E7.3 10862
CR-133451

A STUDY OF THE UTILIZATION OF ERTS-1 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

(E73-10862) A STUDY OF THE UTILIZATION OF ERTS-1 DATA FROM THE WABASH RIVER BASIN Semiannual Progress Report, 1 Jan. - 30 Jun. 1973 (Purdue Univ.) ~~70~~ 68 p HC $5.50 CSCL 05B G3/13 N73-29213 Unclas 00862

July 1973
Type II Report for Period January 1-June 30, 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
July 1973

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

12. Sponsoring Agency Name and Address
Dr. Arthur Fihelly
NASA Goddard Space Flight Center
Greenbelt Road
Greenbelt, Maryland 20771

13. Type of Report and Period Covered
Type II
January 1 To
June 30, 1973

14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

Work performed during the second six month period of the study is described. Nine projects are defined, five ERTS data applications experiments and four supporting technology tasks. The most significant applications results were achieved in the Soil Association Mapping, Earth Surface Feature Identification and Urban Land Use Mapping efforts. Four soil association boundaries were accurately delineated from ERTS-1 imagery. A data bank has been developed to test surface feature classifications obtained from ERTS-1 data. Preliminary forest cover classifications indicated that the number of acres estimated tended to be greater than actually existed by 25%. Urban Land Use analysis of ERTS-1 data indicated highly accurate classification could be obtained for many urban categories. The wooded residential category tended to be misclassified as woods or agricultural land. Further statistical analysis revealed that these classes could be separated using sample variance.

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*

*For sale by the Clearinghouse for Federal Scientific and Technical Information, Springfield, Virginia 22151.

iii PREFACE

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 soil association mapping, earth surface feature identification and urban land use mapping are presented.

D. Summary of Recommendations

Recommendations are made in Section VI for a study of land use change in Marion County, Indiana using geometrically corrected and temporally overlayed data.

TABLE OF CONTENTS

I. Introduction

This report describes work performed during the second six-month period (January 1 - June 30, 1973) of the Purdue University - LARS ERTS-1 Wabash Valley Study Contract. Nine projects are defined in the Data Analysis Plan and this report contains progress and status information for each one of these areas. The required reporting format as specified in GSFC Spec. S-250-P1C has been adhered to as strictly as possible; however, some sections did not lend themselves to the strict format. The first five projects are ERTS-1 data applications experiments and the last four are supporting technology activities in support of the applications projects.

Section II presents progress and results from the Crop Species Identification Project. ERTS data from Northern Illinois and Southeastern Missouri were analyzed to determine the accuracy with which major crop species could be identified using computer techniques. Section III describes valuable new procedures and important results in the Mapping of Soil Associations Project. Photographic overlay procedures were developed which enabled soil maps and ERTS-1 prints to be overlayed enabling precise analysis of the scene with respect to the map. Encouraging results regarding the usefulness of ERTS-1 data for soil association mapping are presented. Section IV discusses Water Resources Research into the usefulness of ERTS Satellite data and aircraft underflight data for recognition of surface water features. Several problems have been encountered including inadequate satellite data spatial resolution and aircraft scanner sun angle effects. Analysis of field spectrometer data from the water test sites is also discussed. Section V details work done on the Earth Surface Features Identification Project and includes progress on a spatial data bank which has been developed. The data bank is being used to test the accuracy of ERTS-1 data derived classifications for a test site in Tippecanoe County, Indiana. Section VI discusses results of a detailed Urban Land Use Analysis of Marion County, Indiana using ERTS-1 data.

Section VII begins the discussion of progress on the supporting technology tasks with a discussion of Analysis Technique developments. Section VIII lists data reformatting and temporal overlay tasks completed in support of digital ERTS-1 data analysis portions of the study. In Section IX the Atmospheric Modelling project is reported on; and Section X discusses problems and presents conclusions on the System vs Scene Corrected ERTS-1 MSS data comparison experiment. Section XI contains the additional items required for Type II reports.

II. Crop Species Identification

Introduction

During the past six months further analyses utilizing the ERTS-1 data collected over Northern Illinois in August 1972 were completed and a cooperative research project with the Statistical Reporting Service (SRS), U. S. Department of Agriculture, was initiated. The primary additional accomplishment with the Northern Illinois data set was the classification of three entire counties and comparison of area estimates to SRS/USDA estimates of corn, soybean, and "other" acreage. The overall objective of the research with SRS is to evaluate the utility of machine analysis of ERTS-1 data for identifying major crop species.

A. Crop Acreage Estimates for Three Northern Illinois Counties

Procedures

Following the classification of the three individual counties, DeKalb, Ogle, and Lee (reported in our previous Type II Report), the statistics from the three areas were combined into one training set and the entire three county area classified. The total number of points classified into each class, corn, soybean, and "other" were tabulated for each county. Using error rates obtained from the classification of test fields the county totals were adjusted for classification bias. The proportion of the area in each class was then computed and compared to estimates made by the USDA.

Results

Comparison of USDA acreage estimates with estimates derived from computer analysis of ERTS-1 data for DeKalb, Ogle, and Lee Counties, Illinois.

	USDA	ERTS
	(Percent of Total Area)	
Corn	40	40
Soybeans	17	23
"Other"	43	37

Conclusions

The first comparisons of area estimates made by conventional methods and those obtained from ERTS data show the feasibility of using remote sensing techniques to obtain crop production information over large geographic areas.

B. Analysis of Southeast Missouri ERTS-1 Data for Crop Species Identification

Procedures

Crop Reporting District No. 9 in Southeastern Missouri is being used in further tests of the feasibility of using ERTS type data to identify major crop species. SRS has supplied ground truth information for 45 segments in this eight county area. Each segment contains two to 20 fields. Primary crops in this area are cotton and soybeans, with some wheat, hay, and corn.

It was found that it was nearly impossible to locate the segments and fields in the data on either gray scale maps of individual wavelength bands or non-supervised classifications (clustering) of the area. This was attributed to two factors: the lack of contrast between fields in the images and the small size of the segments and fields. Most segments are only 1/4 to 3/4 of a square mile and the majority of fields are smaller than 20 acres. Therefore, the ERTS data was geometrically corrected (procedures described elsewhere in this report) and scaled to a 1:24,000 scale. Since the segments were drawn to scale on 1:24,000 scale maps it was then a simple matter to overlay the resulting ERTS images onto the maps and locate the segments. Fields to be used for training and testing the classifier could then be accurately located within the segments by clustering the segment. Data from three dates, August 26, September 13, and October have been registered and multitemporal analyses will be performed.

Results

Classifications of this data set have not been completed at this time.

Plans for Next Bimonthly Period

We will complete the analysis of the data set described above. Mr. William Wigton, SRS/USDA, will be at LARS during this period to assist with the analysis and evaluate the performance.

III. Mapping of Soil Associations

Introduction

During this 6-month period ERTS MSS data were analyzed from two counties in Indiana--Benton County and Tippecanoe County. Most of the analysis was done on data collected on September 30, 1972 (Frame 1069-15585), and on November 24, 1972 (Frame 1124-16052). A lesser amount of analysis was done on data collected May 4, 1973 (Frame 1285-15595). This May data was, however, of superior quality for soils studies because much of the area had been plowed in preparation for planting corn.

The work reported herein involved comparison of general soil maps and single wavelength bands of ERTS MSS imagery photographically reproduced. The studies resulted in significant new conclusions relative to the value of ERTS images for obtaining information on soil and landscape characteristics, including usefulness as base images for producing conventional generalized soil maps.

Procedures

A new technology development was made during this period. The technique for interfacing existing conventional soil maps with ERTS imagery was developed. Generalized soil survey maps (soil association maps) based on more detailed ground surveys of varying vintage have recently been published (November 1971) for each county in Indiana. These general soil maps were compiled by the Agricultural Experiment Station and the Cooperative Extension Service, Purdue University and the Indiana Soil Conservation Service, USDA, using base maps provided by the Indiana Geological Society. These maps are at a scale of 1:190,080 and are intended for general planning purposes. Each delineation contains several soil types as stated in the legend, and, in addition, is known to contain soil types not given in the legend, but only in minor acreages.

To permit more precise and objective comparison of ERTS imagery with soil maps, a photographic overlay capability was developed. Images from individual MSS bands (5 and 7) were photographically enlarged from 70mm negatives directly from NASA (Goddard Space Flight Center) to a scale of approximately 1:190,080.

With these two products, a soil map and an ERTS image, at the same scale, and both relatively free of geometric distortions, direct comparisons were now possible. Transparencies of the soil maps were produced, and overlaid directly on the enlarged ERTS imagery, using landmarks which are independent of soil boundaries to insure proper registration or overlay of the two. Rivers and lakes were especially useful for overlaying, since they appear very dark on MSS band 7 imagery and give good contrast. An important aspect of using this type of overlay procedure for evaluation of imagery is that ability to delineate soil characteristics on imagery is determined after properly registered overlay is established using landmarks which are independent of the spectral characteristics of the soils. This breakthrough in overlay technique resulted in several major conclusions regarding usefulness of ERTS images as a source of significant new information on soil and landscape characteristics of interest to soil surveyors and land use planners. Prior to the development of this technique it was impossible to accomplish the following two tasks.

1. Check all soil boundaries as conventionally mapped to determine whether they agreed or disagreed with boundaries detectable from ERTS imagery.

2. Determine whether mapped soil boundaries on soil association maps were mislocated as a result of transfers from more detailed soil maps made on a photo base.

Obviously, a third task to be undertaken is that of determining whether soil boundaries on conventional maps were mislocated as a result of errors in the original field mapping. This type of error can occur for a variety of reasons including use of black and white base photographs with poor contrast in the original mapping, and use of black and white base photographs on which soil and landscape patterns were obscured by vegetation. Regardless of the reasons why present day conventional soil maps have some soil boundaries mislocated, it is logical to utilize ERTS images to assist in improving the quality of generalized soil maps because the ERTS images give an overall view of the soil landscape.

Results

Thirteen major soil associations are mapped on the conventional general soil map. These associations are shown in Figure III-1. In trying to determine how many and which of these associations can be separated on ERTS imagery we were immediately confronted with a problem in definition. This problem can be understood by an examination of Figure III-2. Some of the soil association boundaries in Figure III-1 have been transferred to the May 4 imagery of MSS band 7. In the upper left hand corner of Figure III-2 a boundary is noted between soil association 73 (Raub-Ragsdale) and soil associations 81 and 66 (Miami-Russell-Fincastle and Fincastle-Ragsdale-Brookston) but soil associations 81 and 66 are not distinguishable from one another. However, just below the boundary and to the right, an area of soil

Cooperative Extension Work in Agriculture and Home Economics, State of Indiana, Purdue University and U. S. Department of Agriculture Cooperating. H. G. Diesslin, Director, Lafayette, Indiana. Issued in furtherance of the Acts of May 8 and June 30, 1914.

AY-50-79

General Soil Map of TIPPECANOE COUNTY

PURDUE UNIVERSITY
Cooperative Extension Service
Lafayette, Indiana

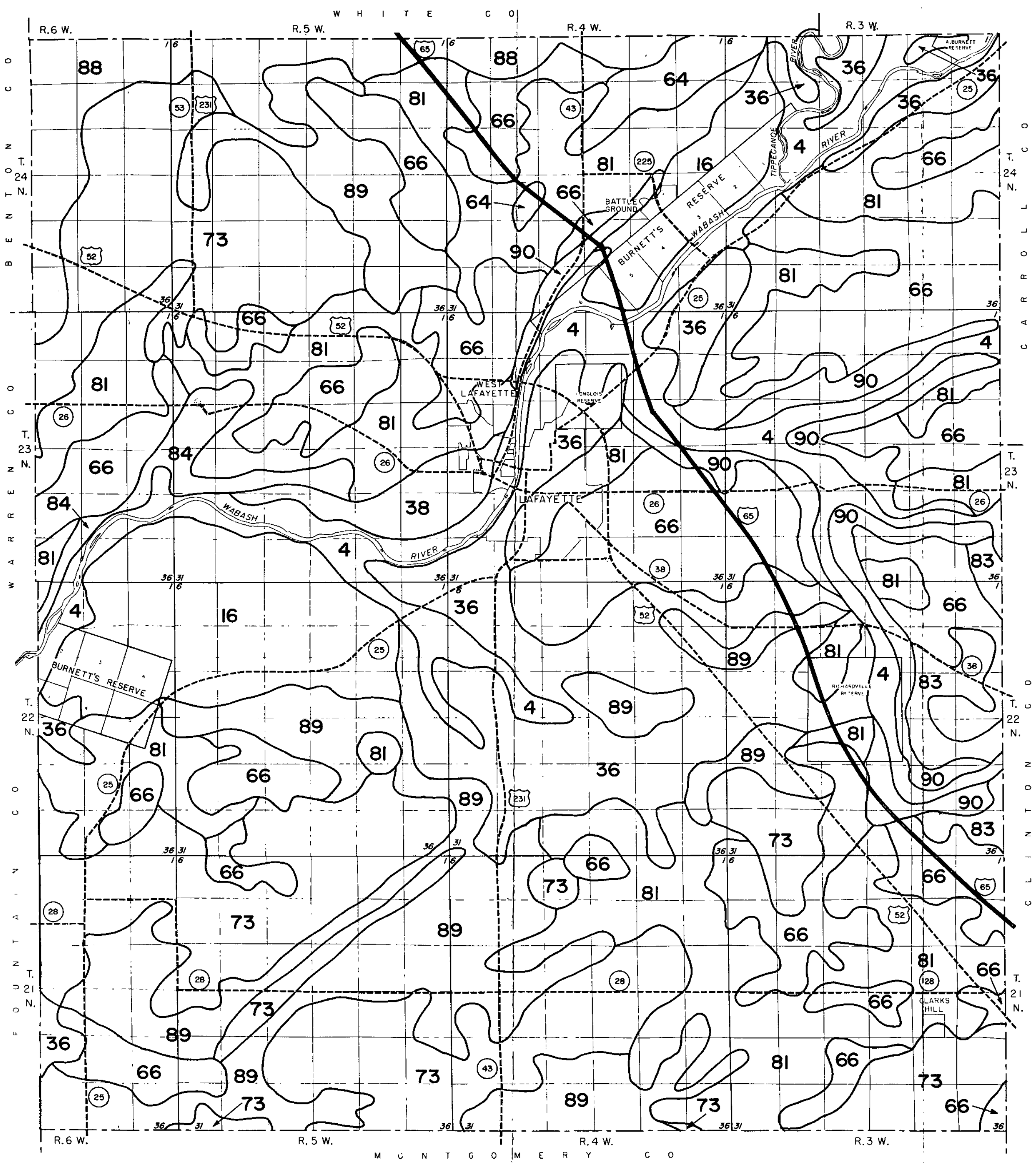


Figure III-1. General Soil Map of Tippecanoe County

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.

