NASA CR-151575
ERIM 122700-35-$F_2$



Final Report

# INVESTIGATION OF TECHNIQUES FOR INVENTORYING FORESTED REGIONS

## Volume II: Forestry Information System Requirements and Joint Use of Remotely Sensed and Anciliary Data

RICHARD C. CICONE, WILLIAM A. MALILA,
and ERIC P. CRIST
Infrared and Optics Division

NOVEMBER 1977

Principal Investigator
Richard F. Nalepka



Prepared for
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Johnson Space Center
Earth Observations Division
Houston, Texas 77058
Contract No. NAS9-14988, Task 5, Vol. II
Technical Monitor: I. Dale Browne/SE8

ENVIRONMENTAL
RESEARCH INSTITUTE OF MICHIGAN
FORMERLY WILLOW RUN LABORATORIES. THE UNIVERSITY OF MICHIGAN
BOX 8618 • ANN ARBOR • MICHIGAN 48107

| 1. Report No. | 2. Government Accession No. | 3. Recipient's Catalog No. |
|---|---|---|
| 122700-35-$F_2$ | | |

4. Title and Subtitle
INVESTIGATIONS OF TECHNIQUES FOR INVENTORYING FORESTED REGIONS
VOLUME II: Forestry Information System Requirements and Joint Use of Remotely Sensed and Ancillary Data

5. Report Date
November 1977

6. Performing Organization Code

7. Author(s)
R. C. Cicone, W. A. Malila, and E. P. Crist

8. Performing Organization Report No.
122700-35-$F_2$

9. Performing Organization Name and Address
Environmental Research Institute of Michigan
Infrared & Optics Division
P. O. Box 8618
Ann Arbor, Michigan 48107

10. Work Unit No.

11. Contract or Grant No.
NAS9-14988

12. Sponsoring Agency Name and Address
National Aeronautics & Space Administration
Johnson Space Center
Houston, Texas 77058

13. Type of Report and Period Covered
Final Technical Report
14 May 76 - 14 Nov 77

14. Sponsoring Agency Code

15. Supplementary Notes: The work was performed for Earth Observations Div./SF as a part of the Nationwide Forestry Applications Program, a joint program of NASA and U.S. Forest Service. Mr. I. Dale Browne/SF3 was Technical Monitor for the contract, Dr. D.L. Amsbury/SF5 was NASA Task Monitor, and Dr. F.P. Weber/SF5 was the cognizant USFS representative for the Task. Vol. I presents results from another subtask. Mr. R.F. Nalepka was ERIM's Principal Investigator on this contract.

16. Abstract

This report documents activities conducted by ERIM under the Nationwide Forestry Applications Program to investigate, develop, and evaluate systems and techniques that make use of remotely sensed data for forestry applications. A two-fold effort was directed at methods for enhancing U.S. Forest Service information systems.

The first activity pertained to the specification of requirements for incorporating remotely sensed data into a USFS forest and rangeland information system. The impact of introducing remotely sensed data as a new layer of spatial data in a Geographical Information System (GIS) was analyzed. A generalized GIS environment was devised that could both facilitate the incorporation of remotely sensed data and accommodate other expected information system requirements. In this context, discussion was carried out that addressed the special problems associated with processing remotely sensed data including: (1) geometry and data volume considerations, (2) a preferred processing environment, and (3) the utility of a data base manager. Further specification and development of the generalized GIS concept is recommended.

The second activity was an investigation of techniques for improving information extracted through the joint use of Landsat data and associated information from non-remote-sensor sources, and by applying and adapting advanced agriculturally oriented information extraction techniques to remotely sensed data from forest regions. Grand County, Colorado, a mountainous region with predominantly coniferous forests, was chosen as the test site. Among others, the investigations carried out included: (1) an analysis of the effects of terrain topography on Landsat signals and simulated forest canopy reflectances, (2) the development of preprocessing techniques that reduce the variability due to topography, (3) an analysis of the utility of spectral and spectral/spatial clustering techniques, and in expansion of the spectral/spatial technique, and (4) an examination of the applicability of the Tasselled Cap transformation of Landsat data, and a related transformation to forestry data.

Preprocessing using the techniques developed was found to substantially reduce topographically induced variability in Landsat signals and improve classification performance. The clustering and data transformation techniques also were found to be beneficial and of potential value for forestry and rangeland applications. Continued development and testing of such improved information extraction techniques plus multitemporal techniques are recommended on other sites and ecosystems.

| 17. Key Words (Suggested by Author(s)) | 18. Distribution Statement |
|---|---|
| Forestry Remote Sensing<br>Landsat Data<br>Terrain Topographic Data<br>Information Systems<br>Preprocessing & Classification | Initial distribution is listed at the end of this document. |

| 19. Security Classif. (of this report) | 20. Security Classif. (of this page) | 21. No. of Pages | 22. Price* |
|---|---|---|---|
| Unclassified | Unclassified | xi + 146 | |

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

# PREFACE

This document reports processing and analysis efforts on one task of a comprehensive and continuing program of research in multispectral remote sensing of the environment. The research is being carried out for NASA's Lyndon B. Johnson Space Center, Houston, Texas, by the Environmental Research Institute of Michigan (ERIM). The basic objective of this program is to develop remote sensing as a practical tool for obtaining extensive environmental information quickly and economically.

The specific focus of the work reported herein was on forestry applications of remote sensing. It constitutes work on one subtask of a two-part effort under the Nationwide Forestry Applications Program (FAP), a joint program of NASA and the U.S. Forest Service, with Headquarters at the Johnson Space Center. The other subtask is reported in Volume I, ERIM 122700-35-$F_1$, entitled, "Reflectance Modeling and Empirical Multispectral Analysis of Forest Canopy Components", by F. G. Sadowski and W. A. Malila.

The research covered in this report was performed under Contract NAS9-14988 during the period 15 May 1976 to 14 Nov 1977. Mr. I. Dale Browne (SF3) served as the NASA Contract Technical Monitor, Dr. David Amsbury (SF5) was NASA Task Monitor, and Dr. F. P. Weber (SF5) was the cognizant USFS Representative. At ERIM, the work was performed within the Infrared and Optics Division, headed by Richard R. Legault, Vice-President of ERIM, in the Information Systems and Analysis Department, headed by Dr. Quentin A. Holmes. Mr. Richard F. Nalepka, Head of the Multispectral Analysis Section served as Principal Investigator and Dr. William A. Malila as Task Leader.

In addition to the support of those sponsor and ERIM personnel noted above, the authors wish to acknowledge the assistance of other ERIM staff members who have been developing techniques in the LACIE agricultural context, some of which we applied here. Also, we acknowledge

use of the LIGMALS computer program package, initiated within The University of Michigan's School of Natural Resources and expanded at ERIM, for our initial Landsat data format conversions. Subsequent processing and analysis was performed within the QLINE/llLINE multispectral processing system developed by the Information Systems & Analysis Department, especially R. H. Hieber. W. Richardson gave helpful advice for programming the ADJOIN algorithm, and R. Kauth encouraged the use of the Tasselled-Cap transformation on forestry data. Ms. Darlene Dickerson, along with Ms. E. Hugg, M. Warren, and J. Watters, provided efficient and accurate typing support throughout the contract period and for this report. We also wish to acknowledge Lockheed Electronics Corporation personnel, Dr. E. Kan and R. Dillman, of Houston, Texas, for their assistance in providing us the data and materials for the Grand County test site.

CONTENTS

CONTENTS (Cont'd)

CONTENTS (Cont'd)

FIGURES

FIGURES (Cont'd)

TABLES

## SUMMARY

This report documents activities carried out by ERIM under the Nationwide Forestry Applications Program to investigate, develop, and evaluate systems and techniques that make use of remotely sensed data for forestry applications. A two-fold effort was directed at methods for enhancing U.S. Forest Service information systems. The first activity pertained to the specification of requirements for incorporating remotely sensed data into a USFS forest and rangeland information system. The second activity was an investigation of techniques for improving the quality of information extracted through joint use of remotely sensed data and associated information (e.g., topographic) from non-remote sensing sources, and by applying and adapting advanced agriculturally oriented information extraction techniques to remotely sensed data from forest regions.

Recommendations for incorporating remotely sensed data in a USFS information system were devised through an analysis of the nature of these data and their impact on a Geographical Information System (GIS) designed to meet the needs of the USFS. The potential value from the joint use of remotely sensed data and other spatial data led to the consideration of remotely sensed data as another layer of spatial data in a GIS. Analysis of existing systems indicated that characteristics of remotely sensed data and their processing could substantially impact the efficiency of a geographical information system. Hence, a generalized geographical information system concept was devised which could both facilitate the incorporation of remotely sensed data and accommodate expected USFS information system requirements.

In this generalized environment, recommendations were made with regard to:

1. remotely sensed data volume and geometry considerations.
2. an optimal processing environment for remotely sensed data that is of general applicability.

3. the use of a data base manager to interface various layers of data and processing functions.

The proposed environment not only accommodates remotely sensed data, but its generalized character provides an overall computer-based geographical information system environment. It is recommended that the feasibility of this approach as well as its further specification, be analyzed in greater depth, since it is not commercially available at present.

To reach its potential value, an information system expanded to include remotely sensed data must have available techniques for the joint use of the remotely sensed data and other spatial data. To investigate such techniques, a test site was chosen in Grand County, Colorado, a mountainous region with predominantly coniferous forest. A more intensive test site chosen therein was the Fraser Experimental Forest in the Arapaho National Forest. Topographic information (digital terrain elevation, slope, and aspect) were registered and merged with Landsat data. A relative insolation factor, representing the degree of solar illumination at the ground, was derived from the topographic data and Landsat acquisition parameters. Analysis of both actual Landsat data and modeled reflectance data for forest canopies indicated that terrain features have a significant effect on the Landsat signal due to the varied illumination on various surface orientations. Preprocessing techniques utilizing the relative insolation factor, and modifications thereof based on modeling results, were therefore developed to normalize the effective illumination at each given scene point. Substantial improvements were achieved in minimizing the variation of Landsat signals due to mountainous terrain topography.

Advanced, agriculturally oriented, information extraction techniques also were applied to these forest data. Included were two data transformations developed for the Large Area Crop Inventory Experiment, and clustering techniques that make use of spatial as well as spectral information.

The two data transformations employed were the Tasselled-Cap transformation and a polar green-angle/brightness-radius transformation. The Tasselled-Cap transformation has been found useful in agricultural applications for data analysis, screening, and compression. It was found that these same advantages are afforded forest data analysis and processing. The transformed data have convenient physical interpretation in spectral space. The Tasselled-Cap brightness variable and brightness-radius were well correlated with aspect and the relative insolation factor. The Tasselled-Cap green axis and polar green-angle were indicators of green vegetative cover. In addition, the first three Tasselled-Cap variables were found to contain 99% of the data variability (96% in green and brightness), indicating that such a transformation provides a viable means for compression of Landsat data from forested scenes, a fact especially important for multitemporal analyses.

Statistical clustering of data is a technique by which natural groupings of data are automatically detected. Both spectral and spectral/spatial clustering techniques were applied. The spectral/spatial algorithm called BLOB, designed for the detection of rectangular field-like shapes in agriculture regions, was not optimized to detect the irregular, elongated shapes encountered in a forest scene. Hence, the algorithm was adapted into a new algorithm called ADJOIN. ADJOIN involves a two-pass procedure in which a grid of spectral/spatial clusters (blobs) is defined using an algorithm similar to BLOB, and then is processed so that spectrally similar neighboring spatial clusters are adjoined. The algorithm successfully delineated forest, aquatic, rangeland, and non-vegetative regions.

Both spectral and spectral/spatial clustering show promise for forest applications. Preprocessing prior to clustering was especially advantageous since it was found to improve the qualitative appearance

of the unsupervised classification maps, eliminating most topographically induced spectral subdivisions.

The results of the analysis conducted on Landsat data for Grand County, Colorado, are of sufficient promise for us to recommend that activities be conducted to:

1. further develop and test preprocessing techniques based on ancillary terrain features, with use of reflectance modeling.
2. investigate multitemporal techniques for forest inventories, including data compression techniques such as the Tasselled-Cap transformation.
3. continue the investigation of improved information extraction techniques for forest and rangeland applications, like clustering.
4. extend processing and analysis studies of the types conducted for the Grand County site to other sites containing deciduous as well as coniferous forests and different ecosystems.

Additional investigations and their recommendations are reported in Volume I [1].

2

# INTRODUCTION

Requirements exist for extensive periodic inventories of the Nation's forests and rangelands, as a result of the Forest and Rangeland Renewable Resources Planning Act of 1974. Remote sensing technology has the potential for providing inputs to help produce inventories efficiently, accurately, and consistently. Remotely sensed information also could be useful in other aspects of the management of these important resources.

However, neither remotely sensed data alone nor other forms of data alone can provide the desired types and levels of information. A need exists for joint information extraction. This can be accomplished through a Geographical Information System (GIS) that combines remotely sensed data with the other types of data on our forest and rangeland resources as well as incorporating the necessary processing capabilities.

One objective of the work reported in this volume was to investigate requirements on forest and rangeland information systems if they are to efficiently and effectively incorporate and utilize remotely sensed data. The other objective was to investigate computer processing and automated classification techniques for using ancillary, geographically based data in combination with remotely sensed data, to improve classification accuracy and the quality of the information extracted.

3

# REQUIREMENTS FOR THE INCORPORATION OF REMOTELY SENSED DATA IN A FOREST AND RANGELAND GEOGRAPHICAL INFORMATION SYSTEM

## 3.1 INTRODUCTION

The incorporation of remotely sensed data in a geographical information system (GIS) requires a thorough understanding of the nature of both geographical information systems and remotely sensed data. The most common forms of remotely sensed data are characteristically spatial, as are the data typical of geographical information systems. The expansion of an information system to incorporate remotely sensed data will require interfacing a variety of data processing strategies and techniques, as well as a variety of data types.

Let us examine, for a moment, one hypothetical but not uncommon approach to the incorporation of remotely sensed data in a GIS. Consider the diagram in Figure 1. Here the remotely sensed data base and processing subsystem are disjoint from the GIS data base and processing

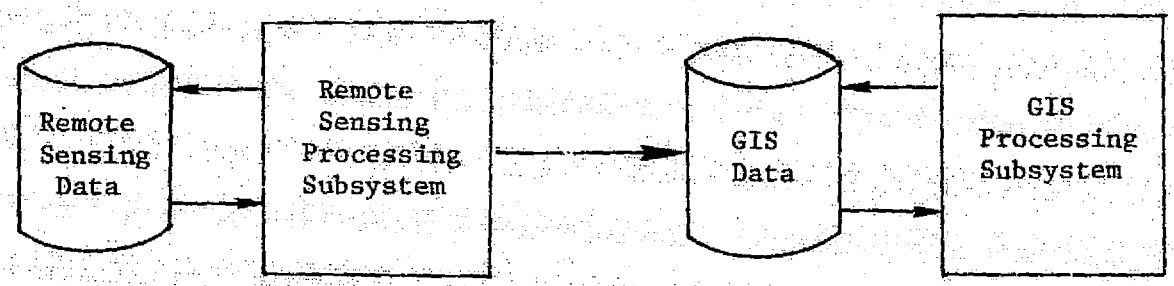



FIGURE 1. NON-OPTIMAL APPROACH TO THE INCORPORATION OF REMOTELY SENSED DATA IN A GEOGRAPHICAL INFORMATION SYSTEM

subsystem. Remotely sensed data would be used, for example, to assign land use categories that would then be loaded into a GIS data base. On the surface, this seems to be an adequate approach to the incorporation of remote sensor data in a GIS. However, it has two inadequacies that will eventually lead to inefficiencies in its use: (1) many of the data structuring and processing requirements in the two subsystems are very similar, and (2) data requirements may overlap extensively as remote sensing applications grow. For example, the processing of remotely sensed data for land use classification may require topographic, climatological, or geological information, or a need may grow for extensive change detection analysis in the GIS environment requiring multiple remote sensor inputs. The result will be the development of two processing systems and two data bases that are quite similar.

A more appropriate approach is to consider remotely sensed data as another layer of spatial data within the GIS data base, and the processing subsystem as an extension of the GIS system. The task becomes, then, one of understanding the features of a GIS in light of the structural and processing requirements imposed by the incorporation of remotely sensed data into the system.

This report addresses requirements for incorporating remotely sensed data into a USFS information system for forest and rangeland applications. For the reasons stated above, the approach taken has been to establish specific requirements in a generalized GIS environment. The task involves a basic understanding of four items: (1) the nature of a geographic information system, (2) specific USFS requirements for a GIS, (3) specific attributes of remotely sensed data and processing systems, and (4) the dynamics of a system design incorporating the first three items.

Sections 3.2 and 3.3 and Appendix I conceptualize geographical information systems, providing several example systems, and proposing a generalized approach to the design of such a system. It is hypothesized

that this generalized approach facilitates the inevitable growth and expansion of a GIS.

Section 3.4 addresses certain USFS requirements for a forestry and rangeland information system.

Section 3.5 provides specific recommendations for the incorporation of remotely sensed data in a USFS geographical information system. Appendix IV is included to complement and further expand upon certain topics brought up in Section 3.5. Various aspects are addressed, including the data elements (3.5.3; IV.3) and processing algorithms (3.5.4; IV.4) required, in light of particular data structures (3.5.1; IV.1) and a specific processing environment (3.5.2; IV.2) that together provide a generalized atmosphere that is adaptable to processing and data base requirements not specific to remotely sensed data.

## 3.2 GEOGRAPHICAL INFORMATION SYSTEMS

Computer-based information systems are found in a variety of forms. Systems have been designed to manage the tedious bookkeeping involved in cataloging such items as library materials, business accounts, and water quality control data. These information systems can be divided primarily into two categories: (1) object-oriented, as a business account system, and (2) spatially or geographically oriented [2]. In a sense, a geographical information system (GIS) is simply an object-oriented system with an added attribute -- geographic location. However, the added complexity in the storage, retrieval and manipulation of these data leads us to distinguish spatial systems from less complex object-oriented systems.

### 3.2.1 DEFINITION OF GEOGRAPHICAL INFORMATION SYSTEMS

Appendix I provides a detailed definition of geographical information systems. The following summary is provided for background.

Spatial data occur in any of three basic forms: (1) point source data, e.g., climatological data, (2) linear data, e.g., a river network, and (3) areal data, e.g., Landsat data.

Data incorporated in a computer-based information system evolves through three stages: (1) raw data structure: the form in which data are acquired, e.g., soil map, (2) computer-external structure: the computer-compatible format in which the data reside on hardware storage devices, e.g., tapes and disks, and (3) computer-internal structure: the format in which the data reside in computer core storage while being actively processed. Computer external and internal formats are of particular interest. Figure 2 illustrates that both internal and external structures can be viewed differently, depending on at what level the data are being interfaced within the system. Whereas a user may be interested only in the attributes of a data structure, a program processing the data is interested in the format structure of the data; at a lower level, the operating system is interested in the data's physical location on storage devices. Our interest in this report is in data structures as viewed by a system user and processing program, not the computer's operating system.

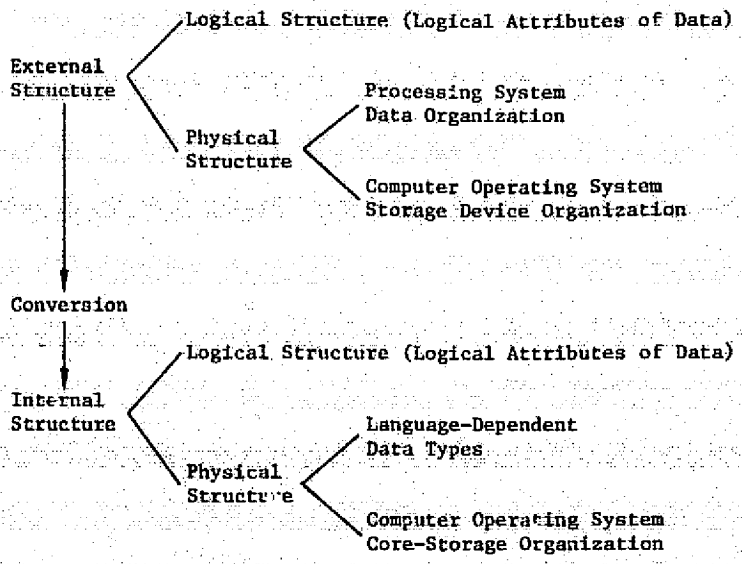



FIGURE 2. COMPUTER INTERNAL AND EXTERNAL STORAGE STRUCTURES

Analysis of existing information systems incorporating spatial data reveals that the storage and processing characteristics of these systems can be categorized into two basic types: (1) irregularly encoded systems and (2) regularly encoded systems.

Irregularly encoded systems describe spatial features by using nodes and connecting line segments. Point form data are described using only nodes, linear form data are described by nodes and connecting line segments, and areal data are described by sets of nodes and line segments which form closed regions, i.e., polygons.

Regularly encoded data utilize a grid of cells. A cell is a special form of areal data. Cells are rectangular polygons or picture elements (pixels) which are generally stored and processed as a contiguous array. Point source data, when stored as cells, require the fabrication of a null cell to indicate pixels for which no data are available. Linear data also require the null cell concept, while areal data are by nature contiguous and easily adapted to cell structuring. Null cells do not necessarily have to appear in the data set, since they can be accounted for by special cellular data structuring.

### 3.2.2 EXAMPLE GEOGRAPHICAL INFORMATION SYSTEMS

Appendix II provides a partial listing of computerized geographical information systems. Appendix III provides an annotated bibliography of computer-based information system literature. Several information systems will be briefly described in this section simply to provide a flavor of the variety of approaches that have been taken. There is no universally agreed upon 'best' approach and, in fact, the various schools of thought are often diametrically opposed.

#### The Canada Geographic Information System (CGIS) [3]

CGIS is a very large-scale system developed by IBM. It currently is in operation at the Department of the Environment, Government of Canada, Ottawa, where it facilitates the use of data gathered by the Canada Land Inventory (CLI) and does not as of yet incorporate remotely

sensed data. CGIS utilizes the irregular encoding approach by specifying polygons to describe features of interest. High density maps are scanned by a digitizing drum scanner and automatically encoded as polygons on digital tapes. Low density polygon data are input using a conventional digitizer table, as are point source data. Descriptive data are manually encoded. CGIS designers chose to collect and map data in polygon rather than grid format in order to maintain exact boundary data and flexibility in data manipulation. It was felt that polygon data could be automatically converted to any size of grid cell summary at any time. Such conversion is in fact employed in certain subsequent forms of spatial analysis. Some acknowledge CGIS as the first and most sophisticated geographical information system [4]. As a pioneering effort, however, others have noted serious deficiencies [5]. CGIS provides no user interface language. This, in conjunction with computational complexity CGIS encounters in processing polygonal data, is largely the reason for the latter point of view. The task of incorporating remotely sensed data in the CGIS will certainly not be straightforward.

Image-Based Information System (IBIS) [6]

IBIS is a system currently in use and under continuing development at the Jet Propulsion Laboratory in Pasadena, California. IBIS makes use of digital image processing techniques to interface existing geocoded data sets and information systems with thematic maps and remotely sensed imagery. Spatial data layers are stored in separate files and can be formatted as either variable-resolution-grid or polygonal structures. The processing of data is carried out in raster format, that is, on a vector of grid cells. Hence, when data are accessed, they are converted into a cell structure, with replications where necessary to accommodate the cell size of the data layer with the finest resolution. Processing capabilities consist of mapping, overlay, classification, and determination of simple sums and averages. IBIS is a

serious effort to supplement a spatial data system comprised of socio-economic and topographic information with remotely sensed imagery. IBIS successfully addresses the mechanics of merging, i.e., overlaying data layers of different external structure into a cellular, or raster, internal structure for processing purposes, but the processing aspects of IBIS, relative to remotely sensed data, are not as yet fully developed.

Earth Resources Interactive Processing System (ERIPS) [7]

ERIPS was developed by IBM for use at NASA/JSC in large-scale processing of remotely sensed data for crop inventory systems. It is currently used as the computer processing system for the Large Area Crop Inventory Experiment (LACIE). ERIPS was developed to interface components of a processing system for remotely sensed data. The processing of spatial data layers from sources other than remote sensing were not considered in the system design. ERIPS under LACIE does address, however, the management problems and processing complexity associated with remotely sensed data. A well defined sampling strategy has been implemented, multitemporal acquisitions of data samples are managed, large volumes of data are stored in a manner to expedite retrieval and analysis, and data update standards have been well defined. Currently over 600 agricultural data segments with an average of four multitemporal Landsat scanner data acquisitions, each of an area of thirty square miles, are being processed to estimate wheat production on a large scale. ERIPS can manage as many as 4800 sites with four acquisitions each. However, whereas specific problems associated with this specific remote sensing question have been carefully addressed, the system has been rigidly defined and lacks the flexibility to be considered a generalized geographical information system.

The Minnesota Land Management Information System (MLMIS) [8]

MLMIS is a state-wide land information system developed by the Minnesota State Planning Agency and the Center for Urban and Regional Affairs of the University of Minnesota. The MLMIS data base contains

state-wide data for nine categories of predominant land use (1969) and seven categories of water orientation. These are manually encoded into forty-acre (16-hectare) cells. In addition to descriptive information, each cell or 'forty' is labeled by the latitude and longitude of the centroid of the forty, a county identification number, and the minor civil division number of the U.S. Bureau of the Census within which the forty falls. MLMIS is very typical of the growing concern of state and county agencies to make resource information available to decision makers in state and local governments in a timely manner. MLMIS faces a problem in updating predominant land use and water information. The need for making use of remotely sensed data, particularly from Landsat, is apparent. The data base demands a rigid structure designed to meet the expected needs in a cost efficient manner. Landsat data structure is not directly compatible. Hence, like other geographic information systems seeking to incorporate remotely sensed data, MLMIS is faced with new and unexpected design and processing considerations.

### 3.3 GENERALIZED GEOGRAPHICAL INFORMATION SYSTEM

The previous discussion, along with Appendix I, was provided to familiarize the reader with terminology as well as with several operational information systems. It is hypothesized that, as a norm, geographical information systems are designed with a particular application and data base in mind. Systems become bounded by that application and data, and thus are not easily adaptable to the incorporation of new processing strategies or data types. It is further hypothesized that this need not be the case. A GIS can be designed independent of any specific application or data base. Hence, this section is provided to address the subject of GIS design on a more analytic basis. That is to say, given the wide variety of approaches characterizing existing information systems, can a generalized approach be constructed that permits versatility enjoyed by no one existing system?

The advantage of such a generalized approach is obvious. The system would be compatible to the introduction of new applications, new data sets and new processing strategies. The incorporation of remotely sensed and associated data and their processing requirements would then be straightforward.

The interesting feature of a generalized approach is that, once the components of a geographical information system are well understood, the dynamics of a generalized system are quite simple. Much of the philosophical groundwork for the design of a generalized system was set in 1971 by the CODASYL Data Base Task Group [9]. This group of computer users set forth a strategy for a generalized Data Base Management System (DBMS) that addressed the special characteristics of object-oriented data. The contention is made that a spatially oriented system can be managed through a general framework that adheres to the philosophical structure of the CODASYL DBMS. The following sections illustrate such a generalized approach to geographical information system design.

### 3.3.1 OVERALL GIS MODEL

The DBMS approach is not generally encountered in existing geographical information systems. It is an approach, however, that lends itself to such systems. Geographical information systems require the management of large quantities of data. These data characteristically are structured as cells or polygons. However, the user's view of the data, for a given application, may require a "look" at the data that is inconsistent with its structure. A grid structure of specific element resolution size may be required, yet the data may be stored as polygons or as a grid of different resolution. A typical solution to this problem is to fix the permitted view of data and so structure it. This is at the expense of generality. A second approach is to duplicate data in different formats. This is at the expense of storage efficiency.

A DBMS, however, separates the data structuring, storage, and retrieval from the data processing. The user can view the data independently of its external storage because responsibility for data management is relegated to an autonomous program, the data base manager. This frees both user and application programs from the responsibility of data management.

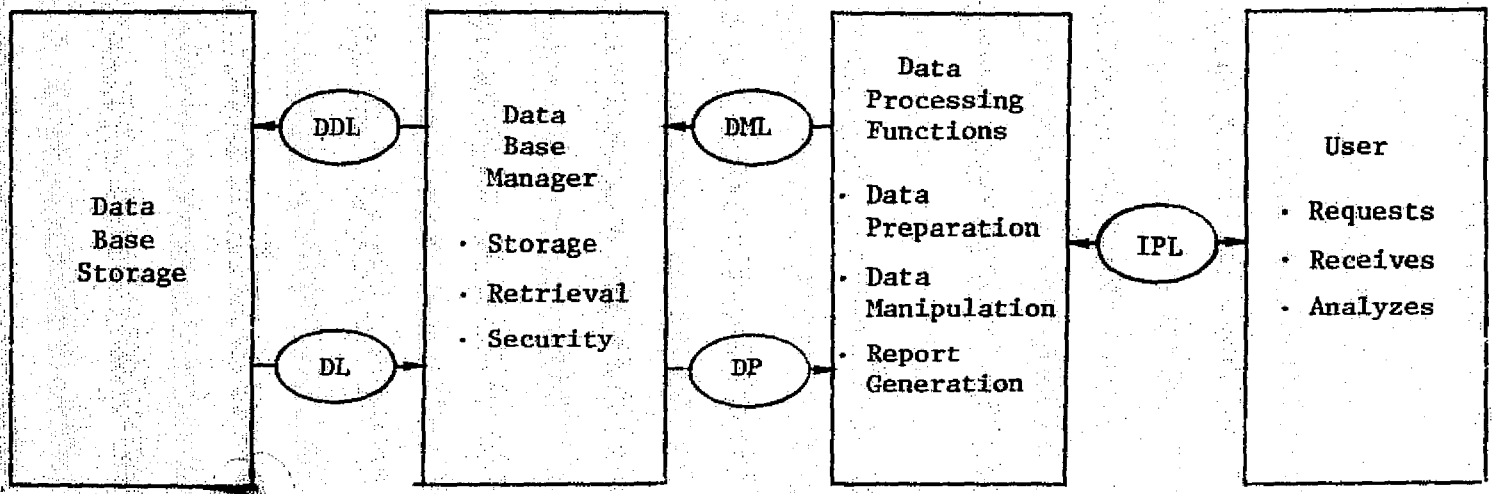
The principles guiding the design of a generalized geographical information system include: (1) provide the GIS user a query environment that is independent of the details of the system structure, (2) provide application programs an environment that is independent of the external structures of the data base, and (3) provide the user the facility to permit multiple views of the data, independent of the external storage characteristics of the data.

