban Land Use Analysis Project. A detailed study of ERTS data from the Gary, Indiana Area is described. This is the third urban area studied; the first two being Milwaukee and Indianapolis. Section V discusses Soil Association Mapping project advances which included computer classification of ERTS data and evaluation of results by overlaying soil association maps on the computer derived map. Section VI describes Earth Surface Features Identification Project results. The test area for the study was reclassified using temporal data and improved agreement was achieved between ERTS results and the ground truth data for forest cover. Section VII presents results for the Atmospheric

Modeling Project, Section VIII states the status of the
Analysis Technique Development Project, Section IX describes
Data Reformatting and Overlay, and X contains the additional
required report items.

The work reported represents the final phase of the
research activities for the study. The following months
will be spent refining results and preparing a comprehensive
final report.

II. CROP SPECIES IDENTIFICATION

During this reporting period work has proceeded on the analysis of ERTS MSS data from three states: Missouri, Idaho, and Indiana. The Missouri and Idaho studies have been carried out in cooperation with the Statistical Reporting Service of the U.S.D.A.

Analysis of Missouri Data

An analysis of ERTS data collected during the Summer and early Fall 1972 over Crop Reporting District No. 9 in Southeast Missouri was conducted. SRS supplied the ground truth data, assisted in the analysis of the MSS data and in cooperation with LARS is evaluating the results. LARS geometrically corrected and overlayed the ERTS MSS data, located the ground truth segments and fields in the data, worked with SRS in analyzing the MSS data, and is assisting in evaluation of the results. A major accomplishment of this particular investigation has been the use of automatic data processing techniques by SRS.

Twenty-nine area segments were located in two ERTS frames which covered Crop Reporting District No. 9 in Southeast Missouri. Data from ERTS passes on August 26, September 14, and October 2, 1972 were overlayed and geometrically corrected. Geometric correction greatly facilitated locating segments and fields. Temporal overlay alleviated the necessity of locating fields in three different data sets as well as permitted a test of the usefulness of temporal data in the classification.

Segments were located in the August ERTS data which had been deskewed and scaled to 1/24,000 scale by overlaying computer printouts onto 1/24,000 scale maps on which the

segments had been drawn. The segments were then clustered and coordinates of the individual fields found on a non-supervised (cluster) classification map.

Statistics for the classes, cotton, soybeans, corn, harvested wheat, grass, and miscellaneous were obtained and the data classified. Nearly all the available crop fields were used as training fields. While training field classification performance is generally higher than test field performance, a major objective is to determine what is the highest level of classification performance which can be expected. The use of training field performance is probably the best indicator of this. Comparisons of discriminant functions with and without weights or prior probabilities were made. The weights were the number of acres of each crop found in the segments.

Preliminary results, considering only training field performance, indicates that it will likely be difficult to identify cotton and soybeans during the growth stages considered in this investigation. There was some confusion between cotton and soybeans as well as between these two crops and other cover types present (Table 1).

Comparison of results in Tables 1 and 2 show that use of prior probability information in the discriminant function improved the performance considerably. Such information is readily available from historic data such as earlier surveys.

Table 1: Classification performance with unequal prior probabilities.*

| Class | No. Points | No. Points Classified As | | | Percent Correct |
		Cotton	Soybeans	"Other"	
Cotton	927	739	137	51	79.7
Soybeans	852	99	612	141	71.8
"Other"	438	68	117	253	57.8
	2217				

Overall Performance $\frac{739+612+253}{2217} = 72.4\%$

Table 2: Classification performance assuming equal prior probabilities.

| Class | No. Points | No. Points Classified As | | | Percent Correct |
		Cotton	Soybeans	"Other"	
Cotton	927	689	83	155	74.3
Soybeans	852	101	338	413	39.7
"Other"	438	52	32	352	80.8
	2217				

Overall Performance $\frac{689+338+354}{2217} = 62.3\%$

* Results in Tables 1 and 2 are for the multitemporal case.

August 26, 1972	Bands 4, 5, 7
September 14, 1972	Bands 5, 7
October 02, 1972	Bands 4, 5, 6, 7

The value of multitemporal information in the classification of cotton and soybeans is shown in Table 3. For cotton, performance was improved 7 to 19 percent by using all bands from three dates compared to each of the dates individually. For soybeans, the highest performance was for the single August 26 ERTS pass.

Table 3. Comparison of unitemporal and multitemporal classification of cotton and soybeans.

	Classification Performance (Percent Correct)			
Crop	August 26[1]/	September 14[2]/	October 2[3]/	All Dates
Cotton	60.6	69.7	73.2	79.7
Soybeans	86.0	67.6	62.4	71.8

1/ Bands 4, 5, and 7

2/ Bands 5 and 7

3/ Bands 4, 5, 6, and 7

Analysis of Idaho Data

An analysis of ERTS data for crop species identification has also been conducted in Southeastern Idaho over a Crop Reporting District area. Procedures similar to those described above have been used in the analysis of ERTS data collected in late August, 1972. The major classes present include: corn, alfalfa, sugar beets, potatoes, beans, pasture, and fallow. Preliminary results are similar to those from Missouri, e.g. performance levels are not as high as has been found in previously reported studies. Possible explanations for this are presented in the Discussion and Conclusions section of this report.

Analysis of Indiana Data

Preparations are being made to analyze ERTS MSS data

collected over the Northwestern Crop Reporting District of Indiana during August 1973 (ERTS scenes 1394-16035 and 1394-16042) for corn, soybeans, and "other". Sufficient ground truth data has been collected for both training and testing purposes. We plan to conduct classifications and make acreage estimates for the entire Crop Reporting District as well as for the individual counties. If time and resources permit, the analysis will be extended to several counties in West Central Illinois.

Discussion

Two experiments have been conducted over geographic areas of 5000 square miles or more. In each case the classification has been less than the 80 to 90 percent accuracies previously reported by LARS and other investigators. Currently we are checking to make sure that the procedures followed were sound, particularly in locating field coordinates. Assuming there are no problems in the field coordinates we have been using, we are left with the alternative that the lower performance is somehow associated with the data being analyzed. The most likely explanation for differences between these results and those found in other experiments is that this time we are working over much larger areas encompassing more variation. To check this hypothesis we plan to subdivide or stratify the areas into smaller, homogeneous parts.

Conclusions

The use of a priori information or weights in the classification improves performance. How much improvement can be expected will be quantified in other experiments. In some cases the use of multitemporal data may also improve classification performance. Further studies will be required

to fully determine under what conditions multitemporal data is superior to single date data for crop classifications. There is some evidence from our experiments that classification performance may decrease when attempts are made to classify large areas. This too will require further study before more definite and quantitative statements as to how big an area can be accurately classified can be made. Undoubtedly this will depend on the characteristics of the area being classified. Important variables are likely to be soil type, topography, and the uniformity of crops being grown.

Plans for Next Bimonthly Period

The analysis of Missouri and Idaho data will be continued with particular emphasis on the questions raised above and classification of the Indiana data will be started.

III. WATER RESOURCES RESEARCH

ERTS-1 MSS Data Analysis

Spectral classes of water in Lake Shafer and Freeman. ERTS-1 MSS data collected over Northern Indiana on May 4, 1973 were analyzed. Six spectral categories were defined by the LARSYS nonsupervised classifier (*CLUSTER) and later used as training classes for the supervised classifier (*CLASSIFYPOINTS). The six spectral classes correspond to:

    (1) Lake Shafer, Indiana water
    (2) Lake Freeman, Indiana water
    (3) Banks (edge of water bodies)
    (4) Crops
    (5) Forest
    (6) Soils   .

The spectral characteristics of the waters from Lake Shafer and Lake Freeman are very similar, as indicated by a separability or divergence value of 146 between these two water bodies. This is a very low value of separability, and we therefore have concluded that the waters of both lakes have spectral responses that are too similar and should be defined as a single spectral class.

In order to improve the acreage estimates of surface water, a spectral class was introduced as "banks" (edge of water bodies). A significant improvement in acreage estimates was observed when the "banks class" was counted as water. This result is illustrated in Table III-1.

TABLE III-1

	Acreage Estimate from ERTS-1	1971 USGS[1] Data	% Error between Estimate & Data
ie water class	1108 acres	1547 acres	28% underestimated
iter class + inks class	1488 acres	1547 acres	4% underestimated

However, the very wet soils of the area were misclassified as banks (shortly before the spacecraft passed overhead, it had rained). Therefore, the original Banks spectral class (edges of water bodies) was subdivided into an Edge class and a water-plus-soils class. Thus, the final classification contained the following spectral classes:

      (1)   Lake Water

      (2)   Water Edge

      (3)   Water plus Soils

      (4)   Crops

      (5)   Forest

      (6)   Bare Soils

Figure III-1 shows a classification map of the lake water class and the edge class, and another map where only the water-edge class has been displayed. Note that because there was ponded water in the bare-soil fields, these areas have been classified as water edge.

---

[1]For several years records of the water-surface elevations of many of the lakes in Indiana have been collected by the Geological Survey under cooperative agreement with the Indiana Department of Natural Resources. Surface area (in acres) of the lakes is that surface area at the established level, where, the established level is that elevation set by the courts to which the average level of the lakes is to be held; it is normally set at about the average level that has prevailed for a number of years prior to the establishment of the level.