N

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.

November, 1971

Figure III-2. May 4, 1973 ERTS-1 MSS Imagery Band 7 (.8-1.1μm) showing location of Tippecanoe County and selected soil boundaries.

association 66 surrounded by an area of soil association 81 is apparent. Near the upper right part of Figure III-2 an area of 66 surrounded by 81 and 90 (Hennepin-Rodman) is readily apparent on the imagery. Therefore, whether or not boundaries are detectable on ERTS imagery is dependent on which two or more soil associations are being considered. The results of this study, therefore, are stated in terms of delineations or separations of soils which could generally be made from the ERTS imagery studied.

In Tippecanoe County four groupings of soil associations could be separated on the September 30, 1972 ERTS imagery (no figure shown). These were:

1. Soil associations 73, 88, and 89 (Raub-Ragsdale, Odell-Chalmers, and Sidell-Parr) from soil association 81 (Miami-Russell-Fincastle).

2. Soil associations 16, 36, and 38 (Elston-Wea, Ockley-Westland, and Ockley-Fox) from soil associations 81 and 89 (Miami-Russell-Fincastle and Sidell-Parr).

3. Soil association 4 (Genesee-Shoals-Eel) from adjacent soil associations except for soil association 90 (Hennepin-Rodman).

These separations relied on both MSS bands 5 and 7 imagery, with band 7 containing more information than band 5. At the time these data were collected, the county was almost completely vegetated, primary crops being corn and soybeans. Soil associations 4, 84, and 90 are often covered with trees.

May 4 imagery over Tippecanoe County permitted the same separations as the September imagery. Additionally, some of the separations showed more contrast in May, and some separations were possible within groupings of soil associations not distinguishable on September imagery. Examples of this are distinguishing soil association 89 from soil associations 73 and 66 (Figure III-2, upper left), and distinguishing association 66 from association 81 (Figure III-2, upper right) in band 7.

Additional soils information, of potentially more value than the above, was noted. In the center of the lower right quadrant of Tippecanoe County, a large two-part elongated dark area was observed in band 7. This area is outlined on band 5 imagery in Figure III-3, and labeled as areas A. It appears light in this band. The area cuts across boundaries between soil associations 66, 36, 89, and 73. Area A, which is apparent on the ERTS imagery, is a large, relatively flat area between major drainageways and many of the soils of that area are similar. This large area of rather similar soils is not indicated on the general soils map in the same manner, because of the procedures which are used in making the general soil map from more detailed field soil maps. This ERTS imagery would provide additional useful information for soil surveyors when producing a general soil map of Tippecanoe County from existing maps and aerial photography.

The study in Benton County, Indiana was done using ERTS data collected November 24, 1972. At this time most of the county was covered with vegetation, primarily corn and soybeans, which in many cases had been harvested. The actual proportion of harvested to unharvested crops is not known, but harvesting had been delayed and very little plowing had been done due to unusually wet fall weather.

The primary soils information provided by the ERTS imagery band 5 (Figure III-4) was with respect to the soil association 70 (Parr-Corwin) and soil association 88 (Odell-Chalmers). The general soil maps and legend are not shown,

Figure III-3. May 4, 1973 ERTS-1 Imagery of MSS Band 5 (.6-.7μm) showing location of Tippecanoe County and selected soil boundaries.

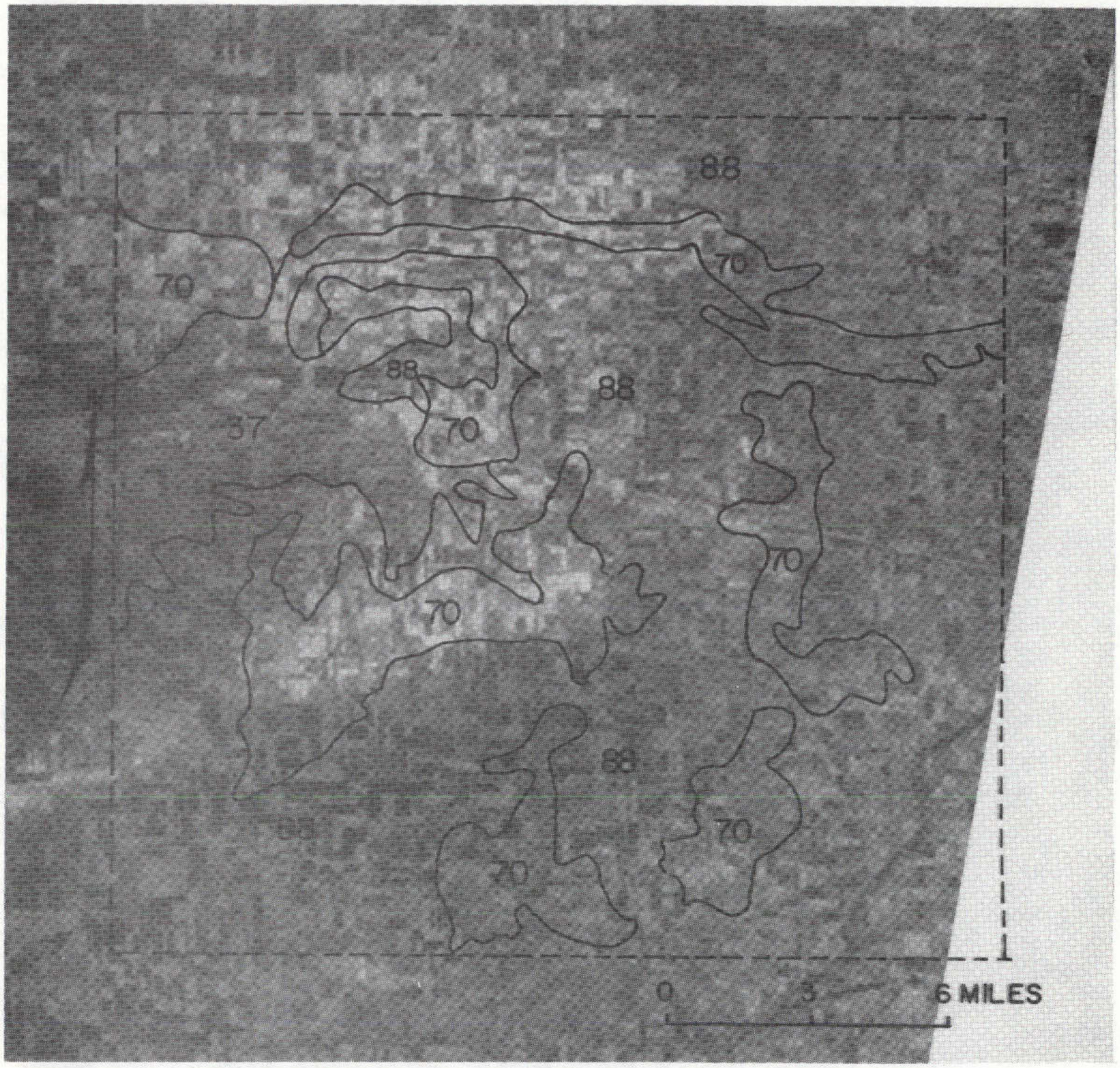


Figure III-4. November 24, 1972 ERTS-1 Imagery of MSS Band 5 (.6-.7μm) showing location of Benton County and selected soil boundaries.

but some of the soil boundaries as conventionally mapped are drawn on the imagery. Benton County is mapped primarily into these two soil associations, with about 10% of the county mapped as soil association 37 (Ockley-Wea). The Par-Corwin association consists of sloping, well-drained, loamy Parr and moderately well-drained, loamy Corwin in glacial till.

In Figure III-4 some of the boundaries between soil associations 70 and 88 are apparent on the ERTS imagery. In the upper left part of Figure III-4 a semicircular area mapped soil association 88 lies between an area mapped soil association 37 and a semicircular area mapped soil association 70. Whereas association 88 usually appears dark on the imagery, as would be expected, this area appears light. Examination of aerial photography shows a good amount of light colored soil in this area, as is typical of the Parr-Corwin association (70). The soils become predominantly dark in the Ockley-Wea (37) area, which is also indicated on the ERTS imagery. The two areas of association 70 in the lower part of the figure do not appear as light on the ERTS imagery as the association 70 areas in the upper left part. Examination of the aerial photography again revealed that these areas contained lesser amounts of light colored soils, indicating that the ERTS imagery represented the actual conditions reasonably well. Similar verification of the new soils information provided by the ERTS imagery were noted in other areas of the county.

Conclusions

A technique was developed for making precise comparisons of ERTS imagery and general soil maps for the first time. This breakthrough greatly advanced the evaluation of ERTS imagery as a potential source of significant new information on soils. It was concluded that ERTS imagery is a useful source of significant new information on soils and landscape characteristics. It was possible to locate boundaries between four or more groups of soil associations in Tippecanoe County, Indiana both using September and May ERTS data. The May data contained more information on soil characteristics and patterns, presumably because much of the county had been plowed. In Benton County, Indiana two soil associations were apparent on the ERTS imagery. When boundaries observed on the imagery were quite different from those mapped on the conventional generalized soil map, examination of aerial photography indicated a strong correlation between ERTS imagery and surface soil characteristics. This gave evidence that the use of ERTS imagery as a photobase during generalized soil mapping would be very helpful in location of boundaries.

The ERTS imagery was also valuable in checking both the accuracy and precision of existing conventional generalized soil maps (soil association maps) for Benton and Tippecanoe Counties. While field checks have not yet been made, there is adequate evidence to believe that ERTS images, because they give an overall view of the soil landscape, are a useful new tool to assist in making general soil maps.

IV. WATER RESOURCES RESEARCH

Introduction

During this past six-month reporting period, water resources research on the Wabash Valley Project have involved the analysis of both ERTS data and aircraft multispectral underflight data. Segments of several frames of ERTS data have been analyzed to determine the ease in identifying surface water features and the spectral characteristics of these water bodies. Most of this work has concentrated on the area around Lafayette, Indiana and to the North of Lafayette, and it has involved the Wabash and Tippecanoe Rivers and Lake Shafer and Freeman.

Procedures

One University of Michigan ERIM aircraft scanner underflight mission was flown on May 4, 1973. This mission included flight lines over the Cayuga Power Plant test site; Lake Freeman; the Wabash-Tippecanoe River junction; and the strip mines in Sullivan, Greene, Knox, and Daviess Counties. In addition, surface observations of water quality (suspected solids and water temperatures) were made at approximately the same time the Michigan aircraft and ERTS passed overhead. One Exotech field spectroradiometer experiment was conducted on July 11, 1973. In this experiment, 28 water spectra were recorded by the instrument at the Tippecanoe-Wabash River junction using a modified Secchi disk. These data were used to study the influence of water depth and turbidity on the spectral characteristics of water bodies.

Results

Analysis results have shown that there is little difference in spectral characteristics between Lake Shafer and Lake Freeman; however, there are some differences within each of the reservoir areas. These spectral differences show up primarily as a change in spectral response from the center to the edges of the reservoirs, and are thought to involve variation in turbidity and depth. A distinct difference in spectral response has been shown to exist between the Wabash River and the Tippecanoe River, largely because the Wabash River has a higher turbidity level than the Tippecanoe River. Both river systems are spectrally different from the reservoirs, but the exact cause for this difference has not been ascertained to date.

Analysis of aircraft underflight data collected by the ERIM aircraft in August of 1972 has shown similar variations in spectral response. However, the aircraft data has shown a much greater range of spectral differences in the water bodies (i.e. a larger number of spectral classes are present). Some rather severe sun angle problems were encountered in the analysis of the aircraft scanner data which were not found to be present in the ERTS data. These results tend to confirm our impression that many geometric illumination problems which are present in multispectral scanner data from aircraft altitudes are not encountered in ERTS data because of the smaller field of view and much greater altitude from which the data is being collected.

Analysis of aircraft data collected in Central Indiana over a series of strip mines has shown that at least 8 spectral classes can be readily defined. The correlation between the spectral response identified in the scanner data and the pH of the water and accompanying differences in algal concentrations are being field checked.

The ability to accurately survey area of surface water bodies with ERTS data appears to offer mixed results thus far. Only preliminary efforts have been carried out in this phase of the study, but we are finding that in some cases very small water bodies are shown clearly on the ERTS imagery and have been enhanced so that they appear to be much larger than the water body actually is, and in other cases small water bodies do not show at all. This seems to be a problem of spatial resolution and variations in spectral response between the water and the surrounding objects.