Figure 3 illustrates a generalized information system model that interfaces various components in a manner that enables user/system, program/data, and user/data independence. Structurally, the system is a Data Base Management System with the addition of a user/system interface through a batch operation or an Interactive Processing Language (IPL). Elements of the diagram will be examined more closely in the following sections.

### 3.3.2 BASIC ELEMENTS OF THE SYSTEM MODEL

The generalized computer-based geographical information system depicted in Figure 3 is characterized by four basic structural components and a number of languages interfacing between those components.

Each basic component (data base, data base manager, data processor, and system user) sequentially perform specific roles in the dynamic operation of the information system. Communication between the components is carried out through languages designed to satisfy a specific need for data or information upon request from an adjoining component.




Notes: DDL = Data Definition Language
DML = Data Manipulation Language
IPL = Interactive Processing Language
DP = Data Primitive
DL = Data Layer

FIGURE 3. SCHEMATIC DIAGRAM OF A GENERALIZED COMPUTER-BASED GEOGRAPHICAL INFORMATION SYSTEM

Data base storage for a geographical information system would be designed to support spatial data in various structural formats.

The data base manager (DBM) is a set of software programs that interfaces between a user or program request for data and the physical representation of those data in external storage. The DBM catalogs not only the data sets that comprise the data base but also the permissible methods of retrieval. This catalog or data dictionary is defined through the data definition language (DDL). The DBM processes a request for data and invokes format service routines (FSR's) designed to convert data from external representation to an internal format designated as the data primitive (DP). This DP is in turn processed by the application programs.

Five basic data processing functions are required: (1) a regularly encoded data processor, (2) an irregularly encoded data processor, (3) object-oriented data processors, (4) a statistics processor, and (5) data display mechanisms, e.g., graphics. The spatial processing software would support both parallel and contextual processing functions as described in Appendix I. This component of the GIS model provides the mechanisms to interpret the data in selected ways so that knowledge about that data, i.e., information, can be extracted for the user. This distinguishes the system as an information system as opposed to a data base management system. A significant point to be stressed is that the data primitive is the internal representation of the data set to be processed. This "active data set" is prepared by the data base manager to satisfy the current view of the data.

The processing system and user interface through batch-mode operation or interactively through an interactive processing language (IPL). This is to say that the typical user may not be a programmer, hence the user is supplied with a very high-level language interface to expedite interaction with the processing system. The supplied vocabulary would depend, of course, on the processing functions available in the system. With such a language interface available, the user is basically free

of the mechanics of the information system providing answers to queries that might be posed. To the user the system is a "black box" that serves a function.

This generalized model provides the background within which recommendation for the incorporation of remotely sensed data into a forest and rangeland information system are set forth. Each structural component will be discussed in further detail as recommendations are made.

## 3.4 REQUIREMENTS FOR FORESTRY AND RANGELAND INFORMATION SYSTEM

The United States Forest Service has identified a growing need for the development of a consolidated information system serving the USFS throughout the United States. This is reflected in the formation of both a Data Management Staff and special groups to analyze existing Forest Service information systems for the purpose of designing a spatially and statistically oriented information system.

The Management Science Staff of the USFS has carried out a detailed analysis of many existing USFS systems [10]. Another task group is approaching the development of a USFS information system in two phases [11]: (1) incorporate the best of existing USFS systems into a single operational data base management and processing system and (2) study the potential needs of the USFS in the next decade in order to establish a set of recommendations with regard to an advanced information system, studying the role of an integrated data base system.

The most pressing needs are simply summarized in two points. First of all, there is a need to provide accurate, up-to-date information useful to USFS personnel and exchange information between personnel in a more timely and efficient manner than is currently possible. This generally leads to 'computerization' of the information extraction and exchange mechanisms. Second, for purposes of efficiency and cost-effectiveness, this automation must be carried out in a manner that minimizes duplication of effort. The support of a single, integrated global system may be more efficient than the maintenance and interfaces required in using a number of distinct locally maintained systems.

### 3.4.1 ADDRESSING USFS INFORMATION NEEDS WITH REMOTE SENSOR DATA

The concern of this document is to address the requirements for the incorporation of remotely sensed data in a USFS information system. The need for timely information is an influencing factor in the incorporation of remotely sensed data. Classic techniques of land use classification, though accurate, are bounded by the end-to-end time required in producing a product. Progress in the field of remote sensing has shown the feasibility of using remotely sensed data for accurate land use inventories, among other applications. The update capacity of a land use classification system incorporating remotely sensed data is greatly enhanced by satellite technology which permits the collection of information at regular intervals.

The incorporation of remotely sensed data into a data base is roughly analogous to the incorporation of any other layer of spatial data. However, two significant differences arise. First, the quantity of data available through remote sensing sources dwarfs other spatial data sources. Secondly, the process of extracting information from remotely sensed data is highly specialized, requiring that a number of sophisticated processing techniques be incorporated into the information system. The incorporation of these data, then, will have ramifications affecting the total information system. Recommendations must be made, therefore, not only in light of current USFS information needs but also expected system design characteristics.

### 3.4.2 USFS GEOGRAPHICAL INFORMATION SYSTEM NEEDS

Particular attributes of a USFS geographical information system have been identified [11]. A few of these attributes will be presented in this section. The important point to note is that whereas these attributes are derived from a broad base of information requirements acquired through experience, they could easily be managed within the context of the generalized geographical information system.

In the long range future, the GIS could be managed through a distributed computer system, with terminals and minicomputers at a regional level interfacing with a centrally maintained software facility and data base that is of general applicability. Software and data of less general applicability would be maintained at a local level.

The information system would require support of object-oriented data, especially of a statistical nature, and geographically oriented data, as well as interfaces between the two data types.

Other requirements include:

1. Maintenance of computer-compatible and other geographical data banks, e.g., thematic maps,
2. Manipulation of geographical data layers in regular or grid formats,
3. Manipulation of all irregular data forms, i.e., polygonal, linear, and point source,
4. Cartographic display of data layers, including overlay processing of data, and
5. Manipulation and display of statistical features of the information extracted from the data.

Adoption of the generalized GIS approach is reasonable in light of the fact that the management of these requirements is facilitated in the generalized GIS environment. Moreso, the incorporation of remotely sensed data could be more easily carried out since processing and data management facilities would be designed and defined in a broad enough sense so as not to be limited in scope.

### 3.4.3 COMMERCIALLY AVAILABLE DBMS

The Management Sciences Staff has documented the implementation of the Generalized Information Management (GIM) data base management system [10]. GIM was developed and marketed by TRW and has been

available to the Forest Service since 1971. CIM, like many other available data base management systems [12,13], is useful for the storage and management of object-oriented data. To the best of our knowledge, however, commercially available DBMS's do not currently provide an environment appropriate for the management of spatial data, especially remotely sensed data. Reasons for this include:

1. Inadequate storage support for extensive spatial systems,
2. Storage and retrieval of data by spatial attributes are not supported since retrieval is nominal in nature,
3. The spatial nature of the data require specialized storage structures, including polygonal and cell encoding which are not supported by available DBMS's, and
4. Functions required to interface a user request for spatial data in a particular form, e.g., in a given resolution size, are not provided.

Hence the employment of a commercially available DBMS for spatial data has to be restricted to small data sets whose layers have common external storage characteristics. Even there, the mechanics of coordinate retrieval must be addressed since provision is not made for this feature in any available DBMS [14].

### 3.5 SPECIFIC RECOMMENDATIONS FOR THE INCORPORATION OF REMOTELY SENSED DATA IN A USFS GEOGRAPHICAL INFORMATION SYSTEM

To this point we have attempted to construct a philosophical framework within which specific recommendations can be made with regard to the incorporation of remotely sensed data in a USFS geographical information system. We emphasize that:

1. A user/system, user/data and program/data independent environment is most desirable,
2. Remotely sensed and associated data can be viewed as different layers of spatial data, inherently grid, linear, or point source in nature,

3. The incorporation of these data in a GIS is in principal no different than the incorporation of any other spatially oriented layer of data,

4. The mechanics of the management of a great volume of data must be addressed, possibly through a careful segmentation of the data,

5. The manipulation of these data requires special processing functions. These could enhance the overall capability for analysis and display of other data layers,

6. The processing of these data results in the need to manipulate non-spatial data, i.e., object-oriented data types, and

7. The display and analysis of these data require interfaces between data layers of both irregular structure and regular structure with varying resolution sizes.

The generalized geographical information system concept introduced earlier serves as the foundation of the recommendations to be set forth. Certain aspects of this concept, specifically the data structure, data management, and data processing aspects, are more fully detailed in this section, as they apply to specific recommendations. It is important to stress that, whereas the generalized approach is well within the capability of state-of-the-art computer data processing technology, the concept has yet to be made commercially available in hardware or software form. Hence, developmental work would be required.

Four categories of recommendations are presented next, regarding: (1) data structural characteristics, (2) data processing environment, 3) data requirements, and (4) processing requirements. Certain topics are discussed in further detail in Appendix IV.

### 3.5.1 EXTERNAL DATA STRUCTURE CHARACTERISTICS

Figure 2 in Section 3.2 illustrated two basic data storage structures for computer-based information systems. The internal structure

pertains to the formatting of data in computer core storage while they are being processed by an application program. Internal structure characteristics for the processing of remotely sensed data are presented in Section 3.5.2. External structures pertain to the formatting of data on hardware storage devices while inactive. This section discusses recommendations for external storage structures for a GIS that utilizes remotely sensed data. Further discussion is carried on in Appendix IV.1.

To remain within the context of a generalized GIS utilizing remotely sensed data, three items are specifically recommended:

1. That external storage support layers of both object-oriented and spatial data (both irregular and regular encoding),
2. That a data dictionary describe the location, structure, and nature of each layer of data, and
3. That all storage and retrieval of data be conducted through a data base manager software subsystem.

External structures of digital data have both physical and logical attributes. Physical attributes include a higher level organization of the data as viewed by the processing system, as well as a lower level organization as viewed by the computer operating system. Our discussion of external structure characteristics concentrates on logical attributes and higher level physical data structuring.

Logical attributes of data incorporated in a GIS refer to the inherent characteristics of each layer of the data base. Remotely sensed data, in particular Landsat data, have certain characteristics that require specific attention, especially when interfacing or overlaying these data with other spatial data layers. The basic logical attribute that is of concern here is the geometry of these data. The geometry of remotely sensed data is a function of the satellite orbit or aircraft flight line and altitude and sensor field of view.

Landsat data currently are collected at an orientation skewed by a few degrees from a North-South orientation in a projection implicit to the satellite. It is advisable, in order to expedite the process of data layer overlay, to geometrically correct these data to a map projection that would be standardized between various layers of spatial data. Also, sample density must be considered. A variety of resolution sizes may be encountered. Landsat data resolution currently is 57x79 m per cell. The Thematic Mapper (Landsat D) is expected to have a 30x30 m resolution size.

The incorporation of remotely sensed data in a GIS imposes the requirement to find a means to accommodate the special geometries encountered. Three approaches are discussed.

The first, and most general approach, would be to reformat the data to suit the current need through the DBM upon retrieval. Hence, Landsat data would remain in raw form until a specific user request for a region of data, in a specific resolution size, and geometrically corrected to some specified map projection, was encountered. The data would be resampled and corrected upon retrieval for the current use.

The second approach would standardize the geometry of all layers in the data base -- a particular projection and a specific cell size for regularly encoded data. Two problems arise. The layers may be geometrically dissimilar so a single cell size may not be practical. In addition, this approach would restrict the GIS user to a particular view of the data, whereas, the user should be able to request these, as well as other layers of data, in a resolution size or map projection that may be different from that characteristic of the original data.

A third approach would be to make certain assumptions with regard to the most likely form of request and convert the data into that format upon loading it into the data base. This approach is a compromise, more restrictive than the first but more general than the second. One could pre-define a basic pixel size, like one meter square, and insist

that each data layer be resolved in storage to some power of two (e.g., 1, 2, 4, 8, etc.) of that size. The advantage of this approach is that the resampling scheme employed upon retrieval would be simplified to an aggregation or splitting of the data cells. The disadvantages being that the user would still be restricted to multiples of the basic pixel size. Resampling of data not originally available in this form would be carried out upon loading. Hence, resampling is not eliminated, simply conducted at loading time as opposed to retrieval time.

It is recommended that one of the more general approaches or a combination of them be employed. There is a tradeoff in cost in these two approaches. One would only correct subsets of data upon use, whereas the other would standardize and correct all the data initially.

Remotely sensed data are commonly available in computer-compatible form structured in grid format. Hence, the physical formatting requirements associated with remotely sensed data are for the most part predefined. However, various other layers of data may be required for the effective use of remotely sensed data. Let us next consider formatting structures as they apply to data in general, both regularly and irregularly structured.

Considerations for the external physical structuring of data in a GIS must be carried out on two levels. The first concerns the technique employed to manage and store data accumulated over large geographical regions, that is, how the data are geographically segmented. The second concerns the structural organization or encoding used for these data, e.g., grid or polygon.

The potential volume of satellite data requires that a careful segmentation strategy be employed. Whereas Landsat is most commonly available in 100-nautical-mile-square frames, this is rarely a convenient storage unit. A feasible strategy may be patterned after that employed in the Land Inventory Mapping (LIM) System of Forest Region 2 [10]. LIM divides a forest into a matrix of quads. Each quad is of maximum

dimension of 210 grid cell rows and 128 grid cell columns. This is not only convenient for storage and retrieval of data, but permits ready joining of data should a request for a broader region be encountered.

Let us now move our discussion to methods of encoding regularly structured data. It was mentioned that satellite data are stored in sequential format. Another form of encoding that is recommended is compact sequential encoding. This technique provides not only particular advantages of cell encoding, but also a substantial savings in the storage requirement. Compact sequential encoding requires that data values be entered cell by cell. However, adjacent repeating cell values are not stored repeatedly, but an associated length attribute is incremented instead. This form of encoding may not necessarily be applicable to remotely sensed data, but is recommended whenever regular encoding is preferred, and data features do not vary rapidly from cell to cell.

Irregular encoding would certainly not be likely for remotely sensed data. However, certain associated data layers, such as training and test areas, as well as many types of ancillary data, may be encoded more naturally in this format. It is suggested that, for these data, the feasibility of employing the chain/node encoding technique be investigated. This technique, developed by the Harvard Laboratory for Computer Graphics, is described in detail in Appendix IV.1. It enables the storage of point source, linear, and areal data in one topological data structure. In addition, attributes of the data, like the area of the polygon, can be managed within the same structure.

### 3.5.2 REMOTE SENSING DATA PROCESSING ENVIRONMENT

This section summarizes a recommended data processing environment that is more specifically detailed in Appendix IV.2. The recommended approach, though developed for the processing and analysis of remotely sensed data, is of more general character. It is designed in a manner that will accommodate either object or spatial data, or both. The prototype system for remotely sensed data is currently under development at ERIM.

Figure 4 is a simplified schematic diagram of the recommended data processing environment for a geographical information system.

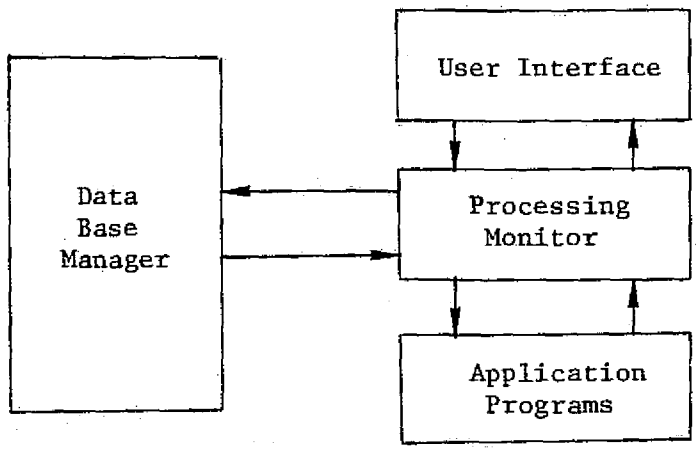



FIGURE 4. GIS DATA PROCESSING ENVIRONMENT

The most prominent feature of this design is the interjection of a processing monitor between the data base manager, the application programs, and the user. The advantage of incorporating a processing monitor is that the DBM, user, and application programs can act independently of one another. The processing monitor insures that all input and output interfaces between these three operators are compatible. The one requirement is that each operator act within certain globally accepted conditions.

The user is responsible for providing adequate information to enable the system to respond to a query. Specifically the user must provide information with regard to: (1) the specific layers of data to be manipulated, (2) the spatial region of data to be processed, (3) the specific process to be carried out, and (4) any appropriate parameters describing the process.

The application programs must be designed for purposes of responding to permitted user queries. To insure external data structure independence, an understanding between the DBM and the programs as to permitted internal structures is required. A data-primitive concept is adopted in order to accomplish this. A data primitive is a globally recognized internal storage structure entity, like a polygon or a raster of cells. Each application program would declare the primitive(s) it requires. The processing monitor would insure that the data requested by the user would be made available to the application programs through the DBM if the requests were compatible. The DBM would convert the data sets required by the user from external format to the internal format requested by the application programs via data primitive declarations (see Apx. IV.2).

The general character of this approach as a processing environment for a GIS that utilizes remotely sensed data stems from two features. First, data primitives can take on any form, object or spatial, regular or irregular. Secondly, the processing monitor acts as a knowledgeable interface between the various operators to insure that communication between them is understood in the proper context. Hence, any form of processing can be carried out, as long as the application programs and data required are available.

### 3.5.3 DATA REQUIREMENTS

Developments in areas of research in remote sensing have identified a need for the joint use of remotely sensed data with layers of data derived from sources other than remote sensors. For example, Volume 1 of this report [1] examined the utility of incorporating remotely sensed data with ancillary data for inferring forest understory conditions. Section 4 of this report discusses in detail the value of incorporating topographic information in the analysis of Landsat data acquired in mountainous terrain. Studies of water balance and evapotransformation processes have indicated a need not only for

Landsat data, but for Metsat data, climatological, geologic, and topographic information as well [15]. Researchers acknowledge remotely sensed data as a valuable source of information, but one that complements or is complemented by other data. It is this requirement for the joint use of data that led initially to our consideration of remotely sensed data as a layer of data in a generalized geographical information system.

Data requirements for incorporating remotely sensed data in a GIS span more than those data derived from remote sensing sources alone. Some of these data layers may be already available in the GIS. Others may not be. Those already available, however, may need to be adapted to remote sensor applications. Appendix IV.3 describes a number of data types that might enhance remotely sensed data information extraction capabilities. The most basic requirements for utilizing remotely sensed data in a USFS information system includes incorporating Landsat digital data as well as associated parameters of acquisition, e.g., sun angle and date. Remotely sensed data from other sources, like aircraft radar or MSS scanner data, may be necessary where finer resolution or more extensive spectral information is required. In addition, spectral information acquired in a controlled setting, e.g., field measurement data, is considered an invaluable aid in the analysis of remotely sensed data. Topographic information is considered critical, especially when terrain features vary rapidly. This is due to the sensitivity of the Landsat signal to the resultant irregular solar illumination. The joint use of Landsat and topographic information to be of use imposes the requirement that the topographic information be sampled at the same resolution as Landsat data, or finer.

Elements of the data base essential to the extraction of information from remotely sensed data are not derived solely from external sources, but also from within the GIS. Certain components of the data base that are felt by us to serve as important agents in this process

of information extraction include remotely sensed data that have been preprocessed to correct for external effects, e.g., haze, as well as data that might have been compressed in some manner. In addition, statistics generated from the data are important to maintain for reference and analysis.

### 3.5.4 DATA PROCESSING REQUIREMENTS

The incorporation of remotely sensed data in a USFS information system will require the incorporation of a set of sophisticated processing functions to enable information extraction and analysis of data from remote sensing sources. Many of these functions are described in Appendix IV.4.

For the purposes of this section, let us simply categorize those functions and indicate certain applications that may effectively utilize remotely sensed data in a forest and rangeland context.

The prerequisite categories for the processing of remotely sensed data include:

1. Corrective Preprocessing
2. Feature Extraction
3. Data Classification
4. Information Display

Corrective preprocessing of remotely sensed data is carried out to eliminate or reduce systematic variations in the data due to external effects such as the sun angle or haze present at time of acquisition. Screening the data for the presence of clouds, cloud shadows, and defective data is also important.

Feature extraction refers to the process of detecting the characteristics of the remotely sensed data signals that enables one to distinguish the classes or information present. Statistical, empirical, and analytical techniques have been employed in feature extraction.

Data classification pertains to the categorization of features derived from remotely sensed data. This may take the form of associating the data elements to specific classes present in the scene, or more generally be thought of as the identification of features.

Information display is the component that acts as interface between the data analyst or user and the algorithms processing the data. The specific algorithms employed are determined, of course, by the application for which remotely sensed data are utilized. Most prominent applications of these data include land use classification and inventory, and change detection. Research in the field of remote sensing is not only improving the effectiveness of these types of analyses, but also broadening the scope of application as well.

# 4

# INFORMATION EXTRACTION TECHNIQUES

Multispectral scanner (MSS) data from a single time period, alone, may not permit sufficiently accurate classification and extraction of information from forest and rangeland scenes. For example, spectral similarities of the scene classes or spectral variability within individual classes from location to location and time to time may preclude adequate performance. Ancillary geographic data or MSS data from additional time periods potentially can be used to improve the quality and quantity of information extracted, provided they can be digitized and associated with the appropriate resolution elements of the original MSS data. Types of ancillary geographic data include terrain elevation, slope, and aspect data, soil type data, geologic data, and land use data, among others.

Given the association or registration of ancillary geographic data with MSS data, it remains to develop effective techniques to exploit the potential of the combined information. One recent investigation [15] gave indications that the use of terrain elevation data with Landsat MSS data could improve classification accuracy in a forestry application. However, only one data set was analyzed so additional verification of improvement is needed. There also is a longer-range need to determine the most effective ways of using ancillary data of this and other types for improving information extraction capabilities from remotely sensed data.

## 4.1 APPROACH

To develop and test various processing and information extraction techniques, we assembled a data set that consisted of Landsat data, acquired on 15 August 1973 over a test site in Grand County, Colorado, plus ancillary terrain data (See Section 4.1.1). This site is one of those being investigated as part of the Forestry Applications

Program's Ten Ecosystem's Study [16]. We concentrated our analysis on the Fraser Experimental Forest, a subset of the site which lies within both Grand County and the Arapaho National Forest; see Figure 5.

The Fraser Forest is a thirty-square-mile experimental watershed, heavily wooded with spruce, fir, and Lodgepole pine. The mountainous terrain varies in elevation from 2400 to 4000 meters (8000 to 13,000 feet) with slopes of up to 45°. The topography of this region lends itself to an investigation of the effects of terrain slope and aspect on one's ability to accurately classify forest stands with approaches using Landsat data both alone and in combination with ancillary data. The terrain slope and aspect have pronounced effects on Landsat signals.

A number of processing and analysis techniques were applied to these data (See Section 4.1.2).

### 4.1.1 DATA SET DESCRIPTION

Data preparation activities preceded our processing and analysis. The Landsat data were geometrically corrected, rotated, and resampled, using a nearest-neighbor technique, in a UTM projection to a scale of 1:24,000 on a line printer display.

Digital terrain elevation data were purchased from the National Cartographic Information Center (NCIC) in Reston, Virginia [17]. These data, previously digitized from 1:250,000-scale topographic quadrangle maps with 200-foot elevation contour intervals, are reportedly accurate to within 100 feet, although interpolated to the nearest foot. A program was developed and employed to convert these data to a format compatible with multispectral data to permit registration with Landsat data.

Three additional types of data were derived from these elevation data -- slope, aspect, and a relative solar insolation factor. The nine-element rules shown in Figure 6 were used to compute slope and aspect for each pixel, using values from its eight neighbors. These formulas had been previously developed for U.S. Forest Service uses [8].

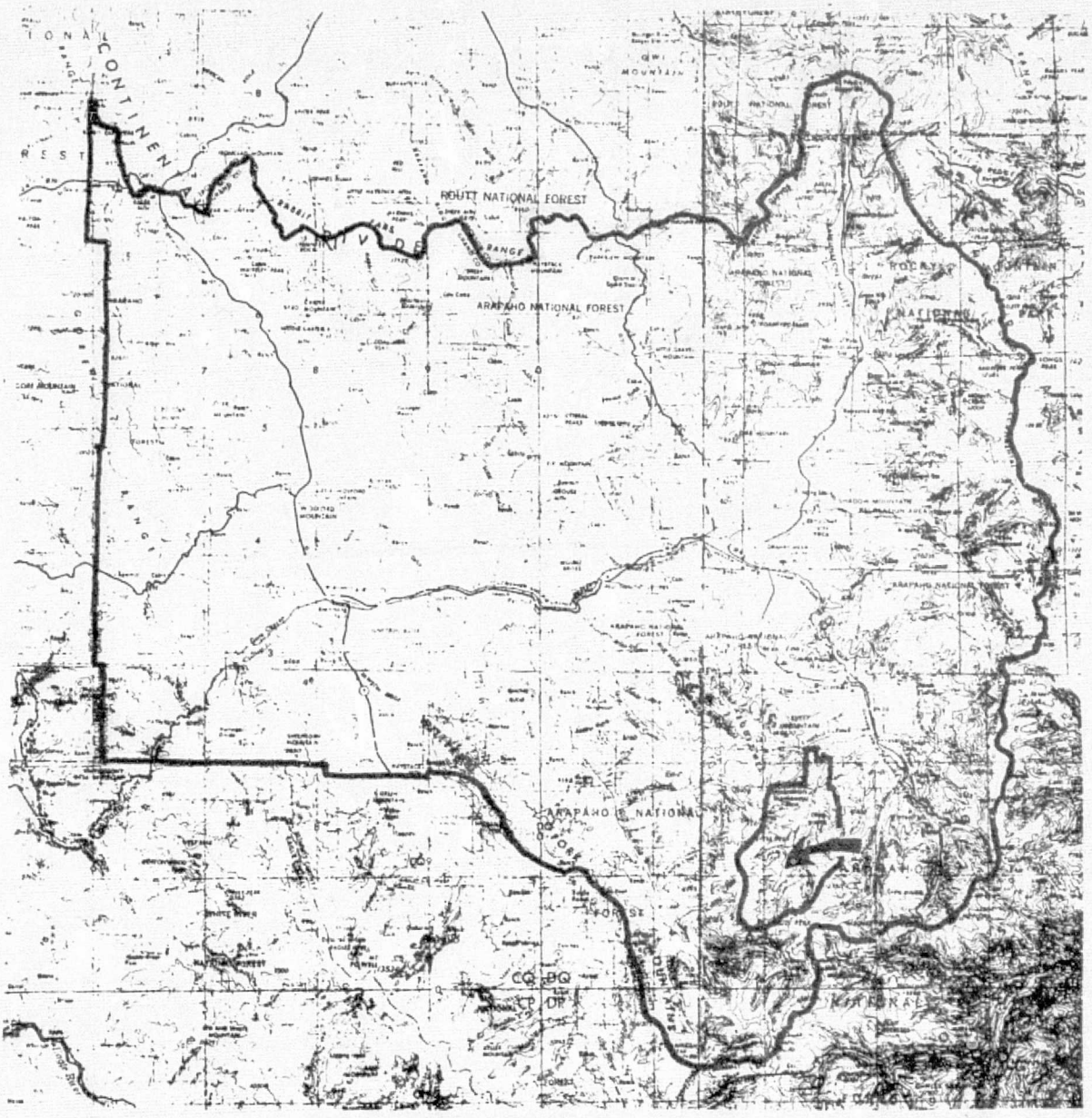
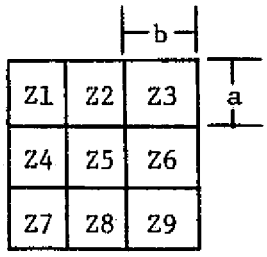



FIGURE 5. LOCATION OF FRASER EXPERIMENTAL FOREST WITHIN GRAND COUNTY, COLORADO




$$B1 = \frac{(Z1 + Z2 + Z3) - (Z7 + Z8 + Z9)}{6a}$$

$$B2 = \frac{(Z3 + Z6 + Z9) - (Z1 + Z4 + Z7)}{6b}$$

$$\text{Slope} = \tan^{-1} \sqrt{B1^2 + B2^2}$$

$$\text{Aspect} = \tan^{-1} \left(\frac{B2}{B1}\right) \quad \text{(converted to azimuth, clockwise from North)}$$

FIGURE 6. FORMULAS FOR TERRAIN SLOPE AND ASPECT CALCULATIONS

We then defined the relative solar insolation factor at a point on the ground as the following function of sun position:

$$F_{RI} = \frac{\bar{n} \cdot \bar{s}}{\cos \phi_s} \qquad (1)$$

where

$F_{RI}$ is the relative insolation factor

$\bar{n}$ is the surface normal unit vector

$\bar{s}$ is the sun direction unit vector

$\phi_s$ is the sun zenith angle.