(a)   (b)
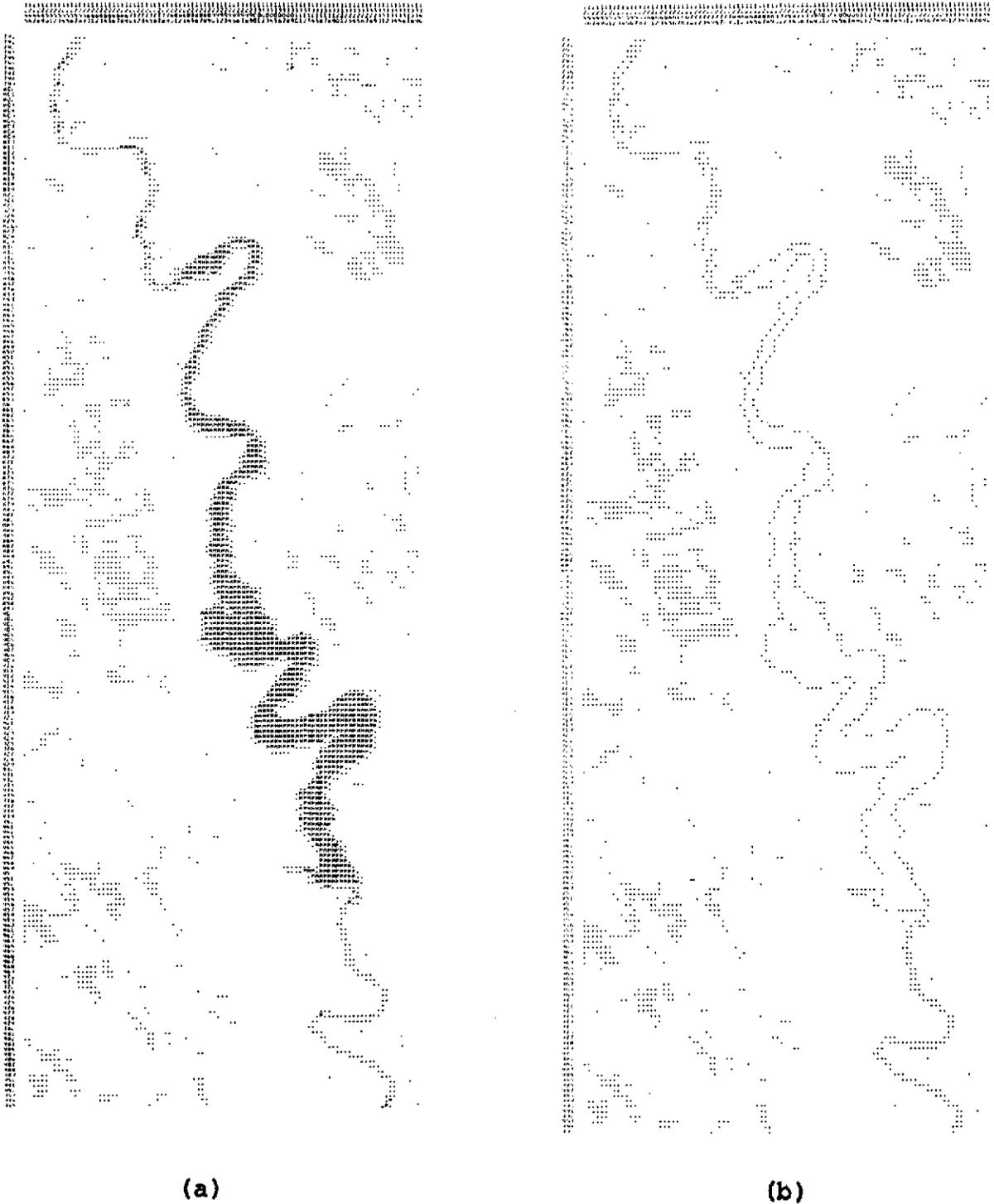
Figure III-1.   Classification map of Lake Freeman, Indiana (a)
                 where the "water" class (M) and the "edge" class (·)
                 are displayed. In (b) only the "edge" class has
                 been displayed.

ERIM[*] MSS Data

Spectral classes of water in Lake Freeman. Multispectral
scanner data was gathered in 12 bands by the ERIM scanner
system over Lake Freeman approximately an hour after the
ERTS-1 overpass on May 4, 1973. These data were collected
at 10,000 feet altitude.

Procedures. The aircraft scanner data were classified using
training fields selected on the basis of photointerpretation
of color, color IR and B & W photography taken simultaneously
with the scanner data. The set of training fields include
forests, crops, and soils. However, in order to define water
spectral classes, it was necessary to perform a nonsupervised
classification of just water. Thus, seven separable spectral
categories of water were defined. The coincident spectral
plot of the seven classes of water, the forest, crops and
soils is shown in Figure III-2.

Discussion. It should be noted that the seven spectral
classes of water in Lake Freeman have an unusual spectral
response, that is, in every reflective channel the seven
classes of water have a response that ranges from high to low;
they don't show a complicated signature as a function of
wavelength band. However, it is interesting to note that
this is not the case with the thermal band (9.3-11.7 μm),
in fact, there seems to be no difference in radiant temperatures
between the seven categories of water.

The result of the classification shows a particular
spatial distribution of the seven classes of water (Figure III-3).
Note that the classes of water in Figure III-3 are distributed
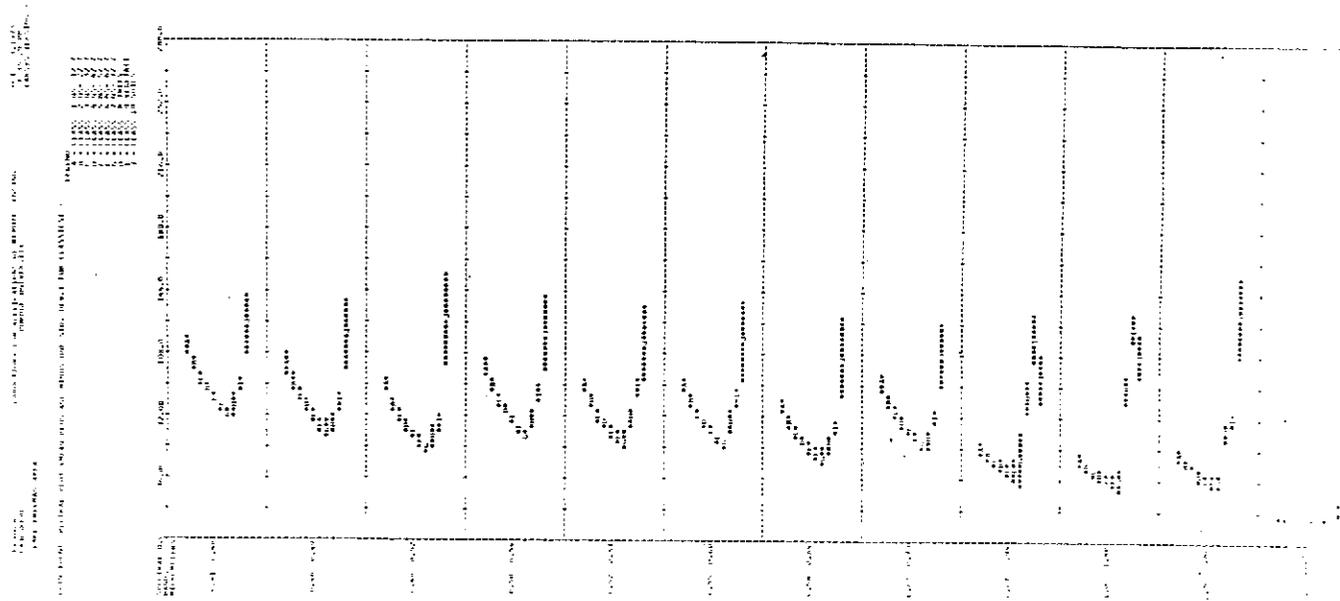
* Environmental Research Institute of Michigan

Figure III-2. Coincident spectral plot of seven spectrally separable classes of water in Lake Freeman, Indiana.
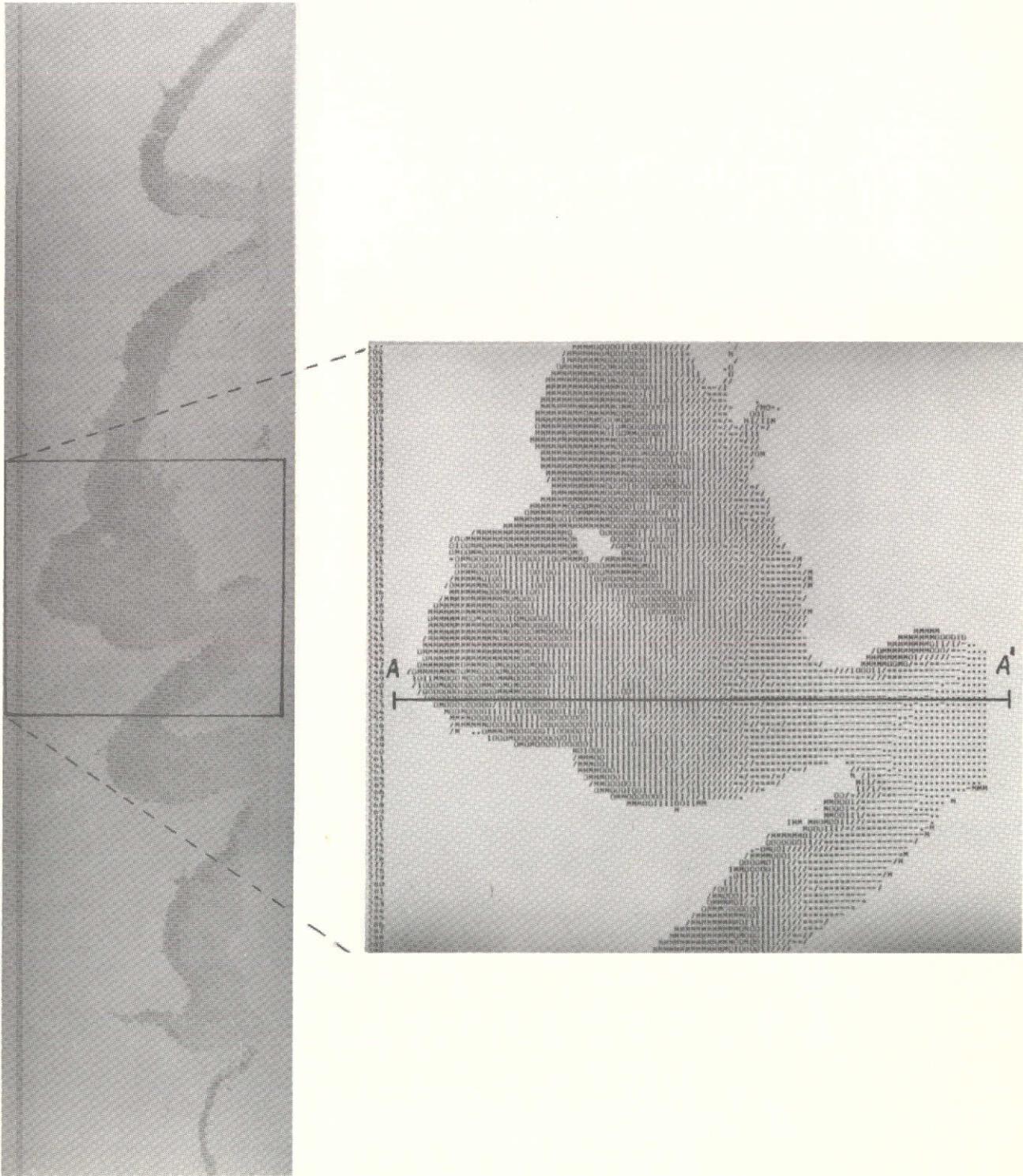
Figure III-3. Seven spectrally separable classes of water in
Lake Freeman, Indiana from Michigan scanner data.