A final phase of the water resources effort during the past six-month reporting period has involved the collection of spectral data with the Exotech field spectrometer on selected water bodies during the past two months. Variations in spectral response due to specular reflectance have required development of a modified method for collecting such field spectral data. Tests are currently underway to determine the best method for collecting accurate, reliable spectral information over a natural water surface using such Exotech field spectrometer techniques. The results of such field spectrometer data collection should allow much more accurate interpretation of the capabilities and limitations of defining spectral differences in water bodies using both aircraft and ERTS data collection sources.

Preliminary results with a layered classifier approach to the mapping of water body characteristics have produced very good results with aircraft data. Further work with the layered classifier for first identifying water bodies and secondly defining the spectral characteristics within the water area will be conducted during the next six-month reporting period.

Conclusions

Results thus far tend to indicate that large water bodies can be mapped with a reasonable degree of accuracy but that small water bodies produce rather questionable results on an individual basis although in a large geographic area the surface water area designated may still be reasonably accurately tabulated with the ERTS data.

Plans for Next Reporting Period

During the next reporting period the correlation between the aircraft data and the spectral response defined on the ERTS data will be examined. During the next two reporting periods the question of accuracy of identifying small surface water bodies will be investigated in more depth.

V. Earth Surface Features Identification

Introduction

A spacial data bank has been developed for a 140 square kilometer (63 square miles) area of Tippecanoe County, Indiana. Although complexities arose, a series of forty-seven natural and cultural resource variables were ultimately formatted such as to be stored on a UTM (universe transverse mercador) grid system. The objective of utilizing such a system of storing data is many-fold. The intent here is to utilize the data as a means of optimizing the land use decision making process as well as comparing the extent of accuracy to which certain critical natural resource variables can be identified from ERTS-1 imagery.

Procedures

Two areas of investigation are being studied to achieve the ultimate goal of attempting to establish a technique whereby ERTS could be analyzed so as to determine its usefullness in extracting resource information critical to the optimization of landuse decisions.

In the first area of study a significant amount of effort was devoted to the classification process of identifying numerous natural and cultural resources from ERTS data.

A detailed land-use classification consisting of 29 spectral classes of an area in Tippecanoe County, Indiana was completed. This classification indicated that a number of urban features can be differentiated including commercial development (shopping centers) two distinct classes of new residential area, and one class of older residential area. This data was collected on September 30, 1972. Some misclassification occurred between old residential areas and row crops.

It was also determined that forest cover can be reliably differentiated from other cover types. There were three different categories of forest cover differentiated (these classes are described in more detail in a latter section of this report) -- a factor that led to the utilization of this resource variable in the second area of this investigation. The other major land-use categories which have been defined included the following: three classes of water (lake and two river classes), pasture/grass and row crops.

A temporal overlay for the test area became available late in the period and a test analysis was performed with this data. Data from September 30, October 19, and November 6 were overlayed and a test analysis to map forest cover was performed using this data set. Very little difference was found between the forest cover map from the overlayed data and that from the map using September 30 data only. An overlay of September and June data will be used in the next reporting period in an attempt to map all land use categories more accurately and to provide more discrete differentiation between such.

In the area of forest cover mapping, attention is concentrated on defining forest cover versus all other types. Differentiation between coniferous and deciduous cover type appears feasible, and a number of different spectral categories of deciduous forest cover can be defined. The intent of this area of investigation for the next reporting period will be to not only assist in the development of more sophisticated classification sets for identifying critical natural and cultural resources from ERTS data but also to expand the current training classes to a larger geographic area in the continued effort to develop a forest

C

cover map for the state of Indiana and to define the accuracy of forest cover mapping in this central U.S. area.

To accomplish Area Two of this investigation, it was determined to focus initial attempts on one specific resource variable -- that of forest cover as classified on ERTS data September 30, 1972. Information was first gathered and stored in a spacial data bank so as to be utilized for ground truth purposes. (This investigation is not to be involved with the constructing of the data bank so its developmental process and problems will not be discussed here).

Since the primary goal of this area of study was that of establishing a correlation of ERTS data to that of an existing spacial data system ERTS imagery was selected, classified, and then reformated so as to be compatible with the existing data analysis system. Of interest here were the forest cover groups assembled as Class T, E, and Y. These classes were manually extracted onto a base grid in an attempt to make them spacially compatible with the existing analysis system. The codes for extraction as well as the class groups are listed below:

Extraction Code:	ERTS Data Symbol:	Classification Groupings:
0	(symbolizes the lack of the three class groupings symbol)	
1	T	Trees 1 Green Agriculture Trees 5 Trees 6
2	E	Trees 3
3	Y	Trees 4 Dense forest

The existing data bank has nine levels of forest classification (based on percent density of specific forest cover) which for purposes of this initial investigation was too detailed. These nine sub-groups were thus agglomerated into three groups as follows:

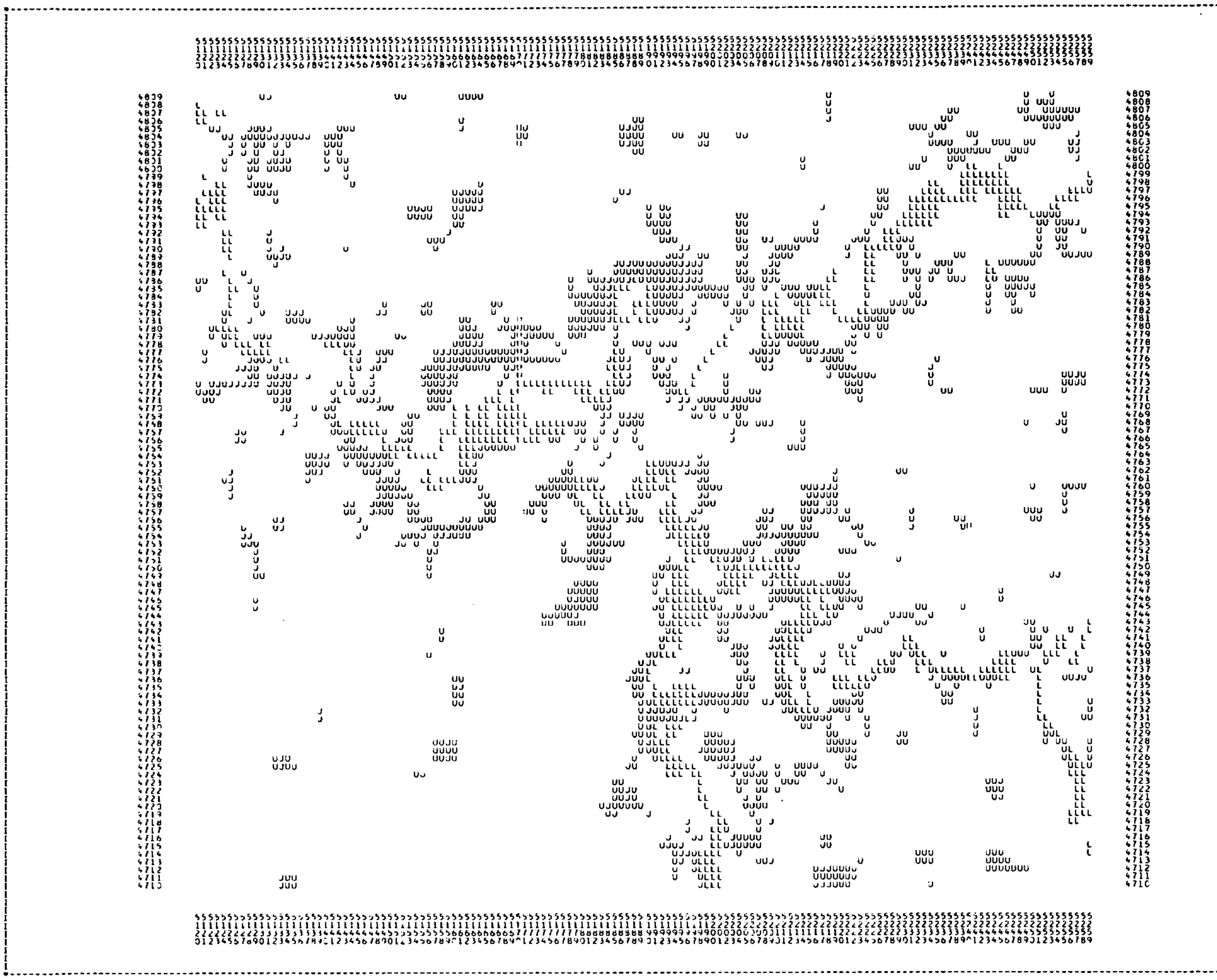
Original Extraction Code	Agglomerated Extraction Code	Map Symbol and Meaning
0	0	Blank (none)
1	0	Blank (none)
2	1	
3	1	L (lowland)
4	1	
5	1	
6	2	
7	2	U (upland forest)
8	2	
9	2	

Figure V-1 illustrates the distribution of forest cover, extracted from low altitude data, in the study area.

The data extracted from ERTS-1 imagery was then spacially related to the data bank. Through basic logic statements and utilizing the manipulative and analytical capabilities of the data bank, preliminary comparative investigations were made studying the accuracy levels of forest cover identification vs. ERTS-1 data.

Area Two Results

Although it is emphasized that these results are represented by a cursory investigation a number of facts became readily apparent. The intent is to ultimately evaluate the potential use of ERTS data in the land use decision making process. In order for this end to be accomplished, the accuracy of data identification was pursued. Figure V-2 is an initial three level classification of ERTS-1 data. After a first level comparison of this



FOREST COVER - LOW ALTITUDE - OCT. 1971

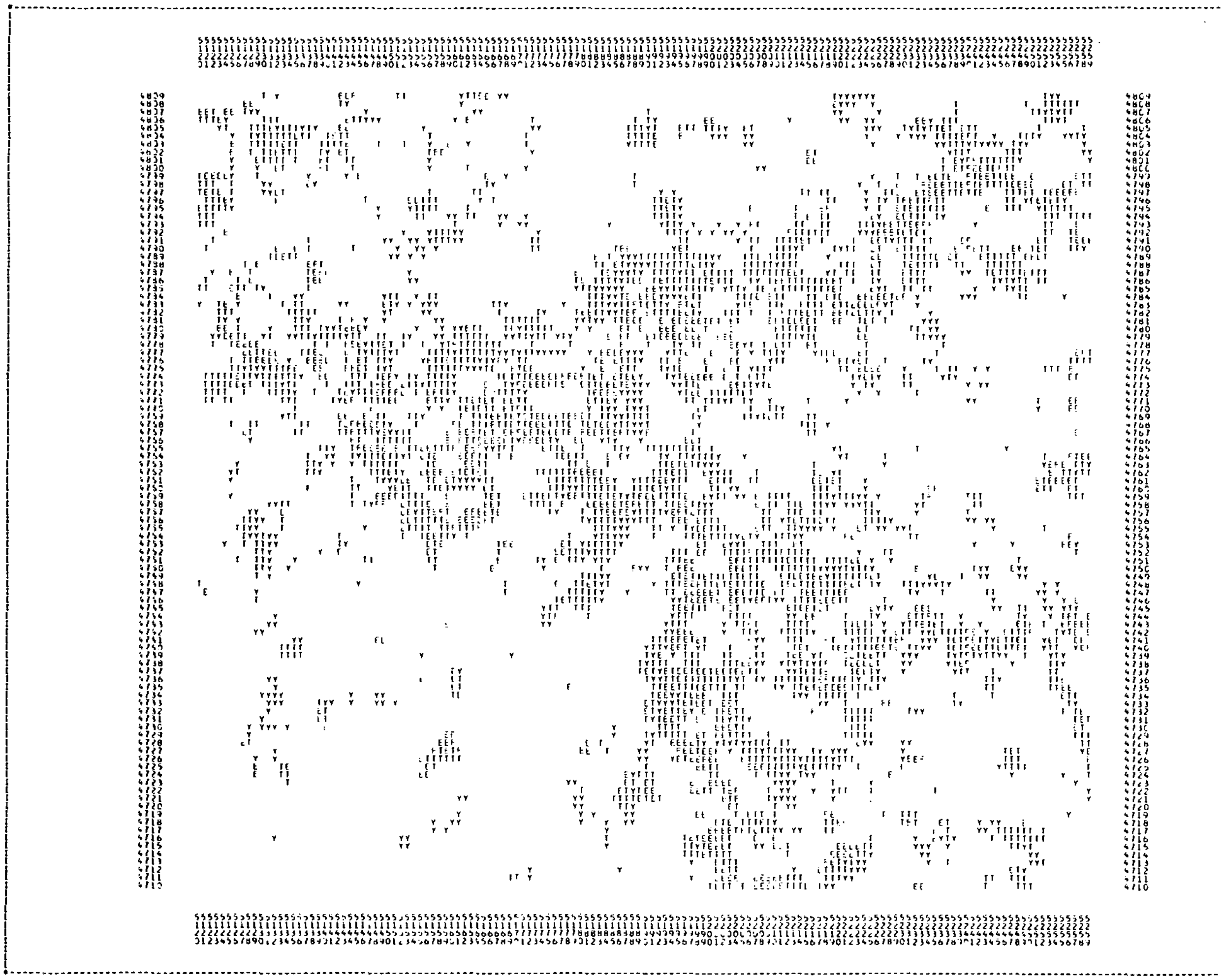
LABORATORY FOR APPLICATIONS OF REMOTE SENSING - PURDUE UNIVERSITY

EAST-CENTRAL SECTION TIPPECANOE COUNTY - INDIANA

L = LOWLAND FOREST COVER
U = UPLAND FOREST COVER

LEVELS	0	1	2	3	4	5	6	7	8	9
SYMBOLS						LLLLLLLLL LLLLLLLLL LLLL LLLL LLLLLLLLL LLLLLLLLL			UUUUUUUUU UUUUUUUUU UUUU UUUU UUUUUUUUU UUUUUUUUU	
FREQUENCY	11282	0	0	0	0	776	0	0	1342	0

Figure V-1. Computer printout of forest cover distribution in the study area extracted from low altitude aerial photography.