This factor, relative to a horizontal diffuse surface, combines the effects of slope and aspect, as well as solar zenith and azimuth angles, into a single factor. In terms of slope angle $\beta$ and relative aspect $\psi$, between the compass bearing of the sun and the horizontal projection of $\bar{n}$, we have:

$$F_{RI} = \tan \phi_s \sin \beta \cos \psi + \cos \beta \qquad (2)$$

The relative insolation of a horizontal surface would equal one. A maximum is reached when the surface faces the sun directly and goes to zero as it approaches a grazing angle with respect to the sun. The algorithm developed does not account for possible shadowing by nearby terrain relief at low sun angles, a capability that should be included in a comprehensive program.

A merged data set was prepared, to enable the mapping and analysis of these data using the same tools available for analysis of Landsat data alone. This ten-channel data set included:

- 4 Landsat channels
- 2 elevation data channels (to encode all levels)
- terrain slope
- terrain aspect
- relative insolation factor
- test site delineation

In addition to this set of Landsat data and associated ancillary data, we made use of simulated reflectance data for pine forest canopies. The simulated data are described in detail in Volume I [1]. In brief, they were generated using the Suits bidirectional reflectance model for vegetation canopies with parameter selections representing several different canopy densities, understory conditions, and terrain slopes and aspects relative to the sun.

4.1.2 PROCESSING TECHNIQUE DEVELOPMENT AND TESTING

The first step in analyzing these data was to produce and examine maps of the various types of data. Next, we studied correlations between Landsat signals and the various ancillary variables. In

addition, correlation analyses were performed on the simulated forest canopy reflectances.

Next, we examined the effects of incomplete training on computer performance in classifying the Landsat data, training being incomplete in the sense that the full range of variation in ancillary terrain variables was not represented in the training data set.

The next step was to investigate different preprocessing algorithms based on ancillary terrain data, to reduce the variability in Landsat signals and improve classification performance. Algorithms were derived both empirically from Landsat data and from analysis of the reflectance model calculations. The studies described up to this point are related to ancillary terrain variables and are described in Section 4.2.

A study also was made of the applicability of selected LACIE-oriented* data processing techniques and algorithms (See Section 4.3). One technique examined was the Tasselled-Cap Transformation which produces linear combinations of the original Landsat data. The first two combinations define a plane that contains the vast majority of variance in the Landsat signals and also have convenient physical interpretations. Another related transformation, a polar green-angle/brightness-radius transformation, was also examined.

Also, unsupervised clustering techniques were investigated -- both pure spectral clustering and joint spectral/spatial clustering. The spectral/spatial technique investigated was the ERIM BLOB algorithm, with derivatives therefrom which were developed as part of this study.

Finally, clustering analyses were conducted on data that had been preprocessed using one of the algorithms developed earlier. These analyses are described in Section 4.4.

Conclusions and recommendations are presented in Section 4.5.

*LACIE is the Large Area Crop Inventory Experiment [19].

## 4.2 INVESTIGATION OF TECHNIQUES USING ANCILLARY TERRAIN DATA

Because of both the pronounced effects of terrain slope and aspect that can be observed in Landsat imagery and the availability of digitized terrain elevation data, it is appropriate to investigate the use of ancillary terrain data in improving the quality of information extracted from Landsat data.

### 4.2.1 ANCILLARY DATA CHARACTERISTICS

Prior to presenting maps of data in the various information channels, the forest stand map of Figure 7 is presented to provide a frame of reference for the Fraser Experimental Forest area. (The upper right-hand corner of this map is not completely filled in.) Interesting insights into both the data and problems faced in processing these data can be gained from a study of computer gray maps of the various Landsat and ancillary data channels.

Figure 8 is a map of Landsat Band 7 in the Fraser Forest. The forest is an experimental watershed draining into St. Louis Creek which runs in a SSQ to NNE pattern through the center of the forest.

As is seen in Figure 9, a map of the terrain elevation data, Fraser Forest is bounded on three sides by mountains. The area to the West of the St. Louis Creek is predominantly made up of slopes with an East-facing aspect, and to the East predominantly West-facing, although North- and South-facing ridges of lower elevation occur throughout (see Figure 10, a map of aspect angle relative to the sun's position).

An obvious feature of the Landsat data is the tendency of the signal magnitudes to correlate visually with the aspect angle. Comparison of the stand map (Figure 8) with Landsat data shows that, even though tree stands of a particular type appear on both sides of St. Louis Creek, the corresponding Landsat signals are very different in intensity. Comparison of slope angle patterns (Figure 11) and Landsat data indicates that the signals are affected by terrain slope as well.

Finally, the relative insolation factor (Figure 12), which incorporates both slope and aspect effects, exhibits a noticeable correlation to the Landsat data.

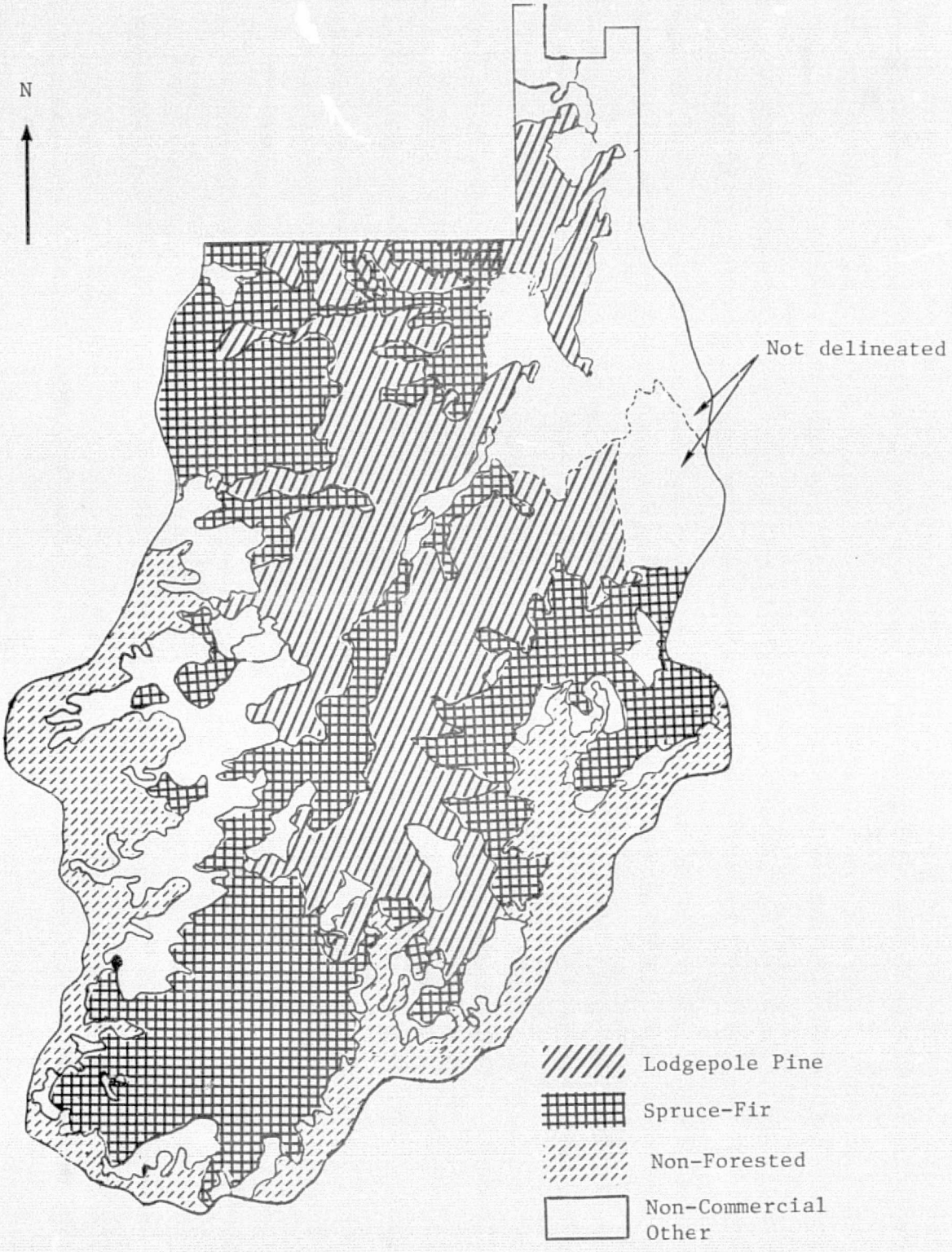
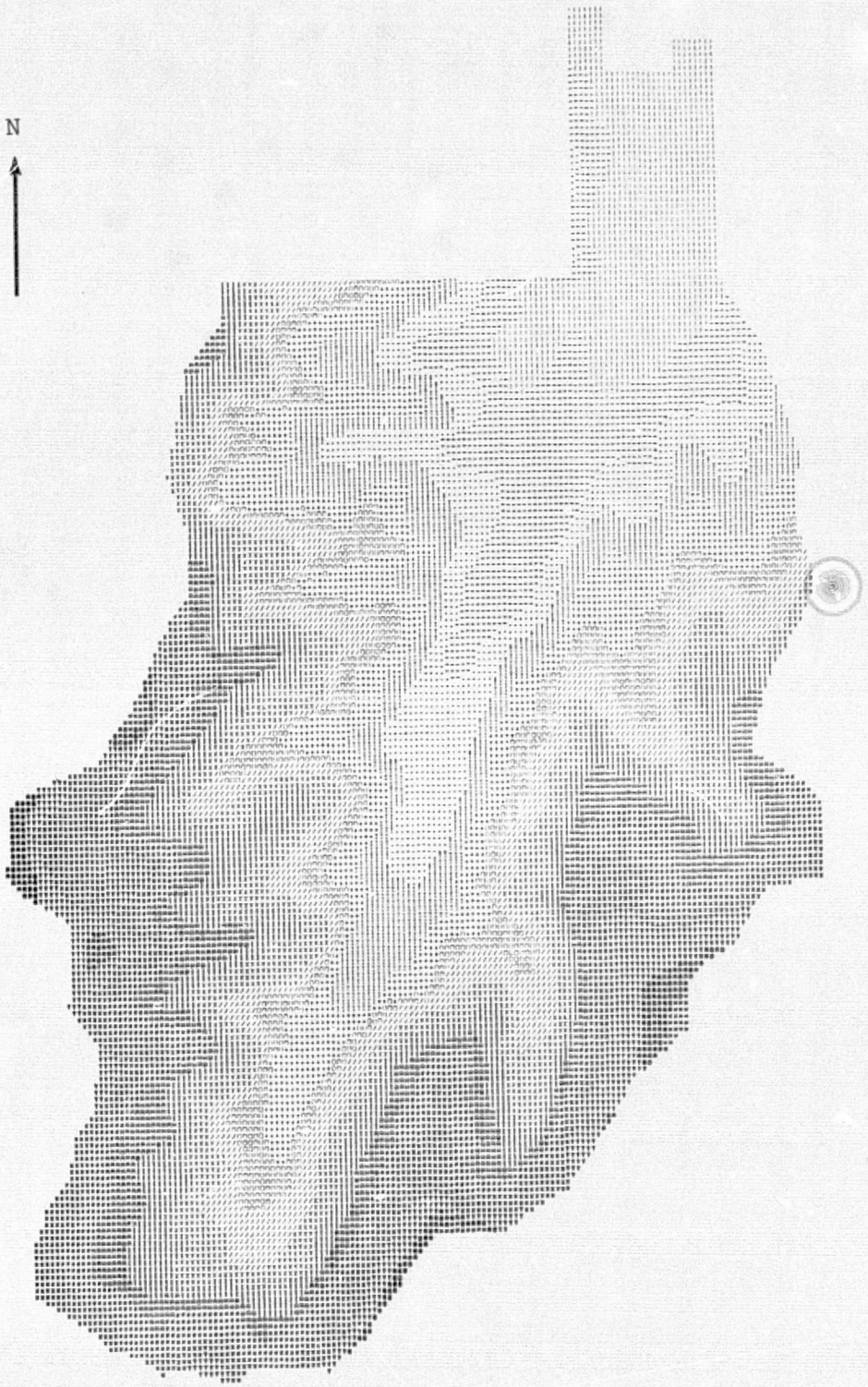



FIGURE 7. FRASER EXPERIMENTAL FOREST STAND MAP (1957 PHOTO BASE)

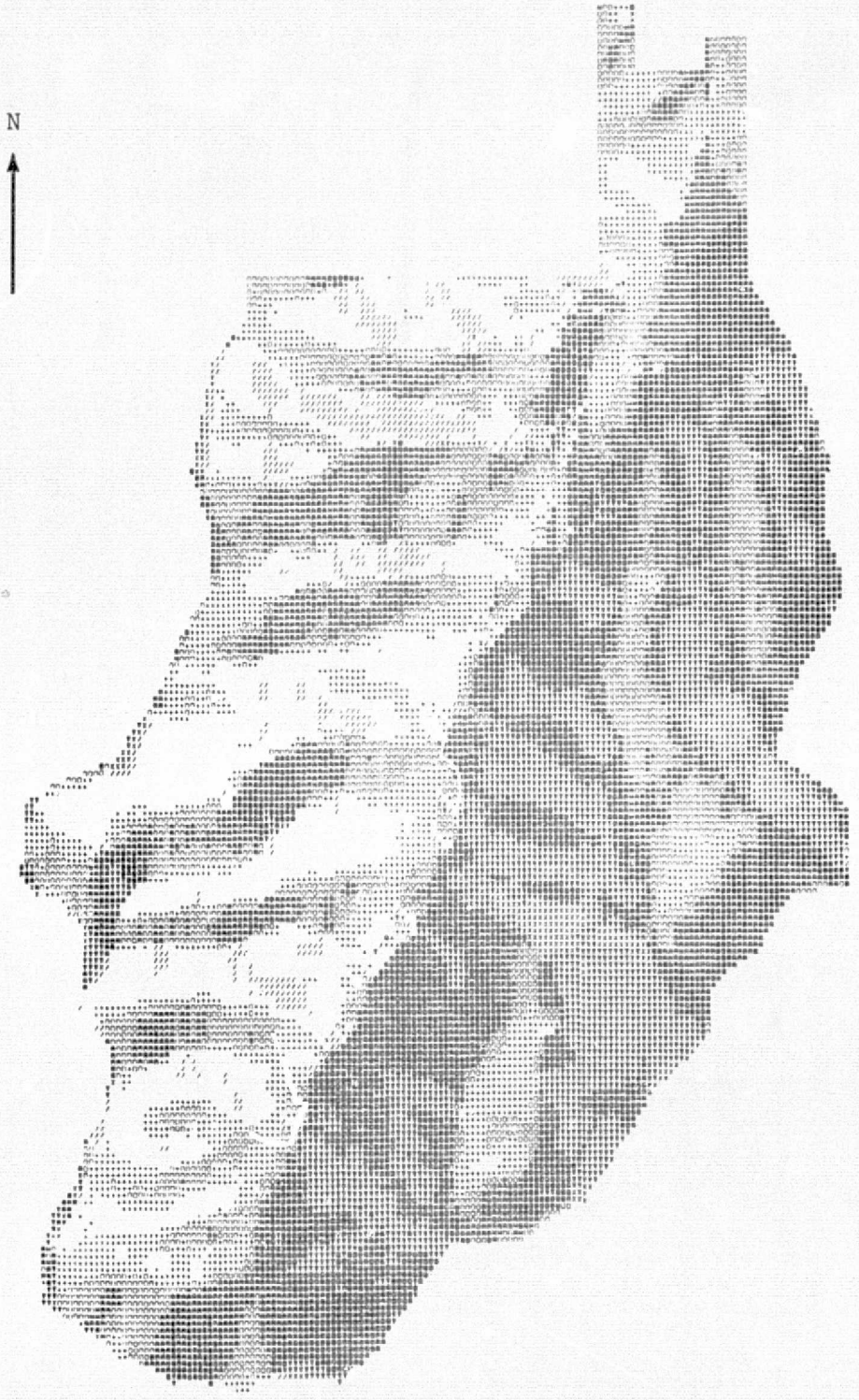

Brighter Tones Indicate Higher Signals

FIGURE 8. FRASER EXPERIMENTAL FOREST, LANDSAT BAND 7, 15 AUG 1973



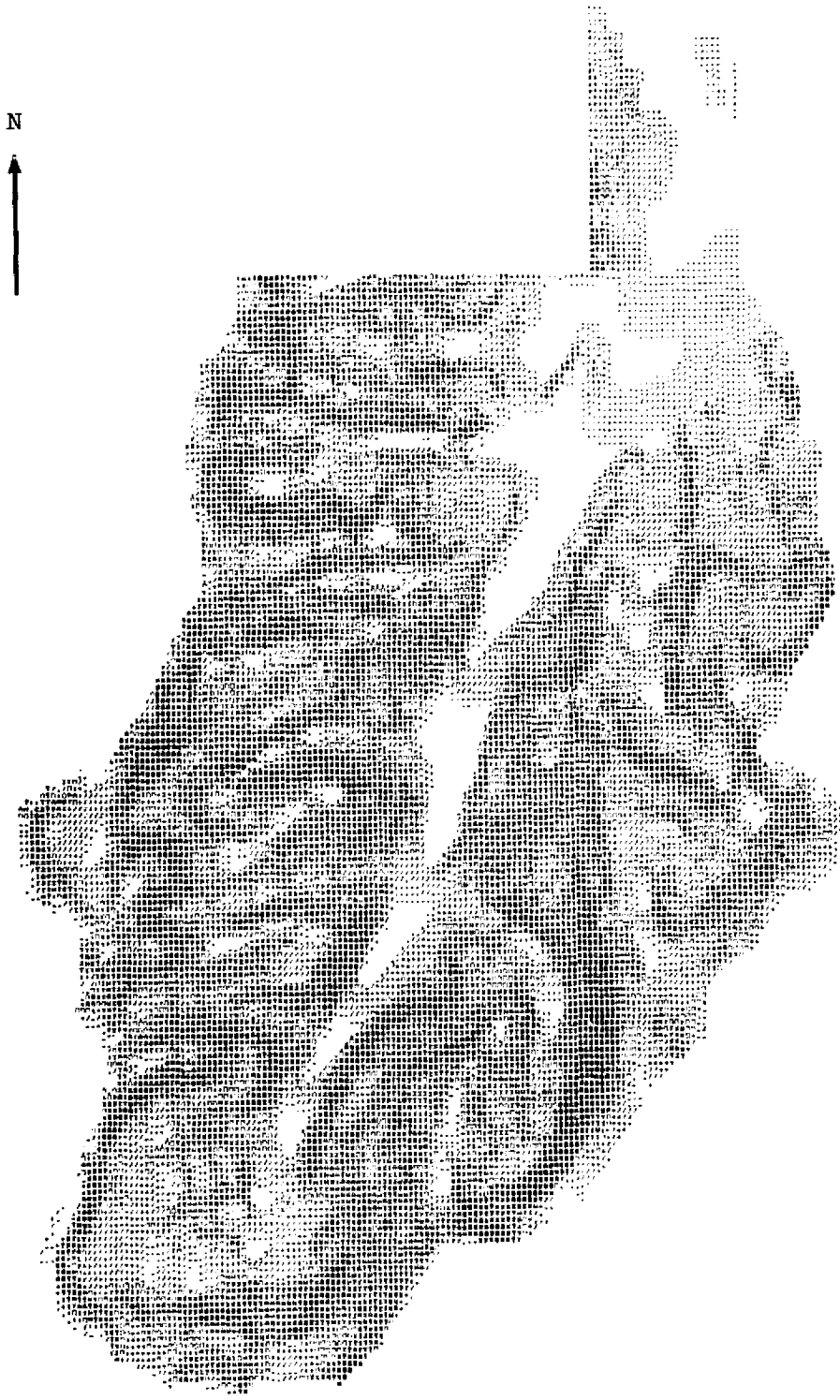

Increasingly Darker Tones Indicate Higher Elevation

FIGURE 9. FRASER EXPERIMENTAL FOREST, TERRAIN ELEVATION



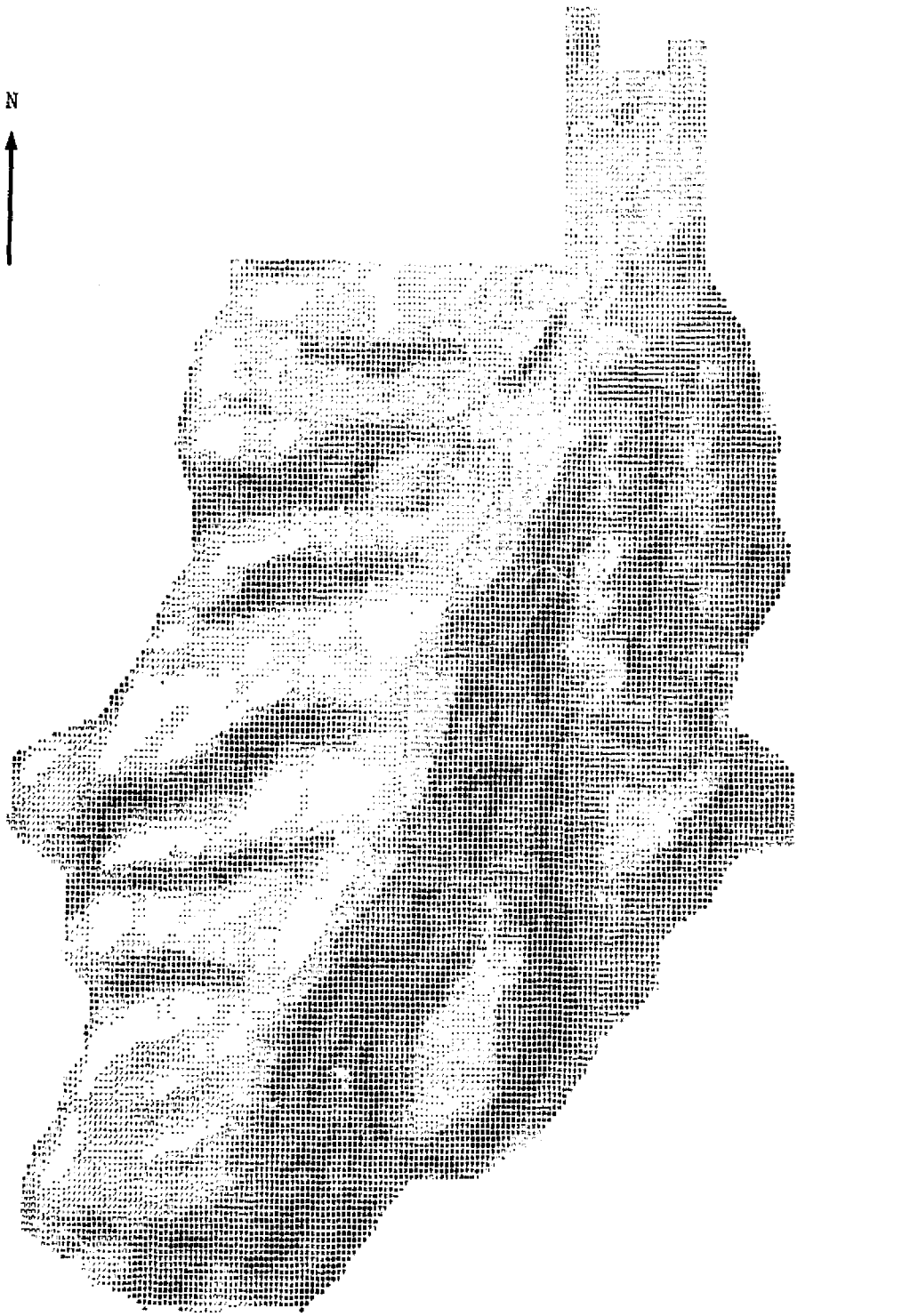

Brighter Tones Indicate Facing Towards the Sun (at 129° azimuth from North)

FIGURE 10. FRASER EXPERIMENTAL FOREST, SLOPE ASPECT ANGLE




Darker Tones Indicate Steeper Slopes (0 to 45° in this scene)

FIGURE 11. FRASER EXPERIMENTAL FOREST, TERRAIN SLOPE ANGLE




Greater Brightness Indicates Higher Levels of Insolation, i.e., Irradiance

FIGURE 12. FRASER EXPERIMENTAL FOREST, RELATIVE INSOLATION FACTOR

4.2.2 ANALYSIS OF ANCILLARY AND LANDSAT DATA

Relationships were examined between ancillary variables and two types of data -- actual Landsat data and simulated forest canopy reflectance data (in Landsat bands).

4.2.2.1 Actual Landsat Data

An analysis was made of the correlation between ancillary variables and the Landsat signals from selected cover types. Results are presented in Table 1. Note that relative aspect had the highest degree of correlation with signals in the forested regions, followed closely by the relative insolation factor. Low correlations were found with elevation and slope. Low correlations with slope are to be expected since the effects of slope depend so highly on the corresponding aspect angle relative to the sun. For example, at a 90° relative aspect, one would not expect a change in slope to have any effect on the signal. The correlations tend to be greater in Bands 6 and 7 than in Bands 4 and 5, especially for spruce-fir which had lower overall correlations than Lodgepole pine. The non-forested data had substantially lower correlations with the ancillary variables, perhaps due to a greater variation in the surface conditions.

Table 2 indicates, for these same data, that a significant portion of the variability in Landsat signals from the forested covers in Band 6 is explainable by the first two ancillary variables. Later, we examine the significance of such effects on the training of a classifier and subsequent classificacion of the data (Section 4.2.3.) and on the possibilities of reducing these effects through preprocessing (Section 4.2.4). In Section 4.3, we present additional analysis of correlations between ancillary data and transformed versions of the Landsat signals.

TABLE 1. CORRELATION BETWEEN LANDSAT SIGNALS AND ANCILLARY VARIABLES

| Landsat Band | Variable | Correlation for Cover Type: Lodgepole Pine | Spruce-Fir | Non-Forested |
|---|---|---|---|---|
| 4 | Aspect (Relative) | - 0.61 | - 0.28 | - 0.14 |
| | *$F_{RI}$ | 0.54 | 0.13 | 0.10 |
| | Slope | - 0.01 | - 0.06 | - 0.10 |
| | Elevation | 0.13 | 0.06 | 0.08 |
| 5 | Aspect | - 0.65 | - 0.23 | - 0.12 |
| | $F_{RI}$ | 0.58 | 0.06 | 0.08 |
| | Slope | - 0.01 | - 0.09 | - 0.08 |
| | Elevation | 0.19 | 0.12 | 0.09 |
| 6 | Aspect | - 0.73 | - 0.51 | - 0.20 |
| | $F_{RI}$ | 0.67 | 0.41 | 0.19 |
| | Slope | 0.02 | 0.06 | - 0.11 |
| | Elevation | 0.27 | 0.01 | 0.09 |
| 7 | Aspect | - 0.73 | - 0.53 | - 0.22 |
| | $F_{RI}$ | 0.67 | 0.44 | 0.23 |
| | Slope | 0.02 | 0.08 | - 0.10 |
| | Elevation | 0.30 | - 0.01 | 0.08 |

*$F_{RI}$ = Relative Insolation Factor

TABLE 2. DEPENDENCE OF LANDSAT SIGNAL ON TERRAIN VARIABLES
$R^2$ -- GOODNESS OF FIT
LINEAR REGRESSION -- LANDSAT BAND 6

| Independent Variable | Cover Type | |
|---|---|---|
| | Lodgepole Pine | Spruce-Fir |
| Relative Aspect | 0.532 | 0.259 |
| Relative Insolation Factor | 0.452 | 0.170 |
| All significant variables | 0.603 | 0.404 |

4.2.2.2 Simulated Forest Canopy Reflectances

A corresponding analysis was made of the simulated forest canopy reflectance data, except that elevation was not a factor. Correlation results are presented in Table 3 for three densities of trees and a combined class of grass and brush. Each class of tree densities was simulated over a variety of understory situations having grass or brush of varied densities over three different surface materials [1]. All densities of grass and brush are included in the fourth class in Table 3.

The results agree qualitatively with those on the actual Landsat data. Correlations with aspect and relative insolation factor are large while those with slope are small. One difference is that, here, correlation with relative insolation factor is higher than with aspect, this may be due to path radiance effects in Landsat data or other factors. Again, Band 6 and 7 correlations are greater than Band 4 and 5 correlations. Note also that correlations decrease as density decreases and that the grass and brush correlations are substantially lower than the tree correlations. Additional correlations are reported in Section 4.3 between ancillary variables and transformed versions of these reflectance values.

TABLE 3. CORRELATION BETWEEN SIMULATED LANDSAT-BAND FOREST CANOPY REFLECTANCES AND ANCILLARY VARIABLES

| Band | Variable | Correlation for Cover Type: Forest — High Density | Medium Den. | Low Den. | Grass & Brush |
|---|---|---|---|---|---|
| 4 | Aspect (Relative) | - 0.77 | - 0.70 | - 0.51 | - 0.33 |
| | *$F_{RI}$ | 0.92 | 0.79 | 0.60 | 0.37 |
| | Slope | - 0.10 | - 0.09 | - 0.08 | - 0.06 |
| 5 | Aspect | - 0.76 | - 0.65 | - 0.51 | - 0.32 |
| | $F_{RI}$ | 0.91 | 0.. + | 0.55 | 0.36 |
| | Slope | - 0.07 | - 0.08 | - 0.09 | - 0.07 |
| 6 | Aspect | - 0.87 | - 0.82 | - 0.71 | - 0.51 |
| | $F_{RI}$ | 0.97 | 0.92 | 0.80 | 0.58 |
| | Slope | - 0.07 | - 0.06 | - 0.07 | - 0.07 |
| 7 | Aspect | - 0.87 | - 0.83 | - 0.73 | - 0.51 |
| | $F_{RI}$ | 0.98 | 0.93 | 0.82 | 0.58 |
| | Slope | - 0.09 | - 0.08 | - 0.07 | - 0.07 |

*$F_{RI}$ = Relative Insolation Factor

4.2.3 EFFECTS OF INCOMPLETE TRAINING ON CLASSIFICATION PERFORMANCE

Inaccessibility to areas due to terrain or availability of resources may limit the amount of training data that can be obtained in practice. Therefore, an experiment was conducted to determine the extent to which limited training (relative to the aspect angles included) would influence classification performance.