from East-West showing the brightest classes at the East side
of the lake and the darkest on the West side. This pattern
is unlikely due to either water depth or water quality. A
close inspection of the photography taken simultaneously with
the scanner data showed that the glare (specular reflection)
due to the surface roughness had an intensity distribution
similar to the pattern shown in the classification. This
suggests that the seven spectral classes of water defined by
the cluster algorithm define different intensities of specular
reflection, which are a function of the "scanner look angle".
Figure III-4 shows the CRT imagery for the 0.52-0.57 μm band.
Note the glare on the East side of the lake. Figure III-5
illustrates the monotonous change in response of a scan line
across the lake in channels 1 (0.41-0.48 μm) and 10 (1.0-1.4 μm).
However, in channel 12 (9.3-11.7) or thermal band the same
scan line does not show any changes as a function of scanner
look angle; thus, suggesting that the scanner look angle
effect is important in the reflective wavelengths, especially
in the visible region of the spectrum, less important in the
near infrared, and it does not affect the thermal response.

It is also interesting to note in (Figure III-4) that due
to the sun-scan-look angle effect, one side of the imagery
appears light and the other dark. However, for the water body
the inverse is true, that is, on the dark side of the imagery
the water appears light and on the light side the water has a
lower response. This phenomena is observed in all the
reflective wavelength bands. The thermal response (9.3-11.7 μm
band) is not affected by the sun-scan-look angle effect.

The light side of the imagery is caused by the scattering
from the atmosphere, while the light tones in the water on the
opposite side of the imagery is caused by specular reflectance
from the lake surface.

Figure III-4. CRT imagery from 10,000 feet altitude in 0.52 -
0.57 µm band. Note the sun-scan look angle effect
on water and other cover types. Water appears
dark on the side where everything else appears light.

SCAN-LINE GRAPH ACROSS LAKE FREEMAN, INDIANA
(May 4, 1973; 10:00 am; 10,000 ft.)

——— THERMAL BAND (9.3-11.7μm)
— — VISIBLE BAND (0.41-0.48μm)
----- NEAR IR BAND (1.00-1.40μm)

SPECTRAL RESPONSE (in relative units)

232.0
192.0
152.0
112.0
72.0
32.0

WEST BANK OF LAKE
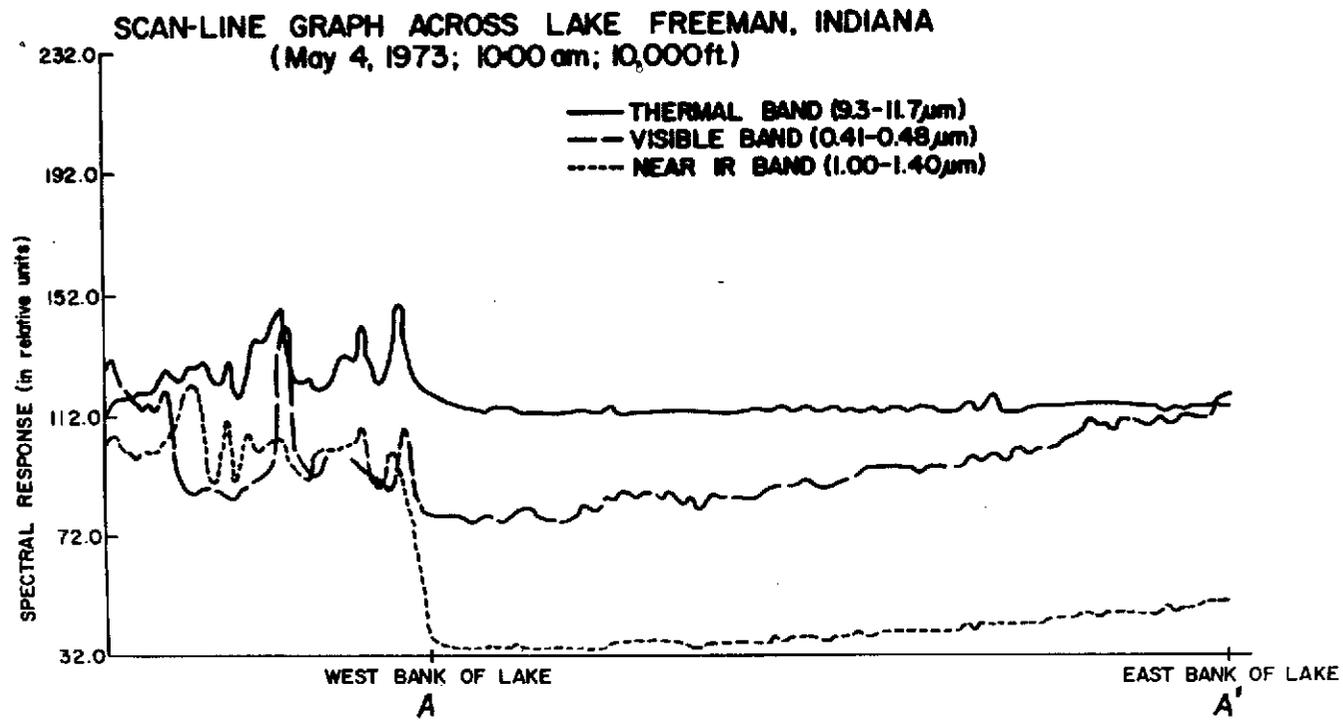A

EAST BANK OF LAKE
A'

Figure III-5. Spectral response graph across Lake Freeman. The location of the
traverse A-A' is illustrated in Figure III-3.

In summary, when strong winds are blowing over water bodies, the effects of the specular reflection from the water wavelets as a function of the scanner look angle override any other spectral characteristic of the water bodies that might be due to either depth or water quality.

Plans for Next Bi-Monthly Period

Analysis of ERTS and aircraft scanner data will be concluded and material for the final report for the contract will be assembled.

IV. Urban Land Use Analysis

As a follow-on to the analysis of the Milwaukee and Indianapolis urban areas reported earlier ERTS data from the Gary, Indiana area was analyzed and quantitative evaluation was made of the results. This area contains a dense urban and industrial complex of great interest to land use analysts.

Multispectral pattern recognition techniques implemented via the LARSYS system were applied to ERTS MSS data gathered on October 1, 1972. The study area includes Gary, portions of the Southern edge of Lake Michigan, and surrounding residential and agricultural areas. A discussion of each of the classes identified is presented and a quantitative evaluation is included.

Categories Identified

Agricultural areas were identified in the Southern part of the study area. This class included crop-land, pasture, and idle land in rural areas, as well as parks, golf courses, and open land in urban areas. Wooded areas are commonly associated with the drainage pattern of the study area. Three principal stands of trees appeared, in conjunction with the Little Calumet River, Deep River, and the Dunes Park area along Lake Michigan. Water, primarily Lake Michigan and other smaller water bodies such as Wolf Lake, was readily identified.

Newer housing developments are located on the fringes of the urbanized area, in the municipalities of Munster, Highland, Griffith, and Merriville. The majority of the structures were built since World War II. Lawns (grass) and streets are the two primary constituents of this spectral class. Therefore, four-lane highways were also classified as newer housing.

Older residentials consists of areas developed prior to World War II. They are found in Hammond, Whiting, East Chicago, and Gary. Closely spaced rooftops, along with mature vegetation (large trees) are the reasons for the spectral separability of this class.

Industrial/commercial areas are usually void of vegetation. They are characterized by the occurrence of rooftops, parking lots, streets, and bare ground. Examples include Inland Steel, U. S. Steel, Standard Oil, Bethlehem Steel, the Gary Central Business District, and Broadway Plaza Shopping Center.

The land use classes identified in this study correspond well with the classes proposed by Anderson, Hardy, and Roach in the U. S. Geological Survey Circular 671 and also with those developed in previous ERTS urban analyses. Further investigations were made, however, into industrial areas in this analysis because of their large areal extent in the Gary-Hammond area.