FOREST COVER - ERTS DATA - SEPT 1972

LABORATORY FOR APPLICATIONS OF REMOTE SENSING - PURDUE UNIVERSITY

EAST-CENTRAL SECTION TIPPECANOE COUNTY - INDIANA

T = TREES 1, TREES 5, TREES 6, GREEN AGRICULTURE
E = TREES 3
Y = TREES 4, DENSE FOREST

LEVELS	0	1	2	3	4	5	6	7	8	9
SYMBOLS				T		E			Y	
FREQUENCY	9797	0	0	2154	0	1115	0	0	934	0

Figure V-2. Three level computer classification of forest types in study area using ERTS-1 data.

data against that of the low altitude data (Figure V-1) was made, it could be seen that the number of acres identified as forest on ERTS was more than that on the low altitude data. Further investigation showed that although more cells indicating forest were present on ERTS data the number of cells common to both low altitude and ERTS data (combination of blank and common forest) was 10433 or 75% (see Figure V-4).

Further comparison of the frequency histograms in Figure V-3 began to illustrate that problems may exist in accurately classifying distinct forest cover from ERTS on a single image analysis. Review of the comparative studies between the low altitude data and each separate classification of ERTS data showed that Class T (Figure V-3 - Chart A) had a total of 1221 cells on the ERTS data that were not present on the low altitude data. This could in fact be due to the broad range which this classification encompasses. More accurate relationships could be established if through layered analysis of ERTS more distinct classes of forests could be generated thus minimizing the inclusion of extemporaneous classes (this study, as was indicated, will be evaluated and analyzed in Area one of this investigation).

Area Two Conclusions and Plans

This first level analysis of the utilization of ERTS-1 imagery for the identification of earth resources having potential usage in the decision making process leads to the belief that statistically the correlation accuracy of quantifying resource variables over large areas from ERTS data is high. It also appears that the interpretation of ERTS-1 data leads to a higher number of acres identified as forest. This appears to be due in part because of inaccuracies in classification, but must also be attributed to the resolution capabilities of ERTS imagery.

```
COMPARISON OF VARIABLES 40 AND 47

VARIABLE 40 IS FOREST COVER EXTRACTED FROM AERIAL PHOTOGRAPHY

VARIABLE 47 IS FOREST COVER EXTRACTED FROM ERTS CLASSIFICATION

                                       LEGEND
                                   = NO EXTRACTED FOREST COVER INFORMATION EXISTED FOR THIS CELL
                                 O = IN THE OLD (AERIAL PHOTO) EXTRACTION, FOREST COVER WAS INDICATED FOR
                                     THIS CELL
                                 N = IN THE NEW (ERTS CLASSIFICATION) EXTRACTION, FOREST COVER WAS
                                     INDICATED FOR THIS CELL
                                 B = IN BOTH EXTRACTIONS, FOREST COVER WAS INDICATED FOR THIS CELL

                                 OLD VS. ERTS 1

LEVELS          0        1        2        3        4        5        6        7        8        9
       ***************************************************************************************************
                     OOOOOOOOO NNNNNNNNN BBBBBBBBB
                     OOOOOOOOO NNNNNNNNN BBBBBBBBB
SYMBOLS              OOOO OOOO NNNN NNNN BBBB BBBB
                     OOOOOOOOO NNNNNNNNN BBBBBBBBB
                     OOOOOOOOO NNNNNNNNN BBBBBBBBB
       ***************************************************************************************************
FREQUENCY  10061     1785     1221      933        0        0        0        0        0        0
```

A.

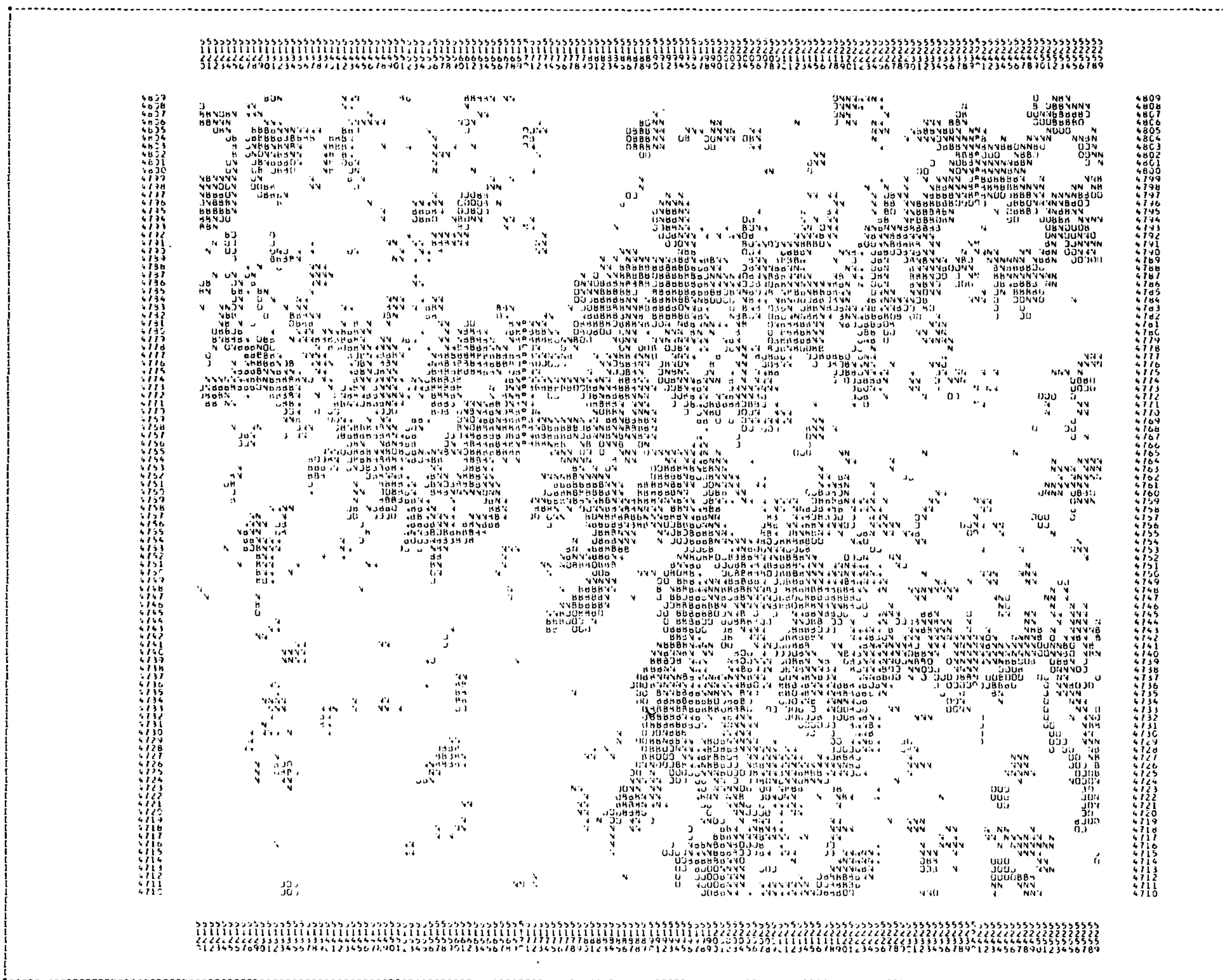
```
COMPARISON OF VARIABLES 40 AND 47

VARIABLE 40 IS FOREST COVER EXTRACTED FROM AERIAL PHOTOGRAPHY

VARIABLE 47 IS FOREST COVER EXTRACTED FROM ERTS CLASSIFICATION

                                       LEGEND
                                   = NO EXTRACTED FOREST COVER INFORMATION EXISTED FOR THIS CELL
                                 O = IN THE OLD (AERIAL PHOTO) EXTRACTION, FOREST COVER WAS INDICATED FOR
                                     THIS CELL
                                 N = IN THE NEW (ERTS CLASSIFICATION) EXTRACTION, FOREST COVER WAS
                                     INDICATED FOR THIS CELL
                                 B = IN BOTH EXTRACTIONS, FOREST COVER WAS INDICATED FOR THIS CELL

                                 OLD VS. ERTS 2

LEVELS          0        1        2        3        4        5        6        7        8        9
       ***************************************************************************************************
                     OOOOOOOOO NNNNNNNNN BBBBBBBBB
                     OOOOOOOOO NNNNNNNNN BBBBBBBBB
SYMBOLS              OOOO OOOO NNNN NNNN BBBB BBBB
                     OOOOOOOOO NNNNNNNNN BBBBBBBBB
                     OOOOOOOOO NNNNNNNNN BBBBBBBBB
       ***************************************************************************************************
FREQUENCY  10038     2247      644      471        0        0        0        0        0        0
```

B.

```
COMPARISON OF VARIABLES 40 AND 47

VARIABLE 40 IS FOREST COVER EXTRACTED FROM AERIAL PHOTOGRAPHY

VARIABLE 47 IS FOREST COVER EXTRACTED FROM ERTS CLASSIFICATION

                                       LEGEND
                                   = NO EXTRACTED FOREST COVER INFORMATION EXISTED FOR THIS CELL
                                 O = IN THE OLD (AERIAL PHOTO) EXTRACTION, FOREST COVER WAS INDICATED FOR
                                     THIS CELL
                                 N = IN THE NEW (ERTS CLASSIFICATION) EXTRACTION, FOREST COVER WAS
                                     INDICATED FOR THIS CELL
                                 B = IN BOTH EXTRACTIONS, FOREST COVER WAS INDICATED FOR THIS CELL

                                 OLD VS. ERTS 3

LEVELS          0        1        2        3        4        5        6        7        8        9
       ***************************************************************************************************
                     OOOOOOOOO NNNNNNNNN BBBBBBBBB
                     OOOOOOOOO NNNNNNNNN BBBBBBBBB
SYMBOLS              OOOO OOOO NNNN NNNN BBBB BBBB
                     OOOOOOOOO NNNNNNNNN BBBBBBBBB
                     OOOOOOOOO NNNNNNNNN BBBBBBBBB
       ***************************************************************************************************
FREQUENCY  10626     2440      656      278        0        0        0        0        0        0
```

C.

Figure V-3. Frequency histograms of cells in each cover type resulting from aerial photography and ERTS classification analysis.



COMPARISON OF VARIABLES 40 AND 47

VARIABLE 40 IS FOREST COVER EXTRACTED FROM AERIAL PHOTOGRAPHY

VARIABLE 47 IS FOREST COVER EXTRACTED FROM ERTS CLASSIFICATION

LEGEND

= NO EXTRACTED FOREST COVER INFORMATION EXISTED FOR THIS CELL
O = IN THE OLD (AERIAL PHOTO) EXTRACTION, FOREST COVER WAS INDICATED FOR THIS CELL
N = IN THE NEW (ERTS CLASSIFICATION) EXTRACTION, FOREST COVER WAS INDICATED FOR THIS CELL
B = IN BOTH EXTRACTIONS, FOREST COVER WAS INDICATED FOR THIS CELL

LEVELS	0	1	2	3	4	5	6	7	8	9
SYMBOLS		OOOOOOOOO	NNNNNNNNN	BBBBBBBBB						
FREQUENCY	8761	1036	2521	1662	0	0	0	0	0	0

Figure V-4. Comparison of cells classified as forest from aerial photography and ERTS-1 data.

Future plans in this area are going to involve the reclassification of ERTS data utilizing imagery acquired at different times so as to define more discretely the various classes of vegetation cover that are not agglomerated into one class.

It was also concluded that in order for ERTS data to be utilized for detailed land use inventory and analysis, more accuracy must be achieved in spacial location of resource variables -- the results of which appear to be achievable through more accurately keying the boundary coordinate points of the study area.

At present a technique for automatic data entry from classified ERTS-1 imagery is being investigated. If this process can be accomplished and spacial accuracy achieved along with accurate ERTS data classification, it appears that a semi-automatic system of data bank development for utilization in land use planning can result.