Classifier training was first limited to data from aspects facing predominantly East (toward the sun). Three cover classes were represented -- forested (Lodgepole pine and spruce-fir), non-forested, and non-commercial. The resultant signatures were used to classify three groupings of the original Landsat data having:

a. the same aspects
b. the opposite aspects
c. a combination of all aspects.

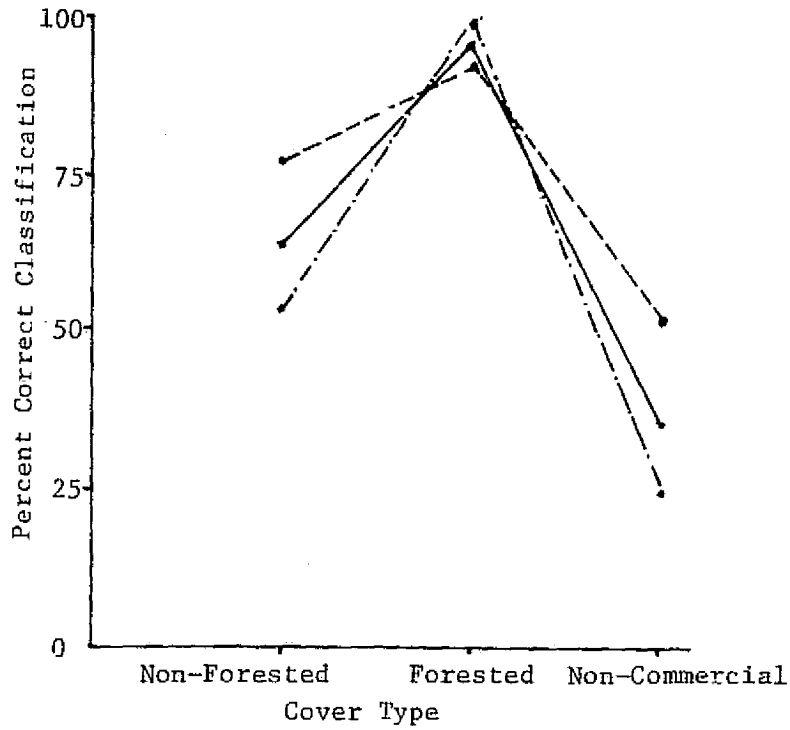
The results presented in Figure 13 exhibit a wide variation in percent correct classification, implying that the range of signal variation for each cover type classified was not always properly represented by the signature used for classification. The effect was most pronounced for the forested cover type in the case where training was performed on West aspects (away from the sun), see Figure 13(b). There, forest classification accuracy was drastically reduced when opposite aspects were classified. Little effect was noted on the classification of forested areas with East-facing signatures.

Spruce-fir and Lodgepole pine are represented as a single class, forest, in this analysis. An attempt was made to discriminate spruce-fir from Lodgepole pine, but due to their high degree of spectral similarity, acceptable separation was not achieved.

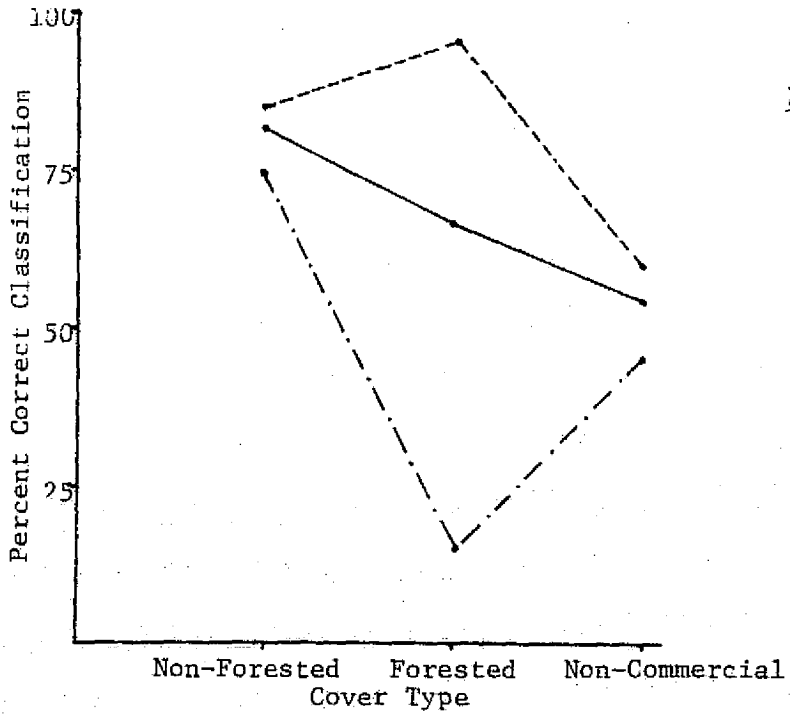
The results indicated in Figure 13(a), on slopes facing the sun, are not as erratic as those in Figure 13(b). A plausible explanation for these results can be hypothesized using known characteristics of the cover types.

Figure 14 shows the relationship between forested and non-forested pixels in a transformed data space. This space is discussed in detail in later sections. For our proposes here, it is simply important to note that forested and non-forested distributions are represented in a manner that closely correlates scene illumination to radial distance from the origin.

When training is carried out on the darker West-facing slopes, (Figure 14(b)), the decision boundary produced would likely result in good classification accuracy for the non-forested type, with little difference related to the slope on which classification was carried out. In contrast, forested classification accuracy could be expected to be high on



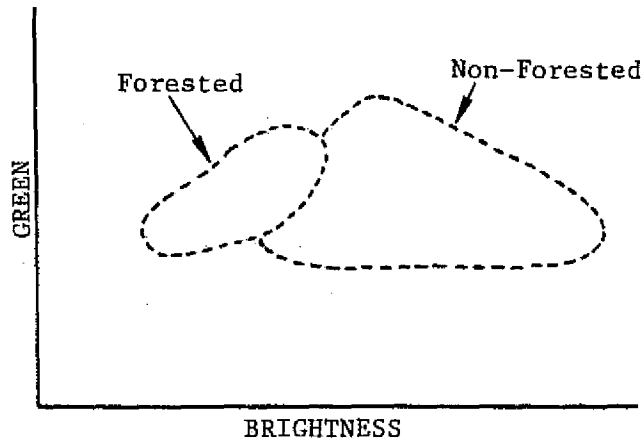

(a) Training on East Aspects (Toward Sun)



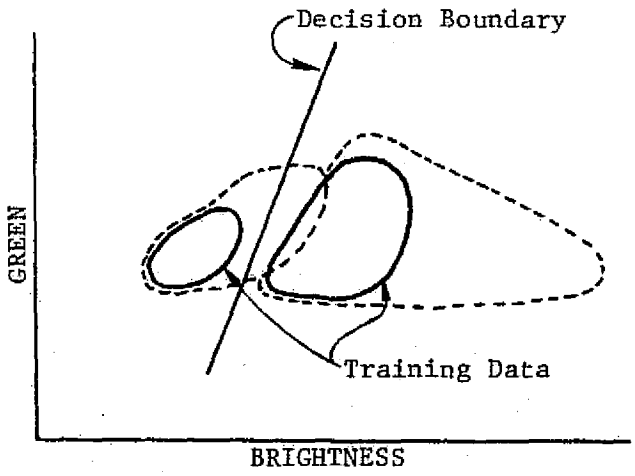

KEY: Figs. (a) and (b)
CLASSIFICATION ON:
— — — Same Aspect
——— All Aspects
• - - - • Opposite Aspect

(b) Training on West Aspects (Away from Sun)

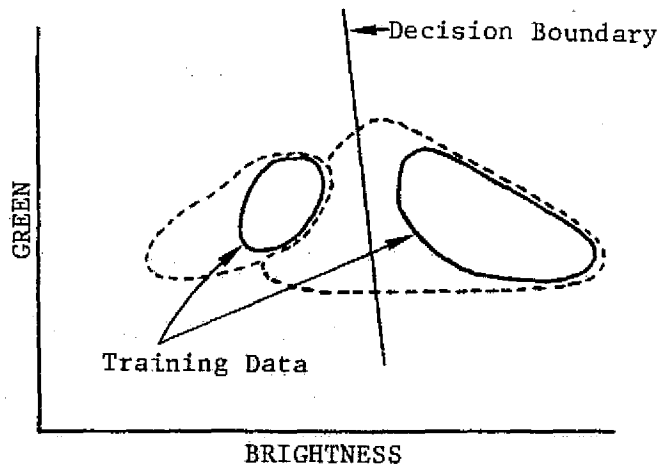
FIGURE 13. EFFECTS OF LIMITED TRAINING ON CLASSIFICATION PERFORMANCE (UN-PREPROCESSED DATA)




(a) Overall Data Distributions
(for All Aspects)




(b) Training on West Aspects
(Away from Sun)




(c) Training on East Aspects
(Toward Sun)

FIGURE 14. ILLUSTRATION OF DECISION BOUNDARIES FOR CASES OF LIMITED TRAINING

West-facing slopes, but very low in East-facing slopes with such a decision boundary. These relationships are apparent in the classification results illustrated in Figure 13(b).

Training on the brighter East-facing slopes produces the opposite effect (Figure 14(c)). The decision boundary now includes all of the forested pixels in the forested class, and classification accuracy should be high, regardless of aspect. Non-forested classification, on the other hand, should be degraded in a manner similar to that for forested classification with West-slope training: high accuracy on the same slopes, low on all slopes, and lowest on opposite slopes. Again, these are the relationships that resulted, as shown in Figure 13(a).

These results indicate a need to approach classifier training cautiously for mountainous terrain. One suggested solution is to stratify the signal space, based on the ancillary variables, such as was done above for aspect, and then train in each stratum. This would help insure that the variability of the signals of each cover type was represented. Such an approach, however, may not be desirable or feasible in mountainous terrain due to the physical constraints imposed on ground training, the complexity of determining appropriate strata, and the number of signatures required.

Another approach investigated was that of preprocessing the data to normalize the terrain effects to those of a fixed condition, e.g., terrain facing the sun or horizontal terrain. This could reduce that variability in the Landsat signals which is attributable to terrain effects and, in turn, reduce the complexity of training and improve classification results.

### 4.2.4 PREPROCESSING TO IMPROVE CLASSIFICATION

Preprocessing, as used here, refers to the application of transformations to MSS signals to remove or reduce systematic variations in the data. Preprocessing algorithms can have either an empirical or a

theoretical basis, or a combination of the two. Both empirical and theoretical preprocessing algorithms were investigated in this study.

The empirical algorithms were based on the previously described statistical analysis of the variability of signals from individual cover types and extent to which it could be explained by ancillary variables. As was shown in Table 2, the relative insolation factor and relative aspect angles both explained a sizeable fraction of the variance in linear regression analyses, especially for the Lodgepole pine cover class.

We used two basic transformation models for preprocessing data prior to classification. The first transformation model was of the form:

$$\vec{S}' = \vec{S} - G\,(\vec{A} - \vec{A}_{Ref}) \qquad (A)$$

where

$\vec{S}'$ is the adjusted signal vector for the scene element,

$\vec{S}$ is the original signal vector,

$\vec{A}$ is a vector describing the relevant ancillary conditions at the element,

$\vec{A}_{Ref}$ describes the reference set of ancillary conditions,

and $G$ is function of $(\vec{A} - \vec{A}_{Ref})$ which can be determined by regressing Landsat signals on one or more ancillary variables.

Thus, Equation (A) defines an additive correction model. We determined G as a function of relative aspect and relative insolation factor for use in separate preprocessing trials.

The other transformation model form used for preprocessing was:

$$\vec{S}' = \frac{1}{F(\vec{A})}\,(\vec{S} - \vec{L}_p) \qquad (B)$$

where $\vec{S}'$ and $\vec{S}$ are as defined previously,

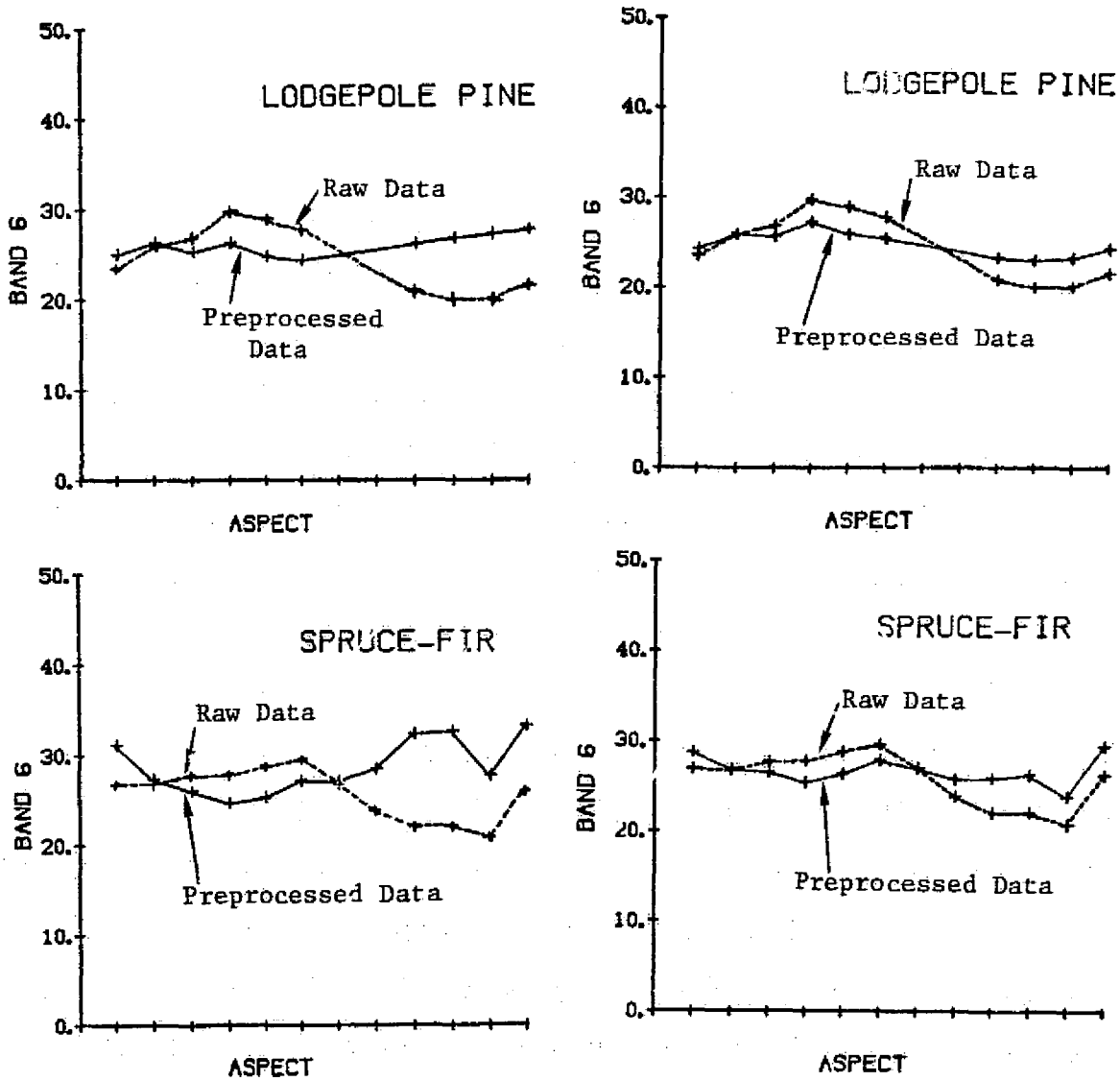
$\vec{L}_p$ is an estimate of the atmospheric path radiance, i.e., extraneous radiation that emanates from the atmosphere rather than from the scene element being viewed,

and $F(\vec{A})$ is a function dependent on ancillary variables.

This type of transformation has a better physical basis than that of Equation (A), in that it first subtracts an additive path radiance term and then performs a multiplicative correction which is appropriate for differences in irradiance.

For one of our trials, we let $\vec{F}(\vec{A})$ be $F_{RI}$, the relative insolation factor, which is proportional to the irradiance on the scene element. For another, we used a modified relative insolation factor $F_{MRI}$ which was based on our work in simulating the reflectance of forested canopies. As described in Volume I [1], the model predicted forest canopy reflectances to be non-diffuse and supporting indications were found in Landsat data from the Fraser Experimental Forest. When $F_{RI}$ was applied to data, it overcorrected for slope and aspect variations. Therefore, we decided to modify the relative insolation factor to make it more closely match the trend in the simulated data. The one modification we tested was to define a new slope angle and recalculate the relative insolation factor using Equation (2). The new slope angle was obtained by multiplying the actual slope angle by the sine of the sun zenith angle. This multiplication factor was determined through analysis of the simulation data which had been produced for two sun zenith angles, $36^{\circ}$ and $51^{\circ}$.

Figure 15 presents comparisons between original and preprocessed data (Landsat Band 6) from the Lodgepole-pine and spruce-fir cover classes. The designated pixels were grouped into bins corresponding to $30^{\circ}$ intervals in aspect, relative to the sun's azimuth. Plots of the means of original data within these bins exhibit sinusoidal variations as a function of aspect. When preprocessed using the transformation of Equation (B) and the ordinary relative insolation factor, the curves in Figure 15(a) show that the data were overcorrected, as previously noted. The modified relative insolation factor produces a much better transformation, with only a slight residual variation present, as shown in Figure 15(b).




(a) Preprocessing With Relative Insolation Factor

(b) Improved Preprocessing with Modified Relative Insolation

NOTES: (1) Aspect scale designates intervals of $30^{o}$ from $0^{o}$ through $360^{o}$, relative to sun's azimuth.

(2) Preprocessing was according to Equation (B) with $L_p \neq 0$.

FIGURE 15. COMPARISON OF ASPECT DEPENDENCE OF LANDSAT BAND 6 SIGNALS BEFORE AND AFTER PREPROCESSING BY TWO DIFFERENT TRANSFORMATIONS

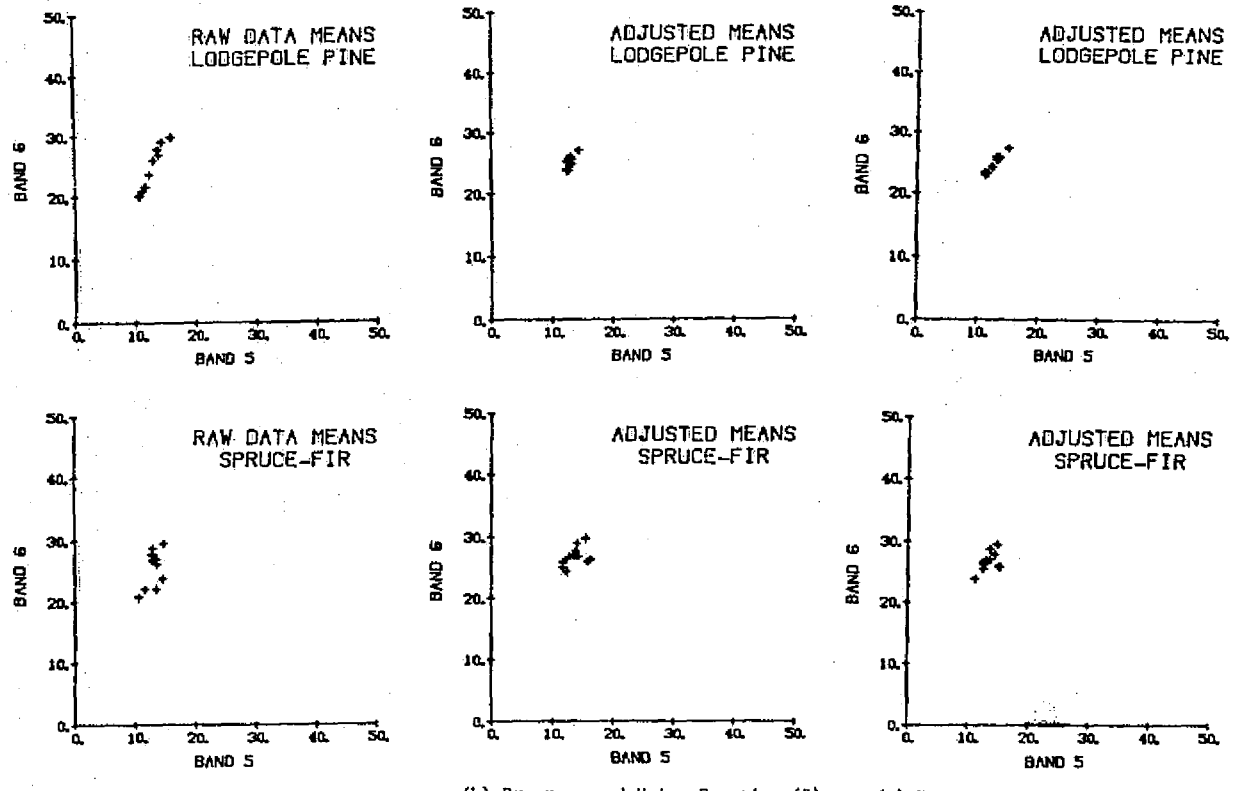
The two-channel scatter diagrams of Figure 16 provide another display of the effectiveness of two preprocessing algorithms in reducing the variance of the signals from the Lodgepole-pine and spruce-fir cover types.

Next, we turn to classification of the data to evaluate the possible beneficial effects of preprocessing. Results to be compared with those presented earlier in Figure 13 for unpreprocessed data are presented in Figures 17 and 18 for two different preprocessing algorithms. Both figures exhibit marked improvement over the unpreprocessed results, especially for the case where training was limited to West-facing slopes. Here, there is little if any dependence on the orientations of the data that were classified, in contrast to the drastic drop in performance for forested pixels on opposite-facing slopes with unpreprocessed data.

Figure 17 represents preprocessing by an empirical transformation having the form of Equation (A) and being based on a regression of signal values on the relative insolation factor. On the other hand, the transformation of Figure 18 had a theoretical basis (the modified relattive insolation factor) and utilized Equation (B). The classification of non-forested areas in Figure 18(b) is somewhat less accurate than without preprocessing, but there is an improvement in results for the other-than-forested categories in Part(a), i.e., training on East-facing slopes. Overall, preprocessing appears advantageous in these results.

### 4.2.5 USE OF ANCILLARY DATA AS CLASSIFICATION VARIABLES

Another topic investigated in exploratory fashion was whether or not the inclusion of ancillary variables as extra data channels, with Landsat channels, in classification would enhance the separability of scene classes. This approach showed promise, under the condition that variability in the signals was adequately represented in the signatures. The inclusion of variables which primarily affect the relative amount of solar irradiance reaching an element improved classification performance (see Figure 19).




(a) Original Data

(b) Preprocessed Using Equation (B) With $F = F_{MRI}$ and $L_p = 0$

(c) Preprocessed Using Equation (B) With $F = F_{MRI}$ and $L_p \neq 0$

FIGURE 16. EXAMPLES OF REDUCED SCATTER IN LANDSAT DATA VALUES ACHIEVED BY PREPROCESSING

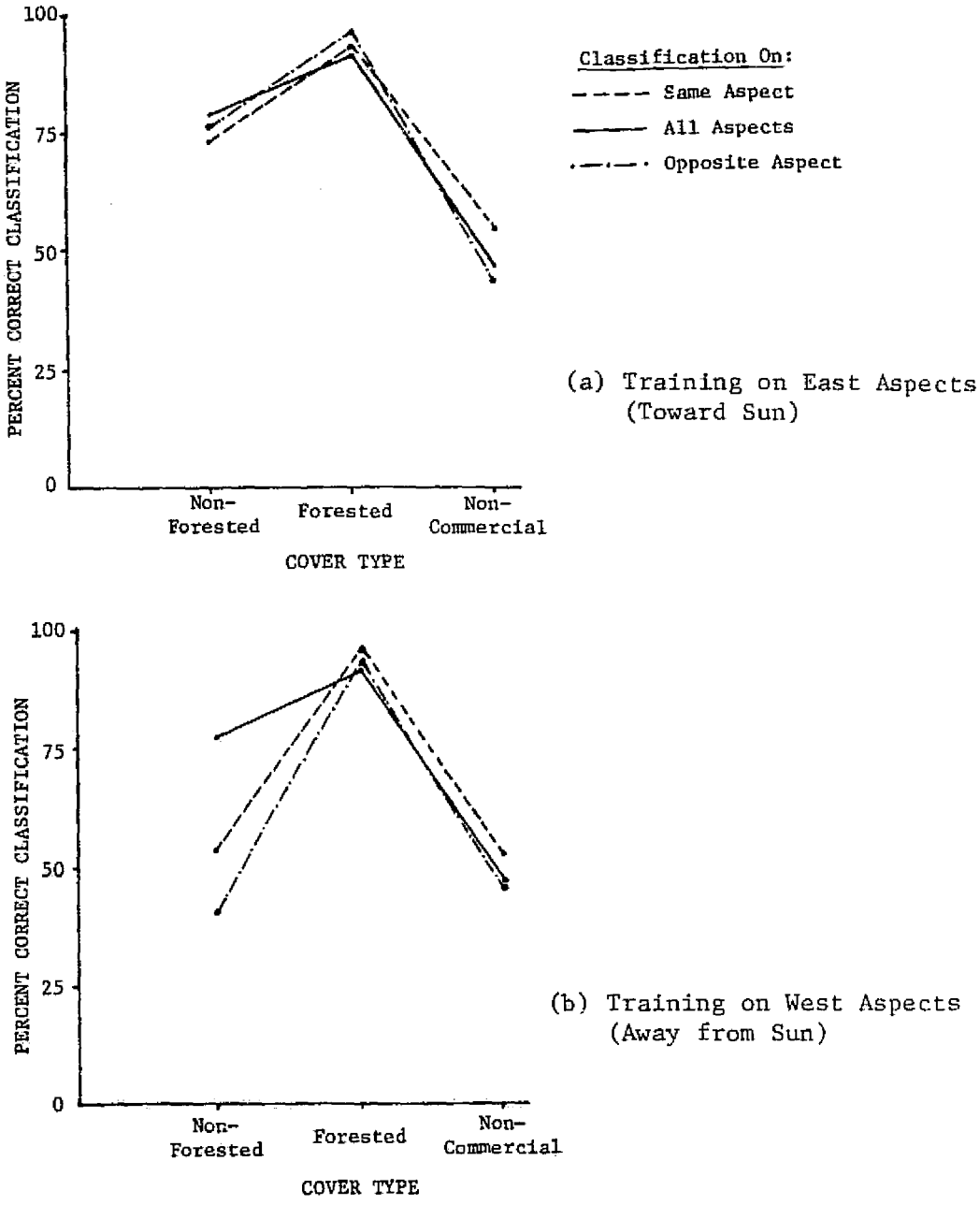
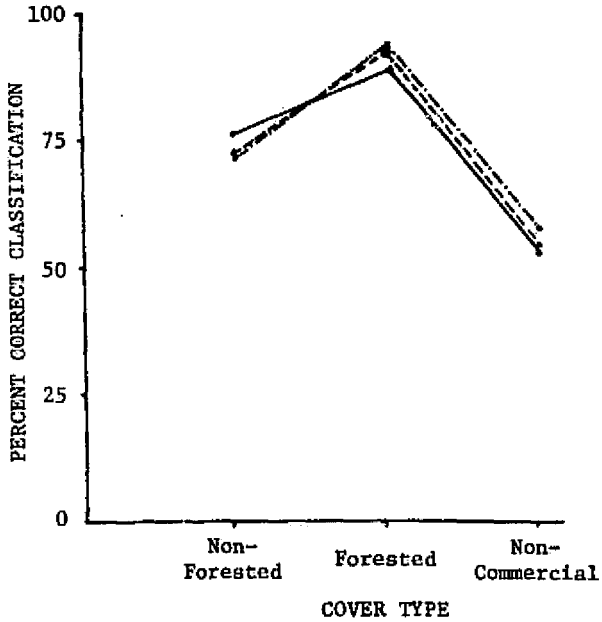
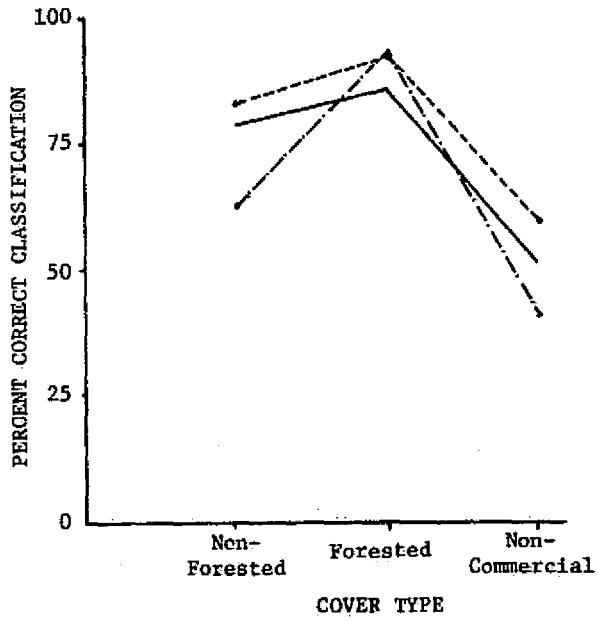



FIGURE 17. EFFECTS OF PREPROCESSING ON CLASSIFICATION PERFORMANCE WITH LIMITED TRAINING -- PREPROCESSING BY EQUATION (A) USING REGRESSION ON $F_{RI}$



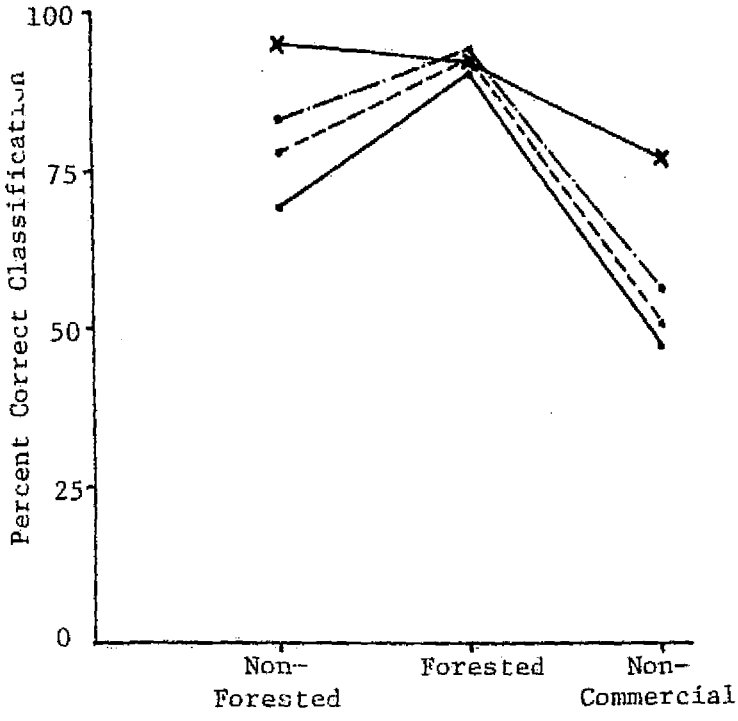

Classification On:
- Same Aspect
- All Aspects
- Opposite Aspect

(a) Training on East Aspects (Toward Sun)




(b) Training on West Aspects (Away from Sun)

FIGURE 18. EFFECTS OF PREPROCESSING ON CLASSIFICATION PERFORMANCE WITH LIMITED TRAINING -- PREPROCESSING BY EQUATION (B) USING $F_{MRI}$




NOTE: Adequate Training
No Preprocessing

KEY: ———— Four-Channel Classifier

---------- + Relative Aspect

.——.——. + Relative Aspect + Slope + Rel. Insolation Factor

x————x + Relative Aspect + Elevation

FIGURE 19. INCLUSION OF ANCILLARY DATA CHANNELS IN CLASSIFIER

Even more significant, in this data set, however, was elevation, which apparently had residual correlation with particular cover types. The separability of the non-commercial cover type, which is a mixture of non-harvestable forest, brush, and grasses, was especially enhanced due to its appearance at higher elevations than forested covers and lower elevations than non-vegetative cover. The difficulty of this type of coincidence approach is in correctly representing the patterns of terrain conditions that favor particular covers.