Five spectral classes of commercial/industrial land use were developed. Two of the classes are associated with closely spaced rooftops; the other three are associated with gravel or sandy areas in industrial areas, adjacent to the rooftop classes. The first rooftop class is associated primarily with dark roofing material, but also with large coal piles. Large areas of this spectral class are associated with the three large steel firms in the study area -- Inland, U.S., and Bethlehem. The other rooftop class is associated with brighter reflecting rooftops. Reasons for the three spectral categories of gravel/sandy areas are not entirely clear at this time but they probably relate to both the color of the material and the presence/lack of sparse vegetative cover. Two large areas of gravel/sandy material are located in the Northwest part of the study area, one between the large

building complexes of Inland and U. S. Steel companies and the second in the large oil refining district in the Whiting-East Chicago area. In the Northeastern part of the study area, another large area of gravel/sand is located West of the buildings of Bethlehem Steel. This area was dominated by one of the three classes of gravel/sand, and had a particularly high spectral reflectance in both the visible and infrared portions of the spectrum. The ground cover in this area is the dune sand typical of this locale.

The final class used in this classification scheme, certain white colored areas in Lake Michigan, is speculated to be smoke coming from coastal industrial establishments. Spectrally, the class is similar to water, having a very dark reflectance in the infrared. Several facts, however, when considered as a whole, lead one to conclude it is smoke. The linear, parallel arrangements of the data points, extending some 30 miles into Lake Michigan, are contrary to the circulation patterns in the lake. Moreover, the meterological records report that the wind was out of the Southwest on the morning of the ERTS pass.

While smoke was probably identified in the Western part of the study area, the smoke data points along the coast in the East were probably water. Moreover, the large area classified as smoke Northwest of Bethlehem Steel was probably a thin cloud. Despite the spectral confusion in these areas, the partial separability does warrant further investigation of the phenomena of smoke located over water bodies.

Classification Accuracy

An attempt was made to determine the classification accuracy by a sampling method. A number of rectangular test areas were determined for each land use, and the class accuracy

determined (Table IV-2). Water, wooded areas, older housing, and newer housing were all identified with over 90 percent accuracy. Trouble was encountered in industrial/commercial areas, most of the misclassification being attributed to older housing. The poorest classification accuracy was in grassy and agricultural areas, where less than 70 percent of the data points were accurately classified. Misclassification of these areas was of two major types. One, areas in agricultural regions associated with darker colored soils proved difficult to separate from older housing. One such area was located south of Little Calumet River, in Munster and Highland. The other type of misclassification was in undeveloped marshland adjacent to industrial areas. Large areas South of U. S. Steel and along U. S. Highway 12 were misclassified as older housing.

Conclusions

Important tabular data can be generated from the machine processing of ERTS data. Table IV-3 contains an estimate of the proportion of the study area (excluding Lake Michigan) allocated to the various land uses, obtained by a simple tallying of the numbers of data points in each land use. Adjustments were made for agricultural/grassy areas, commercial/industrial, and older housing, relative to the misclassification between these three land uses. Acreages were obtained by multiplying the number of data points by 1.1, the approximate resolution of ERTS. The data in Table IV-3 could have been reported by smaller areal units, such as municipalities, townships, or census tracts, by storing the desired boundaries in the computer.

Plans for Next Bi-Monthly Period

This analysis essentially concludes work on this project. Refinements will be made in results and materials assembled for the final report in the following months.

Table IV-2. Classification accuracy for test samples

Land Use	Percentage of Data Points Classified As:					
	C/I[1]	OHg[2]	NHg[3]	Wod[4]	A/G[5]	Wat[6]
Cmrc/Inds.[1]	<u>89.8</u>	7.3	0.3	-	0.2	2.5
Old Hsng.[2]	0.9	<u>97.9</u>	0.9	0.3	-	-
New Hsng.[3]	0.6	4.0	<u>94.0</u>	-	1.4	-
Wooded	-	2.0	-	<u>94.4</u>	3.5	-
Agrc/Grsy[5]	0.8	27.4	3.2	3.1	<u>65.5</u>	-
Water	0.8	-	-	-	-	<u>99.2</u>

$\bar{X}$ Classification accuracy by class = 90.3%.

[1]Commerce/Industry

[2]Older Housing

[3]Newer Housing

[4]Wooded

[5]Agricultural/Grassy

[6]Water

Table IV-3. Land use area calculations for study area
            (excluding Lake Michigan).

Land Use	Number of Data PTs.	Number of Acres	Number of Hectares	% of Study Area
Cmrce/Indstry[1]	25766	28343	11479	8.2
Older Housing[1]	56528	62181	25183	18.0
Newer Housing	28540	31394	12714	9.1
Wooded	52346	57581	23320	16.6
Agric/Grassy[1]	150982	166080	67262	48.0
Water	499	549	222	0.2
TOTAL	314661	346127	140181	100.0

[1]Adjustments made in accordance with test classification
accuracy (see Table IV-2).

V. Soil Association Mapping

Introduction

During this reporting period ERTS data of improved quality
was obtained. This data set was collected on June 9, 1973,
over Tippecanoe County, Indiana, when 55% of the area was
cultivated, permitting an improved capability for mapping soil
characteristics. Improved techniques for computer analysis were
used in conjunction with improved imagery output (color coded
computer classifications and false color images produced using
the digital image display). We extracted much more information
relating to soil associations during this reporting period.

Procedures

The task being pursued in this ERTS-1 study is concerned
primarily with comparison of generalized county soil maps with
multispectral maps produced by computer analysis of ERTS MSS data.
Initial investigation of discriminability of individual soil
types for more detailed mapping are also being conducted.

The procedure being followed in determining the utility of
ERTS data involves direct comparisons of ERTS images of various
types to existing soil association maps. Techniques were
developed during the course of the project to facilitate making
such comparisons independent of the spectral characteristics of
the soils. Essentially, this is accomplished by the digital
geometric correction of ERTS MSS data where necessary (in the
computer analysis phases) and direct overlay of conventional soil
maps onto the ERTS images.

In each case an attempt was made to draw boundaries between
generalized soil associations onto the image. This was followed
by comparison with a soil association map in transparent overlay
form. The overlay was photographically manipulated to the same

scale as the ERTS image containing the interpreted boundaries.
The map of soil associations is given in Figure V-1.

Results

An inspection and test of data from various dates has verified
the original assumption that data gathered at a time of maximum
soil exposure provides the best results. This occurs in late
Spring in the Wabash Valley. For example on June 9, 1973 it was
found that 55% of Tippecanoe County, Indiana was nonvegetated.

The computer classification was produced using a nonsupervised
classification of the four band data. An example is shown in
Figure V-2 which includes the soil map overlay. This image was
produced in color, and much contrast is lost in the black and
white rendition shown in Figure V-2. It was found in this case
that the data could be partitioned into 10 multispectral classes
such that none were spectrally similar to one another.

In Figure V-2 the very dark colors represent vegetated areas
while intermediate and light colors show nonvegetated soils. A
close inspection of the image relative to the map boundaries
reveals that some generalized soil boundaries are mislocated on
the conventional map. An example is seen in the upper left part
of the image, where a light colored area extends from soil
association 81 (Miami-Russell-Fincastle) across the area mapped
as soil association 73 (Raub-Ragsdale) and into soil association
89 (Sidell-Parr). The Raub-Ragsdale soil association would not
typically contain such large delineations of these well-drained
soils.

Several soil associations can be delineated very well in
Figure V-2. Soil association 16 (Elston-Wea) is very distinctive
in the left center of the image. Soil association 66 (Fincastle-
Ragsdale-Brookston) has a distinctive appearance in the upper
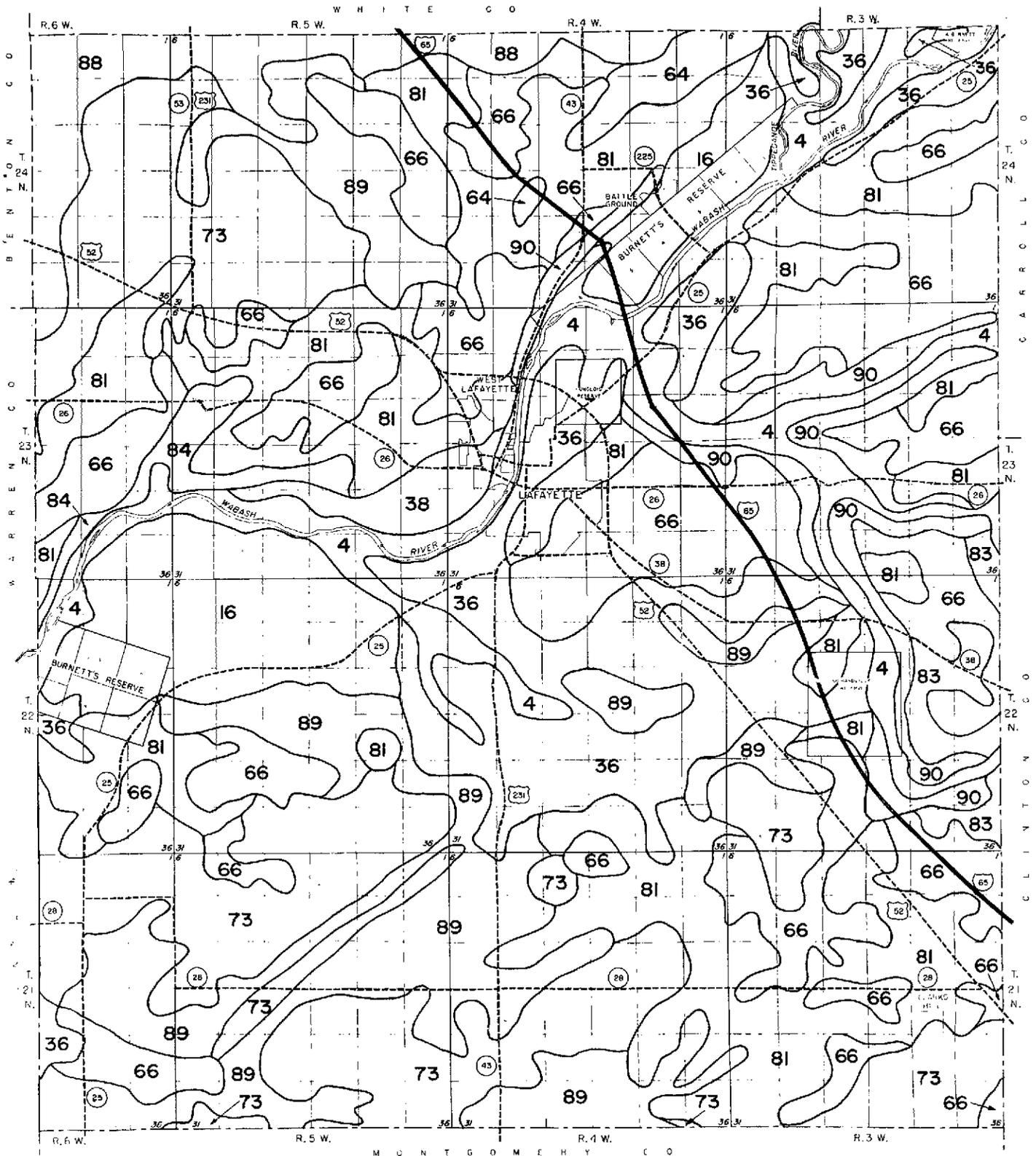right portion of the image as well as the other areas. Soil