VI. URBAN LAND USE ANALYSIS

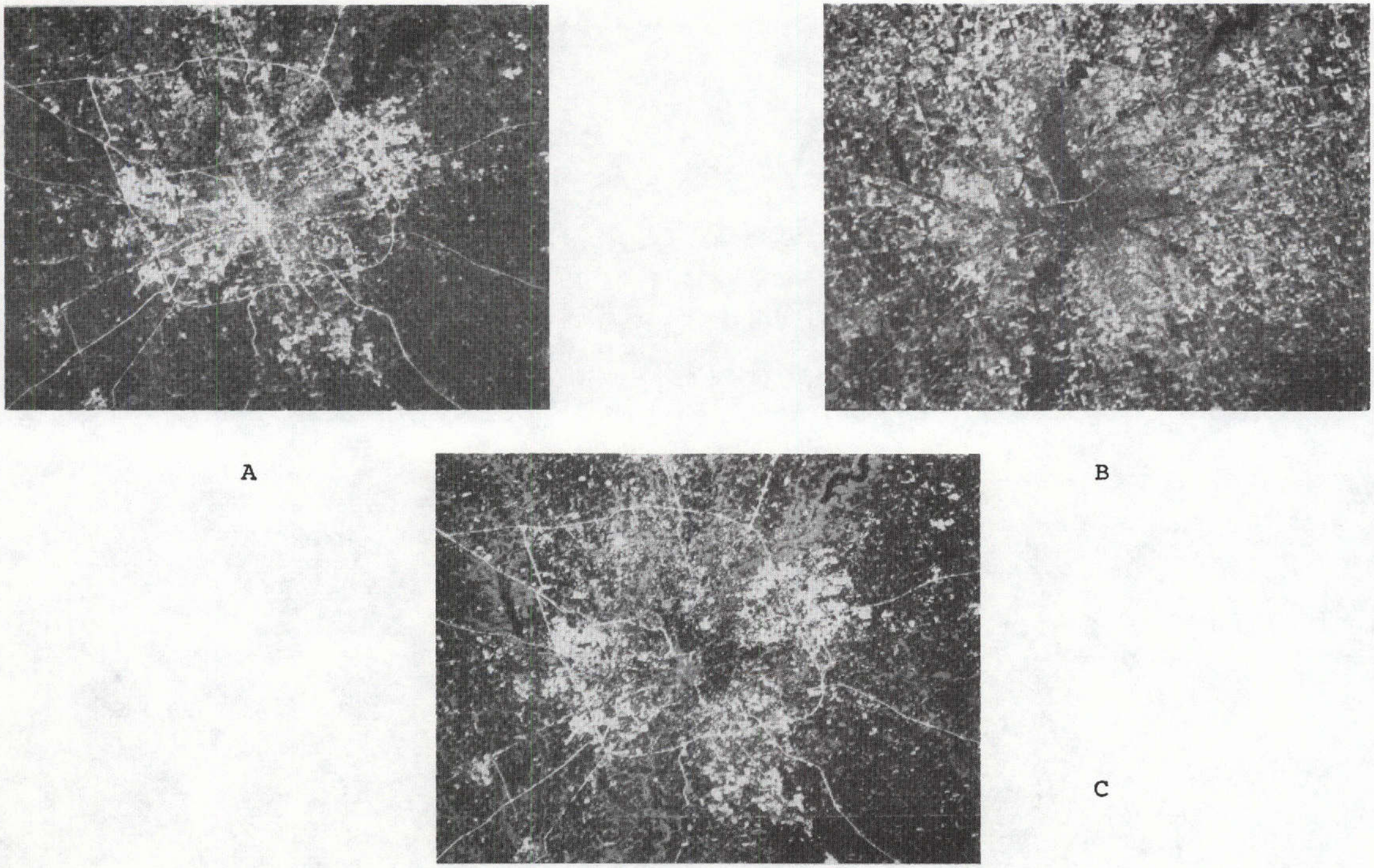
Introduction

City, county, and regional planning officials have a definite need for rapid, accurate land use information. Inventories of urban land use are required in order to make decisions concerning the best allocation of an area's resources. Marion County (Indianapolis), Indiana was selected as the area to test the applicability of ERTS data in collecting land use information.

Four bands of digitized multispectral data collected on September 30, 1972 (Scene ID 106915585) were analyzed by computer processing to obtain a land use classification of the county.

Procedures

Initially, the data were viewed on the LARS digital display system for purposes of orientation. Figure VI-1 shows two of the four spectral bands; a is from the visible portion of the spectrum (Band 4, 0.5-0.6 μm) and b is from the reflectance infrared (Band 6, 0.7-0.8 μm). At that time, several "pre-analysis" activities were accomplished. Several small areas were chosen for input into a clustering algorithm. Moreover, several small areas were chosen for histogramming to allow more controlled viewing of the digital display in the future. Finally, the study area (Marion County, Indiana) was located, along with a number of important land marks (Central Business District, interstate highways, larger water bodies, airports, rivers, etc.), to ease subsequent analysis of maps from the line printer with alpha-numeric symbols.



A

B

C

Figure VI-1. Photos from digital display. Figure VI-1-A is a graytone image from the visible portion of the spectrum (Band 4, 0.5-0.6μm), B is from the reflective infrared (Band 6, 0.7-0.8μm). Figure VI-1-C is a land use classification (see text for association of brightness levels with land use category).

The three areas chosen for clustering were submitted to the clustering algorithm, asking the program to find fourteen spectral classes, using all four bands of data. Statistics were then calculated for the clusters delimited, and all data points in the study area were classified into one of the fourteen cluster classes. The resulting "cluster map" of the area was then used as a base map, from which training samples (small, rectangular areas) could be chosen by hand for the desired land use classes. Statistics were calculated for these classes, and the area was reclassified.

Results

A. Explanation of Classification Image Figure VI-1c shows the classification results. The graylevels for the various classes are as follows:

Older, higher density housing	- black
Newer, lower density housing	- white
Industry/commerce	- medium gray
Woods	- light gray
Agricultural, open areas, parks	- dark gray
Water	- black
Clouds	- white
Cloud shadows	- black

Water, older housing, and cloud shadows have the same graylevel, but the cover types will not be confused in the image if their spatial distribution is considered. Water occurs in two reservoirs, Geist and Eagle Creek, located in the Northeastern corner and East-central parts of the county, respectively, and in several small ponds. There are, moreover, approximately ten small cumulus clouds, with their associated shadows (about one-half mile in diameter), which have a characteristic, easily-distinguishable areal distribution. Older housing is centrally located in the county, surrounding the central business district.

The above principle also applys to newer housing and clouds, both of which are white in the classification image.

B. Areal Distribution of Spectral Classes

The class "industrial/commercial" is characterized by the occurrence of rooftops and concrete, in varying proportions. The principal location is the Central Business District and surrounding industrial complexes, an area bounded roughly by 20th Street on the North, Morris Street on the South, West Street on the West, and College Avenue on the East. Other smaller areas are found where large industrial plants or shopping centers are located.

Older, higher-density housing surrounds the Central Business District, in a circular fashion. Most of the structures in this area were built prior to World War II and house two or more families. This area is roughly bounded by 56th Street on the North, Troy Avenue on the South, Tibbs Avenue on the West, and Arlington Avenue on the East.

The class which has been designated newer, lower-density housing consists of single-family dwellings, most of which were built after World War II. Lawns (grass) and roads (concrete) are the primary constituent of this spectral class. Not unusually, interstate highways also were classified as suburban housing. Three primary areas were classified as newer housing, each of which begins approximately five miles from the Central Business District:

Approximate Boundaries

AREA	NORTH	SOUTH	EAST	WEST
West	46th St	10th St	Tibbs Ave	I-465
East	62nd St	Washington St	Church Rd	Arlington Ave
South	Edgewood Rd	County Line Rd	McFarland Rd	Bluff Rd

Wooded areas are found in all parts of the county, but the two largest areas are located around the two reservoirs. The spectral class "agricultural, grassy, open" included cropland, pasture, open areas, golf courses, parks, and cemeteries. A number of small cumulus clouds (all classified as "cloud") were identified in the Northern part of the study area, along with their associated shadows (a separate spectral class). The distribution of water was discussed above.

C. Accuracy of Classification

Classification results were good to excellent for the following spectral classes: industrial/commercial, newer housing, water, cloud, cloud shadow, and trees. Agricultural, open areas were defined fairly well, but numerous, scattered data points in these areas were misclassified as older housing. Conversely, scattered data points of certain areas of older housing were misclassified as agricultural areas.

Two large residential areas comprised the greatest problem in the classification. One is on in North-central part of the county, bounded by County Line Road on the North, 56th Street on the South, Northwestern Avenue on the West, and I-465 on the East. The other is in the South-central part of the county, bordered by Troy Avenue on the North, Edgewood Road on the South, Bluff Road on the West, and Sherman Drive on the East. The Northern area consists of non-continuous, upper-income clusters of residences, most of which were built after 1940. Data points in these areas were usually classified as wooded or agricultural (grassy, open). The proportion of structures built prior to 1940 in the Southern area ranges from 25 to 50 percent. Incomes of its inhabitants are in the middle range of the distribution. Data points in this area were classified as either newer housing or agricultural (grassy, open).

Specific investigations were then made into the spectral characteristics of three types of land use -- older, higher density housing, agricultural (grassy, open), and a class herein termed "wooded residential", consisting of the aforementioned upper-income residential clusters in the Northern part of the county (for which no single spectral class was able to be developed). A number of training samples of each of these cover types were chosen off the digital display, with a 1971 color air-photo mosaic as a guide. Statistics were then calculated for the samples.

All three cover types had very similar ranges of spectral response in the two visible bands. In the infrared bands, the classes older housing and wooded residential were separable (the latter was brighter), both classes were comfused with agricultural (grassy, open) areas. Two parameters, however, showed that the samples could be separated spectrally. One, the agricultural (grassy, open) areas had standard deviations in bands 6 and 7, which were at least 15 percent and usually 40 percent higher than that of older housing and wooded residential samples. Two, the correlation between bands 6 and 7 (r_{67}) was almost always +0.95 or greater, while the r_{67} for the older housing and wooded residential classes was typically below +0.80.

Conclusions

A number of important land use classes were mapped in Marion County by application of a Gaussian maximum likelihood Classifier with four channels of ERTS-1 MSS data. By investigation of other parameters, moreover, such as standard deviations and correlation coefficients, a better classification is achieved. It is believed that the technique can be useful to urban/regional planners for attaining/monitoring urban land use data.

Plans for Next Bi-Monthly Period

Work in the next period will consist of preparing illustrations for the Marion County study, and readying a detailed manuscript for presentation October 16-18, 1973 at the LARS/Purdue University "Symposium on Machine Processing of Remotely Sensed Data".

Problems and Recommendations

It was a great disappointment to view January ERTS data of Marion County (Scene ID 117715593) because of bad data lines. Band 5 had one bad data line, Band 7 one, and Band 6 over twenty-five (apparently every sixth line). Consequently, no classification work has been attempted yet. Fortunately, however, the bad data lines in Band 6 appear to extend from Arlington Avenue (in the Eastern part of the county) Eastward. Preliminary (cluster method) classification work may be attempted. If successful, a recommended study would utilize rotation (for North-South orientation) and the "temporal overlay" of the September and January data. Changes in land use, as well as data comparability, would be the foci of such analysis.

VII. Analysis Technique Development

Data-Based Criteria for Defining Training Classes

The scale and resolution of the digital ERTS-1 MSS data have, as expected, led to multispectral data analysis techniques which differ significantly in detail (if not in concept) from the techniques developed for data collected by aircraft. Interdisciplinary efforts at LARS involving both data processing and applications scientists have evolved techniques applicable to a wide range of earth survey problems (many of which are evidenced in this report).

Two extreme situations may be identified which require somewhat different analysis approaches. In one case no ground observation data are available -- or at best ground observation data of a very general nature (such as might be obtained from interpretation of high altitude underflight photography). At the other extreme is the case in which detailed ground observation data are available (for instance, from field visitation). The analysis techniques for these cases differ primarily in the extent to which training class definition depends on spectral variability inherent in the multispectral data. Following are outlines of the analysis procedures applicable to the two extreme situations.

Case I: Limited ground observation data.

1. Apply cluster analysis to randomly selected areas and/or areas expected to contain cover types of interest.
2. Associate the clusters (= spectral classes) with general ground cover types. Use relative response in two or more channels to indicate cover type.
3. Classify the image using the spectral class definition obtained in the previous step.
4. Refine the spectral class association if possible based on any available information about the scene.

5. Perform qualitative evaluation of the results.

Case II: Detailed ground observation data

1. Unsupervised multichannel image enhancement
   a. Apply cluster analysis to areas containing ground observations.
   b. Use the spectral classes produced by the cluster analysis as a basis for classifying the scene. Use grey scale symbols to represent the spectral classes.

      Result: enhanced field and object boundaries.
2. Locate training and test samples in the enhanced imagery.
3. Use cluster analysis to refine training field selection.
4. Classify the image using the training class definitions.
5. Perform quantitative results evaluation using test samples.

The analysis approach developed for Case I is basically unsupervised classification whereas for Case II supervised classification is utilized. Of course the two methods are usually blended appropriately according to the amount and detail of the ground observations available.

Use of Data Features of a Spatial Nature

Investigation has continued into two aspects of data classification involving the spatial information content of the data in addition to the spectral information.

Adaptive Classification.

A model has been formalized for an adaptive classifier in which adaptation to slowly varying image characteristics is performed in an unsupervised manner, i.e., without benefit of ground observations.

The usual approach to adaptive classification is to assume supervised adaptation; i.e., it is assumed that the true classification of data to be used for adaptation is known. This could require a substantial amount of ground observation data, however, so it was decided to develop an adaptive model based on unsupervised adaptation.

Under the assumption of the multivariate Gaussian statistics which we normally use for classification, it is easy to show that the maximum likelihood classifier using simple unsupervised adaptation will usually be unstable in the sense that adaptation will eventually lead it to classify everything into the same class. This will be the case unless the pattern classes are highly separable so that the error rate is essentially zero -- which ordinarily can not be expected. An adaptive classifier model has therefore been developed which assumes a special case of Gaussian statistics and avoids the instability problem.

It is assumed that n-dimensional measurement vectors $X = [x_1, x_2, \ldots, x_n]^T$ are to be classified into m classes $\omega_1, \omega_2, \ldots, \omega_m$. The classes are assumed to be characterized by multivariate Gaussian probability density functions with unequal means and identical covariance matrices:

$$X \varepsilon \omega_i \rightarrow X \sim N(M_i, S)$$

where the M_i, $i = 1, 2, \ldots, m$ are the class mean vectors and S is the common covariance matrix. For simplicity at this stage, a zero - one loss function and equal prior probabilities of the classes are assumed.

Under these assumptions, the classifier is linear with discriminant functions of the form

$$D_i(X) = W_i^T X + c_i \quad i = 1, 2, \ldots, m$$

where

$$W_i = S^{-1} M_i \text{ and } c_i = -\frac{1}{2} M_i^T S^{-1} M_i$$

The covariance matrix is assumed constant, not subject to adaptation. For each observation X to be used for adaptation, the components of the mean vector of class ω_i into which X has been classified is updated according to

$$m'_{ik} = (1-\alpha_k)\, m_{ik} + \alpha_k x_k, \quad k = 1, 2, \ldots, n$$

where α_k is an adaptation parameter which may have different values for different components of the multidimensional measurement space.