## 4.3 APPLICABILITY OF LACIE-ORIENTED INFORMATION EXTRACTION TECHNIQUES

Two types of information extraction techniques, developed at ERIM in the course of research and development for agricultural applications of Landsat data for LACIE and judged by us to be of potential value for forestry and rangeland applications, were investigated. The first type included techniques of transforming original Landsat signals which have potential for data compression and improved interpretability of the extracted features. Both the Tasselled-Cap transformation, which produces linear combinations of the original Landsat signals, and a non-linear polar-coordinate transformation of the signals were investigated.

The second type included clustering techniques, both spectral-only and spectral-spatial techniques which offer potential advantages for classification and data compression.

Both types of techniques were applied to Landsat data acquired over the Grand County, Colorado, test site on 15 August 1973. In addition, the transformation techniques were applied to the simulated forest canopy reflectance data previously described.

### 4.3.1 TASSELLED-CAP TRANSFORMATION

The Tasselled-Cap transformation is a linear transformation of Landsat signals which was developed at ERIM [20] to exploit the high degree of correlation that exists between signals in certain pairs of Landsat bands in data from agricultural scenes. It produces four

linear combinations of the original signals, combinations which have the following physical interpretations:

a. Brightness - aligned generally with the direction of variations in the brightness of bare soil (and exposed rock).

b. Green - orthogonal to brightness in the plane of principal variation and in the direction of signals from healthy green plants.

c. Yellow - orthogonal to the green-brightness plane, in a direction specified by an example of yellowed vegetation, sensitive to existing atmospheric conditions, and serving as a haze diagnostic.

d. Nonsuch - orthogonal to the other directions, containing residual variation, and having no direct physical interpretation.

We sought to determine whether or not the Tasselled-Cap transformation is appropriate and useful for forest scenes, in addition to the agricultural and rangeland scenes for which it was developed.

The Tasselled-Cap transformation was applied to Landsat data from specific cover types in the Fraser Experimental Forest. Analysis of total signal variability revealed that nearly 80% to 90% of the variance was in the brightness component, 10% to 20% in the green component, only 1% to 2% in the yellow component, and less in the nonsuch component, as shown in Table 4. This result is similar to experiences we have had with agricultural data.

To test the generality of this conclusion, we next analyzed data acquired over the larger Grand County, Colorado, test area. A systematic sample (every 10$^{th}$ point on every 10$^{th}$ line) was analyzed, with results as shown in Figure 20(a). Again, the vast majority (96%) of the variance lies within the green-brightness plane. Thus, the data

TABLE 4. TASSELLED-CAP TRANSFORM RESULTS FOR FRASER FOREST LANDSAT DATA

| | % OF TOTAL VARIANCE | | | |
|---|---|---|---|---|
| COVER TYPE | BRIGHTNESS | GREEN | YELLOW | NONSUCH |
| NON-FORESTED (NF) | 87 | 11 | 2.0 | 0.2 |
| LODGEPOLE PINE (LP) | 78 | 18 | 1.7 | 1.5 |
| SPRUCE-FIR (SF) | 77 | 21 | 1.1 | 1.0 |
| NON-COMMERCIAL (NO) | 84 | 15 | 0.7 | 0.4 |

% TOTAL VARIANCE IN THE PLANE OF BRIGHTNESS AND GREEN, BY COVER TYPE

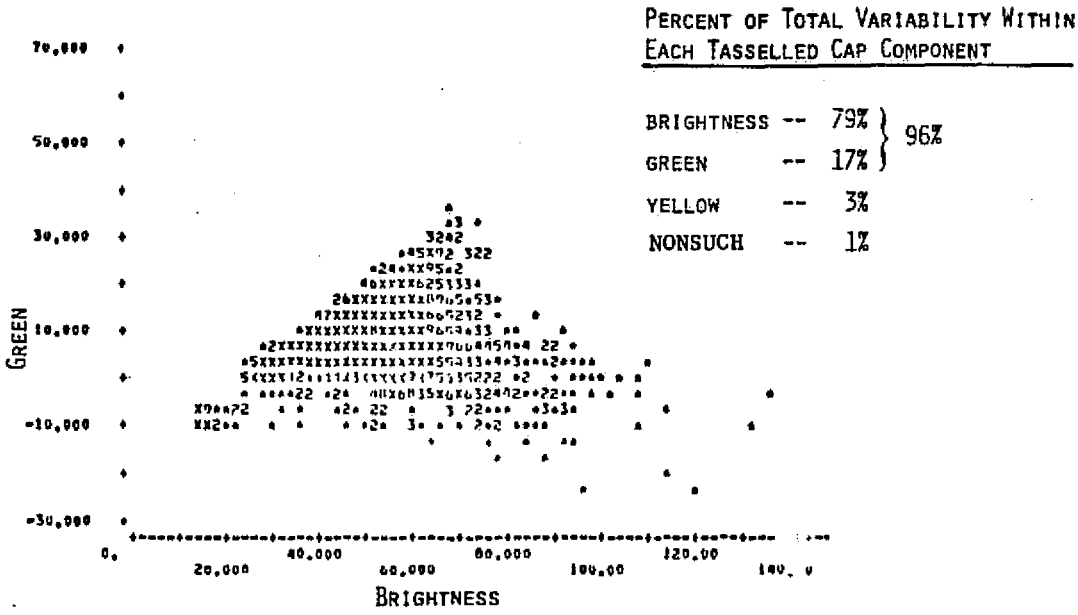
| NF | LP | SF | NO |
|---|---|---|---|
| 98 | 97 | 98 | 99 |

could be reduced to three or possibly two channels with little or negligible loss of information, especially for multitemporal analyses.

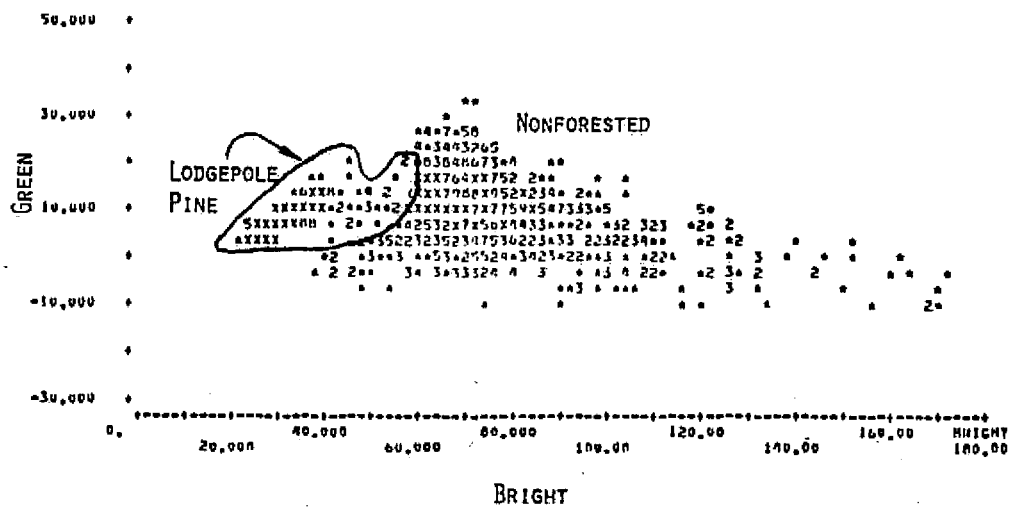
The scatter diagram in Figure 20(a) covers a wide range of cover types, including water, exposed rock, and various kinds of vegetation. A more selective display of data is presented in Figure 20(b); pixels from Lodgepole pine areas are identified and the remainder are from non-forested areas. Note the wide dispersion in the brightness direction for the non-forested pixels, representing bright to dark exposed rock and various amounts of vegetative cover, including grasses which are brighter green than pines. The dispersion in the green direction indicates that some areas have rather extensive amounts of vegetative cover. Note that the pine trees are both dark in brightness and relatively low in greenness. To avoid negative green values as found on these figures, 32 counts are usually added to each Tasselled-Cap component.

### 4.3.2 POLAR GREEN-ANGLE/BRIGHTNESS-RADIUS TRANSFORMATION

In displays of Landsat and simulated reflectance data from agricultural scenes in the Tasselled-Cap plane, it was noted that a (non-linear) polar coordinate transformation might provide a useful alternative pair of variables [21]. Figure 21 illustrates the way in which




(a) In Systematic Sample over Grand County, Colorado




(b) From Selected Cover Classes in Fraser Forest

FIGURE 20. TASSELLED-CAP TRANSFORMATION OF LANDSAT DATA

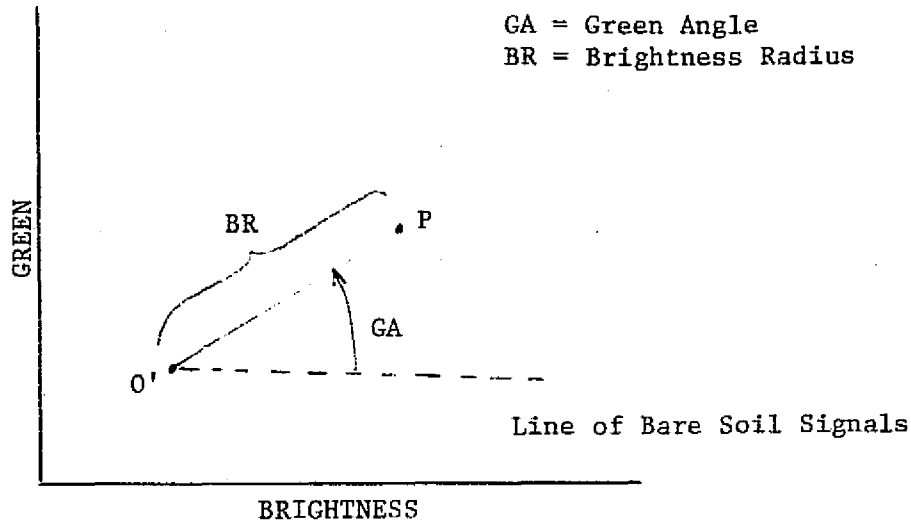



FIGURE 21. DEFINITION OF POLAR GREEN-ANGLE/ BRIGHTNESS-RADIUS TRANSFORMATION

the two new variables are defined in the Tasselled-Cap plane. First, a displaced origin (O') must be selected for the transformation; the location of O' is influenced largely by the atmospheric conditions, particularly the amount of path radiance. Locating O' accurately and consistently is one of the major obstacles to the routine use of the transformation. The "brightness radius" is the radial distance from the displaced origin. Bare soils of varied brightness tend to lie along a line approximately parallel to the brightness axis. That line serves as the starting ray for the "green angle" of point P which is measured counter-clockwise from it.

In selected agricultural data, the angular component has been found to be indicative of the amount of green vegetation present and largely independent of soil brightness effects [21], these latter effects being reflected mainly in variations of the brightness radius.

The utility of this polar transformation has not been thoroughly assessed for agricultural data. Nevertheless, we made exploratory

calculations to examine its characteristics in the forestry data assembled under this contract -- both Landsat and simulated reflectance data.

Table 5(a) presents correlations that were computed between ancillary variables and the green-angle and brightness-radius components of the simulated reflectances, as well as the Tasselled-Cap green and brightness components. Part (b) of this table presents similar quantities computed from Landsat data for selected cover types in the Fraser Experimental Forest. The reflectances and pixels analyzed spanned a wide range of slope and aspect variations. These correlations tend to support the prior observations that signal variations associated with surface orientations are most highly correlated with brightness and brightness-radius components. In addition, note how the green angle is much less correlated with these variations than is the Tasselled-Cap green component.

The Lodgepole pine correlations differ somewhat from the others, having greater correlations for all four components and having appreciable correlation with green angle. The latter may be a result of the darkness of the signals from this cover class, placing the values near the displaced origin and making the angle measure more sensitive to noise and other variations (refer back to Figure 20). Note that deciduous trees and grasses would be expected to have greater brightness radii for comparable green angles.

### 4.3.3 SPECTRAL CLUSTERING TECHNIQUES

Clustering techniques were investigated because they provide a potential means for: (a) unsupervised classification and mapping of scenes, (b) stratification and identification of areas for training, and (c) data compression. Clustering techniques associate pixels that are similar to each other in the feature space, according to one of several possible similarity measures. A drawback of virtually all clustering techniques is the need for user-provided parameter settings.

TABLE 5. CORRELATIONS BETWEEN ANCILLARY VARIABLES AND TRANSFORMED LANDSAT VARIABLES

(a) Simulated Landsat Inband Reflectance Values

| | Correlation With: | |
|---|---|---|
| Variable | Relative Aspect | Relative Insolation Factor |
| Brightness (Tasselled-Cap) | -0.53 | 0.60 |
| Green (Tasselled-Cap) | -0.46 | 0.51 |
| Brightness Radius | -0.56 | 0.63 |
| Green Angle | 0.04 | -0.05 |

(b) Landsat Data from Fraser Experimental Forest

(i) Spruce-Fir Cover Class

| | Correlation With: | |
|---|---|---|
| Variable | Relative Aspect | Relative Insolation Factor |
| Brightness (Tasselled-Cap) | -0.43 | 0.30 |
| Green (Tasselled-Cap) | -0.55 | 0.54 |
| Brightness Radius | -0.48 | 0.37 |
| Green Angle | 0.11 | 0.09 |

(ii) Lodgepole Pine Cover Class

| | Correlation With: | |
|---|---|---|
| Variable | Relative Aspect | Relative Insolation Factor |
| Brightness (Tasselled-Cap) | -0.73 | 0.66 |
| Green (Tasselled-Cap) | -0.66 | 0.62 |
| Brightness Radius | -0.74 | 0.67 |
| Green Angle | 0.43 | -0.36 |

(iii) Non-Commercial Cover Class

| | Correlation With: | |
|---|---|---|
| Variable | Relative Aspect | Relative Insolation Factor |
| Brightness (Tasselled-Cap) | -0.34 | 0.33 |
| Green (Tasselled-Cap) | -0.40 | 0.36 |
| Brightness Radius | -0.39 | 0.38 |
| Green Angle | -0.00 | -0.02 |

The lack of clearly-defined criteria for determining the right levels for any given scene forces the user to experiment until a suitable set of values to use is determined.

In pure spectral clustering, the only criterion used in forming clusters is spectral similarity. ERIM's CLUSTR algorithm [22], used in this study, considers a modified chi-squared distance from each pixel to each cluster previously created. The pixel is assigned to the nearest cluster, held for later classification, or used as the initial element of a new cluster (depending on its distance from the nearest cluster).

As seen by comparison of Figures 22 and 23, spectral clustering of preprocessed data successfully distinguished, on a coarse level, forested regions in the Fraser Experimental Forest from those areas classified as non-forested or non-commercial. The same parameter settings caused both water pixels and grass/rangeland pixels to be separated from the rest of the scene in the larger Grand County subset of data.

### 4.3.4 SPECTRAL/SPATIAL CLUSTERING TECHNIQUES

A second type of clustering technique that has recently been under extensive development and use at ERIM is a joint spectral/spatial algorithm called BLOB [23]. This algorithm includes pixel coordinate channels along with spectral channels in the clustering process, thereby combining spatial "nearness" with spectral similarity in the clustering metric. The spectral and spatial channels are weighted independently so that the best combination can be used for a given scene.

BLOB calculates a modified chi-squared distance from each pixel to an array of previously existing spectral/spatial clusters (called blobs) and compares that distance to a user-provided limit. Each pixel is added to the closest blob, or is used to initiate a new blob if none of the existing blobs are within the required distance. The end product of this algorithm is an irregular grid of blobs which tends to conform well to field patterns in agricultural areas.

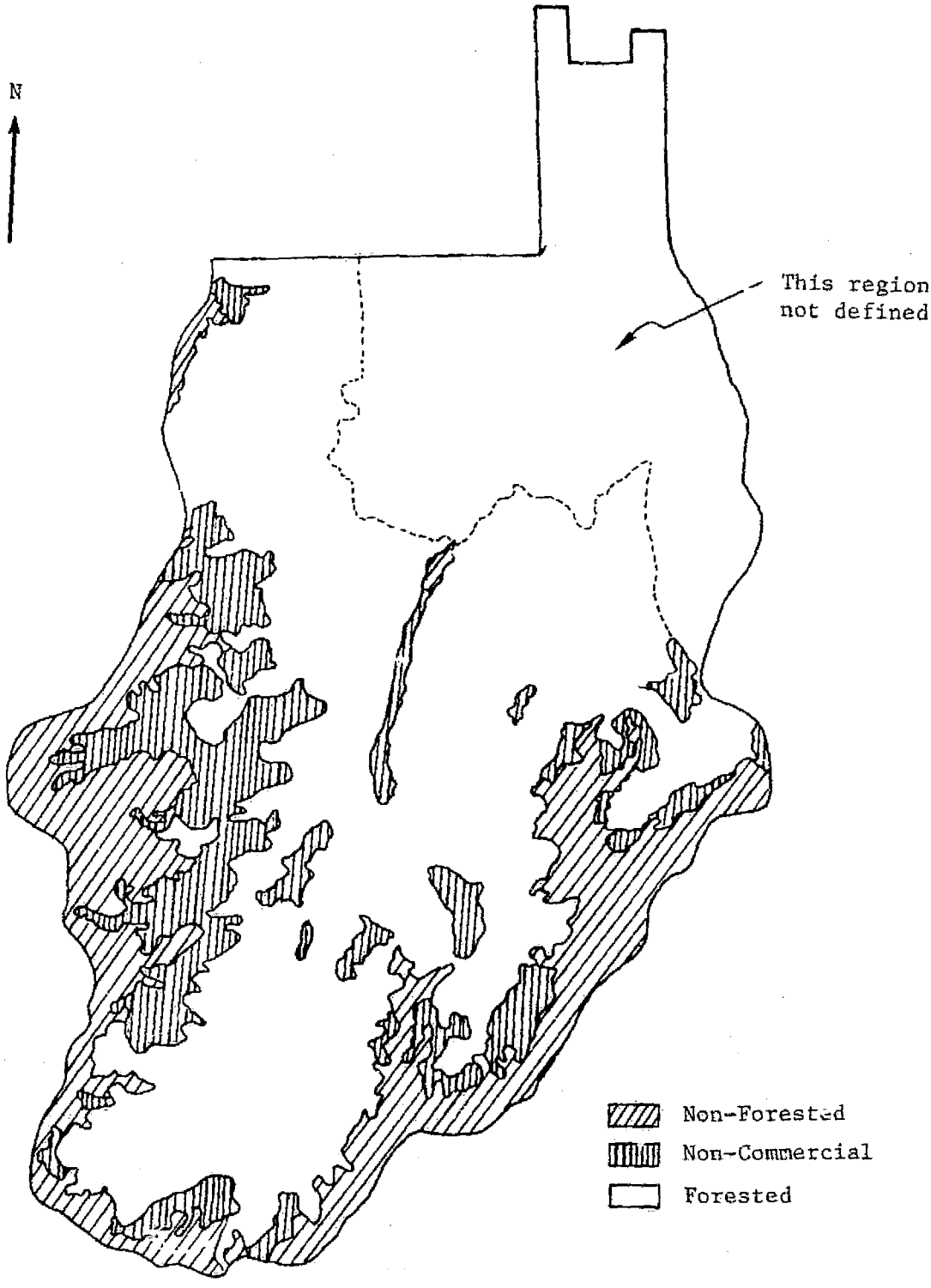



FIGURE 22. GENERAL COVER MAP FOR FRASER EXPERIMENTAL FOREST
(1957 Photo Base)

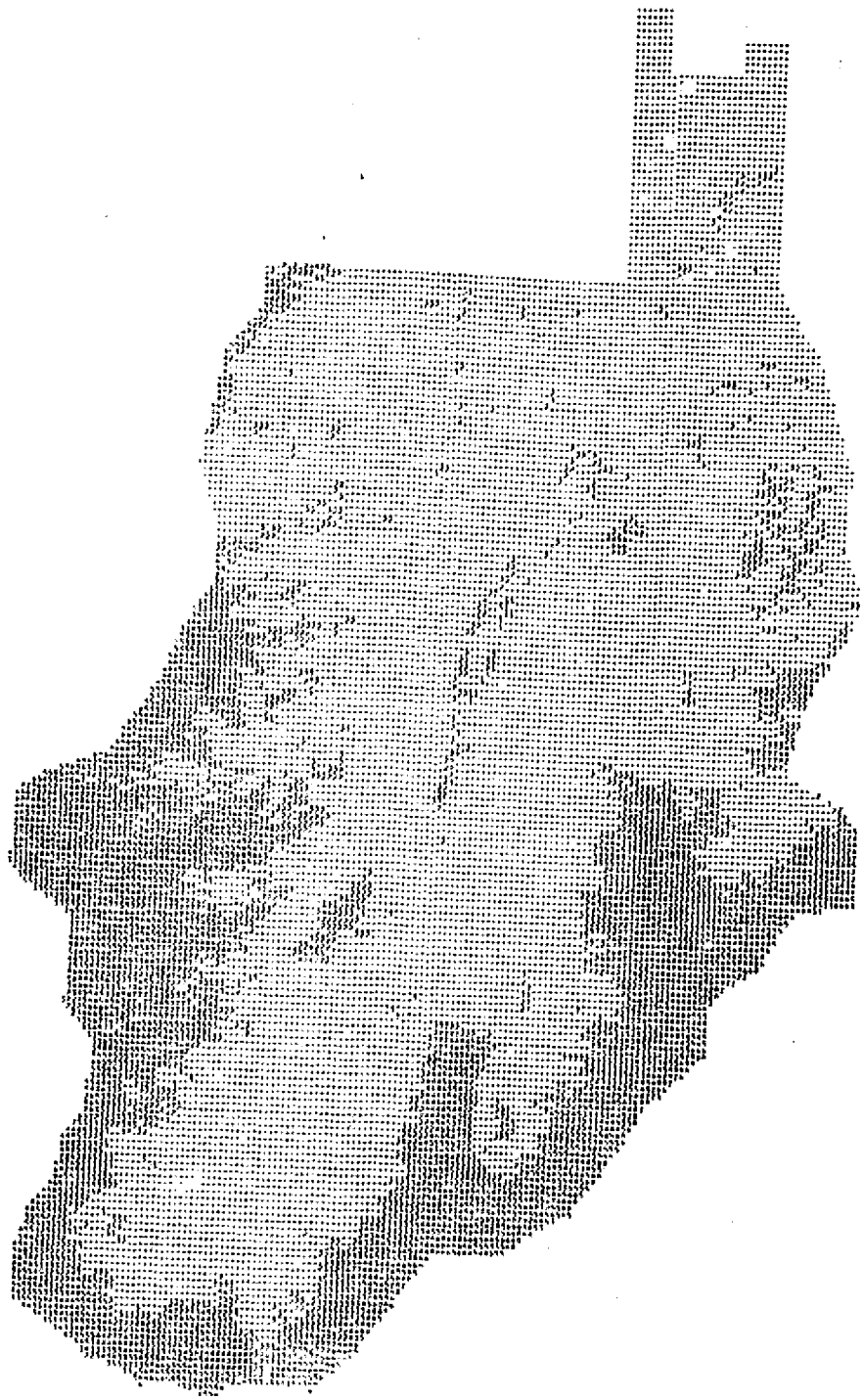


FIGURE 23. MAP OF CLUSTR OUTPUT (Preprocessed Data)

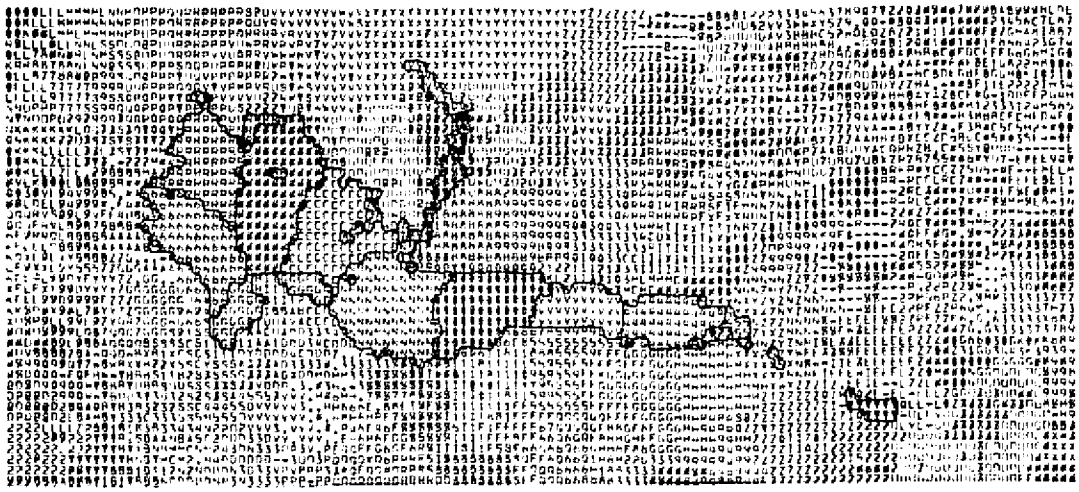
Several potential advantages of spectral/spatial clustering over pure spectral clustering are presented in Table 6.

The spatial constraint in the BLOB algorithm favors a small average size of the blobs. Such a feature is a distinct advantage for agricultural scenes, where fields tend to be relatively small, and is generally advantageous in that it helps maintain a high level of spectral similarity within blobs. Under these circumstances, however, large uniform areas tend to be divided into several semi-regularly shaped blobs of a size determined by the spatial weighting factors. An example of this phenomenon is presented in Figure 24(a).

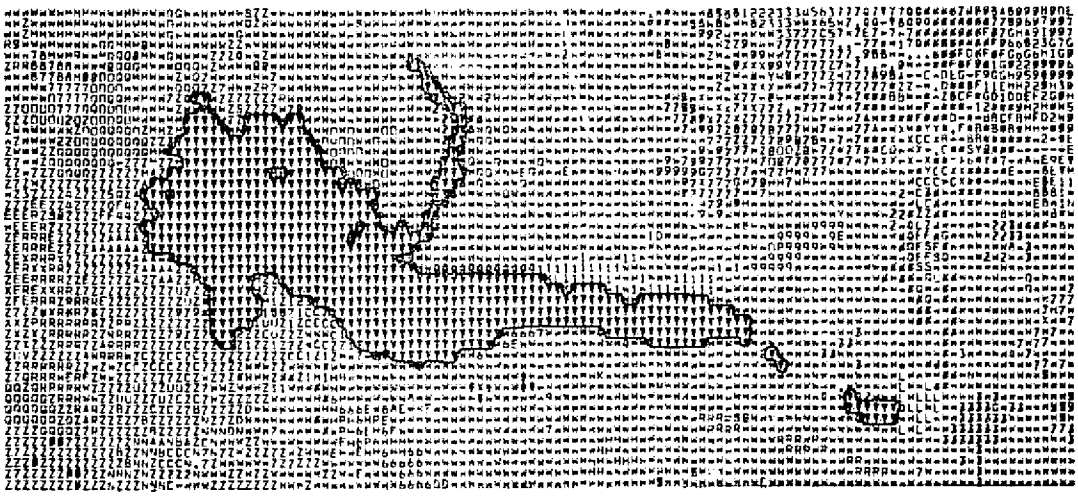
Since natural areas such as forest stands tend to be irregular in shape and spatially extensive, we concluded that efficient utilization of the BLOB technique would be enhanced through the inclusion of an additional processing step. A technique called ADJOIN was therefore developed as a part of this task.