Figure V-1.   General Soil Map of Tippecanoe County, Indiana.
Soil Association legend is on following page.

TIPPECANOE COUNTY

General Soil Map

AGRICULTURAL EXPERIMENT STATION AND COOPERATIVE
EXTENSION SERVICE, PURDUE UNIVERSITY; AND THE SOIL
CONSERVATION SERVICE, U.S. DEPARTMENT OF AGRICULTURE

Scale 1 : 190 080

1   0   1   2   3   4 Miles
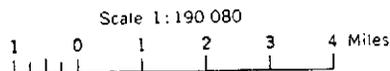
Base map from the Indiana Geological Society

*Note: This map is intended for general planning. Each delineation contains
soils different from those shown in the legend. For operational planning,
use detailed soil maps that may be available in published or unpublished
form at the local Soil and Water Conservation District Office.*

SOIL ASSOCIATIONS

4. *Genesee-Shoals-Eel:* Nearly level, well drained, loamy Genesee,
moderately well drained, loamy Eel, and somewhat poorly drained,
loamy Shoals in alluvial deposits.

16. *Elston-Wea:* Nearly level, well drained, loamy soils on outwash sand
and gravel.

36. *Ockley-Westland:* Nearly level, well drained, loamy Ockley on
outwash sand and gravel.

38. *Ockley-Fox:* Nearly level, well drained, loamy soils on outwash sand
and gravel.

64. *Crosby-Brookston:* Nearly level, somewhat poorly drained, clayey
Crosby and very poorly drained, loamy Brookston in glacial till.

66. *Fincastle-Ragsdale-Brookston:* Nearly level, somewhat poorly drained,
silty Fincastle in wind-blown silts and glacial till, very poorly
drained, silty Ragsdale in wind-blown silts and loamy Brookston in
glacial till.

73. *Raub-Ragsdale:* Nearly level, somewhat poorly drained, silty Raub in
wind-blown silts and glacial till and very poorly drained, silty
Ragsdale in wind-blown silts.

81. *Miami-Russell-Fincastle:* Sloping, well drained, loamy Miami in glacial
till and silty Russell in wind-blown silts and glacial till and nearly
level somewhat poorly drained, silty Fincastle in wind-blown silts
and glacial till.

83. *Miami-Crosby:* Sloping, well drained, loamy Miami and nearly level,
somewhat poorly drained, clayey Crosby in glacial till.

84. *Miami-Hennepin:* Sloping, well drained, loamy Miami and steep, well
drained, shallow, loamy Hennepin in glacial till.

88. *Odell-Chalmers:* Nearly level, somewhat poorly drained, loamy Odell
and very poorly drained, loamy Chalmers in glacial till.

89. *Sidell-Parr:* Sloping, well drained, silty Sidell in wind-blown silts and
glacial till and loamy Parr in glacial till.

90. *Hennepin-Rodman:* Steep, well drained, shallow, loamy Hennepin in
glacial till and excessively drained, shallow, sandy Rodman on sand
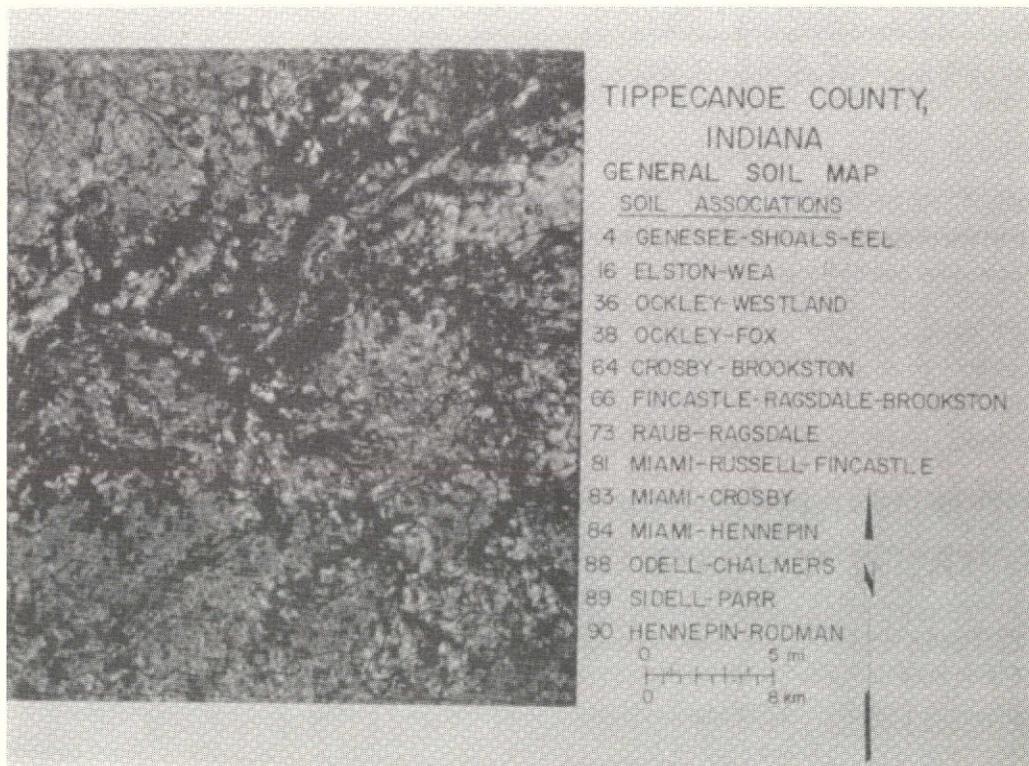and gravel.

N

Figure V-2.   Computer soil association classification
reproduced in gray scale form with soil
association map drawn on computer produced
map. Dark areas are vegetated. Intermediate
and light colors are soils. (Original in color).

associations 73 and 89 are easily delineated from soil associations
66 and 81 in several areas.

Table V-1 shows results of comparing a false color electronic
image and a computer classification image with the soil association
map. The computer classification most closely agreed with the
soil map, and the discrepancies noted are associated with errors
in the conventional map to a large extent.

Table V-2 gives results of the investigation of spectral
separability of individual soil types. The overall accuracy
of 89% indicates that ERTS data can be used to assist in
identification and mapping of individual soil types.

Conclusions

Although these studies have not been completed, it seems
that several positive conclusions will be forthcoming. Geometric
correction of ERTS MSS data and direct overlay of existing
conventional soil maps facilitated specific and precise compar-
isons. Used together, these two sources provide increased
information in soil characteristics which should be extremely
useful in the National Cooperative Soil Survey. Computer
classification images contained more soils information than any
other images investigated. This advantage of computer analysis,
coupled with the advantages of being compatible with other data
storage and retrieval systems for soils and land use data make
the computer analysis approach highly desirable. Since soils
are a major consideration in land use and land use planning,
this type of analysis will likely become of even greater utility
in the next few years.

Plans for Next Period

Improvement of illustration quality and refinement of
results will be carried out in preparation for final report
material generation.

Table V-1.   Soil Association Mapping Comparison

Soil Association	Percent Difference Within Association	
	Color Electronic Image	Computer Classification Image
73 Raub-Ragsdale	3.5	8.5
89 Sidell-Parr	1.1	2.7
16 Elston-Wea	0	0
81 Miami-Russell-Fincastle	2.0	3.8
66 Fincastle-Ragsdale-Brookston	3.0	3.1
Average (Weighted)	15.0	7.3

Table V-2.   Soil Series Separability Test

Soil Series	Percent Correct (Training Samples)	
Ragsdale	90.6	
Fincastle	81.8	
Genesee	100.0	
Wea	98.6	
Sidell	84.0	
Miami	77.3	
Ockley	76.9	
Overall	89.0	246 Data Points

VI. Earth Surface Features Identification

Introduction

In the previous reporting period it was determined that relatively good accuracy for identifying earth surface features from ERTS could be achieved. Of major concern was the forest cover classification and since it was determined that some inaccuracies may have been present a new classification of ERTS was acquired. During this past reporting period, along with the acquisition of a reclassified frame for forest cover from ERTS imagery, the forest cover information in the data base that is being utilized for comparative purposes was updated.