However, to avoid the instability problem noted earlier, not every classified vector is used for adaptation. Instead, only those which are "confidently classified" are used, which is accomplished by "thresholding" the discriminant values: an observation X is used to update the mean vector of the class into which it was classified only if the value of the discriminant lies within a specified range, say $D_i \pm T_i$, where D_i is the mean value of the discriminant for class ω_i and T_i is the threshold for class ω_i.

Thus, to completely define the classifier model it must be possible to specify

(1) the adaptation parameters, α_k

(2) the discriminant threshold values, T_i. This can be accomplished by a well-defined procedure, the details of which will be provided in a subsequent report. Experiments on limited amounts of data have been encouraging and attention is currently being directed toward quantifying the benefits (in terms of classifier accuracy) which may be

derived from this approach as applied to multispectral scanner imagery.

Use of Context

Contextual information in image data can be utilized in a number of ways. Two that have been studied in the present investigation are (1) improved analysis results through sample classification (rather than classification of individual data points), and (2) compression of results storage through object isolation, classification, and coding. The approach is outlined below; theoretical and procedural details appear in [1].

An algorithm known as Recursive Image Partitioning (RIMPAR) has been developed for the decomposition of multispectral images into "objects". In this context, an object is a region in the data which (1) exhibits internal homogeneity and (2) contrasts with its surroundings. Basically, RIMPAR divides an image into successively smaller rectangular blocks. At each step in the algorithm, a block is subdivided unless

(1) the size of the block (largest side) is less than a parameter MINSIZE; or

(2) it is likely that the image points within the block are from a single object.

Thus, RIMPAR partitions an image into blocks such that each object is approximated by a set of adjacent blocks. The algorithm has been found to work well with noisy or fuzzy object boundaries and objects of varying texture.

Figure VII-1 is a flow diagram of the basic RIMPAR algorithm. Although many of the details are beyond the scope of this report (see [1]), a brief explanation of some important aspects of the algorithm follow.

---

[1] Thomas V. Robertson, "Multispectral Image Partitioning", submitted to the 1973 Telecommunications Conference, Atlanta, Georgia, November 1973.

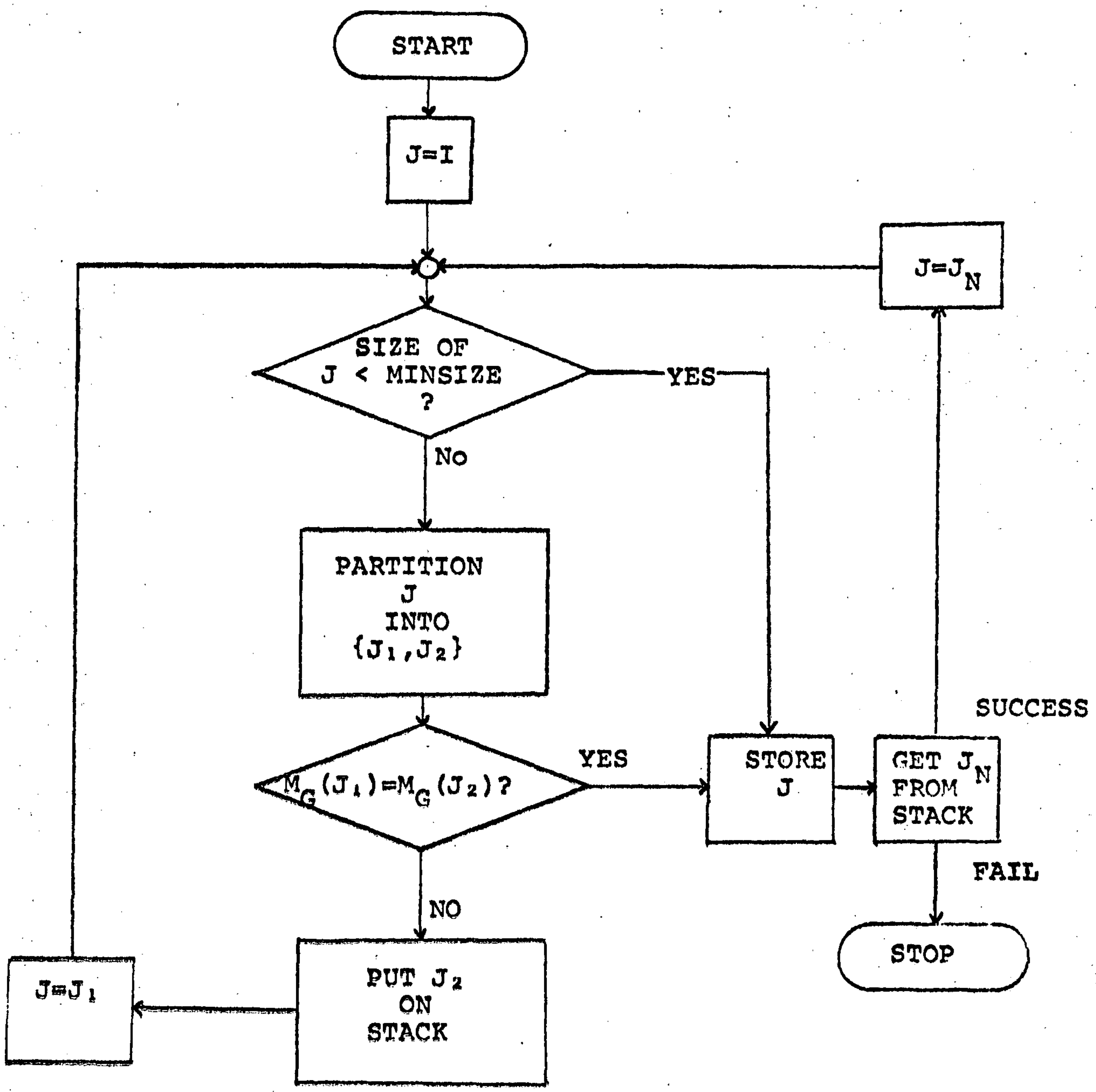


Figure VII-1. Basic RIMPAR flow chart.

The value of the parameter MINSIZE which sets a lower limit on the size of the blocks produced is not critical but allows the user to input prior knowledge he may have about the anticipated size of objects in the image. In this way, the user can reduce the total processing time and storage requirements by avoiding subdivision into unnecessarily small blocks.

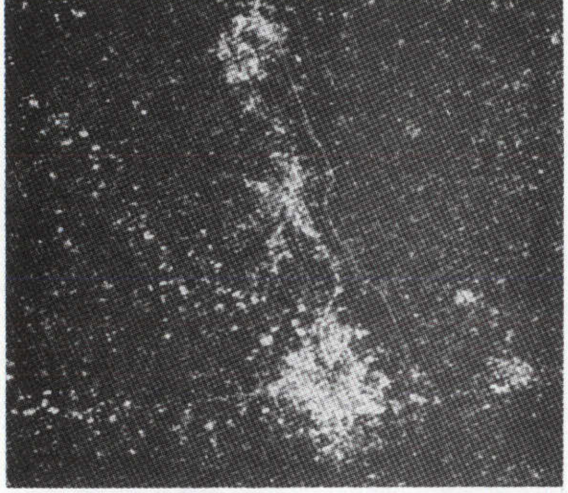
At each step, several trial partitions of a block are attempted and only the "best" one is tested to see if the partition should stand. The trial partitions are evaluated using the partitioning criterion introduced in an earlier report, namely:

$$V_P = \frac{1}{N_T} \sum_{j=1}^{n} N_j \sigma_j^2$$

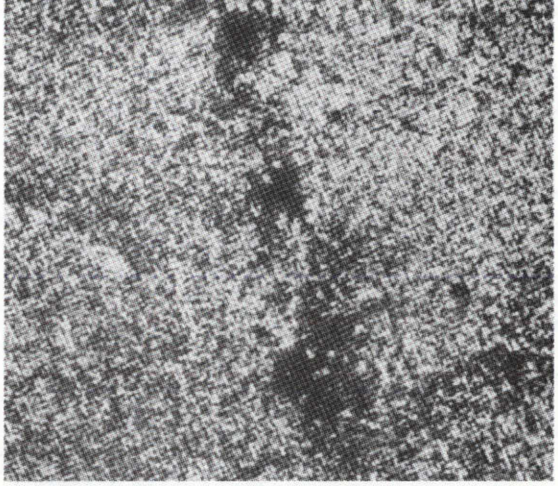
where the block to be partitioned, containing N_T image points, has been subdivided into n sub-blocks, the jth sub-block containing N_j image points and having a feature variance σ_j^2. The optimal partition P* is the one in which the blocks correspond exactly to objects in the image, and it can be proven that $V_{P*} \leq V_P$ for any partition P. Thus, the "best" trial partition is the one with minimal V_P.

$M_G(J_i)$ is the mean vector associated with subimage J_i. The test for equality of subimage mean vectors shown in the flow diagram is in fact a statistical test which takes into account the data variability within the sub-blocks.

Detailed results of the experimental application of RIMPAR will be presented and discussed in a future report. As an example of its performance, Figure VII-2 illustrates its ability to isolate urban areas in ERTS-1 MSS imagery.

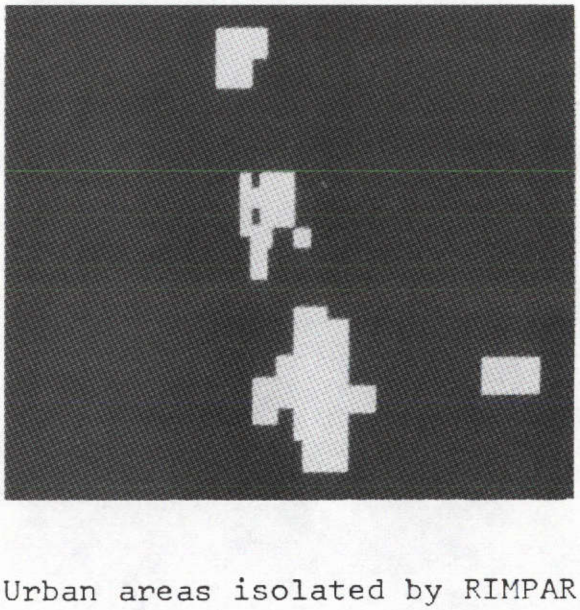


Channel 2



Channel 4

(a) ERTS Imagery Containing Urban Areas.



(b) Urban areas isolated by RIMPAR.

Figure VII-2. An Application of RIMPAR.

Layered Classifiers

Layered classifier decision logic provides a capability for making maximal use of available multispectral information at minimal data processing cost. A hypothetical example illustrating this concept was provided in our previous 6-month interim report. The potential advantage of the layered decision method arises from the fact that it can be "cheaper" (in terms of processing time) to make n decisions each based on m variables than it would be to make a single decision based on n x m variables (this is the case for multivariate maximum likelihood classifiers, for instance).

Determining the optimal layered decision logic is a considerable problem in combinatorics: One must select the best decision procedure from a very large class of possibilities.

In seeking an optimal design one must have a criterion for measuring the goodness of any given design. For the purpose of this investigation we use the following general optimality criterion for evaluating a decision procedure defined in terms of a decision logic tree structure.

$$V(D_i) = -[E(D_i) + kT(D_i)]$$

where $V(D_i)$ is referred to as the performance resulting from decision logic tree D_i; $E(D_i)$ is the expected classification error resulting from application of D_i; $T(D_i)$ is the expected implementation cost of D_i; and k is a proportionality constant. Under this criterion performance is optimized (V is maximized) by the decision logic which minimizes a combination of the expected probability of error and the cost of implementation of the logic. In general, it will be necessary to particularize V to suit a given problem. For example, E could be chosen as the Bayesian error criterion; T could be chosen to reflect the cost of hardware implementation together with the expected time to classify an observation. The value of k reflects the relative importance of implementation cost versus classification accuracy.

To be specific, we restate the problem as follows: Given a set of multiclass and multivariate Gaussianly distributed data the task is to construct a multistage classifier D which maximizes the optimality criterion given above with E taken as the Bayesian error criterion and T the total computation time.

A typical decision tree might be a shown in Figure VII-3. Notice the following aspects of this tree structure:

(1) several classes may be classified in a single decision step.

(2) the feature subset used at each decision step will vary.

(3) individual classes may be paritioned into subclasses at any decision step.

The number of trees that can be so constructed for any given problem is in general, extremely large. We propose to use an heuristic procedure to reduce the number of trees which must be evaluated.