TABLE 6. POSSIBLE ADVANTAGES OF SPECTRAL/SPATIAL CLUSTERING OVER PURE SPECTRAL CLUSTERING

- LOCALIZED SPECTRAL SIMILARITY TESTS
  - Not As Susceptible to Scene-Wide Variations
  - Can Relax Spectral Criterion to Include Some Spectral Inclusions on Localized Basis, Retaining Major Scene Features
- PRODUCES LOCALIZED SPATIALLY DISTINCT ENTITIES SUITABLE FOR TRAINING AND GROUND TRUTHING
- DATA COMPRESSION ADVANTAGES IN ALLOWING PREPROCESSING OF BLOB CHARACTERISTICS PRIOR TO THEIR CLASSIFICATION, FOLLOWED BY MAPPING OF PIXELS



(a) BLOB Map



(b) ADJOIN Map (Using Above Blobs as Input)

NOTE: Map symbols denote blob numbers only, not gray scale levels.

FIGURE 24. ILLUSTRATION OF BLOB AND ADJOIN MAP CHARACTERISTICS; LAKE GRANBY, GRAND COUNTY, COLORADO

ADJOIN operates on blobs obtained in the manner explained above. Beginning with a list of these blobs and their spectral means, and a list of neighboring (adjacent) blobs for each blob, ADJOIN uses a modified chi-squared distance function to compare each blob to its neighbors. Whenever the distance is less than a user-defined limit (which is independent of the limit set in BLOB), the two blobs are combined, a revised mean vector is computed, and the neighbor lists of the two blobs are merged. This comparison process continues for the newly defined blob, and progresses in like fashion for all the existing blobs. The result of applying the ADJOIN process is a pattern of blobs which more closely depicts the spectral patterns present (Figure 24(b)).

When ADJOIN was used for the Fraser Experimental Forest test site, the overall separation of forested from other areas was in good agreement with the stand map (compare Figures 22 and 25).

Another way of viewing the BLOB/ADJOIN process is as follows. BLOB applies a localized test which can vary the relative importance of spatial and spectral parameters. One might wish, for example, to relax the spectral criterion somewhat to minimize the "salt and pepper" appearance of some scenes. Next, ADJOIN applies a more regional spectral test of all neighboring (adjacent) blobs. The resultant product is one step removed from the global spectral similarity test of conventional spectral clustering which is more susceptible to scene-wide variations. This global test could be approximated by spectral clustering of the blob or adjoined-blob means.

It should be noted that, in spite of the theoretical advantages of the BLOB/ADJOIN process as compared with simple spectral clustering, the results of the two processes when applied to the Fraser Experimental Forest were very similar in terms of major type delineation on the mapped output. Additional studies on varied scenes would be needed to fully compare the two techniques.

NOTES: Map symbols denote blob numbers only, not gray scale levels.
Preprocessed data were used, with TAU = 8.

FIGURE 25. ADJOINed BLOB MAP OF FRASER EXPERIMENTAL FOREST

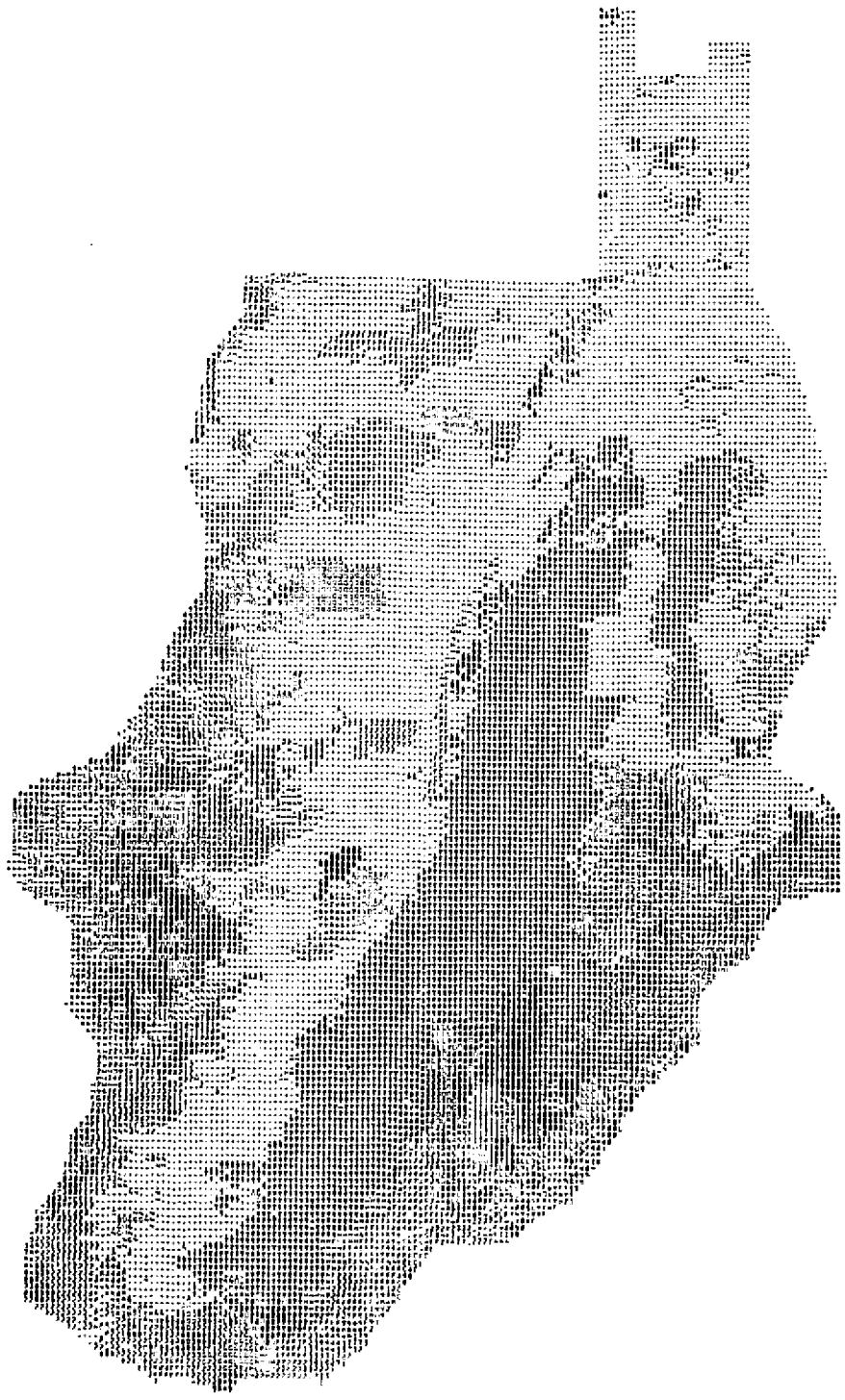
## 4.4 EXAMPLE CLUSTERING OF DATA AFTER PREPROCESSING BASED ON ANCILLARY VARIABLES

The objective of the preprocessing discussed in Section 4.2.3 was to use ancillary variables such as terrain slope and aspect and sun zenith and azimuth in a transformation to reduce the variability in signals that is associated with those variables. Using the theoretical preprocessing model based on the modified relative insolation factor (with $L_p = 0$), followed by the BLOB/ADJOIN procedure for joint spectral/spatial clustering, a demonstration of the effect of the preprocessing transformation was carried out on the Fraser Experimental Forest test site.

Figure 26 illustrates the ADJOIN output using untransformed signal values, and clearly exhibits the influence of aspect. The two major blobs in the center both contain the two forest types, differing primarily in aspect. In Figure 27, a map of the ADJOINed blobs created using the same parameter settings but operating on preprocessed data, there is one major blob instead of two, and it includes both East- and West-facing slopes. Preprocessing allows the use of tighter parameter settings without generating extra subclasses that are due only to the topographic variations. This would be of maximum benefit in scenes where one needs to separate rather small differences in spectral properties -- not really tested in this data set. The two coniferous classes, Lodgepole pine and spruce-fir, proved to be very similar spectrally with no reliable spectral discrimination. On the other hand, they were readily separable from the non-forested portions of the scene.

## 4.5 CONCLUSIONS AND RECOMMENDATIONS REGARDING INFORMATION EXTRACTION TECHNIQUES

From the foregoing analysis and processing of data that combine spectral characteristics from remote sensing with ancillary geographic data from other sources, we draw the following conclusions and recommendations.



NOTE: Map symbols denote blob numbers only, not gray scale levels.
TAU = 4.

FIGURE 26. ADJOINed MAP OF FRASER EXPERIMENTAL FOREST (Raw Data)

NOTE: Map symbols denote blob numbers only, not gray scale levels.
TAU = 4.

FIGURE 27. ADJOINed MAP OF FRASER EXPERIMENTAL FOREST (Preprocessed Data)

4.5.1 CONCLUSIONS

1. Effects of terrain topography in mountainous forested regions on Landsat signals and classifier training are significant.
   a. The aspect of sloping terrain relative to the sun's azimuth is the major cause of variability,
   b. A relative insolation factor can be defined which, in a single variable, represents the joint effects of slope and aspect and solar geometry on irradiance, and
   c. Forest canopy reflectances were found, both through simulation and empirically, to have non-diffuse reflectance characteristics; they exhibit a lesser total reflectance change in the presence of terrain slope and aspect variations than would a diffuse surface, suggesting use of a modified relative insolation factor.
2. Training procedures can be improved by stratifying in the space of ancillary variables and training in each stratum or, alternatively, training can be simplified by first transforming the data using preprocessing models based on ancillary variables.
3. Preliminary indications are that the inclusion of ancillary variables in the classification process can improve the performance of the classifier; however, the generality of this observation is yet to be established.
4. Application of the Tasselled-Cap transformation for Landsat data acquired over forested terrain should provide a viable technique for data compression and convenient physical interpretations.
   a. The dimensionality of Landsat data over forested regions appears to be the same as over agricultural areas -- two

major dimensions, Brightness and Green which have convenient physical interpretations, and a third (yellow) which may serve as an indicator of haze conditions of observation.

b. A polar-coordinate transformation to a green angle and a brightness radius in the plane of principal variation may have a better and more meaningful physical interpretation, with better decoupling of brightness and vegetation density effects, but has not been fully evaluated.

5. Clustering techniques appear useful for analysis of forest scenes.

   a. Both pure spectral and spectral/spatial technique were applied successfully to the test site data.

   b. We identified potential advantages for spectral/spatial clustering over pure spectral clustering, but the data set did not enable an adequate comparison to be made -- few differences were observed.

6. Preprocessing techniques which take advantage of ancillary data on conditions (i.e., topographic orientation) affecting the illumination of a Landsat pixel scene element can reduce the effects of irregular illumination.

   a. Scene class variability was substantially reduced, and

   b. Classification performance was improved.

4.5.2 RECOMMENDATIONS

As a result of this investigation we recommend that:

1. Further development and testing of preprocessing techniques based on topographic variables be conducted, including inputs from additional modeling of forest canopy reflectance and the effects of terrain shadowing at low sun angles.

2. Investigation and development of improved information extraction techniques for forest and rangeland applications be continued, including:
   a. Data transformation techniques for data compression and improved interpretation,
   b. Data screening and atmospheric correction techniques,
   c. Clustering techniques, both spectral and spectral/spatial, and
   d. Multitemporal techniques which take into account seasonal changes.
3. The continued investigations be conducted on other sites beyond the Grand County site which contain deciduous as well as coniferous forests and different ecosystems, to increase the base of applicability.

# APPENDIX I

## COMPO S OF A GEOGRAPHICAL INFORMATION SYSTEM

A geographical information system has been simply defined as "a data base management system with the addition of analytical capabilities"[4]. Behind this simple depiction one finds a dynamic and elaborate organization of data, information, and information extraction techniques. Figure I-1 diagrams six components of an information system illustrating their interrelationship [25].

Four basic functions required of any GIS have been divided into four categories [4]: (1) data entry and data cleaning, (2) data summarization and manipulation, (3) data analysis, and (4) data output. Examining the components of a GIS, "data specification" and "acquisition" address the first function; "data management" and data base" pertain to the first two and specifically to the maintenance and retrieval of data in a computer-based environment. The information extraction processes are carried out at the "data processing" an "dissemination" stages.

Each of these components will be addressed in detail in the following discussion.

### Data Specification

Data specification involves four basic processes [25]:

1. The establishment of specific data needs. These data needs may span a variety of data types including: land, environment, population, and administration. The selection of specific data types would be based on the system application.
2. The establishment of cross-level data needs, as well as cross-functional data needs.
3. Categorization of data types and interrelationships by topic and feature.
4. Determination of data update standards, based on the rate of data change and data growth through processing.

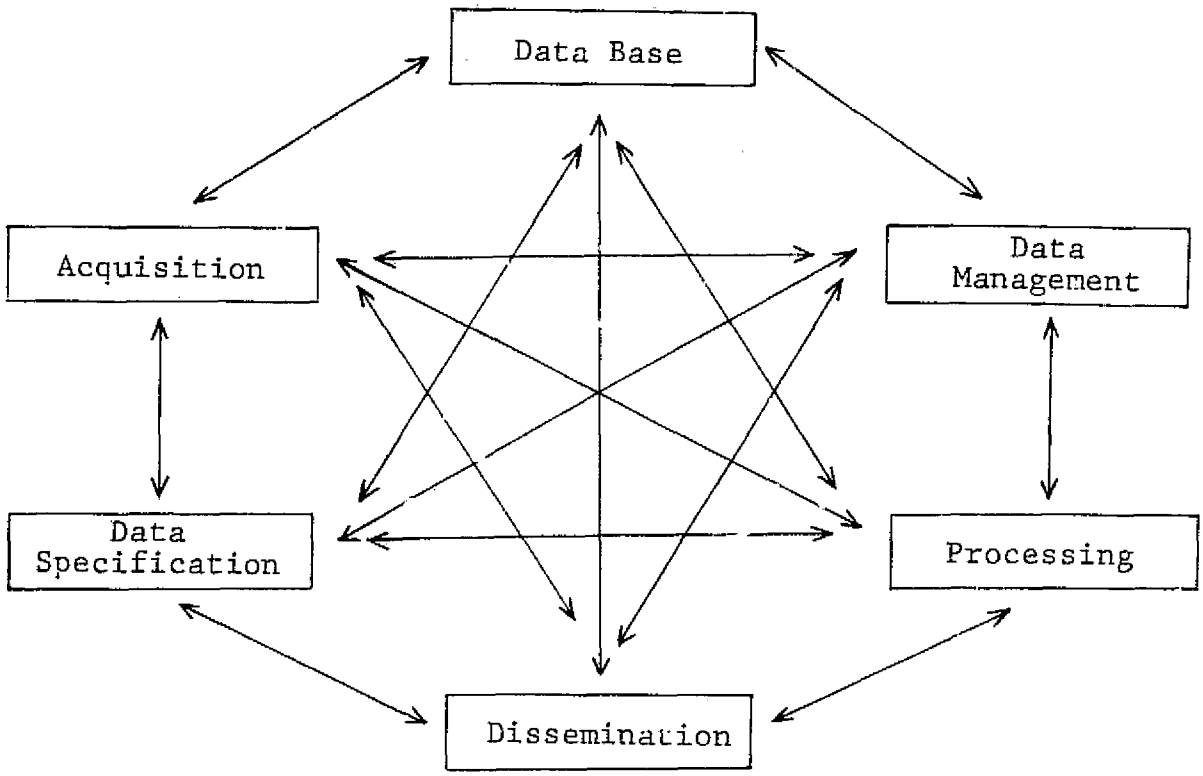



FIGURE I-1. SCHEMATIC REPRESENTATION OF THE RELATIONSHIPS BETWEEN COMPONENTS OF AN INFORMATION SYSTEM

A unique characteristic of spatially oriented data is encountered once one initiates the process of data specifications. That characteristic is the "layered" nature of geographical data. That is, every location on the ground can have associated with it a wide variety of characteristics. One system employed at the Environmental Research Institute of Michigan defined 23 variables (like land use, soil and topography) to characterize a location of approximately one hectare in size. The same coordinate references many layers of information [26].

Data Acquisition

Probably the most awesome task confronting the implementation of any information system is the gathering of data in a computer-compatible

format. This process takes on special problems when the data are geographically oriented. The tasks at hand include [27]:

1. Establishing data sources, i.e., determining which data are currently available and which data must be measured.
2. Establishing strategies for sampling.
3. Determining data computer compatibility. This may require a complicated digitization process to a standardized coordinate referencing system.

Here we are confronted with a second important spatial data characteristic. The volume of data required for even the small applications may be enormous. Spatially oriented data can be dimensioned not only by their spatial resolution, but also by their temporal resolution, i.e., rate of change as reflected in the frequency of measurement. As a familiar example, remotely sensed data gathered by Landsat are segmented into frames. Each frame is 100 nautical miles on a side containing over 28,000,000 bytes of data. These data are measured every 18 days. Approximately 20 data sets are gathered over a given site in a year, representing over 0.5 billion bytes of data. Associated with these data, one may require a variety of other information: elevation from sea level at a point or land use category.

Data Base

By "data base" we mean the collection of pieces of quantitative and qualitative information, in a retrievable format, that measures or describes features of interest. The term "data base" is often misused, as is "data bank", for the information system itself. The information content of each piece of datum is three dimensional [25]: (1) thematic, what is being measured, (2) spatial, where it is being measured, and (3) temporal, when it is measured. Each datum can function either as an analytical variable (i.e., a measurement that can take on any numerical value over a continuous or interval scale) or

a categorical variable (i.e., a descriptor or attribute that can take on a limited number of values on a discrete or nominal scale) or both. For example, multispectral scanner data are analytical, soil type data are categorical, and topographic information could be either or both.

The logical design of a spatially oriented data base includes the determination of the data layers or attributes, data interrelationships, and, due to the potential volume of data, a data sampling and segmentation strategy.

Spatial data occurs in any of three basic forms: (1) point source data, e.g., water quality data collected at discrete points along a river, (2) linear data, e.g., a street network, and (3) areal data, e.g., thematic maps over contiguous regions.

Physical storage characteristics of geographically oriented data include two basic types: (1) regular cells or grid encoded data and (2) irregular cells or linearly encoded data, though each can be encoded in a variety of ways [28]. Traditionally, systems are of one type or the other. The fact that should not be compromised, however, is that certain layers of information fall naturally into one storage type or the other. The optimum system can manage both forms of data. Let us discuss the concept of the spatial data structure a little more fully.

1. Raw Data Structure: The form in which data are acquired, e.g., soil map or MSS CCT format.
2. Computer-External Data Stoage Structure: The computer-compatible format in which the data base resides outside the computer processing unit.
3. The Computer-Internal Data Storage Structure: The format in which the data reside within the computer processing system.

The external storage structure is the format from which data are initially retrieved before processing and by which data can be disseminated to various users. As mentioned, the two basic geographic

information system external data structure organizations are line encoding and cell encoding.

In line encoding, spatial features are defined using nodes and connecting line segments. Point form data are described using only nodes; linear form data consist of nodes and connecting line segments; and areal data consist of nodes and line segments forming closed regions, i.e., polygons. Polygons need not be contiguous nor completely cover the scene of interest. The organization of the encoded nodes and line segments is generally handled through lists. Linear encoding techniques include: (1) location lists [29], (2) point-dictionaries [29], (3) DIME files [30], and (4) chain/node encoding [31]. Line encoding offers the most general type of geographic data representation [32] and is particularly advantageous in terms of computer storage requirements in describing: (1) large uniform regions of data such as state or county boundaries, i.e., regions that are large in area in comparison with the basic data cell size, (2) regions of irregular shape, and (3) features that are characteristically linear.

Cell encoding is a special form of line encoding of areal data. Cells are rectangular polygons and are usually square. Because of the regularity of the shapes, and since they generally cover an entire scene, cells can be stored as an array, rather than in a list. This form of encoding can permit an efficient way of retrieving certain kinds of data since access is done through coordinate referencing, i.e., indexing into the array, rather than searching through a list.

Grid structures include three encoding techniques [25]:

1. Sequential -- Data values are entered into cell after cell along rows or columns.
2. Compact Sequential -- Repeating data values are not stored for each cell, but stored along with a length attribute.
3. Complete Coding -- Each data value has a locational vector associated with it.

A third data storage structure type that is not always considered integral to the geographic data base are data that are not necessarily geographically oriented, but list oriented. Yet these data are so integrally related to the processing of geographical data that they should not be separated from it. Most systems manage these data in associated flat files. As previously mentioned, these data could include statistical characterizations of a particular layer of geographical data. Tables of aggregated statistical information that correspond to features of interest to the user of the system form another integral part of the geographical data base.

Data Management

We have seen, so far, that geographical systems are characterized by: (1) the spatial orientation of the data, (2) the layered characteristics of the data, (3) the potential volume of data, and (4) the variety of optimal data structures.

These characteristics especially affect the management aspect of data in a GIS. This aspect concerns all forms of data storages, raw, computer external and computer internal. Two important considerations in the management of spatial data include: (1) a segmentation strategy to divide data into manageable regions and (2) the storage and retrieval of these data segments. The latter requirement poses a twofold problem. First of all, a particular data requirement may result in the need to mosaic different segments together. Secondly, a requirement for multiple layers of data reflects a need to merge layers of data.

Whereas the need to properly manage data in its raw form is most likely the responsibility of an intervening human, the management of data in computer external into internal form and vice versa is the responsibility of a computer data base management subsystem. Let us discuss this briefly in the terms proposed by CODASYL Data Base Task

group [19]. In 1971 a Data Base Task Group (DBTG) defined the components of a Data Base Management System (DBMS). A DBMS is an integrated package of software designed to manage large amounts of various types for all phases of input, updating, retrieval and output by a number of users [33,34]. A DBMS was meant to be a middleman between a collection of data and users of that data. The data base management sybsystem of a GIS would be aware of the data structuring, and is responsible for the retrieval of these data in a manner that assures data integrity and provides an atmosphere within which application programs are independent of the external data structure.

A DBMS for spatial data must be able to perform the functions defined by the DBTG on the spatial data. It is likely that the retrieval demands of a GIS preclude the direct employment of commercially available DBMS. This was discussed more fully in Section 3.4.3. Retrieval requirements of a GIS include retrieval based on nominal data characteristics, coordinate data designation and on relational techniques. Another complexity in the retrieval of data is the requirement of accessing layers of data stored in a variety of formats. For example, whereas one layer of data is stored as a set of linear features, another may be grid in structure. It is necessary, then, to invoke a point-in-polygon algorithm which will convert the irregularly shaped data into a grid cell matrix. The need to insure compatibility in resolution size of associated grid cells is also an important concern at a data management level.

The important concept to retain at this point is that a request for a data set may not structurally correspond directly with the external structures of the data elements that comprise that data set. A data base management system carries the responsibility of providing the appropriate mosaicing and structural conversions from external to internal form.

Data Processing

The processing of spatial data is often analytic in nature and can generate new layers of spatial information that must be maintained by the data base management subsystem. Data base growth, therefore, comes not only from the specification and encoding of raw data types, but also from the processing of encoded data. Whereas a data management system attempts to preserve data base and application program independence, the nature of the application programs employed may affect the data supported in the data base.

The processing of spatial data falls into two basic categories. Processing procedures are either parallel, i.e., can be employed to all spatial elements simultaneously after determination of specific global parameters, or contextual, that is, the spatial context of a datum influences the outcome of the result.

Processing of spatial data is carried out in three steps: a preparatory or preprocessing step, a primary or analytical processing step, and a descriptive or graphic phase.

The intent of data preprocessing is not to extract information from the encoded data, but to modify the data in such a manner as to make the extraction of information more feasible or efficient.

Data preprocessing generally deals with such items as transforming raw data into some standard coordinate referencing system like Universal Transverse Mercator. This activity is termed geometric correction. A second preparatory activity might involve the analytical transformation of data. For example, many forms of spatial data, in particular those measured using remote sensors, are multivariate in nature. A principal component analysis of the data may warrant a transformation to compress the data into fewer dimensions with axes oriented in the direction of the principal components. In effect, a new layer of data is produced.

Primary processing of the spatial data pertains to the information extraction process. This involves three basic operations:

1. Feature extraction in response to a user's query within a layer of data, e.g., a discriminant analysis to determine physical characteristics of the data.
2. Feature extraction between layers of data; commonly this is accomplished through co-occurrence analysis or overlay processing, here layers of data are "intersected" to determine geographic regions that satisfy a user specified query that may be algebraic in nature.
3. Inference modeling; the information content of the data is used in conjunction with a mathematical model to project changes that may occur, e.g., an ecological system over a period of time under a set of circumstances.

Descriptive processing pertains to the aggregation of information extracted in the primary processing stage. Statistics are gathered into a format compatible to some report or tabular display. Oftentimes the sequence of data processing efforts is an interactive one. The post-processing of the data may warrant another processing approach to extract new or different forms of information.

Dissemination

Dissemination pertains to the delivery and maintenance of data and information extracted from data to the users of the system. At a local level, data are disseminated to the users through some sort of hard or softcopy interface in the form of a map or a report. That report may be generated as a response to a query language interface with the system user. It may take the form of a table, chart or graph. The standard vehicles designed to transport these data would include a line printer, a table plotter, or a video terminal with associated hardcopy unit.

Dissemination of spatial data and information both begins and ends the cycle of a geographic information system. The information extracted may be the computation of a new layer of data which is in

turn re-entered into the system for further processing and analysis, or may be a final report describing the results of data processing and analysis.

# APPENDIX II

## PARTIAL LISTING OF OTHER COMPUTERIZED GEOGRAPHICAL INFORMATION SYSTEMS

LUDA - Land Use and Land Cover Data and Analysis Program, U.S. Geological Survey, Geography Program, Mail Stop 710, Reston, Va., 22092

MLMIS - Minnesota Land Management Information System, Center for Urban Affairs, 311 Walter Library, Minneapolis, Minn. 55455

South Carolina Information and Mapping Program. S.C. Wildlife and Marine Resources Department, Marine Resources Center, Office of Coastal Planning, Charleston, S.C., 29407

PLUM - Planning and Land Use Management System, Land Use Planning Section, Department of Natural and Economic Resources, P.O. Box 27687, Raleigh, N.C., 27611

ORRMIS - Oak Ridge Regional Modeling Information System, Oak Ridge National Laboratory, Oak Ridge, Tenn., 37830

MAGI - Maryland Automated Geographic Information System, Maryland Department of State Planning, 301 West Preston Street, Baltimore, Md., 21201

REAP - North Dakota Regional Environmental Assessment Program, North Dakota State Planning Division, Bismarck, N.D., 58501

LUNR - Land Use Natural Resources Inventory, State of New York, Office of Planning Services, Albany, N.Y., 12201

PIOS - Polygon Information Overlay System, Environmental Systems Research Institute, Redlands, California, 92373

CGIS - The Canada Geographic Information System, Land Management Information Systems, Div. Land Directorate, Environment Canada, Ottawa, Canada

TNRIS - Texas Natural Resources Information System

ERIPS - Earth Resources Interactive Processing System, National Aeronautics and Space Administration, Johnson Space Center, Houston, Texas

CLUIS - South Carolina Computerized Land Use Information System, South Carolina Land Resources Conservation Commission, Clemson, S.C.

GRDSR - The Geographically Referenced Data Storage and Retrieval System, Methodology and Systems Branch, Statistics Canada, Ottawa, Canada, K1A0T6

QLINE - Image-Based Data Processing System, Environmental Research Institute of Michigan, Ann Arbor, Mich., 48107

NARIS - Natural Resource Information System, Center for Advanced Computation, University of Illinois at Urbana-Champaign, Urbana, Ill., 61801

IBIS - Image Based Information System, Jet Propulsion Laboratory, San Francisco, California

MIDAS - Maine Information Display and Analysis System, Maine Bureau of Inland Fisheries and Game

LARSYS - Laboratory for Application of Remote Sensing, Purdue University, W. Lafayette, Ind., 47906

STORET - Storage and Retrieval of Water Quality Control Data, Environmental Protection Agency, Washington, D.C.

CMS-II - Computer Mapping System, Federation of Rocky Mountain States, Denver, Colo., 80211

INTURMAP - Interactive Urban Mapping System, Department of Computer Science, University of British Columbia, Vancouver, B.C., Canada

# APPENDIX III

## ANNOTATED BIBLIOGRAPHY OF COMPUTER-BASED INFORMATION SYSTEM LITERATURE

(R. C. Cicone)

Abbitt, C. W., Earth Resources Interactive Processing System Requirements, Aeronutronic Ford Corporation, November 1975.

The Earth Resources Interactive Processing System (ERIPS) is the system currently used in processing the LACIE data base; $4.8 \times 10^9$ bytes of data are managed. This report describes data structures and formats with an emphasis on the menu directed interaction between computer and system user. The method for storage and retrieval of the data is of particular interest. Landsat multitemporal-multispectral data are stored by sample segments. Data ancillary to the Landsat data segment are stored in a historical file which also contains pointers to the location of the data segment. Retrieval of the data is done nominally by segment title. The chain of pointers is followed and the data stored on disc are read into core when processing takes place.