Procedures

The test site classified and discussed in the previous Type II report was reclassified using overlayed data from two times. Table VI-1 shows the percentage classification of the test area in four classes: agriculture, forest, urban and water. The table contains results obtained from use of the September 30, 1972 data temporally overlayed with the June 9, 1973 data. This data was geometrically corrected to the 1:24,000 scale for use with 7 1/2 minute quad topographic maps. The analysis procedure used to obtain this classification was as follows: Two relatively small areas were clustered which gave representative samples of urban, commercial, forest, agriculture and water areas. The clustering was done using the complete eight channel set thereby including the temporal information in the definition of the training set. The classes used in the final classifications were selected from the cluster maps on the basis of photointerpretation using high altitude

aircraft infrared photography and relating the area being clustered to the identical area as shown in the infrared photography.

Results

This classification was found to produce a fairly good overlay with the topographic map and showed excellent correlation between the water areas, forest areas, agriculture and urban and commercial areas in the portion of the Lafayette metropolitan area.

Unfortunately we have no good means of evaluating the accuracy of the classification since no reliable data is available which gives information on the amount of acres in each of the classification classes. The results have been evaluated subjectively by comparison of the classification to the information available on the USGS topographic maps and aerial photography. Figure VI-1 contains a manually extracted forest cover map from the high altitude aerial photography which was enlarged to 1:24000 scale. The areas classified as forest from the ERTS data then manually extracted from the computer printout and this is presented in Figure VI-2. The two extractions were then overlayed and a comparison illustration was made, Figure VI-3, showing areas of disagreement as a dark shade and areas of forest agreement as cross hatch. Greatly improved agreement was observed in this result compared to the previous result presented in the June 73 Type II report. Table VI-1 gives the percentage of the area classified and the various classes for four separate classifications using the same data set. The first classification used all eight data channels and produced the best overlay with the topographic map. Two additional classifications were made using the same statistics and the same data set.
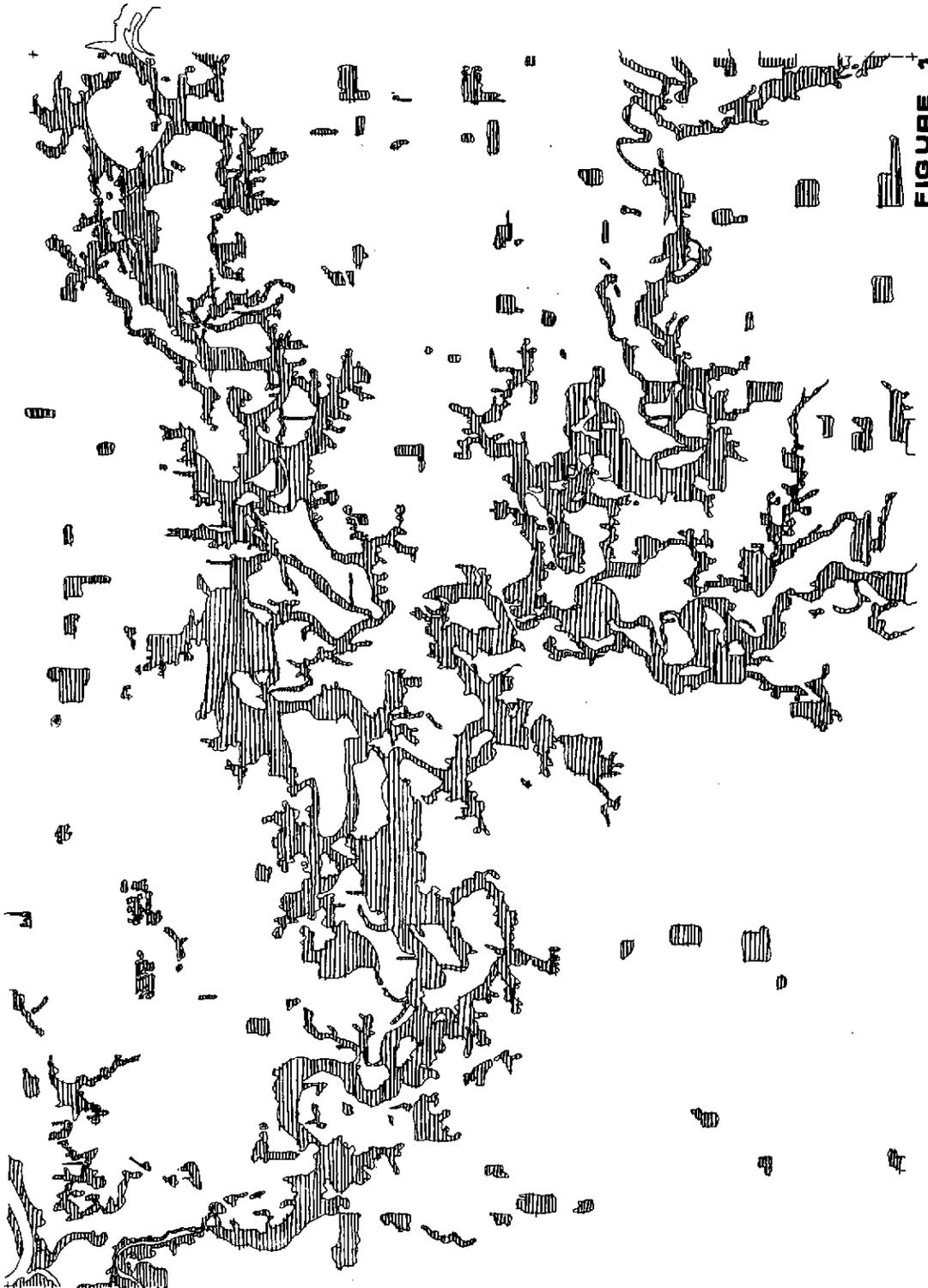
Figure VI-1.   Forest cover in the test site extracted from 1971
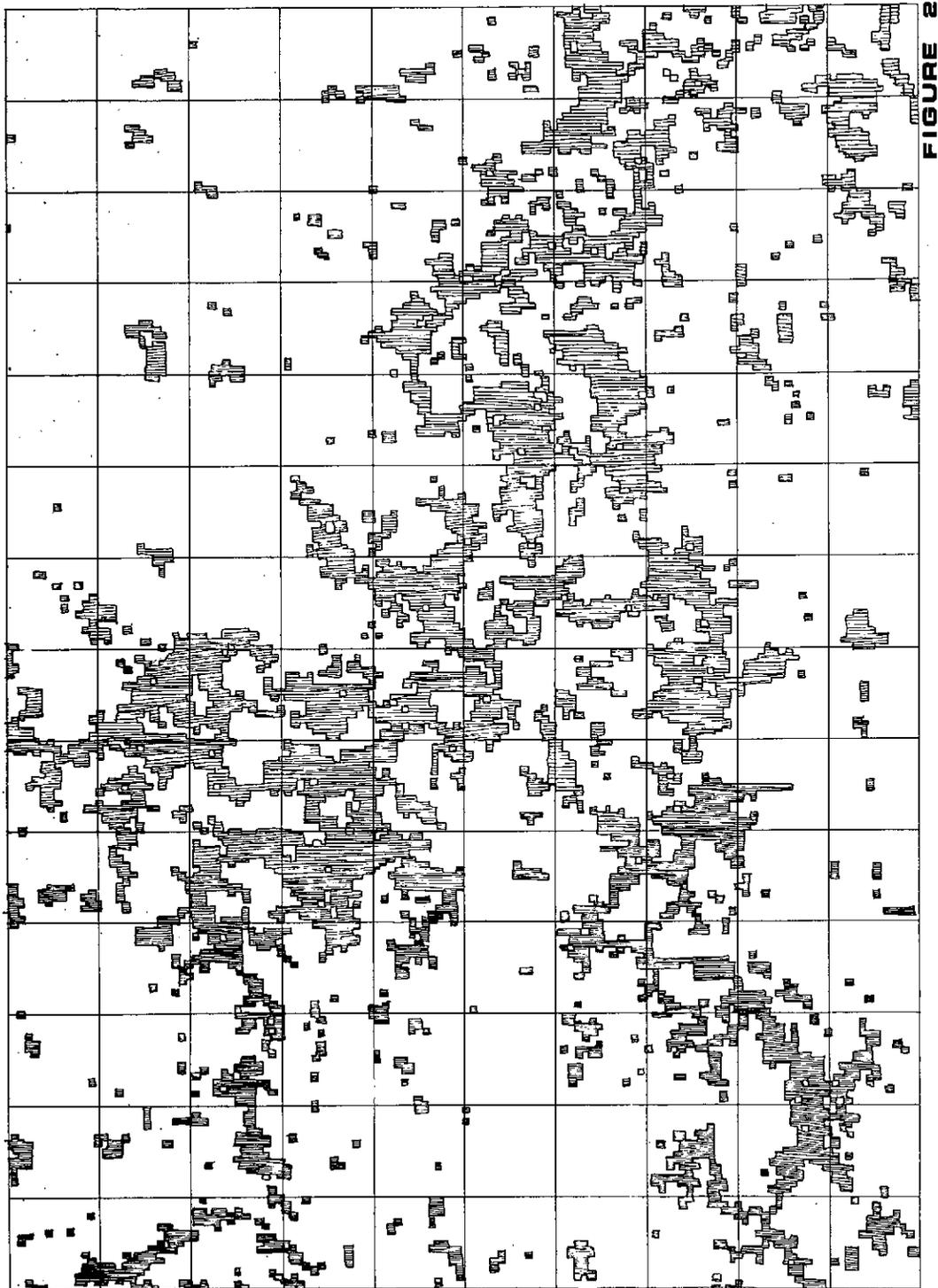   high altitude aerial photography.