One of the methods proposed here is referred to as a "best-two pruning with second level cut-off". That is, in deciding on the analysis procedure to be applied at any mode in the tree, only the best two possiblities are retained for further investigation (the criterion noted above is always used for this type of evaluation), and one of these two possible procedures is discarded after the tree has been expanded by two more levels. The details of the heuristic

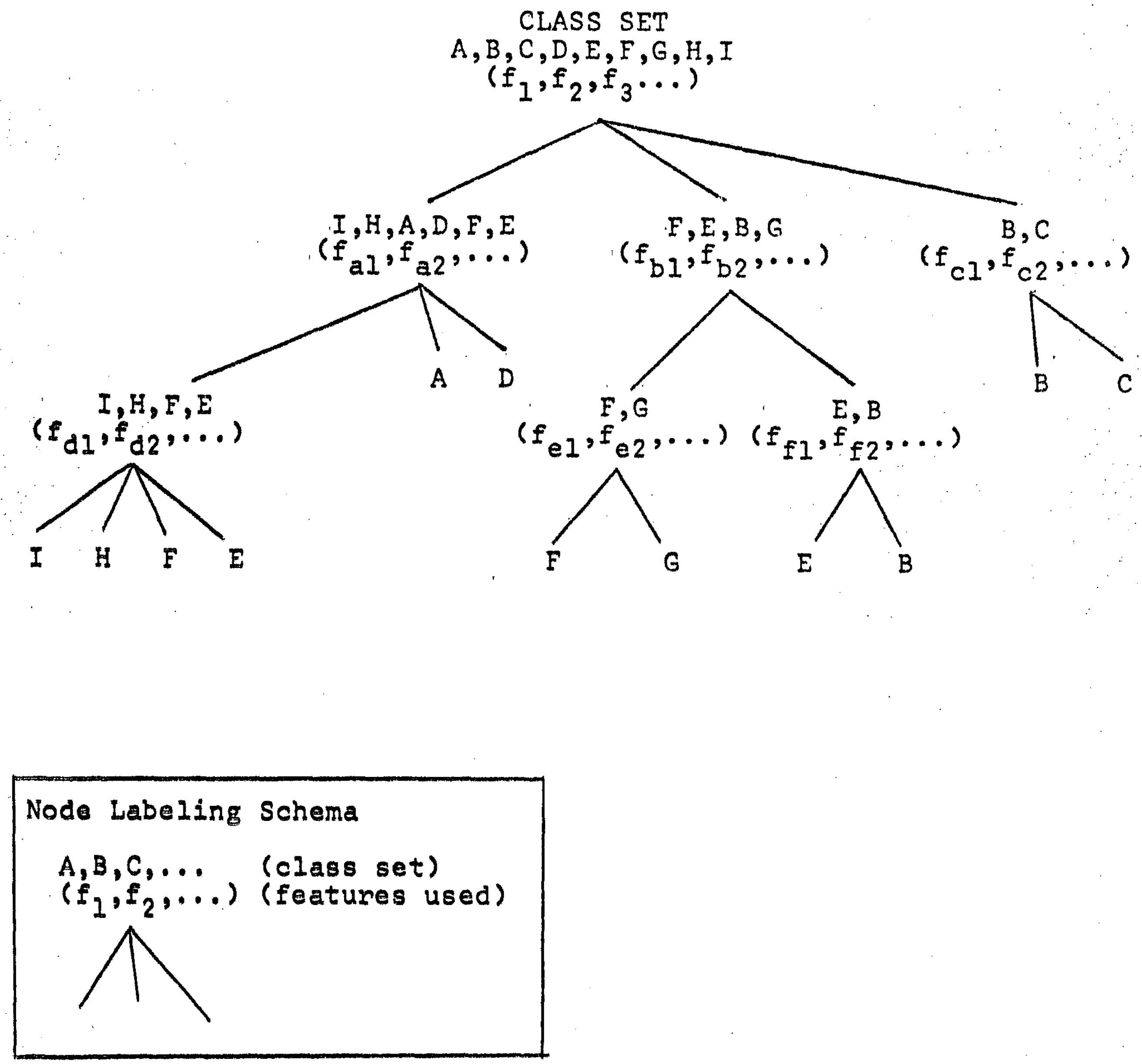


Figure VII-3. A Tree Describing Layered Decision Logic.

procedure are beyond the scope of this report and will be the subject of a later detailed report. Figure VIII-4 shows a simplified block diagram of the search heuristic.

Experiments using this procedure to design multilevel decision logic for typical remote sensing problems have produced decision structures capable of yielding classification accuracy competitive with the best possible accuracy obtained from single level decision classifiers and exhibiting classifier speed much faster than the single level procedure. Details of these experiments will also be presented in a later report.

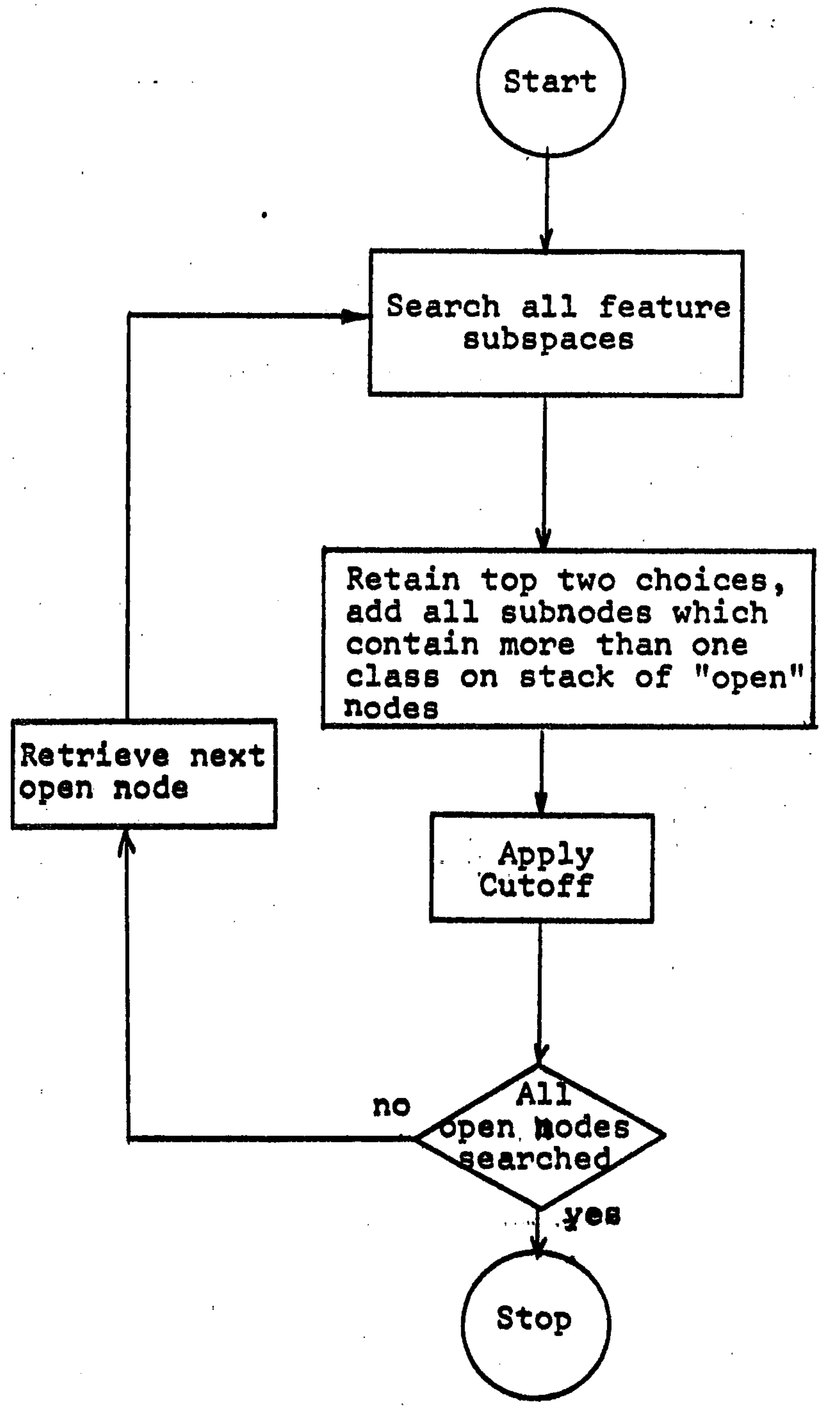


Figure VII-4. Simplified block diagram of search heuristic.

E

VIII. DATA REFORMATTING AND OVERLAY

Introduction

The reformatting and overlay activity proceeded in a relatively operational manner during the reporting period with relatively few data or processor problems occurring. Data quality was generally acceptable and the existing reformatting and overlay software was capable of satisfying LARS user needs for digital ERTS data on an acceptable time scale. A new technology development effort was begun to seek means of performing all digital geometric transformation of sensor processed MSS data. A developmental program is presently in operation which can produce approximately corrected data which matches map geometry over small areas.

Reformatting

Satellite Data Reformatting

During the report period 307 ERTS MSS frames were received, 67 of which were duplicates of frames previously received. Also, 90 frames were returned to NASA Goddard. Frames were returned for the following reasons: Duplicate frame, insufficient test site coverage, or completion of work on the frame. Complete or portions of 115 ERTS MSS digital frames have been reformatted into LARSYS III format and placed in the LARS Multispectral Image Storage Tape library.

No serious problems have been encountered in handling and reformatting this data. Minor problems included storage space, procedure for returning tapes, and improperly labeled or written ERTS SYCI tapes.

ERIM Reformatting

Analog multispectral scanner data collected by the Environmental Research Institute of Michigan Aircraft

System on 2 missions were processed. Wabash Valley flightlines 4, 4a, and 5 collected on October 10, 1972 during mission 67M were digitized and reformatted. Flightline 2b of mission 71M of January 2, 1973 was also digitized and reformatted.

New Programs

A data reformatting program to convert ERTS Scene Corrected digital frames to LARSYS III format was completed. This reformatting task was much more difficult than that of the SYCI data because of the severe incompatibility of the NASA SCCI data tape format and LARSYS tape format. The approach to the problem was through the use of an 8 megabyte virtual machine. Using the very large virtual memory, the program loads one-eighth of the full SCCI frame all channels; reformats and writes the data in LARSYS format; then loads the next eight and so on. Data from 5 SCCI frames have been reformatted.

Transfer of Technology

Miss Lottie Brown of NASA Goddard requested the LARS ERTS SYCI data reformatting program. As a result of that request, the reformatting program and all LARS subroutines used by the programs were transferred to Miss Brown over the Goddard remote terminal.

Temporal Overlay

The LARS image registration system is continuing to produce acceptable overlay results for sequential ERTS MSS frames over relatively short time intervals. Data from four areas have been overlayed during the reporting period. 1.) Lynn County, Texas, three times: October 9, 1972 - frame 1078-16524, November 15, 1972 - frame 1114-16532 and

December 2, 1972 - frame 1132-16532; 2.) Northern Illinois, three times: August 9, 1972 - frame 1017-16093, September 19, 1972 - frame 1053-16095, October 2, 1972 - frame 1017-1609501; 3.) Eastern Missouri, three times, two areas: September 13, 1972 - frames 1052-1605500, and 1052-1605200, August 26, 1972 - frames 1034-1605500 and 1034-1605200, October 1, 1972 - frames 1070-1605501 and 1070-1605201; 4.) Tippecanoe County, Indiana, three times, September 30, 1972 - frame 1064-1558501, October 19, 1972 - frame 1088-1605001 and November 24, 1972 - frame 1124-1605201. In each case, the overlay area is approximately 1600 columns and 1200 lines or about 25% of a frame.

Temporally overlayed data is being analyzed in three project areas: Crop Species Identification, Soil Association Mapping, and Data Reformatting and Overlay. The benefits of overlay fall into two general categories: 1.) Dimensionality Expansion, 2.) Coordinate System Standardization. Work on evaluation of the temporal dimension for classification is beginning and results should be available in the January 1974 report. The value of coordinate standardization has been demonstrated in cases where a large number of test and training fields had been selected from data from a given date. The fields were automatically located in new data overlayed on the previous data thereby eliminating the need for redefining the fields. Overlayed data also allows the user to find features of interest in data at one time when they may be more visible and to analyze the same points from another time when contrast may be poor. Another capability made available by temporal overlay is change detection via image differentiating. Work is in progress as part of the overlay technology development project to evaluate temporal overlay for change detection. No results are available at this time.

Geometric Correction

Many of the users of digital ERTS MSS data at LARS expressed interest in obtaining pictorial data output formats which would be easily relatable to maps and aerial photography of the same area. Digital system corrected data reproduced in pictorial form on a computer line printer or digital display has different scale factors in the X and Y directions, is rotated nominally 12° from North in North America, and contains skew distortion due to earth rotation. Also, many second order spatial distortions exist due to sensor and satellite induced errors. Scene corrected digital data has not been available in quantity and it apparently contains severe geometric and radiometric errors which have restricted interest in this data form (see Section X). Thus an experiment was begun to determine the value of performing a simple linear transformation on the system corrected digital data to remove the major geometrical problems.

Software was brought on line which enabled relatively small blocks of data to be rotated, deskewed, and rescaled such that the image output product would be North-oriented and have given scale factor. Unknown higher order errors were expected to remain in the data; however, it was assumed that the approximate correction would be very valuable over small areas. The correction program uses a "nearest neighbor" rule in transforming the data thus the spectral value of the data is not altered; however, points may be skipped or duplicated and any point in the transformed data may be up to 50 meters away from its true position in the scene. The benefits of the corrected form of the data were thought to outweigh the errors introduced in this process. Figure VIII-1 contains digital display photographs of data before and after correction. The ERTS data when reproduced on a line printer has a scale near 1"=24000" and since the topographic maps in use by researchers were 1:24000 scale the preferred line printer scale for correction of the data was 1:24000. Figure VIII-1a contains uncorrected data and Figure VIII-1b is a portion of a corrected run scaled to produce a North-oriented 1:24000 scale line printer output. (Note that since the line printer has a .8 to 1 aspect ratio and the digital display aspect ratio is 1 to 1 the display image will appear stretched in the horizontal direction.) Availability of corrected data has been extremely helpful in many of the applications research projects.

During the next bi-monthly period overlay and geometric transformation programs will continue to be improved and made more efficient and convenient to use. Also, methods of improving the accuracy of the all digital geometric correction technique will be studied; however, since this work is outside the scope of the present study, it will not be pursued in detail.