Alvarey, D. T., M. L. Taylor, Cartographic Data Base Hierarchy, Vol. 1: Systems Analysis and Design, PRC Information Sciences Co., McLean, Virginia, September 1974.

A system was designed and implemented that employs a hierarchical encoding scheme for the detailed description of cartographic features. Variable data resolution was achieved through the storage of geographic coordinates in a compact incremental format.

Anuta, P. E., Computer-Aided Analysis Techniques for Remote Sensing Data Interpretation, Laboratory for Applications of Remote Sensing, LARS Information Note 100675, Purdue University, West Lafayette, Indiana.

An overview of the LARSYS system is presented with a description of clustering, geometric correction and classification techniques employed in the processing of multispectral data.

Cooke, Maxfield, The Dime Geocoding System, Bureau of the Census, Washington, D.C., Census Use Study Report 4, 1967.

The Dual Independent Mapping Encoding System used by the Census Use Study stores information about a geographic entity, not only in terms of its spatial whereabouts but also with its topological relationship to other features of the map scene.

Craven, C.W., F. P. Baxter, C. R. Meyers, R. J. Olsen, C. R. Schuller, A. H. Voelker, Regional Environmental Systems Analysis, Progress Report June 15, 1971 - June 15, 1972, Oak Ridge National Laboratory, Oak Ridge, Tennessee, March 1973.

and

Durfee, R. C., ORRMIS: Oak Ridge Regional Modeling Information System, Part I, Oak Ridge National Laboratory, Oak Ridge, Tennessee, September 1974.

The Oak Ridge Regional Modeling Information System is described. Aspects covered include a thorough description of the data base, the data structures and the variety of modeling efforts that are carried out. System applications emphasized include socioeconomic, land use, ecological and sociopolitical analyses that are carried out.

DeCourcy, D. J., The Development of a Natural Resource Information System (NRIS), Vol. 1, Raytheon Corporation, August 1973.

NRIS exemplifies a typical polygon-oriented information system. The data base for this system is derived from a variety of Thematic Maps from which polygonal regions are digitized, point by point, so that they can be reconstructed within computer memory as n-sided polygons. A digital plotter is used to produce high quality maps. However, the system is non-interactive.

Elliott, M. (ed.), Proceedings of the International Conference on Automation in Cartography, Library of Congress Number 76-2164, Reston, Virginia, December 1974.

Advances in computer hardware and software technology in the area of cartography are covered. Attention is paid in the software section to geographic information system display capabilities. Polygonal data structures are reviewed. Of particular interest is a discussion by Professor Nicholas Chrisman of Harvard University wherein he discusses a chain structure for storing polygonal data elements.

Hsu, M., L. Mahi, G. Owing, The Minnesota Information System and Land Use Map, Department of Geography and Center for Urban and Regional Affairs, University of Minnesota, Minneapolis, Minnesota, ASCM, Fall 1972.

and

Hsu, M., et al, "Computer Applications in Land-Use Mapping and the Minnesota Land Management Information System", pp. 298-310 of J. Davis and M. McCullagh, Display and Analysis of Spatial Data, New York, J. Wiley, 1975.

MLMIS is a statewide information system for land-related information. These references describe the aims of the project, and the approach used in collection, assimilation, analysis and display of data. MLMIS is of particular interest in that it is an attempt to collect data from a variety of sources in a standardized cellular format.

Marble, D. F., (Committee Chairman), Computer Software for Spatial Data Handling, Review draft prepared for USGS under Grant #14-08-0001-G-215, March 1976.

A compilation of computer software available throughout the U.S. that is pertinent to the task of manipulating and displaying spatially distributed data that are stored in polygonal or cellular format. This volume is not yet in publication, though it is expected to be approved for publication within a year. It is available on microfilm through the Resources and Land Investigation Program at USGS National Center in Reston, Virginia.

NASA Earth Resources Survey Symposium, Volume 1.B: Geology, Information Systems and Services, Johnson Space Center, NASA, June 1975.

Several hardware and software solutions to typical problems encountered in the design of remote sensor-oriented, geographically-based information systems are presented in the Information Systems and Services Session. The complexity of some of these problems require the development of specialized hardware. A sampling of systems is found among the following papers:

Schaller, E. S., "Image 100 - The Multispectral Image Processing System", I-11.

Quinn, M. J., "Screwworm Eradication Data System (SEDS)", I-13.

Johnson, R. H., "M-DAS SYSTEM for Multispectral Data Analysis", I-16.

Henze, J., and R. DeZur, "Interactive Digital Image Manipulation System (IDIMS)", I-23.

NASA Earth Resources Survey Symposium, Volume 1-A: Agriculture, Environment, Johnson Space, NASA, June 1975.

Several presentations made on the behalf of state and local users of geographically-based information systems are of interest especially in depicting the kinds of information the user at this level is interested in and in the variety of ways their respective systems have been designed to meet their needs. Of interest are the following articles:

Shoeder, R. H., R. H. Cartmill, "South Louisiana Environmental Information System", C-5.

Sizer, J. E., "Remote Sensing in Minnesota Evaluation of Programs and Current Needs", S-9.

Baldridge, P., "Remote Sensing in the State of Ohio", S-10.

Nez, G., A Regional Land Use Survey Based on Remote Sensing and Other Data, Quarterly Report, Federation of Rocky Mountain States, Inc., 10 April 1976.

The Federation of Rocky Mountain States organizational structure is described and a description of the progress being made in the development of CMS-II, the Composite Mapping System, is documented. This system is a grid-oriented system which features line printer mapping of a variety of geographically-based information with the capability of merging remotely sensed data and ancillary data. Available documentation includes the following:

CMS-II Users Manual, Federation of Rocky Mountain States, Inc., March 1976.

CMS-II System Documentation, System Version II, Federation of Rocky Mountain States, Inc., August 1975.

Philips, R. L.,"Computer Graphics in Urban and Environmental Systems", Proceedings of the IEEE, Vol. 62, No. 4, April 1974, pp. 437-452.

An excellent overview of the key role played by computer graphics in urban, regional and environmental information systems. Specific examples are cited with recommendations given for research in the area of data base design and display techniques.

Proceedings of the Tenth International Symposium on Remote Sensing of Environment, Environmental Research Institute of Michigan, Ann Arbor, Michigan, October 1975.

Hardy, E. E., L. E. Hun, "Testing Low Cost Interpretation Systems for Updating Land Use Inventories", Department of Natural Resources, Cornell University, Ithaca, New York, Session 7.

Halpern, J. A., L. D. Alexander, D. M. O'Regan, "Application of Remote Sensing Data to Geographic-Based Information Management Systems", Session 7.

Kriegler, F. J., M. F. Gordon, R. H. McLaughlin, R. E. Marshall, "The MIDAS Processor", Environmental Research Institute of Michigan, Ann Arbor, Michigan, Session 11.

Crouch, R. G., J. P. Dangermond, "The Use of Remote Sensing Imagery and the PIOS System in Land Use Studies at the Southern California Edison Company", Session 13.

Guinn, C., "Development of a Land Information System -- Practical Guide to Concept Formulation and Design", New York State Division of State Planning, Session 13.

Schwartz, E. L., "Louisiana Comprehensive Planning Information System: Compilation on Utilization of the Data Base", Louisiana State Planning Office and Governmental Services Institute, Louisiana State University, Baton Rouge, La., Session 13.

Sayn-Wittgenstein, "Landsat Applications in Canadian Forestry", Forest Management Institute, Canadian Forestry Service, Ottawa, Ontario, Canada, Session 16.

Peuquet, Donna J., Raster Data Handling in Geographical Information Systems, Geographic Information Systems Laboratory, State University of New York at Buffalo, July 1977.

A discussion of the raster grid approach to processing of spatial data is discussed. This paper is especially conscious of the importance of a DBMS approach to the management of geographical data.

Ray III, R. M., Summary of Illiac IV - ARPA Network Multispectral Image Processing Research Activities, Final Report, CAC Document No. 168, NASA Grant NGR 14-005-202, Center for Advanced Computation, University of Illinois at Urbana-Champaign, August 1975.

The research reported in this document focuses on the implementation of Illiac IV - ARPA Network software systems for computer assisted interpretation of multispectral earth-resources imagery.

The report overviews multispectral image interpretation algorithms implemented and describes peripheral software systems developed to enable decentralized access to Illiac IV image analysis capabilities for a community of users including NASA, USGS, and USDA.

Swain, P. H., (ed.), Symposium on Machine Processing of Remotely Sensed Data, June 29 - July 1, 1976, Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana, 1976.

This volume of symposium proceedings includes several informative presentations of systems that have been designed to handle remotely sensed data. The articles of interest include the following:

Bryant, N. A., A. L. Zobrist, "IBIS: A Geographic Information System Based on Digital Image Processing and Image Raster Data Type", Jet Propulsion Laboratories, 1A-1.

Haralick, R. M., W. F. Bryant, G. J. Minden, A. Singh, C. A. Paul, D. R. Johnson, "KANDIDATS Image Processing System", University of Kansas, 1A-8.

Pope, A. E., D. L. Truitt, "The Earth Resources Interactive Processing System (ERIPS) Image Data Access Method (IDAM)", IBM, 1A-18.

Erickson, J. D., R. F. Nalepka, "PROCAMS: A Second Generation Multispectral-Multitemporal Data Processing System for Agricultural Mensuration", Environmental Research Institute of Michigan, Ann Arbor, Michigan.

Tomlinson, R. F., (ed.), Geographical Data Handling, International Geographical Union, Ottawa, Canada, August 1972.

A publication of the International Geographical Union Commission on geographical data sensing and processing for the UNESCO/IGU Second Symposium on Geographical Information Systems. This two-volume work describes geographical information systems both conceptually and structurally using state-of-the-art information systems as examples in setting guidelines for future systems and identifying needed research.

Tomlinson, R. F., (ed.), Data Handling Guidebook, Office of Land Use and Water Planning and International Geographical Union, Reston, Virginia, July 1975.

The IGU was commissioned through Argon National Laboratories to produce manuals reviewing the establishment of geographic information systems for state and local governmental users. This work

represents a thorough overview of geographically based information systems as researched by Prof. Tomlinson and five other contributing authors. Only 500 copies were originally published though future editions are envisioned.

Tomlinson, R. F., H. W. Calkins, D. F. Marble, Computer Handling of Geographic Data, Unesco Press, Paris, France, 1976. ISBN 92-3-101340-8

An examination of five geographic information systems is carried out. These systems include the Canada Geographic Information System (CGIS), Polygon Information Overlay System (PIOS), Minnesota Land Management Information System (MLMIS), Land Use and Natural Resource Inventory of New York State (LUNR), and Oak Ridge Regional Modeling Information System (ORRMIS). The purposes and objectives of each system are described as well as a history of the systems implementation. System organizational structures, data structures and output products are presented.

An experiment was conducted employing each system to determine its comparative efficiency.

Warntz, W., The Redbook, Harvard University Press, Cambridge, Massachusetts, 1971.

Selected projects of the Laboratory for Computer Graphics and Spatial Analysis are described including the mapping systems SYMAP and CALFORM. SYMAP offers a particularly good example of a polygon encoding system termed 'location lists', wherein a geographic entity is described by specifying coordinates around its perimeter. CALFORM encodes polygons in a point dictionary format. This format contains a list of coordinate values for the whole map and polygons are built by referencing the dictionary of points in an appropriate sequence.

Additional References:

Abram, P., The Role of ERTS in the Establishment of a Nationwide Land Cover Information System, NASA-CR-141045, ECON Incorporated, Princeton, N. J., 31 October 1974.

Barr, B. G., et al, The Application of Remote Sensing to Resource Management and Environmental Quality Programs in Kansas, Annual Report, NASA-CR-148325, July 1976.

Denker, K. J., "A Framework for Encoding Spatial Data," Geographical Analysis, Vol. 4, 1972, pp. 98-105.

Gale, L., "Recommendations on the Type of Coordinate System for Correlating Files in a Data Bank System", Canadian Surveyor, Vol. 23, 1969.

Additional References (Cont'd):

Gardner, J., A Study of Environmental Monitoring and Information Systems, Iowa City, Institute of Urban and Regional Research, 1972.

Herring, B. E., Development of Alabama Resources Information System, ARIS, NASA-CR-144342, Auburn University, 1 May 1976.

Reeves, R. G., (editor), Manual of Remote Sensing, "Remote Sensor Data Systems, Processing and Management", Chapter XII by D. Steiner, pp. 611-786, American Society of Photogrammetry, Falls Church, Virginia, 1974.

Rogers, R. H., et al, Computer Mapping of Landsat Data for Environmental Applications, Special Report, Bendix Corporation, November 1975.

Schell, J. A., On the Development of an Interactive Resource Information Management System for Analysis and Display of Spatiotemporal Data, NASA-CR-142241, Texas A&M University, December 1974.

Tobler, W., "Geographical Ordering of Information", The Canadian Geographic, 7, 4(1963), pp. 203-205.

# APPENDIX IV

## DATA BASE RECOMMENDATIONS

This appendix provides a supplemental discussion of recommendations to complement Section 3.5, Data Base Requirements for the Incorporation of Remotely Sensed Data in a USFS Forest and Rangeland Information System. Four topics are addressed: data structural characteristics, data processing environment, data requirements, and processing requirements.

### IV.1 DATA STRUCTURAL CHARACTERISICS

Two features of the generalized geographical information system relate to the structural characteristics of the data base:

1. A variety of spatial and object-oriented data structures are supported.
2. The storage and retrieval of these data is carried out not by a system user or application program, but by an intervening data base manager software subsystem responding to user and application program requests.

Various aspects of the data base structure are described in the next two sections.

#### IV.1.1 DATA STORAGE MODEL

Figure IV-1 models the basic computer-external data structures of a geographical information system and their interaction with the data base manager. Most elements of this diagram have been described in the main body. Our purpose here is to discuss formatting techniques recommended in Section 3.5.1. First let us distinguish hardware formats from software formats. Hardware formatting pertains to the physical structuring of data on drum, disc, and tape. That is, the physical arrangement of these data on hardware storage devices. It is beyond the scope of this document to discuss hardware formatting. Software formatting pertains to the

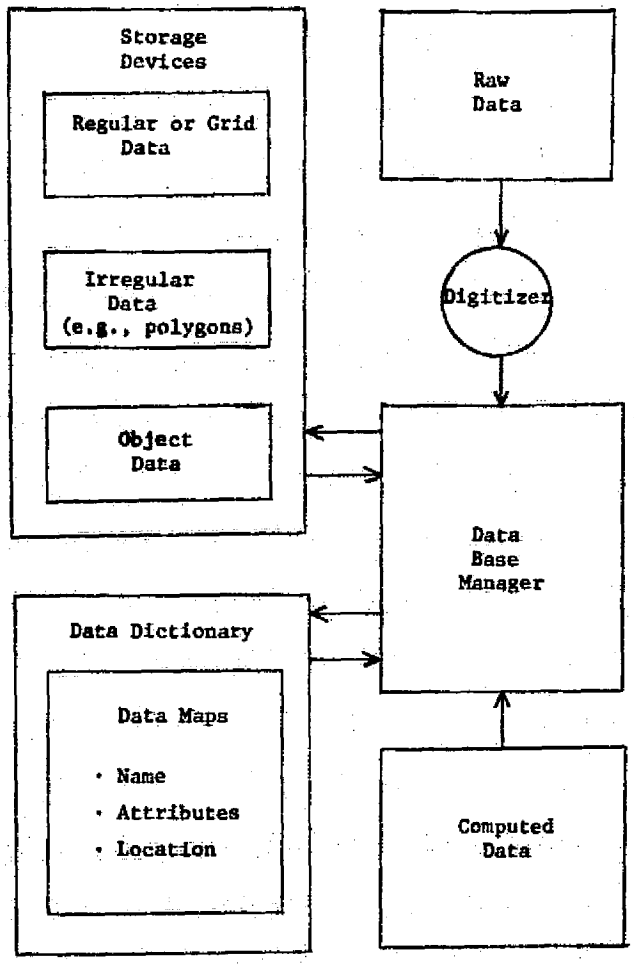



FIGURE IV-1. DATA STRUCTURE SCHEMATIC FOR GEOGRAPHICAL INFORMATION SYSTEM

the logical structuring of the data. That is to say, the inherent data characteristics (e.g., geometric projection) and structural layout (e.g., polygon or grid) that are important to a user or program irregardless of its actual physical location on hardware storage devices.

A particularly important feature of Figure IV-1 is the data dictionary. This dictionary provides a data map for each element of the data base to the data base manager. This data map provides the basic information required so the data base manager can carry out the storage adn retrieval of the data. A data map would be created each time a new data element was loaded into the data base. Important attributes include:

data type, specific formats, spatial features like resolution and coverage, and temporal attributes.

IV.1.2 DATA STRUCTURAL ELEMENTS

As has been mentioned, the incorporation of remotely sensed data in a USFS GIS will result in the processing of these data with layers of data that are derived from sources other than remote sensors. A careful organization of the layers of data to expedite the overlay process is essential. The following is a discussion that pertains to a number of important considerations.

Computer Compatibility

Processing in a computer environment requires data in digital format. Information collected by remote sensors, for example, Landsat 2, is available in both digital and image form. Other data, however, are not available in digital form, requiring a digitization process. Numerous examples of digitization techniques are available. The National Cartographic Information Center (NCIC) employs a system involving polygonal techniques utilizing table digitizers [18]. There are examples of manual digitization techniques wherein a grid is overlayed on a thematic map and values within each grid are coded by an analyst [26]. The Wildland Resource Information System developed by R. M. Russell, D. A. Sharpnack and E. I. Amidon in Forest Region 5 employs and utilizes an automated scanner polygon line input system [10]. This system might serve as a model for digitization requirements for a USFS geographic information system.

Data Attributes

Four important data attributes must be established as standards common to each layer of spatial data. These include geometric orientation, geometric projection or map rectification, data registration, and, in the case of cellular data, resolution size.

1. Geometric Correction and Registration of Layers

In order to carry out simultaneous analyses, the various layers of data must be accurately registered to each other, preferably a

standardized geometric projection and orientation. These operations could be performed either when the data were entered into the data bank or upon retrieval for processing. In the latter case, each data layer would be stored in its own particular format. The former procedure may be more efficient where the same data are used frequently.

To illustrate some of the differences to be encountered, we compare the formats of Landsat data and the digital terrain elevation data currently available through NCIC [18]. Landsat data are skewed by Earth's rotation, scan lines are rotated by several degrees from a true West to East ground scan, and scan lines progress from North to South over a frame of data. The terrain elevation data are stored in profiles, analogous to scan lines, that have South to North orientation and progress from East to West. Thus, correction and registration of these data layers involves a complicated resampling. In the future, Landsat data may become available in geometrically corrected form.

Spatial data provide planar representations of the Earth's surface. Implicit in any such geometric projection are distortions and locational errors. Numerous map projections are available to chose from including Universal Transverse Mercator (UTM), latitude-longitude, space-oblique-mercator (SOM) and others.

2. Pixel Resolution

Cellular data formats assume that each element or pixel spans a rectangular area on the earth's surface. The pixel's size is termed pixel resolution. Landsat data are provided in pixels of size 57 x 79 meters. Most available topographic information from NCIC was digitized with 63 x 63 meter resolution. It is advantageous in an integrated data base to standardize the available resolution sizes. One approach taken by the Oak Ridge National Laboratory in the system ORRMIS [35] was to insist that layers of data be resolved in sizes that were even multiples of one another. This expedites the ability to respond to a user request for a particular resolution size. Pixels are either aggregated or split into

the appropriate resolution. The user is restricted, however, to requests that are integer multiples of a basic size. Other systems like the Minnesota Land Management Information System [8] permit only one size element (40 acre resolution for MLMIS).

The incorporation of remotely sensed data introduces yet another complication. Whereas data collected locally can be digitized to a specific size resolution, data gathered elsewhere carry their own implicit resolution, size, e.g., current Landsat at 57 x 79 meter and Thematic Mapper projected to be 30 x 30 meter. No guarantee can be made that the pizel sized will be multiples of one another. Hence, resampling of one or more of the different data types would likely be required.

Data Segmentation

Spatial data coverage has both a spatial and temporal component. Landsat data, for example, are available through the EROS Data Center in frames that are 100 nautical miles square. Data can be collected over a geographical location every 18 days. A careful strategy must be determined in order to establish which geographical regions and which time periods are required for purposes of information extraction. The incorporation of spatial data in an operational geographical information system may be carried out in a manner that results in wall-to-wall spatial spatial coverage (that is, all of a geographical region like a state) or by applying some sampling criterion that selects particular subsets of interest. In establishing an overall storage strategy for these data, two questions must be considered. First of all, what will be the retrieval demands. Secondly, what is a reasonable unit of storage.

To address the first question, the processing of spatial data in general, and remote sensor data in particular, requires retrieval of polygonal windows of data. For example, a retrieval request may be for a particular county or a particular stand of trees within a forest. Access to this data would be through a specification of the coordinates of the polygon's vertices.

On the other hand, spatial data over extensive regions cannot be stored contiguously over the whole region. Device limitations preclude this. Hence, it is more appropriate to store these data in segments. As an example, the Earth Resources Information Processing System (ERIPS) employed in the Large Area Crop Inventory Experiment (LACIE) samples data in disjoint 5 x 6-mile segments [20]. Retrieval of these data from the data base is conducted by loading any desired segment in its entirety into core storage. Hence, arbitrary polygonal retrieval from the data base is not supported. However, polygonal retrieval of the data in core can be carried out. As another example, the USFS maintains a data base in Region 2 supported by the Land Inventory Mapping System (LIM) which employs a segmentation strategy for grid-orietned data that is particularly designed for typical forestry applications [10]. LIM divides a forest map a priori into a matrix of 'quads' with permanently assigned quad numbers. Each quad is at most 210 grid cell rows and 128 grid cell columns in size. This feature permits ready joining of data, i.e., mosaicing, on geographical fragments of the forest.

Segmentation of remote sensor data into quads would be an advantageous approach for purposes of retrieval. Retrieval of data from arbitrary polygonal boundaries could be carried out by mosaicing quads and flagging cells within a polygon. This activity must be carried out with caution. Landsat data are affected by atmospheric conditions, sun angle, and viewing angle. These conditions vary from frame to frame and even within a frame. The mosaicing of Landsat data with varying conditions will lead to difficulties in data analysis unless careful preprocessing of these data to minimize angular and atmospheric effects are carried out. At a minimum, a data set composed of several quads should contain information specifying each cell's Landsat frame and associated attributes.

Segmentation of linear data can be carried out in a similar fashion to grid data. Mosaicing of linear data is termed edge matching. As in cellular data, edge matching must account for errors implicit in the map projection employed.

Spatial Data Encoding

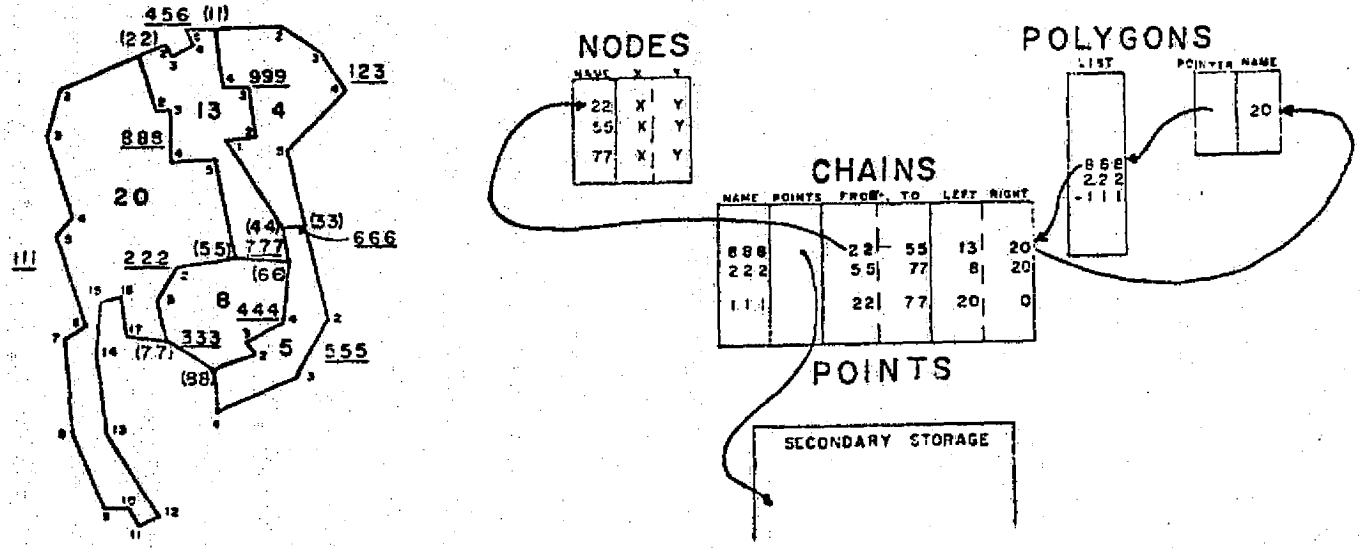
Encoding techniques employed for spatial data were discussed in Appendix I. Encoding refers to the format structures employed upon digitization of raw data. Remotely sensed data is most often provided in digital form. However, other layers of spatial data may need to be digitized. Two encoding techniques of particular interest are described in fuller detail here. The advantages afforded by each warrant their careful consideration when devising spatial data structures.

1. Compact Sequential Coding

Compact sequential coding is a cellular encoding technique that provides not only the particular computational advantages of cell encoding but also a substantial savings in the storage requirement. With compact sequential coding, data values are also entered into cell after cell along rows or columns; except that repeating data values in adjacent cells are not stored; instead, the repeated value is stored along with its number of occurrences (i.e., a length attribute). Since pixels can be accessed by coordinate referencing, retrieval of data can be carried out efficiently. This is done while simultaneously minimizing stoage requirements that can make sequential cell encoding impractical.

2. Chain/Node Encoding

The recommended Chain/Node Polygon Storage Concept was developed at the Harvard Laboratory for Computer Graphics [31] and is illustrated in Figure IV-2. A traditional approach to the digitization of polygons is to place each vertex of every polygon in a list format. This is the so-called location list. In contrast, the chain/node storage




**DEFINITIONS:**

Node -- Junction of 3 or more line segments

Chain -- Vertices between 2 nodes

Polygons defined by a list of linking nodes to chains

**ADVANTAGES:**

Efficient manual digitization

Compact storage

Topological information can be conveyed

Linear and point source data can be included by adding new pointer lists

**FIGURE IV-2. THE CHAIN/NODE POLYGON STORAGE CONCEPT (Developed at Harvard Laboratory for Computer Graphics)**

method defines polygons by chains and nodes and a list that associates or links the chains and nodes from each polygon.

A node is a point of junction of three or more line segments; for example, two specific nodes on Figure IV-2 are labeled (22) and 55). A chain consists of the vertices between a pair of nodes, for example, Chain 888 between (22) and (55). The polygon labelled 20 would consist of Chains 888, 222, and 111. These are linked together with corresponding nodes, as is illustrated on the chain structure diagram on the right of Figure IV-2. A seemingly complicated structure, the chain/node storage concept offers several distinct advantages over other storage approaches. Since Chain 888 is a boundary shared by Polygon 20 and Polygon 13, one need digitize and store only one occurrence of this chain. Hence digitization effort is cut in half, as is the computer storage requirement. Furthermore, as is illustrated, topological information can be conveyed by storing attributes in the chain file indicating what is on the "left" and "right" of the chain. In a forestry application, for example, stand type polygons may be resolved to describe not only stand types but also stand density. If stand density is not of interest, one need only ignore the chains that separate similar types into different densities.

Another advantage of the storage concept is the ability to describe not only data that are polygonal in nature, but also linear and point source data in the same logical structure. Linear data, e.g., river networks, are simply unclosed polygons. Point source data, e.g., water quality samples, are nodes. Hence, utilizing the same structure one can describe a watershed through a description of the polygonal soil patterns, linear drainage patterns, and point-source spring locations.

## IV.2 DATA PROCESSING ENVIRONMENT

Processing of remotely sensed data can be carried out in a manner that insures user/data structure and program/data structure independence. This section will address a set of desirable design components

of a data processing environment for remote sensing data. Implementation of a similar system has begun at the Environmental Research Institute of Michigan.* A unique characteristic of the design is that the same environment can be employed in processing of both spatial and object-oriented data layers. This gives the system the generalized character discussed in the main body of this report. The intent of this section is not to provide a detailed system design, but simply to discuss the essential design components.

### IV.2.1 MODEL PROCESSING ENVIRONMENT

Figure IV-3 is a schematic model of a modular data processing environment for a geographical information system. A processing monitor acts

* See Acknowledgements.