Figure VI-2. Forest cover extracted from ERTS data computer
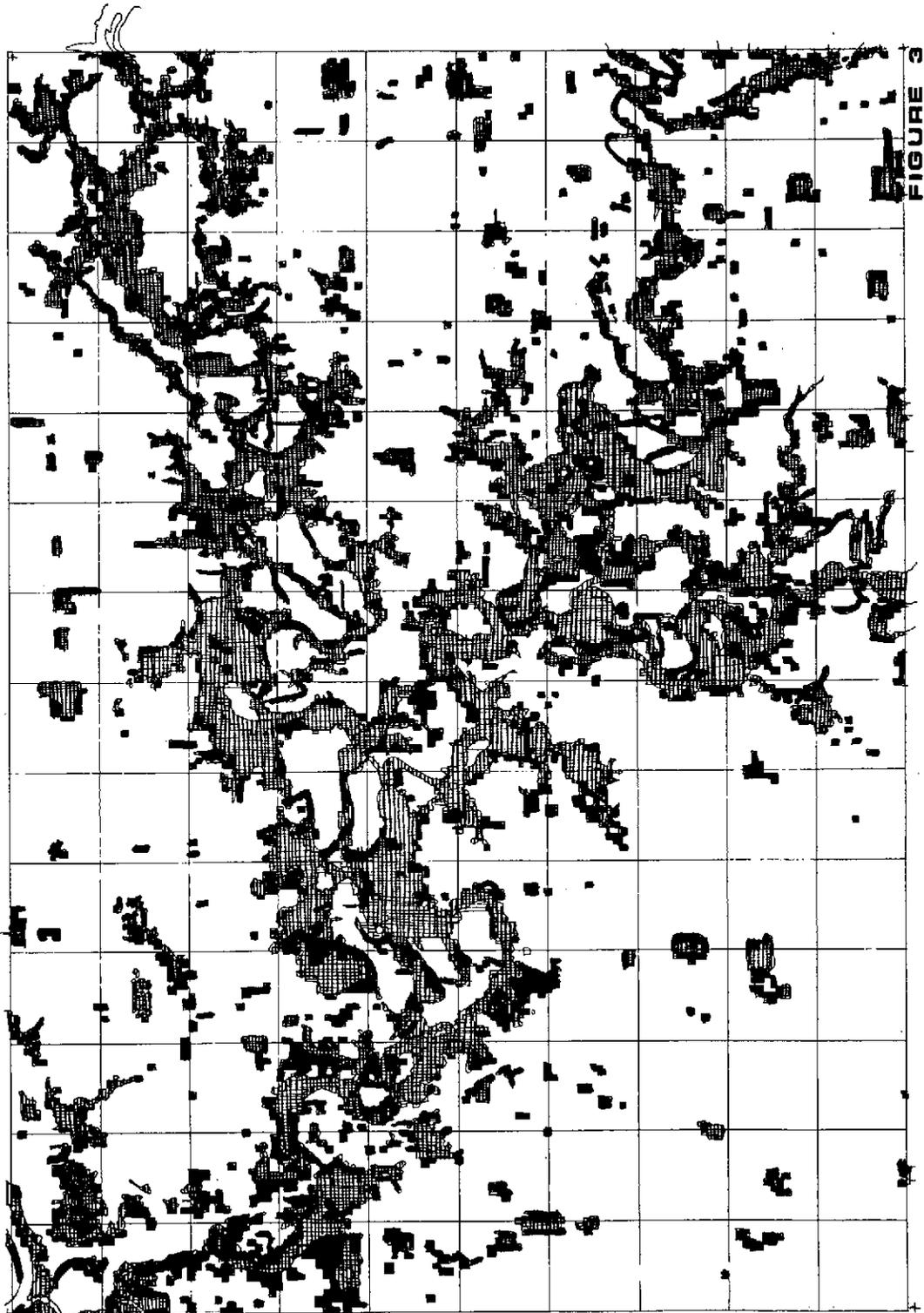classification printout.

Figure VI-3. Comparison of Figure VI-1 and 2 showing areas of
discrepancy as a dark shade and areas of forest
agreement as a cross-hatched pattern.

However in this case the channels representing September data were classified in a second classification. An additional classification was made using the "best four" channels as selected by use of the SEPARABILITY processor. The "best four" set of channels was selected from examination of output from this processor and it was found that the majority of the top rated 10 to 15 channels consisted of an infrared and a visible channel selected from each date. The relative evaluation of these channels indicated that they were essentially equivalent in their ability to separate the classes existing in the training sets.

The interesting point of this table is that while some loss in accuracy in agriculture and urban and forest classes existed in September data alone and also in the June data alone, the results for the best four channels are essentially identical to those of the complete eight channel data set. This suggests the possibility that temporal overlays might be used on small areas to produce training sets which could then be used for classification using data from only one date. Also the success of the classification using only four channels is encouraging since the computer time required for four channels is far less than that required for eight.

Table VI-2 shows the results of separate classification made using only September data but in this case the entire classification process used only September data, therefore excluding any temporal information from the definition of training sets. The same geographic area was classified and evaluation was made from aerial photography in the same manner as the previous classification. However, the classification contains a different number of classes and the results are shown to be quite different from any of the previous four. This classification exhibits the problems which have been

Table VI-1.   Percentage of Area Classification Using Complete
Eight Channel Overlay Data Set and Subsets of
Data

Data Set	Agriculture	Forest	Urban	Water
Sept. & June	75.7	18.9	4.9	0.5
Sept.	71.0	23.3	5.2	0.5
June	70.8	20.4	8.5	0.3
"Best 4"	75.8	18.5	5.3	0.4

Table VI-2.   Percentage of Area Classification Using Only
September Data for Entire Analysis Procedure
(no temporal information in training set
selection)

Agriculture	Forest	Urban	Water
62.8	22.6	11.5	3.1

encountered with many classifications in which agriculture
and urban areas have been mixed together.

Conclusion and Plans

The comparison of the two tables indicates the value
of the temporal information in both definition of training
sets and in classification. It is believed that the
greatest value is in the definition of training sets since
no great differences are seen in the urban and agriculture
areas in either September data or June data when classified
alone using training sets defined with temporal information
available.

Although no percentage accuracies were arrived at, the
forest classification agreement appeared to be well beyond
eighty-five percent. In the next period both the reclassified
data from ERTS and that from the high altitude data base
will be extracted and put into cell format for accuracy
comparisons. Upon conclusion of this phase the classified
data from ERTS will be subjected to a process of automatic
data entry. By this it is proposed to automatically put ERTS
data into the data bank after classification and then compare
the spatial accuracy of ERTS entry into the data bank against
the existing high altitude data base. An attempt is also
being made at comparing two scales of data entry - a one
tenth kilometer square and a one fifth kilometer square. By
this it is hoped to determine that if by increasing the cell
size utilized for earth surface identification on a regional
scale, an increased accuracy of identification can be
achieved. Thus, as stated previously, if this process can
be accomplished and spatial accuracy achieved along with
accurate ERTS data classification, a semi-automatic system
of data bank development for utilization in land use planning
can result.

VII. Atmospheric Modeling

Introduction

Meteorological conditions at the time of an ERTS overpass determine the amount of radiant energy that is transmitted through the atmosphere to the sensor. The atmospheric model in use at LARS must have realistic measures of these meteorological variables if it is to be of use in improving the LARSYS classification accuracy.

Procedures

Three of the required inputs are the absorption coefficient for ozone, the absorption coefficient for water vapor, and the Rayleigh scattering optical thickness. These are wavelength dependent parameters. Since the model requires a single value of wavelength as input, a spectrally averaged absorption coefficient for ozone, $\overline{\alpha}_{O_3}$, and water vapor, $\overline{\alpha}_{H_2O}$,

and a spectrally averaged Rayleigh scattering optical thickness, $\overline{\tau}_b$, must be derived. In direct correspondence with the averaging effect of the sensor's optical system on the signal produced in each band, $\overline{\alpha}_{O_3}$, $\overline{\alpha}_{H_2O}$, and $\overline{\tau}_b$ were found using:

$$\overline{\alpha}_{O_3}(i) = \frac{\int_0^{\infty} \alpha_{O_3}(\lambda) \; E_i(\lambda) \; d\lambda}{\int_0^{\infty} E_i(\lambda) \; d\lambda} \qquad i = 1, \; 2, \; 3, \; 4$$

$$\bar{\alpha}_{H_2O}(i) = \frac{\int_0^\infty \alpha_{H_2O}(\lambda)\, E_i(\lambda)\, d\lambda}{\int_0^\infty E_i(\lambda)\, d\lambda} \qquad i = 1,\, 2,\, 3,\, 4$$

$$\bar{\tau}_b(i) = \frac{\int_0^\infty \tau_b(\lambda)\, E_i(\lambda)\, d\lambda}{\int_0^\infty E_i(\lambda)\, d\lambda} \qquad i = 1,\, 2,\, 3,\, 4$$

where $E_i(\lambda)$ is the optical filter response of the ith ERTS band. These computations were made using the average filter response curves shown in Figure VII-1. The results are tabulated in Table VII-2A.