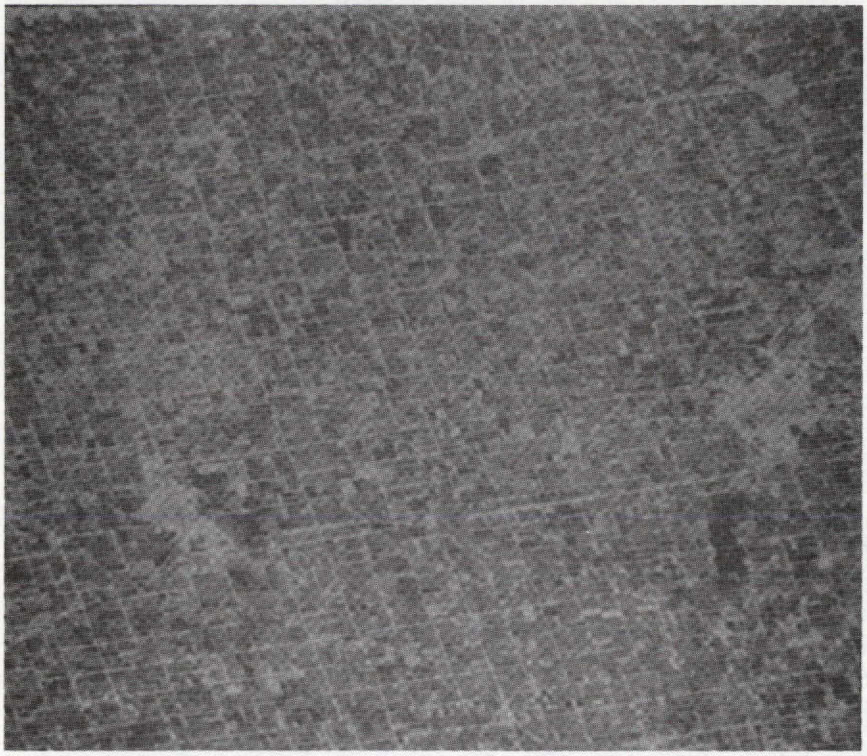


Figure VIII-1a. Digital Display Image of ERTS Frame August 9, 1972, Northern Illinois, MSS Band 5 (.6-.7μm)

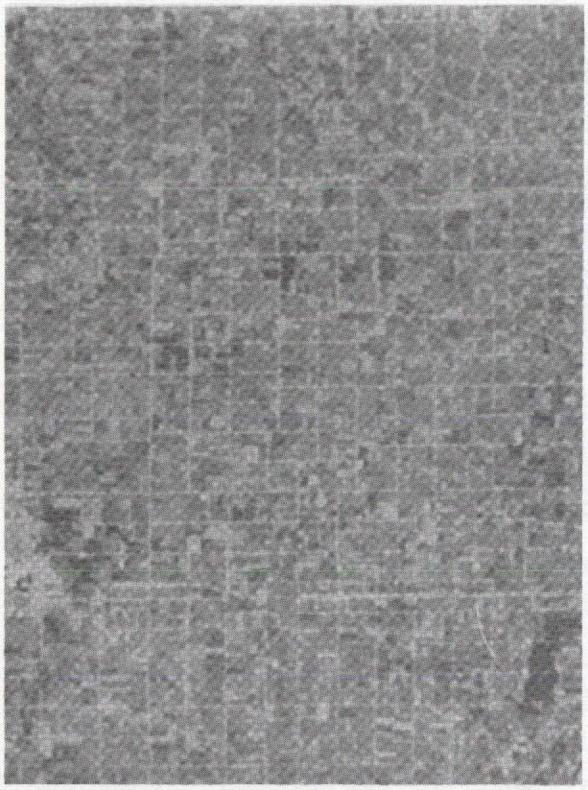


Figure VIII-1b. Geometrically Transformed Version of above data rescaled to produce 1:24000 scale image on computer line printer.

IX. ATMOSPHERIC MODELING

Introduction

Activity during this period has continued to be directed at preparation of the atmospheric model for application to test sets of ERTS data.

Procedures

Several aspects of the computational results have been studied. These are: (1) selection of the best computational parameters to obtain an acceptable level of model accuracy, (2) tests of model results for possible meteorlogical conditions, and, (3) selection of atmospheric parameters for subsequent computations.

Results

Accurate estimates of measured radiance values requires proper representation of scattering by aerosol particles of different sizes. This is influenced by the increment used when summing over the range in particle sizes and the maximum size of particle included. The size increment is given by the dimensionless parameter $\Delta x = 2\pi\Delta r/\lambda$ where r is particle radius and λ is wavelength, and the maximum particle size is given by r_{max}. The computational time required is greatly influenced by the values used for Δx and r_{max}, increasing for smaller values of Δx and larger values of r_{max}. Comparison of several test cases demonstrated that intensity estimates with $\Delta x = 0.5$ and $r_{max} = 2\mu m$ departed less than 1% from a more precise, and much more time consuming, case. All subsequent calculations are being performed with these parameter values.

The various test cases of model accuracy have been performed for the mid-latitude Summer and Winter standard atmospheres. These have been chosen to simulate expected conditions when ERTS data sets are analyzed.

An important factor in correctly determining the atmosphere's effect upon ERTS data is the amount of absorption by water vapor in the visible and near infrared spectrum. Consequently, the model has been modified to accommodate both ozone and water vapor. However, values of water vapor absorption coefficients in the visible spectrum are not well known. Results thus far indicate that water vapor should certainly be included in ERTS bands 6 and 7. Also, it may be necessary to include water vapor effects in ERTS bands 4 and 5; this result is still tentative, subject to further efforts to determine water vapor absorption coefficients for those bands.

The multispectral scanner (MSS) on ERTS measures total radiance over broad spectral intervals, while the model performs calculations at a given wavelength. A procedure for comparing these different quantities has been developed.

Conclusions

The atmospheric model has been modified and tested and is nearly ready for application to ERTS data sets. It is believed that estimates of measured radiances can be made to an accuracy few percent.

Plans for Next Bi-Monthly Period

Final determination of water vapor absorption coefficients must be made. Then the model will be applied to read data for analysis.

Problems and Recommendations

The importance of water vapor in the different ERTS bands has been difficult to assess. This has delayed progress somewhat.

X. COMPARISON OF SYSTEM CORRECTED AND SCENE CORRECTED CCT DATA

Introduction

This study was initiated to investigate the quality of the scene corrected MSS data products from the ERTS-1 satellite. Since the scene corrected data has been geometrically corrected, it was thought that this data could be utilized in a data overlay scheme to geometrically correct the system corrected digital data products. This procedure would allow a researcher to maintain the radiometric quality of the bulk product while gaining the geometric accuracy of the scene corrected data products. The geometric accuracy would be of considerable use in improving acreage measurements of different crop types by utilizing the overlayed (system corrected upon scene corrected) data set for the classification.

Procedures

The original plan of attack was to use data from the Wabash Valley area. Scene corrected data from these areas was ordered in October 1972 and February 1973. Neither of these data products were received in time for this investigator to use. However, one frame of scene corrected data was received (via USDA) from the Missouri area South of St. Louis (Scene I. D. E-1071-16111-601). This scene was thus used to complete this study comparing the system and scene corrected data sets.

Another aspect of this project was the investigation of the processing equipment hardware and software used in generating the scene corrected data products. None of the references located gave any information that was not already found in the ERTS Data Users Handbook.

Data Product Description

Table X-1 lists the basic characteristics of the data products after they were reformatted to LARS data storage tapes.

Table X-1

Data Form	LARS Run No.	Lines	Columns	Data Channels	Information Bits
System Corrected	72063200	2340	3232	4	CHAN 1,2,3 7 Bits CHAN 4 6 Bits
Scene Corrected	72063201	4096	2204	4	All CHANNELS 7 Bits
	72063202	4096	2204	4	All CHANNELS 7 Bits

Since the scene corrected data is digitized to many more data points, it was reformatted onto two tapes (the left and right halves of the frame). The only other data products available for this frame were the system processed photographic images.

Data Quality Study

Two main areas of quality were studied in this investigation. First, the geometric quality was investigated by looking for any obvious discrepancies in the scene corrected data on the digital display. Secondly, the radiometric quality was also investigated by using the digital display and the LARSYS $STAT program to compare the scene corrected data to the system corrected data. Some results of these investigations are contained in the accompanying photographs.

Results

In studying the geometric quality of the scene processed data, attention was directed towards any obvious geometric anomalies or discontinuities. Shown within the digital display field outlines in Figure X-1 are horizontal or lateral shifts in the data. Also, the right-most box in the figure contains what appears to be a vertical shift of the data.

The radiometric quality was investigated by two means. First, obvious discrepancies were located visually using the LARS digital display on both the system corrected and the scene corrected data. Secondly, the scene corrected data was compared to the system corrected data by choosing "training" fields in the same areas on both data sets. By performing $STAT computer analysis on these fields in each data type, a comparison can be made. In choosing the "training" fields, water was used mainly because of its large size and ease of identification. Also, two agricultural fields were chosen that could be identified in both data sets.

Referring to Figure X-1, it is obvious that there are many points in the scene controlled data which are saturated (i.e. the white spots). This fact is also indicated in the histograms. These saturations are indicated by the large peak at level 127 on the histograms for run #72063201. The system corrected data is pictured in Figure X-2. Note that this data appears consistant and free from saturated points. Also histograms of this run do not indicate any irregularities. Results of statistical analysis of "fields" in the data are not yet available and will be reported in future documents.

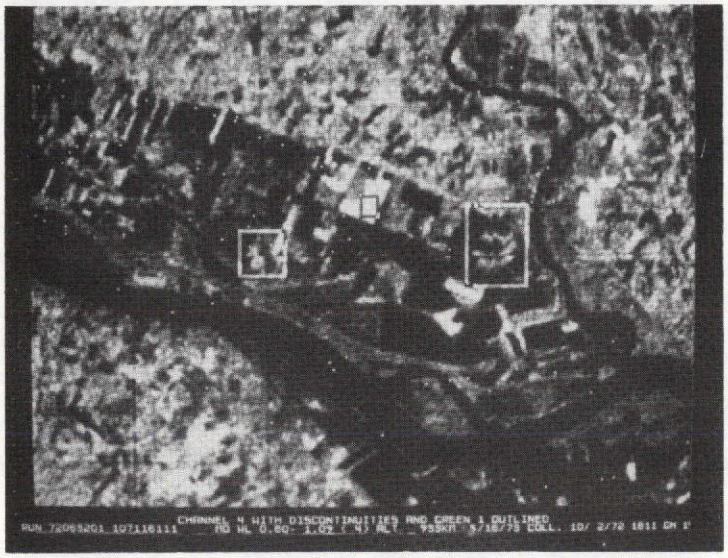


Figure X-1. Digital Display Image of Scene Corrected ERTS MSS Data from frame 1071-16111-601, Channel 4 (.8-1.1μm).

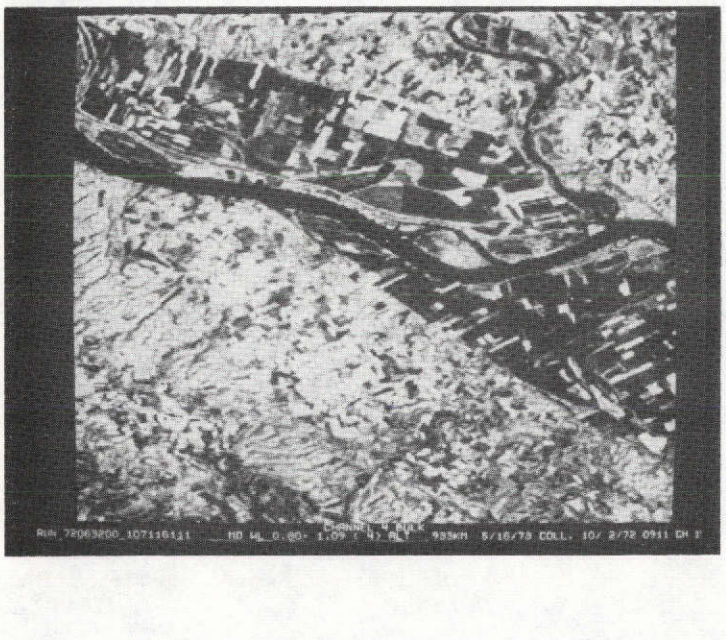


Figure X-2. Digital Display Image of System Corrected ERTS MSS Data, Channel 4 (.8-1.1μm).

Conclusions

The results presented are incomplete and inconclusive; however, sufficient evidence is presented to conclude that serious problems may exist in the particular scene corrected ERTS MSS data studied. The data exhibits both geometric and radiometric distortion observable both visually and statistically. Although other data sets may not contain these distortions, it was decided not to pursue the sensor corrected - scene corrected overlay project further until scene corrected data of adequate quality became available.

XI. Additional Type II Report Items

A. New Technology

A procedure for photographically overlaying maps and ERTS-1 imagery was developed to enable comparison of soil boundaries in the two data forms. A new algorithm is reported on which analyses multispectral images for contextual information. The algorithm (called RIMPAR for Recessive Image Partitioning) decomposes images into blocks such that each object is approximated by a set of adjacent blocks. An approximate digital geometric correction procedure was developed to rotate and rescale ERTS-1 MSS system processed CCT data such that reconstructed images are North-oriented and have a given scale factor.

B. Program for Next Reporting Period

Plans for next period are included in each project section.

C. Conclusions

Conclusions have been drawn in each project reporting section where appropriate.

D. Recommendations

Recommendations are made in Section VI for a study of land use change in Marion County, Indiana using geometrically corrected and temporally overlayed data.

E. Type I Items

1. No operational changes are recommended.
2. No changes in the standing order form are requested.
3. No image descriptor forms are included.

F