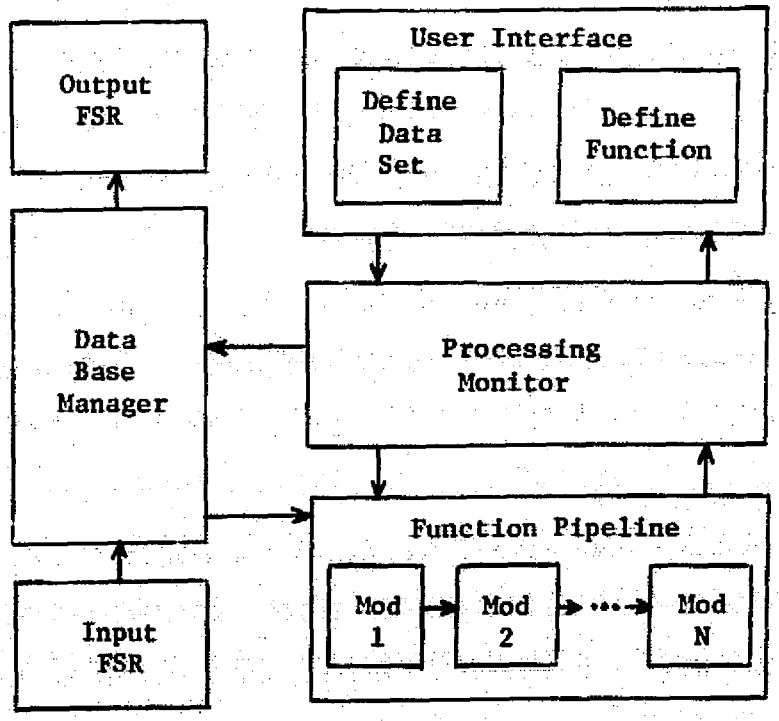



FIGURE IV-3. MODULAR GIS DATA PROCESSING ENVIRONMENT

in the role of job controller interfacing with the user, the data base manager, and application programs. User requests for information would be directed to the processing monitor through a command language processor. Data storage and retrieval is through the data base manager which invokes appropriate input and output format service routines (FSR's) to interface with data stored on external devices. The processing of the data is carried out by function modules. Functions can be performed in series, i.e., can be 'pipelined', to minimize the overhead of data retrieval. The following sections define and discuss the interfaces between these design components.

IV.2.2 DISCUSSION

User/Monitor Interface

The user ideally communicates with the processing monitor through a command language interface. A command language processor would parse a user query and invoke the monitor appropriately. A user query is comprised of the following four basic pieces of information.

1. The Specific Layers of Data to be Manipulated

Data layers would be specified nominally employing the name specified when data is loaded.

2. The Spatial Region of Data to be Processed

Specification of the processing region could be done nominally or by specification of an arbitrary polygonal boundary. For spatial data an example of nominal reference may be the declaration of a particular quad, county, or forest. Internal tables for such mapping would have to be available in the Data Dictionary available through the Data Base Manager. Subsets of object oriented data layers could be specified by some algebraic expression, for example, "all wheat signatures" or "all corn signatures and fields over 40 acres".

3. The Specific Process to be Carried Out

The user need for information requires a specification of what kind of information is required or what processing functions

must be invoked. Examples may include CLASIFY, OVERLAY, LOAD, GRAPH and so on.

4. Modifiers

Both the processing region and information required may be modified by the user. For example, a region may be invoked at a user specified resolution size. A process like CLASIFY may require more specific information with regard to the specific technique employed, e.g., quadratic rule, linear rule, etc.

Construction of a Function Module

An application program, or function module is constructed in a manner so as to carry out a specific task on the specified data set. The function module has two basic design features.

1. Data Primitive Declaration

Every application program written will require data in some computer-internal format. However, modules do not access data in computer-external form. This is the function of the Data Base Manager. Each function does declare, however, one or more data primitives. Data primitives are globally recognized entities. They define the data type required by the module. Each data primitive would carry an implicit storage structure. Modules that are pipelined would define compatible data primitives. These primitives are declared to and managed by the processing monitor. Example primitives include:

a. Cell Raster -- A cellular structure, one line of pixels composed of the various data layers.
b. Cell Region -- A rectangular region of pixels.
c. A Signature -- The statistical means vector and covariance matrix of signals from a specific spatial feature.
d. A Polygon -- A linear structure defining an enclosed spatial region.

A data primitive is not necessarily a complete data set. However, it is a unit that defines a basic feature of a particular layer of data. For

example, Landsat data usually are composed of a series of cell rasters, hence a module can perform a function on each raster until the job is completed for all of them.

2. Processing Monitor Interface

Each function module makes the basic assumption that when it requires data, it is provided the appropriate data in the expected computer-internal format. The processing monitor's role is to initiate each module, insuring that the user-declared data requirements and the function data requirements are compatible. Each module would, therefore, be composed of a set of steps. Each step is initiated by the monitor and constitutes a particular phase to which the monitor or data base manager may have to respond. Four important phases include: (1) the module's declaration of data primitives, (2) initialization of global data-set parameters before the processing of a data region, (3) repeated processing of data primitives until the entire region is processed, and (4) post-processing required after the global data region has been processed. At each phase the monitor would need to respond appropriately. For example, at the third phase the monitor repeatedly invokes the DBM to retrieve the next data primitive to be processed.

Data Management Interface

The data primitives, defined in the previous section, are controlled by the data base manager through input and output format service routines. The processing monitor invokes the data base manager with information corresponding to the user data requirements and the function module data requirements.

The first function of the data base manager is to insure data type compatibility. A user may have requested information that cannot be determined from the data set specified. For example, the user may have queried, "What is the determinant of a polygon?" The requested data type is a polygon, the required data primitive could be the covariance matrix of a signature.

The next function of the data base manager is to invoke the appropriate FSR's. Format service routines are a set of programs at the DBM level that interface computer-external data structures with computer-internal data primitive structures. Input FSR's supply data from the data base to a module in the structure of a data primitive. Output FSR's convert data from internal form to external format. Output FSR's in effect load or replace data in the data base or produce an output product for dissemination.

The data base manager must manage requests for data types that differ from the external storage characteristics of the data layer in the data base. For example should data, say a soil association map, be stored as linear features, the request for soil association as a cell raster would require the invocation of an FSR capable of carrying out a conversion through some point-in-polygon routine. Another situation may involve a user request for cellular data at a resolution different than the data layer resolution size. If the request is an acceptable one, the appropriate aggregation or splitting procedure must be carried out before passing the data into the storage area assigned to the data primitive.

## IV.3 DATA REQUIREMENTS

This section will describe certain data requirements for the incorporation of remotely sensed data in a USFS geographical information system. The potential volume of remote sensor data available results in a need to define a data sampling strategy. Research has been carried out in an attempt to analyze and recommend sampling strategies [36]. This subject is beyond the scope of this report but of enough critical concern to warrant mention. As mentioned previously, the acquisition of data for a geographical information system demands not only specification of the data types, but also of its spatial and temporal characteristics.

A data base incorporating remotely sensed data is constructed not only of data acquired from external sources, as with satellites, but also from internal sources. That is, data from internal sources are manipulated within the processing system to create additional data. Requirements for data from external and internal sources will be discussed.

IV.3.1 DATA ELEMENTS ACQUIRED FROM EXTERNAL SOURCES

Two groups of data are presented. The first represents a minimum set of requirements for effective information extraction utilizing remotely sensed data. The second group includes a number of data elements that could provide greater versatility, enhance and simplify information extraction capabilities.

Group 1

1. Landsat Digital Data:

Landsat represents state-of-the-art in remote sensor data collection. As an arbitrary satellite, it can collect and transmit data at regular intervals in a resolution size that may satisfy some forest inventory needs. The launch of the Thematic Mapper will open another invaluable source of data.

2. Topographic Data:

Topographic information is required wherever terrain features vary rapidly. Varying terrain results in irregular solar illumination of the ground even at high sun angles and in shadowing at lower sun angles. This irregular illumination and shadowing affects Landsat data so as to complicate the inventory information extraction process. Methods have been employed to preprocess Landsat data to minimize effects of irregular illumination (Section 4). Other studies have resulted in terrain models that can be employed to flag shadowed regions. Studies have shown that terrain information enhances one's ability to infer forest overstory classes when used in conjunction with Landsat data [37]. The topographic features required are slope, aspect and solar

insolation. These can be derived from terrain elevation [19]. Digital terrain data has been made available through the National Cartographic Information Center. It should be noted that terrain data are an integral element of applications like soil erosion modeling, independent of its usefulness in conjunction with remote sensing data. This applies to all of the recommended data layers that are derived from other than remote sensor sources.

Group 2

1. Field Measurement Data:

The analysis of Landsat data is greatly enriched when prefaced by the analysis of the reflectances of features of interest. The modeling of forest canopy for fuller understanding of the physical components affecting the signals measured by remote sensors relies on the availability of reflectance measurements for forest components as well as atmospheric parameters [See Volume I]. The acquisition of these data would require extensive field work. Although the availability of these data is in question, the importance of it as an analytical tool influences us to list it first among ancillary data requirements.

2. Geological Features:

A priori knowledge of predominant geologic features would provide additional information in the classification of forest and rangeland data using remote sensing techniques.

3. Soil Types and Condition:

Soil information provides an integral component of forest understory inference strategies employing remote sensor data [see Volume I].

4. Climatological Data:

Timely climatological information is required for purposes of change detection analysis employing Landsat data. Varying weather conditions affect Landsat signals from varying ground cover differently. Climatological effects may lead to incorrect observation in change if not accounted for. Climatological information is an integral component of inference and predictive information modeling systems.

5. Hydrological Information:

Information with regard to the hydrological conditions within a region of analysis provides a dimension in the layered classification of Landsat data. This information where available provides a priori knowledge of location of river networks and drainage patterns.

6. Ownership, Jurisdictional, and Political Boundaries:

Information which specifies geographic boundaries are valuable for the aggregation of inventory statistics for reporting purposes.

7. Current or Projected Land Use:

Existing land use information is valuable in conjunction with remote sensor data for change detection analysis as well as layered classification of Landsat data.

8. Aircraft Multispectral Scanner Data:

Certain regions of interest may require analysis at a resolution finer than Landsat can provide. The selective collection of data in these regions by the employment of aircraft multispectral scanners will help meet the need for a finer data resolution.

9. Active Remote Sensor Data:

Active remote sensors, including both laser scanners [38] and multichannel imaging radar [39], are now available and may provide useful information on forest and rangeland resources.

10. Meteorological Information:

Meteorological satellites can provide layers of information that can enhance an analyst's knowledge of climatological conditions.

11. Non-Digital Data:

Maps, aerial photography, and Landsat false-color imagery provide visual aids to an analyst in the interpretation of information extracted from remotely sensed data.

IV.3.2 DATA ELEMENTS ACQUIRED FROM INTERNAL SOURCES

The digital analysis of remote sensor data results in the creation of additional layers of information that may be in turn stored in the data base for future reference. The data elements to be set forth in this section do not incorporate all those data derived through the normal course of analysis. There are certain data elements derived from internal sources however, that enable and enhance the analysis of remote sensor data. The following serve as examples of data derived from the basic Landsat data source.

1. Sun Angle Corrected Landsat Data

For uniform analysis of Landsat data that have been collected in different frames but representing the same temporal acquisition, it is necessary to employ a technique to normalize effects of the solar angle, the specific place and date of acquisition. Simple cosine corrections are sometimes employed for sun angle normalization [40].

2. Tasselled Cap Transformed Data

The Tasselled Cap transformation of Landsat data is described in Section 4. It provides a compression of the data such that the data is arranged in a more physically meaningful way along axes representing brightness, green and yellow components.

3. Haze Diagnostics

Landsat signals are calibrated radiance measures of ground reflectance through an atmosphere. The same reflectance measured through a variety of atmospheric conditions will result in just as many signal values. The effects of atmosphere have been well documented [e.g., 41,42]. Techniques have also been devised to determine haze diagnostics in order to normalize atmospheric effects from place to place [42]. These diagnostics at a minimum as well as the haze corrected equivalent of Landsat data should be available to the analyst.

4. Landsat Cloud, Cloud Shadow and Water, Data Quality

Techniques have been devised and are being refined [42] to screen Landsat data to detect the location of clouds, cloud shadow and water as well as suspect data points.

5. Landsat Feature Signatures

The analysis of Landsat data requires the derivation of statistics describing features of interest. These statistics represent the means and variance/covariance of data features. Often these statistics are employed in the classification of the data and then most often discarded. It had proven that the maintenance of this information along with descriptive attributes is invaluable as an analytical tool used to compare difference features from scene to scene.

### IV.3.3 DATA/GIS INTERFACES

Table IV-1 illustrates the interrelationship between system storage characteristics and some possible data formats for the generalized data storage management facilities. For example, a soil data file could be stored in two physical storage formats, e.g., storage in both compact sequential grid format and in polygonal chain/node format. Such duplication could be beneficial under certain circumstances. If, for example, soils data were stored only in polygonal format, frequent requests for grid format would require repeated invocation of a point-in-polygon process.

TABLE IV-1. POSSIBLE DATA/INFORMATION SYSTEM INTERFACES

| System Storage Characteristics | Data: Landsat | Topo Data | Soil Data | Climatological | Boundaries | Calculated Signatures | Hydrology | Land Use |
|---|---|---|---|---|---|---|---|---|
| Grid Storage | SEQUENTIAL | SEQUENTIAL | COMPACT SEQUENTIAL | COMPLETE CODING | NO | NO | NO | SEQUENTIAL OR C. SEQ. |
| Polygon Storage | NO | NO | POLY CHAIN NODE | POINT CHAIN NODE | POLY CHAIN NODE | NO | LINEAR CHAIN NODE | NO |
| List Storage | NO | NO | NO | NO | LOCATION LISTS | YES | NO | NO |

## IV.4 DATA PROCESSING REQUIREMENTS

The incorporation of remotely sensed and associated data in a USFS geographical information system results in the need to incorporate a set of sophisticated processing functions to facilitate information extraction and analysis of data from remote sensor sources. These functions can be complex and not typical of the functions most commonly employed by geographical information systems [see Table IV-2]. Much effort has been expended to this point in establishing a processing and data base environment that can accommodate remotely sensed data while at the same time providing a generalized environment for the wide variety of other system processing functions. The objective of this section is to display a representative list of processing algorithms currently employed in the analysis of Landsat data and recommended for incorporation into a forest and rangeland system utilizing remotely sensed data. First, a categorization of typical processing functions is provided. These categories must be in some way represented in the GIS processing system in order to insure adequate tools to analyze remotely sensed data. Secondly, a list of specific information extraction techniques are provided to provide sample approaches to each processing category as well as an array of additional applications.

TABLE IV-2*. COMMON SPATIAL DATA HANDLING CAPABILITIES OF A GEOGRAPHICAL INFORMATION SYSTEM

| Entry & Cleaning | Generalization & Manipulation | Output |
|---|---|---|
| - format conversion | - polygon merger | - windowing |
| - spike & gap removal | - overlaying | - clipping |
| - geometric correction rotation, translation) | - edge matching | - printing and plotting |
| - map rectification | - point-in-polygon | |
| - error estimation | - scale change | |
| | - projection conversion | |
| | - parsing & smoothing | |
| | - perimeter & area calculation | |
| | - centroid calculation | |
| | - simple sums & averages | |
| | - classing | |

| Data Maintenance | Analysis |
|---|---|
| - geographical update | - (special-purpose procedures) |

*Source - D. Peuquet [4]

IV.4.1 CATEGORIZATION OF REMOTE SENSING DATA PROCESSING

Figure IV-4 illustrates four categories for the processing of remotely sensed data.

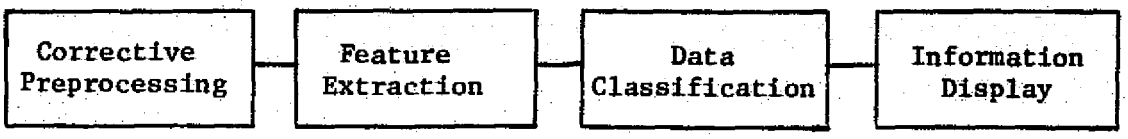



FIGURE IV-4. REMOTE SENSING DATA PROCESSING CATEGORIES

Corrective preprocessing of remotely sensed data is carried out to eliminate or reduce inconsistencies in the data due to the parameters and variables of data collection. A number of external effects influence the nature of the data. Consistency in analysis from one acquired data region to another demands that certain preprocessing functions be carried out. Landsat data, for example, may require sun angle and atmospheric haze correction factors. Aircraft remotely sensed data are often strongly affected by the scan angle, requiring an across-scan correction. Radar digital imagery is characterized by speckles that should be smoothed from the data before analysis.

Feature extraction refers to the process of determining the characteristics of remotely sensed data that enables one to distinguish classes of information. Feature extraction techniques include data channel selection, data compression, training data selection and signature extraction. Signature extraction may require the identification of the statistical, analytical or modeled characteristics of spectral, temporal, spatial or textural features of data. Clustering is a commonly employed statistical feature extraction technique.

Data classification pertains to the identification of features detected using remotely sensed data. It is the decision-making process that associates data elements to the classes present in a scene as

predicted by the training procedure. The basic data element classed may be a pixel or a group of pixels, that is, a field or stand. Individual element labeling may not be the primary goal of data classification. One may be interested in proportion estimates of various classes over an entire scene, or even a prediction of crop yield, e.g., wheat yield or timber volume. Numerous classification strategies are available. The most common employ statistical techniques like the maximum likelihood decision rule. Techniques have been developed that employ spatial, textural, and temporal as well as spectral information (see Section 4).

Information display acts as interface between the data analyst and the algorithms processing the data. Tabular and graphical display are the two basic means the analyst has at his/her disposal to view the information extracted. Common strategies employed with remotely sensed data include false-color or graymap display, data histograms, and simple tabular displays. It is recommended that the statistical analysis of any signature data be aided by the graphical display of contour (e.g., one standard deviation) ellipses representing a signature in two dimensions.

### IV.4.2 INFORMATION EXTRACTION TECHNIQUES

The following list of algorithms applicable to the processing of remote sensor data is not exhaustive, but represents each of the data processing categories described in the previous section.

Preprocessing or Data Preparatory Functions

| Function | Description |
|---|---|
| 1. Sun Angle Correction | Normalize differences in illumination due to sun angle |
| 2. Scan Angle Correction | Eliminate variability in signal attributed to scan angle, especially with aircraft data |

| | | |
|---|---|---|
| 3. | Haze Effect Correction | Minimize variability in signal attributed to atmospheric effects |
| 4. | Tasselled Cap Transformation | Linear transformation to orient data along axes containing meaningful information |
| 5. | Data Quality Analysis | Flag clouds and cloud shadow, deviant data points |
| 6. | Geometric Correction | To deskew and rotate data |
| 7. | Map Rectification | Projection of the data into UTM coordinates |
| 8. | Terrain Effects Correction | Correction of effects of irregular illumination due to terrain and flagging shadowed points in rugged terrain at low sun angles |

Feature Extraction

| | Function | Description |
|---|---|---|
| 1. | Spectral and Temporal Feature Selection | To define the optimum subset or linear combination of a set of spectral (dimensions) and temporal (date of acquisition) features for classification |
| 2. | Spatial Feature Selection | Determination of spatial field or stand structure for training selection and data compression purposes |
| 3. | Statistics for a Fixed Region | Direct determination of mean vector and covariance matrices, i.e., single signature, for sets of data vectors, fields or regions defined by user |
| 4. | Clustering | Automatic determination of the set of signatures representing a region of interest |

Classification Functions

| Function | Description |
|---|---|
| 1. Maximum Likelihood Classification of Pixels | Assignment of individual multivariate samples to user-defined classes |
| 2. Maximum Likelihood Classification of Compressed Features | Assignment of compressed features to user-defined classes |
| 3. Layered Classifier | Classify points using a hierarchical sequence of decisions leading from a general class definition to specific classes |
| 4. Spatial Classification Rules | Classing of a sample is based on its association to spatial neighbors in addition to spectral characteristics |

Descriptive Functions

| Function | Description |
|---|---|
| 1. Tabular Results | Display of aggregated results of classification |
| 2. Image Display | Graymap or false color image displays on a line printer or CRT |
| 3. Statistical Displays | Histograms and scatter diagrams depicted data distributions |
| 4. Signature Display | One, two and three dimension displays of Gaussian distribution representing a data class |

Analytical Functions

The following lists a set of functions whose utility is to expand the analytical tools available to the system user.

| Function | Description |
|---|---|
| 1. Overlay Processing | Many user queries can simply be resolved by the union or intersection of information in multiple layers of data. Such analysis is conjunction with statistical tabulation and mapping routines |
| 2. Signature Manipulation | The determination of adequate signature for classification requires analysis of those computed through statistical and clustering techniques requiring a facility to display (see Descriptive Functions) and manipulate these signatures statistically |
| 3. Distance Function | Often overlooked is theneed for simple measures of distance between sample points and statistical distributions. Requirements would include Euclidean distance between a point and a line, print or plane; $\chi^2$ distance and Bhattacharyya distance between two distributions |
| 4. Polygonal Manipulation | Computation of area, perimeter, intersection, centroid |
| 5. Statistical Manipulation | Statistical analysis function provide the most powerful analytical tool available -- linear and multiple regression analysis, analysis of variance |
| 6. Contour Mapping | To display data stored as grid cells in standard map notation |

Data Base Manager Functions

Interfacing a data base composed of a variety of data types will require, at a minimum, the following format service routines. The DBM philosophy is not to reformat data but to accommodate it in its most straightforward form, providing FSR's to read the data at time of storage and retrieval.

| | FSR | Description |
|---|---|---|
| 1. | Universal Landsat Format | A standard NASA remotely sensed data format structure |
| 2. | Landsat Format | Standard EROS Data Center Landsat format |
| 3. | NCIC Digital Terrain Tape | To read digital terrain data available through the National Cartographic Information Center |
| 4. | Polygon to Grid Conversion | Employment of a point-in-polygonal algorithm to convert linear data to cellular structure |
| 5. | Other | FSR's should be designed to manage any nonstandard data layer format |

## REFERENCES

1. Sadowski, F. G., and W. A. Malila, "Investigation of Techniques for Inventorying Forested Region", Vol. I, Report No. 112700-35-$F_1$, Environmental Research Institute of Michigan, Ann Arbor, Michigan, November 1977.

2. Steiner, D. and A. Salerno, "Remote Sensor Data Systems, Processing and Management", Chapter 12, Vol. I, Manual of Remote Sensing, American Society of Photogrammetry, 1975.

3. Tomlinson, R. F., H. W. Calkins, and D. F. Marble, Computer Handling of Geographical Data, The UNESCO Press. 1975, pp. 27-73.

4. Peuquet, D. J., Raster Handling in Geographic Information Systems, Geographic Information Systems Laboratory, State University of New York at Buffalo, July 1977.

5. Smedes, H., "IRIS" Feasibility Study, Final Report, Center for Advanced Computation, University of Illinois at Urbana-Champaign, April 30, 1972.

6. Zobrist, Albert L., "Elements of an Image-Based Information System", Institute of Electrical and Electronics Engineers, Workshop on Image Data Base Management, 1976, pp. 55-63.

7. Abbitt, C. W., Earth Resources Interactive Processing System Requirements, Aeronutronic Ford Corporation, November 1975.

8. Center for Urban and Regional Affairs and State Planning Agency (CURA, MSPA), "Minnesota Land Use, The Minnesota Land Management Information System", University of Minnesota, Minneapolis, and the Minnesota State Planning Agency, St. Paul, 1972.

9. CODASYL Data Base Task Group, ACM, New York, April 1971 report.

10. Management Sciences Staff, Analysis of Computer Support Systems for Multi-Functional Planning, USFS, USDA, Berkeley, California, June 1976.

11. Personal Communication during April, 1977, with the following: T. George, D. Loff, E. Tolin, USFS Washington, D. C., and N. Allen, R. Driscoll, USFS, Fort Collins, Colorado.

REFERENCES (Cont'd)

12. A Buyer's Guide to Data Base Management Systems, Data Pro Research Corporation, 1975.

13. Fry, J. P., E. H. Sibley, Evolution of Data-Base Management Systems, Computing Surveys, Vol. 8, No. 1, March 1976.

14. Philips, R. L., "A Query Language for a Network Data Base with Geographical Entities", Department of Aerospace Engineering, The University of Michigan, Ann Arbor, Michigan, 1977.

15. Algazi, V. R., M. Suk, "Satellite Land Use Acquisition and Application to Hydrologic Planning Models," Proceedings of Eleventh Internation Symposium Remote Sensing of Environment, Ann Arbor, Michigan, April, 1977.

16. Hoffer, R. M., and Staff, Computer-Aided Analysis of Skylab Multispectral Scanner Data in Mountainous Terrain for Land Use, Forestry, Water Resource, and Geologic Applications, Final Report, Laboratory for Application of Remote Sensing, Purdue University, W. Lafeyette, Indiana, December 1975.

17. Kan, E. P., Ten Ecosystems Study Interim Report. Nationwide Forestry Applications Program, JSC-12924, LEC-10539 by Lockheed Electronics Corporation for NASA Johnson Space Center, Houston, Texas, September, 1977.

18. National Cartographic Information Center, Digital Terrain Tapes: NCIC User guide, U.S. Department of the Interior, Geological Survey, Reston, Virginia.

19. Sharpnack, David A., and Garth Akin, An Algorithm for Computing Slope and Aspect from Elevations, Photogrammetric Review, March 1969, pp. 247-248.

20. MacDonald, R. B., R. B. Erb, and F. G. Hall, 1977, "LACIE: A Look Into the Future", Proc. of Eleventh International Symposium on Remote Sensing of Environment, Environmental Research Institute of Michigan and the University of Michigan Extension Service, Ann Arbor, Michigan.

21. Kauth, R., and G. Thomas, "The Tasselled Cap--A Graphic Description of the Spectral-Temporal Development of Agricultural Crops as Seen by Landsat", Proc. of 1976 Symposium on Machine Processing of Remotely Sensed Data, Purdue University, W. Lafayette, Indiana, 1976.

REFERENCES (Cont'd)

22. Malila, W. A., and J. M. Gleason, Investigations of Spectral Separability of Small Grains, Early Season Wheat Detection and Multicrop Inventory Planning, NASA CR-_____, ERIM 122700-34-F, Environmental Research Institute of Michigan, Ann Arbor, Michigan, 1977.

23. Horwitz, H. M., J. T. Lewis and A. P. Pentland, Estimating Proportions of Objects from Multispectral Scanner Data, NASA CR-_____, ERIM 109600-13-F, Environmental Research Institute of Michigan, Ann Arbor, Michigan, May 1975.

24. Kauth, R., A. Pentland, and G. Thomas, "BLOB, An Unsupervised Clustering Approach to Spatial Preprocessing of MSS Imagery", Proc. of Eleventh International Symposium on Remote Sensing of Environment, Environmental Research Institute of Michigan and the University of Michigan Extension Service, Ann Arbor, Michigan, 1977.

25. Adapted from: International Geographical Union, Geographical Data Handling, R. F. Tomlison (editor), UNESCO/IGU Second Symposium on Geographical Information Systems, Ottawa, Canada, August 1972.

26. Istvan, L. B., Land Use Interpretation from Aerial Photography for Input to the Section 208 Water Quality Management Program of the Toledo Metropolitan Area Council of Governments, Environmental Research Institute of Michigan, Ann Arbor, Michigan, February 1976.

27. U. S. Department of the Interior, Information/Data Handling: A Guidebook for Development of State Programs, Office of Land Use and Water Planning and U. S. Geological Survey Resource and Land Investigations Program, July 1975.

28. Deuker, K. J., "A Framework for Encoding Spatial Data", Geographical Analysis, Vol. 4, 1972, pp. 98-105.

29. Warntz, W., The Redbook, Harvard University Press, Cambridge, Massachusetts, 1971.

30. Cooke, Maxfield, The Dime Encoding System, Bureau of the Census, Washington, D. C., Census Use Study Report 4, 1967.

31. Chrisman, Nicholas, Cartographic Data Structures Panel, Proceedings of the International Conference on Automation in Cartography, "Auto-Carto I", Reston, Virginia, December 9-12, 1974, pp. 165-177.

REFERENCES (Cont'd)

32. Philips, R. L. "Computer Graphics in Urban and Environmental Systems", Proceedings of the IEEE, Vol. 62, No. 4, April 1974, pp. 437-452.

33. Taylor, Robert W., R. L. Frank, "CODASYL Data-Base Management Systems", Computing Surveys, Vol. 8, No. 1, March 1976.

34. Tsichritzis, D. C., F. H. Lochovsky, "Hierarchical Data-Base Management: A Survey", Computing Surveys, Vol. 8., No. 1, March 1976.

35. Durfee, R. C., ORRMIS: Oak Ridge Regional Modeling Information System, Part I, Oak Ridge National Laboratory, Oak Ridge, Tennessee, September 1974.

36. Colwell, R. N., S. J. Titus, Forest Applications Project/Timber Resource: Sam Houston National Forest Inventory and Development of a Survey Planning Model, Final Report for NASA contract 9-14452, University of California, Berkeley, California, 14 July 1976.

37. Hoffer, R. M., Natural Resource Mapping in Mountainous Terrain by Computer Analysis of ERTS-1 Satellite Data, Laboratory for Applications of Remote Sensing and the Agricultural Experiment Station, Purdue University, W. Lafayette, Indiana, September 1974.

38. Hasell, P. G., Jr., L. M. Peterson, F. J. Thomson, E. A. Work, and F. J. Kriegler, Active and Passive Multispectral Scanner for Earth Resources Applications: An Advanced Applications Flight Experiment, NASA CR-_____, ERIM 115800-49-F, Environmental Research Institute of Michigan, Ann Arbor, Michigan, June 1977.

39. Larson, R., P. Jackson, R. Dallaire, R. Schuchman, and R. Rawson, "Interpretation and Measurement of Multichannel SAR Imagery", Proceedings of the Tenth International Symposium on Remote Sensing of Environment, Environmental Research Institute of Michigan, Ann Arbor, Michigan, 1975.

40. Kauth, R. J., and W. Richardson, Procedure B: A Multisegment Training Selection and Proportion Estimation Procedure for Processing Landsat Agricultural Data, ERIM 122700-31-F, Environmental Research Institute of Michigan, Ann Arbor, Michigan, November 1977.

REFERENCES (Cont'd)

41. Malila, W. A., J. M. Gleason and R. C. Cicone, Atmospheric Modeling Related to Thematic Mapper Scan Geometry, ERIM 119300-5-F, Environmental Research Institute of Michigan, Ann Arbor, Michigan, April 1976.

42. Lambeck, Peter F., Signature Extension Preprocessing for Landsat MSS Data, ERIM 122700-32-F, Environmental Research Institute of Michigan, Ann Arbor, Michigan, November 1977.