Lee, Ogle and DeKalb Counties in Northern Illinois were chosen as a case study to investigate the atmosphere's effect on classification accuracy. Meteorological data were gathered from the National Meteorological Center for the area at the time of the overpass on August 9, 1972. An upper-air radiosonde sounding from Peoria and ground observations from Peoria, Rockford, and O'Hare Airport were all received. From them, vertical distributions of ozone, water vapor, and aerosol concentration were deduced. Based on climatological data for Northern Illinois, a complex index of refraction of 1.33-0.0i corresponding to a water-base aerosol was chosen. Land-use information and spectro-photometer data for the prominent surface crops in the three county area were found. By weighting the crop reflectance in a band with the percentage of land occupied by that crop in August 1972, a spectral surface reflectivity was computed. With the average absorption coefficients, the average Rayleigh scattering optical thickness, the
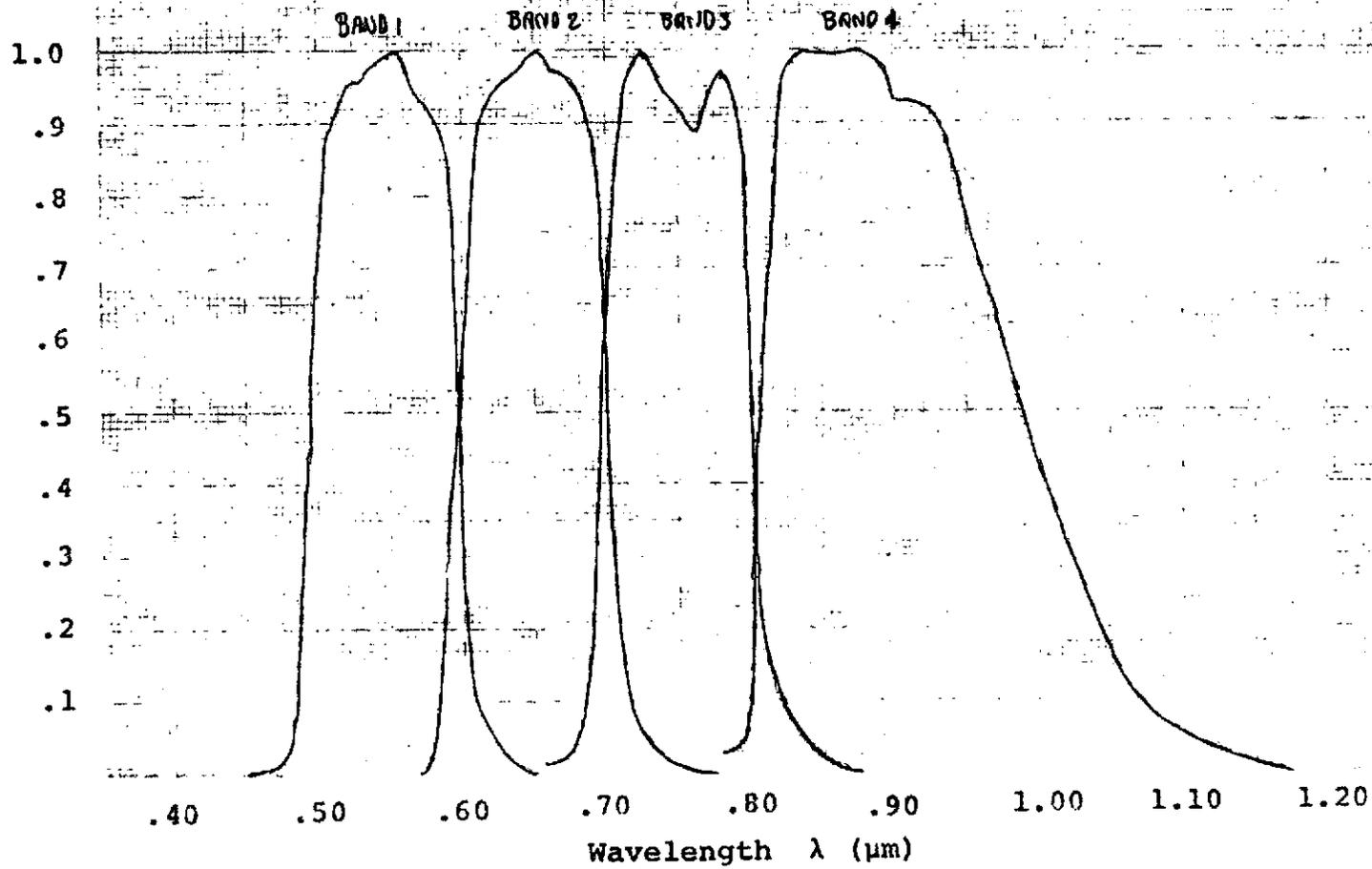
Figure VII-1. Response of ERTS spectral bands based on average filter response curves.

Figure VII-2A. Calculated values for atmospheric
parameters for the four ERTS bands.

ERTS Band	$\overline{\alpha}_{O_3}$	$\overline{\alpha}_{H_2O}$	$\overline{\tau}_b$
1	0.0826	0.0	0.1028
2	0.0689	0.0	0.0514
3	0.0	0.0184	0.0298
4	0.0	0.1255	0.0147

Figure VII-2B. Results of atmospheric model evaluation.

ERTS Band	Surface Reflectivity	Atmospheric Transmission (%)
1	0.1435	105.69
2	0.1096	89.88
3	0.4135	75.52
4	0.4718	61.90

surface reflectivities, and the vertical distributions, the atmospheric model was run for each of the four ERTS bands.

Results and Plans

The results of the model evaluation are tabulated in Table 2B. In band 1, 105.69% of the radiant energy leaving the surface reached the sensor. With this data in hand, work is proceeding to determine what the impact actually is on the accuracy with which surface features can be identified. In the next period classification evaluation will be carried out and plans for the material to be included in the final report will be made.

VIII. ANALYSIS TECHNIQUE DEVELOPMENT

Work on the use of context for classification of objects in multispectral images was completed and documented in the reporting period. A PhD Thesis was prepared entitled, "Multispectral Image Partitioning" and a paper describing the work entitled "Extraction and Classification of Objects in Multispectral Images" was presented at the Purdue Conference on Machine Processing of Remotely Sensed Data. Work continued on layered classifier and adaptive classifier research throughout the period and will be reported in the final report of the ERTS study.

IX. REFORMATTING AND OVERLAY

Reformatting and maintenance of the ERTS-1 CCT data for the LARS data bank proceeded without problem during the period. Overlay of sequential frames of ERTS data was supplied as a data support service as described in earlier reports. Over 201 LARSYS data sets were generated from MSS CCT's during the period.

A useful advancement in all digital geometric correction of ERTS data was achieved during the period. Previous developments reported in the June 1973 Type II report have enabled rotation, rescaling and deskewing operations to be performed such that images generated on LARS output devices would have a specific scale factor and be North-oriented. This correction is approximate and produces results which can be in error by several hundred meters over a distance of several kilometers. Correction of the image geometry to an accuracy of one sample requires the measurement of matching points in a reference and in the data. These ground control

points are then used to recorrect the image geometry.

An experimental precision correction was carried out
in conjunction with a project funded by the U. S. Geological
Survey and the results were excellent as determined by visual
inspection. CCT data from ERTS frame 1003-18175 was
approximately corrected for scale, rotation, and skew using
previously discussed techniques. The data was scaled so
that when printed in pictorial form on a computer line
printer the scale is approximately 1"=24000". Easily
identifiable features such as schoolyards and parks were
located on 1:24000 topographic maps by USGS personnel. The
corresponding areas were located in the ERTS data printouts.
The map used was USGS 7 1/2 minute quad-San Jose West.
Thirty-six matching points were found covering a 10 x 7 1/2
mile area. The coordinate system used for the map points
was the UTM system. Vertical and horizontal coordinates
were measured to the nearest 10 meters and punched in standard
LARS checkpoint format on cards along with the line and
column coordinates for the same point in the data. These
coordinates were processed by a geometric distortion function
estimation program and parameters were computed to correct
the remaining geometric error in the data for the given area.
The data was then re-geometrically corrected to produce the
final version. The results were overlayed on the topographic
map to inspect the accuracy of the fit. No error could be
visually observed over the 7 1/2 x 10 mile area although it
is extremely difficult to estimate locations to better than one
or two pixels in ERTS-1 data. Further work will be done to
develop a procedure for making all digital precision geometric
corrections when necessary.

X. Summary Items

A. New Technology

No new technology items were developed during the period. The context algorithm referred to in Part VIII was developed in the previous six month period and reported on in the June 73 Type II Report. The precision correction scheme referred to in Part IX is experimental and not considered a new technology development.

B. Program for Next Period

Results refinements and preparation of a final report will be pursued in the final six month period.

C. Conclusions

Conclusions are presented in the individual sections.

D. Recommendations

No recommendations are made at this time.

E. Type I Items

1. No operational changes are recommended.
2. No changes in the standing order form are requested.
3. One image descriptor form is attached.

ERTS IMAGE DESCRIPTOR FORM
(See Instructions on Back)

DATE **January 30, 1974**

PRINCIPAL INVESTIGATOR **David A. Landgrebe**   ← PLEASE FILL IN

GSFC **UN127**

ORGANIZATION **LARS - Purdue University**

Idaho →

⊂Illinois →

PRODUCT ID (INCLUDE BAND AND PRODUCT)	FREQUENTLY USED DESCRIPTORS*			DESCRIPTORS
Scene Id				
1034-17473	Cropland			
1035-17525	"			
1017-16093	Cropland	Corn	Soybeans	
1017-16100	"	"	"	

*FOR DESCRIPTORS WHICH WILL OCCUR FREQUENTLY, WRITE THE DESCRIPTOR TERMS IN THESE COLUMN HEADING SPACES NOW AND USE A CHECK (✓) MARK IN THE APPROPRIATE PRODUCT ID LINES. (FOR OTHER DESCRIPTORS, WRITE THE TERM UNDER THE DESCRIPTORS COLUMN).

MAIL TO   ERTS USER SERVICES
CODE 563
BLDG 23 ROOM E413
NASA GSFC
GREENBELT, MD. 20771
301-982-5406

GSFC 37 2 (7/72)